

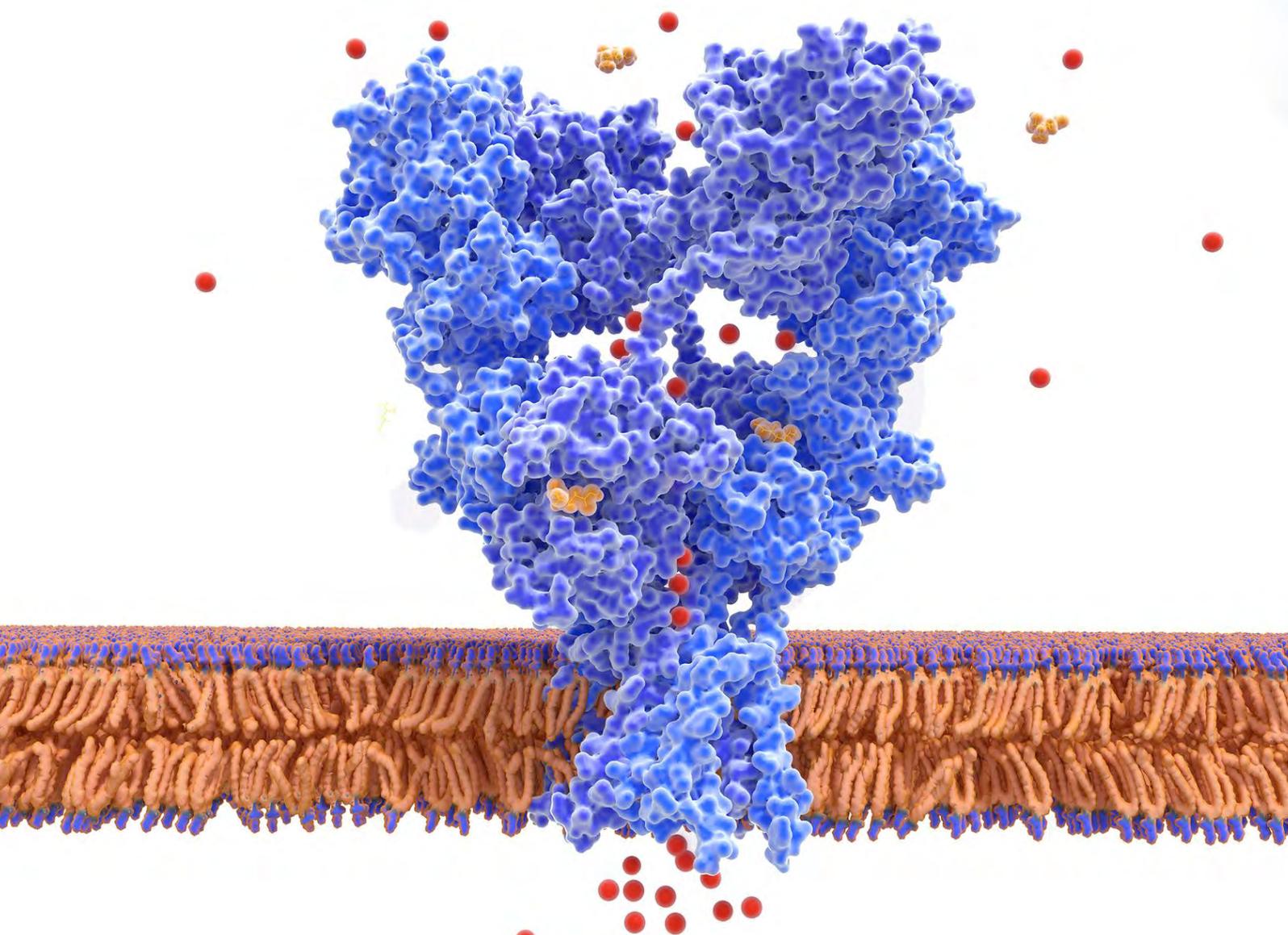
CONICET



Trabajo de tesis doctoral

DESCUBRIMIENTO DE NUEVOS ANTICONVULSIVOS QUE ACTÚAN MEDIANTE INTERACCIONES CON CANALES IÓNICOS

Tesista: Farm. Manuel A. Llanos
Directora: Dra. Luciana Gavernet
Co-director: Dr. Pedro Martín
Año 2022





UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS
DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Tesis Doctoral:

DESCUBRIMIENTO DE NUEVOS ANTICONVULSIVOS QUE ACTÚAN MEDIANTE INTERACCIONES CON CANALES IÓNICOS

Tesista: Farm. Manuel A. Llanos

Director: Dra. Luciana Gavernet

Co-director: Dr. Pedro Martín

Año 2022

A la memoria de mi vieja.

Agradecimientos

En primer lugar, quisiera agradecer a mis directores **Dra. Luciana Gavernet** y **Dr. Pedro Martín**, por la confianza que depositaron en mí, por su paciencia y su apoyo.

A la **UNLP**, la **Facultad de Ciencias Exactas** y el **CONICET**, por la formación académica y por brindarme los medios necesarios para llevar a cabo esta tesis.

A las **Dras. María Laura Sbaraglini** y **Sofía Goicoechea** por su colaboración en la realización de los ensayos *in vivo* en el LIDeB.

A la **Dra. Dolores López García** y el **Dr. David Sánchez Benito** por la realización de los ensayos *in vivo* en la Universidad de Salamanca, España.

A los **Dres. Pedro Martín, Verónica Milesi, Nicolás Enrique y Valentina Pastore** por su colaboración en la realización de los ensayos *in vitro*.

Al **Dr. León Islas** (Universidad Nacional Autónoma de México, México) y la **Dra. Jesica Raingo** (IMBICE, Argentina) quienes nos donaron amablemente los clones de rTRPV1 y eGFP, respectivamente.

Al **Dr. Laureano Sabatier** y el **Lic. Federico Garofalo** por su colaboración en la síntesis y purificación de los compuestos.

A **Laboratorios Bagó** por la donación del Montelukast sódico utilizado en las evaluaciones experimentales, y a **NVIDIA Corporation** por la donación de las GPUs utilizadas en los cálculos computacionales.

A mis **compañeros del LIDeB** por tantas historias, aprendizajes y momentos compartidos.

A mi **familia**, por el sacrificio y por su apoyo constante que me permitieron llegar hasta acá.

A mis **amigos**, por la incondicionalidad.

Abstract

Epilepsy is a disease characterized by the recurrent presence of seizures, with the neurobiological, cognitive, psychological and social consequences that this implies. It affects more than 50 million people worldwide, which makes it the second most common neurological disease globally. The vast majority of those affected live in low-income countries and as a consequence nearly 75 % of them do not receive appropriate treatment. Pharmacotherapy is the first-line treatment for this pathology. However, approximately 30 % of patients do not respond to existing pharmacological therapies. This motivates the constant search for safer and better tolerated anticonvulsant drugs (ACDs) that overcome the drug resistance problem. In this regard, this doctoral thesis aims to find multitarget compounds that act simultaneously on TRPV1 and NaV1.2 channels, with potential anticonvulsant activity *in vivo*, through a computer-aided drug repositioning strategy. For both molecular targets ligand- (QSAR) and structure-based (docking and molecular dynamics) predictive models were developed. The ensemble of models was applied in a virtual screening campaign over the DrugBank database, aiming to repurpose already approved drugs as ACDs, and three candidates were selected for experimental testing: Montelukast, Novobiocin and Cinnarizine. All of them demonstrated a potent inhibitory activity on both targets, measured by the patch clamp technique on a heterologous expression system in HEK293 cells. Additionally, the candidates were tested in four animal models of seizures: MES, 6-hz, PTZ, and GASH:Sal. All drugs exhibited anticonvulsant activity in at least one of these models, and none of them showed signs of neurotoxicity in the Rotorod test. The combination of *in silico* methodologies, based on the structure of the ligands and the receptor, proved to be a useful approach for the identification of multitarget compounds in the context of a disease such as Epilepsy. Moreover, the joint modulation of TRPV1 and NaV1.2 channels emerge

as a promising strategy for the development of novel ACDs.

Como leer esta tesis

Esta tesis se encuentra dividida en siete capítulos, que sintetizan el recorrido científico y personal que derivó en el presente trabajo de investigación doctoral. El documento fue escrito íntegramente en L^AT_EX, un lenguaje de programación orientado a la generación de textos. Por lo tanto, la versión digital en formato PDF posee enlaces de hipertexto, que vinculan el índice general con las diferentes secciones del documento e interconectan todos los elementos referenciados en el texto, como tablas, figuras o referencias bibliográficas. A su vez, el documento dispone de un *outline* que refleja la estructura global de la tesis y le permitirá navegar por la misma con mayor facilidad.

El **Capítulo 1** introduce a la Epilepsia como patología, haciendo foco en el tratamiento farmacológico de la misma y sus limitaciones, lo cual nos motiva a continuar la búsqueda de nuevas alternativas terapéuticas para dar respuesta a esta necesidad médica no resuelta. Luego de describir los distintos blancos moleculares asociados a la Epilepsia, se profundiza en detalle sobre los que fueron abordados en esta tesis: el receptor de potencial transitorio vaniloide tipo 1 (TRPV1) y el canal de sodio operado por voltaje isoforma 1.2 (NaV1.2).

En el **Capítulo 2**, luego de una breve descripción del proceso de descubrimiento de fármacos y las metodologías clásicas que se emplean con dicho propósito, se detallan varias metodologías computacionales, haciendo hincapié en las que fueron utilizadas a lo largo de la tesis.

El **Capítulo 3** describe el desarrollo y la validación de los modelos basados en la estructura de los ligandos y del receptor, con el objetivo de identificar compuestos inhibidores del canal TRPV1, actuando en el sitio de unión de los capsaicinoides. De manera análoga, en el **Capítulo 4** se detalla la construcción de ambos tipos de

modelos para identificar, esta vez, compuestos que interaccionan con la región del poro en los canales NaV1.2.

El **Capítulo 5** comienza con la aplicación de los modelos desarrollados previamente en un campaña de cribado virtual, con el objetivo de reposicionar fármacos conocidos como drogas anticonvulsivas, actuando simultáneamente en ambos canales iónicos. Luego de una introducción general a la técnica de patch clamp y su aplicación al estudio de estos canales iónicos, se presentan los resultados de la evaluación experimental *in vitro* de los *hits* seleccionados. En el **Capítulo 6** se reportan los resultados sobre la evaluación *in vivo* de la actividad anticonvulsiva en diversos modelos animales.

Finalmente, en el **Capítulo 7** se resumen las conclusiones principales que surgen de este trabajo de tesis doctoral y se plantean también las perspectivas a futuro para esta línea de investigación.

El **Material suplementario Algoritmos** describe el fundamento de los principales algoritmos desarrollados y aplicados a lo largo de esta tesis, mientras que en el **Material suplementario Métricas** se detallan las distintas métricas de clasificación, regresión y ranking que fueron utilizadas.

ÍNDICE

Acrónimos	3
Objetivos e hipótesis	5
1 EPILEPSIA	6
1.1 Introducción	7
1.2 Causas	8
1.3 Tratamientos	10
1.3.1 Tratamiento Farmacológico	10
1.3.2 Mecanismos de acción de fármacos anticonvulsivos	12
1.3.2.1 Modulación de canales iónicos operados por voltaje	12
1.3.2.2 Aumento de la neurotransmisión inhibitoria mediada por el ácido γ -aminobutírico	14
1.3.2.3 Disminución de la neurotransmisión excitatoria mediada por glutamato	15
1.3.2.4 Modulación de la actividad de proteínas presinápticas	16
1.4 El receptor de potencial transitorio V1	17
1.4.1 TRPV1 y Epilepsia	19
1.4.2 TRPV1 y el sistema endocannabinoide	21
1.4.3 Antagonistas TRPV1 e hipertermia	22
1.5 Canales de sodio operados por voltaje	23
1.5.1 Canales NaV y Epilepsia	27
1.5.2 Sitios de unión en canales NaV	28
1.6 Epilepsia refractaria	30
1.7 Fármacos multiobjetivo y enfermedades complejas	32
Referencias Capítulo 1	34
2 DESCUBRIMIENTO DE FÁRMACOS	50

2.1	Introducción	51
2.2	Metodologías clásicas	52
2.2.1	Serendipia	52
2.2.2	Análisis de la medicina tradicional	53
2.2.3	Aislamiento de productos naturales	53
2.2.4	Cribado de alta eficiencia	54
2.3	Metodologías <i>in silico</i>	55
2.3.1	Cribado virtual	55
2.3.2	Metodologías basadas en la estructura de los ligandos	57
2.3.2.1	Búsqueda por similaridad	57
2.3.2.2	Farmacóforos basados en los ligandos	58
2.3.2.3	Modelado de relaciones estructura-actividad (QSAR)	59
2.3.2.4	Estrategia desarrollada en esta tesis	67
2.3.3	Metodologías basadas en la estructura del receptor	73
2.3.3.1	Modelado por homología	74
2.3.3.2	Docking molecular	77
2.3.3.3	Dinámica molecular	79
2.3.4	Metodologías combinadas	82
2.4	Reposicionamiento de fármacos	83
2.4.1	Reposicionamiento de fármacos en Epilepsia	84
	Referencias Capítulo 2	85
3	BÚSQUEDA DE INHIBIDORES DEL CANAL TRPV1	96
3.1	Introducción	97
3.2	Metodologías	97
3.2.1	Compilado y partición de la base de datos	98
3.2.2	Desde los ligandos	99
3.2.2.1	Desarrollo y validación de un modelo clasificador	99
3.2.3	Desde el receptor	100
3.2.3.1	Modelado del canal hTRPV1	100
3.2.3.2	Desarrollo y validación de un modelo de docking	102
3.2.3.3	Simulaciones de dinámica molecular	105
3.3	Resultados	107
3.3.1	Compilado y partición de la base de datos	107

3.3.2	Desde los ligandos	108
3.3.2.1	Desarrollo y validación de un modelo clasificador	108
3.3.3	Desde el receptor	115
3.3.3.1	Modelado del canal hTRPV1	115
3.3.3.2	Desarrollo y validación de un modelo de docking	117
3.3.3.3	Simulaciones de dinámica molecular	122
Referencias Capítulo 3		125
4	BÚSQUEDA DE INHIBIDORES DEL CANAL NAV1.2	128
4.1	Introducción	129
4.2	Metodologías	129
4.2.1	Compilado y partición de la base de datos	129
4.2.2	Desde los ligandos	132
4.2.2.1	Desarrollo y validación de un modelo de regresión	132
4.2.3	Desde el receptor	133
4.2.3.1	Obtención de un <i>ensemble</i> estructural del canal NaV1.2	133
4.2.3.2	Desarrollo y validación de un modelo de docking	134
4.3	Resultados	134
4.3.1	Compilado y partición de la base de datos	134
4.3.2	Desde los ligandos	135
4.3.2.1	Desarrollo y validación de un modelo de regresión	136
4.3.3	Desde el receptor	144
4.3.3.1	Obtención de un <i>ensemble</i> estructural del canal NaV1.2	144
4.3.3.2	Desarrollo y validación de un modelo de docking	146
Referencias Capítulo 4		150
5	CRIBADO VIRTUAL Y ENSAYOS <i>IN VITRO</i>	152
5.1	Campaña de cribado virtual	153
5.1.1	Canal TRPV1	153
5.1.2	Canal NaV1.2	154
5.1.3	Selección de los candidatos	155
5.2	La técnica de patch clamp	159
5.2.1	Generalidades	159
5.2.2	Condiciones experimentales de registro	161
5.3	Registros en canales TRPV1	162

5.3.1	Métodos	162
5.3.2	Resultados	163
5.4	Registros en canales NaV1.2	165
5.4.1	Métodos	165
5.4.2	Resultados	167
Referencias Capítulo 5		171
6	ACTIVIDAD ANTICONVULSIVA. ENSAYOS <i>IN VIVO</i>	173
6.1	Introducción	174
6.2	Condiciones experimentales de los ensayos <i>in vivo</i>	175
6.2.1	Ensayo MES	176
6.2.2	Ensayo de 6 Hz	177
6.2.3	Ensayo de PTZsc	178
6.2.4	Ensayo audiogénico GASH:Sal	178
6.2.5	Ensayo Rotorod	180
6.3	Resultados	181
Referencias Capítulo 6		184
7	CONCLUSIONES	188
7.1	Conclusiones generales	189
7.2	Búsqueda de nuevos inhibidores del canal TRPV1	191
7.3	Búsqueda de nuevos inhibidores del canal NaV1.2	193
7.4	Actividad anticonvulsiva <i>in vivo</i>	195
Referencias Capítulo 7		197
Publicaciones, congresos y becas		202
A	Material suplementario Capítulo 3	204
B	Material suplementario Capítulo 4	213
C	Material suplementario Algoritmos	222
C.1	Silhouette Optimized Molecular Clustering (SOMoC)	223
C.2	Métodos de optimización	227
C.2.1	Optimización bayesiana	227
C.2.2	Algoritmos genéticos	230
C.3	Algoritmos de aprendizaje automático	231

C.3.1	K-nearest neighbors	231
C.3.2	ExtraTrees	233
C.3.3	Support Vector Machine	235
Referencias Material suplementario Algoritmos		238
D Material suplementario Métricas		243
D.1	Métricas de regresión	244
D.2	Métricas de ranking	246
D.3	Métricas de clasificación	247
D.4	Métricas específicas del cribado virtual	250
Referencias Material suplementario Métricas		252

Acrónimos

ACDs: Drogas anticonvulsivas (del inglés, Anticonvulsant Drugs)

AI: Inteligencia Artificial (del inglés, Artificial Intelligence)

AMPA: Ácido α -amino-3-hidroxi-5-metilo-4-isoxazolpropiónico

AUC-ROC: Área bajo la curva ROC (del inglés, Area Under the Receiver Operating Characteristic Curve)

CA: cell-attached

CAP: Capsaicina

CBD: Cannabidiol

CBDV: Canabidivarina

CNZ: Cinarizina

CPZ: Capsazepina

Cryo-EM: Crio-microscopía electrónica (del inglés, Cryogenic Electron Microscopy)

CVIs: Índices de validez de *cluster* (del inglés, Cluster Validity Indexes)

DA: Dominio de Aplicación

DMSO: Dimetil sulfóxido

eGFP: Proteína verde fluorescente (del inglés, enhanced Green Fluorescent Protein)

FF: Campo de fuerza (del inglés, Force Field)

GA: Algoritmo Genético (del inglés, Genetic Algorithm)

HTS: Cribado de alto rendimiento (del inglés, High-Throughput Screening)

HCS: Cribado de alto contenido (del inglés, High-Content Screening)

ILAE: Liga internacional contra la Epilepsia (del inglés, International League Against Epilepsy)

IO: inside-out

i.p.: Intraperitoneal

KNN: K-vecinos más cercanos (del inglés, K-nearest neighbors)

MD: Dinámica molecular (del inglés, Molecular Dynamics)

MES: Convulsión máxima inducida por electrochoque (del inglés, Maximal Electroshock Seizure)

ML: Aprendizaje automático (del inglés, Machine Learning)

MTK: Montelukast

NaV: Canales de sodio operados por voltaje

NMDA: N-metil-D-aspartato

NVB: Novobiocina

OO: outside-out

PTZsc: Ensayo de Pentilentetrazol subcutáneo

QSAR: Relaciones cuantitativas de estructura-actividad (del inglés, Quantitative Structure-Activity Relationships)

RMSD: Raíz del error cuadrático medio (del inglés, Root-Mean-Square Deviation)

ROC: Característica Operativa del Receptor (del inglés, Receiver Operating Characteristic)

s.c.: Subcutáneo

SEM: Error estándar de la media (del inglés, Standard Error of the Mean)

SF: Solución Fisiológica

SFB: Suero fetal bovino

SNC: Sistema Nervioso Central

SNP: Sistema Nervioso Periférico

SOMoC: Silhouette Optimized Molecular Clustering

SVM: Máquina de vector soporte (del inglés, Support Vector Machine)

SLR: Regresión lineal simple (del inglés, Simple Linear Regression)

THC: Tetrahidrocannabinol

TLA: Epilepsia del lóbulo temporal (del inglés, Temporal Lobe Epilepsy)

TRPV-1: Receptor de Potencial Transitorio Vaniloide tipo 1 (del inglés, Transient Receptor Potential Cation channel subfamily V member 1)

VS: Cribado Virtual, Tamizado Virtual (del inglés, Virtual Screening)

VSD: Dominio del sensor de voltaje (del inglés, Voltage Sensor Domain)

WC: whole-cell

Objetivos e hipótesis

Objetivo general

El objetivo general de la presente tesis de investigación es el descubrimiento racional de nuevos fármacos anticonvulsivos, cuyo mecanismo de acción principal sea la interacción con canales iónicos, mediante la aplicación de diversas técnicas computacionales.

Como blanco molecular clásico de fármacos anticonvulsivos se abordará el canal de sodio voltaje operado isoforma 1.2 (NaV1.2), y como blanco novedoso al Receptor de Potencial Transitorio Vaniloide tipo 1 (TRPV1).

Se construirán modelos computacionales predictivos, capaces de identificar moduladores de dichos canales iónicos, los cuales se aplicarán en posteriores campañas de cribado virtual con el fin de detectar compuestos con actividad anticonvulsiva. Se espera que estos candidatos resulten el punto de partida para el descubrimiento de fármacos más eficaces y mejor tolerados que los actualmente disponibles, ofreciendo así una alternativa terapéutica a la problemática de la epilepsia refractaria.

Objetivos específicos

- Compilar una base de datos de compuestos evaluados experimentalmente frente a cada uno de los blancos moleculares propuestos.
- Utilizar este conjunto de datos para derivar modelos matemáticos basados en la estructura de los ligandos, capaces de predecir de manera cualitativa (modelos clasificadores) y cuantitativa (modelos de regresión) la actividad biológica de

nuevas moléculas sobre estos blancos moleculares.

- Obtener mediante metodologías de modelado molecular la estructura tridimensional de los blancos en cuestión, para los casos en que la misma no haya sido determinada experimentalmente.
- Desarrollar modelos predictivos basados en la estructura del receptor, capaces de identificar nuevos moduladores de dichos canales iónicos, mediante simulaciones de docking molecular.
- Aplicar los modelos desarrollados de manera conjunta en campañas de cribado virtual, a fin de identificar potenciales compuestos moduladores.
- Simular la interacción de los candidatos más promisorios con el blanco molecular en cuestión - mediante técnicas de dinámica molecular - con el objetivo de estudiar las interacciones droga-receptor responsables de la actividad biológica predicha.
- Evaluar los compuestos seleccionados mediante la técnica de patch clamp en modelos celulares de expresión heteróloga, con el fin de validar experimentalmente las predicciones computacionales.
- Investigar la actividad anticonvulsiva de los candidatos en diversos modelos animales de crisis convulsivas.

Hipótesis de trabajo

La hipótesis general que motiva el presente trabajo de investigación doctoral postula que es posible encontrar compuestos *multitarget* que actúen simultáneamente sobre los canales TRPV1 y NaV1.2, con actividad anticonvulsiva *in vivo*, a través de una estrategia de reposicionamiento de fármacos asistida por computadora. En particular, para el canal TRPV1 se postula que, mediante esta estrategia, es posible identificar nuevos antagonistas sin efectos adversos relacionados con la modulación de la temperatura corporal.



Capítulo 1

Epilepsia

1.1. Introducción

En la actualidad, la Epilepsia se define como una enfermedad caracterizada por la presencia recurrente de crisis convulsivas, con las consecuencias neurobiológicas, cognitivas, psicológicas y sociales que esto implica. Por otra parte, una convulsión es la manifestación transiente de una actividad neuronal anómala o excesiva en el cerebro (**Fisher** et al. 2005, 2014). La manifestación única y aislada de una convulsión no es motivo suficiente para hablar de Epilepsia. De esta manera, se considera que una persona tiene Epilepsia si se cumple al menos uno de los siguientes criterios:

- Sufre al menos dos convulsiones no provocadas o reflejas, en un intervalo de tiempo mayor a 24 horas.
- Sufre una convulsión no provocada o refleja y presenta una probabilidad de ocurrencia de futuras convulsiones en los próximos 10 años, similar al riesgo general de ocurrencia (al menos del 60%) luego de dos convulsiones no provocadas.
- Fue diagnosticada con algún síndrome epiléptico.

El término “refleja” hace referencia a las convulsiones desencadenadas como respuesta a un estímulo externo, que ocurren de manera no intencional. Un ejemplo es la Epilepsia fotosensible, donde las convulsiones son provocadas por estímulos luminosos, pero no de manera intencional.

La Epilepsia se considera resuelta (aunque no curada) en aquellos individuos diagnosticados con un síndrome epiléptico cuya aparición es dependiente de la edad, y que han superado esa edad sin presentar crisis; o en aquellos individuos que no han presentado crisis en los últimos 10 años y no han necesitado medicinas antiepilépticas en los últimos 5 años (**Fisher** et al. 2014). Esta patología afecta aproximadamente a 50 millones de personas en el mundo, lo que la convierte en la segunda enfermedad neurológica más común a nivel global, luego de las enfermedades cerebrovasculares (Figura 1.1). La gran mayoría de los afectados (80%) se concentra en países de bajos ingresos y como consecuencia un 75% de esos pacientes no recibe un tratamiento

apropiado. Con un diagnóstico y tratamiento adecuado, se estima que el 70 % de las personas diagnosticadas con Epilepsia podría vivir libre de convulsiones (Bell, Neligan y Sander 2014; Forsgren et al. 2005).



Figura 1.1: La Epilepsia en números. Fuente: WHO 2018.

1.2. Causas

En su actualización más reciente sobre la etiología de la Epilepsia, la International League Against Epilepsy (ILAE) propone seis categorías no jerarquizadas ni autoexcluyentes que clasifican las causas conocidas de la enfermedad. Las mismas se encuentran representadas en la Figura 1.2.

Estructurales: Existen alteraciones estructurales a nivel del sistema nervioso central (SNC) capaces de generar Epilepsia. Dichos cambios se evidencian mediante técnicas de neuroimagen y esta información, a su vez, debe complementarse con registros de electroencefalograma (EEG) y el análisis de las manifestaciones clínicas de la enfermedad (Lapalme-Remis y Cascino 2016).

Genéticas: Variantes en genes específicos (o variaciones en el número de copias) pueden causar diversos tipos de Epilepsia. Sin embargo, determinar cuando una variante genética es la causante de la enfermedad, es demasiado complejo en



Figura 1.2: Categorías que clasifican las causas conocidas de la Epilepsia, según la ILAE.

muchos casos. Por lo tanto, tener un historial familiar asociado a la enfermedad, junto con registros de EEG y manifestaciones clínicas, es suficiente para categorizar una Epilepsia como genética. Las epilepsias asociadas a variantes génicas son muy diversas, y se cree que entre un 70 % y un 80 % de las epilepsias diagnosticadas se deben a uno o más factores genéticos (Steinlein 2008).

Un ejemplo de particular importancia dentro de esta categoría son las canalopatías, donde una o más mutaciones en un canal iónico afectan su funcionalidad, generando así un desbalance entre las sinapsis inhibitorias y excitatorias (Mulley et al. 2003; Spillane, Kullmann y Hanna 2016).

Infecciosas: Procesos infecciosos a nivel del SNC pueden derivar en alteraciones estructurales del encéfalo, que finalmente causan Epilepsia. Las infecciones

pueden ser de origen parasitario (Toxoplasmosis), viral (VIH) o bacteriano (Meningitis). Esta es una de las causas más comunes de epilepsias adquiridas, de particular incidencia en poblaciones marginadas o de bajos recursos (**Vezzani et al. 2016**).

Metabólicas: Algunas alteraciones metabólicas pueden dar origen a la Epilepsia, como por ejemplo las deficiencias de Piridoxina (Vitamina B6) o Folato. Aunque en muchos casos la causa de la alteración metabólica es genética, una correcta clasificación tiene implicancias en la elección del tratamiento (**Papetti et al. 2013**).

Inmunes: Desórdenes autoinmunes pueden ser la causa de la Epilepsia. En estos casos la misma puede diagnosticarse mediante la presencia de anticuerpos o marcadores inflamatorios específicos en el líquido cefalorraquídeo (**Britton 2016**).

Desconocida: Pese a la evolución diagnóstica, en algunos casos no se conoce la causa o no está clara. Para esas situaciones se reserva esta clasificación.

1.3. Tratamientos

Existen diversas opciones terapéuticas para el tratamiento de la Epilepsia. Según el tipo de intervención, éstas pueden clasificarse como terapias farmacológicas o no farmacológicas. Dentro de estas últimas, podemos mencionar a los distintos tratamientos quirúrgicos (**Rasul et al. 2017**), dispositivos implantables (**Wong, Mani y Danish 2019**), estimulación magnética transcraneal (**Tsuboyama, Kaye y Rotenberg 2020**) y a las dietas cetogénicas (**D'Andrea Meira et al. 2019**). Naturalmente estos tratamientos no son la primera opción, y se reservan para los casos donde las alternativas farmacológicas disponibles no logran controlar las crisis de manera efectiva (**Saxena y Nadkarni 2011**).

1.3.1. Tratamiento Farmacológico

La terapia farmacológica es el tratamiento de primera elección y el más utilizado hoy en día en pacientes con Epilepsia (**Perucca y Tomson 2011**). Desde el descubrimiento de las propiedades anticonvulsivas del bromuro de potasio en 1853 hasta la actualidad, una gran variedad de drogas anticonvulsivas (ACDs, del inglés

anticonvulsant drugs) han sido desarrollados con el objetivo de encontrar alternativas terapéuticas superadoras. Aunque algunos autores difieren en el agrupamiento de las distintas drogas dentro de cada generación, el consenso actual distingue tres generaciones de ACDs ([Cotterman-Hart 2015b](#); [Löscher 2017](#)).

Los fármacos de **primera generación** son aquellos introducidos al mercado entre 1857 y 1958, desarrollados mediante modificaciones estructurales de los barbituratos. Dentro de esta generación los exponentes principales son el Fenobarbital, la Fenitoína, la Primidona, la Trimetadiona y la Etosuximida ([Krall et al. 1978a](#); [Löscher y Schmidt 2011](#)).

Los fármacos de **segunda generación** fueron introducidos al mercado entre los años 1960 y 1975, y comprenden diversas estructuras químicas que difieren de los barbituratos. En general, estos fármacos son mejor tolerados y poseen un perfil de efectos adversos más favorable. Algunos ejemplos dentro de esta generación son la Carbamazepina, el Valproato y las benzodiazepinas ([Löscher et al. 2013](#)).

La **tercera generación** comprende a las ACDs introducidos al mercado a partir de los años noventa, y cuyos exponentes fueron desarrollados de manera racional basándose en el creciente conocimiento científico sobre la fisiopatología de la enfermedad. Dentro de esta generación podemos encontrar al Brivaracetam, la Eslicarbazepina, la Pregabalina, el Estiripentol y el Perampanel, entre otros ([Łuszczki 2009](#)).

En la Figura 1.3 se detallan los distintos exponentes de cada generación, en orden cronológico de aparición. Pueden observarse dos hitos principales en el diagrama, que de alguna manera han contribuido a delimitar las diferentes generaciones. El primero de ellos es la introducción de modelos animales de crisis convulsivas como método de *screening* de nuevas ACDs en 1936. El segundo hito se da en 1975 con la implementación - por parte del “U.S. National Institutes of Health (NIH)” - del programa de *screening* denominado “Anticonvulsant Screening Project (ASP)” ([Krall et al. 1978a,b](#)). Por último, en el año 2016 este programa fue revisado y actualizado, transformándose en el “Epilepsy Therapy Screening Program (ETSP)” ([Kehne et al. 2017](#)).

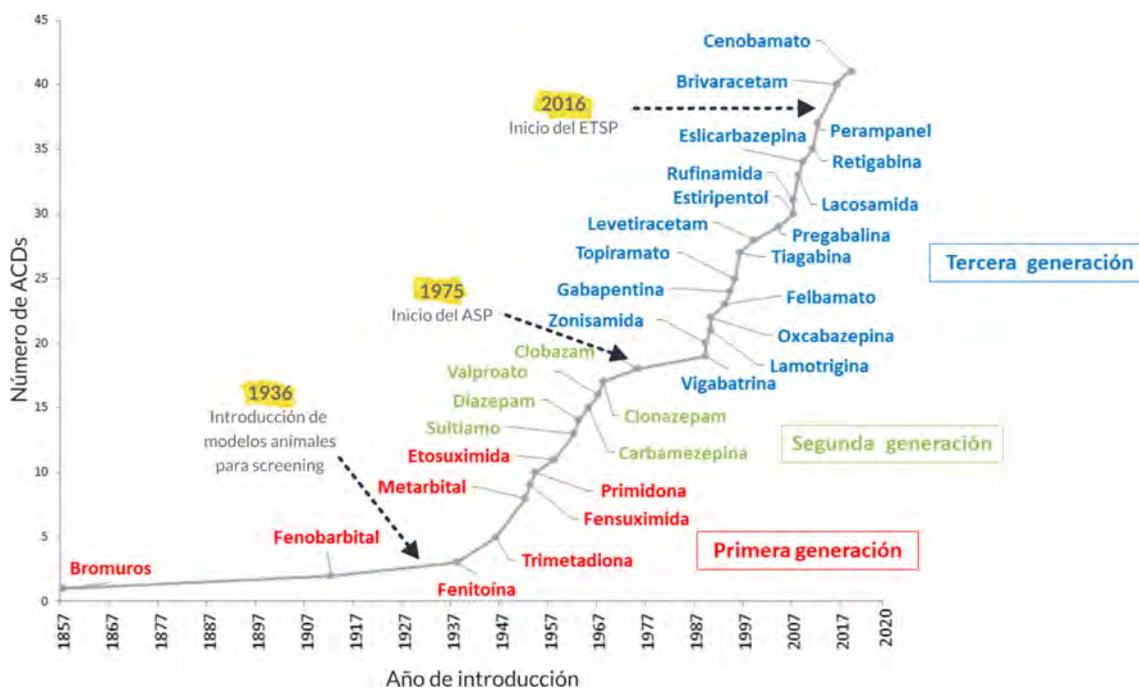


Figura 1.3: Introducción de las distintas ACDs en el mercado. Adaptado de Löscher et al. 2020.

Cabe destacar que la mayoría de los fármacos disponibles hoy en día no ejercen una acción sobre la patología en sí, sino que solo logran tratar los síntomas de la enfermedad (convulsiones). Por lo tanto, la denominación de “drogas antiepilépticas” ha caído en desuso y hoy en día se prefiere el término “drogas anticonvulsivas”.

1.3.2. Mecanismos de acción de fármacos anticonvulsivos

Las ACDs tienen como objetivo modificar la excitabilidad neuronal, de modo de interferir con la generación y propagación de la actividad cerebral epileptiforme. Dentro de este grupo de fármacos, pueden reconocerse 4 mecanismos de acción principales, detallados a continuación.

1.3.2.1. Modulación de canales iónicos operados por voltaje

Los canales iónicos voltaje-operados juegan un papel fundamental en la regulación de la excitabilidad neuronal, permitiendo la generación y transmisión de los impulsos nerviosos. Es por ello que el mecanismo de acción de la mayoría de las

ACDs de uso clínico comprende la modulación de estos canales ([Armijo et al. 2005](#); [Lerche, Jurkat-Rott y Lehmann-Horn 2001](#)). En particular, los mecanismos más comunes implican el bloqueo de los canales de sodio, el bloqueo de canales de calcio, y la activación de los canales de potasio.

Bloqueo de canales de sodio

Los canales de sodio operados pro voltaje (NaV), son los responsables de la despolarización de la membrana en células excitables, por lo que juegan un papel fundamental en el disparo y propagación de los potenciales de acción a través de las neuronas. Estos son considerados un blanco molecular clásico para el desarrollo de nuevas ACDs, ya que existe una gran cantidad de fármacos de uso clínico cuyo mecanismo de acción principal o secundario se debe a su capacidad de modular estos canales, como por ejemplo: Fenitoína, Carbamazepina, Fenobarbital, Valproato, Lamotrigina, Gabapentina, Vigabatrina, Retigabina, Perampanel, Felbamato, Topiramato, Oxcarbazepina, Zonisamida, Rufinamida, Lacosamida, Eslicarbazepina Acetato y Benzodiazepinas ([Brodie 2017](#); [Löscher 2021](#); [Valentina, Marco y Mirko 2012](#)).

Los canales NaV, y en particular la isoforma NaV1.2, serán descritos en profundidad en la Sección 1.5, ya que constituyen uno de los blancos moleculares abordados durante esta tesis.

Bloqueo de canales de calcio

Los canales de calcio operados por voltaje se encuentran implicados en el disparo de los potenciales de acción, y en la liberación de los neurotransmisores a nivel de las terminales presinápticas, por lo que su bloqueo permite una disminución de la excitabilidad neuronal y la liberación de neurotransmisores.

En base a sus propiedades biofísicas y farmacológicas, estos pueden clasificarse en tres grandes familias: CaV1 (de tipo L), CaV2 (de tipo P/Q, N o R) y CaV3 (de tipo T). Los canales de tipo L se relacionan con la generación de crisis febriles y la activación de vías dependientes de calcio implicadas en la regulación génica ([Zhang et al. 2019](#)). Los tipo T son aquellos que se activan a valores bajos de potencial de membrana, y se expresan en gran medida a nivel del tálamo, donde generan

corrientes transitorias asociadas a las crisis de ausencia generalizadas. Los canales P/Q se encuentran directamente implicados en la liberación de neurotransmisores, y una disminución en su función puede generar distonía y ataxia. Los de tipo N, por su parte, se hallan ampliamente distribuidos en los núcleos de diferentes neuronas y en células de la glía, donde se puede observar una expresión aumentada luego del *status epilepticus* (Xu et al. 2010).

Algunos ejemplos de ACDs que actúan sobre estos canales son la Pregabalina, Gabapentina, Etosuximida, Zonisamida, Topiramato, Valproato y Lamotrigina (Rajakulendran y Hanna 2016).

Potenciación de canales de potasio

Contrarrestando la despolarización producida principalmente por el ingreso de iones sodio a la célula, los canales de potasio operados por voltaje (Kv) juegan un rol fundamental en la re-polarización neuronal, modulando así el potencial de membrana en reposo y la excitabilidad. Los canales Kv son un grupo muy heterogéneo, con más de 40 subtipos conocidos hoy en día (Gutman et al. 2005; Miller 2000).

Mutaciones del gen KCNA1 que codifica canales Kv1.1 defectuosos pueden dar origen a crisis parciales complejas (Wei et al. 2017), mientras que variantes mutacionales en el gen KCNQ2 pueden originar una serie de trastornos epilépticos, desde la Epilepsia neonatal familiar benigna hasta la encefalopatía epiléptica (Miceli et al. 1993). Dentro de las ACDs de uso clínico, la Retigabina actúa como un potenciador de las corrientes de tipo M mediadas por canales de potasio de la familia KCNQ (Ihara et al. 2016; Tatulian et al. 2001).

1.3.2.2. Aumento de la neurotransmisión inhibitoria mediada por el ácido γ -aminobutírico

El ácido γ -aminobutírico (GABA, del inglés Gamma Aminobutyric Acid) es el principal neurotransmisor inhibitorio a nivel de la corteza cerebral y las neuronas que lo producen, denominadas GABAérgicas, se encuentran ampliamente distribuidas en todo el SNC. Las vías GABAérgicas están involucradas en diversas funciones fisiológicas, tales como el aprendizaje, la memoria y el control motor (Huang y Paul

2019).

Las ACDs que actúan a nivel de las vías GABAérgicas pueden ser clasificadas según tengan una acción agonista directa sobre el receptor GABA-A (barbitúricos, benzodiazepinas, Topiramato y Felbamato) o modulen el metabolismo del GABA, bloqueando su recaptación (Tiagabina) o inhibiendo su degradación (Vigabatrina) (**Cotterman-Hart 2015a**).

1.3.2.3. Disminución de la neurotransmisión excitatoria mediada por glutamato

El glutamato es un neurotransmisor excitatorio que ejerce su acción sobre receptores localizados en la membrana neuronal, los cuales pueden ser metabotrópicos o ionotrópicos. Dentro de estos últimos, también denominados canales catiónicos asociados a ligandos, podemos encontrar a los receptores postsinápticos del ácido α -amino-3-hidroxi-5-metilo-4-isoxazolpropiónico (AMPA) y los del ácido N-metil-D-aspartico (NMDA), además de los receptores de kainato. Todos los receptores ionotrópicos de glutamato son permeables tanto al sodio como al potasio, aunque difieren en cuanto a su permeabilidad frente al calcio. Los receptores NMDA son altamente permeables al calcio, mientras que los de kainato y AMPA son prácticamente impermeables (**Gaidin et al. 2019**).

Debido a que el glutamato es el neurotransmisor predominante en las sinapsis excitatorias, luego de una crisis epiléptica su concentración extracelular aumenta en gran medida, lo que incrementa el daño neuronal asociado a la crisis, a la vez que altera la expresión neuronal de sus receptores y recaptadores (**Barker-Haliski y White 2015**). Por lo tanto, el bloqueo farmacológico de estos receptores constituye una estrategia prometedora para el desarrollo de nuevas ACDs. Hasta el momento la única ACDs aprobada para uso clínico que modula selectivamente a los receptores AMPA es el Perampanel, un antagonista no competitivo del receptor AMPA (**Tsai et al. 2018**).

Existen otras ACDs cuyo mecanismo de acción principal no es la modulación de estos receptores, pero su actividad anticonvulsiva se debe en parte a este mecanismo. El Topiramato, por ejemplo, se comporta como antagonista de los receptores AMPA

y kainato, aunque este no sea su único mecanismo de acción (**Shank** et al. 2000). Por su parte, el Felbamato se comporta como un antagonista de los receptores NMDA, al bloquear el sitio de reconocimiento de la glicina en dicho blanco molecular (**Rho, Donevan y Rogawski** 1994).

1.3.2.4. Modulación de la actividad de proteínas presinápticas

La proteína de la vesícula sináptica 2 (SV2A) es una glicoproteína integral de membrana que se encuentra ampliamente distribuida en las vesículas de las terminales nerviosas. Aún cuando esta proteína no es indispensable para la liberación de neurotransmisores, es necesaria para una neurotransmisión normal (**Mendoza-Torreblanca** et al. 2013).

Las bases moleculares de la actividad de esta proteína no se encuentran totalmente dilucidadas aún. Sin embargo, la evidencia actual sugiere un rol regulador en las sinapsis GABAérgicas, vinculado a la expresión de la proteína sinaptotagmina 1 (**Tokudome** et al. 2016). A su vez, se encontró una marcada disminución de esta proteína durante la epileptogénesis y en modelos de Epilepsia crónica (**Hanaya** et al. 2012).

La proteína SV2A es el blanco molecular para los fármacos pertenecientes a la familia de los racetam: Levetiracetam, Brevetiracetam y Seletacetam. En modelos animales de crisis convulsivas, como el modelo de 6 Hz, Levetiracetam y sus análogos suprimen las convulsiones, pero a diferencia de otras ACDs no presentan actividad protectora en la prueba de convulsión máxima inducida por electrochoque (MES test, del inglés Maximal Electroshock Seizure Test) o PTZ (Pentilentetrazol) (**Barton** et al. 2001). Levetiracetam se utiliza principalmente como tratamiento complementario para las crisis de inicio parcial en adultos, con o sin generalización secundaria (**Lyseng-Williamson** 2011).

1.4. El receptor de potencial transitorio V1

Nota: En esta tesis nos referiremos al TRPV1 como un canal o un receptor indistintamente, aún cuando estos términos no sean necesariamente sinónimos en otros contextos.

El receptor de potencial transitorio vaniloide tipo 1 (TRPV-1, del inglés Transient Potential Receptor Vanilloid Type 1) es un canal iónico perteneciente a la familia de los canales *Transient Receptor Potential* (TRP), la cual comprende 28 canales clasificados en 6 subfamilias: *canonical* (TRPC), *melastatin* (TRPM), *vanilloid* (TRPV), *ankyrin* (TRPA), *mucolipin* (TRPML) y *polycystin* (TRPP).

El poro del canal está delimitado por cuatro copias de una subunidad α , formando un homotetrámero (**Kedei et al. 2001**). Cada subunidad está formada por 6 segmentos transmembrana (S1-S6) y un *loop* hidrofóbico entre los segmentos S5 y S6 (P-loop). Al igual que otros miembros de la familia TRP, este posee un dominio intracelular formado por 3 repeticiones de ankyrina (*ankyrin repeats*), el cual sirve como sitio de interacción con otras proteínas celulares, como la calmodulina (**Rosenbaum et al. 2004; Sedgwick y Smerdon 1999**)(Figura 1.4).

Una particularidad de este canal es su capacidad de responder a distintos estímulos como el voltaje de la membrana, la temperatura, la concentración de protones, y distintos ligandos tanto endógenos como exógenos (**Caterina et al. 1997; Gunthorpe et al. 2000; Piper et al. 1999; Tominaga et al. 1998**). Es por ello que juega un rol fundamental en la integración de diversos estímulos tanto intra como extracelulares (**Clapham 2003; Szallasi y Blumberg 1999**), y es ampliamente estudiado como blanco de fármacos para el tratamiento del dolor (**Aghazadeh Tabrizi et al. 2017; Almeida, Bernardes y Trevisan 2021; Fakhri et al. 2021; Jara-Oseguera, Simon y Rosenbaum 2008; Premkumar y Sikand 2008**). Otras investigaciones proponen el uso de antagonistas del TRPV1 para tratar la ansiedad (**Terzian et al. 2014**), depresión (**Manna y Umathe 2012a**), la adicción a sustancias como la cocaína (**Adamczyk et al. 2012**) o las metanfetaminas (**Tian et al. 2018**), o incluso trastornos psiquiátricos severos como la esquizofrenia o los desórdenes bipolares (**Escelsior et al. 2020**).

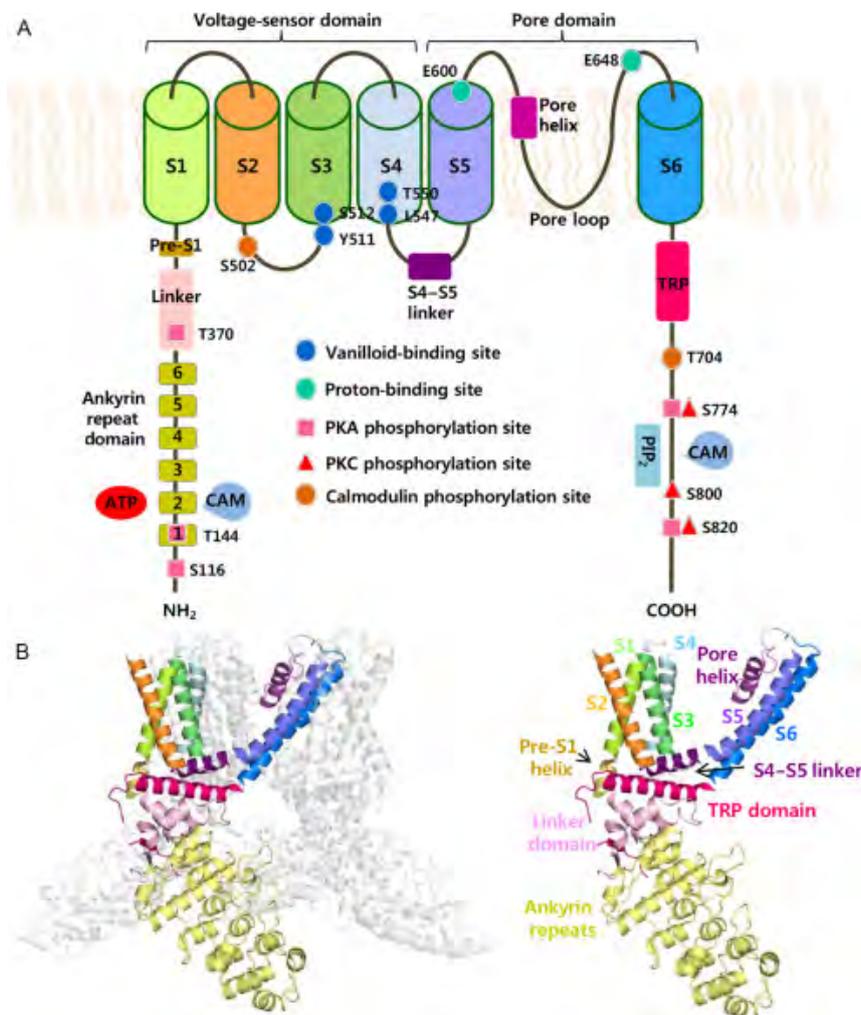


Figura 1.4: Arquitectura general del canal TRPV1. A) Representación bidimensional de los diferentes dominios de canal, incluyendo el dominio repetitivo de anquirina, *linker*, hélice pre-S1, dominios transmembrana (S1–S6), *linker* S4–S5, hélice del poro, *loop* del poro (Loop-P), dominio TRP y dominio C-terminal. Los residuos críticos para la unión de ligandos, calmodulina, PIP-2, ATP y sitios de fosforilación se muestran con diferentes símbolos. B) Representación de la forma apo del canal TRPV1, donde pueden observarse 4 copias idénticas de la subunidad principal formando el homotetrámero (PDB-ID: 3J5P) y una de las subunidades se encuentra coloreada detallando las distintas regiones estructurales que la componen. Imagen tomada de **Cui et al. 2016** y reproducida con autorización de Elsevier.

Por ser un canal catiónico no selectivo, el TRPV1 permite la entrada de calcio, magnesio y sodio a la célula (en ese orden de permeabilidad) (**Caterina et al. 1997**), lo que produce una despolarización de la neurona que finalmente culmina con la liberación de neurotransmisores al espacio sináptico. La entrada de calcio como consecuencia de la activación del canal desencadena un complejo proceso de retroalimentación negativa, que conduce a la desensibilización del mismo (**Carnevale y Rohacs 2016**). Esto implica que el canal se vuelve menos sensible a posteriores estímulos, autolimitando la activación neuronal y restringiendo la entrada de calcio

a la célula, lo que pudiera resultar tóxico (Vyklícký et al. 2008).

1.4.1. TRPV1 y Epilepsia

A diferencia de los canales de sodio, el TRPV1 es un blanco molecular novedoso en lo que respecta a la Epilepsia. Basándose en la creciente evidencia científica, varios estudios proponen al TRPV1 como un blanco promisorio para el desarrollo de nuevas ACDs (Asth et al. 2019; Cho et al. 2018; Fu, Xie y Zuo 2009; Gonzalez-Reyes et al. 2013; Nazıroglu 2015; Socala et al. 2015).

Los ratones *knockout* para el gen TRPV1 son menos susceptibles a las convulsiones inducidas por PTZ (Jia et al. 2015; Kong et al. 2014). Además, en modelos animales de *status epilepticus* inducido por Pilocarpina, se encontró una sobreexpresión del TRPV1 a nivel del hipocampo. En este mismo modelo, la activación del canal por Capsaicina (CAP) produjo un incremento de la transmisión sináptica en las regiones CA1 y CA3, mientras que la inhibición por Capsazepina (CPZ) disminuyó la actividad neuronal en las mismas regiones del hipocampo (Saffarzadeh et al. 2016). En modelos murinos de Epilepsia del lóbulo temporal (TLE, del inglés Temporal Lobe Epilepsy), se encontró una sobreexpresión significativa del TRPV1 en la región del giro dentado (Bhaskaran y Smith 2010). Del mismo modo, en pacientes con TLE mesial, se encontró una sobreexpresión del TRPV1 a nivel de la corteza temporal y el hipocampo (Sun et al. 2013).

En el modelo de convulsiones inducidas por sonido en ratas *Genetically Eepilepsy-Prone Rat* (GEPR-3), CPZ redujo considerablemente la severidad de las convulsiones (Cho et al. 2018). A su vez, en el modelo de Epilepsia genética inducida por sonido *Wistar Audiogenic Rat* (WAR), se observó una sobreexpresión del canal TRPV1 en distintas regiones del cerebro de los animales control y de aquellos sometidos a un protocolo de *kindling* sonoro (Lazarini-Lopes et al. 2022).

En líneas generales los antagonistas del TRPV1 han demostrado actividad anticonvulsiva en varios modelos animales de Epilepsia, mientras que los agonistas han mostrado ser pro-convulsivos (Bhaskaran y Smith 2010; Carletti et al. 2017; Cho et al. 2018; Gonzalez-Reyes et al. 2013; Jia et al. 2015; Shirazi et al. 2014; Socala et al. 2015). Sin embargo, existen algunos casos controversiales en este sentido,

donde algunos agonistas presentan un perfil farmacológico *in vivo* que se asemeja más al de un antagonista.

CAP, el componente activo de los pimientos picantes del género *Capsicum*, es un ligando exógeno del canal ampliamente estudiado. Esta molécula es capaz de activarlo con una potencia nanomolar, uniéndose a su sitio específico delimitado por los segmentos S3, S4 y el Loop-P entre los segmentos S4-S5 (Cao et al. 2013; Gavva et al. 2004; Jordt y Julius 2002). Mientras que en algunos estudios CAP ha demostrado ser proconvulsivante (Carletti et al. 2017; Gonzalez-Reyes et al. 2013; Jia et al. 2015; Manna y Umathe 2012b), otros autores han reportado su actividad anticonvulsiva en modelos de crisis inducida por Ácido Kaínico y PTZ (Abdel-Salam et al. 2020; T.-H. Lee et al. 2011).

En neuronas piramidales de la corteza prefrontal, CAP disminuye la excitabilidad neuronal suprimiendo la actividad neuronal epileptiforme, probablemente mediante la inhibición de las corrientes de sodio operadas por voltaje (Pasierski y Szulczyk 2020). En cultivos neuronales del ganglio trigémino, CAP inhibe las corrientes de sodio operadas por voltaje y su efecto es antagonizado por CPZ, un antagonista competitivo clásico del TRPV1 que comparte el sitio de unión de CAP (Liu et al. 2001). A su vez, debido a sus propiedades anfifílicas CAP y CPZ son capaces de modular la actividad de estos canales modificando las propiedades biofísicas de la membrana, como por ejemplo su elasticidad (Lundbaek et al. 2005).

Otro caso particular es el de la Piperina, el compuesto activo de la pimienta negra o *Piper nigrum*. A pesar de ser un agonista débil del TRPV1 *in vitro* (McNamara, Randall y Gunthorpe 2005; Okumura et al. 2010), este compuesto ha demostrado propiedades anticonvulsivas *in vivo* en modelos de crisis inducidas por Pilocarpina, Picrotoxina, Estricnina, BAYK-8644, PTZ y por aquellos basados en estímulos eléctricos como el MES test (Chen et al. 2013; Cruz et al. 2013; Mao et al. 2017; Mishra et al. 2015). Existen distintas hipótesis que intentan explicar este fenómeno. Mientras que algunos autores sugieren que este efecto es mediado directamente por el canal TRPV1 (Chen et al. 2013; Dong et al. 2019), otros proponen un mecanismo mediado por receptores opioides, gabaérgicos e incluso canales de sodio (Bukhari et al. 2013; Cruz et al. 2013; Mishra et al. 2015). Así mismo, aún cuando la Piperina es un agonista menos potente que CAP, este ligando presenta

una eficacia mucho mayor frente al TRPV1 y es capaz de inducir una desensibilización del receptor mucho más pronunciada (McNamara, Randall y Gunthorpe 2005). Este fenómeno de desensibilización permitiría también explicar la actividad analgésica y anticonvulsiva de estos agonistas. En particular, CAP se utiliza hoy en día como analgésico tópico en el tratamiento del dolor neuropático (Basith et al. 2016; Groninger y Schisler 2012).

1.4.2. TRPV1 y el sistema endocannabinoide

Existe una estrecha relación entre los receptores TRPV1 y el sistema endocannabinoide, el cual comprende a los receptores cannabinoides CB1 y CB2; a los ligandos endocannabinoides N-araquidonoiletanolamida (Anandamida) y 2-araquidonoilglicerol (2-AG); y a las enzimas implicadas en el metabolismo de los mismos: *Fatty Acid Amide Hydrolase* (FAAH) y *Monoacylglyceride Lipase* (MAGL) (Lu y Mackie 2016; Muller, Morales y Reggio 2019; Pacher, Bátakai y Kunos 2006).

La Anandamida es un activador endógeno del receptor metabotrópico CB1, que promueve la disminución de la actividad neuronal (Dasilva et al. 2014; Williams y Kirkham 1999). Aunque con una menor afinidad, este compuesto también es capaz de activar al canal TRPV1, lo que conduce a un incremento de la actividad neuronal (Van Der Stelt y Di Marzo 2004). En preparadas *in vitro*, se observó que la Anandamida produce un aumento de las corrientes postsinápticas excitatorias (EPSC) a través de la activación de los canales TRPV1 (Ross 2003; Tóth, Blumberg y Boczán 2009).

Ciertos fitocannabinoides como el Cannabidiol (CBD) y la Canabidivarina (CBDV), han demostrado propiedades anticonvulsivas en distintos modelos animales (Perucca 2017). El CBD, luego de demostrar ser seguro y eficaz como terapia *add-on* en distintos ensayos clínicos (Lattanzi et al. 2018; Silvestro et al. 2019), fue recientemente aprobado por la FDA para el tratamiento de los síndromes de Dravet y Lennox-Gastaut (Commissioner 2020).

El mecanismo anticonvulsivo de la CBDV es independiente de los receptores CB1 (Hill et al. 2013). Sin embargo, al igual que el CBD, ambos cannabinoides son capaces de activar al TRPV1 e inducir la desensibilización del receptor (Iannotti

et al. 2014). Nuevamente, la desensibilización del TRPV1 mediada por agonistas nos permitiría explicar la actividad observada *in vivo*. Sin embargo, el mecanismo anticonvulsivo de los cannabinoides no involucra a un único *target*, ni ha sido completamente elucidado aún (Gray y Whalley 2020).

1.4.3. Antagonistas TRPV1 e hipertermia

Como se discutió anteriormente, la modulación de los receptores TRPV1 es una estrategia ampliamente explorada en el desarrollo de fármacos para distintas patologías, particularmente para el tratamiento del dolor (Jara-Oseguera, Simon y Rosenbaum 2008). Sin embargo, varios antagonistas de este canal han presentado efectos adversos relacionados con la regulación de la temperatura corporal en etapas de ensayos preclínicos y clínicos (Gavva et al. 2008).

Dentro de la superfamilia de canales TRP, existen 7 isoformas capaces de activarse en diferentes rangos de temperatura denominadas *thermo-TRPs*, implicadas en la regulación de la temperatura corporal (Naziroğlu y Braidy 2017). El canal TRPV1 se activa en un rango de temperatura intermedio ($>42^{\circ}\text{C}$), si lo comparamos con otros canales como el TRPV2 ($>53^{\circ}\text{C}$) o el TRPM8 ($<30^{\circ}\text{C}$) (Dhaka, Viswanath y Patapoutian 2006).

Analizando el perfil de los antagonistas TRPV1 frente a los diferentes modos de activación (temperatura, pH, CAP), Garami y colaboradores establecieron una correlación entre el efecto de los antagonistas sobre los distintos modos de activación, y el resultado que este produce sobre la temperatura corporal. En roedores, los compuestos que bloquean, potencian o no modifican la activación por pH producen hipertermia, hipotermia o no modifican la temperatura corporal, respectivamente (Garami et al. 2018; Garami et al. 2010). En humanos sin embargo, la relación es más compleja ya que el bloqueo de la activación por temperatura también se encuentra relacionado con el efecto sobre la temperatura corporal (Garami et al. 2020).

Los antagonistas modo-selectivos son aquellos que bloquean potentemente la activación por CAP pero ejercen diferentes efectos sobre los otros modos de activación, y constituyen la denominada segunda generación de antagonistas TRPV1. Por

lo tanto, el diseño de antagonistas modo-selectivos constituye una de las estrategias posibles para abordar el problema de la hipertermia, aunque no la única (**Garami et al. 2020**).

1.5. Canales de sodio operados por voltaje

Existen 9 genes que codifican para la subunidad α de estos canales (Nav1.1-Nav1.9), y una décima (Nax) que podría actuar como un sensor de salinidad aunque no es sensible al voltaje (**Dolivo et al. 2021**). Las isoformas NaV1.1, NaV1.2, NaV1.3 y NaV1.6 se expresan principalmente a nivel del SNC, mientras que NaV1.7, NaV1.8 y NaV1.9 se encuentran mayoritariamente en el Sistema Nervioso Periférico (SNP). A nivel celular se distribuyen principalmente a lo largo de la membrana neuronal, en las dendritas, somas, axones y terminales nerviosas. La isoforma NaV1.4 es la predominante a nivel del músculo esquelético y la NaV1.5 a nivel del músculo cardíaco (aunque también puede encontrarse en SNC) (**J. Wang, Ou e Y.-J. Wang 2017**).

La subunidad α está formada por cuatro dominios homólogos (DI-DIV) plegados formando un poro central. Cada dominio consta de seis segmentos transmembrana (S1-S6), siendo los segmentos S5-S6 los que forman la región del poro o vía de permeación, mientras que los segmentos S1-S4 constituyen la región del sensor de voltaje (VSD, del inglés Voltage Sensor Domain) (Figura 1.5). Aun cuando estas regiones se encuentran separadas espacialmente en la arquitectura modular de este canal, ambas se encuentran íntimamente conectadas a nivel funcional. Dentro del VSD, el segmento S4 posee residuos de Arginina cargados positivamente, que constituyen el denominado sensor de voltaje, capaces de responder a cambios en el potencial de la membrana, promoviendo así la activación e inactivación del canal. Este movimiento en el sensor de voltaje se transmite a la región del poro, permitiendo la apertura del canal y el flujo de iones. Los *loops* que conectan los segmentos S5 y S6 en cada uno de los cuatro dominios (Loop-P) forman el filtro de selectividad del canal, donde cada *loop* contribuye con un aminoácido al denominado anillo DEKA: Aspartato (D) D-I, Glutamato (E) D-II, Lisina (K) D-III y Alanina (A) D-IV.

Junto a la subunidad α comúnmente se expresan subunidades accesorias β , que pueden modificar la cinética del canal y su sensibilidad al voltaje, pero no son

esenciales para la conducción de los iones. Las subunidades β -1 y β -3 se asocian de manera no covalente al canal y presentan una mayor similitud de secuencia entre ellas, mientras que β -2 y β -4 se unen de forma covalente a la subunidad α . A nivel del SNC, se expresan principalmente las subunidades β -1 y β -2 (Hull e Isom 2018).

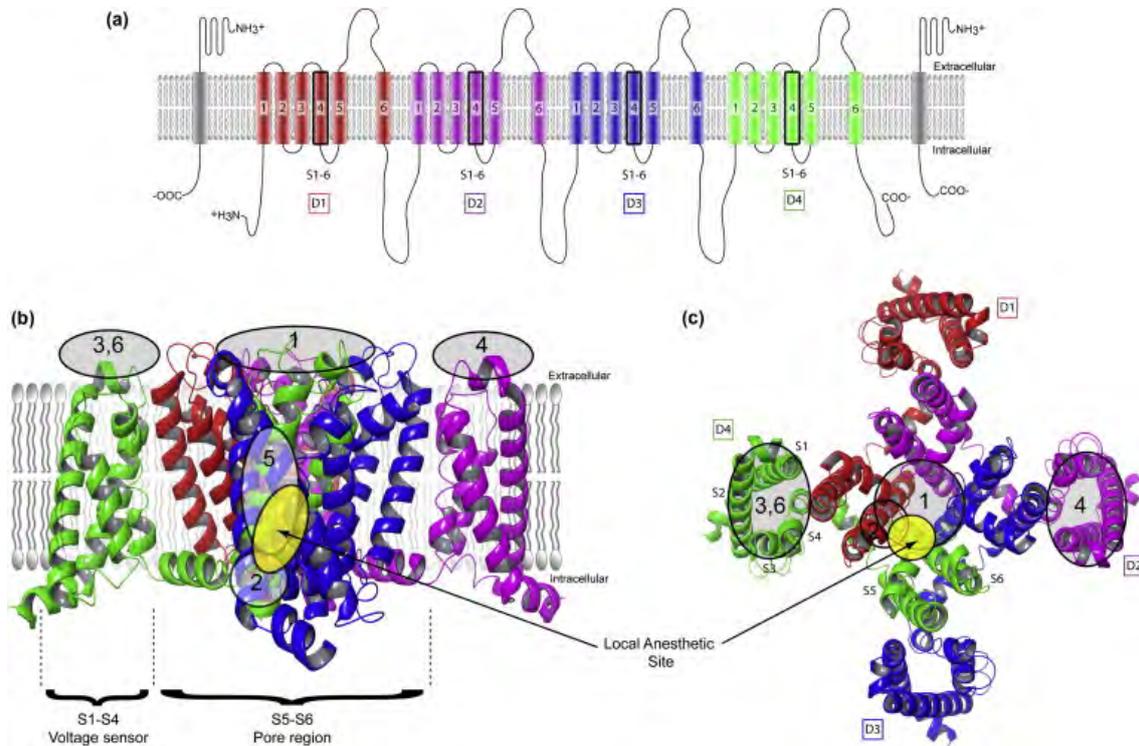


Figura 1.5: (a) Topología estructural de los canales NaV. Los dominios D1–D4 se representan en diferentes colores, mientras que las subunidades β se muestran en gris. Dentro de cada dominio se observan los segmentos transmembrana S1–S6 junto con la representación gráfica de los Loop-P. Vistas lateral (b) y superior (c) del canal NaV de la bacteria *Arcobacter butzleri* (PDB-ID: 3RVY), donde se resaltan los distintos sitios de unión de toxinas y moléculas pequeñas. Imagen tomada de Bagal et al. 2014 y reproducida con autorización de Elsevier.

En los NaV pueden distinguirse al menos tres estados funcionales diferentes: abierto, cerrado e inactivado (Figura 1.6). Como respuesta a la despolarización de la membrana celular, se produce un cambio conformacional y el canal pasa del estado cerrado al abierto, permitiendo así el ingreso de sodio a la célula a favor de su gradiente. Luego de unos microsegundos, el canal transita hacia un estado abierto no conductor, al que se denomina inactivado. Mientras el canal se encuentra en este estado inactivado, no es capaz de responder nuevamente a una despolarización de la

membrana. Esta refractariedad a sucesivos estímulos permite regular la frecuencia de disparos de los potenciales de acción.

Normalmente nos referimos a la inactivación como un único proceso. Sin embargo, en esta pueden distinguirse al menos dos fases con cinéticas muy diferentes, denominadas inactivación rápida e inactivación lenta.

La inactivación rápida ocurre mediante un mecanismo conocido como bola y cadena (*ball-and-chain*), donde una partícula de inactivación en forma de bola ocluye la región citoplasmática del poro. Esta partícula la constituye el *loop* intracelular que conecta los dominios D-III y D-IV, donde puede distinguirse un motivo característico (IFMT) que interacciona con su sitio de acoplamiento, sellando así el poro. El extremo C-terminal también se encuentra implicado en la fase de inactivación rápida, a la vez que contiene sitios de unión para otras proteínas que interaccionan con el canal (**Abriel y Kass 2005; Mantegazza et al. 2001**).

La inactivación lenta ocurre cuando el canal es expuesto a una despolarización prolongada o a una estimulación de alta frecuencia. Este proceso no depende del *loop* intracelular entre los dominios III y IV, sino que implica un cambio conformacional a nivel del poro. Por lo tanto, la cinética de este proceso es aproximadamente cuatro órdenes de magnitud más lenta que la inactivación rápida.

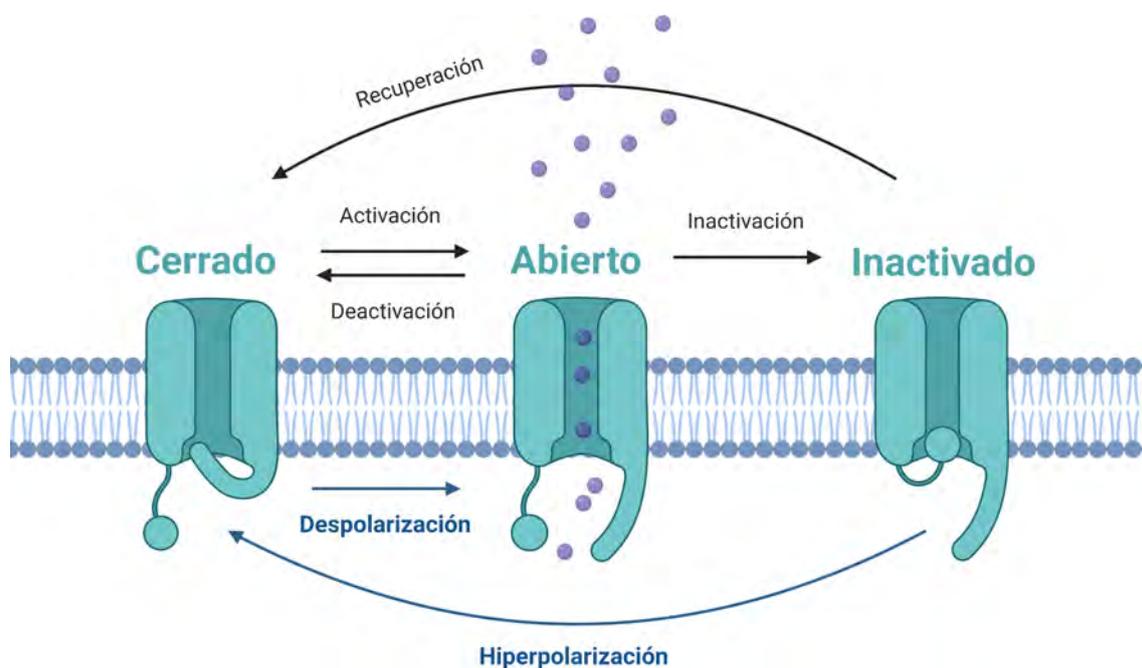


Figura 1.6: Esquema general los estados funcionales en los canales de sodio operados por voltaje.

La recuperación de la inactivación es un proceso voltaje y tiempo-dependiente, necesiéndose la repolarización del potencial de membrana para tener nuevamente canales NaV disponibles para disparar un nuevo potencial de acción. De esta manera, la proporción de canales en el estado cerrado e inactivado es una función del potencial de membrana en reposo.

La selectividad funcional en los bloqueantes de canales NaV es un determinante fundamental de su índice terapéutico. Estos compuestos, idealmente, alcanzarán su máxima potencia sólo si los canales funcionan bajo condiciones patológicas. De esta manera, un anestésico local puede impedir la excitabilidad neuronal, un analgésico o anticonvulsivo puede limitar la hiperexcitabilidad, mientras que un antiarrítmico es capaz de ajustar la excitabilidad del músculo cardíaco (Nardi et al. 2012).

La gran mayoría de los bloqueantes de canales de sodio actúa estabilizando el estado inactivado de estos canales. Este tipo de inhibición se la conoce como dependiente de estado (*state-dependent*) o dependiente del voltaje (*voltage-dependent*). En neuronas con una alta frecuencia de disparo de potenciales de acción o *firing*, como aquellas con un comportamiento patológico, los canales pasan frecuentemente por el estado inactivado, por lo que este efecto inhibitorio será acumulativo y se volverá

más importante a medida que aumente la frecuencia de estimulación (*use-dependent*) (Qiao et al. 2014).

1.5.1. Canales NaV y Epilepsia

Debido a su capacidad la responder a las variaciones en el potencial de la membrana plasmática, los canales de sodio operados por voltaje juegan un rol fundamental en la generación y transmisión coordinada de los potenciales de acción a nivel del sistema nervioso (Yu y Catterall 2003). La Epilepsia puede entenderse como un desbalance entre las sinapsis excitatorias e inhibitorias, por lo que las alteraciones en la funcionalidad de estos canales se encuentran íntimamente relacionadas a esta patología (Kaplan, Isom y Petrou 2016). Hasta el momento, mutaciones en las isoformas NaV1.1, NaV1.2, NaV1.3, NaV1.6 y NaV1.7 han sido relacionadas con la Epilepsia, mientras que no se encontró correlación con las mutaciones de las isoformas NaV1.4, NaV1.5, NaV1.8 y NaV1.9. A su vez, la subunidad regulatoria β -1 también ha sido asociada a la Epilepsia (Ademuwagun et al. 2021; Menezes et al. 2020).

En lo que respecta a la isoforma NaV1.2, mas de 100 mutaciones han sido identificadas en pacientes con Epilepsia (Reynolds, King y Gorman 2020). Esta isoforma se expresa principalmente a nivel de las neuronas exitatorias y glutamatérgicas (Sanders et al. 2018), a diferencia de la isoforma NaV1.1 que se expresa principalmente a nivel de las interneuronas GABAérgicas inhibitorias (Catterall, Kalume y Oakley 2010; Ogiwara et al. 2007).

En lineas generales, las variantes del gen SCN2A (hNaV1.2) asociadas a las epilepsias de inicio temprano, implican una ganancia de función y por lo tanto contribuyen a la hiperexcitabilidad neuronal. En estos casos, las crisis son bien controladas con bloqueantes de canales de sodio. Por otra parte en las epilepsias de inicio tardío, las variantes génicas han sido asociadas a una perdida de función, y los bloqueantes de canales de sodio no logran controlar las crisis de manera efectiva (Musto, Gardella y Møller 2020).

1.5.2. Sitios de unión en canales NaV

Se conocen hoy en día una gran variedad de moléculas capaces de modular a los NaV, dentro de las cuales podemos distinguir varias toxinas y moléculas orgánicas pequeñas. Utilizando diversas técnicas como la ingeniería genética, registros electrofisiológicos y la determinación estructural mediante rayos-X o Cryo-EM, se han identificado varios sitios de unión distribuidos sobre toda la superficie de estos canales (Nardi et al. 2012; S.-Y. Wang y G. K. Wang 2003). En la Figura 1.5 se muestran esquemáticamente los distintos sitios de unión, los cuales se describen a continuación.

Sitio 1

Este sitio se encuentra en la porción extracelular del poro, formado por los *Loop-P* de los cuatro dominios homólogos. Las moléculas que se unen a este sitio se conocen como bloqueantes de poro, ya que obstruyen físicamente el paso de los iones a través del canal. Un exponente característico de este sitio es la Tetrodotoxina (TTX), una potente neurotoxina de origen marino. Según su afinidad por esta toxina los NaV pueden clasificarse como sensibles a TTX (NaV1.1-1.4, 1.6 y 1.7) o insensibles a TTX (NaV1.5, 1.8 y 1.9) (C. H. Lee y Ruben 2008). Existen también otras toxinas que se unen en esta región, como la Saxitoxina extraída de dinoflagelados marinos o la μ -conotoxina, presente en caracoles marinos del género *Conus*.

Sitio 2

Formado por los segmentos S6 de los cuatro dominios DI-DIV, este es el sitio de unión de neurotoxinas no peptídicas como la Batracotoxina (BTX), presente en algunas especies de ranas y aves, o la Veratrina, un alcaloide esteroideo de origen vegetal. Estas toxinas se unen al canal en su conformación abierta, estabilizandola y generando una activación persistente. Esta región se solapa parcialmente con el sitio de unión de moléculas pequeñas, donde actúan los anestésicos locales y las ACDs.

Sitio 3

Este sitio comprende al *loop* extracelular que conecta los segmentos S3-S4 en

el dominio DIV. Debido a la presencia de varios residuos con carga negativa a pH fisiológico, este sitio es capaz de unir péptidos conteniendo residuos básicos, como las α -toxinas de escorpión o las toxinas de anémona de mar. La unión de estas toxinas al sitio 3 impide la inactivación rápida del canal.

Sitio 4

Las moléculas que interaccionan con este sitio ejercen su acción como modificadores de la apertura o *gating modifiers*, es decir que corren el umbral de activación del canal y modifican también la amplitud máxima de la corriente. El sitio 4 está formado por los *loops* extracelulares que conectan los segmentos S1–S2 y S3–S4 del dominio DII, aunque el DIII también se encuentra parcialmente implicado. Las β -toxinas de escorpiones corren el umbral de activación hacia potenciales más negativos, promoviendo así la activación del canal. Por su parte, la Huwentoxina-IV (HwTx-IV) y la Protoxina-II (ProTx-II), causan un corrimiento del umbral de activación hacia potenciales más despolarizados, por lo que actúan como inhibidores de los NaV.

Sitio 5

Este sitio comprende los segmentos S6 del dominio DI y S5 del DIV, y es blanco de toxinas altamente lipofílicas producidas por dinoflagelados marinos, como las Brevetoxinas (*Karenia brevis*) o Ciguatoxinas (*Gambierdiscus toxicus*). Ambas toxinas presentan una mayor afinidad por el estado activado del canal, y corren el umbral de activación e inactivación hacia potenciales más hiperpolarizados, promoviendo una hiperactivación de estos canales (Stevens, Peigneur y Tytgat 2011).

Sitio 6

Este puede considerarse como un sub-sitio dentro del sitio 3, y de manera similar a las toxinas del sitio 3 generalmente impiden la inactivación, aunque sin la típica dependencia del voltaje. Ciertas Δ -conotoxinas de caracoles cónicos como Δ -SVIE, TxVIA y GmVIA interaccionan con este sitio, causando una acción farmacológica sinérgica con las toxinas α de escorpión, posiblemente capturando de manera similar al sensor de voltaje S4-DIV. Este sitio se ubica en el *loop* extracelular que conecta los segmentos S3-S4 del dominio DIV.

Sitio 7

También denominado sitio de unión de piretroides, el sitio 7 se ubica en el segmento S6 del dominio DIII, y es el sitio de unión de algunos insecticidas como el dicloro difenil tricloroetano (DDT) y los piretroides. Se ha demostrado que estos compuestos inhiben la inactivación del canal, y provocan una activación persistente.

Sitio de moléculas pequeñas

Las moléculas pequeñas como ciertas ACDs (Fenitoína, Carbamazepina, Lamotrigina y Lacosamida) o anestésicos locales (Lidocaína, Tetracaína y Bupivacaína) se unen a este sitio localizado en la cavidad interna del poro, formado principalmente por las hélices transmembrana S6 de los dominios DI–DIV. A su vez, como se describió anteriormente, este sitio se superpone en gran medida con el sitio de unión de BTX (sitio 2). La secuencia de aminoácidos en la región del poro es casi idéntica para todos los subtipos de canales NaV, por lo que ha sido difícil identificar compuestos que se unan a este sitio y presenten selectividad de isoformas (**Oelstrom et al. 2014**).

1.6. Epilepsia refractaria

La definición de Epilepsia refractaria más reciente fue elaborada por la ILAE, y establece que se considera Epilepsia refractaria cuando dos o más esquemas de tratamiento farmacológico no logran controlar las crisis de manera sostenida en el tiempo. Estos esquemas de tratamiento pueden incluir una o varias ACDs y se asume que fueron bien elegidos por el especialista y tolerados por el paciente (**Kwan et al. 2010**).

De acuerdo con esta definición se estima que aproximadamente un 30% de los pacientes diagnosticados con Epilepsia son refractarios, lo que nos conduce a una necesidad imperiosa de desarrollar nuevos tratamientos farmacológicos que den respuesta a un gran sector de la población con necesidades médicas no resueltas (**Vreman et al. 2019**).

Existen varias hipótesis que intentan explicar el fenómeno de refractariedad o

farmacorresistencia en la Epilepsia (**Tang, Hartz y Bauer 2017**), dentro de las cuales podemos mencionar:

Hipótesis de los transportadores

Esta hipótesis plantea la existencia de una sobreexpresión de transportadores de eflujo a nivel de los distintos componentes del SNC, como las neuronas, los astrocitos o las células endoteliales de la Barrera Hemato Encefálica (BHE). Estos transportadores de eflujo son proteínas de membrana pertenecientes a la familia ABC (del inglés, ATP-binding cassette), que actúan expulsando las moléculas de fármaco hacia el exterior de la célula. De este modo, aun cuando se alcancen concentraciones terapéuticas de los fármacos en plasma, las concentraciones en el líquido cefalorraquídeo son subterapéuticas. Existen varios transportadores de eflujo implicados en este mecanismo, dentro de los cuales la Glicoproteína P (PgP) y las proteínas de multirresistencia a medicamentos (MRP) como la MRP1, MRP2 y MRP5 son las más estudiadas (**Kwan y Brodie 2006; Sisodiya et al. 2002**).

Hipótesis farmacocinética

El planteo aquí es similar al anterior, solo que la sobreexpresión de estos transportadores se produce a nivel de los órganos periféricos como el intestino, hígado o riñones. Como consecuencia, los fármacos no logran alcanzar concentraciones plasmáticas en el rango terapéutico (**Lazarowski et al. 2007**).

Hipótesis del blanco molecular

De naturaleza farmacodinámica, esta hipótesis plantea que una alteración en los blancos moleculares de las ACDs puede disminuir la efectividad de los mismos. Esta alteración puede ser genética, como es el caso de las mutaciones en canales de sodio que dan origen a las Epilepsias febriles del tipo GEFS+ (**Escayg et al. 2000; Wallace et al. 1998**); o adquiridas, como por ejemplo la internalización de receptores GABA durante la evolución del *status epilepticus* (**Goodkin, Yeh y Kapur 2005**).

Hipótesis de la gravedad intrínseca

Según esta hipótesis, la farmacacorresistencia es inherente a la severidad de la enfermedad. Es decir, los factores neurobiológicos que determinan la severidad de

la Epilepsia son los que le confieren la refractariedad al tratamiento farmacológico (Rogawski y Johnson 2008).

Hipótesis de las redes neuronales

Postula que la ocurrencia de convulsiones afecta la plasticidad neuronal, eliminando conexiones neuronales y generando otras nuevas, lo que impide a los mecanismos fisiológicos - así como también a las ACDs - controlar la aparición de nuevas convulsiones (Fang et al. 2011).

Cabe destacar que estas hipótesis no son mutuamente excluyentes, con lo que más de una puede coexistir dando origen a la Epilepsia refractaria.

1.7. Fármacos multiobjetivo y enfermedades complejas

Las enfermedades complejas son aquellas causadas por la combinación de distintos factores genéticos, ambientales y conductuales. Como se describió anteriormente, las causas de la Epilepsia pueden ser muy diversas y mutuamente no excluyentes, es decir que varios factores etiológicos pueden contribuir al desarrollo de la patología, convirtiéndola en una enfermedad compleja (Ottman 2005).

El paradigma clásico del desarrollo de fármacos se basa en el concepto de “*one target, one drug*”, donde las drogas altamente potentes y selectivas son preferidas por sobre aquellas “*drogas promiscuas*”, que interactúan con múltiples dianas biológicas. Sin embargo, la evidencia actual nos ha llevado a un cambio de paradigma en este sentido, adoptando el principio de “*one drug, multiple targets*” (Morphy, Kay y Rankovic 2004; Roth, Sheffler y Kroeze 2004). Los fármacos *multitarget* han demostrado ser seguros e incluso más eficaces que aquellos *monotarget* en el tratamiento de diversas enfermedades complejas (Bolognesi 2013; Bolognesi y Cavalli 2016; Cavalli et al. 2008; Gentile et al. 2017; Löscher 2021).

Como se describió en la sección 1.3.2, varias de las ACDs más efectivas de uso actual actúan sobre más de una diana biológica. Esto soporta el objetivo general

de esta tesis, de encontrar compuestos *multitarget* que modulen simultáneamente a los canales NaV1.2 y TRPV1, como estrategia para desarrollar nuevas ACDs más eficaces, seguras y mejor toleradas.

Referencias Capítulo 1

- Abdel-Salam, O. M. E. et al.** (2020). “Capsaicin Exerts Anti-convulsant and Neuroprotective Effects in Pentylentetrazole-Induced Seizures”. *Neurochemical Research* 45.5, págs. 1045-1061. DOI: [10.1007/s11064-020-02979-3](https://doi.org/10.1007/s11064-020-02979-3).
- Abriel, H. y R. S. Kass** (2005). “Regulation of the voltage-gated cardiac sodium channel Nav1.5 by interacting proteins”. *Trends in Cardiovascular Medicine* 15.1, págs. 35-40. DOI: [10.1016/j.tcm.2005.01.001](https://doi.org/10.1016/j.tcm.2005.01.001).
- Adamczyk, P. et al.** (2012). “The effects of cannabinoid CB1, CB2 and vanilloid TRPV1 receptor antagonists on cocaine addictive behavior in rats”. *Brain Research* 1444, págs. 45-54. DOI: [10.1016/j.brainres.2012.01.030](https://doi.org/10.1016/j.brainres.2012.01.030).
- Ademuwagun, I. A. et al.** (2021). “Voltage Gated Sodium Channel Genes in Epilepsy: Mutations, Functional Studies, and Treatment Dimensions”. *Frontiers in Neurology* 12.
- Aghazadeh Tabrizi, M. et al.** (2017). “Medicinal Chemistry, Pharmacology, and Clinical Implications of TRPV1 Receptor Antagonists: TRPV1 RECEPTOR ANTAGONISTS”. *Medicinal Research Reviews* 37.4, págs. 936-983. DOI: [10.1002/med.21427](https://doi.org/10.1002/med.21427).
- Almeida, A. S. de, L. d. B. Bernardes y G. Trevisan** (2021). “TRP channels in cancer pain”. *European Journal of Pharmacology* 904, págs. 174185. DOI: [10.1016/j.ejphar.2021.174185](https://doi.org/10.1016/j.ejphar.2021.174185).
- Armijo, J. A. et al.** (2005). “Ion Channels and Epilepsy”. *Current Pharmaceutical Design* 11.15, págs. 1975-2003. DOI: [10.2174/1381612054021006](https://doi.org/10.2174/1381612054021006).
- Asth, L. et al.** (2019). “Exploiting cannabinoid and vanilloid mechanisms for epilepsy treatment”. *Epilepsy & Behavior*, págs. 106832. DOI: [10.1016/j.yebeh.2019.106832](https://doi.org/10.1016/j.yebeh.2019.106832).

- Bagal**, S. K. et al. (2014). “Recent progress in sodium channel modulators for pain”. *Bioorganic & Medicinal Chemistry Letters* 24.16, págs. 3690-3699. DOI: [10.1016/j.bmcl.2014.06.038](https://doi.org/10.1016/j.bmcl.2014.06.038).
- Barker-Haliski**, M. y H. S. **White** (2015). “Glutamatergic Mechanisms Associated with Seizures and Epilepsy”. *Cold Spring Harbor Perspectives in Medicine* 5.8, a022863. DOI: [10.1101/cshperspect.a022863](https://doi.org/10.1101/cshperspect.a022863).
- Barton**, M. E. et al. (2001). “Pharmacological characterization of the 6 Hz psychomotor seizure model of partial epilepsy”. *Epilepsy Research* 47.3, págs. 217-227. DOI: [10.1016/S0920-1211\(01\)00302-3](https://doi.org/10.1016/S0920-1211(01)00302-3).
- Basith**, S. et al. (2016). “Harnessing the Therapeutic Potential of Capsaicin and Its Analogues in Pain and Other Diseases”. *Molecules* 21.8. DOI: [10.3390/molecules21080966](https://doi.org/10.3390/molecules21080966).
- Bell**, G. S., A. **Neligan** y J. W. **Sander** (2014). “An unknown quantity—the worldwide prevalence of epilepsy”. *Epilepsia* 55.7, págs. 958-962. DOI: [10.1111/epi.12605](https://doi.org/10.1111/epi.12605).
- Bhaskaran**, M. D. y B. N. **Smith** (2010). “Effects of TRPV1 activation on synaptic excitation in the dentate gyrus of a mouse model of temporal lobe epilepsy”. *Experimental Neurology*. Beta-amyloid and tau protein abnormalities in Alzheimer’s disease 223.2, págs. 529-536. DOI: [10.1016/j.expneurol.2010.01.021](https://doi.org/10.1016/j.expneurol.2010.01.021).
- Bolognesi**, M. L. (2013). “Polypharmacology in a single drug: multitarget drugs”. *Current Medicinal Chemistry* 20.13, págs. 1639-1645. DOI: [10.2174/0929867311320130004](https://doi.org/10.2174/0929867311320130004).
- Bolognesi**, M. L. y A. **Cavalli** (2016). “Multitarget Drug Discovery and Polypharmacology”. *ChemMedChem* 11.12, págs. 1190-1192. DOI: [10.1002/cmdc.201600161](https://doi.org/10.1002/cmdc.201600161).
- Britton**, J. (2016). “Autoimmune epilepsy”. *Handbook of Clinical Neurology* 133, págs. 219-245. DOI: [10.1016/B978-0-444-63432-0.00013-X](https://doi.org/10.1016/B978-0-444-63432-0.00013-X).
- Brodie**, M. J. (2017). “Sodium Channel Blockers in the Treatment of Epilepsy”. *CNS drugs* 31.7, págs. 527-534. DOI: [10.1007/s40263-017-0441-0](https://doi.org/10.1007/s40263-017-0441-0).
- Bukhari**, I. A. et al. (2013). “The analgesic and anticonvulsant effects of piperine in mice”. *Journal of Physiology and Pharmacology: An Official Journal of the Polish Physiological Society* 64.6, págs. 789-794.
- Cao**, E. et al. (2013). “TRPV1 structures in distinct conformations reveal activation mechanisms”. *Nature* 504.7478, págs. 113-118. DOI: [10.1038/nature12823](https://doi.org/10.1038/nature12823).

- Carletti**, F. et al. (2017). “Neuronal nitric oxide synthase is involved in CB/TRPV1 signalling: Focus on control of hippocampal hyperexcitability”. *Epilepsy Research* 138, págs. 18-25. DOI: [10.1016/j.eplesyres.2017.09.018](https://doi.org/10.1016/j.eplesyres.2017.09.018).
- Carnevale**, V. y T. **Rohacs** (2016). “TRPV1: A Target for Rational Drug Design”. *Pharmaceuticals* 9.3. DOI: [10.3390/ph9030052](https://doi.org/10.3390/ph9030052).
- Caterina**, M. J. et al. (1997). “The capsaicin receptor: a heat-activated ion channel in the pain pathway”. *Nature* 389.6653, págs. 816-824. DOI: [10.1038/39807](https://doi.org/10.1038/39807).
- Catterall**, W. A., F. **Kalume** y J. C. **Oakley** (2010). “NaV1.1 channels and epilepsy”. *The Journal of Physiology* 588.Pt 11, págs. 1849-1859. DOI: [10.1113/jphysiol.2010.187484](https://doi.org/10.1113/jphysiol.2010.187484).
- Cavalli**, A. et al. (2008). “Multi-target-Directed Ligands To Combat Neurodegenerative Diseases”. *Journal of Medicinal Chemistry* 51.3, págs. 347-372. DOI: [10.1021/jm7009364](https://doi.org/10.1021/jm7009364).
- Chen**, C.-Y. et al. (2013). “Piperine exerts anti-seizure effects via the TRPV1 receptor in mice”. *European Journal of Pharmacology* 714.1, págs. 288-294. DOI: [10.1016/j.ejphar.2013.07.041](https://doi.org/10.1016/j.ejphar.2013.07.041).
- Cho**, S. J. et al. (2018). “Inhibition of transient potential receptor vanilloid type 1 suppresses seizure susceptibility in the genetically epilepsy-prone rat”. *CNS Neuroscience & Therapeutics* 24.1, págs. 18-28. DOI: [10.1111/cns.12770](https://doi.org/10.1111/cns.12770).
- Clapham**, D. E. (2003). “TRP channels as cellular sensors”. *Nature* 426.6966, págs. 517-524. DOI: [10.1038/nature02196](https://doi.org/10.1038/nature02196).
- Commissioner**, O. o. t. (2020). *FDA Approves First Drug Comprised of an Active Ingredient Derived from Marijuana to Treat Rare, Severe Forms of Epilepsy*.
- Cotterman-Hart**, S. (2015a). “Antiepileptic Drugs”. *Epilepsy and Brain Tumors*. Elsevier, págs. 171-193. DOI: [10.1016/B978-0-12-417043-8.00012-2](https://doi.org/10.1016/B978-0-12-417043-8.00012-2).
- (2015b). “Chapter 11 - Antiepileptic Drugs: First Generation”. *Epilepsy and Brain Tumors*. Ed. por H. B. **Newton** y M. **Maschio**. Boston: Academic Press, págs. 159-169. DOI: [10.1016/B978-0-12-417043-8.00011-0](https://doi.org/10.1016/B978-0-12-417043-8.00011-0).
- Cruz**, G. M. P. da et al. (2013). “Piperine decreases pilocarpine-induced convulsions by GABAergic mechanisms”. *Pharmacology Biochemistry and Behavior* 104, págs. 144-153. DOI: [10.1016/j.pbb.2013.01.002](https://doi.org/10.1016/j.pbb.2013.01.002).
- Cui**, M. et al. (2016). “Chapter Three - Polymodal Transient Receptor Potential Vanilloid Type 1 Nociceptor: Structure, Modulators, and Therapeutic Applications”. *Advances in Protein Chemistry and Structural Biology*. Ed. por R.

- Donev**. Vol. 104. Ion channels as therapeutic targets, part B. Academic Press, págs. 81-125. DOI: [10.1016/bs.apcsb.2015.11.005](https://doi.org/10.1016/bs.apcsb.2015.11.005).
- D'Andrea Meira, I. et al.** (2019). “Ketogenic Diet and Epilepsy: What We Know So Far”. *Frontiers in Neuroscience* 13. DOI: [10.3389/fnins.2019.00005](https://doi.org/10.3389/fnins.2019.00005).
- Dasilva, M. et al.** (2014). “Anandamide activation of CB1 receptors increases spontaneous bursting and oscillatory activity in the thalamus”. *Neuroscience* 265, págs. 72-82. DOI: [10.1016/j.neuroscience.2014.01.049](https://doi.org/10.1016/j.neuroscience.2014.01.049).
- Dhaka, A., V. Viswanath y A. Patapoutian** (2006). *TRP Ion Channels And Temperature Sensation*. review-article. DOI: [10.1146/annurev.neuro.29.051605.112958](https://doi.org/10.1146/annurev.neuro.29.051605.112958).
- Dolivo, D. et al.** (2021). “The Nax (SCN7A) channel: an atypical regulator of tissue homeostasis and disease”. *Cellular and molecular life sciences: CMLS*. DOI: [10.1007/s00018-021-03854-2](https://doi.org/10.1007/s00018-021-03854-2).
- Dong, Y. et al.** (2019). “A distinct structural mechanism underlies TRPV1 activation by piperine”. *Biochemical and Biophysical Research Communications* 516.2, págs. 365-372. DOI: [10.1016/j.bbrc.2019.06.039](https://doi.org/10.1016/j.bbrc.2019.06.039).
- Escayg, A. et al.** (2000). “Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2”. *Nature Genetics* 24.4, págs. 343-345. DOI: [10.1038/74159](https://doi.org/10.1038/74159).
- Escelsior, A. et al.** (2020). “Transient receptor potential vanilloid 1 antagonism in neuroinflammation, neuroprotection and epigenetic regulation: potential therapeutic implications for severe psychiatric disorders treatment”. *Psychiatric Genetics* 30.2, págs. 39-48. DOI: [10.1097/YPG.000000000000249](https://doi.org/10.1097/YPG.000000000000249).
- Fakih, D. et al.** (2021). “Capsazepine decreases corneal pain syndrome in severe dry eye disease”. *Journal of Neuroinflammation* 18.1, pág. 111. DOI: [10.1186/s12974-021-02162-7](https://doi.org/10.1186/s12974-021-02162-7).
- Fang, M. et al.** (2011). “A new hypothesis of drug refractory epilepsy: neural network hypothesis”. *Medical Hypotheses* 76.6, págs. 871-876. DOI: [10.1016/j.mehy.2011.02.039](https://doi.org/10.1016/j.mehy.2011.02.039).
- Fisher, R. S. et al.** (2005). “Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)”. *Epilepsia* 46.4, págs. 470-472. DOI: [10.1111/j.0013-9580.2005.66104.x](https://doi.org/10.1111/j.0013-9580.2005.66104.x).

- Fisher, R. S. et al.** (2014). “ILAE Official Report: A practical clinical definition of epilepsy”. *Epilepsia* 55.4, págs. 475-482. DOI: [10.1111/epi.12550](https://doi.org/10.1111/epi.12550).
- Forsgren, L. et al.** (2005). “The epidemiology of epilepsy in Europe – a systematic review”. *European Journal of Neurology* 12.4, págs. 245-253. DOI: [10.1111/j.1468-1331.2004.00992.x](https://doi.org/10.1111/j.1468-1331.2004.00992.x).
- Fu, M., Z. Xie y H. Zuo** (2009). “TRPV1: A potential target for antiepileptogenesis”. *Medical Hypotheses* 73.1, págs. 100-102. DOI: [10.1016/j.mehy.2009.01.005](https://doi.org/10.1016/j.mehy.2009.01.005).
- Gaidin, S. G. et al.** (2019). “Epileptiform activity promotes decreasing of Ca²⁺ conductivity of NMDARs, AMPARs, KARs, and voltage-gated calcium channels in Mg²⁺-free model”. *Epilepsy Research* 158, pág. 106224. DOI: [10.1016/j.eplepsyres.2019.106224](https://doi.org/10.1016/j.eplepsyres.2019.106224).
- Garami, A. et al.** (2018). “TRPV1 antagonists that cause hypothermia, instead of hyperthermia, in rodents: Compounds’ pharmacological profiles, in vivo targets, thermoeffectors recruited and implications for drug development”. *Acta Physiologica (Oxford, England)* 223.3. DOI: [10.1111/apha.13038](https://doi.org/10.1111/apha.13038).
- Garami, A. et al.** (2010). “Contributions of Different Modes of TRPV1 Activation to TRPV1 Antagonist-Induced Hyperthermia”. *Journal of Neuroscience* 30.4, págs. 1435-1440. DOI: [10.1523/JNEUROSCI.5150-09.2010](https://doi.org/10.1523/JNEUROSCI.5150-09.2010).
- Garami, A. et al.** (2020). “Hyperthermia induced by transient receptor potential vanilloid-1 (TRPV1) antagonists in human clinical trials: Insights from mathematical modeling and meta-analysis”. *Pharmacology & Therapeutics* 208, pág. 107474. DOI: [10.1016/j.pharmthera.2020.107474](https://doi.org/10.1016/j.pharmthera.2020.107474).
- Gavva, N. R. et al.** (2004). “Molecular determinants of vanilloid sensitivity in TRPV1”. *The Journal of Biological Chemistry* 279.19, págs. 20283-20295. DOI: [10.1074/jbc.M312577200](https://doi.org/10.1074/jbc.M312577200).
- Gavva, N. R. et al.** (2008). “Pharmacological blockade of the vanilloid receptor TRPV1 elicits marked hyperthermia in humans”. *PAIN* 136.1, págs. 202-210. DOI: [10.1016/j.pain.2008.01.024](https://doi.org/10.1016/j.pain.2008.01.024).
- Gentile, C. et al.** (2017). “Kinase Inhibitors in Multitargeted Cancer Therapy”. *Current Medicinal Chemistry* 24.16, págs. 1671-1686. DOI: [10.2174/0929867324666170112112734](https://doi.org/10.2174/0929867324666170112112734).
- Gonzalez-Reyes, L. E. et al.** (2013). “TRPV1 antagonist capsazepine suppresses 4-AP-induced epileptiform activity in vitro and electrographic seizures in vivo”.

- Experimental Neurology* 250, págs. 321-332. DOI: [10.1016/j.expneurol.2013.10.010](https://doi.org/10.1016/j.expneurol.2013.10.010).
- Goodkin, H. P., J.-L. Yeh y J. Kapur** (2005). “Status Epilepticus Increases the Intracellular Accumulation of GABAA Receptors”. *The Journal of Neuroscience* 25.23, págs. 5511-5520. DOI: [10.1523/JNEUROSCI.0900-05.2005](https://doi.org/10.1523/JNEUROSCI.0900-05.2005).
- Gray, R. A. y B. J. Whalley** (2020). “The proposed mechanisms of action of CBD in epilepsy”. *Epileptic Disorders* 22.S1, S10-S15. DOI: [10.1684/epd.2020.1135](https://doi.org/10.1684/epd.2020.1135).
- Groninger, H. y R. E. Schisler** (2012). “Topical Capsaicin for Neuropathic Pain #255”. *Journal of Palliative Medicine* 15.8, págs. 946-947. DOI: [10.1089/jpm.2012.9571](https://doi.org/10.1089/jpm.2012.9571).
- Gunthorpe, M. J. et al.** (2000). “Voltage- and time-dependent properties of the recombinant rat vanilloid receptor (rVR1)”. *The Journal of Physiology* 525.Pt 3, págs. 747-759. DOI: [10.1111/j.1469-7793.2000.t01-1-00747.x](https://doi.org/10.1111/j.1469-7793.2000.t01-1-00747.x).
- Gutman, G. A. et al.** (2005). “International Union of Pharmacology. LIII. Nomenclature and Molecular Relationships of Voltage-Gated Potassium Channels”. *Pharmacological Reviews* 57.4, págs. 473-508. DOI: [10.1124/pr.57.4.10](https://doi.org/10.1124/pr.57.4.10).
- Hanaya, R. et al.** (2012). “Low distribution of synaptic vesicle protein 2A and synaptotagmin-1 in the cerebral cortex and hippocampus of spontaneously epileptic rats exhibiting both tonic convulsion and absence seizure”. *Neuroscience* 221, págs. 12-20. DOI: [10.1016/j.neuroscience.2012.06.058](https://doi.org/10.1016/j.neuroscience.2012.06.058).
- Hill, T. D. M. et al.** (2013). “Cannabidiol-rich cannabis extracts are anticonvulsant in mouse and rat via a CB1 receptor-independent mechanism”. *British Journal of Pharmacology* 170.3, págs. 679-692. DOI: [10.1111/bph.12321](https://doi.org/10.1111/bph.12321).
- Huang, Z. J. y A. Paul** (2019). “The diversity of GABAergic neurons and neural communication elements”. *Nature Reviews Neuroscience* 20.9, págs. 563-572. DOI: [10.1038/s41583-019-0195-4](https://doi.org/10.1038/s41583-019-0195-4).
- Hull, J. M. y L. L. Isom** (2018). “Voltage-gated sodium channel subunits: The power outside the pore in brain development and disease”. *Neuropharmacology. Channelopathies* 132, págs. 43-57. DOI: [10.1016/j.neuropharm.2017.09.018](https://doi.org/10.1016/j.neuropharm.2017.09.018).
- Iannotti, F. A. et al.** (2014). “Nonpsychotropic Plant Cannabinoids, Cannabidiol (CBDV) and Cannabidiol (CBD), Activate and Desensitize Transient Receptor Potential Vanilloid 1 (TRPV1) Channels in Vitro: Potential for the Treatment of Neuronal Hyperexcitability”. *ACS Chemical Neuroscience* 5.11, págs. 1131-1141. DOI: [10.1021/cn5000524](https://doi.org/10.1021/cn5000524).

- Ihara, Y.** et al. (2016). “Retigabine, a Kv7.2/Kv7.3-Channel Opener, Attenuates Drug-Induced Seizures in Knock-In Mice Harboring Kcnq2 Mutations”. *PLOS ONE* 11.2, e0150095. DOI: [10.1371/journal.pone.0150095](https://doi.org/10.1371/journal.pone.0150095).
- Jara-Oseguera, A., S. A. Simon y T. Rosenbaum** (2008). “TRPV1: ON THE ROAD TO PAIN RELIEF”. *Current molecular pharmacology* 1.3, págs. 255-269.
- Jia, Y.-F.** et al. (2015). “Interference of TRPV1 function altered the susceptibility of PTZ-induced seizures”. *Frontiers in Cellular Neuroscience* 9, pág. 20. DOI: [10.3389/fncel.2015.00020](https://doi.org/10.3389/fncel.2015.00020).
- Jordt, S.-E. y D. Julius** (2002). “Molecular basis for species-specific sensitivity to ”hotçhili peppers”. *Cell* 108.3, págs. 421-430. DOI: [10.1016/s0092-8674\(02\)00637-2](https://doi.org/10.1016/s0092-8674(02)00637-2).
- Kaplan, D. I., L. L. Isom y S. Petrou** (2016). “Role of Sodium Channels in Epilepsy”. *Cold Spring Harbor Perspectives in Medicine* 6.6, a022814. DOI: [10.1101/cshperspect.a022814](https://doi.org/10.1101/cshperspect.a022814).
- Kedei, N.** et al. (2001). “Analysis of the native quaternary structure of vanilloid receptor 1”. *The Journal of Biological Chemistry* 276.30, págs. 28613-28619. DOI: [10.1074/jbc.M103272200](https://doi.org/10.1074/jbc.M103272200).
- Kehne, J. H.** et al. (2017). “The National Institute of Neurological Disorders and Stroke (NINDS) Epilepsy Therapy Screening Program (ETSP)”. *Neurochemical Research* 42.7, págs. 1894-1903. DOI: [10.1007/s11064-017-2275-z](https://doi.org/10.1007/s11064-017-2275-z).
- Kong, W.-L.** et al. (2014). “Role of TRPV1 in susceptibility to PTZ-induced seizure following repeated hyperthermia challenges in neonatal mice”. *Epilepsy & Behavior* 31, págs. 276-280. DOI: [10.1016/j.yebeh.2013.10.022](https://doi.org/10.1016/j.yebeh.2013.10.022).
- Krall, R. L.** et al. (1978a). “Antiepileptic drug development: I. History and a program for progress”. *Epilepsia* 19.4, págs. 393-408. DOI: [10.1111/j.1528-1157.1978.tb04506.x](https://doi.org/10.1111/j.1528-1157.1978.tb04506.x).
- Krall, R. L.** et al. (1978b). “Antiepileptic Drug Development: II. Anticonvulsant Drug Screening”. *Epilepsia* 19.4, págs. 409-428. DOI: [10.1111/j.1528-1157.1978.tb04507.x](https://doi.org/10.1111/j.1528-1157.1978.tb04507.x).
- Kwan, P. y M. J. Brodie** (2006). “Refractory epilepsy: mechanisms and solutions”. *Expert Review of Neurotherapeutics* 6.3, págs. 397-406. DOI: [10.1586/14737175.6.3.397](https://doi.org/10.1586/14737175.6.3.397).

- Kwan, P.** et al. (2010). “Definition of drug resistant epilepsy: Consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies”. *Epilepsia* 51.6, págs. 1069-1077. DOI: [10.1111/j.1528-1167.2009.02397.x](https://doi.org/10.1111/j.1528-1167.2009.02397.x).
- Lapalme-Remis, S.** y **G. D. Cascino** (2016). “Imaging for Adults With Seizures and Epilepsy”. *Seizures and Epilepsy*, pág. 29.
- Lattanzi, S.** et al. (2018). “Efficacy and Safety of Cannabidiol in Epilepsy: A Systematic Review and Meta-Analysis”. *Drugs* 78.17, págs. 1791-1804. DOI: [10.1007/s40265-018-0992-5](https://doi.org/10.1007/s40265-018-0992-5).
- Lazarini-Lopes, W.** et al. (2022). “Increased TRPV1 Channels and FosB Protein Expression Are Associated with Chronic Epileptic Seizures and Anxiogenic-like Behaviors in a Preclinical Model of Temporal Lobe Epilepsy”. *Biomedicines* 10.2, pág. 416. DOI: [10.3390/biomedicines10020416](https://doi.org/10.3390/biomedicines10020416).
- Lazarowski, A.** et al. (2007). “ABC transporters during epilepsy and mechanisms underlying multidrug resistance in refractory epilepsy”. *Epilepsia* 48 Suppl 5, págs. 140-149. DOI: [10.1111/j.1528-1167.2007.01302.x](https://doi.org/10.1111/j.1528-1167.2007.01302.x).
- Lee, C. H.** y **P. C. Ruben** (2008). “Interaction between voltage-gated sodium channels and the neurotoxin, tetrodotoxin”. *Channels (Austin, Tex.)* 2.6, págs. 407-412. DOI: [10.4161/chan.2.6.7429](https://doi.org/10.4161/chan.2.6.7429).
- Lee, T.-H.** et al. (2011). “Capsaicin prevents kainic acid-induced epileptogenesis in mice”. *Neurochemistry International* 58.6, págs. 634-640. DOI: [10.1016/j.neuint.2011.01.027](https://doi.org/10.1016/j.neuint.2011.01.027).
- Lerche, H., K. Jurkat-Rott** y **F. Lehmann-Horn** (2001). “Ion channels and epilepsy”. *American Journal of Medical Genetics* 106.2, págs. 146-159. DOI: [10.1002/ajmg.1582](https://doi.org/10.1002/ajmg.1582).
- Liu, L.** et al. (2001). “Capsaicin inhibits activation of voltage-gated sodium currents in capsaicin-sensitive trigeminal ganglion neurons”. *Journal of Neurophysiology* 85.2, págs. 745-758. DOI: [10.1152/jn.2001.85.2.745](https://doi.org/10.1152/jn.2001.85.2.745).
- Löscher, W.** (2017). “Animal Models of Seizures and Epilepsy: Past, Present, and Future Role for the Discovery of Antiseizure Drugs”. *Neurochemical Research* 42.7, págs. 1873-1888. DOI: [10.1007/s11064-017-2222-z](https://doi.org/10.1007/s11064-017-2222-z).
- (2021). “Single-Target Versus Multi-Target Drugs Versus Combinations of Drugs With Multiple Targets: Preclinical and Clinical Evidence for the Treatment or Prevention of Epilepsy”. *Frontiers in Pharmacology* 12.

- Löscher, W.** y **D. Schmidt** (2011). “Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma”. *Epilepsia* 52.4, págs. 657-678. DOI: [10.1111/j.1528-1167.2011.03024.x](https://doi.org/10.1111/j.1528-1167.2011.03024.x).
- Löscher, W.** et al. (2013). “New avenues for anti-epileptic drug discovery and development”. *Nature Reviews Drug Discovery* 12.10, págs. 757-776. DOI: [10.1038/nrd4126](https://doi.org/10.1038/nrd4126).
- Löscher, W.** et al. (2020). “Drug Resistance in Epilepsy: Clinical Impact, Potential Mechanisms, and New Innovative Treatment Options”. *Pharmacological Reviews* 72.3. Ed. por E. L. **Barker**, págs. 606-638. DOI: [10.1124/pr.120.019539](https://doi.org/10.1124/pr.120.019539).
- Lu, H.-C.** y **K. Mackie** (2016). “An introduction to the endogenous cannabinoid system”. *Biological psychiatry* 79.7, págs. 516-525. DOI: [10.1016/j.biopsych.2015.07.028](https://doi.org/10.1016/j.biopsych.2015.07.028).
- Lundbaek, J. A.** et al. (2005). “Capsaicin regulates voltage-dependent sodium channels by altering lipid bilayer elasticity”. *Molecular Pharmacology* 68.3, págs. 680-689. DOI: [10.1124/mol.105.013573](https://doi.org/10.1124/mol.105.013573).
- Luszczki, J. J.** (2009). “Third-generation antiepileptic drugs: mechanisms of action, pharmacokinetics and interactions”. *Pharmacological Reports* 61.2, págs. 197-216. DOI: [10.1016/S1734-1140\(09\)70024-6](https://doi.org/10.1016/S1734-1140(09)70024-6).
- Lyseng-Williamson, K. A.** (2011). “Levetiracetam: a review of its use in epilepsy”. *Drugs* 71.4, págs. 489-514. DOI: [10.2165/11204490-000000000-00000](https://doi.org/10.2165/11204490-000000000-00000).
- Manna, S. S. S.** y **S. N. Umathe** (2012a). “A possible participation of transient receptor potential vanilloid type 1 channels in the antidepressant effect of fluoxetine”. *European Journal of Pharmacology* 685.1-3, págs. 81-90. DOI: [10.1016/j.ejphar.2012.04.023](https://doi.org/10.1016/j.ejphar.2012.04.023).
- (2012b). “Involvement of transient receptor potential vanilloid type 1 channels in the pro-convulsant effect of anandamide in pentylenetetrazole-induced seizures”. *Epilepsy Research* 100.1, págs. 113-124. DOI: [10.1016/j.eplepsyres.2012.02.003](https://doi.org/10.1016/j.eplepsyres.2012.02.003).
- Mantegazza, M.** et al. (2001). “Role of the C-terminal domain in inactivation of brain and cardiac sodium channels”. *Proceedings of the National Academy of Sciences of the United States of America* 98.26, págs. 15348-15353. DOI: [10.1073/pnas.211563298](https://doi.org/10.1073/pnas.211563298).
- Mao, K.** et al. (2017). “Anticonvulsant effect of piperine ameliorates memory impairment, inflammation and oxidative stress in a rat model of pilocarpine-induced

- epilepsy”. *Experimental and Therapeutic Medicine* 13.2, págs. 695-700. DOI: [10.3892/etm.2016.4001](https://doi.org/10.3892/etm.2016.4001).
- McNamara, F. N., A. Randall y M. J. Gunthorpe** (2005). “Effects of piperine, the pungent component of black pepper, at the human vanilloid receptor (TRPV1)”. *British Journal of Pharmacology* 144.6, págs. 781-790. DOI: [10.1038/sj.bjp.0706040](https://doi.org/10.1038/sj.bjp.0706040).
- Mendoza-Torreblanca, J. G. et al.** (2013). “Synaptic vesicle protein 2A: basic facts and role in synaptic function”. *The European Journal of Neuroscience* 38.11, págs. 3529-3539. DOI: [10.1111/ejn.12360](https://doi.org/10.1111/ejn.12360).
- Menezes, L. F. S. et al.** (2020). “Epilepsy-Related Voltage-Gated Sodium Channelopathies: A Review”. *Frontiers in Pharmacology* 11.
- Miceli, F. et al.** (1993). “KCNQ2-Related Disorders”. *GeneReviews*®. Ed. por M. P. **Adam** et al. Seattle (WA): University of Washington, Seattle.
- Miller, C.** (2000). “An overview of the potassium channel family”. *Genome Biology* 1.4, REVIEWS0004. DOI: [10.1186/gb-2000-1-4-reviews0004](https://doi.org/10.1186/gb-2000-1-4-reviews0004).
- Mishra, A. et al.** (2015). “Anticonvulsant mechanisms of piperine, a piperidine alkaloid”. *Channels* 9.5, págs. 317-323. DOI: [10.1080/19336950.2015.1092836](https://doi.org/10.1080/19336950.2015.1092836).
- Morphy, R., C. Kay y Z. Rankovic** (2004). “From magic bullets to designed multiple ligands”. *Drug Discovery Today* 9.15, págs. 641-651. DOI: [10.1016/S1359-6446\(04\)03163-0](https://doi.org/10.1016/S1359-6446(04)03163-0).
- Muller, C., P. Morales y P. H. Reggio** (2019). “Cannabinoid Ligands Targeting TRP Channels”. *Frontiers in Molecular Neuroscience* 11. DOI: [10.3389/fnmol.2018.00487](https://doi.org/10.3389/fnmol.2018.00487).
- Mulley, J. C. et al.** (2003). “Channelopathies as a genetic cause of epilepsy”. *Current Opinion in Neurology* 16.2, págs. 171-176.
- Musto, E., E. Gardella y R. S. Møller** (2020). “Recent advances in treatment of epilepsy-related sodium channelopathies”. *European Journal of Paediatric Neurology* 24, págs. 123-128. DOI: [10.1016/j.ejpn.2019.12.009](https://doi.org/10.1016/j.ejpn.2019.12.009).
- Nardi, A. et al.** (2012). “Advances in Targeting Voltage-Gated Sodium Channels with Small Molecules”. *ChemMedChem* 7.10, págs. 1712-1740. DOI: [10.1002/cmdc.201200298](https://doi.org/10.1002/cmdc.201200298).
- Naziroglu, M.** (2015). “TRPV1 Channel: A Potential Drug Target for Treating Epilepsy”. *Current Neuropharmacology* 13.2, págs. 239-247. DOI: [10.2174/1570159X13666150216222543](https://doi.org/10.2174/1570159X13666150216222543).

- Naziroğlu, M. y N. Braidy** (2017). “Thermo-Sensitive TRP Channels: Novel Targets for Treating Chemotherapy-Induced Peripheral Pain”. *Frontiers in Physiology* 8. DOI: [10.3389/fphys.2017.01040](https://doi.org/10.3389/fphys.2017.01040).
- Oelstrom, K. et al.** (2014). “Evolutionarily conserved intracellular gate of voltage-dependent sodium channels”. *Nature Communications* 5.1, pág. 3420. DOI: [10.1038/ncomms4420](https://doi.org/10.1038/ncomms4420).
- Ogiwara, I. et al.** (2007). “Nav1.1 localizes to axons of parvalbumin-positive inhibitory interneurons: a circuit basis for epileptic seizures in mice carrying an Scn1a gene mutation”. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27.22, págs. 5903-5914. DOI: [10.1523/JNEUROSCI.5270-06.2007](https://doi.org/10.1523/JNEUROSCI.5270-06.2007).
- Okumura, Y. et al.** (2010). “Activation of TRPV1 and TRPA1 by black pepper components”. *Bioscience, Biotechnology, and Biochemistry* 74.5, págs. 1068-1072. DOI: [10.1271/bbb.90964](https://doi.org/10.1271/bbb.90964).
- Ottman, R.** (2005). “Analysis of Genetically Complex Epilepsies”. *Epilepsia* 46.Suppl 10, págs. 7-14. DOI: [10.1111/j.1528-1167.2005.00350.x](https://doi.org/10.1111/j.1528-1167.2005.00350.x).
- Pacher, P., S. Bátkai y G. Kunos** (2006). “The Endocannabinoid System as an Emerging Target of Pharmacotherapy”. *Pharmacological reviews* 58.3, págs. 389-462. DOI: [10.1124/pr.58.3.2](https://doi.org/10.1124/pr.58.3.2).
- Papetti, L. et al.** (2013). “Metabolic epilepsy: An update”. *Brain and Development* 35.9, págs. 827-841. DOI: [10.1016/j.braindev.2012.11.010](https://doi.org/10.1016/j.braindev.2012.11.010).
- Pasierski, M. y B. Szulczyk** (2020). “Capsaicin inhibits sodium currents and epileptiform activity in prefrontal cortex pyramidal neurons”. *Neurochemistry International* 135, pág. 104709. DOI: [10.1016/j.neuint.2020.104709](https://doi.org/10.1016/j.neuint.2020.104709).
- Perucca, E.** (2017). “Cannabinoids in the Treatment of Epilepsy: Hard Evidence at Last?” *Journal of Epilepsy Research* 7.2, págs. 61-76. DOI: [10.14581/jer.17012](https://doi.org/10.14581/jer.17012).
- Perucca, E. y T. Tomson** (2011). “The pharmacological treatment of epilepsy in adults”. *The Lancet. Neurology* 10.5, págs. 446-456. DOI: [10.1016/S1474-4422\(11\)70047-3](https://doi.org/10.1016/S1474-4422(11)70047-3).
- Piper, A. S. et al.** (1999). “A study of the voltage dependence of capsaicin-activated membrane currents in rat sensory neurones before and after acute desensitization”. *The Journal of Physiology* 518.Pt 3, págs. 721-733. DOI: [10.1111/j.1469-7793.1999.0721p.x](https://doi.org/10.1111/j.1469-7793.1999.0721p.x).

- Premkumar**, L. S. y **P. Sikand** (2008). “TRPV1: A Target for Next Generation Analgesics”. *Current Neuropharmacology* 6.2, págs. 151-163. DOI: [10.2174/157015908784533888](https://doi.org/10.2174/157015908784533888).
- Qiao**, X. et al. (2014). “Properties of human brain sodium channel α -subunits expressed in HEK293 cells and their modulation by carbamazepine, phenytoin and lamotrigine”. *British Journal of Pharmacology* 171.4, págs. 1054-1067. DOI: [10.1111/bph.12534](https://doi.org/10.1111/bph.12534).
- Rajakulendran**, S. y **M. G. Hanna** (2016). “The Role of Calcium Channels in Epilepsy”. *Cold Spring Harbor Perspectives in Medicine* 6.1, a022723. DOI: [10.1101/cshperspect.a022723](https://doi.org/10.1101/cshperspect.a022723).
- Rasul**, F. T. et al. (2017). “Current Surgical Options for Patients with Epilepsy”. *Current Pharmaceutical Design* 23.42, págs. 6508-6523. DOI: [10.2174/1381612823666171031101](https://doi.org/10.2174/1381612823666171031101)
- Reynolds**, C., **M. D. King** y **K. M. Gorman** (2020). “The phenotypic spectrum of SCN2A-related epilepsy”. *European Journal of Paediatric Neurology* 24, págs. 117-122. DOI: [10.1016/j.ejpn.2019.12.016](https://doi.org/10.1016/j.ejpn.2019.12.016).
- Rho**, J. M., **S. D. Donevan** y **M. A. Rogawski** (1994). “Mechanism of action of the anticonvulsant felbamate: opposing effects on N-methyl-D-aspartate and gamma-aminobutyric acidA receptors”. *Annals of Neurology* 35.2, págs. 229-234. DOI: [10.1002/ana.410350216](https://doi.org/10.1002/ana.410350216).
- Rogawski**, M. A. y **M. R. Johnson** (2008). “Intrinsic severity as a determinant of antiepileptic drug refractoriness”. *Epilepsy Currents* 8.5, págs. 127-130. DOI: [10.1111/j.1535-7511.2008.00272.x](https://doi.org/10.1111/j.1535-7511.2008.00272.x).
- Rosenbaum**, T. et al. (2004). “Ca²⁺/calmodulin modulates TRPV1 activation by capsaicin”. *The Journal of General Physiology* 123.1, págs. 53-62. DOI: [10.1085/jgp.200308906](https://doi.org/10.1085/jgp.200308906).
- Ross**, R. A. (2003). “Anandamide and vanilloid TRPV1 receptors”. *British Journal of Pharmacology* 140.5, págs. 790-801. DOI: [10.1038/sj.bjp.0705467](https://doi.org/10.1038/sj.bjp.0705467).
- Roth**, B. L., **D. J. Sheffler** y **W. K. Kroeze** (2004). “Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia”. *Nature Reviews. Drug Discovery* 3.4, págs. 353-359. DOI: [10.1038/nrd1346](https://doi.org/10.1038/nrd1346).
- Saffarzadeh**, F. et al. (2016). “TRPV1 receptors augment basal synaptic transmission in CA1 and CA3 pyramidal neurons in epilepsy”. *Neuroscience* 314, págs. 170-178. DOI: [10.1016/j.neuroscience.2015.11.045](https://doi.org/10.1016/j.neuroscience.2015.11.045).

- Sanders**, S. J. et al. (2018). “Progress in Understanding and Treating SCN2A-Mediated Disorders”. *Trends in Neurosciences* 41.7, págs. 442-456. DOI: [10.1016/j.tins.2018.03.011](https://doi.org/10.1016/j.tins.2018.03.011).
- Saxena**, V. S. y V. V. **Nadkarni** (2011). “Nonpharmacological treatment of epilepsy”. *Annals of Indian Academy of Neurology* 14.3, págs. 148-152. DOI: [10.4103/0972-2327.85870](https://doi.org/10.4103/0972-2327.85870).
- Sedgwick**, S. G. y S. J. **Smerdon** (1999). “The ankyrin repeat: a diversity of interactions on a common structural framework”. *Trends in Biochemical Sciences* 24.8, págs. 311-316. DOI: [10.1016/s0968-0004\(99\)01426-7](https://doi.org/10.1016/s0968-0004(99)01426-7).
- Shank**, R. P. et al. (2000). “An Overview of the Preclinical Aspects of Topiramate: Pharmacology, Pharmacokinetics, and Mechanism of Action”. *Epilepsia* 41.s1, págs. 3-9. DOI: [10.1111/j.1528-1157.2000.tb02163.x](https://doi.org/10.1111/j.1528-1157.2000.tb02163.x).
- Shirazi**, M. et al. (2014). “Involvement of central TRPV1 receptors in pentylene-tetrazole and amygdala-induced kindling in male rats”. *Neurological Sciences* 35.8, págs. 1235-1241. DOI: [10.1007/s10072-014-1689-5](https://doi.org/10.1007/s10072-014-1689-5).
- Silvestro**, S. et al. (2019). “Use of Cannabidiol in the Treatment of Epilepsy: Efficacy and Security in Clinical Trials”. *Molecules* 24.8. DOI: [10.3390/molecules24081459](https://doi.org/10.3390/molecules24081459).
- Sisodiya**, S. M. et al. (2002). “Drug resistance in epilepsy: expression of drug resistance proteins in common causes of refractory epilepsy”. *Brain* 125.1, págs. 22-31. DOI: [10.1093/brain/awf002](https://doi.org/10.1093/brain/awf002).
- Socala**, K. et al. (2015). “-Spinasterol, a TRPV1 receptor antagonist, elevates the seizure threshold in three acute seizure tests in mice”. *Journal of Neural Transmission* 122.9, págs. 1239-1247. DOI: [10.1007/s00702-015-1391-7](https://doi.org/10.1007/s00702-015-1391-7).
- Spillane**, J., D. M. **Kullmann** y M. G. **Hanna** (2016). “Genetic neurological channelopathies: molecular genetics and clinical phenotypes”. *Journal of Neurology, Neurosurgery & Psychiatry* 87.1, págs. 37-48. DOI: [10.1136/jnnp-2015-311233](https://doi.org/10.1136/jnnp-2015-311233).
- Steinlein**, O. K. (2008). “Genetics and epilepsy”. *Dialogues in Clinical Neuroscience* 10.1, págs. 29-38. DOI: [10.31887/DCNS.2008.10.1/oksteinlein](https://doi.org/10.31887/DCNS.2008.10.1/oksteinlein).
- Stevens**, M., S. **Peigneur** y J. **Tytgat** (2011). “Neurotoxins and Their Binding Areas on Voltage-Gated Sodium Channels”. *Frontiers in Pharmacology* 2. DOI: [10.3389/fphar.2011.00071](https://doi.org/10.3389/fphar.2011.00071).

- Sun, F.-J.** et al. (2013). “Increased expression of TRPV1 in the cortex and hippocampus from patients with mesial temporal lobe epilepsy”. *Journal of molecular neuroscience: MN* 49.1, págs. 182-193. DOI: [10.1007/s12031-012-9878-2](https://doi.org/10.1007/s12031-012-9878-2).
- Szallasi, A.** y **P. M. Blumberg** (1999). “Vanilloid (Capsaicin) receptors and mechanisms”. *Pharmacological Reviews* 51.2, págs. 159-212.
- Tang, F., A. M. S. Hartz** y **B. Bauer** (2017). “Drug-Resistant Epilepsy: Multiple Hypotheses, Few Answers”. *Frontiers in Neurology* 8. DOI: [10.3389/fneur.2017.00301](https://doi.org/10.3389/fneur.2017.00301).
- Tatulian, L.** et al. (2001). “Activation of Expressed KCNQ Potassium Currents and Native Neuronal M-Type Potassium Currents by the Anti-Convulsant Drug Retigabine”. *Journal of Neuroscience* 21.15, págs. 5535-5545. DOI: [10.1523/JNEUROSCI.21-15-05535.2001](https://doi.org/10.1523/JNEUROSCI.21-15-05535.2001).
- Terzian, A. L. B.** et al. (2014). “Medial prefrontal cortex Transient Receptor Potential Vanilloid Type 1 (TRPV1) in the expression of contextual fear conditioning in Wistar rats”. *Psychopharmacology* 231.1, págs. 149-157. DOI: [10.1007/s00213-013-3211-9](https://doi.org/10.1007/s00213-013-3211-9).
- Tian, Y.-H.** et al. (2018). “Blockade of TRPV1 Inhibits Methamphetamine-induced Rewarding Effects”. *Scientific Reports* 8.1, pág. 882. DOI: [10.1038/s41598-018-19207-2](https://doi.org/10.1038/s41598-018-19207-2).
- Tokudome, K.** et al. (2016). “Synaptic vesicle glycoprotein 2A (SV2A) regulates kindling epileptogenesis via GABAergic neurotransmission”. *Scientific Reports* 6, pág. 27420. DOI: [10.1038/srep27420](https://doi.org/10.1038/srep27420).
- Tominaga, M.** et al. (1998). “The cloned capsaicin receptor integrates multiple pain-producing stimuli”. *Neuron* 21.3, págs. 531-543. DOI: [10.1016/s0896-6273\(00\)80564-4](https://doi.org/10.1016/s0896-6273(00)80564-4).
- Tóth, A., P. M. Blumberg** y **J. Boczán** (2009). “Anandamide and the vanilloid receptor (TRPV1)”. *Vitamins and Hormones* 81, págs. 389-419. DOI: [10.1016/S0083-6729\(09\)81015-7](https://doi.org/10.1016/S0083-6729(09)81015-7).
- Tsai, J.-J.** et al. (2018). “Perampanel, an AMPA receptor antagonist: From clinical research to practice in clinical settings”. *Acta Neurologica Scandinavica* 137.4, págs. 378-391. DOI: [10.1111/ane.12879](https://doi.org/10.1111/ane.12879).
- Tsuboyama, M., H. L. Kaye** y **A. Rotenberg** (2020). “Review of Transcranial Magnetic Stimulation in Epilepsy”. *Clinical Therapeutics* 42.7, págs. 1155-1168. DOI: [10.1016/j.clinthera.2020.05.016](https://doi.org/10.1016/j.clinthera.2020.05.016).

- Valentina, Z., F. Marco y R. Mirko** (2012). “Sodium Channel Blockers as Therapeutic Target for Treating Epilepsy: Recent Updates”. *Current Topics in Medicinal Chemistry* 12.9, págs. 962-970.
- Van Der Stelt, M. y V. Di Marzo** (2004). “Endovanilloids. Putative endogenous ligands of transient receptor potential vanilloid 1 channels”. *European Journal of Biochemistry* 271.10, págs. 1827-1834. DOI: [10.1111/j.1432-1033.2004.04081.x](https://doi.org/10.1111/j.1432-1033.2004.04081.x).
- Vezzani, A. et al.** (2016). “Infections, inflammation and epilepsy”. *Acta Neuropathologica* 131.2, págs. 211-234. DOI: [10.1007/s00401-015-1481-5](https://doi.org/10.1007/s00401-015-1481-5).
- Vreman, R. A. et al.** (2019). “Unmet Medical Need: An Introduction to Definitions and Stakeholder Perceptions”. *Value in Health* 22.11, págs. 1275-1282. DOI: [10.1016/j.jval.2019.07.007](https://doi.org/10.1016/j.jval.2019.07.007).
- Vyklický, L. et al.** (2008). “Calcium-dependent desensitization of vanilloid receptor TRPV1: a mechanism possibly involved in analgesia induced by topical application of capsaicin”. *Physiological Research* 57 Suppl 3, S59-68.
- Wallace, R. H. et al.** (1998). “Febrile seizures and generalized epilepsy associated with a mutation in the Na⁺-channel beta1 subunit gene SCN1B”. *Nature Genetics* 19.4, págs. 366-370. DOI: [10.1038/1252](https://doi.org/10.1038/1252).
- Wang, J., S.-W. Ou e Y.-J. Wang** (2017). “Distribution and function of voltage-gated sodium channels in the nervous system”. *Channels* 11.6, págs. 534-554. DOI: [10.1080/19336950.2017.1380758](https://doi.org/10.1080/19336950.2017.1380758).
- Wang, S.-Y. y G. K. Wang** (2003). “Voltage-gated sodium channels as primary targets of diverse lipid-soluble neurotoxins”. *Cellular Signalling* 15.2, págs. 151-159. DOI: [10.1016/s0898-6568\(02\)00085-2](https://doi.org/10.1016/s0898-6568(02)00085-2).
- Wei, F. et al.** (2017). “Ion Channel Genes and Epilepsy: Functional Alteration, Pathogenic Potential, and Mechanism of Epilepsy”. *Neuroscience Bulletin* 33.4, págs. 455-477. DOI: [10.1007/s12264-017-0134-1](https://doi.org/10.1007/s12264-017-0134-1).
- WHO** (2018). “Organización Mundial de la Salud”. DOI: <https://www.who.int/es/news-room/fact-sheets/detail/epilepsy>.
- Williams, C. M. y T. C. Kirkham** (1999). “Anandamide induces overeating: mediation by central cannabinoid (CB1) receptors”. *Psychopharmacology* 143.3, págs. 315-317. DOI: [10.1007/s002130050953](https://doi.org/10.1007/s002130050953).

- Wong, S., R. Mani y S. Danish** (2019). “Comparison and Selection of Current Implantable Anti-Epileptic Devices”. *Neurotherapeutics* 16.2, págs. 369-380. DOI: [10.1007/s13311-019-00727-2](https://doi.org/10.1007/s13311-019-00727-2).
- Xu, J. H. et al.** (2010). “Nuclear localization of Ca(v)2.2 and its distribution in the mouse central nervous system, and changes in the hippocampus during and after pilocarpine-induced status epilepticus”. *Neuropathology and Applied Neurobiology* 36.1, págs. 71-85. DOI: [10.1111/j.1365-2990.2009.01044.x](https://doi.org/10.1111/j.1365-2990.2009.01044.x).
- Yu, F. H. y W. A. Catterall** (2003). “Overview of the voltage-gated sodium channel family”. *Genome Biology* 4.3, pág. 207. DOI: [10.1186/gb-2003-4-3-207](https://doi.org/10.1186/gb-2003-4-3-207).
- Zhang, S. et al.** (2019). *Ion Channels in Epilepsy: Blasting Fuse for Neuronal Hyperexcitability*. IntechOpen. DOI: [10.5772/intechopen.83698](https://doi.org/10.5772/intechopen.83698).

Capítulo 2

Descubrimiento de fármacos



2.1. Introducción

El descubrimiento y desarrollo de un nuevo fármaco es un proceso lento, costoso y desafiante, que involucra diversas áreas del conocimiento como la química, la biología, la matemática, la farmacología y la informática, entre otras. En números, un estudio reciente indica que el costo de introducir un nuevo fármaco en el mercado ronda en promedio los USD 500 millones (**Jayasundara et al. 2019**), mientras que según DiMasi y colaboradores esta cifra ascendería a los USD 2.600 millones (**DiMasi, Grabowski y Hansen 2016**). Por otra parte, el tiempo que transcurre hasta que la droga finalmente alcanza el mercado es variable y se estima entre 12 y 15 años (**Hughes et al. 2011**).

Es posible identificar distintas etapas dentro de este proceso. La primera, denominada **descubrimiento de fármacos**, comprende la identificación y validación de nuevos *targets* y la posterior identificación de compuestos que presentan actividad frente al blanco molecular en cuestión, los cuales suelen denominarse prototipos activos o *hits*. Estos *hits* son posteriormente optimizados, dando origen a los compuestos cabeza de serie o *leads*. La etapa de **ensayos pre-clínicos** involucra pruebas *in vitro* e *in vivo* para optimizar la potencia, evaluar la potencial toxicidad y comenzar la caracterización farmacocinética. Tras su **aprobación**, los candidatos más promisorios entran en la fase de ensayos clínicos realizados en humanos. En esta etapa se determinan una serie de parámetros como el perfil de efectos adversos, la eficacia, seguridad y el perfil farmacocinético en humanos. Finalmente, luego de su aprobación por la autoridad sanitaria la droga es introducida en el mercado, y comienza la etapa de **vigilancia post-comercialización** (**Hughes et al. 2011**). El esquema de la Figura 2.1 resume las etapas principales del descubrimiento y desarrollo de un nuevo fármaco.

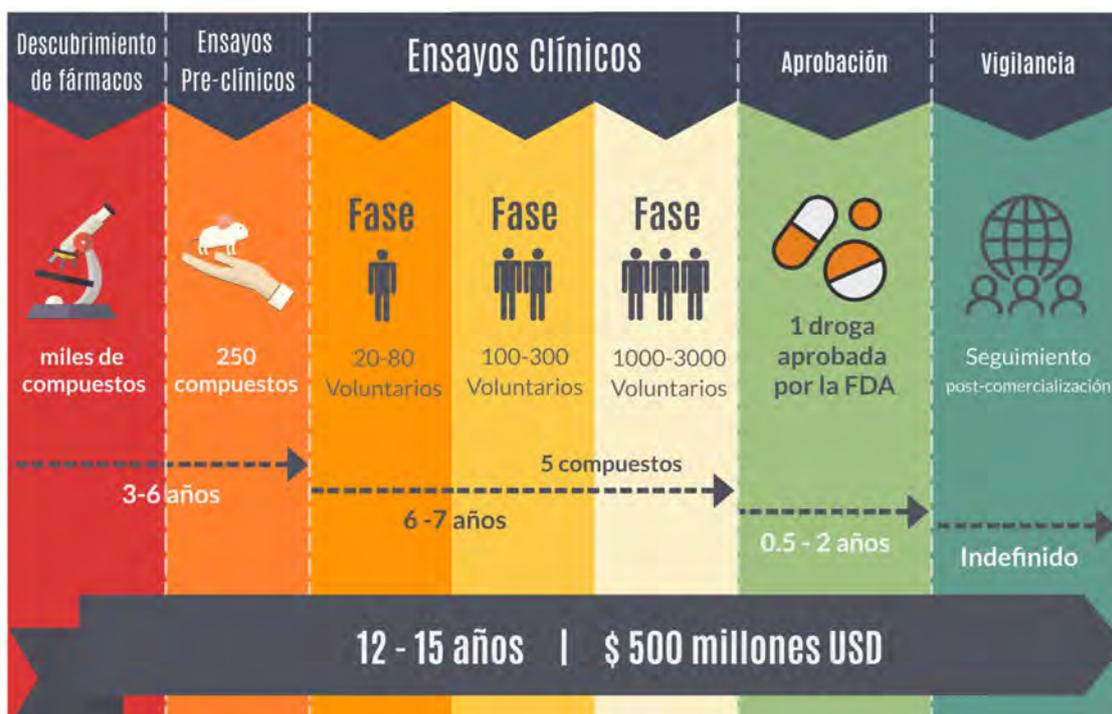


Figura 2.1: Esquema general del proceso de descubrimiento y desarrollo de un fármaco. Adaptado de *Pharmaceutical Research and manufacturers of America* y **Jaya-sundara** et al. 2019

Existen varias estrategias que pueden ser empleadas para identificar nuevos compuestos activos, dentro de las cuales podemos distinguir dos grandes categorías: Las metodologías clásicas y las metodologías computacionales o *in silico*.

2.2. Metodologías clásicas

2.2.1. Serendipia

El descubrimiento al azar o “por casualidad” es quizás uno de los métodos más intuitivos por el que es posible descubrir un nuevo fármaco. Existen varios fármacos utilizados hoy en día que fueron descubiertos de manera no intencional. Un ejemplo clásico es el caso de la Penicilina, cuya acción antibacteriana fue hallada fortuitamente por Alexander Fleming a comienzos del siglo XX, lo que le valdría más tarde un premio Nobel (**Tan** y **Tatsumura** 2015). Otro ejemplo más actual lo constituye el Sildenafil, comercializado para el tratamiento de la disfunción eréctil por Pfizer

bajo el nombre de “Viagra”, que fue inicialmente desarrollado para el tratamiento de la hipertensión y la angina de pecho (**Ghofrani, Osterloh y Grimminger 2006**).

2.2.2. Análisis de la medicina tradicional

El estudio de cómo nuestros ancestros curaban enfermedades es también una forma de encontrar nuevos fármacos. Este conocimiento forma parte del patrimonio cultural de las distintas civilizaciones que han existido o existen hoy en día. Quizás la más conocida sea la medicina tradicional china, pero se encuentran otros ejemplos como la Ayurveda (India), Kampo (Japón) o Unani (Grecia) (**Yuan et al. 2016**). Dado que las antiguas civilizaciones dependían fuertemente de su entorno para subsistir, hay una marcada relación entre esta estrategia y el aislamiento de productos naturales para el descubrimiento de fármacos.

2.2.3. Aislamiento de productos naturales

La naturaleza es una fuente inagotable de nuevos compuestos tipo droga, cuya diversidad estructural supera cualquier librería sintética conocida. Estas moléculas han sido optimizadas a través de la evolución natural, para interaccionar con macromoléculas biológicas (**Hong 2011**). La medicina moderna se ha enfocado en la identificación, aislamiento y purificación de los principios activos presentes en extractos naturales con propiedades terapéuticas, con el objetivo de desarrollar fármacos más seguros y eficaces. Los avances en áreas como la genética y la biología molecular, permiten hoy en día producir algunos de estos compuestos a gran escala utilizando bacterias o levaduras modificadas genéticamente (**Newman y Cragg 2012**). Alternativamente, la semisíntesis y derivatización de productos naturales constituye una estrategia ampliamente utilizada para generar nuevos compuestos activos (**Majhi y D. Das 2021**).

Un ejemplo de especial relevancia en el campo de la Epilepsia es el aceite de cannabis. Los primeros registros de la utilización del cannabis en el tratamiento de esta enfermedad datan de miles de años atrás (**Russo et al. 2008**). Los extractos de cannabis poseen una gran cantidad de compuestos activos denominados cannabinoi-

des, dentro de los cuales se ha logrado identificar al Tetrahidrocannabinol (THC) como el responsable de los efectos psicoactivos y no deseados del cannabis (**Murray et al. 2016**), mientras que el Cannabidiol (CBD) y sus compuestos relacionados serían los responsables de los efectos terapéuticos (**Rong et al. 2017**). Los extractos enriquecidos en CBD han demostrado ser efectivos en distintos tipos de epilepsias, principalmente en niños y en pacientes que presentan refractariedad a los fármacos antiepilépticos convencionales (**Gaston y Szaflarski 2018; Perucca 2017; Zaheer et al. 2018**).

2.2.4. Cribado de alta eficiencia

El cribado de alta eficiencia (HTS, del inglés High-throughput Screening), también conocido como “*screening* de alta *performance*” o “cribado de alta *performance*”, permite la evaluación experimental de grandes librerías de compuestos frente a un ensayo particular de manera veloz y eficaz (**Macarron et al. 2011**). Esta metodología combina varias tecnologías como la robótica, la miniaturización de ensayos, los detectores de alta sensibilidad y el manejo de líquidos, entre otras. El ensayo puede realizarse sobre un blanco molecular específico o sobre sistemas complejos, en lo que se denomina *screening* fenotípico. Estos sistemas complejos pueden ser organoides, células completas u organelas, en orden decreciente de complejidad (**Hou et al. 2018**). Analizando los cambios en el fenotipo del sistema en estudio, es posible evaluar el efecto de las drogas sobre el mismo. El cribado de alto contenido (HCS, del inglés high content screening), es una variante del *screening* fenotípico que permite identificar *hits* que generan cambios espacio-temporales en el sistema, detectables mediante técnicas de imagen (morfología celular, estructura subcelular, etc.) (**Zanella, Lorens y Link 2010**). Esto es posible gracias a las nuevas tecnologías en detección de imágenes y también a los avances en áreas de la informática como la visión por computadora y el *deep learning*, que permiten el procesamiento y análisis de las imágenes obtenidas (*image segmentation*) (**Kraus y Frey 2016; Sommer et al. 2017**).

Cabe destacar que, si bien el tiempo efectivo que se tarda en evaluar los compuestos es relativamente corto, la etapa previa de diseño del *screening* y el posterior análisis de grandes cantidades de datos no es una tarea sencilla.

2.3. Metodologías *in silico*

En los últimos años, las metodologías computacionales o *in silico* han cobrado un rol fundamental en las distintas etapas del descubrimiento de fármacos. Algunos ejemplos incluyen la identificación y validación de nuevos *targets* (Dezső y Ceccarelli 2020), la identificación de nuevos *hits* (X. C. Yan et al. 2020) y su posterior optimización (Cozza 2017) o incluso la selección de pacientes para ensayos clínicos (Woo 2019). A continuación, se resumen algunas de las técnicas computacionales más utilizadas en las etapas tempranas del descubrimiento de fármacos, haciendo hincapié en las que fueron aplicadas en la presente tesis de investigación.

2.3.1. Cribado virtual

El cribado o tamizado virtual (VS, del inglés Virtual Screening), consiste en la aplicación de una o más técnicas computacionales para cribar una base de datos, con el objetivo de identificar nuevos *hits* o compuestos activos. Esta estrategia ha ganado gran visibilidad en los últimos años, ya que como puede observarse en la Figura 2.2, el número de publicaciones científicas que contienen la palabra “*virtual screening*” en su título o *abstract*, indexadas en el buscador PubMed, ha tenido un gran crecimiento durante los últimos 20 años.

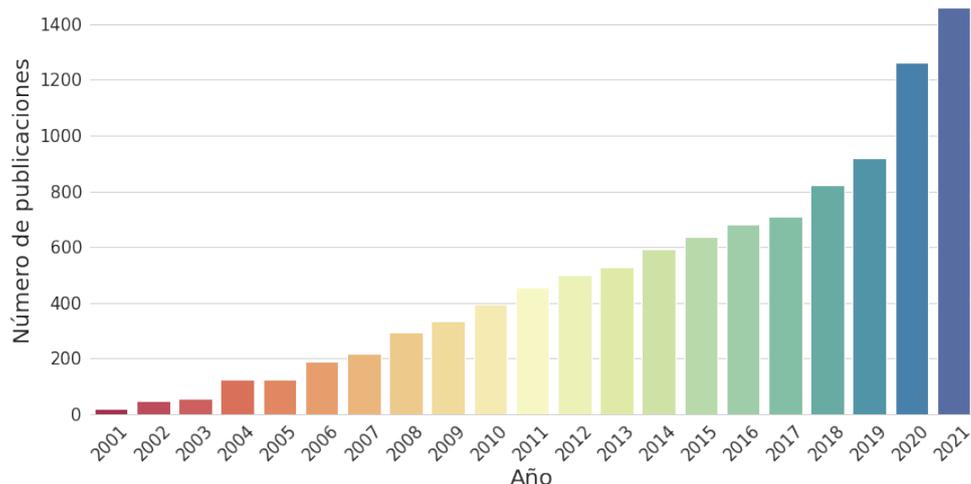


Figura 2.2: Número de publicaciones científicas que contienen la palabra "virtual screening" en su título o *abstract*, indexadas en el buscador PubMed. La búsqueda se acotó a los últimos 20 años. Último acceso 27/05/2022.

El VS presupone varias ventajas con respecto a las técnicas convencionales de HTS, tales como:

Menor inversión

El HTS requiere de equipamiento especializado altamente complejo y costoso. Además, se utilizan grandes cantidades de reactivos e insumos para realizar los ensayos. Por su parte, el VS solo requiere de equipamiento computacional o capacidad de cálculo, la cual puede incluso adquirirse a proveedores especializados (*cloud computing*).

Mayor eficiencia

Mientras que el *hit-rate* (proporción de activos identificados en relación a la cantidad de compuestos ensayados) en un *screening* convencional varía entre 0.01 % y 0.14 %, para el caso del VS este ratio varía entre 1 % y 40 %. Esto supone una reducción en los costos y el tiempo asociados al proceso de identificación de nuevos *hits* (Babaoglu et al. 2008; Doman et al. 2002; Ferreira et al. 2010).

Mayor exploración del espacio químico

Según Bohacek (**Bohacek, McMartin y Guida 1996**), el espacio químico que comprende a las moléculas tipo droga contiene aproximadamente 10^{63} moléculas. Este número es inabarcable para cualquier técnica de cribado *in vitro* disponible en la actualidad. Incluso cuando fuera posible evaluar ese número de compuestos experimentalmente, la etapa previa de síntesis y purificación de los mismos sería una tarea titánica y extremadamente costosa. Las técnicas de cribado *in silico* permiten evaluar compuestos potenciales o teóricos, incluso antes de ser sintetizados. Considerando que, aproximadamente cada 2 años se duplica el número de transistores en un microprocesador (de acuerdo a la Ley de Moor (**Liddle 2006**), de carácter empírico pero perfectamente válida hasta la actualidad), los métodos *in silico* serán cada vez más accesibles y poderosos. Recientemente, Irwin y colaboradores reportaron un VS de 170 millones de compuestos, el cual demoró aproximadamente un día y medio de cálculo en un *cluster* de 1.000 núcleos (**Lyu et al. 2019**).

Para realizar el VS pueden aplicarse una o varias metodologías computacionales, las cuales pueden clasificarse en dos grandes categorías: las basadas en la estructura de los ligandos y las basadas en la estructura del receptor.

2.3.2. Metodologías basadas en la estructura de los ligandos

Como su nombre lo indica, este conjunto de metodologías requiere conocer la estructura de una o varias moléculas que presenten la actividad biológica de interés. Partiendo del supuesto de que moléculas similares exhibirán propiedades de unión similares con respecto a un determinado blanco molecular, o desencadenarán respuestas biológicas similares, es posible identificar las características electrónico-estructurales comunes a uno o varios ligandos activos, responsables de la actividad biológica observada (**Macalino et al. 2015**).

2.3.2.1. Búsqueda por similitud

La búsqueda por similitud consiste en cuantificar el grado de similitud entre una o varias estructuras que presentan la actividad biológica deseada, y los diferentes compuestos que integran la base de datos sometida al tamizado (X. **Yan et al.**

2016). Existen diversas formas de evaluar el grado de similitud entre dos moléculas, siendo el cálculo de *fingerprints* moleculares y su posterior comparación mediante algún coeficiente de similitud la más utilizada. Algunos ejemplos de coeficientes de similitud son el coeficiente de Tanimoto, Jaccard, Dice, Cosine, etc. (Bero et al. 2017).

2.3.2.2. Farmacóforos basados en los ligandos

La International Union of Pure and Applied Chemistry (IUPAC) define un farmacóforo como “un conjunto de rasgos estéricos y electrónicos necesarios para asegurar las óptimas interacciones supramoleculares con un blanco biológico específico y desencadenar (o bloquear) su respuesta biológica” (Wermuth et al. 1998). Es decir, un farmacóforo representa el arreglo espacial de las características moleculares que deben cumplir los ligandos activos frente a un *target* específico. Estas características comúnmente se refieren a grupos capaces de establecer interacciones del tipo electrostáticas, aromáticas o hidrofóbicas.

La construcción de un farmacóforo basado en los ligandos requiere conocer la pose de unión de cada molécula activa en la diana biológica en estudio. Esta información tridimensional puede extraerse de los ligandos que han sido cristalizados con el *target* en cuestión, si existieran. Alternativamente, es posible aproximar esta pose con una conformación de mínima energía, calculada de manera teórica. Sin embargo, esta última puede no coincidir con la conformación biológicamente activa. Las conformaciones consideradas activas son superpuestas en el espacio, y se extraen las características comunes para generar el farmacóforo.

Para realizar un cribado con el modelo farmacofórico obtenido, los compuestos de la base de datos deben alinearse uno a uno con el mismo. Evaluando el grado de superposición o *matching*, es posible predecir si una molécula presentará o no la actividad biológica que se está modelando.

También es posible derivar modelos farmacofóricos desde la estructura del receptor, considerando el arreglo tridimensional de los residuos en el sitio activo de la diana biológica (Udayakumar et al. 2013).

2.3.2.3. Modelado de relaciones estructura-actividad (QSAR)

Los modelos de relación cuantitativa entre estructura y actividad (QSAR, del inglés Quantitative Structure-Activity Relationship) son modelos matemáticos que relacionan la estructura química de las moléculas con alguna propiedad biológica asociada a las mismas, como por ejemplo IC₅₀, EC₅₀, K_i, Solubilidad, etc. (Cherkasov et al. 2014). Estos modelos pueden utilizarse para predecir la actividad biológica de interés en otros compuestos, por lo que resultan muy útiles en las etapas pre-clínicas del descubrimiento de fármacos, particularmente en la identificación de nuevos *hits* y su posterior optimización (Neves et al. 2018).

El modelado QSAR puede describirse como un proceso que involucra 6 etapas principales, resumidas en el esquema e la Figura 2.3:



Figura 2.3: Etapas principales del desarrollo y la validación de modelos QSAR.

Compilado de un conjunto de datos

Este es un paso crítico en la construcción de cualquier modelo predictivo, ya que la calidad de los datos utilizados para generar el modelo será el factor determinante de la capacidad predictiva del mismo (Roy, Kar y R. N. Das 2015a). Es necesario compilar un conjunto de datos (comúnmente denominado *dataset*) que incluya la estructura química de cada compuesto y la actividad o propiedad biológica asociada que se desea modelar. En el caso de los modelos de regresión, se necesita un valor numérico de actividad biológica asociado a cada molécula, mientras que para el caso de modelos clasificadores la variable independiente será categórica (por ejemplo, activo/inactivo).

La diversidad química de las moléculas del conjunto de datos es también un aspecto fundamental a tener en cuenta, ya que de ella dependerá la aplicabilidad de los modelos generados. Si uno o más quimiotipos se encuentran sobrerrepresentados en el conjunto de datos, obtendremos modelos sesgados, y luego sus predicciones sólo serán confiables para esa región del espacio químico. Si lo que deseamos es obtener

modelos QSAR para luego optimizar compuestos que presentan este quimiotipo en particular, el sesgo o *bias* quizás no sea un problema mayor. Sin embargo, para aplicaciones del tipo de VS e identificación de nuevos *hits*, es deseable que las predicciones de los modelos sean confiables para una amplia región del espacio químico.

Partición del conjunto de datos

Este paso no es indispensable, y puede omitirse cuando el tamaño del *dataset* es muy pequeño, realizando así solo una validación interna ([Hawkins, Basak y Mills 2003](#)). Sin embargo, esta partición nos asegura contar con uno o más conjuntos de datos externos para validar los modelos y es recomendable realizarla siempre que sea posible. Idealmente, el conjunto inicial de datos debe dividirse en tres grupos ([Bonnin 2017](#)):

Conjunto de entrenamiento: Serán los datos que intentaremos ajustar mediante algún algoritmo de aprendizaje automático (ML, de inglés Machine Learning). Este conjunto representa aproximadamente el 70 % de los datos.

Conjunto de validación: Este conjunto de datos tiene varias utilidades posibles y generalmente constituye un 10 % de los datos. Puede ser utilizado para evaluar el grado de ajuste del modelo a lo largo del proceso de entrenamiento, y así saber cuándo finalizar el entrenamiento. También puede utilizarse para optimizar los hiperparámetros del algoritmo, mediante algún método de optimización apropiado. Otra utilidad es comparar la *performance* de distintos modelos construidos con el mismo conjunto de entrenamiento. Sin embargo, en el campo de la quimioinformática es común contar con conjuntos de datos muy pequeños, por lo cual suele obviarse la partición de validación y en su lugar se utilizan otras técnicas con dichos propósitos, como la validación cruzada.

Conjunto de testeo: Este conjunto se utilizará pura y exclusivamente para evaluar la *performance* del modelo final. Debido a que estos datos nunca fueron parte del proceso de entrenamiento, son datos “no vistos” por el modelo, y por lo tanto permiten evaluar la capacidad de generalizar del mismo. Usualmente un 20 % de los datos son reservados con este fin.

La partición de los datos puede realizarse de forma aleatoria o racional. Esta

última opción a menudo comprende una etapa previa de agrupamiento o *clustering*, donde cada compuesto es asignado a un *cluster* particular en base a su estructura química. Luego, los datos son particionados de manera estratificada, es decir conservando la misma proporción inicial de *clusters* en cada partición, lo que permite explotar al máximo la diversidad química del conjunto inicial de datos.

Cálculo de descriptores moleculares

Los descriptores moleculares son variables numéricas que codifican la información química contenida en una molécula. Es decir, mediante la aplicación de procedimientos lógico-matemáticos sobre una representación de la molécula, es posible extraer la información química y expresarla como variables numéricas (**Todeschini** y **Consonni** 2008). De acuerdo a las representaciones utilizadas se originarán distintos descriptores moleculares, que pueden clasificarse en categorías según sus dimensionalidad:

0-Dimensionales: Son aquellos cuyo cálculo se basa exclusivamente en la fórmula química de una molécula (**Balaban** 1998). Debido a que existen muchos compuestos que presentan la misma fórmula química, estos descriptores son altamente degenerados y su contenido de información es bajo.

1-Dimensionales: Su cálculo se basa en la presencia o ausencia de ciertas subestructuras dentro de una molécula, pero sin considerar cómo estas subestructuras se interconectan entre sí (**Roy, Kar** y **R. N. Das** 2015b).

2-Dimensionales: Dentro de este grupo se encuentran los descriptores del tipo topológico, los cuales consideran la conectividad de los átomos que conforman una molécula teniendo en cuenta la presencia y naturaleza de los enlaces químicos. Su cálculo se deriva de la teoría de grafos, donde cada molécula es representada como un grafo, cuyos vértices son los átomos y las aristas son los enlaces (**Ivanciuc** 2013).

3-Dimensionales: A diferencia de los descriptores 0-2D, estos descriptores son dependientes de la conformación, es decir que su valor para un ligando determinado depende del conformero que se esté considerando. Esta información tridimensional puede obtenerse experimentalmente o aproximarse de la misma forma que para la construcción de farmacóforos.

4-dimensionales: Estos descriptores son similares a los 3D, con la diferencia que es posible considerar distintos conformeros y protómeros para cada ligando (Vedani y Dobler 2000).

Cabe destacar que, a medida que avanzamos en la dimensionalidad de los descriptores, su cálculo se vuelve computacionalmente más costoso, limitando así su aplicación a grandes conjuntos de moléculas.

El desarrollo y la validación de nuevos descriptores moleculares es un área en constante crecimiento dentro de la quimioinformática, y existen aún más categorías de descriptores que las mencionadas anteriormente (Lo et al. 2018; J. Wang et al. 2021). Sin embargo, una descripción exhaustiva de las mismas excede el objetivo de esta tesis.

Selección de descriptores

Esta etapa tiene como objetivo eliminar aquellos descriptores con bajo contenido de información, es decir que son poco relevantes para predecir la propiedad biológica de interés. Estos a menudo suelen ‘confundir’ al algoritmo disminuyendo así su capacidad predictiva, a la vez que incrementan el costo computacional asociado al proceso de entrenamiento. Dependiendo del algoritmo en cuestión, el tamaño del conjunto de entrenamiento y la relación descriptores/número de ejemplos (*p:n ratio*), este efecto puede ser más o menos importante (Goodarzi, Dejaegher y Heyden 2012).

A las variables independientes que se utilizan para predecir la variable dependiente se las denomina comúnmente predictores o *features*. En el contexto de la quimioinformática, los predictores que se utilizan suelen ser descriptores moleculares, aunque estos términos no son equivalentes, ya que existen otras alternativas para describir a las moléculas, como por ejemplo los *fingerprints* moleculares.

Los métodos de **selección de predictores** se encuentran englobados dentro de un conjunto más amplio de métodos al que se denomina **reducción de dimensionalidad**, y cuyo objetivo es disminuir el número de variables independientes con las que se entrenará el modelo (Figura 2.4). Este conjunto también comprende a los métodos de **extracción o proyección de predictores**, los cuales aplican

transformaciones matemáticas para reducir en número de variables independientes, reteniendo la mayor cantidad de información posible durante el proceso. Según el tipo de transformación que realizan sobre los datos, estos pueden clasificarse en **lineales** o **no lineales**. A su vez, según incorporen o no información sobre la variable dependiente durante el proceso, estos métodos pueden ser **supervisados** o **no supervisados**, respectivamente. Los métodos **semi-supervisados** representan una categoría intermedia entre las anteriormente nombradas, donde solo se cuenta con información sobre la variable dependiente para algunos ejemplos en el conjunto de datos.

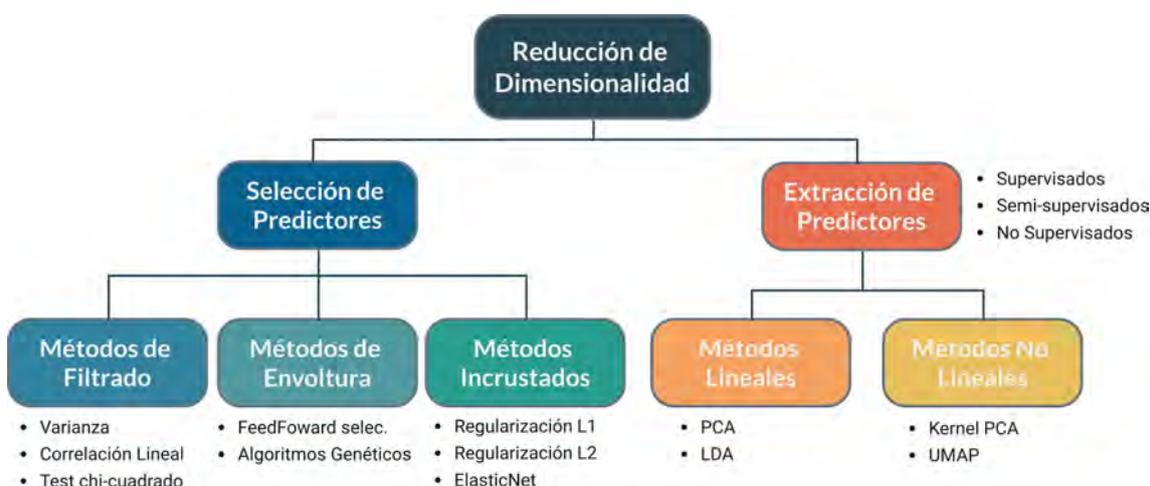


Figura 2.4: Clasificación de los métodos de reducción de dimensionalidad. PCA: Principal Component Analysis; LDA: Linear Discriminant Analysis; UMAP: Uniform Manifold Approximation and Projection.

Dentro de las metodologías de **selección de predictores**, pueden distinguirse al menos tres categorías principales:

Métodos de Filtrado (*Filter methods*): Los métodos de filtrado seleccionan predictores basados en alguna propiedad intrínseca de los mismos, la cual se mide mediante parámetros estadísticos. Uno de los más utilizados es el método de varianza, mediante el cual se excluyen los predictores invariantes o constantes a lo largo de todo el conjunto de entrenamiento (aunque puede aplicarse otro umbral de varianza distinto de cero).

Métodos de Envoltura (*Wrapper methods*): Este conjunto de métodos “envuelve” distintos subconjuntos de predictores alrededor de un ciclo completo de entrena-

miento y evaluación del modelo. Es decir, sobre el conjunto inicial de predictores, se selecciona un subconjunto de los mismos para entrenar el algoritmo y posteriormente se evalúa la *performance* del modelo obtenido. El objetivo es encontrar un subconjunto de predictores que minimice alguna función de pérdida o *loss function*, que debe definirse de antemano. Los distintos métodos dentro de esta categoría difieren en cómo se seleccionan estos subconjuntos de predictores.

Métodos Incrustados (*Embedded methods*): Al igual que en los métodos de envoltura, este conjunto de métodos también minimiza una función de pérdida. Sin embargo, se los denomina incrustados ya que la selección de predictores es intrínseca del algoritmo y por lo tanto se realiza *on-the-fly*, es decir, durante el proceso de entrenamiento del modelo.

Aplicando el concepto de la navaja de Ockham o principio de parsimonia, los modelos más parsimoniosos o simples son preferidos por sobre los más complejos. Se consideran modelos simples a aquellos que incorporan un menor número de predictores o que tienen un menor número de hiperparámetros. Esto tiene implicancias directas sobre la interpretación de los modelos. Un modelo fácilmente interpretable, es aquel que podemos comprender mejor los seres humanos. Sin embargo, este puede no coincidir con el mejor modelo en términos de capacidad predictiva.

Entrenamiento del modelo

El objetivo de esta etapa es encontrar un modelo matemático que describa cómo las variables independientes (descriptores) pueden relacionarse con la variable dependiente (propiedad biológica) que se desea modelar. La forma matemática que adopta el modelo es muy variable, dependiendo del algoritmo que se aplique. Por ejemplo, para caso de un modelo de regresión lineal con un único predictor, será una ecuación del tipo:

$$y = a \times x + b \tag{2.1}$$

Ecuación que describe una recta. Donde y es la variable dependiente que se desea predecir, x es la variable independiente o predictor, a es el coeficiente o pendiente de la recta y b es la ordenada al origen. Mediante el entrenamiento del modelo, se intenta aprender o ajustar los coeficientes a y b que completan esta ecuación, minimizando alguna función de pérdida.

Validación del modelo

La validación de un modelo es una etapa fundamental, ya que nos permite evaluar ciertas características que hacen a la calidad del mismo. En general, se suele validar la capacidad predictiva de un modelo, pero también es posible evaluar otros parámetros como la robustez o la estabilidad ([Tropsha 2010](#)). Las estrategias de validación pueden clasificarse en internas o externas, según se apliquen sobre el conjunto de entrenamiento o sobre un conjunto de testeo, respectivamente.

Validación interna

- **Validación cruzada de K iteraciones (K-fold CV)**

En esta técnica de validación cruzada, el conjunto de datos se divide en K particiones iguales. Una de estas particiones se utiliza como conjunto de prueba o testeo, mientras que las restantes (K-1) se utilizan para entrenar el modelo. Este proceso constituye una iteración y es repetido K veces. En cada iteración es posible calcular alguna métrica de interés relacionada con la *performance* del modelo para luego, al concluir las K iteraciones, reportar el valor medio y su desvío estándar asociado. El valor de K generalmente varía entre 3 y 10, ya que existe una relación de compromiso entre el sesgo o *bias* y la varianza de los mismos, para distintos valores de K ([Fushiki 2011](#)).

- **Aleatorización de Fisher (y-Randomization)**

Este método consiste en aleatorizar la variable dependiente para el conjunto de entrenamiento, y luego ajustar el modelo con esos datos. Al hacer esto, se elimina la correlación que existe entre las variables independientes (descriptores) y la variable dependiente (actividad biológica). Por lo tanto, el modelo que se obtiene debe presentar un desempeño muy pobre; cercano al azar para el caso de modelos clasificadores, o con correlaciones prácticamente nulas para los modelos de regresión ([Wold, Eriksson y Clementi 2008](#)).

- **Bootstrapping**

Esta técnica de re-muestreo permite hacer inferencia estadística sobre la distribución de una población, a partir de una única muestra ([Mooney et al. 1993](#)). Considerando un conjunto de datos de tamaño N , se toman B mues-

tras con reemplazo de tamaño n sobre el conjunto inicial. Para cada muestra podemos calcular distintas métricas, y luego analizando la distribución de \mathbf{B} es posible inferir parámetros estadísticos tales como la media, el desvío estándar o calcular intervalos de confianza.

Debido a que el muestreo se realiza con reemplazo, para cada muestra \mathbf{B} existe una cierta proporción de los datos que no fueron incluidos en la misma, a los que comúnmente se denomina *out-of-bag*. Si se entrena un algoritmo con los datos en la muestra \mathbf{B} , es posible estimar la capacidad predictiva del mismo utilizando los datos remanentes en el *out-of-bag*, ya que estos datos no fueron utilizados para entrenar el modelo. Esta técnica es particularmente útil cuando se cuenta con muy pocos datos para generar el modelo, y por lo tanto no es posible reservar un conjunto de datos externo para testeo.

Validación externa

La validación externa permite evaluar la capacidad predictiva de un modelo, es decir su capacidad de generalizar sobre nuevos datos. Esta consiste aplicar los modelos para predecir un conjunto de datos que no haya sido utilizado en la generación de los mismos (**Chirico y Gramatica 2011**). Según la composición y el tamaño de este conjunto externo, es posible validar distintos aspectos relacionados al poder predictivo de los modelos. Para los dos blancos moleculares seleccionados en esta tesis, se lograron compilar *datasets* relativamente grandes, por lo tanto en ambos casos se reservaron particiones de testeo que se utilizaron durante la validación externa.

A su vez, se recurrió también al VS retrospectivo como estrategia alternativa de validación externa. Esta técnica consiste en aplicar los modelos sobre quimiotecas o bases de datos con una baja proporción de compuestos activos conocidos, dispersados en una gran cantidad de inactivos o *decoys*. De este modo se emula una situación que se asemeja a un caso de aplicación real, donde la base de datos a cribar de manera prospectiva tendrá un rendimiento de activos muy bajo. Esto a su vez permite mitigar el “efecto de saturación” que ocurre cuando la proporción de activos es alta y estos saturan los primeros lugares del ranking, limitando así la sensibilidad de las métricas de enriquecimiento para reflejar diferencias entre los distintos modelos (**Truchon y Bayly 2007**).

Con el objetivo de evaluar el desempeño de los modelos en campañas de VS retrospectivo, es posible recurrir a la herramienta DUD-E (Database of Useful Decoys - Enhanced) (Mysinger et al. 2012) para generar bases de datos simuladas. Partiendo de uno o varios compuestos activos denominados “semilla”, este servidor permite generar un conjunto de presuntos compuestos inactivos o *decoys*, que son fisicoquímicamente similares pero topológicamente diferentes a las semillas utilizadas. Para ello, el algoritmo compara estas semillas con las moléculas de la base de datos ZINC (Sterling e Irwin 2015), considerando 6 propiedades fisicoquímicas: peso molecular, una variante del coeficiente de partición octanol-agua (mLogP), el número de enlaces rotables, la carga molecular neta, y la cantidad de dadores y aceptores de puentes de hidrógeno. De este modo, obtiene entre 3000 y 6000 *decoys* por semilla, de los cuales se retienen sólo aquellos que presentan una similitud menor a 0,25 con la semilla, calculada mediante *fingerprints* ECFP4 (Rogers y Hahn 2010). Si existiesen más de 50 compuestos, se retienen sólo los 50 más disímiles.

Dominio de aplicación

El dominio de aplicación (AD, del inglés Applicability Domain), puede entenderse como una región teórica del espacio químico para la cual las predicciones de un modelo QSAR serán confiables, y estará definido por la naturaleza de los compuestos con los que se entrenó el modelo (*training set*). Si el DA de un modelo es limitado, su utilidad también lo será, ya que las predicciones del mismo sólo serán confiables para un grupo reducido de moléculas (Roy, Kar y Ambure 2015).

Existen diversas metodologías para estimar el DA de un modelo, las cuales difieren principalmente en cómo se define este subespacio químico. Dentro de ellas podemos mencionar a los métodos basados en rangos, métodos geométricos, métodos basados en distancias y los basados en distribución de densidad de probabilidad (Sahigara et al. 2012).

2.3.2.4. Estrategia desarrollada en esta tesis

La estrategia de modelado QSAR que se utilizó, desarrollada durante la presente tesis de investigación, consiste de varias etapas, las cuales se resumen en el esquema de la Figura 2.5. Esta metodología se aplicó en forma independiente para los dos

blancos moleculares estudiados, el canal TRPV1 y el NaV1.2.

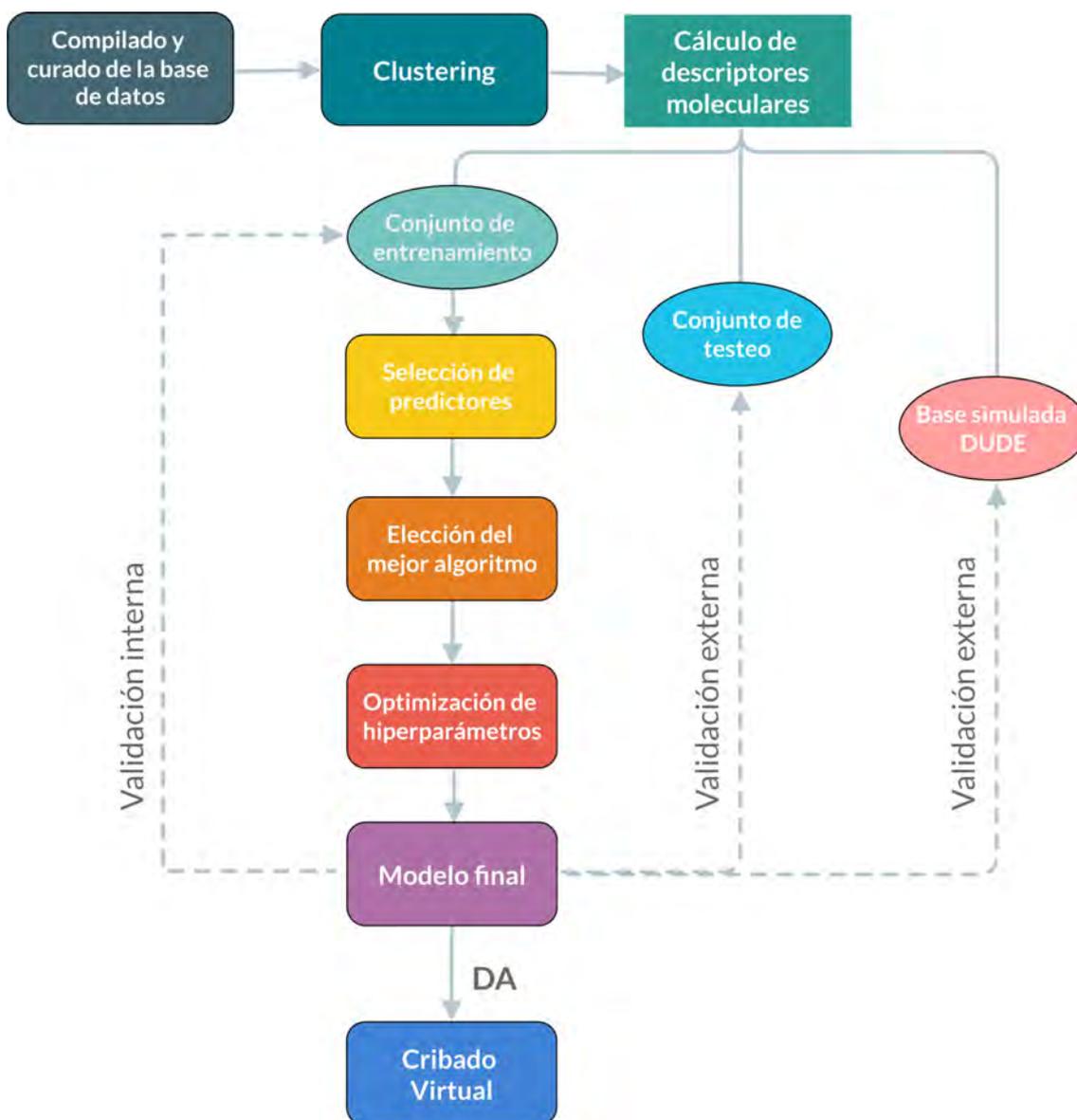


Figura 2.5: Flujo de trabajo general que se utilizó para el desarrollo de los modelos QSAR de regresión y clasificación. DA: Dominio de aplicación.

Compilado y curado de la base de datos

Como primer paso, se procedió a compilar un conjunto de datos de compuestos evaluados experimentalmente frente a cada blanco molecular, los que serán utilizados para la construcción y validación de los modelos predictivos. Para ello, se recurrió a la base de datos ChEMBL (Davies et al. 2015; Mendez et al. 2019), que contiene datos de actividad biológica pre-curados de la literatura científica, para una gran cantidad de blancos moleculares. Los mismos fueron curados teniendo en cuenta las características particulares de cada diana biológica y de los ensayos de actividad pertinentes.

Las moléculas en código SMILES (del inglés, Simplified Molecular Input Line Entry Specification) fueron estandarizadas eliminando solventes, contraiones y metales; explicitando hidrógenos y reparando colisiones entre átomos o enlaces. Las estructuras fueron convertidas desde el formato SMILES al formato InChI, para luego agruparlas por su código InChI y eliminar así los compuestos repetidos. Esto se realizó mediante un *workflow* propio desarrollado en la plataforma KNIME (**Berthold et al. 2008**).

Es importante destacar que, errores en la representación estructural de las moléculas se traducen en un cálculo incorrecto de los descriptores moleculares, lo que conlleva a una interpretación incorrecta de las potenciales relaciones estructura-actividad (**Fourches, Muratov y Tropsha 2010**).

Clustering

En esta etapa, las moléculas en el conjunto inicial de datos fueron agrupadas según su similitud química/estructural, para luego ser muestreadas de manera estratificada a la hora de generar las distintas particiones de entrenamiento y testeo. Se ha demostrado que, un muestreo racional de los datos en subconjuntos representativos, conduce a modelos con mejor capacidad predictiva (**Leonard y Roy 2006**; **Martin et al. 2012**).

La metodología de *clustering* utilizada, denominada Silhouette Optimized Molecular Clustering (SOMoC), fue desarrollada durante la presente tesis de investigación y se describe en detalle en la sección suplementaria Algoritmos - [C.1](#).

Cálculo de descriptores moleculares

Para el cálculo de los descriptores moleculares se utilizó el software Mordred (**Moriwaki et al. 2018**). Implementado como una librería de Python, este software gratuito y de código abierto permite el cálculo de 1613 descriptores 0-2D. El programa asigna el modelo de hidrógenos y aromaticidad para cada compuesto según los requerimientos de cada descriptor en particular, lo cual resulta muy conveniente.

Selección de predictores

El proceso de selección de predictores consistió de 3 etapas, en cada una de las

cuales se aplicó una de las estrategias antes mencionadas:

1. Método de filtrado (Filter)

En esta etapa, los descriptores cuyo cálculo falló para los compuestos del conjunto de entrenamiento fueron eliminados de este conjunto y del de testeo. Posteriormente se aplicaron dos filtros. El primero es un filtro de varianza, donde se eliminaron los descriptores con varianza menor o igual que 0.01. El segundo es un filtro de correlación, mediante el cual se eliminaron descriptores cuyo coeficiente de determinación (R^2) para una correlación lineal es mayor o igual a 0.9.

2. Método de envoltura (Wrapper)

En esta etapa se aplicó una metaheurística del tipo evolutiva, combinada con una estrategia de validación cruzada en el conjunto de entrenamiento. Aplicando 10-Fold CV en el conjunto de entrenamiento, sobre la partición o *fold* mayoritario (90 % de los datos) se utilizó un algoritmo genético (GA, del inglés Genetic Algorithm) para seleccionar subconjuntos de predictores con los cuales entrenar un algoritmo de ML, el cual se utilizó luego para predecir los compuestos remanentes en la partición minoritaria (10 % de los datos). El número de individuos se fijó en 5000, los cuales se evolucionaron durante 1000 generaciones. El criterio de finalización fue que el error en la partición de testeo se mantenga constante durante 5 generaciones. El fundamento de los GA se describe en profundidad en la sección suplementaria Algoritmos [C.2.2](#).

En el esquema de la Figura [2.6](#) se ilustra una iteración completa del método propuesto. El algoritmo de ML y la métrica de pérdida se seleccionaron en base al tipo de modelo (clasificación o regresión). En cada iteración, los descriptores fueron estandarizados utilizando una normalización del tipo gaussiana o *z-score* (centrados en la media y con desviación estándar unitaria). El tamaño del subconjunto de descriptores (n) se ajustó en cada caso según el tamaño del conjunto de entrenamiento, manteniendo aproximadamente una relación ejemplos:predictores de 10:1.

Debe tenerse en cuenta que, si solo consideramos la mejor solución de todas al finalizar las 10 iteraciones, esa solución corresponde a una distribución particular de los ejemplos, con lo cual estaríamos ignorando la información que

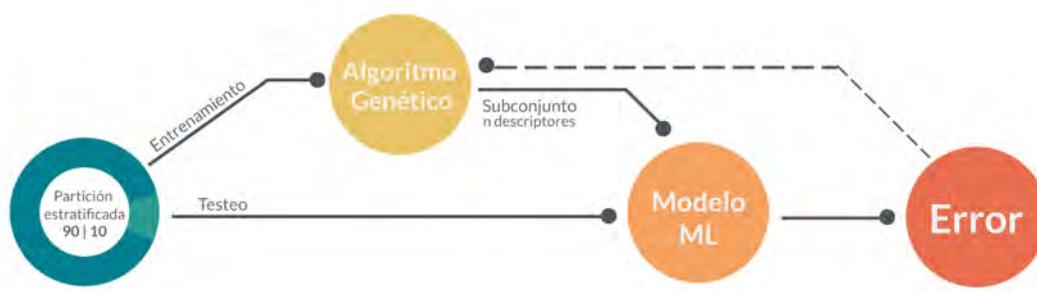


Figura 2.6: Esquematización de una iteración completa de la etapa de selección de descriptores mediante el algoritmo genético.

aportan todas las demás particiones. Por tal motivo, se agruparon las 100 mejores soluciones obtenidas en cada una de las 10 iteraciones, para luego generar un ranking de importancia de variables considerando la frecuencia de aparición de cada descriptor en el conjunto de 1000 soluciones. Este ranking, se utilizó como criterio para elegir los descriptores con los que finalmente se entrenaron los modelos, seleccionando un número de predictores acorde con el tamaño del conjunto de entrenamiento.

3. Método incrustado (Embedded)

En esta etapa, la selección de los descriptores es llevada a cabo por el mismo algoritmo de ML mediante estrategias internas de regularización, como por ejemplo las penalidades del tipo L1 o la selección interna que realizan los árboles de decisión. Por lo tanto, esta etapa de selección solo se llevó a cabo en los algoritmos que así lo admiten.

Elección del mejor algoritmo

Existe una amplia gama de algoritmos de ML que pueden aplicarse a problemas de regresión o clasificación, cada uno de ellos con sus ventajas y desventajas particulares. Con el objetivo de hallar el algoritmo que mejor ajuste nuestros datos, se consideraron diversas opciones implementadas en la librería de Python Scikit-learn (Pedregosa et al. 2011). Los algoritmos evaluados para cada blanco molecular se detallan en los capítulos correspondientes. En todos los casos, se utilizó la estrategia de validación interna K-fold CV para evaluar el desempeño de los algoritmos, utilizando los hiperparámetros por defecto para cada uno de ellos.

Optimización de los hiperparámetros

La *performance* de cualquier algoritmo se verá afectada en gran medida por los hiperparámetros del mismo. En líneas generales, los hiperparámetros por defecto funcionan bien, pero cada conjunto de datos a modelar tiene sus particularidades, por lo que siempre es recomendable hacer un *fine-tuning* de los mismos, con el objetivo de optimizar su desempeño. Cuanto mayor sea el número de hiperparámetros a optimizar, y mayor sea el tamaño del conjunto de datos, el proceso de optimización se vuelve más complejo y costoso computacionalmente (L. Yang y Shami 2020; Yu y Zhu 2020). Por lo tanto, solo se avanzó en la optimización de los algoritmos que tuvieron una buena *performance* en la etapa anterior, utilizando los valores por defecto.

Los hiperparámetros de cada algoritmo se optimizaron mediante optimización bayesiana, una técnica ampliamente utilizada en el campo del ML (Wu et al. 2019), que resulta particularmente útil para funciones cuya evaluación implica un alto costo computacional y/o económico (Hutter, Hoos y Leyton-Brown 2013). El fundamento de esta metodología se describe en el material suplementario Algoritmos - Sección C.2.1. En particular se utilizó la librería de Python HyperOpt (Bergstra, Yamins y Cox 2013), la cual emplea el Tree-of-Parzen-Estimators (TPE) como método de búsqueda. Los parámetros a optimizar y los rangos de búsqueda fueron cuidadosamente seleccionados para cada algoritmo. Para cada combinación de hiperparámetros muestreada se realizó 10-Fold CV en el conjunto de entrenamiento, registrando la media de las 10 iteraciones.

Validación del modelo

Finalmente, los mejores modelos obtenidos para cada blanco molecular fueron sometidos a las diferentes validaciones *in silico* descritas anteriormente.

La validación interna consistió en 100 iteraciones de 10-fold CV, a la vez que se realizaron 100 iteraciones de *y-randomization*. Por su parte, la validación externa consistió en predecir los compuestos remanentes en el conjunto de testeo, y también en un *screening* retrospectivo utilizando la base simulada DUD-E.

Como se describió anteriormente, el DA constituye una herramienta fundamen-

tal para estimar el grado de interpolación que el modelo está haciendo en cada una de sus predicciones, y determinar así si estas serán confiables o no (**Sahigara et al. 2012**). En este caso, se utilizó un método simple basado en distancias euclidianas. Para cada compuesto de prueba (*query*) se calcula la distancia a su vecino más cercano en el conjunto de entrenamiento y se la compara con un umbral de dominio de aplicación (APD) predefinido. Si la distancia supera este umbral, la predicción se considera poco confiable (S. **Zhang et al. 2006**). El APD puede definirse como:

$$\text{APD} = d + Z\sigma$$

Inicialmente, se calcula la media de todas las distancias entre pares de moléculas del conjunto de entrenamiento. Aquellas distancias entre pares que sean mayores a la media son descartadas, y sobre el conjunto remanente se re-calcula la media (d) y su desvío estándar (σ). El parámetro Z es un valor de corte empírico, cuyo valor por defecto es 0.5. Debido a que la distancia calculada es del tipo euclidiana, se aplica una normalización del *z-score* a los descriptores antes de calcular las distancias.

2.3.3. Metodologías basadas en la estructura del receptor

Este conjunto de metodologías, como su nombre lo indica, requiere conocer la estructura tridimensional del *target* o diana biológica. El término *target* generalmente se refiere a una proteína, pero también puede tratarse de ácidos nucleicos como ADN o ARN. Asumiendo que la actividad biológica de un compuesto sobre una diana biológica específica estará determinada por su capacidad de generar interacciones favorables con la misma, es posible simular dichas interacciones e inferir así la actividad biológica de una molécula pequeña. De esta manera, se espera que moléculas que interaccionan de manera similar con un *target* tengan efectos biológicos similares.

Idealmente, la estructura tridimensional de la proteína se obtiene mediante metodologías experimentales, como por ejemplo la cristalografía de rayos X, la resonancia magnética nuclear (RMN) o la Crio-microscopía electrónica (Cryo-EM, del inglés cryogenic electron microscopy). En los últimos años, la Cryo-EM ha tenido un fuerte impacto en la elucidación estructural de proteínas de membrana, debido

principalmente a que esta técnica no requiere de un cristal ordenado y difractante de la proteína, difícil de lograr en este tipo de macroestructuras (**Nogales y Scheres 2015**).

Cuando la estructura de la proteína de interés no ha sido elucidada aún en forma experimental, es posible generar predicciones o modelos de la misma mediante técnicas de modelado molecular. Existen diversas técnicas para modelar la estructura de una proteína, dentro de las cuales podemos mencionar al modelado *ab initio* o de primeros principios, el modelado por *threading* o reconocimiento de plegamiento y el modelado por homología (**Agnihotry et al. 2022**). A continuación se detalla esta última herramienta por ser la utilizada en este trabajo de tesis.

2.3.3.1. Modelado por homología

El modelado por homología o modelado comparativo requiere de un molde o *template* estructural que será utilizado como referencia para modelar la secuencia *target*. El fundamento de esta técnica se basa en que las secuencias evolutivamente relacionadas tendrán una estructura tridimensional similar (**Ginalski 2006; Lesk y Chothia 1980**). Aunque existen varias excepciones a esta regla, esta técnica ha demostrado ser de gran utilidad a través de los años.

El modelado por homología puede describirse en 4 etapas principales, resumidas en el esquema de la Figura 2.7.

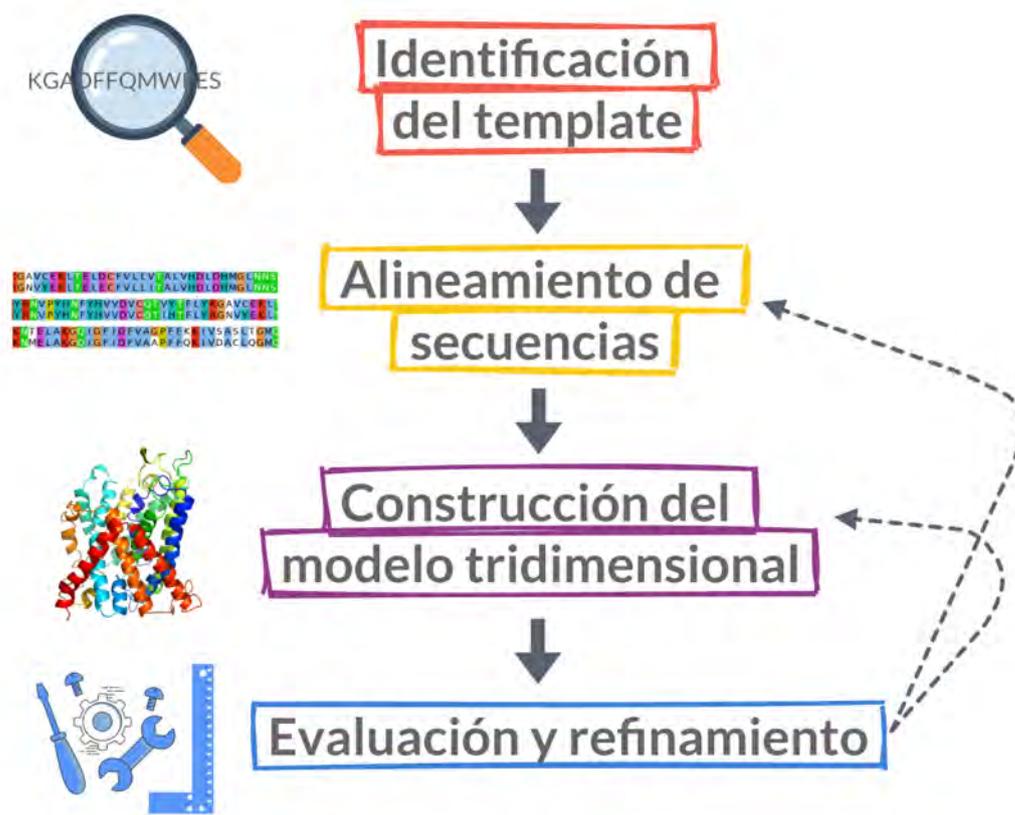


Figura 2.7: Etapas principales del modelado por homología.

1. Identificación del *template*

En esta primera etapa, utilizando la secuencia del blanco molecular a modelar (secuencia problema) como objeto para la búsqueda, se buscan secuencias similares que posean una estructura tridimensional conocida (*templates*). Existen varios algoritmos que nos permiten buscar en grandes bibliotecas de secuencias de manera rápida y efectiva, como por ejemplo BLAST (**Johnson et al. 2008**) o HHBlits (**Remmert et al. 2012**). En cuanto a fuente de *templates*, usualmente se recurre a la base de datos Protein Data Bank (PDB), que recopila información sobre la estructura tridimensional de macromoléculas (como proteínas y ácidos nucleicos) elucidadas experimentalmente y constituye una herramienta de consulta habitual para la comunidad científica relacionada con la biología estructural (**Berman et al. 2000**). Adicionalmente, en este repositorio se encuentran complejos formados entre macromoléculas y moléculas pequeñas (inhibidores, activadores, ligandos naturales), lo que proporciona información muy valiosa para el diseño de fármacos ya que permite reconocer interacciones intermoleculares que justifiquen la actividad biológica.

2. Alineamiento de secuencias

Este es un paso fundamental en el modelado comparativo de proteínas, ya que el alineamiento entre las secuencias será el determinante fundamental de la calidad de los modelos obtenidos. Es posible utilizar el alineamiento generado durante la etapa de búsqueda, o pueden re-alinearse las secuencias utilizando diferentes parámetros o algoritmos específicos. La inspección visual y corrección manual de los mismos, guiada por la propia *expertise* y/o información experimental, es de gran importancia para lograr un buen alineamiento, sobre todo en aquellos casos donde el porcentaje de identidad entre las secuencias es bajo (Forrest, Tang y Honig 2006; Ginalski 2006).

3. Construcción del modelo tridimensional

Los métodos para construir un modelo estructural pueden clasificarse en tres grandes grupos: ensamblado de cuerpos rígidos, coincidencia de segmentos y satisfacción de restricciones espaciales (Martí-Renom et al. 2000). En este último método, utilizado por el software Modeller (Webb y Sali 2016), las restricciones se obtienen del *template* estructural, asumiendo que las distancias y ángulos entre residuos alineados son similares. Luego, estas restricciones son complementadas con restricciones estereoquímicas de enlaces (longitudes y ángulos), ángulos diedros y contactos interatómicos derivados de algún campo de fuerza (FF, del inglés Force Field). Finalmente el modelo se construye minimizando las violaciones de todas estas restricciones.

Las regiones de la secuencia *target* que no se encuentran alineadas con el *template* se denominan inserciones o *loops*. Existen diversas técnicas para modelarlos, como por ejemplo la utilización de librerías estructurales de fragmentos, el modelado *ab initio* o la búsqueda conformacional guiada por funciones de puntuación. A su vez, también existen metodologías híbridas que combinan las técnicas antes mencionadas (Barozet, Chacón y Cortés 2021).

4. Evaluación y refinamiento

Considerando el proceso descrito anteriormente para la construcción de un modelo por homología, es evidente que existirán ciertas zonas “problemáticas” en la estructura, donde las restricciones espaciales no pudieron ser satisfechas. Las mismas a menudo coinciden con aquellas regiones donde el alineamiento

no fue bueno, o no pudieron ser alineadas en absoluto.

Desde un punto de vista estructural, el refinamiento se refiere al proceso de permitir cierto movimiento en los átomos que componen el sistema, en presencia de un FF, de modo de optimizar las distancias interatómicas generando interacciones estables, liberando tensiones y minimizando los *clashes* o colisiones entre átomos cercanos. Para este fin pueden utilizarse distintas metodologías, desde una simple minimización energética hasta simulaciones más complejas como la dinámica molecular o las trayectorias de Montecarlo (**Fan y Mark 2004**). En otro sentido, también es posible refinar un modelo modificando el alineamiento de secuencias que dio origen al mismo, generando así nuevos modelos de manera iterativa (**Webb y Sali 2016**).

2.3.3.2. Docking molecular

El docking molecular es una técnica de simulación cuyo objetivo es predecir, simultáneamente, la orientación espacial relativa de un ligando en un receptor (modo de unión) y la afinidad o energía de unión asociada a la formación de dicho complejo, a partir de las estructuras tridimensionales de los mismos. Como resultado, el programa devuelve una serie de poses o conformaciones del ligando en el sitio de unión del receptor, ordenadas o rankeadas según su energía de unión predicha, siendo aquellas poses mejor rankeadas las más probables (**Gaitonde, Karmakar y Trivedi 2020**). Generalmente, el ligando es una molécula orgánica pequeña y el receptor es una proteína, aunque también puede tratarse de un docking proteína-proteína (**Porter et al. 2019**) o de una molécula pequeña y un ácido nucleico (**Luo et al. 2019**).

Un programa de docking puede entenderse como la combinación de dos algoritmos, uno que explora o propone distintas conformaciones del ligando (algoritmo de búsqueda) y otro que evalúa cada una de esas poses y les asigna un valor de *score* o energía de unión, basándose en las interacciones con el receptor y la conformación del ligando (función de evaluación). Por lo tanto, la *performance* del docking puede evaluarse desde dos perspectivas diferentes:

El *docking power* se refiere a la capacidad de un programa de reproducir correc-

tamente el modo de unión de un ligando en un receptor, y puede estimarse mediante experimentos de *re-docking* y *cross-docking*. El *re-docking* consiste en extraer el ligando de un complejo ligando-receptor determinado experimentalmente, para luego dockearlo en el mismo sitio. En el *cross-docking*, el ligando es extraído de un complejo pero es dockeado en otra estructura del receptor, la cual puede provenir de otra estructura holo o ser una estructura apo (Figura 2.8) Finalmente, la pose predicha se compara con el modo de unión determinado experimentalmente utilizando alguna métrica, como la raíz cuadrada de la desviación cuadrática media (RMSD, del inglés root-mean-square deviation) de las posiciones atómicas calculadas respecto de las experimentales (Kirchmair et al. 2008).

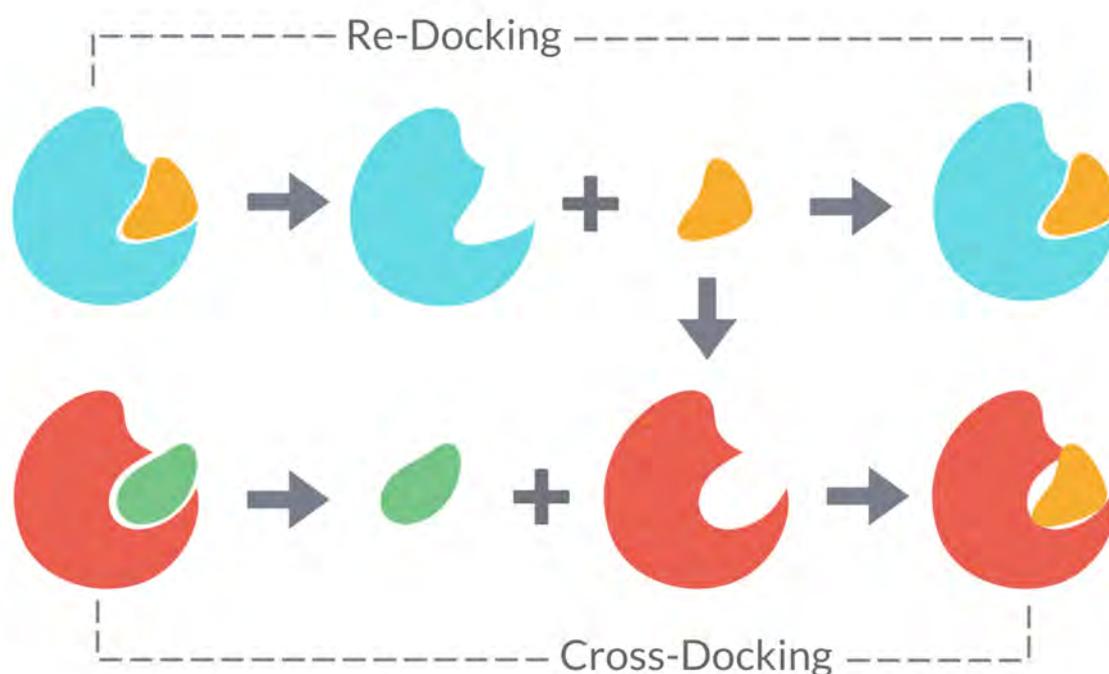


Figura 2.8: Representación esquemática de un experimento de *re-docking* (superior) y uno de *cross-docking* (inferior).

El *scoring power* por su parte, se refiere a la capacidad de una función de evaluación de predecir correctamente la afinidad o energía de formación del complejo ligando-receptor, y puede evaluarse mediante el cribado retrospectivo de un conjunto de validación. Dependiendo de la naturaleza de la actividad biológica asociada a cada compuesto dentro de dicho conjunto, podemos utilizar diferentes métricas. Si la variable dependiente es del tipo continua (IC₅₀, K_i, Inhibición), entonces es posible calcular métricas de correlación o de ranking entre la afinidad predicha y la determinada experimentalmente (Llanos et al. 2020). Por el contrario, si la variable dependiente es del tipo binaria (activo/inactivo), podemos evaluar la *performance*

del docking como un clasificador binario utilizando el valor de afinidad predicho. A su vez, cuando la proporción de la clase positiva (activo) es pequeña en comparación con la clase negativa (inactivo o *decoy*), podemos hablar de *screening power* (**Llanos et al. 2021**; **C. Wang e Y. Zhang 2017**). En la sección suplementaria **D** - Métricas se encuentran detalladas las métricas de correlación, clasificación y cribado virtual más comunes, haciendo foco en las que fueron utilizadas en esta tesis.

En cuanto al grado de flexibilidad con el que se considera al ligando y al receptor durante la simulación, se han desarrollado distintas aproximaciones. En el docking rígido, ambos son considerados como cuerpos rígidos y solo se permite al ligando rotar y trasladarse dentro del sitio de unión. El docking flexible por su parte, comprende diversas variantes según se exploren distintas conformaciones del ligando, de la proteína o de ambos simultáneamente (**Andrusier et al. 2008**).

Existen diversas estrategias que permiten considerar la flexibilidad del *target* en las simulaciones de docking (**Lexa y Carlson 2012**), como el uso de potenciales suavizados o *soft-docking* (**Ferrari et al. 2004**), el uso de librerías de rotámeros para modelar la flexibilidad de las cadenas laterales de ciertos residuos (**Alberts, Todorov y Dean 2005**), considerar distintas conformaciones del receptor (*ensemble docking*) (**Amaro et al. 2018**), o incluso muestrear conformaciones del complejo ligando-receptor *on-the-fly* (**DeLuca, Khar y Meiler 2015**).

A medida que se incluyen más grados de libertad en la simulación (mayor flexibilidad), el costo computacional escala considerablemente. Por lo tanto, la aproximación más utilizada es considerar el ligando totalmente flexible y la proteína rígida (**Meng et al. 2011**).

2.3.3.3. Dinámica molecular

La dinámica molecular (MD, del inglés Molecular Dynamics) es una técnica de simulación que permite estudiar el comportamiento dinámico de las moléculas (biológicas, orgánicas o inorgánicas) a lo largo del tiempo (**Karplus y McCammon 2002**).

En la MD clásica, las distintas configuraciones de un sistema (entendidas como

las posiciones relativas de todos los átomos que lo componen) se generan mediante la resolución de las ecuaciones de movimiento. Por integración numérica de la ecuación que representa la segunda ley de Newton (Ecuación 2.2) es posible calcular la trayectoria de cada partícula, es decir, cómo varían su posición y velocidad a través del tiempo. Para ello, es necesario conocer las fuerzas que actúan sobre cada átomo o partícula del sistema, las cuales naturalmente varían cada vez que estas cambian de posición. Estas fuerzas derivan de la función de energía potencial, comúnmente conocida como Hamiltoniano (Tuckerman y Martyna 2000). A modo de ejemplo, en la Ecuación 2.3 se muestra el Hamiltoniano de Amber, el paquete simulación que fue utilizado en este trabajo de tesis.

$$\mathbf{F} = \frac{\partial \mathbf{p}}{\partial t} \quad (2.2)$$

Ecuación 2.2 Segunda ley de Newton. La fuerza \mathbf{F} que impulsa el movimiento de una partícula es proporcional al cambio de *momentum* \mathbf{p} con respecto al tiempo \mathbf{t} . El término \mathbf{p} es el producto de la masa de una partícula y su velocidad.

$$E_{\text{total}} = \sum_{\text{enlaces}} K_r (r - r_{eq})^2 + \sum_{\text{angulos}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedros}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

Ecuación 2.3. Función de energía potencial que utiliza el programa de simulación Amber (Ponder y Case 2003).

La Ecuación 2.3 puede entenderse como una suma ponderada de las distintas interacciones entre pares de átomos, dentro de la cuales podemos distinguir dos categorías: enlazantes y no enlazantes. Las contribución de las interacciones enlazantes incluye a la energía de los enlaces, los ángulos de torsión y los ángulos diedros (expresadas matemáticamente en los tres primeros términos). Por su parte, las no enlazantes incluyen a las interacciones del tipo Van der Waals y coulombicas, representadas en el último término de la ecuación.

Calcular la energía potencial del sistema para cada configuración requiere resolver la Ecuación 2.3 a cada instante de tiempo, utilizando la posición de los átomos

así como también una serie de parámetros o constantes específicas relacionadas con cada uno de los términos. En el contexto de la MD, este Hamiltoniano y su conjunto de parámetros asociados se conoce como campo de fuerza (FF).

Según el grado de separación entre el sistema y su entorno, pueden generarse distintos *ensembles* de simulación. En el *ensemble* microcanónico (NVE), el sistema se encuentra totalmente aislado de su entorno y no existe intercambio de energía o materia con el exterior. Es decir, el número de partículas (N), el volumen (V) y la energía total (E) permanecen constantes. Esto se corresponde con un proceso adiabático, donde no hay intercambio de calor y no hay una temperatura definida para el sistema. En el *ensemble* canónico (NVT), el volumen del sistema permanece constante y solo puede existir una transferencia de energía (en forma de temperatura) a través de los límites del sistema, pero no de masa. Podemos pensar el sistema como si estuviera inmerso en un gran baño termostatizado de temperatura (T), por lo que se requiere de un algoritmo que controle la temperatura exterior, denominado termostato. En este *ensemble*, la energía del sistema no es constante, sino que puede tomar diferentes valores dependiendo de la temperatura. Finalmente, en el *ensemble* isotérmico–isobárico (NPT), el número de partículas (N), la presión (P) y la temperatura (T) permanecen constantes. Por lo tanto, además de un termostato también se requiere un barostato para controlar la presión. Esta configuración se asemeja más a una experiencia real en el laboratorio, donde el sistema es un contenedor abierto a la atmósfera y a temperatura ambiente.

Una simulación de MD puede describirse en 3 etapas principales (**Katiyar y Jha 2018**). Durante la etapa de **pre-procesamiento**, la estructura o complejo macromolecular es cuidadosamente analizado y reparado, de manera que esta no contenga átomos faltantes o residuos con más de una conformación. A su vez, debe establecerse el estado de protonación de los residuos que sean protonables, ya sea de manera automática o manual. En esta etapa se deben elegir el/los FF que se utilizarán para describir el sistema, según los componentes del mismo (proteína, lípidos, ácidos nucleicos, metales, moléculas orgánicas, etc.). El sistema es ensamblado y solvatado, estableciendo la forma y tamaño de la caja de simulación, y las condiciones periódicas de contorno. Finalmente, el sistema completo se minimiza, para aliviar las tensiones y solapamientos entre átomos que pudieran resultar en grandes

fuerzas interatómicas, lo cual genera inestabilidades durante la simulación. La etapa de **simulación** propiamente dicha comprende una fase inicial de calentamiento y equilibración, hasta que las propiedades de interés (temperatura, presión, volumen, densidad, etc.) converjan, y una etapa de producción durante la cual se muestrean distintas configuraciones del sistema en equilibrio. La etapa final es la de **análisis de resultados**, durante la cual se evalúan distintas propiedades o características a lo largo de la trayectoria simulada.

La MD permite considerar el entorno del blanco molecular de manera explícita, es decir las moléculas de soluto, solvente o la membrana plasmática. Por lo tanto, esta herramienta resulta muy útil para estudiar aquellos fenómenos que involucran al blanco molecular y su interacción con el entorno, como los canales iónicos.

2.3.4. Metodologías combinadas

Este conjunto de métodos supone la aplicación conjunta de dos o más metodologías diferentes a la hora de cribar una base de datos, las cuales pueden basarse en la estructura de los ligandos, del receptor o de ambas (Sotriffer 2011). Debido a que ambos tipos de metodologías son ortogonales entre sí, es decir que se basan en aproximaciones diferentes, su combinación permite compensar las debilidades inherentes a cada una de ellas. Las mismas pueden aplicarse en paralelo o de manera secuencial, siendo esta última la forma más común ya que implica un menor costo computacional (Kumar y K. Y. J. Zhang 2015; Perez-Pineiro et al. 2009; Xu, Song y Q. Yang 2017).

Mediante la aplicación inicial de métodos menos costosos computacionalmente (más rápidos), como los basados en la estructura de los ligandos, es posible reducir el número de moléculas que serán sometidas a simulaciones de mayor costo computacional, como el docking molecular (Alberca et al. 2018). Sin embargo, esta aproximación secuencial de VS puede limitar el número de estructuras novedosas halladas en el cribado, ya que los métodos basados en los ligandos suponen un *bias* intrínseco asociado a los ligandos conocidos actualmente para un determinado blanco molecular. A su vez, aquellos compuestos activos erróneamente clasificados como inactivos (falsos negativos), no podrán ser sometidos a las predicciones del segundo

método.

En los métodos paralelos de VS, la librería completa es evaluada y rankeada por ambos tipos de modelos, y finalmente los *hits* mejor rankeados por el conjunto de modelos son seleccionados para su posterior evaluación experimental (**Kumar** y K. Y. J. **Zhang** 2015).

2.4. Reposicionamiento de fármacos

El reposicionamiento de fármacos (*drug repositioning* o *drug repurposing*), se refiere a encontrar nuevos usos terapéuticos para medicamentos ya existentes, incluyendo aquellos que se encuentran actualmente en el mercado farmacéutico, fármacos discontinuados o retirados, o los que se encuentran aún en etapas de desarrollo (**Jarada, Rokne** y **Alhajj** 2020).

Esta estrategia supone ciertas ventajas con respecto a las técnicas convencionales de descubrimiento de fármacos, y es por ello que ha ganado gran visibilidad en los últimos años. El reposicionamiento reduce considerablemente el tiempo y los costos asociados a la aprobación de una droga para una nueva indicación (**Pushpakom** et al. 2019).

Los candidatos a ser reposicionados han atravesado previamente estudios toxicológicos pre-clínicos y poseen un perfil farmacocinético caracterizado. Cuando el candidato a ser reposicionado ya ha sido aprobado para su uso y comercialización, significa que la droga ha atravesado uno o varios ensayos clínicos, y también existen datos de ensayos de fase 4 o post-comercialización. Por otra parte, la síntesis de este principio activo ya ha sido optimizada, y existen estudios de pre-formulación a nivel farmacotécnico.

Un ejemplo reciente es la pandemia de SARS-CoV-2 (del inglés, severe acute respiratory syndrome coronavirus 2), originada en China en 2019. Ante la urgente necesidad de encontrar terapias farmacológicas para tratar el virus, el reposicionamiento de fármacos fue una de las estrategias más exploradas por la comunidad científica para dar una respuesta rápida ante la alarmante situación (**Bellera** et al.

2021).

2.4.1. Reposicionamiento de fármacos en Epilepsia

El reposicionamiento de fármacos en el campo de la Epilepsia es una estrategia prometedora considerando que, como se describió en el Capítulo 1, esta es una enfermedad compleja donde más de un blanco molecular puede estar implicado en el desarrollo de la patología, a la vez que algunas ACDs presentan un alto grado de promiscuidad y múltiples mecanismos de acción (**Mehndiratta et al. 2016**).

En un estudio reciente donde se analizaron casos de reposicionamiento reportados para patologías del SNC, Caban y colaboradores encontraron que las ACDs son una de las mayores fuentes de candidatos a reposicionamiento para otras patologías del SNC. Cabe destacar que, para que un fármaco sea reposicionado como anti-epiléptico, este debe ser capaz de atravesar la barrera hematoencefálica. Este no es un requisito fácil de cumplimentar, y es por ello que la mayor proporción de casos de reposicionamiento para patologías del SNC provienen de fármacos cuya indicación original es otra patología del SNC (**Caban et al. 2017**).

En los últimos años, gracias a los avances en áreas como la genómica, los esfuerzos por encontrar nuevos antiepilépticos a través del reposicionamiento de fármacos se han volcado también hacia el análisis de genes y su expresión diferencial en esta patología (**Brueggeman et al. 2019; Ko et al. 2019; Mirza et al. 2017**).

Referencias Capítulo 2

- Agnihotry**, S. et al. (2022). “Chapter 11 - Protein structure prediction”. *Bioinformatics*. Ed. por D. B. **Singh** y R. K. **Pathak**. Academic Press, págs. 177-188. DOI: [10.1016/B978-0-323-89775-4.00023-7](https://doi.org/10.1016/B978-0-323-89775-4.00023-7).
- Alberca**, L. N. et al. (2018). “Cascade Ligand- and Structure-Based Virtual Screening to Identify New Trypanocidal Compounds Inhibiting Putrescine Uptake”. *Frontiers in Cellular and Infection Microbiology* 8, pág. 173. DOI: [10.3389/fcimb.2018.00173](https://doi.org/10.3389/fcimb.2018.00173).
- Alberts**, I. L., N. P. **Todorov** y P. M. **Dean** (2005). “Receptor flexibility in de novo ligand design and docking”. *Journal of Medicinal Chemistry* 48.21, págs. 6585-6596. DOI: [10.1021/jm050196j](https://doi.org/10.1021/jm050196j).
- Amaro**, R. E. et al. (2018). “Ensemble Docking in Drug Discovery”. *Biophysical Journal* 114.10, págs. 2271-2278. DOI: [10.1016/j.bpj.2018.02.038](https://doi.org/10.1016/j.bpj.2018.02.038).
- Andrusier**, N. et al. (2008). “Principles of Flexible Protein-Protein Docking”. *Proteins* 73.2, págs. 271-289. DOI: [10.1002/prot.22170](https://doi.org/10.1002/prot.22170).
- Babaoglu**, K. et al. (2008). “Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against -Lactamase”. *Journal of medicinal chemistry* 51.8, págs. 2502-2511. DOI: [10.1021/jm701500e](https://doi.org/10.1021/jm701500e).
- Balaban**, A. T. (1998). “Topological and Stereochemical Molecular Descriptors for Databases Useful in QSAR, Similarity/Dissimilarity and Drug Design”. *SAR and QSAR in Environmental Research* 8.1-2, págs. 1-21. DOI: [10.1080/10629369808033259](https://doi.org/10.1080/10629369808033259).
- Barozet**, A., P. **Chacón** y J. **Cortés** (2021). “Current approaches to flexible loop modeling”. *Current Research in Structural Biology* 3, págs. 187-191. DOI: [10.1016/j.crstbi.2021.07.002](https://doi.org/10.1016/j.crstbi.2021.07.002).

- Bellera, C. L. et al.** (2021). “Can drug repurposing strategies be the solution to the COVID-19 crisis?” *Expert Opinion on Drug Discovery* 16.6, págs. 605-612. DOI: [10.1080/17460441.2021.1863943](https://doi.org/10.1080/17460441.2021.1863943).
- Bergstra, J., D. Yamins y D. Cox** (2013). “Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms”. Austin, Texas, págs. 13-19. DOI: [10.25080/Majora-8b375195-003](https://doi.org/10.25080/Majora-8b375195-003).
- Berman, H. M. et al.** (2000). “The Protein Data Bank”. *Nucleic Acids Research* 28.1, págs. 235-242. DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).
- Bero, S. A. et al.** (2017). “Similarity Measure for Molecular Structure: A Brief Review”. *Journal of Physics: Conference Series* 892, págs. 012015. DOI: [10.1088/1742-6596/892/1/012015](https://doi.org/10.1088/1742-6596/892/1/012015).
- Berthold, M. R. et al.** (2008). “KNIME: The Konstanz Information Miner”. *Data Analysis, Machine Learning and Applications*. Ed. por C. **Preisach** et al. Studies in Classification, Data Analysis, and Knowledge Organization. Berlin, Heidelberg: Springer, págs. 319-326. DOI: [10.1007/978-3-540-78246-9_38](https://doi.org/10.1007/978-3-540-78246-9_38).
- Bohacek, R. S., C. McMartin y W. C. Guida** (1996). “The art and practice of structure-based drug design: a molecular modeling perspective”. *Medicinal Research Reviews* 16.1, págs. 3-50. DOI: [10.1002/\(SICI\)1098-1128\(199601\)16:1<3::AID-MED1>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6).
- Bonnin, R.** (2017). *Machine Learning for Developers*. Packt Publishing.
- Brueggeman, L. et al.** (2019). “Drug repositioning in epilepsy reveals novel antiseizure candidates”. *Annals of Clinical and Translational Neurology* 6.2, págs. 295-309. DOI: [10.1002/acn3.703](https://doi.org/10.1002/acn3.703).
- Caban, A. et al.** (2017). “Filling the gap in CNS drug development: evaluation of the role of drug repurposing”. *Journal of Market Access & Health Policy* 5.1, págs. 1299833. DOI: [10.1080/20016689.2017.1299833](https://doi.org/10.1080/20016689.2017.1299833).
- Cherkasov, A. et al.** (2014). “QSAR Modeling: Where Have You Been? Where Are You Going To?” *Journal of Medicinal Chemistry* 57.12, págs. 4977-5010. DOI: [10.1021/jm4004285](https://doi.org/10.1021/jm4004285).
- Chirico, N. y P. Gramatica** (2011). “Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient”. *Journal of Chemical Information and Modeling* 51.9, págs. 2320-2335. DOI: [10.1021/ci200211n](https://doi.org/10.1021/ci200211n).

- Cozza**, G. (2017). “The Development of CK2 Inhibitors: From Traditional Pharmacology to in Silico Rational Drug Design”. *Pharmaceuticals (Basel, Switzerland)* 10.1. DOI: [10.3390/ph10010026](https://doi.org/10.3390/ph10010026).
- Davies**, M. et al. (2015). “ChEMBL web services: streamlining access to drug discovery data and utilities”. *Nucleic Acids Research* 43.Web Server issue, W612-W620. DOI: [10.1093/nar/gkv352](https://doi.org/10.1093/nar/gkv352).
- DeLuca**, S., K. **Khar** y J. **Meiler** (2015). “Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand”. *PLOS ONE* 10.7, e0132508. DOI: [10.1371/journal.pone.0132508](https://doi.org/10.1371/journal.pone.0132508).
- Dezső**, Z. y M. **Ceccarelli** (2020). “Machine learning prediction of oncology drug targets based on protein and network properties”. *BMC Bioinformatics* 21.1, pág. 104. DOI: [10.1186/s12859-020-3442-9](https://doi.org/10.1186/s12859-020-3442-9).
- DiMasi**, J. A., H. G. **Grabowski** y R. W. **Hansen** (2016). “Innovation in the pharmaceutical industry: New estimates of R&D costs”. *Journal of Health Economics* 47, págs. 20-33. DOI: [10.1016/j.jhealeco.2016.01.012](https://doi.org/10.1016/j.jhealeco.2016.01.012).
- Doman**, T. N. et al. (2002). “Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B”. *Journal of Medicinal Chemistry* 45.11, págs. 2213-2221. DOI: [10.1021/jm010548w](https://doi.org/10.1021/jm010548w).
- Fan**, H. y A. E. **Mark** (2004). “Refinement of homology-based protein structures by molecular dynamics simulation techniques”. *Protein Science* 13.1, págs. 211-220. DOI: [10.1110/ps.03381404](https://doi.org/10.1110/ps.03381404).
- Ferrari**, A. M. et al. (2004). “Soft Docking and Multiple Receptor Conformations in Virtual Screening”. *Journal of medicinal chemistry* 47.21, págs. 5076-5084. DOI: [10.1021/jm049756p](https://doi.org/10.1021/jm049756p).
- Ferreira**, R. S. et al. (2010). “Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors”. *Journal of Medicinal Chemistry* 53.13, págs. 4891-4905. DOI: [10.1021/jm100488w](https://doi.org/10.1021/jm100488w).
- Forrest**, L. R., C. L. **Tang** y B. **Honig** (2006). “On the Accuracy of Homology Modeling and Sequence Alignment Methods Applied to Membrane Proteins”. *Biophysical Journal* 91.2, págs. 508-517. DOI: [10.1529/biophysj.106.082313](https://doi.org/10.1529/biophysj.106.082313).
- Fourches**, D., E. **Muratov** y A. **Tropsha** (2010). “Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research”. *Journal of Chemical Information and Modeling* 50.7, págs. 1189-1204. DOI: [10.1021/ci100176x](https://doi.org/10.1021/ci100176x).

- Fushiki**, T. (2011). “Estimation of prediction error by using K-fold cross-validation”. *Statistics and Computing* 21.2, págs. 137-146. DOI: [10.1007/s11222-009-9153-8](https://doi.org/10.1007/s11222-009-9153-8).
- Gaitonde**, V., P. **Karmakar** y A. **Trivedi** (2020). *Drug Discovery and Development*. Rijeka: IntechOpen. DOI: [10.5772/intechopen.77685](https://doi.org/10.5772/intechopen.77685).
- Gaston**, T. E. y J. P. **Szaflarski** (2018). “Cannabis for the Treatment of Epilepsy: an Update”. *Current Neurology and Neuroscience Reports* 18.11, pág. 73. DOI: [10.1007/s11910-018-0882-y](https://doi.org/10.1007/s11910-018-0882-y).
- Ghofrani**, H. A., I. H. **Osterloh** y F. **Grimminger** (2006). “Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond”. *Nature Reviews Drug Discovery* 5.8, págs. 689-702. DOI: [10.1038/nrd2030](https://doi.org/10.1038/nrd2030).
- Ginalski**, K. (2006). “Comparative modeling for protein structure prediction”. *Current Opinion in Structural Biology. Theory and simulation/Macromolecular assemblages* 16.2, págs. 172-177. DOI: [10.1016/j.sbi.2006.02.003](https://doi.org/10.1016/j.sbi.2006.02.003).
- Goodarzi**, M., B. **Dejaegher** e Y. V. **Heyden** (2012). “Feature Selection Methods in QSAR Studies”. *Journal of AOAC INTERNATIONAL* 95.3, págs. 636-651. DOI: [10.5740/jaoacint.SGE_Goodarzi](https://doi.org/10.5740/jaoacint.SGE_Goodarzi).
- Hawkins**, D. M., S. C. **Basak** y D. **Mills** (2003). “Assessing Model Fit by Cross-Validation”. *Journal of Chemical Information and Computer Sciences* 43.2, págs. 579-586. DOI: [10.1021/ci025626i](https://doi.org/10.1021/ci025626i).
- Hong**, J. (2011). “Natural product diversity and its role in chemical biology and drug discovery”. *Current opinion in chemical biology* 15.3, págs. 350-354. DOI: [10.1016/j.cbpa.2011.03.004](https://doi.org/10.1016/j.cbpa.2011.03.004).
- Hou**, S. et al. (2018). “Advanced Development of Primary Pancreatic Organoid Tumor Models for High-Throughput Phenotypic Drug Screening”. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* 23.6, págs. 574-584. DOI: [10.1177/2472555218766842](https://doi.org/10.1177/2472555218766842).
- Hughes**, J. P. et al. (2011). “Principles of early drug discovery”. *British Journal of Pharmacology* 162.6, págs. 1239-1249. DOI: [10.1111/j.1476-5381.2010.01127.x](https://doi.org/10.1111/j.1476-5381.2010.01127.x).
- Hutter**, F., H. **Hoos** y K. **Leyton-Brown** (2013). “An evaluation of sequential model-based optimization for expensive blackbox functions”. *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*.

- GECCO '13 Companion. Amsterdam, The Netherlands: Association for Computing Machinery, págs. 1209-1216. DOI: [10.1145/2464576.2501592](https://doi.org/10.1145/2464576.2501592).
- Ivanciuc**, O. (2013). “Chemical graphs, molecular matrices and topological indices in chemoinformatics and quantitative structure-activity relationships”. *Current Computer-Aided Drug Design* 9.2, págs. 153-163. DOI: [10.2174/1573409911309020002](https://doi.org/10.2174/1573409911309020002).
- Jarada**, T. N., J. G. **Rokne** y R. **Alhajj** (2020). “A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions”. *Journal of Cheminformatics* 12.1, pág. 46. DOI: [10.1186/s13321-020-00450-7](https://doi.org/10.1186/s13321-020-00450-7).
- Jayasundara**, K. et al. (2019). “Estimating the clinical cost of drug development for orphan versus non-orphan drugs”. *Orphanet Journal of Rare Diseases* 14.1, pág. 12. DOI: [10.1186/s13023-018-0990-4](https://doi.org/10.1186/s13023-018-0990-4).
- Johnson**, M. et al. (2008). “NCBI BLAST: a better web interface”. *Nucleic Acids Research* 36.Web Server issue, W5-9. DOI: [10.1093/nar/gkn201](https://doi.org/10.1093/nar/gkn201).
- Karplus**, M. y J. A. **McCammon** (2002). “Molecular dynamics simulations of biomolecules”. *Nature Structural Biology* 9.9, págs. 646-652. DOI: [10.1038/nsb0902-646](https://doi.org/10.1038/nsb0902-646).
- Katiyar**, R. S. y P. K. **Jha** (2018). “Molecular simulations in drug delivery: Opportunities and challenges”. *WIREs Computational Molecular Science* 8.4, e1358. DOI: [10.1002/wcms.1358](https://doi.org/10.1002/wcms.1358).
- Kirchmair**, J. et al. (2008). “Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes?” *Journal of Computer-Aided Molecular Design* 22.3, págs. 213-228. DOI: [10.1007/s10822-007-9163-6](https://doi.org/10.1007/s10822-007-9163-6).
- Ko**, Y. et al. (2019). “Systematic Approach for Drug Repositioning of Anti-Epileptic Drugs”. *Diagnostics (Basel, Switzerland)* 9.4. DOI: [10.3390/diagnostics9040208](https://doi.org/10.3390/diagnostics9040208).
- Kraus**, O. Z. y B. J. **Frey** (2016). “Computer vision for high content screening”. *Critical Reviews in Biochemistry and Molecular Biology* 51.2, págs. 102-109. DOI: [10.3109/10409238.2015.1135868](https://doi.org/10.3109/10409238.2015.1135868).
- Kumar**, A. y K. Y. J. **Zhang** (2015). “Hierarchical virtual screening approaches in small molecule drug discovery”. *Methods (San Diego, Calif.)* 71, págs. 26-37. DOI: [10.1016/j.ymeth.2014.07.007](https://doi.org/10.1016/j.ymeth.2014.07.007).

- Leonard**, J. T. y **K. Roy** (2006). "On Selection of Training and Test Sets for the Development of Predictive QSAR models". *QSAR & Combinatorial Science* 25.3, págs. 235-251. DOI: [10.1002/qsar.200510161](https://doi.org/10.1002/qsar.200510161).
- Lesk**, A. M. y **C. Chothia** (1980). "How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins". *Journal of Molecular Biology* 136.3, págs. 225-270. DOI: [10.1016/0022-2836\(80\)90373-3](https://doi.org/10.1016/0022-2836(80)90373-3).
- Lexa**, K. W. y **H. A. Carlson** (2012). "Protein Flexibility in Docking and Surface Mapping". *Quarterly reviews of biophysics* 45.3, págs. 301-343. DOI: [10.1017/S0033583512000066](https://doi.org/10.1017/S0033583512000066).
- Liddle**, D. E. (2006). "The Wider Impact of Moore's Law". *IEEE Solid-State Circuits Society Newsletter* 11.3, págs. 28-30. DOI: [10.1109/N-SSC.2006.4785858](https://doi.org/10.1109/N-SSC.2006.4785858).
- Llanos**, M. A. et al. (2020). "A structure-based approach towards the identification of novel antichagasic compounds: Trypanosoma cruzi carbonic anhydrase inhibitors". *Journal of Enzyme Inhibition and Medicinal Chemistry* 35.1, págs. 21-30. DOI: [10.1080/14756366.2019.1677638](https://doi.org/10.1080/14756366.2019.1677638).
- Llanos**, M. A. et al. (2021). "Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study". *Journal of Chemical Information and Modeling* 61.8, págs. 3758-3770. DOI: [10.1021/acs.jcim.1c00404](https://doi.org/10.1021/acs.jcim.1c00404).
- Lo**, Y.-C. et al. (2018). "Machine learning in chemoinformatics and drug discovery". *Drug Discovery Today* 23.8, págs. 1538-1546. DOI: [10.1016/j.drudis.2018.05.010](https://doi.org/10.1016/j.drudis.2018.05.010).
- Luo**, J. et al. (2019). "Challenges and current status of computational methods for docking small molecules to nucleic acids". *European Journal of Medicinal Chemistry* 168, págs. 414-425. DOI: [10.1016/j.ejmech.2019.02.046](https://doi.org/10.1016/j.ejmech.2019.02.046).
- Lyu**, J. et al. (2019). "Ultra-large library docking for discovering new chemotypes". *Nature* 566.7743, págs. 224-229. DOI: [10.1038/s41586-019-0917-9](https://doi.org/10.1038/s41586-019-0917-9).
- Macalino**, S. J. Y. et al. (2015). "Role of computer-aided drug design in modern drug discovery". *Archives of Pharmacal Research* 38.9, págs. 1686-1701. DOI: [10.1007/s12272-015-0640-5](https://doi.org/10.1007/s12272-015-0640-5).
- Macarron**, R. et al. (2011). "Impact of high-throughput screening in biomedical research". *Nature Reviews Drug Discovery* 10.3, págs. 188-195. DOI: [10.1038/nrd3368](https://doi.org/10.1038/nrd3368).

- Majhi, S. y D. Das** (2021). “Chemical derivatization of natural products: Semisynthesis and pharmacological aspects- A decade update”. *Tetrahedron* 78, pág. 131801. DOI: [10.1016/j.tet.2020.131801](https://doi.org/10.1016/j.tet.2020.131801).
- Martí-Renom, M. A. et al.** (2000). “Comparative Protein Structure Modeling of Genes and Genomes”. *Annual Review of Biophysics and Biomolecular Structure* 29.1, págs. 291-325. DOI: [10.1146/annurev.biophys.29.1.291](https://doi.org/10.1146/annurev.biophys.29.1.291).
- Martin, T. M. et al.** (2012). “Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?” *Journal of Chemical Information and Modeling* 52.10, págs. 2570-2578. DOI: [10.1021/ci300338w](https://doi.org/10.1021/ci300338w).
- Mehndiratta, M. M. et al.** (2016). “Drug repositioning”. *International Journal of Epilepsy* 03.2, págs. 91-94. DOI: [10.1016/j.ijep.2016.09.002](https://doi.org/10.1016/j.ijep.2016.09.002).
- Mendez, D. et al.** (2019). “ChEMBL: towards direct deposition of bioassay data”. *Nucleic Acids Research* 47.D1, págs. D930-D940. DOI: [10.1093/nar/gky1075](https://doi.org/10.1093/nar/gky1075).
- Meng, X.-Y. et al.** (2011). “Molecular Docking: A powerful approach for structure-based drug discovery”. *Current computer-aided drug design* 7.2, págs. 146-157.
- Mirza, N. et al.** (2017). “Identifying new antiepileptic drugs through genomics-based drug repurposing”. *Human Molecular Genetics* 26.3, págs. 527-537. DOI: [10.1093/hmg/ddw410](https://doi.org/10.1093/hmg/ddw410).
- Mooney, C. Z. et al.** (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference*. SAGE.
- Moriwaki, H. et al.** (2018). “Mordred: a molecular descriptor calculator”. *Journal of Cheminformatics* 10.1, pág. 4. DOI: [10.1186/s13321-018-0258-y](https://doi.org/10.1186/s13321-018-0258-y).
- Murray, R. M. et al.** (2016). “Traditional marijuana, high-potency cannabis and synthetic cannabinoids: increasing risk for psychosis”. *World psychiatry: official journal of the World Psychiatric Association (WPA)* 15.3, págs. 195-204. DOI: [10.1002/wps.20341](https://doi.org/10.1002/wps.20341).
- Mysinger, M. M. et al.** (2012). “Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking”. *Journal of Medicinal Chemistry* 55.14, págs. 6582-6594. DOI: [10.1021/jm300687e](https://doi.org/10.1021/jm300687e).
- Neves, B. J. et al.** (2018). “QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery”. *Frontiers in Pharmacology* 9. DOI: [10.3389/fphar.2018.01275](https://doi.org/10.3389/fphar.2018.01275).

- Newman, D. J. y G. M. Cragg** (2012). “Natural products as sources of new drugs over the 30 years from 1981 to 2010”. *Journal of Natural Products* 75.3, págs. 311-335. DOI: [10.1021/np200906s](https://doi.org/10.1021/np200906s).
- Nogales, E. y S. H. W. Scheres** (2015). “Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity”. *Molecular Cell* 58.4, págs. 677-689. DOI: [10.1016/j.molcel.2015.02.019](https://doi.org/10.1016/j.molcel.2015.02.019).
- Pedregosa, F. et al.** (2011). “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research* 12.85, págs. 2825-2830.
- Perez-Pineiro, R. et al.** (2009). “Development of a Novel Virtual Screening Cascade Protocol to Identify Potential Trypanothione Reductase Inhibitors”. *Journal of Medicinal Chemistry* 52.6, págs. 1670-1680. DOI: [10.1021/jm801306g](https://doi.org/10.1021/jm801306g).
- Perucca, E.** (2017). “Cannabinoids in the Treatment of Epilepsy: Hard Evidence at Last?” *Journal of Epilepsy Research* 7.2, págs. 61-76. DOI: [10.14581/jer.17012](https://doi.org/10.14581/jer.17012).
- Ponder, J. W. y D. A. Case** (2003). “Force fields for protein simulations”. *Advances in Protein Chemistry* 66, págs. 27-85. DOI: [10.1016/s0065-3233\(03\)66002-x](https://doi.org/10.1016/s0065-3233(03)66002-x).
- Porter, K. A. et al.** (2019). “What method to use for protein–protein docking?” *Current Opinion in Structural Biology. Macromolecular Assemblies Theory and simulation: Demystifying GPCRs* 55, págs. 1-7. DOI: [10.1016/j.sbi.2018.12.010](https://doi.org/10.1016/j.sbi.2018.12.010).
- Pushpakom, S. et al.** (2019). “Drug repurposing: progress, challenges and recommendations”. *Nature Reviews. Drug Discovery* 18.1, págs. 41-58. DOI: [10.1038/nrd.2018.168](https://doi.org/10.1038/nrd.2018.168).
- Remmert, M. et al.** (2012). “HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment”. *Nature Methods* 9.2, págs. 173-175. DOI: [10.1038/nmeth.1818](https://doi.org/10.1038/nmeth.1818).
- Rogers, D. y M. Hahn** (2010). “Extended-Connectivity Fingerprints”. *Journal of Chemical Information and Modeling* 50.5, págs. 742-754. DOI: [10.1021/ci100050t](https://doi.org/10.1021/ci100050t).
- Rong, C. et al.** (2017). “Cannabidiol in medical marijuana: Research vistas and potential opportunities”. *Pharmacological Research* 121, págs. 213-218. DOI: [10.1016/j.phrs.2017.05.005](https://doi.org/10.1016/j.phrs.2017.05.005).

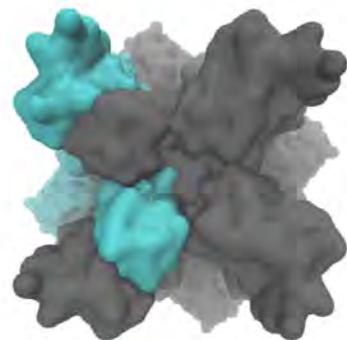
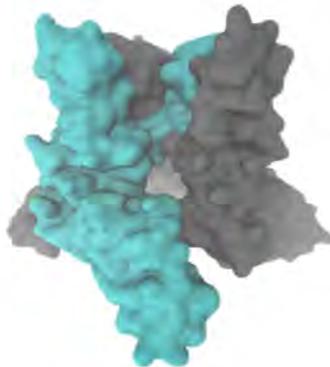
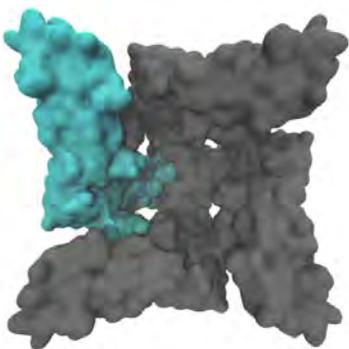
- Roy, K., S. Kar y P. Ambure** (2015). “On a simple approach for determining applicability domain of QSAR models”. *Chemometrics and Intelligent Laboratory Systems* 145, págs. 22-29. DOI: [10.1016/j.chemolab.2015.04.013](https://doi.org/10.1016/j.chemolab.2015.04.013).
- Roy, K., S. Kar y R. N. Das** (2015a). “Chapter 2 - Chemical Information and Descriptors”. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Ed. por K. **Roy**, S. **Kar** y R. N. **Das**. Boston: Academic Press, págs. 47-80. DOI: [10.1016/B978-0-12-801505-6.00002-8](https://doi.org/10.1016/B978-0-12-801505-6.00002-8).
- (2015b). “Statistical Methods in QSAR/QSPR”. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*. Ed. por K. **Roy**, S. **Kar** y R. N. **Das**. SpringerBriefs in Molecular Science. Cham: Springer International Publishing, págs. 37-59. DOI: [10.1007/978-3-319-17281-1_2](https://doi.org/10.1007/978-3-319-17281-1_2).
- Russo, E. B. et al.** (2008). “Phytochemical and genetic analyses of ancient cannabis from Central Asia”. *Journal of Experimental Botany* 59.15, págs. 4171-4182. DOI: [10.1093/jxb/ern260](https://doi.org/10.1093/jxb/ern260).
- Sahigara, F. et al.** (2012). “Comparison of Different Approaches to Define the Applicability Domain of QSAR Models”. *Molecules* 17.5, págs. 4791-4810. DOI: [10.3390/molecules17054791](https://doi.org/10.3390/molecules17054791).
- Sommer, C. et al.** (2017). “A deep learning and novelty detection framework for rapid phenotyping in high-content screening”. *Molecular Biology of the Cell* 28.23, págs. 3428-3436. DOI: [10.1091/mbc.e17-05-0333](https://doi.org/10.1091/mbc.e17-05-0333).
- Sottriffer, C. A.** (2011). “Virtual screening : principles, challenges, and practical guidelines”.
- Sterling, T. y J. J. Irwin** (2015). “ZINC 15 – Ligand Discovery for Everyone”. *Journal of Chemical Information and Modeling* 55.11, págs. 2324-2337. DOI: [10.1021/acs.jcim.5b00559](https://doi.org/10.1021/acs.jcim.5b00559).
- Tan, S. Y. e Y. Tatsumura** (2015). “Alexander Fleming (1881–1955): Discoverer of penicillin”. *Singapore Medical Journal* 56.7, págs. 366-367. DOI: [10.11622/smedj.2015105](https://doi.org/10.11622/smedj.2015105).
- Todeschini, R. y V. Consonni** (2008). *Handbook of Molecular Descriptors*. John Wiley & Sons.
- Tropsha, A.** (2010). “Best Practices for QSAR Model Development, Validation, and Exploitation”. *Molecular Informatics* 29.6-7, págs. 476-488. DOI: [10.1002/minf.201000061](https://doi.org/10.1002/minf.201000061).

- Truchon, J.-F. y C. I. Bayly** (2007). “Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem”. *Journal of Chemical Information and Modeling* 47.2, págs. 488-508. DOI: [10.1021/ci600426e](https://doi.org/10.1021/ci600426e).
- Tuckerman, M. E. y G. J. Martyna** (2000). “Understanding Modern Molecular Dynamics: Techniques and Applications”. *The Journal of Physical Chemistry B* 104.2, págs. 159-178. DOI: [10.1021/jp992433y](https://doi.org/10.1021/jp992433y).
- Udayakumar, M. et al.** (2013). “Receptor-based pharmacophore tool for design and development of next-generation drugs”. *International Journal of Bioinformatics Research and Applications* 9.5, págs. 487-516. DOI: [10.1504/IJBRA.2013.056076](https://doi.org/10.1504/IJBRA.2013.056076).
- Vedani, A. y M. Dobler** (2000). “Multi-dimensional QSAR in drug research. Predicting binding affinities, toxicity and pharmacokinetic parameters”. *Progress in Drug Research. Fortschritte Der Arzneimittelforschung. Progres Des Recherches Pharmaceutiques* 55, págs. 105-135.
- Wang, C. e Y. Zhang** (2017). “Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest”. *Journal of Computational Chemistry* 38.3, págs. 169-177. DOI: [10.1002/jcc.24667](https://doi.org/10.1002/jcc.24667).
- Wang, J. et al.** (2021). “Gaussian accelerated molecular dynamics: Principles and applications”. *WIREs Computational Molecular Science* n/a.n/a, e1521. DOI: [10.1002/wcms.1521](https://doi.org/10.1002/wcms.1521).
- Webb, B. y A. Sali** (2016). “Comparative Protein Structure Modeling Using MODELLER”. *Current Protocols in Bioinformatics* 54.1, págs. 5.6.1-5.6.37. DOI: <https://doi.org/10.1002/cpbi.3>.
- Wermuth, C. G. et al.** (1998). “Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)”. *Pure and Applied Chemistry* 70.5, págs. 1129-1143. DOI: [10.1351/pac199870051129](https://doi.org/10.1351/pac199870051129).
- Wold, S., L. Eriksson y S. Clementi** (2008). “Statistical Validation of QSAR Results”. *Chemometric Methods in Molecular Design*. John Wiley & Sons, Ltd, págs. 309-338. DOI: [10.1002/9783527615452.ch5](https://doi.org/10.1002/9783527615452.ch5).
- Woo, M.** (2019). “An AI boost for clinical trials”. *Nature* 573.7775, S100-S102. DOI: [10.1038/d41586-019-02871-3](https://doi.org/10.1038/d41586-019-02871-3).
- Wu, J. et al.** (2019). “Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb”. *Journal of Electronic Science and Technology* 17.1, págs. 26-40. DOI: [10.11989/JEST.1674-862X.80904120](https://doi.org/10.11989/JEST.1674-862X.80904120).

- Xu, H., Y. Song y Q. Yang** (2017). “A Novel High Throughput Virtual Screening Protocol to Discover New Indoleamine 2,3-Dioxygenase 1 (IDO1) Inhibitors”. *Chemical & Pharmaceutical Bulletin* 65.8, págs. 714-717. DOI: [10.1248/cpb.c16-01010](https://doi.org/10.1248/cpb.c16-01010).
- Yan, X. et al.** (2016). “Chemical Structure Similarity Search for Ligand-based Virtual Screening: Methods and Computational Resources”. *Current Drug Targets* 17.14, págs. 1580-1585. DOI: [10.2174/1389450116666151102095555](https://doi.org/10.2174/1389450116666151102095555).
- Yan, X. C. et al.** (2020). “Augmenting Hit Identification by Virtual Screening Techniques in Small Molecule Drug Discovery”. *Journal of Chemical Information and Modeling*. DOI: [10.1021/acs.jcim.0c00113](https://doi.org/10.1021/acs.jcim.0c00113).
- Yang, L. y A. Shami** (2020). “On hyperparameter optimization of machine learning algorithms: Theory and practice”. *Neurocomputing* 415, págs. 295-316. DOI: [10.1016/j.neucom.2020.07.061](https://doi.org/10.1016/j.neucom.2020.07.061).
- Yu, T. y H. Zhu** (2020). *Hyper-Parameter Optimization: A Review of Algorithms and Applications*. DOI: [10.48550/arXiv.2003.05689](https://doi.org/10.48550/arXiv.2003.05689).
- Yuan, H. et al.** (2016). “The Traditional Medicine and Modern Medicine from Natural Products”. *Molecules* 21.5. DOI: [10.3390/molecules21050559](https://doi.org/10.3390/molecules21050559).
- Zaheer, S. et al.** (2018). “Epilepsy and Cannabis: A Literature Review”. *Cureus* 10.9. DOI: [10.7759/cureus.3278](https://doi.org/10.7759/cureus.3278).
- Zanella, F., J. B. Lorens y W. Link** (2010). “High content screening: seeing is believing”. *Trends in Biotechnology* 28.5, págs. 237-245. DOI: [10.1016/j.tibtech.2010.02.005](https://doi.org/10.1016/j.tibtech.2010.02.005).
- Zhang, S. et al.** (2006). “A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models”. *Journal of Chemical Information and Modeling* 46.5, págs. 1984-1995. DOI: [10.1021/ci060132x](https://doi.org/10.1021/ci060132x).

Capítulo 3

Búsqueda de inhibidores
del canal TRPV1



3.1. Introducción

Como se describió en el Capítulo 1, la modulación de los canales TRPV1 constituye una estrategia novedosa en el desarrollo de nuevos fármacos para el tratamiento de la Epilepsia. Sin embargo, los efectos adversos relacionados con la regulación de la temperatura corporal, asociados a los denominados antagonistas de primera generación, han dificultado la aprobación de estos compuestos para su uso en pacientes (Garami et al. 2020).

En este sentido, el abordaje propuesto en esta tesis de investigación consiste en recurrir al reposicionamiento de fármacos como estrategia para encontrar moduladores del TRPV1 que no modifiquen la temperatura corporal. Los fármacos ya aprobados y que son utilizados actualmente en la clínica, han demostrado ser seguros y se conoce su perfil de efectos adversos, al menos en el rango terapéutico de su indicación original.

Luego de compilar una base de datos de compuestos evaluados frente al blanco molecular, se desarrolló y validó un modelo QSAR de tipo clasificador, capaz de identificar compuestos que antagonizan la activación del canal por CAP.

Debido a que la estructura tridimensional de la isoforma humana hTRPV1 no fue determinada experimentalmente, se recurrió al modelado por homología para poder obtener modelos del *target*. Estos modelos, junto con el *dataset* compilado previamente, se utilizaron para desarrollar y validar un modelo de docking capaz de identificar compuestos que se unen al sitio de los capsaicinoides del canal.

Finalmente, se describen las simulaciones de MD que se llevaron a cabo para los *hits* identificados durante el VS (descrito en el Capítulo 5).

3.2. Metodologías

3.2.1. Compilado y partición de la base de datos

Los datos de actividad biológica fueron recopilados de la base de datos ChEMBL26. En un principio, se consideraron únicamente datos de K_i . Sin embargo, debido a que la gran mayoría de los compuestos reportados son inhibidores relativamente potentes del canal TRPV1 ($K_i \leq 10 \mu\text{M}$), se decidió incluir también datos de IC_{50} para poder encontrar compuestos inactivos o de baja potencia inhibitoria. De este modo, se consideraron como activos aquellas moléculas con $K_i \leq 1 \mu\text{M}$ y como inactivos aquellas reportadas como “No Activos” o cuyo valor de $\text{IC}_{50} \geq 10 \mu\text{M}$. Solo se incluyeron ensayos realizados en la isoforma humana de este canal, descartando determinaciones en otras especies como cerdo, ratón o conejo. Tampoco se consideraron ensayos de inhibición donde la activación previa a la aplicación del inhibidor fuere por otro estímulo como cambios en el pH, la temperatura o el empleo de otros agonistas distintos de CAP.

Partición de la base de datos

Con el objetivo de agrupar los compuestos en *clusters*, se utilizó el algoritmo SOMoC, el cual se encuentra descrito en la sección suplementaria Algoritmos - C.1. En este caso cada clase (activos/inactivos) fue procesada por separado.

Debido a que la proporción de compuestos inactivos es considerablemente menor que la de activos, sobre estos últimos se muestrearon dos subconjuntos. El primero fue del mismo tamaño que el conjunto de inactivos, de modo de generar un *dataset* balanceado en cuanto a la distribución de clases. El segundo consistió en una muestra más pequeña de los activos, que fueron luego utilizados para crear la biblioteca simulada DUD-E. En ambos casos el muestreo se llevó a cabo de manera estratificada utilizando los *clusters* previamente identificados, de modo que preservar la diversidad química presente en el conjunto inicial.

El *dataset* balanceado fue particionado nuevamente, reservando un 80% de los datos para entrenamiento y un 20% para testeo. Esta partición también fue estratificada pero utilizando la clase (activo/inactivo), de manera de conservar la distribución de clases en ambos conjuntos.

3.2.2. Desde los ligandos

La secuencia de metodologías utilizada para el desarrollo y validación de los modelos QSAR se resume en el esquema de la Figura 2.5 y se encuentra detallada en el Capítulo 2, Sección 2.3.2.4. Por lo tanto, en este capítulo no se ahondará nuevamente en el fundamento de dichas herramientas y sólo se describirán las particularidades pertinentes al modelo desarrollado para el canal TRPV1, que fue diseñado como un modelo clasificador.

3.2.2.1. Desarrollo y validación de un modelo clasificador

Cálculo y selección de descriptores

Se utilizó el software Mordred para el cálculo de los descriptores moleculares independientes de la conformación. El proceso de selección de descriptores aplicado se encuentra detallado en el Capítulo 2, Sección 2.3.2.4. En este caso, el algoritmo de ML utilizado fue K vecinos cercanos (KNN, del inglés K-nearest Neighbors). El número de vecinos se fijó en 10 y se utilizó el método de ponderamiento por distancia, donde aquellos vecinos más cercanos contribuyen más a la predicción. Este algoritmo es simple (pocos hiperparámetros) y rápido, por lo que resulta conveniente en este caso. Su funcionamiento se detalla en material suplementario Algoritmos - C.4. La métrica a minimizar fue la pérdida logarítmica (LogLoss, del inglés Logarithmic Loss).

Elección del mejor algoritmo

Por tratarse de un problema de clasificación binaria, en este caso se evaluaron los algoritmos KNN, Light Gradient Boosting Machine (LGBM), Extra Trees Classifier (ExtraTrees), Stochastic Gradient Descent (SGD), Support Vector Machine Classifier (SVC), Logistic Regression (LR), Decision Tree Classifier (DT), Random Forest Classifier (RF), Multi Layer Perceptron Classifier (MLP) y AdaBoost Classifier (AdaBoost). Para cada uno de ellos se realizaron 100 iteraciones de 10-Fold CV en el conjunto de entrenamiento, utilizando los hiperparámetros por defecto.

3.2.3. Desde el receptor

En lo que respecta a las metodologías basadas en la estructura del blanco molecular, se combinaron diversas técnicas, resumidas en el esquema de la Figura 3.1.



Figura 3.1: Flujo de trabajo que describe las distintas metodologías, basadas en la estructura del receptor, que fueron aplicadas con el objetivo de identificar antagonistas del canal TRPV1. Reproducido y adaptado con autorización de **Llanos et al. 2022**. Copyright 2022 American Chemical Society.

3.2.3.1. Modelado del canal hTRPV1

Debido a que la estructura tridimensional de la isoforma humana del canal TRPV1 (hTRPV1) no ha sido determinada experimentalmente, se recurrió a la metodología de modelado comparativo para obtener un modelo tridimensional del blanco molecular.

La secuencia canónica del hTRPV1 se obtuvo de Uniprot (Uniprot-ID: Q8NER1). Tres moldes o *templates* fueron seleccionados para realizar el modelado, los cuales representan tres estados biológicamente relevantes del canal: la forma apo (PDB-ID: 5IRZ, resolución 3.28 Å); una conformación abierta unida al agonista Resiniferatoxina (RTX) y a una toxina peptídica (DkTx) (PDB-ID: 5IRX, resolución 2,95 Å); y una conformación cerrada en complejo con el antagonista CPZ (PDB-ID: 5IS0, resolución 3,43 Å). Todas ellas corresponden a la especie *Rattus norvegicus* (rTRPV1) y fueron determinadas en nanodiscos lipídicos mediante Cryo-EM (**Cao et al. 2013**). Adicionalmente, en los tres modelos se incluyó el dominio repetitivo de Ankirina (PDB-ID: 2PNN, resolución 2.7 Å) (**Lishko et al. 2007**) para incrementar la cobertura del modelo.

El alineamiento inicial se realizó mediante programación dinámica utilizando el algoritmo de Needleman-Wunsch (**Needleman** y **Wunsch** 1970) y luego se editó manualmente de la siguiente manera: Los segmentos comprendidos entre las posiciones 1-111, 604-627 y 753-839 fueron removidos de la secuencia *target*, debido a que los moldes no cubren estas regiones y no se encuentran en la vecindad del sitio de unión de CAP. El dominio de Ankirina se utilizó para modelar la región intracelular comprendida entre las posiciones 112-359, mientras que las estructuras de rTRPV1 cubren aproximadamente la región transmembrana. Solo una subunidad del canal fue modelada en esta etapa.

A partir de cada molde (apo, agonista, antagonista) se generaron 100 modelos con Modeller (**Eswar** et al. 2006), manteniendo los ligandos en el sitio activo durante el modelado en los dos *templates* que los contienen (agonista y antagonista). En cada caso, el mejor modelo se seleccionó en base al potencial zDOPE informado por Modeller (**Shen** y **Sali** 2006), así como también en base a la evaluación provista por el servidor Molprobit (Chen et al. 2010). Finalmente, cada modelo homotetramérico del canal se ensambló utilizando el molde correspondiente.

Cada uno de los modelos generados fue refinado con Rosetta, utilizando un protocolo adaptado para proteínas de membrana, RosettaMP (**Koehler** **Leman**, **Mueller** y **Gray** 2017). Este protocolo combina el clásico algoritmo FastRelax de Rosetta con una función de puntuación que incluye términos específicos para modelar la interacción con la membrana (*mpframework_smooth_fa.2012*), a la vez que optimiza la inserción de la proteína en una representación implícita de la membrana lipídica. El algoritmo FastRelax consiste en ciclos alternados de reconstrucción de las cadenas laterales utilizando una librería de rotámeros (*re-packing*) y minimizaciones de energía guiadas por gradiente (**Conway** et al. 2014). Los residuos embebidos en la membrana se calcularon a partir de la estructura tridimensional utilizando el servidor PPM (**Lomize** et al. 2012) y luego se convirtieron al formato span.

Las simulaciones se llevaron a cabo utilizando el *framework* de Rosetta para trabajar con sistemas simétricos, donde solo una subunidad es simulada de manera explícita, mientras que las restantes son representaciones virtuales, conectadas entre sí a través de saltos o *jumps*. En cada caso, los elementos de simetría se extrajeron del modelo inicial, las cuales a su vez provienen del *template* que se utilizó en cada caso

para ensamblar los homotetrámeros. Durante la simulación, la posición relativa de las subunidades así como también la orientación del canal con respecto a la membrana se optimizaron de manera conjunta. Esto permitió disminuir considerablemente el tiempo requerido para refinar un complejo proteico tan grande, como es el caso de este canal iónico.

A partir de cada estructura inicial se generaron 100 modelos, y en cada caso se seleccionó el que presentaba una menor energía según la función de puntuación de Rosetta. A su vez, se realizaron diversas validaciones para evaluar la calidad de los modelos obtenidos en términos geométricos y energéticos: QMEANBrane (**Studer, Biasini y Schwede 2014**) es una función de puntuación basada en potenciales estadísticos, entrenada para los tres segmentos presentes en proteínas de membrana: soluble, interfase e inserto en la membrana. Molprobit es un servidor que realiza diversas evaluaciones sobre los modelos y las combina en un *score* final (**Chen et al. 2010**).

3.2.3.2. Desarrollo y validación de un modelo de docking

Dos algoritmos de docking fueron evaluados para dockear los *datasets* en los tres modelos del hTRPV1 previamente generados: Quick Vina 2 (**Alhossary et al. 2015**) (QVina2) y Autodock4-GPU (**Santos-Martins et al. 2019**) (AD4-GPU) (Figura 3.2). QVina2 es un programa basado en Autodock Vina (**Trott y Olson 2010**), que incorpora una serie de optimizaciones basadas en heurísticas para incrementar su velocidad. Esta cualidad lo hace particularmente interesante para aplicaciones del tipo VS, donde se desean dockear grandes bibliotecas de compuestos. AD4-GPU es una nueva implementación para GPU del clásico Autodock4 (**Morris et al. 2009**), que resulta en un incremento sustancial de velocidad con respecto a la versión original (aproximadamente 40 veces más rápido), utilizando la misma función de puntuación.

Para QVina2, el parámetro de *exhaustiveness* se incrementó de 8 (*default*) a 32, y el número de conformaciones a generar se limitó a 3. El resto de los parámetros se establecieron en los valores por defecto. Para el caso de AD4-GPU, el número de corridas se aumentó de 100 (*default*) a 300. Gracias a una heurística implemen-

tada en el nuevo código de AD4-GPU, el número de evaluaciones que realiza el algoritmo genético, así como también el método de optimización local (Solis-Wet o ADADELTA) es seleccionado *on-the-fly* para cada ligando, lo que permite reducir considerablemente el tiempo de cálculo sin sacrificar precisión. A su vez, también se utilizó la opción de autostop, la cual finaliza la corrida de docking cuando los valores de energía de las mejores conformaciones del ranking no varían significativamente (Santos-Martins et al. 2019). El sitio de docking se definió como una caja rectangular de 22x22x25 Å centrada en el sitio de unión de CAP.

Como post-procesamiento de los complejos ligando-receptor obtenidos, se realizó un *re-scoring* de los mismos utilizando otras funciones de puntuación. Para ello, se utilizó el programa Smina, un software basado en el anteriormente mencionado Vina (Koes, Baumgartner y Camacho 2013). Los complejos fueron minimizados con tres funciones de puntuación diferentes: Vina, Vinardo y Ad4 (Figura 3.2). Vina se refiere a la función de puntuación original de Autodock Vina, Vinardo (Vina Optimized Radii) es una función de puntuación re-optimizada basada en Vina (Quiroga y Villarreal 2016), y Ad4 es una función de puntuación que se asemeja a la de Autodock4. Durante la minimización, las cadenas laterales de los residuos que se encuentren a una distancia menor o igual a 3 Å de cada ligando fueron consideradas como flexibles. Este procedimiento permite considerar, al menos en parte y de una manera muy simplista, la flexibilidad del sitio activo y el denominado efecto de *induced fit* asociado al evento de unión de un ligando.



Figura 3.2: Esquema de las diferentes combinaciones de modelo estructural del canal TRPV1, programa de docking y función de puntuación utilizada para el *re-scoring*, evaluadas durante el desarrollo y la validación del modelo de docking. Reproducido y adaptado con autorización de **Llanos** et al. 2022. Copyright 2022 American Chemical Society.

Docking power

Debido a que no existen estructuras del canal hTRPV1 determinadas experimentalmente, los experimentos de *re-docking* y *cross-docking* se realizaron con las estructuras de rTRPV1 utilizadas como moldes: 5IS0, 5IRX y 5IRZ. Los ligandos RTX y CPZ fueron extraídos de sus respectivos complejos, protonados a pH 7.4 con OpenBabel y luego convertidos al formato pdbqt mediante el *script* `prepare_ligand4.py` incluido en Autodock Tools, aleatorizando sus coordenadas. Posteriormente, los mismos fueron dockeados en las 3 estructuras mencionadas anteriormente, utilizando ambos protocolos de docking (QVina2 y AD4-GPU).

Scoring power

El *scoring power* o *scoring accuracy* de los modelos se estimó utilizando el *dataset* balanceado, dockeándolo en los tres modelos estructurales del hTRPV1, utilizando los distintos protocolos de docking anteriormente descritos. Para cada combinación se calculó el AUC-ROC como métrica para juzgar la *performance* de los modelos.

Screening power

El *screening power* se evaluó mediante un VS retrospectivo sobre la base simulada DUD-E. En este caso, además del AUC-ROC también se calcularon otras métricas de enriquecimiento como BEDROC ($\alpha=100$), RIE ($\alpha=100$), EF1 % y EF10 %.

3.2.3.3. Simulaciones de dinámica molecular

Luego de realizado el VS y con el objetivo de evaluar la interacción de los *hits* seleccionados (Ver Capítulo 5 Sección 5.1.3) con el blanco molecular, se llevaron a cabo simulaciones de MD tomando como punto de partida los complejos ligando-receptor obtenidos mediante el docking.

En la estructura homotetramérica del canal TRPV1 existen 4 sitios de unión de CAP, los cuales son idénticos entre sí ya que están formados por la interfaz entre 2 subunidades contiguas. Gracias a que nuestro modelo del hTRPV1 fue refinado con Rosetta de manera simétrica, se logró preservar la simetría de estos sitios. Sin embargo, las simulaciones de docking se realizaron sobre uno solo de ellos. Por lo tanto, las coordenadas de cada ligando (correspondientes a la pose de menor energía generada mediante el docking) fueron replicadas en los restantes sitios unión de CAP, de manera de ocupar los 4 sitios homólogos. Este complejo de estequiometría 1:4 se utilizó como punto de partida para las simulaciones de MD, lo cual nos permitió tener 4 réplicas de la trayectoria de un ligando dentro de cada simulación (Figura 3.3).

Simular múltiples trayectorias cortas ha demostrado mejorar la capacidad de muestreo o *sampling* de la MD, a la vez que permite estimar diversos parámetros o constantes con mayor precisión, comparado con una única trayectoria larga (Perez, Tomas y Rubio-Martinez 2016; Pranami y Lamm 2015). En este caso, se realizaron 4 corridas independientes de 25 ns cada una, totalizando unos 400 ns por ligando (16 x 25 ns).

El sistema se ensambló utilizando el servidor CHARMM-GUI (Jo et al. 2008; Wu et al. 2014). El complejo ligandos-receptor se insertó en un parche de 1-palmitoil-2-oleoil-sn-glicero-3-fosfolina (POPC), se solvató con una caja rectangular de aguas

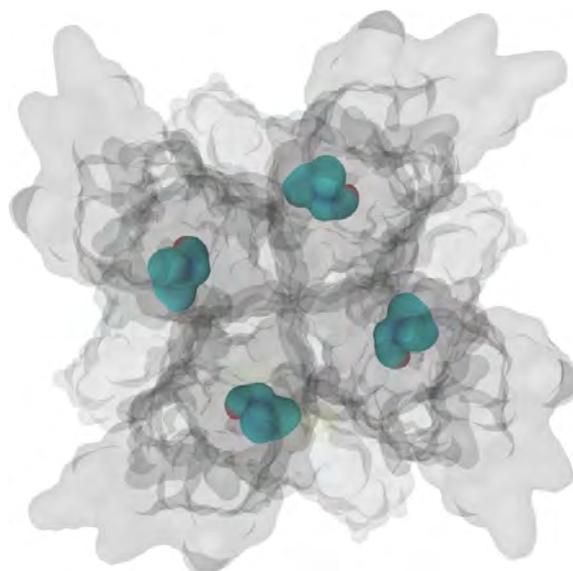


Figura 3.3: Vista superior del modelo homotetramérico del canal hTRPV1 basado en el molde apo (5IRZ), donde pueden observarse las 4 copias del ligando MTK (cyan) ocupando los 4 sitios de unión de CAP. Reproducido y adaptado con autorización de **Llanos** et al. [2022](#). Copyright 2022 American Chemical Society.

OCP y se adicionó suficiente NaCl hasta emular aproximadamente una concentración de 0.15 mM. El tamaño final de cada sistema ensamblado fue de aproximadamente 372.000 átomos. Las simulaciones se realizaron con el software Amber20, utilizando el campo de fuerza ff19SB para la proteína, Lipids17 para los lípidos de membrana y GAFF2 para los ligandos. Inicialmente el sistema se relajó mediante 2500 pasos de minimización utilizando el algoritmo de *steepest descent*, seguidos por 2500 de gradiente conjugado. Luego, las fases de equilibración y calentamiento se llevaron a cabo en seis etapas consecutivas, iniciando en un ensamble NVT y luego cambiando a un NPT, a medida que se liberan progresivamente los *restraints* sobre la proteína y los lípidos. Las simulaciones se realizaron a una temperatura de 303.15 K utilizando un termostato de acople débil, mientras que la presión del sistema se controló mediante un barostato de Berendsen con acople semi-isotrópico. Las interacciones electrostáticas de largo alcance se calcularon mediante el algoritmo PME con un *cutoff* de 10 Å. En la fase de producción, se utilizó un tiempo de integración de 4 fs, combinando los algoritmos de SHAKE y Hydrogen Mass Repartitioning (**Balusek** et al. [2019](#); **Hopkins** et al. [2015](#)).

3.3. Resultados

3.3.1. Compilado y partición de la base de datos

Inicialmente, se compilaron un total de 4545 datos de actividad, incluyendo determinaciones de IC₅₀ y Ki. Luego del curado, se generó un *dataset* que consta de 989 moléculas, 664 clasificadas como activas y 325 como inactivas (Figura 3.5).

Este conjunto de datos fue agrupado utilizando el algoritmo de *clustering* SOMoC, analizando cada clase por separado. En la Figura 3.4-A se muestra el gráfico de *elbow* generado por SOMoC durante la etapa de *clustering*, el cual fue utilizado para seleccionar el número óptimo de *clusters* (K) dentro de las moléculas activas. El valor de K se varió entre 2 y 25, mientras que los demás parámetros se mantuvieron en sus valores por defecto. Se seleccionó finalmente un K=9, lo cual resulta en un SIL=0.978. El SIL es una métrica que varía entre -1 y 1, siendo este último valor el que indica un agrupamiento óptimo (Ver Material suplementario Algoritmos - C.1). Todos los CVIs (del inglés, Cluster Validity Indexes) calculados por SOMoC se resumen en la Tabla A.1. En base a los mismos, podemos decir que se obtuvieron *clusters* densos y bien separados.

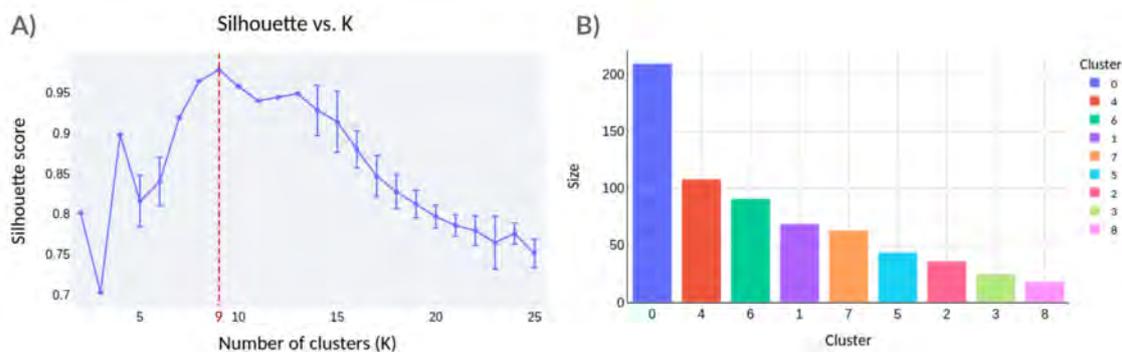


Figura 3.4: Validación interna del *clustering*. A) Gráfico de *elbow* generado por SOMoC, utilizado para seleccionar el número óptimo de *clusters* (K) en el conjunto de moléculas activas. En el eje X se representa el número K, mientras que en el eje Y se muestra el SIL *score*. Las barras verticales representan la desviación estándar de 10 iteraciones independientes de *clustering*. B) Distribución de tamaño para los *clusters* generados por SOMoC (K=9). El eje X muestra cada uno de los clusters identificados y el eje Y su tamaño.

Utilizando este *clustering*, se tomaron dos muestras estratificadas de los compuestos activos. La primera (325) se agrupó con el total de inactivos para generar

así el *dataset* balanceado, que contiene un total de 650 estructuras. La segunda muestra fue de 50 compuestos activos, los cuales se utilizaron como semillas o *seeds* para generar la biblioteca simulada DUD-E, que finalmente contiene 2600 *decoys* y 50 activos verdaderos (Figura 3.5).

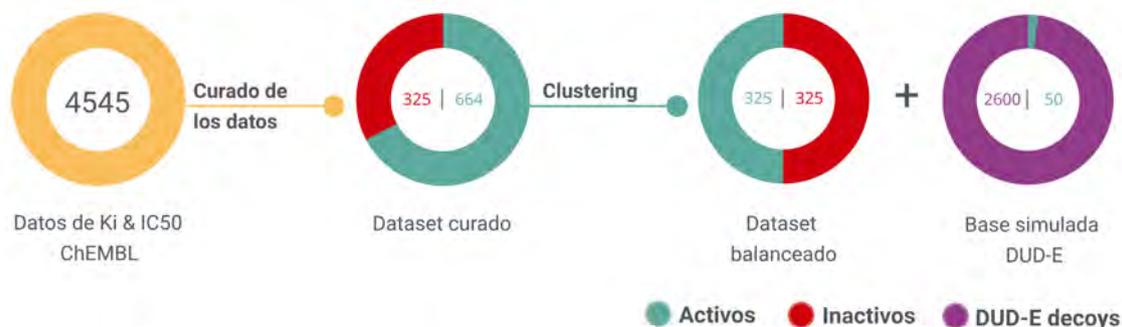


Figura 3.5: Tamaño y composición de los distintos conjuntos de datos compilados para desarrollar y validar los modelos QSAR y de docking para el canal TRPV1. Reproducido y adaptado con autorización de **Llanos** et al. 2022. Copyright 2022 American Chemical Society.

Las estructuras que componen el *dataset* balanceado y la base simulada DUD-E, en formato SMILES, pueden encontrarse en el material suplementario de la publicación correspondiente (**Llanos** et al. 2022).

3.3.2. Desde los ligandos

3.3.2.1. Desarrollo y validación de un modelo clasificador

Cálculo y selección de descriptores

Luego de particionar el *dataset* balanceado de manera estratificada, se obtuvo un conjunto de prueba de 130 estructuras y un conjunto de entrenamiento de 520 moléculas. Sobre este último, se calcularon 1613 descriptores independientes de la conformación mediante el software Mordred. Luego de aplicar los filtros de varianza y correlación lineal, un total de 371 descriptores fueron sometidos a la etapa de selección mediante el método de envoltura.

Al finalizar las 10 iteraciones de K-fold CV se evaluaron un total de 882.707

combinaciones de entre 10 y 30 descriptores, y finalmente el ranking de importancia de variables se construyó considerando las mejores 1000 soluciones en términos de *LogLoss*. En el gráfico de la Figura 3.6 se muestran los primeros 50 descriptores de este ranking, siendo estos los que se utilizaron para entrenar los modelos en la siguiente etapa. Este número de descriptores representa una relación ejemplos/predictores de aproximadamente 10, con lo cual se disminuye la probabilidad de sobreajuste.

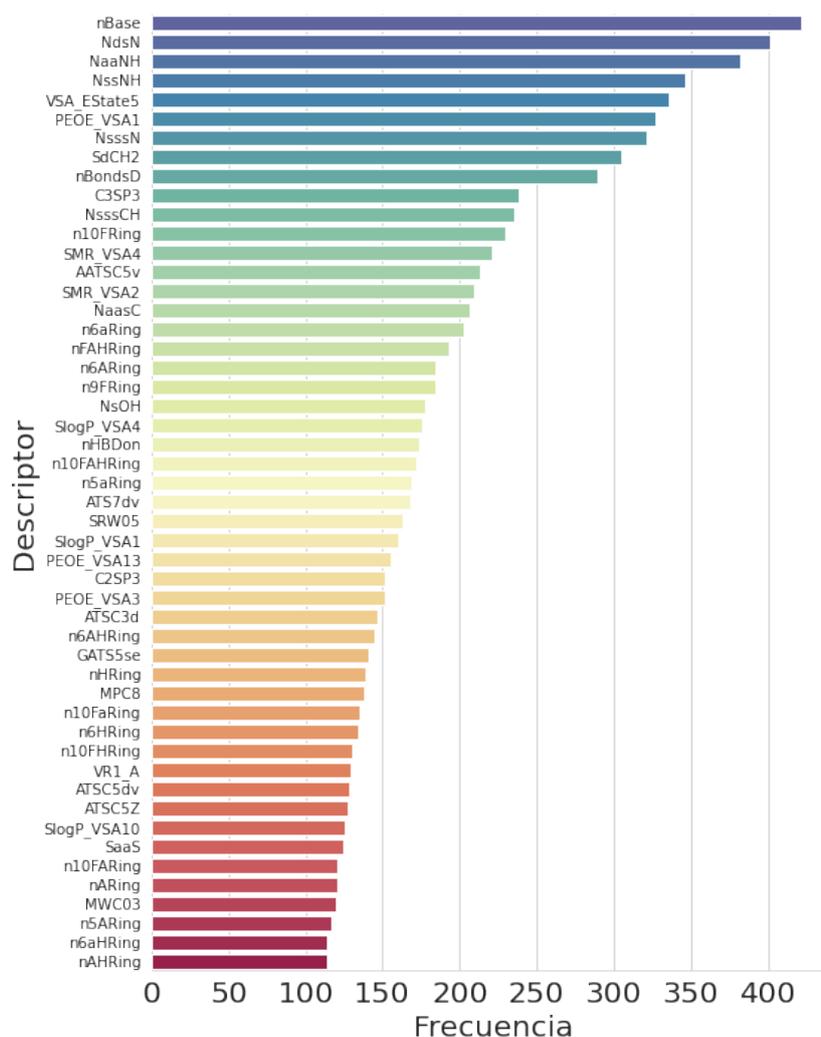


Figura 3.6: Ranking de importancia de variables. Primeros 50 descriptores del ranking generado mediante la metodología de selección desarrollada, basada en una metaheurística de tipo evolutiva. La escala de color representa la importancia de cada descriptor, desde el azul (más importante) al rojo (menos importante).

Selección del mejor algoritmo

El proceso de selección consistió en realizar 100 iteraciones de 10-fold CV sobre

el conjunto de entrenamiento, utilizando diversos algoritmos de ML. En la Figura 3.7 se muestran los resultados de dicha validación donde puede observarse que, en líneas generales, todos los algoritmos presentan un buen desempeño, medido en términos de AUC-ROC. En la Tabla A.2 se detalla, para cada uno de los algoritmos, los resultados numéricos obtenidos para diversas métricas de clasificación como ROC-AUC, Accuracy, Kappa, Precision, Recall y F1, donde puede observarse una tendencia similar en cuanto al desempeño de los mismos.

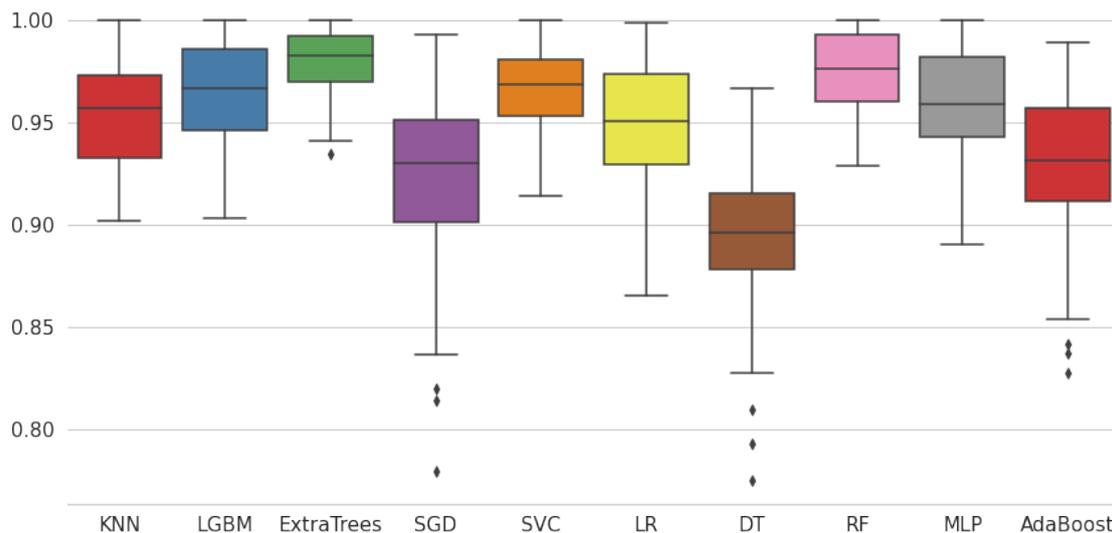


Figura 3.7: Selección del mejor algoritmo. Gráfico de cajas que muestra los valores de AUC-ROC obtenidos mediante 100 iteraciones de 10-fold CV sobre el conjunto de entrenamiento, para todos los algoritmos evaluados, utilizando los hiperparámetros por defecto en cada caso. El límite inferior y superior de cada caja representa el primer y tercer cuartil, respectivamente, mientras que la línea media representa la mediana o segundo cuartil. Las líneas que se continúan de las cajas (*whiskers*) contienen el resto de la distribución, a excepción de los valores atípicos u *outliers* que se representan como puntos.

A su vez, se realizaron 1000 iteraciones de aleatorización de Fisher y se observó que todos los modelos generados presentan un desempeño igual o peor que el azar (Tabla A.3).

Optimización de hiperparámetros

En base a los resultados obtenidos en la etapa exploratoria previa, se decidió

avanzar en la optimización de los algoritmos SVC, RF, ExtraTrees y KNN. Para cada uno de ellos, se realizaron 3 x 1000 iteraciones de búsqueda, reiniciando la semilla o *seed* cada vez para evitar que el algoritmo de búsqueda quede estancado en alguna región particular del espacio.

A modo de ejemplo, en la Figura 3.8 se resume el resultado de la optimización para el algoritmo ExtraTrees. Puede observarse que, con el transcurso de las sucesivas iteraciones, el algoritmo tiende a encontrar mejores soluciones (menor LogLoss). A su vez, a partir de las distribuciones obtenidas para los hiperparámetros, puede observarse que la búsqueda se centra preferentemente sobre ciertas regiones del hiperespacio, las cuales generan mejores soluciones. Los resultados obtenidos para los algoritmos SVC, RF y KNN pueden encontrarse en el material suplementario, Figuras A.1, A.2 y A.3, respectivamente.

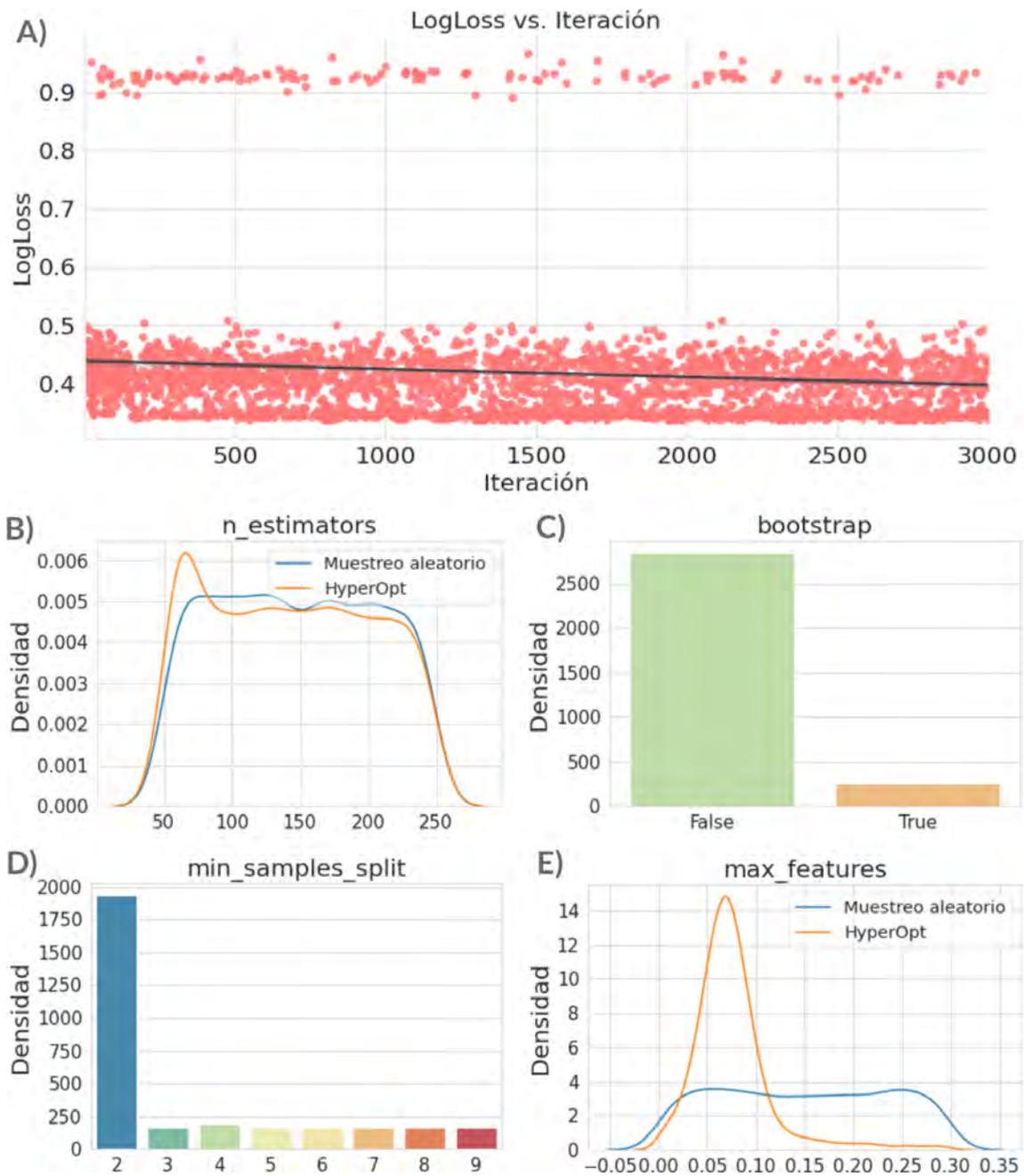


Figura 3.8: Optimización del algoritmo ExtraTrees. En el panel superior (A) se muestra como varía la métrica de pérdida (LogLoss) a medida que avanza la búsqueda. La línea negra representa una regresión lineal entre los valores de x e y , mientras que el sombreado representa un intervalo de confianza del 95 %. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: `n_estimators` (B), `bootstrap` (C), `min_samples_split` (D) y `max_features` (E). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

Validación *in silico*

Luego de optimizar cada uno de los algoritmos, los mismos fueron sometidos a las distintas validaciones internas y externas descritas en el Capítulo 2 - Sección 2.3.2.3.

Tabla 3.1: Validación interna K-fold CV de los algoritmos optimizados. Para cada métrica se informa la media y el desvío estándar (SD) de 100 x 10-Fold CV.

Modelo	ROC-AUC		Accuracy		Kappa		Precision		Recall		F1	
	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
ExtraTrees	0.979	0.017	0.943	0.029	0.887	0.057	0.932	0.040	0.963	0.033	0.946	0.027
KNN	0.962	0.023	0.909	0.035	0.820	0.070	0.872	0.046	0.972	0.029	0.918	0.030
RF	0.978	0.019	0.943	0.029	0.888	0.057	0.930	0.041	0.966	0.032	0.947	0.027
SVC	0.972	0.022	0.941	0.029	0.884	0.057	0.926	0.040	0.967	0.032	0.945	0.026

Los resultados de la validación interna K-fold CV (Tabla 3.1) muestran que, en líneas generales, todos los modelos presentan un buen desempeño, a juzgar por el conjunto de métricas calculadas y la pequeña SD asociada a cada una de ellas. En cuanto al test de aleatorización de Fisher, nuevamente se observa que la capacidad predictiva de los modelos aleatorizados es prácticamente nula, con lo que se minimiza la posibilidad de generación de buenos modelos por azar. Su desempeño es incluso peor que el de los modelos iniciales, sin optimización de hiperparámetros (Tabla A.4).

Con respecto a las validaciones externas podemos observar que los modelos presentan un buen desempeño en ambos conjuntos externos, el de testeo y la base simulada DUD-E. Se incluyó adicionalmente, a fines comparativos, la evaluación de los mismos en el conjunto de entrenamiento (*Train*) (Tabla 3.2). En un escenario ideal, libre de sobreajuste, la *performance* de los modelos en los conjuntos de entrenamiento y validación externa no debería variar sustancialmente. En este caso, el desempeño de los modelos en la validación interna K-fold CV y las validaciones externas es muy similar, lo cual evidencia un bajo grado de sobreajuste.

Tabla 3.2: Validación externa de los algoritmos optimizados. Para cada métrica se expresa la media y el desvío estándar (SD), calculados mediante 1000 iteraciones de *bootstrapping*. Adicionalmente se incluye el desempeño en el conjunto de entrenamiento (*Train*).

Model	Set	Accuracy		ROC-AUC		F1		Kappa		Precision		Recall	
		Media	SD	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
ExtraTrees	DUD-E	0.701	0.006	0.988	0.005	0.568	0.012	0.557	0.013	0.962	0.011	0.404	0.012
ExtraTrees	Test	0.968	0.014	0.986	0.009	0.968	0.014	0.933	0.029	0.986	0.013	0.951	0.023
ExtraTrees	Train	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
KNN	DUD-E	0.518	0.000	0.921	0.009	0.071	0.001	0.036	0.001	0.969	0.010	0.037	0.000
KNN	Test	0.939	0.018	0.964	0.019	0.938	0.018	0.865	0.041	0.986	0.014	0.895	0.030
KNN	Train	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
RF	DUD-E	0.679	0.005	0.987	0.005	0.523	0.011	0.510	0.012	0.969	0.010	0.359	0.010
RF	Test	0.976	0.011	0.977	0.014	0.975	0.012	0.947	0.026	1.000	0.000	0.951	0.022
RF	Train	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
SVC	DUD-E	0.699	0.006	0.994	0.002	0.565	0.012	0.554	0.013	0.972	0.009	0.399	0.012
SVC	Test	0.981	0.010	0.987	0.012	0.981	0.010	0.960	0.023	1.000	0.000	0.963	0.020
SVC	Train	0.980	0.006	0.998	0.002	0.980	0.006	0.958	0.012	0.993	0.005	0.968	0.010

Finalmente, a partir de las predicciones sobre la base simulada DUD-E se calcularon varias métricas propias del VS, para evaluar el *screening power* de los modelos. Con excepción de KNN, todos tuvieron un excelente desempeño en términos de enriquecimiento de activos, a juzgar por las métricas específicas como BEDROC($\alpha=100$), RIE($\alpha=100$) y EF1 % (Tabla 3.3).

Tabla 3.3: Evaluación del *screening power* para los algoritmos optimizados. La campaña de VS retrospectivo se realizó sobre la base simulada DUD-E. La media y el desvío estándar (SD) se calcularon mediante 1000 iteraciones de *bootstrapping*.

Model	Set	ROC-AUC		BEDROC($\alpha=100$)		RIE($\alpha=100$)		EF1 %	
		Media	SD	Media	SD	Media	SD	Media	SD
ExtraTrees	DUD-E	0.988	0.005	0.982	0.005	44.776	0.207	54.114	0.000
KNN	DUD-E	0.921	0.009	0.445	0.028	20.295	1.273	21.337	1.983
RF	DUD-E	0.987	0.005	0.980	0.005	44.694	0.213	54.114	0.000
SVC	DUD-E	0.994	0.002	0.949	0.009	43.274	0.399	52.097	0.827

En la Figura 3.9 se muestran las curvas ROC obtenidas para los modelos optimizados sobre ambos conjuntos externos, el de testeo y la base simulada DUD-E.

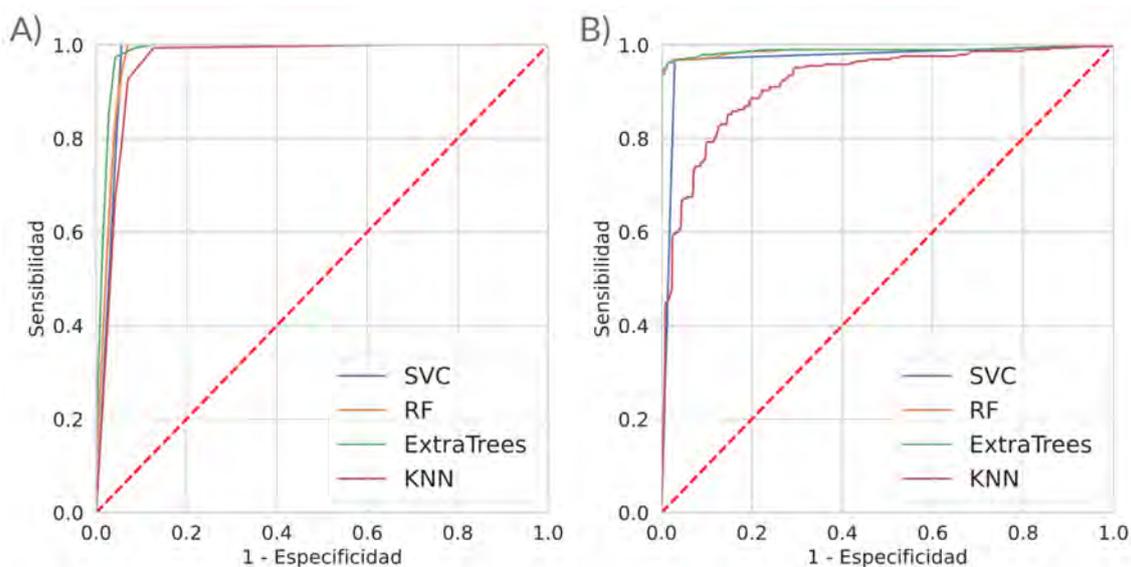


Figura 3.9: Curvas ROC obtenidas para los distintos modelos optimizados. A) Conjunto de testeo. B) Base simulada DUD-E.

En base a los resultados obtenidos, se seleccionó el algoritmo ExtraTrees para realizar el VS prospectivo, debido al buen desempeño que demostró en todas las validaciones realizadas. El fundamento de este algoritmo se detalla en la material suplementario Algoritmos - [C.3.2](#).

3.3.3. Desde el receptor

3.3.3.1. Modelado del canal hTRPV1

En la Figura [3.10](#) se observa la arquitectura final de los modelos obtenidos, los cuales comprenden la región transmembrana del canal así como también el dominio intracelular de ankirina, dispuestos en un arreglo homotetramérico. Luego del refinamiento en Rosetta, se obtuvieron modelos de alta calidad en términos geométricos y energéticos, evidenciada por las distintas evaluaciones como Molprobit y QMEAN-Brane.

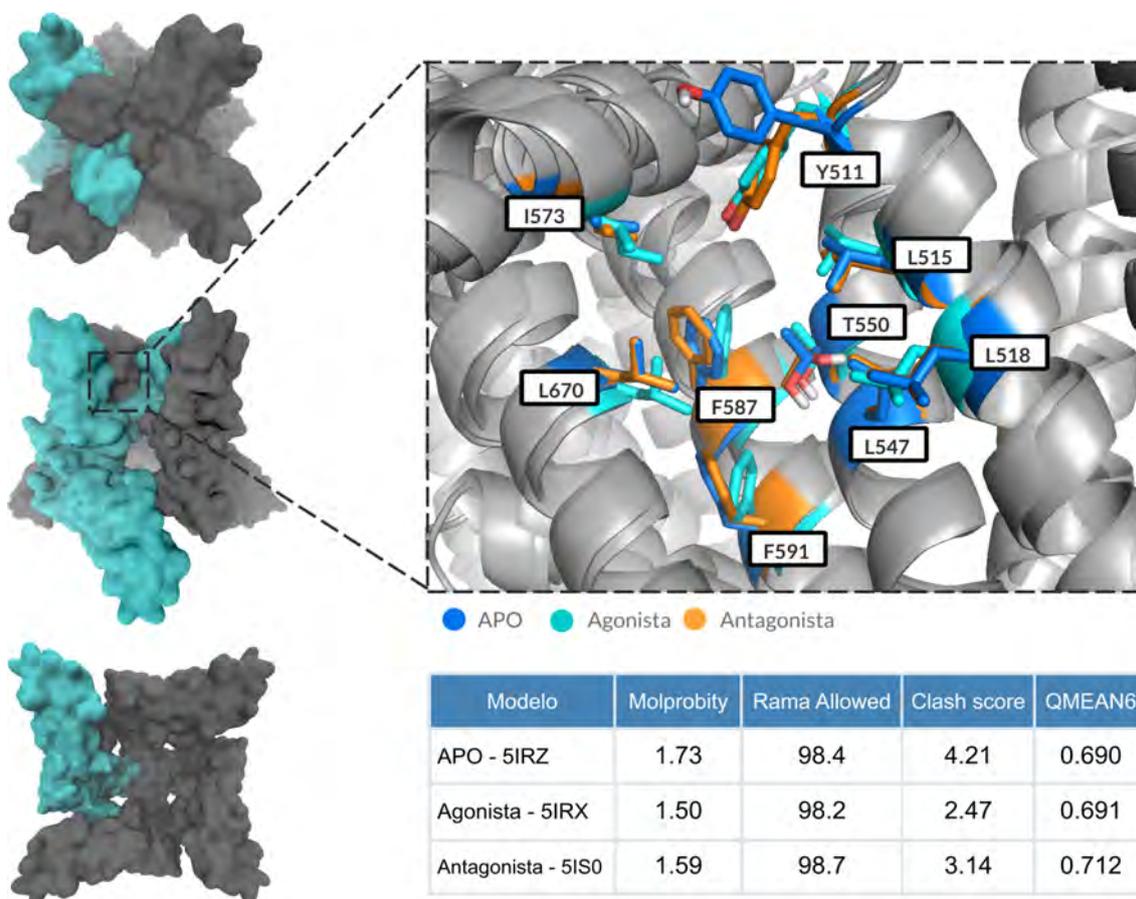


Figura 3.10: Vistas superior, lateral e inferior del modelo apo hTRPV1 en su forma de homotetrámero, donde la subunidad monomérica es destacada en cian mientras que las restantes se muestran en color gris. A la derecha, un zoom dentro del sitio de unión de CAP muestra los residuos identificados como determinantes para la unión de capsaicinoides, destacando las diferencias que existen entre los modelos apo, agonista y antagonista. Solo se muestran los hidrógenos polares. En la tabla inferior se resumen las métricas calculadas para la validación de los distintos modelos tridimensionales. La métrica QMEAN6 score oscila entre 0 y 1, siendo este último valor el que indica un modelo de mejor calidad. Para el caso del Molprobity score, valores cercanos a 0 indican un mejor modelo. Reproducido y adaptado con autorización de **Llanos et al. 2022**. Copyright 2022 American Chemical Society.

A través de estudios mutacionales en canales TRPV1 de rata (**Lee et al. 2011**) y humano (**Ohbuchi et al. 2016**), se han identificado varios residuos considerados importantes para la unión de ligandos en el sitio de CAP: Y511, L515, L518, L547, T550, I573, F587, F591 y L670. En el recuadro de la Figura 3.10, puede observarse una superposición de los tres modelos obtenidos, resaltando estos residuos. La mayor

diferencia se observa en la orientación de la Y511, la cual apunta hacia afuera del sitio activo en el modelo apo (al igual que en su respectivo molde, 5IRZ) o hacia adentro en ambos modelos holo (molde 5IS0 y 5IRX). Adicionalmente, en el modelo derivado del complejo experimental con el antagonista CPZ (molde 5IS0), la T550 presenta un rotámetro diferente en comparación con el modelo derivado del complejo con RTX (molde 5IRX) y el apo (5IRZ).

3.3.3.2. Desarrollo y validación de un modelo de docking

Docking power

Los resultados de los experimentos de re-docking se muestran en la Figura 3.11. Para el caso de AD4-GPU, se analizaron la pose de mejor energía (LE) y la pose de menor energía del cluster más numeroso (LC_LE), mientras que para QVina2 se analizaron las 3 primeras poses del ranking. Puede observarse que la *performance* de QVina2 es ligeramente superior a la de AD4-GPU en términos de *docking power*, aunque ambos programas presentan valores de RMSD menores a 2 Å, un valor de umbral comúnmente utilizado para definir dos poses como equivalentes. En los experimentos de cross-docking, las predicciones para las estructuras holo siguen siendo buenas, aunque para la estructura apo los valores de RMSD son considerablemente mayores en ambos algoritmos (Figura 3.11-A). Este comportamiento puede ser atribuido al efecto de *induced fit* desencadenado por la unión de ciertos ligandos a un receptor. Un claro ejemplo de este fenómeno se observa en la Y511, la cual presenta una conformación diferente entre las estructuras apo y holo (Figura 3.10).

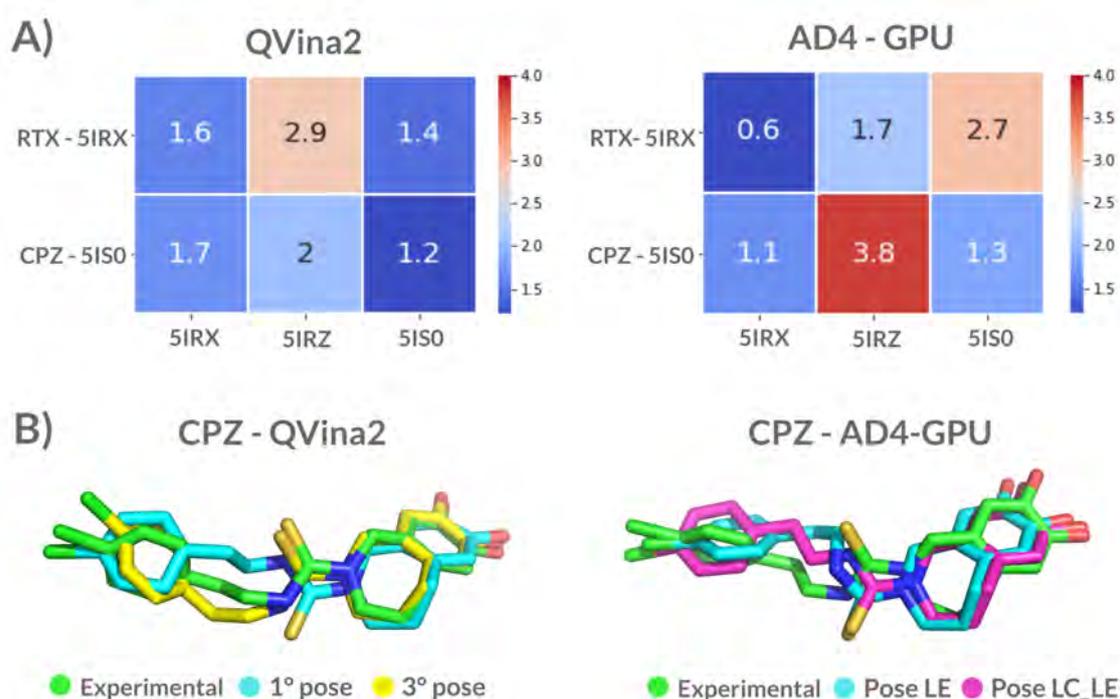


Figura 3.11: Evaluación del *docking power*. A) Valores de RMSD (Å) expresados como *heatmaps* para los experimentos de *re-docking* y *cross-docking* utilizando QVina2 o AD4-GPU. B) Poses obtenidas para el re-dockeo del antagonista CPZ con QVina2 (izquierda) o AD4-GPU (derecha), superpuestas con la pose de unión determinada experimentalmente. Reproducido y adaptado con autorización de **Llanos et al. 2022**. Copyright 2022 American Chemical Society.

Cabe destacar que, basándonos únicamente en los valores de RMSD obtenidos, el re-docking del antagonista CPZ utilizando QVina2 (1.2 Å) o AD4-GPU (1.3 Å) parece ser muy bueno (Figura 3.11-A). Sin embargo, en una inspección visual de esas poses se puede observar que el grupo tiourea, el cual interacciona con la T550 en la estructura experimental, se encuentra mal orientado. Para el caso de AD4-GPU, el grupo tiourea se encuentra rotado en ambas poses LE y LC.LE, aun cuando la orientación global de la molécula es correcta. Esto pone de manifiesto ciertas falencias que tiene el RMSD como métrica para evaluar el *docking power* de un programa, cuando se lo utiliza como único criterio. Con respecto a Qvina2, en la tercera pose generada este grupo se orienta correctamente, lo que demuestra que aún cuando el programa es capaz de encontrar la pose correcta, la función de puntuación falla en asignarle el menor *score* global (Figura 3.11-B). Por lo tanto, las 3 poses generadas por QVina2 se sometieron al *re-scoring* con Smina. Esta minimización

flexible produjo movimientos muy pequeños ($< 0.5 \text{ \AA}$) en las poses iniciales de los ligandos.

Scoring power

Utilizando el *dataset* balanceado de 350 compuestos activos y 350 compuestos inactivos (Figura 3.5), se construyeron curvas ROC para cada combinación de modelo tridimensional del canal, programa de docking y función de puntuación utilizada para re-scorear los complejos ligando-receptor.

Analizando los valores de AUC-ROC obtenidos para cada combinación, puede observarse que el modelo derivado del agonista (5IRX) tiene el peor desempeño, sin importar el programa que se utilice para dockear los compuestos. Luego existen algunas diferencias entre los modelos 5IRZ y 5IS0, según se utilice QVina2 o AD4-GPU para dockear, y según qué función de puntuación se utilice para el *re-scoring*. Sin embargo, en todos los casos puede observarse que el *re-scoring* con la función ad4 mejora la capacidad de los modelos para diferenciar compuestos activos de inactivos (Figura 3.12).

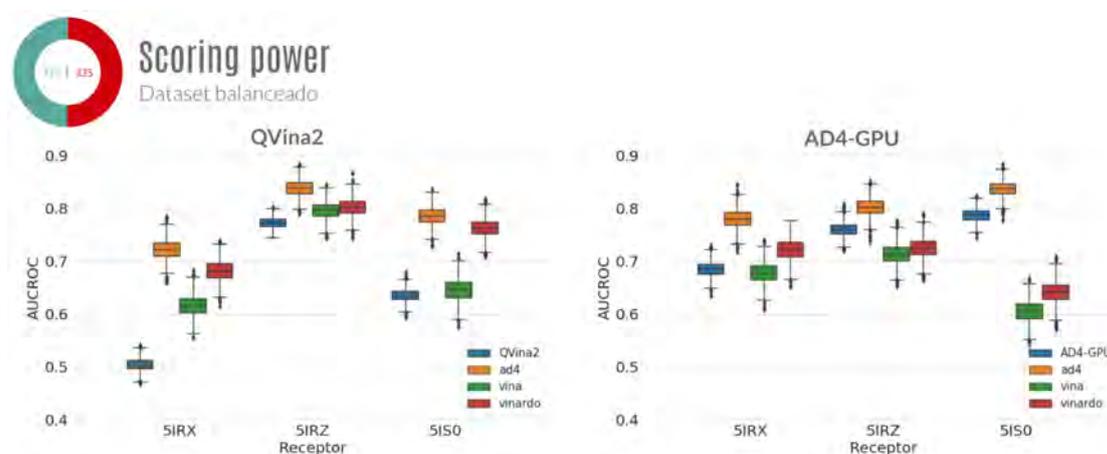


Figura 3.12: Evaluación del *scoring power*. Valores de AUC-ROC obtenidos luego de dockear el *dataset* balanceado en cada modelo de canal, utilizando ambos programas de docking (QVina2 y AD4-GPU), y su posterior *re-scoring* con las funciones de puntuación: vina, vinardo y ad4. La línea horizontal dentro de cada caja representa la mediana de la distribución, mientras que los puntos fuera del rango son considerados *outliers*. Reproducido y adaptado con autorización de Llanos et al. 2022. Copyright 2022 American Chemical Society.

Con respecto al post procesamiento, el *re-scoring* de los complejos precedido por una minimización con cadenas laterales flexibles, en líneas generales, mejora la capacidad de los modelos para diferenciar compuestos activos de inactivos. En particular, la función de puntuación ad4 permite, en todos los casos, mejorar el *scoring power* de los modelos.

Los valores de AUC-ROC para todas las combinaciones pueden encontrarse en el material suplementario (Tabla A.5).

Screening power

Utilizando la biblioteca simulada DUD-E, se realizaron cribados retrospectivos con cada modelo de docking para evaluar su comportamiento en un escenario que intenta modelar un caso de aplicación real (pocos compuestos activos en una gran cantidad de compuestos inactivos), al cribar una base de datos de manera prospectiva.

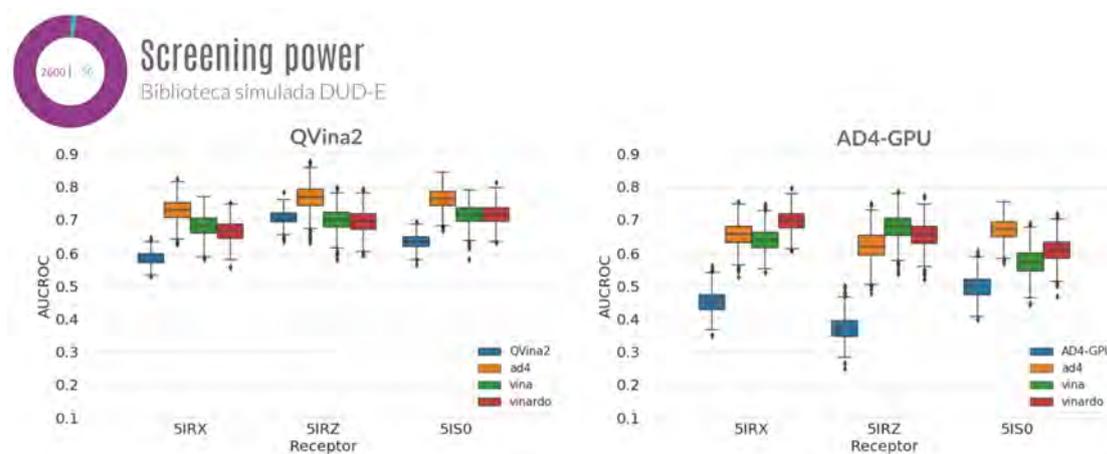


Figura 3.13: Evaluación del *screening power*. Valores de AUC-ROC obtenidos luego de dockear el la biblioteca simulada DUD-E en cada modelo de canal, utilizando ambos programas de docking (QVina2 y AD4-GPU), y su posterior *re-scoring* con las funciones de puntuación: vina, vinardo y ad4. La línea horizontal dentro de cada caja representa la mediana de la distribución, mientras que los puntos fuera del rango son considerados *outliers*. Reproducido y adaptado con autorización de Llanos et al. 2022. Copyright 2022 American Chemical Society.

Con respecto al modelo estructural del hTRPV1 se observa una tendencia similar, donde el modelo derivado de agonista (5IRX) presenta un pobre desempeño en

términos de AUC-ROC. Analizando la influencia del programa utilizado para dockear los compuestos, podemos observar que QVina2 funciona mejor que AD4-GPU, sin importar que modelo de hTRPV1 se utilice. En todos los casos el *re-scoring* con cadenas flexibles mejora el screening power de los modelos, siendo una vez más ad4 la función de puntuación que mejores resultados arroja (Figura 3.13).

Las métricas de enriquecimiento, calculadas para todas las combinaciones, pueden encontrarse en el material suplementario (Tabla A.6).

Analizando conjuntamente las validaciones de *docking power*, *scoring power* y *screening power*, se decidió que el mejor modelo consiste el dockear los compuestos con el programa QVina2, utilizando el modelo por homología derivado del molde apo (5IRZ), y luego como post-procesamiento el *re-scoring* con Smina utilizando la función de puntuación ad4, considerando las cadenas laterales a menos de 3 Å de cada ligando como flexibles. Las coordenadas atómicas de este modelo en formato PDB se encuentran disponibles en la sección suplementaria de la publicación correspondiente (Llanos et al. 2022).

Utilizando este modelo, se procedió a analizar secuencialmente distintos valores de docking *score* para clasificar a los compuestos como activos/inactivos, calculando para cada uno de ellos diferentes métricas de clasificación binaria (sobre el *dataset* balanceado). De esta manera, es posible encontrar un *cutoff* óptimo que nos permita utilizar este modelo como un clasificador binario en campañas prospectivas de VS. A partir de este análisis, se eligió 46 como valor de docking *score* para clasificar a un compuesto *query* como activo, para el cual se obtiene un F1=0.776, Recall=0.762, Precision=0.789, Matthews (MCC)=0.535, Kappa=0.535 (Figura 3.14).

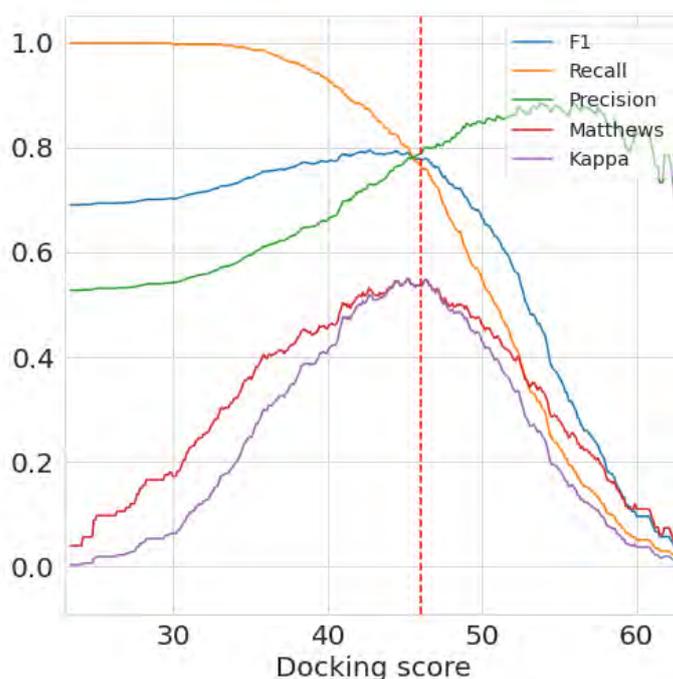


Figura 3.14: Selección del *cutoff*. Gráfica utilizada para seleccionar el docking *score* óptimo para utilizar el modelo como un clasificador binario. En el eje Y se muestran diferentes métricas de clasificación binaria como F1, Recall, Precision, Matthews (MCC) y Kappa, calculadas para cada valor de *score* (eje X). La línea roja punteada señala el *cutoff* elegido (46).

3.3.3.3. Simulaciones de dinámica molecular

A continuación se muestran los resultados más relevantes de las simulaciones, resumidos en la Figura 3.15 para cada uno de los candidatos. Mediante las gráficas de RMSD se confirma que los complejos se mantienen estables durante toda la producción, etapa de la simulación que proporciona la información de las interacciones entre los candidatos y el blanco molecular.

Montelukast

Inicialmente, el grupo carboxilato de MTK se encuentra formando un puente de hidrógeno con la Ser512, muy cercano a un puente salino entre los residuos Arg557 y Glu570. Durante la simulación, este carboxilato es capaz de tomar el lugar de Glu570 y formar un puente salino con la proteína, el cual se mantiene estable durante gran

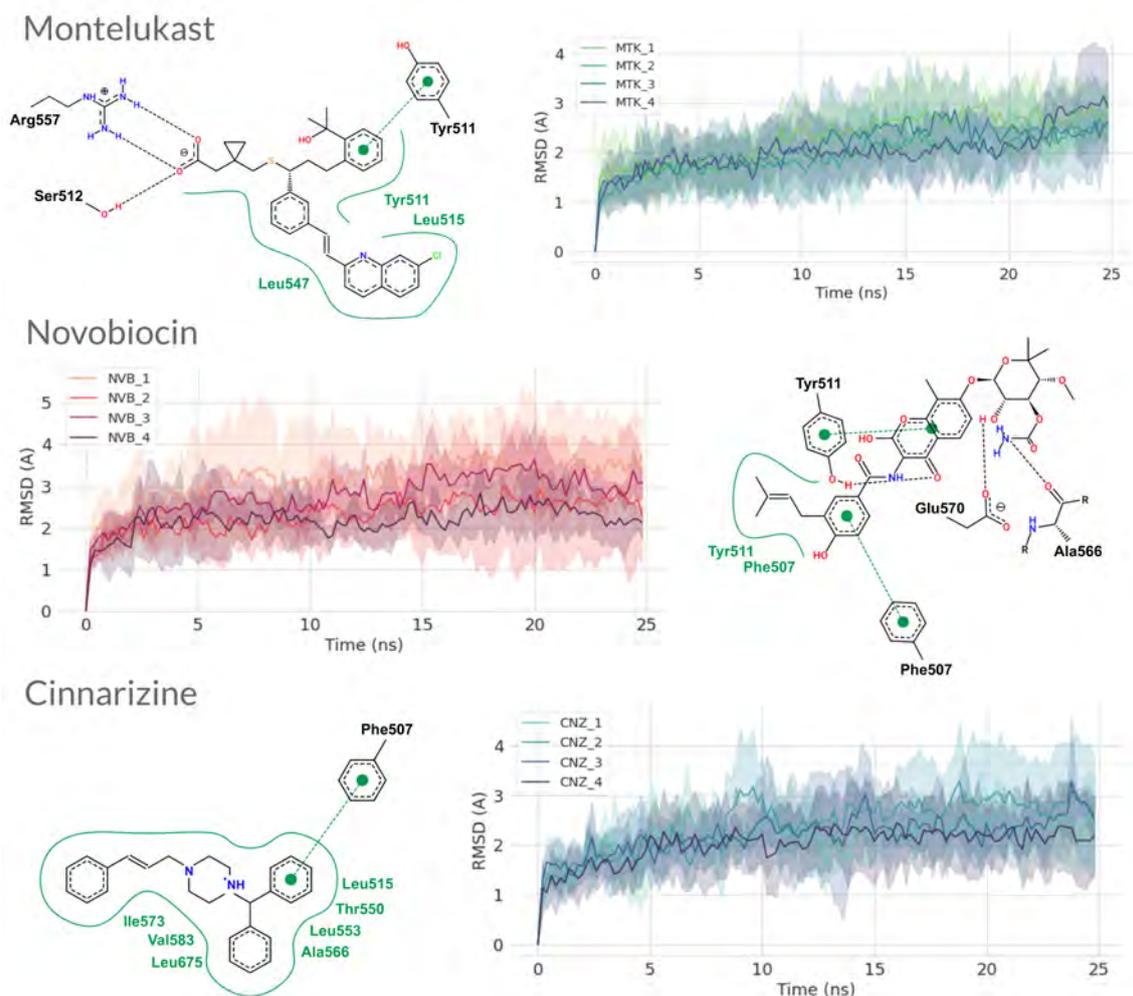


Figura 3.15: Simulaciones de dinámica molecular. Para cada uno de los candidatos seleccionados, se muestra el gráfico de RMSD (Å) vs. tiempo (ns) y un diagrama 2D de interacciones, generado a partir de un *snapshot* representativo extraído de la simulación. Las líneas verdes sólidas muestran las interacciones de tipo hidrofóbicas, y las de tipo π -stacking se representan como círculos unidos por líneas punteadas. En los gráficos de RMSD, la línea sólida representa la media y el sombreado la desviación estándar, calculadas para cada copia del ligando agrupando las 4 trayectorias independientes. Reproducido y adaptado con autorización de **Llanos et al.** 2022. Copyright 2022 American Chemical Society.

parte de la simulación. La distancia media entre el carbono Cz de la Arg557 y el carbono carbonílico del MTK es de 4.07 ± 0.21 Å, calculada sobre toda la simulación y para todas las copias del ligando (Figura A.4). A su vez, MTK forma un π -stacking paralelo con el residuo de Tyr511 en el sitio activo (Figura 3.15).

Novobiocina

Durante la simulación, puede observarse un puente de hidrógeno persistente entre el grupo -OH de la Tyr511 y el oxígeno cumarínico de la NVB que actúa

como receptor (Figura 3.15). La distancia media entre ambos átomos de oxígeno a lo largo de toda la simulación es de 4.41 ± 1.48 Å (Figura A.5-A). A su vez, también se observa otro puente de hidrógeno entre el oxhidrilo libre del azúcar y el residuo Glu570, cuya distancia media es de 5.28 ± 1.10 Å (Figura A.5-B). Adicionalmente los residuos aromáticos Tyr511 y Phe507 forman interacciones del tipo π -stacking con el núcleo cumarínico y el fragmento benzamida de NVB, respectivamente.

Cinarizina

Para el caso de CNZ, la mayoría de las interacciones que se observan durante la simulación son del tipo hidrofóbicas. Varios residuos no polares como Leu515, Leu553, Ala556, Ile573, Val583, Leu675 interaccionan de forma no específica con CNZ. Adicionalmente, la Phe507 forma un π -stacking con uno de los anillos aromáticos del grupo bencidrilo.

Referencias Capítulo 3

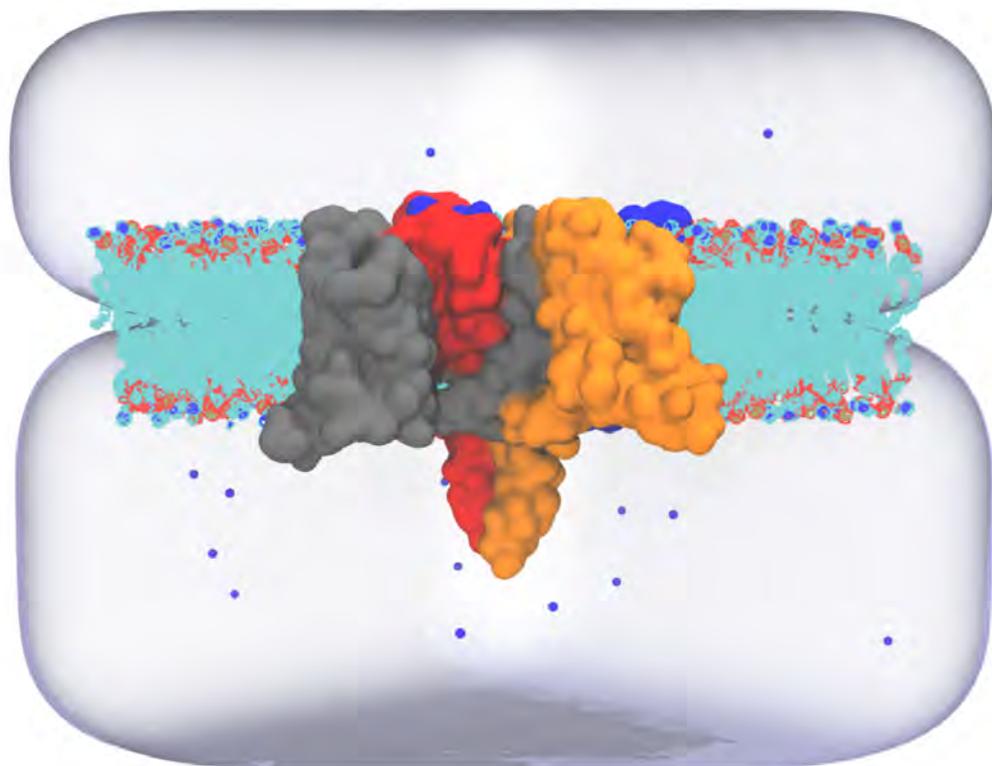
- Alhossary, A.** et al. (2015). “Fast, accurate, and reliable molecular docking with QuickVina 2”. *Bioinformatics* 31.13, págs. 2214-2216. DOI: [10.1093/bioinformatics/btv082](https://doi.org/10.1093/bioinformatics/btv082).
- Balusek, C.** et al. (2019). “Accelerating Membrane Simulations with Hydrogen Mass Repartitioning”. *Journal of Chemical Theory and Computation* 15.8, págs. 4673-4686. DOI: [10.1021/acs.jctc.9b00160](https://doi.org/10.1021/acs.jctc.9b00160).
- Cao, E.** et al. (2013). “TRPV1 structures in distinct conformations reveal activation mechanisms”. *Nature* 504.7478, págs. 113-118. DOI: [10.1038/nature12823](https://doi.org/10.1038/nature12823).
- Chen, V. B.** et al. (2010). “MolProbity: all-atom structure validation for macromolecular crystallography”. *Acta Crystallographica Section D: Biological Crystallography* 66.Pt 1, págs. 12-21. DOI: [10.1107/S0907444909042073](https://doi.org/10.1107/S0907444909042073).
- Conway, P.** et al. (2014). “Relaxation of backbone bond geometry improves protein energy landscape modeling”. *Protein Science : A Publication of the Protein Society* 23.1, págs. 47-55. DOI: [10.1002/pro.2389](https://doi.org/10.1002/pro.2389).
- Eswar, N.** et al. (2006). “Comparative Protein Structure Modeling Using Modeller”. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]* 0 5, Unit-5.6. DOI: [10.1002/0471250953.bi0506s15](https://doi.org/10.1002/0471250953.bi0506s15).
- Garami, A.** et al. (2020). “Hyperthermia induced by transient receptor potential vanilloid-1 (TRPV1) antagonists in human clinical trials: Insights from mathematical modeling and meta-analysis”. *Pharmacology & Therapeutics* 208, pág. 107474. DOI: [10.1016/j.pharmthera.2020.107474](https://doi.org/10.1016/j.pharmthera.2020.107474).
- Hopkins, C. W.** et al. (2015). “Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning”. *Journal of Chemical Theory and Computation* 11.4, págs. 1864-1874. DOI: [10.1021/ct5010406](https://doi.org/10.1021/ct5010406).

- Jo**, S. et al. (2008). “CHARMM-GUI: A web-based graphical user interface for CHARMM”. *Journal of Computational Chemistry* 29.11, págs. 1859-1865. DOI: [10.1002/jcc.20945](https://doi.org/10.1002/jcc.20945).
- Koehler Leman**, J., **B. K. Mueller** y **J. J. Gray** (2017). “Expanding the toolkit for membrane protein modeling in Rosetta”. *Bioinformatics* 33.5, págs. 754-756. DOI: [10.1093/bioinformatics/btw716](https://doi.org/10.1093/bioinformatics/btw716).
- Koes**, D. R., **M. P. Baumgartner** y **C. J. Camacho** (2013). “Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise”. *Journal of chemical information and modeling* 53.8, págs. 1893-1904. DOI: [10.1021/ci300604z](https://doi.org/10.1021/ci300604z).
- Lee**, J. H. et al. (2011). “Structural insights into transient receptor potential vanilloid type 1 (TRPV1) from homology modeling, flexible docking, and mutational studies”. *Journal of Computer-Aided Molecular Design* 25.4, págs. 317-327. DOI: [10.1007/s10822-011-9421-5](https://doi.org/10.1007/s10822-011-9421-5).
- Lishko**, P. V. et al. (2007). “The Ankyrin Repeats of TRPV1 Bind Multiple Ligands and Modulate Channel Sensitivity”. *Neuron* 54.6, págs. 905-918. DOI: [10.1016/j.neuron.2007.05.027](https://doi.org/10.1016/j.neuron.2007.05.027).
- Llanos**, M. A. et al. (2022). “Structure-Based Virtual Screening Identifies Novobiocin, Montelukast, and Cinnarizine as TRPV1 Modulators with Anticonvulsant Activity In Vivo”. *Journal of Chemical Information and Modeling*. DOI: [10.1021/acs.jcim.2c00312](https://doi.org/10.1021/acs.jcim.2c00312).
- Lomize**, M. A. et al. (2012). “OPM database and PPM web server: resources for positioning of proteins in membranes”. *Nucleic Acids Research* 40.Database issue, págs. D370-D376. DOI: [10.1093/nar/gkr703](https://doi.org/10.1093/nar/gkr703).
- Morris**, G. M. et al. (2009). “AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility”. *Journal of computational chemistry* 30.16, págs. 2785-2791. DOI: [10.1002/jcc.21256](https://doi.org/10.1002/jcc.21256).
- Needleman**, S. B. y **C. D. Wunsch** (1970). “A general method applicable to the search for similarities in the amino acid sequence of two proteins”. *Journal of Molecular Biology* 48.3, págs. 443-453. DOI: [10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4).
- Ohbuchi**, K. et al. (2016). “Detailed Analysis of the Binding Mode of Vanilloids to Transient Receptor Potential Vanilloid Type I (TRPV1) by a Mutational and Computational Study”. *PLOS ONE* 11.9. Ed. por A. A. **Romanovsky**, e0162543. DOI: [10.1371/journal.pone.0162543](https://doi.org/10.1371/journal.pone.0162543).

- Perez, J. J., M. S. Tomas y J. Rubio-Martinez** (2016). “Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory”. *Journal of Chemical Information and Modeling* 56.10, págs. 1950-1962. DOI: [10.1021/acs.jcim.6b00347](https://doi.org/10.1021/acs.jcim.6b00347).
- Pranami, G. y M. H. Lamm** (2015). “Estimating error in diffusion coefficients derived from molecular dynamics simulations”. *Journal of Chemical Theory and Computation* 11.10, págs. 4586-4592. DOI: [10.1021/acs.jctc.5b00574](https://doi.org/10.1021/acs.jctc.5b00574).
- Quiroga, R. y M. A. Villarreal** (2016). “Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening”. *PLOS ONE* 11.5, e0155183. DOI: [10.1371/journal.pone.0155183](https://doi.org/10.1371/journal.pone.0155183).
- Santos-Martins, D. et al.** (2019). “Accelerating AutoDock4 with GPUs and Gradient-Based Local Search”. DOI: [10.26434/chemrxiv.9702389.v1](https://doi.org/10.26434/chemrxiv.9702389.v1).
- Shen, M.-y. y A. Sali** (2006). “Statistical potential for assessment and prediction of protein structures”. *Protein Science : A Publication of the Protein Society* 15.11, págs. 2507-2524. DOI: [10.1110/ps.062416606](https://doi.org/10.1110/ps.062416606).
- Studer, G., M. Biasini y T. Schwede** (2014). “Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane)”. *Bioinformatics* 30.17, págs. i505-i511. DOI: [10.1093/bioinformatics/btu457](https://doi.org/10.1093/bioinformatics/btu457).
- Trott, O. y A. J. Olson** (2010). “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading”. *Journal of Computational Chemistry* 31.2, págs. 455-461. DOI: [10.1002/jcc.21334](https://doi.org/10.1002/jcc.21334).
- Wu, E. L. et al.** (2014). “CHARMM-GUI Membrane Builder toward realistic biological membrane simulations”. *Journal of Computational Chemistry* 35.27, págs. 1997-2004. DOI: [10.1002/jcc.23702](https://doi.org/10.1002/jcc.23702).

Capítulo 4

Búsqueda de inhibidores
del canal NaV1.2



4.1. Introducción

Este capítulo describe el desarrollo y la validación de modelos *in silico*, basados en la estructura de los ligandos (QSAR) y del receptor (docking), con el objetivo de identificar nuevos inhibidores del canal NaV1.2, con potencial actividad anticonvulsiva.

En una primera etapa, se compiló y curó un *dataset* de compuestos evaluados en el ensayo de desplazamiento de BTX. Este conjunto de datos se utilizó para desarrollar un modelo basado en la estructura de los ligandos capaz de predecir, de manera cuantitativa, la actividad de un compuesto en este ensayo, lo que representa una medida indirecta de su afinidad al sitio, en isoformas cerebrales de canales NaV.

Luego, a partir de la estructura tridimensional del canal hNaV1.2 determinada experimentalmente, se generó un *ensemble* conformacional, el cual se utilizó para desarrollar un modelo de docking capaz de identificar compuestos que se unen a la región del poro.

4.2. Metodologías

4.2.1. Compilado y partición de la base de datos

El conjunto de datos se compiló a partir de la base de datos ChEMBL26. Solo se consideraron datos provenientes del ensayo de desplazamiento de Batracotoxina (BTX) marcada. Esta toxina, aislada de las ranas venenosas del género *Phylllobates*, se une al poro del canal activándolo de manera irreversible (**Daly y Witkop 1971; Wang et al. 2006**). El sitio de unión de la BTX (sitio 2) se superpone parcialmente con el de moléculas pequeñas o anestésicos locales (**Linford et al. 1998**). Utilizando un derivado tritiado de esta toxina (H-BTX) es posible medir cuánta H-BTX es desplazada de su sitio de unión por el agregado de un compuesto, infiriendo así la afinidad relativa del mismo por este sitio.

Este ensayo se realiza comúnmente en sinaptosomas (terminal o botón sináptico aislado de una neurona), por lo que incluye a todas las isoformas cerebrales de los

canales de sodio operados por voltaje (NaV1.1, NaV1.2, NaV1.3 y NaV1.6). Debido a que la región del poro es prácticamente invariable entre estas isoformas (**Oelstrom et al. 2014**), este *dataset* puede ser utilizado en la construcción y validación de modelos predictivos para las isoformas antes mencionadas.

La elección de este ensayo en particular se fundamenta en que, como se describió en el Capítulo 2, este tipo de canales posee más de un sitio de unión de moléculas pequeñas. Si consideramos datos de ensayos funcionales, no tendremos certeza alguna sobre el sitio de unión de los compuestos activos. En el modelado QSAR, esto no representa un problema mayor ya que pueden construirse modelos fenotípicos, para predecir compuestos activos frente a un ensayo biológico o *target* particular. Sin embargo, a la hora de validar una metodología de docking es necesario que los compuestos activos o *binders* interaccionen con la región del receptor donde ocurre la simulación.

En un principio, solo se consideraron los datos reportados como concentración inhibitoria 50 % (IC50) en el ensayo de desplazamiento de H-BTX, realizados en sinaptosomas de rata (*Rattus norvegicus*). Los compuestos duplicados fueron eliminados agrupándolos por su código InChI y calculando para cada uno de ellos la mediana y la desviación estándar de las pIC50 (-Log(IC50)) reportadas. Si esta última fuera mayor a 0.3, el dato es descartado. Si la desviación estándar es menor a 0.3, el dato se incluye como la mediana de todas las actividades reportadas. Debido a la homogeneidad de los datos recopilados y a la naturaleza continua de la variable dependiente (pIC50), este conjunto de datos (**Dataset A**) se utilizó para desarrollar modelos QSAR de regresión, así como también en la validación de los modelos de docking.

Adicionalmente, se decidió compilar otro conjunto de datos (**Dataset B**) el cual fue utilizado durante la validación externa de los modelos QSAR desarrollados a partir del **Dataset A**. En este *dataset*, la variable dependiente es de tipo categórica, es decir que los compuestos se encuentran clasificados como activos o inactivos, según los siguientes criterios:

Activos:

- El valor de pIC50 reportado es mayor a 6 (1 μ M).

- El porcentaje de inhibición en el ensayo de desplazamiento de H-BTX a una concentración de 10 μM es mayor o igual a 60 %.

Inactivos:

- El valor de pIC50 reportado es menor o igual a 5 (10 μM).
- La actividad reportada es del tipo categórica, es decir ' $\geq 10 \mu\text{M}$ ' o 'Not Active'.
- El porcentaje de inhibición en el ensayo de desplazamiento de H-BTX a una concentración de 10 μM es menor o igual a 50 %.

Se consideraron también ensayos de porcentaje de inhibición en cobayo (*Cavia Porcellus*), realizados a una concentración fija (10 μM). Los compuestos duplicados dentro de cada categoría fueron eliminados utilizando el procedimiento anteriormente descrito. Así mismo, aquellas moléculas que - por haber sido reportadas con valores de actividad muy diferentes - caen en ambas categorías, fueron eliminadas del conjunto de datos para evitar ambigüedades. Aquellas moléculas ya incluidas en el **Dataset A** también fueron removidas de este conjunto de datos.

Finalmente, se construyó una base simulada utilizando la herramienta DUD-E para generar *decoys* o presuntos inactivos. En este caso, se utilizaron como activos semilla los compuestos dentro del **Dataset A** cuyo pIC50 ≥ 6.0 (1 μM), aunque solo aquellos no incluidos en el conjunto de entrenamiento fueron sembrados en esta base simulada **DUD-E**.

Partición del conjunto de datos

Con el objetivo de particionar los datos de manera representativa, es decir que la diversidad química presente en el conjunto inicial de datos (**Dataset A**) se vea representada tanto en el conjunto de entrenamiento como en el de testeo, se realizó una partición estratificada de los mismos. Para ello, los datos fueron agrupados utilizando la metodología de *clustering* SOMoC, detallada en el material suplementario Algoritmos - [C.1](#).

En lo que refiere a modelos de regresión, la distribución de la variable dependiente es un factor fundamental a tener en cuenta (**Kiralj y Ferreira 2009**). Realizar una partición estratificada considerando dos variables a estratificar, donde una es continua (pIC50) y la otra categórica (*cluster*), es un problema complejo. Por tal

motivo, se procedió de la siguiente manera: Se generaron particiones estratificadas aleatorias utilizando la pertenencia a cada *cluster*, todas ellas del mismo tamaño: 80 % para entrenamiento y 20 % para testeo. Para cada una de ellas evaluó la similitud entre las distribuciones de pIC50 del conjunto de entrenamiento y de testeo, calculando distintos parámetros característicos de una distribución como su media, desviación estándar, el Kurtosis y el Skewness. Finalmente, se eligió la partición estratificada que genera distribuciones similares para ambos conjuntos, de acuerdo a estos estadísticos.

4.2.2. Desde los ligandos

La estrategia de modelado utilizada para desarrollar y validar los modelos QSAR de regresión para el canal NaV1.2 es la misma que se utilizó para el canal TRPV1, y se encuentra detallada en el Capítulo 2, Sección 2.3.2.4. De esta manera, a lo largo de este capítulo solo se hará hincapié en las diferencias particulares por tratarse de un modelo de regresión.

4.2.2.1. Desarrollo y validación de un modelo de regresión

Cálculo y selección de descriptores

Los descriptores moleculares se calcularon mediante el software Mordred, y la selección de los mismos se llevó a cabo mediante *pipeline* descrito en el Capítulo 2, Sección 2.3.2.4. El algoritmo utilizado durante la selección de descriptores fue la regresión lineal (LR, del inglés Linear Regression), debido a su simplicidad y bajo costo computacional. Como función de pérdida se utilizó el coeficiente Tau de Kendall, de modo de maximizar la correlación entre el ranking predicho y el experimental.

Elección del mejor algoritmo

Los siguientes algoritmos de regresión fueron evaluados: LR, Light Gradient Boosting Machine Regressor (LGBM), Stochastic Gradient Descent Regressor (SGD), Ridge Regression (Ridge), Support Vector Machine Regression (SVR) y Gradient-

BoostingRegressor (GBT). Nuevamente, para cada uno de ellos se realizaron 100 iteraciones de 10-Fold CV en el conjunto de entrenamiento, utilizando los hiperparámetros por defecto.

4.2.3. Desde el receptor

En el año 2019, Pan y colaboradores publicaron la primer estructura tridimensional de un canal NaV1.2 humano (**Pan et al. 2019**). Mediante Cryo-EM, se logró capturar al canal junto con su subunidad moduladora β -2 y la μ -conotoxina KIIIA, un péptido pequeño con capacidad de bloquear los canales NaV.

Previamente, varios modelos de esta (y otras isoformas de la familia NaV), fueron desarrollados recurriendo a técnicas de modelado comparativo. En el año 2021, publicamos una revisión donde se analizan comparativamente estos modelos y su aplicación en metodologías basadas en la estructura del receptor para identificar nuevas ACDs (**Gantner et al. 2021**). En particular, el modelo de la isoforma hNaV1.2 desarrollado por nuestro grupo de investigación se construyó utilizando como *template* un canal NaV bacteriano. El mismo se aplicó en una campaña de VS, donde se lograron identificar 4 inhibidores de este canal, 3 de ellos con actividad anticonvulsiva *in vivo* (**Palestro et al. 2018**).

4.2.3.1. Obtención de un *ensemble* estructural del canal NaV1.2

La estructura del canal hNaV1.2 (PDB-ID: 6J8E, resolución 3.00 Å) fue procesada, removiendo la μ -conotoxina KIIIA y la subunidad β -2. Luego, el segmento no resuelto entre los residuos 285-313 fue modelado utilizando el programa Modeller (**Eswar et al. 2006**). Si bien esta región se encuentra lejos del poro, donde posteriormente se llevarán a cabo las simulaciones de docking, a la hora de realizar simulaciones que implican a la proteína en su totalidad -como la dinámica molecular- resulta conveniente que la misma no presente este tipo de discontinuidades.

Esta estructura fue sometida a una etapa de refinamiento, utilizando el protocolo RosettaMP (**Koehler Leman, Mueller y Gray 2017**), descrito en el Capítulo 3, Sección 3.2.3.1. El servidor PPM (**Lomize et al. 2012**) fue utilizado para calcu-

lar la porción del canal inserto en la membrana, la cual fue corroborada y refinada utilizando la estructura experimental. De esta manera, se generaron 100 modelos refinados de la proteína, obteniendo así un *ensemble* conformacional alrededor de la estructura inicial.

4.2.3.2. Desarrollo y validación de un modelo de docking

Selección de la mejor estructura

El **Dataset A** fue dockeado en las 100 conformaciones de la proteína generadas previamente. Adicionalmente se incluyó la estructura inicial, antes y después del primer paso de refinamiento, con el objeto de establecer una línea de base o modelo estándar contra el cual comparar los resultados. Por lo tanto, fueron 102 las estructuras evaluadas.

Como criterio para evaluar la capacidad predictiva de cada conformación, bajo determinadas condiciones de docking, se utilizó la correlación entre el ranking obtenido producto de ordenar los compuestos por su *score* de docking, y el ranking real producto de ordenarlos por el valor de pIC50 determinado experimentalmente. Distintas métricas de ranking fueron calculadas para cada una de ellas, como lo son el coeficiente de Spearman (Rho), el coeficiente de Kendall (TauB) y el de Pearson (R). Las mismas se encuentran detalladas en el Material Suplementario [D](#) - Métricas.

4.3. Resultados

4.3.1. Compilado y partición de la base de datos

Luego de procesar y curar los datos de actividad biológica en base a los criterios descritos anteriormente, se generaron 3 conjuntos de datos, representados en la Figura [4.1](#).

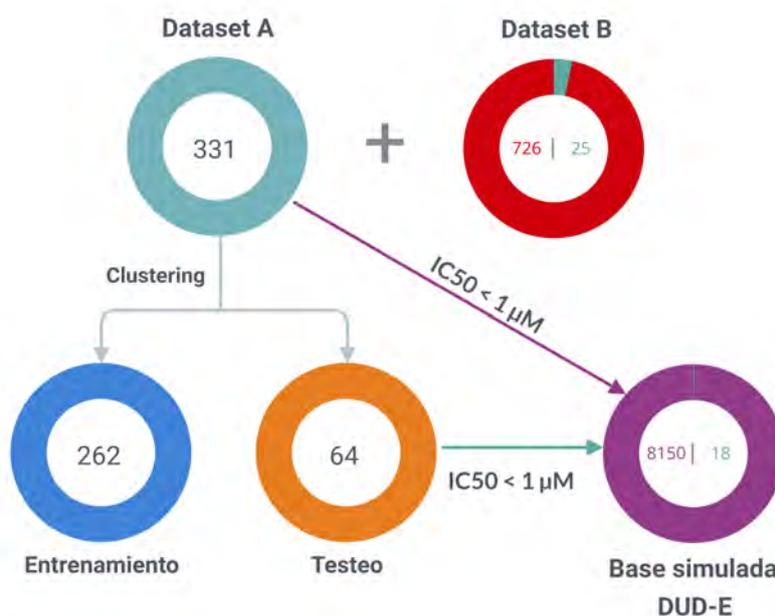


Figura 4.1: Tamaño y composición de los distintos conjuntos de datos compilados para desarrollar y validar los modelos QSAR y de docking para el canal NaV1.2

El **Dataset A** consta de 331 compuestos y su respectivo valor de pIC_{50} en el ensayo de desplazamiento de H-BTX (Ver material suplementario B). Todas las determinaciones se realizaron en la misma especie (*Rattus Norvegicus*) y el valor de actividad reportado es numérico en todos los casos, por lo que será utilizado para generar modelos QSAR de regresión, cuyo objetivo es predecir el valor de pIC_{50} . Por su parte, el **Dataset B** comprende 751 compuestos clasificados como activos (25) o inactivos (726). Finalmente, la **Biblioteca simulada DUD-E** consta de 18 activos y 8150 *decoys* o presuntos inactivos.

El **Dataset A** fue agrupado utilizando la metodología SOMoC a través de la aplicación web. El número de *clusters* K se varió entre 2 y 25. Durante la reducción de dimensionalidad con UMAP, el número de vecinos se fijó en 12. Todos los demás parámetros se mantuvieron en sus valores por defecto. Observando el *elbow plot* generado (Figura 4.2) se seleccionó un valor de $K=8$, con el que se obtiene un $SIL=0.871$. En base a las distintas métricas calculadas, podemos observar una vez más que se logran *clusters* densos y bien separados (Tabla B.1).

4.3.2. Desde los ligandos

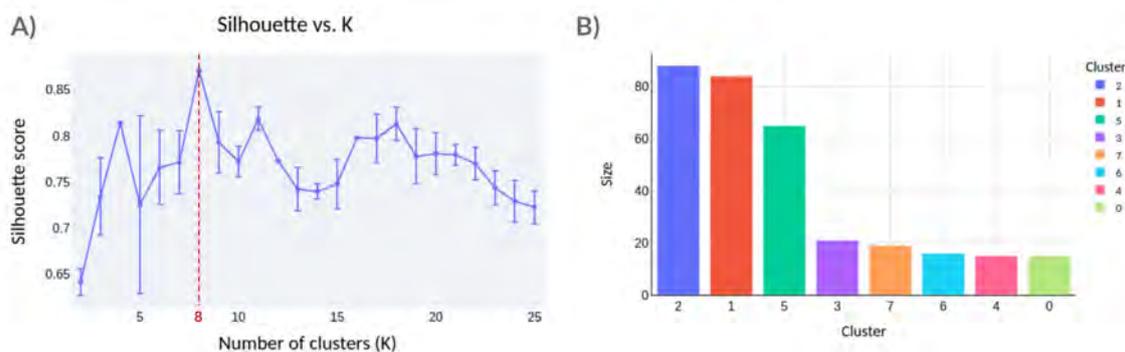


Figura 4.2: Validación interna del *clustering*. A) Gráfico de *elbow* generado por SOMoC, utilizado para seleccionar el número óptimo de *clusters* (K) en el conjunto de datos. En el eje X se representan los valores de K, mientras que en el eje Y se muestra el SIL *score*. Las barras de error representan la desviación estándar calculada sobre 10 iteraciones independientes de *clustering*. B) Distribución de tamaño para los *clusters* generados por SOMoC (K=8). El eje X muestra cada uno de los *clusters* identificados y en el eje Y su tamaño.

4.3.2.1. Desarrollo y validación de un modelo de regresión

Cálculo y selección de descriptores

A partir del **Dataset A** se generaron 5.000 particiones estratificadas considerando la pertenencia a cada *cluster*, todas ellas del mismo tamaño: 80 % para entrenamiento (262 moléculas) y 20 % para testeo (64 moléculas). Finalmente, se seleccionó la partición que genera distribuciones similares en cuanto a su media, desviación estándar, Kurtosis y Skewness.

El conjunto de datos compilados presenta una gran diversidad química, y comprende un amplio rango de actividad biológica. A su vez, se logró una distribución de pIC50 aproximadamente normal y muy similar en ambos conjuntos de entrenamiento y testeo (Figura 4.3).

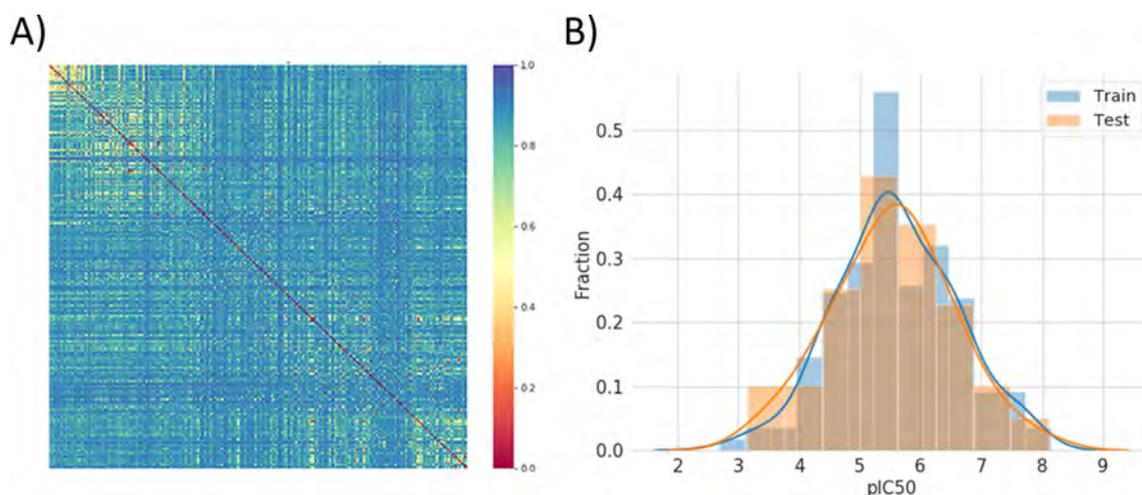


Figura 4.3: A) *Heatmap* de disimilaridad para todos los compuestos del conjunto de datos. La similitud se calculó utilizando ECFP4 y el coeficiente de Tanimoto como medida de distancia. Un valor de 0 significa compuestos idénticos, mientras que un valor de 1 representa compuestos muy diferentes. Las moléculas están ordenadas en orden decreciente de potencia. B) Distribución de la variable dependiente (pIC50) para las moléculas del conjunto de datos. La distribución del conjunto de entrenamiento está representada en azul y la del conjunto de testeo en naranja.

Se calcularon inicialmente 1613 descriptores independientes de la conformación implementados en el software Mordred. Luego de aplicar los filtros de varianza y correlación lineal, 317 fueron sometidos a la etapa de selección mediante el algoritmo genético.

Al finalizar las 10 iteraciones de K-Fold CV, un total de 989.361 soluciones fueron generadas y evaluadas por el algoritmo genético. De cada iteración se tomaron las mejores 100 soluciones, generando así un conjunto de 1000 “buenas soluciones”. El ranking final se estableció contabilizando la frecuencia de aparición de cada descriptor en este conjunto de 1000 soluciones. (Figura 4.4). Finalmente, se seleccionaron los primeros 20 descriptores del ranking para entrenar los algoritmos, lo que representa una relación ejemplos/descriptores de aproximadamente 13.

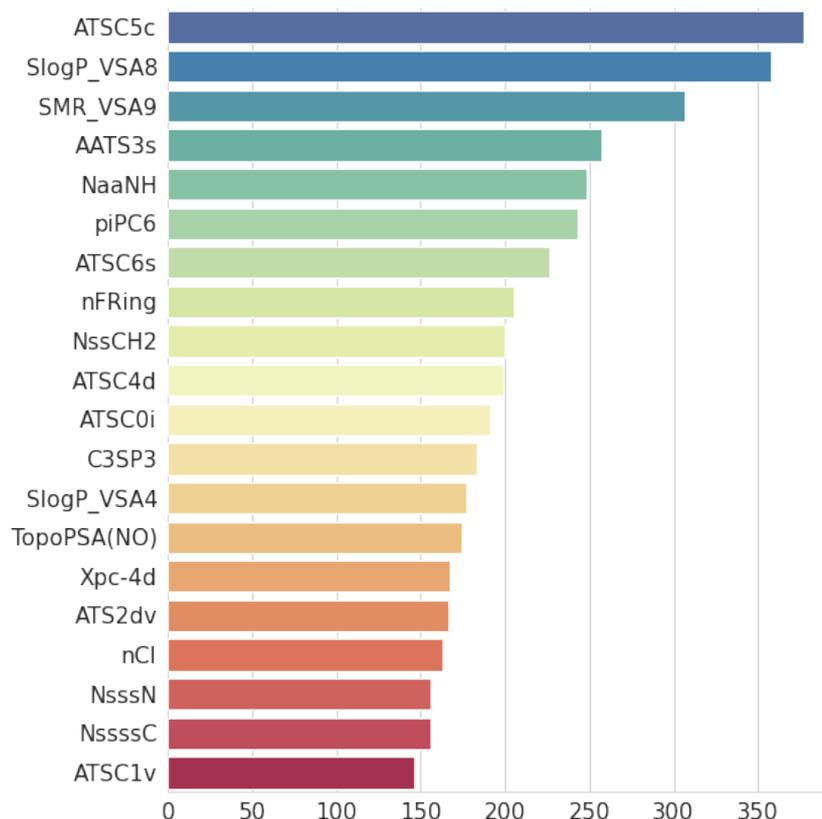


Figura 4.4: Ranking de importancia de variables generado mediante el método de envoltura. Solo se muestran los primeros 20 descriptores del ranking, que fueron seleccionados para la etapa posterior de entrenamiento de los modelos. La escala de color representa la importancia de cada descriptor, desde el azul (más importante) al rojo (menos importante).

Elección del mejor algoritmo

Como puede observarse en la Figura 4.5, los 6 algoritmos evaluados presentan un desempeño similar en la validación interna 100 x 10-Fold CV. Para todos ellos, el coeficiente TauB de Kendall toma valores mayores a 0.6, lo que demuestra una buena correlación entre ambos rankings, mientras que el MAE oscila alrededor de 0.4 en todos los casos. Sin embargo, se observa una leve superioridad de SVR con respecto a los demás. Los resultados numéricos para las métricas MAE, R^2 , Rho y TauB se detallan en la Tabla B.2.

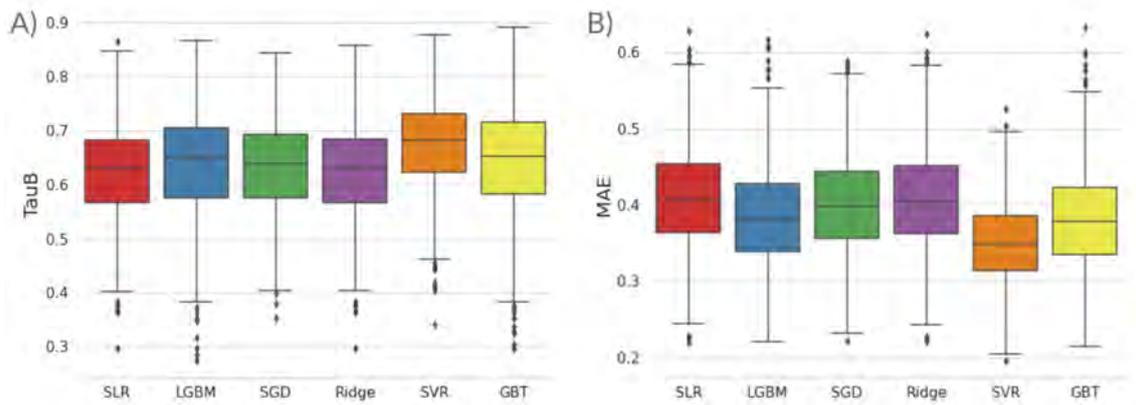


Figura 4.5: Selección del mejor algoritmo. Validación interna 100 x 10-Fold CV en el conjunto de entrenamiento para todos los algoritmos evaluados, utilizando los hiperparámetros por defecto en cada caso. En el gráfico de cajas A) se muestran los valores de TauB, mientras que en el B) se representan los valores de MAE obtenidos. El límite inferior y superior de cada caja representa el primer y tercer cuartil, respectivamente, mientras que la línea media representa la mediana o segundo cuartil. Las líneas que se continúan de las cajas (*whiskers*) contienen el resto de la distribución, a excepción de los valores atípicos u *outliers* que se representan como puntos.

Luego de realizar 1000 iteraciones de aleatorización de Fisher, se observó que todos los modelos generados presentan un pobre desempeño, obteniéndose valores de R^2 , TauB y Rho cercanos a 0 o incluso negativos, y grandes errores en términos absolutos (MAE) (Tabla B.3).

Optimización de hiperparámetros

En base a los resultados obtenidos, se avanzó en la optimización de los algoritmos SGD, SVR y LGBM. La metodología de optimización se encuentra detallada en el Capítulo 2, Sección 2.3.2.4. Para cada algoritmo se corrieron 3 x 1000 iteraciones de optimización bayesiana y como métrica a minimizar se escogió una vez más el coeficiente TauB.

Analizando la evolución del coeficiente TauB a lo largo del proceso de optimización, se observa un aumento del mismo a medida que la búsqueda avanza. Esta tendencia se vuelve menos pronunciada sobre el final, lo que implica cierto grado de convergencia. La gran mayoría de las soluciones exploradas son “buenas soluciones”, es decir que tienen valores de TauB altos (Figura B.2). Al igual que para los modelos

del canal TRPV1, el análisis de la distribución de valores muestreados para cada hiperparámetro revela que la búsqueda bayesiana se focaliza en ciertas regiones del dominio de búsqueda, donde la densidad de muestreo es considerablemente mayor (Figura 4.6). Los resultados obtenidos para los algoritmos SGD y LGBM pueden encontrarse en el material suplementario, Figuras B.1 y B.2, respectivamente.

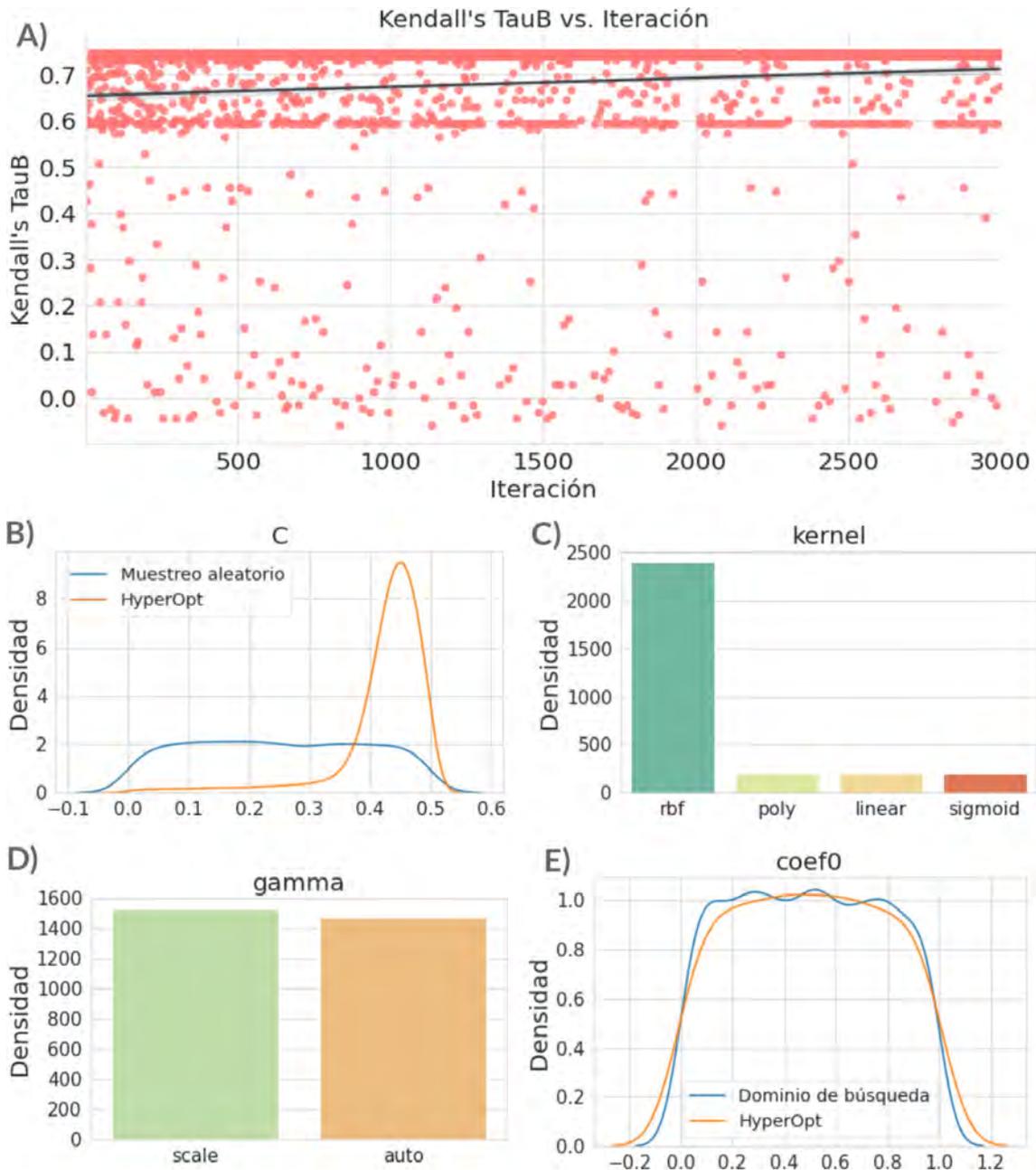


Figura 4.6: Optimización de hiperparámetros para el algoritmo SVR. A) Evolución del coeficiente TauB a lo largo del proceso de optimización. La línea negra representa una regresión lineal entre los valores de x e y, mientras que el sombreado representa un intervalo de confianza del 95 %. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: C (B), kernel (C), gamma (D), coef0 (E). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

Tabla 4.1: Validación interna 100 x 10-Fold CV para los algoritmos optimizados. Para cada métrica se expresa la media acompañada de su desvío estándar (SD).

Modelo	MAE		R^2		Rho		TauB	
	Media	SD	Media	SD	Media	SD	Media	SD
SGD	0.403	0.065	0.625	0.155	0.803	0.077	0.629	0.084
SVR	0.360	0.059	0.755	0.099	0.860	0.063	0.706	0.076
LGBM	0.385	0.066	0.712	0.138	0.826	0.078	0.667	0.086

Validación *in silico*

Utilizando los modelos optimizados previamente, se procedió a realizar las validaciones internas y externas de los mismos.

Con respecto a la validación interna, los algoritmos optimizados presentaron una buena *performance* en la validación 100 x 10-Fold CV (Tabla 4.1) y un desempeño muy pobre en la aleatorización de Fisher, con coeficientes de correlación nulos o negativos y valores de MAE superiores a 0.8 en todos los casos (Tabla B.4).

Analizando la Figura 4.7, podemos observar que los algoritmos no lineales como SVR y LGBM tienen un mejor desempeño que SGD, a juzgar por las métricas TauB y MAE.

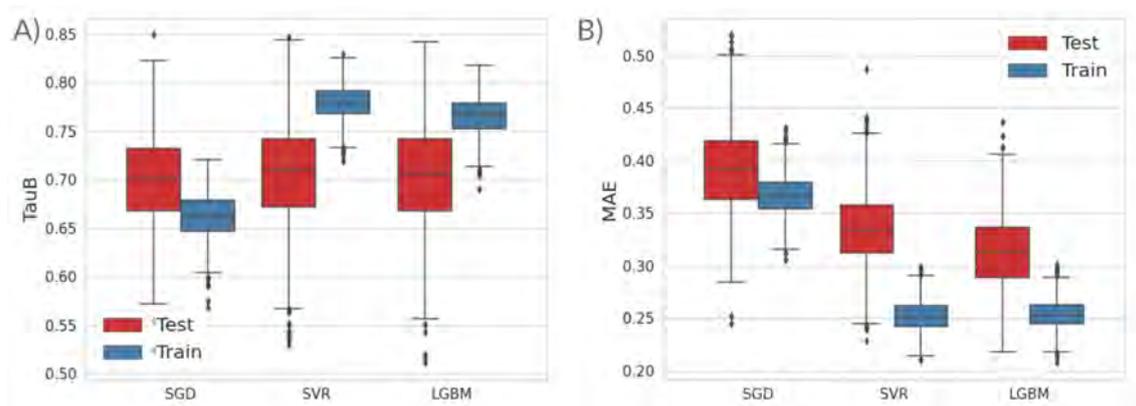


Figura 4.7: Validación externa de los algoritmos optimizados. *Performance* de los algoritmos SGD, SVR y LGBM sobre los conjuntos de entrenamiento (*Train*) y testeo (*Test*). En el gráfico de cajas A) se muestran los valores de TauB, mientras que en el B) se representan los valores de MAE obtenidos. El límite inferior y superior de cada caja representa el primer y tercer cuartil, respectivamente, mientras que la línea media representa la mediana o segundo cuartil. Las líneas que se continúan de las cajas (*whiskers*) contienen el resto de la distribución, a excepción de los valores atípicos u *outliers* que se representan como puntos.

Finalmente, se procedió a evaluar los modelos como clasificadores binarios, utilizando el **Dataset B** y la **base simulada DUD-E**, ambos conjuntos externos compuestos por moléculas clasificadas como activas o inactivas.

Para el **Dataset B**, los 3 modelos demostraron un desempeño similar en términos de AUC-ROC, pero existen algunas diferencias en la métrica BEDROC, donde se observa que SVR y LGBM presentan valores ligeramente mayores. Para el caso de la **base simulada DUD-E**, esta tendencia se acentúa ya que la *performance* de SVR y LGBM es considerablemente mejor que SDG en términos de AUC-ROC y BEDROC (Figura 4.8).

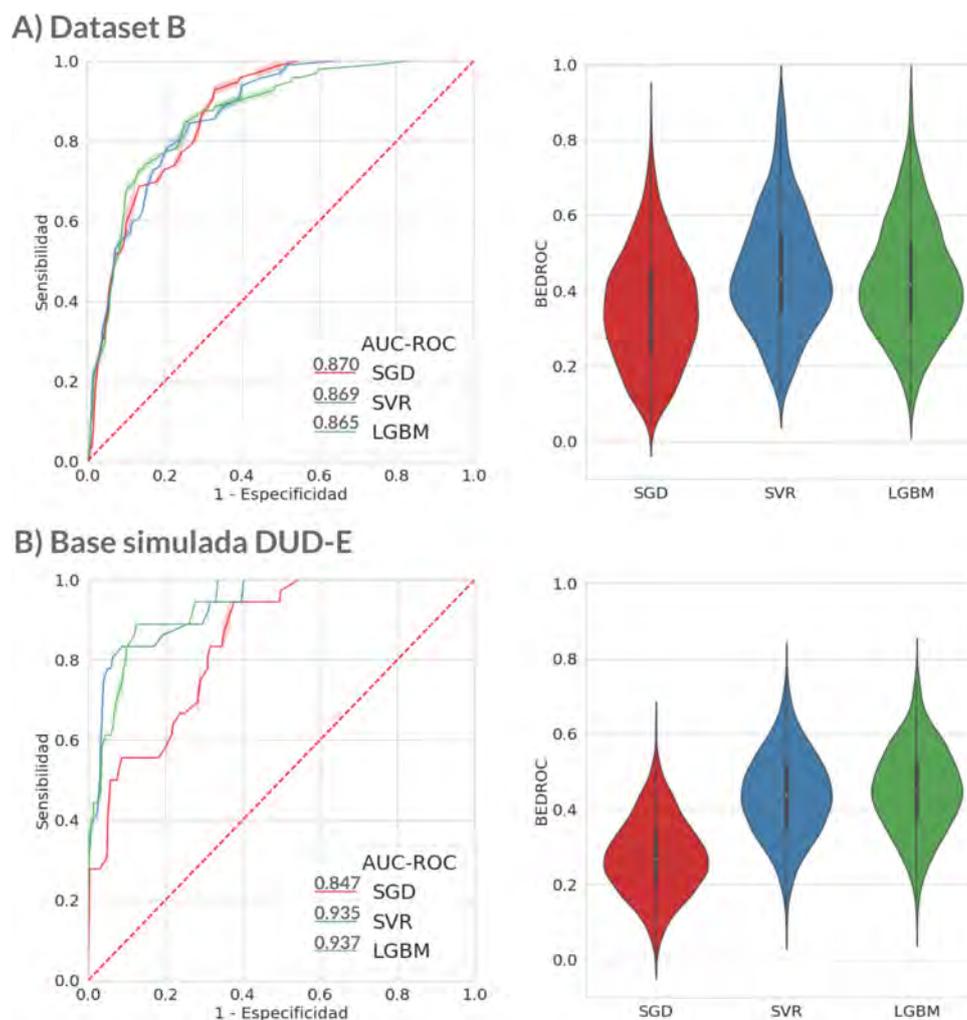


Figura 4.8: Evaluación de los modelos como clasificadores binarios. Para los modelos optimizados SGD, SVR y LGBM se muestran las curvas ROC y la métrica BEDROC($\alpha=100$), calculadas para el **Dataset B** (A) y la **base simulada DUD-E** (B). El sombreado de las curvas representa un intervalo de confianza del 95 % obtenido mediante 1000 iteraciones de *bootstrapping*. La misma técnica de re-muestreo se utilizó para generar los gráficos de violín.

A diferencia de la biblioteca generada para el canal TRPV1, donde los *decoys* se generaron a partir de la totalidad de las semillas activas sembradas en la **base simulada DUD-E**, en este caso solo se incluyeron aquellos activos que no están presentes en el conjunto de entrenamiento. Por lo tanto, la proporción de activos en este caso es aún más baja, lo que representa un desafío mayor para el modelo.

Una ventaja de los modelos SGD y LGBM es que permiten estimar la importancia relativa de las variables que componen el modelo. Para SGD esta se estima a partir de los coeficientes que completan la ecuación lineal del modelo, mientras que

para LGBM se calcula en base al número de veces que un descriptor es utilizado para generar un *split*. En la Figura 4.9 pueden observarse los rankings de importancia de variables generados por ambos modelos, donde se encuentran ciertas coincidencias, aunque cada uno relaciona estas variables entre sí de una manera diferente, lineal para SGD y no lineal para LGBM.

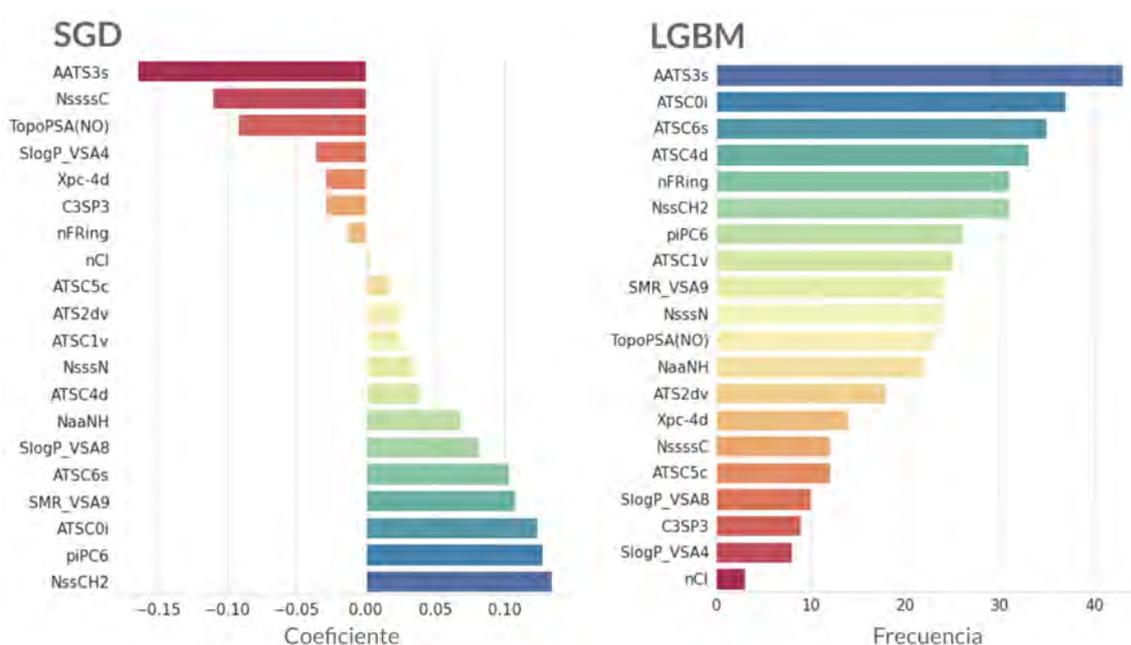


Figura 4.9: Rankings de importancia de variables generados por los algoritmos SGD y LGBM. La escala de color representa la importancia de cada descriptor, desde el azul (más importante) al rojo (menos importante).

En base al conjunto de validaciones *in silico* realizadas, se observa que los modelos SVR y LGBM tienen un desempeño muy similar. Sin embargo, SVR es un algoritmo mucho más simple (menos hiperparámetros) y por lo tanto fue el elegido para ser utilizado en futuras campañas prospectivas de VS. El fundamento de este algoritmo se detalla en el material suplementario Algoritmos - C.3.3.

4.3.3. Desde el receptor

4.3.3.1. Obtención de un *ensemble* estructural del canal NaV1.2

Luego de reparar la región faltante entre los residuos 285-313, la estructura fue sometida a 8 ciclos de MPFastRelax, aplicando fuertes restricciones armónicas en

todos los átomos pesados, para evitar que ésta se desvíe de su conformación inicial. Este paso es recomendado antes de aplicar cualquier otro protocolo de simulación, y permite que la proteína se adapte al campo de fuerza a utilizar.

Posteriormente, el mismo protocolo fue utilizado para generar 100 modelos de la proteína, pero esta vez relajando las restricciones impuestas sobre la estructura, de modo de muestrear el espacio conformacional alrededor de la conformación determinada experimentalmente.

Tomando como ejemplo el modelo 53 del *ensemble* y comparándolo con la estructura inicial determinada experimentalmente (PDB-ID: 6J8E), puede observarse que el refinamiento con Rosetta permite mejorar la calidad de los modelos en términos energéticos y geométricos, evaluados mediante los servidores Molprobability y QMEANBrane. Naturalmente, la zona más problemática del modelo es el segmento no resuelto entre los residuos 285-313, que fue reparado con Modeller antes del refinamiento. Sin embargo, esta zona se encuentra lejos del poro del canal y por lo tanto no representa un problema mayor en las simulaciones de docking (Figura 4.10).

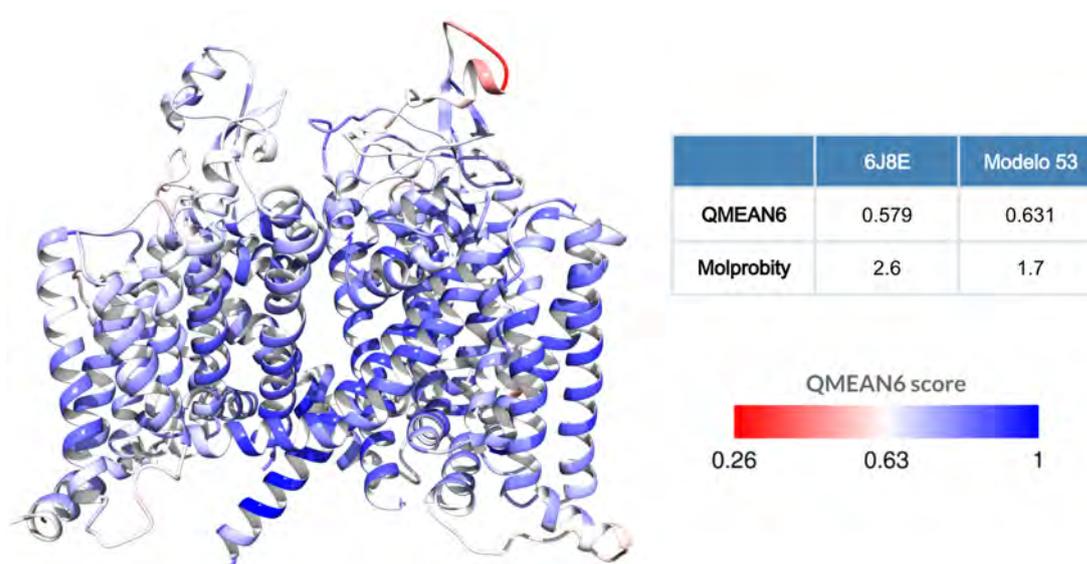


Figura 4.10: Vista lateral del modelo 53, coloreado según el QMEAN6 score, calculado para cada residuo a través del servidor QMEANBrane. En la tabla se muestran los valores de QMEAN6 score normalizado y Molprobability obtenidos para la estructura original (6J8E) y el modelo 53 generado por Rosetta. La métrica QMEAN6 score oscila entre 0 y 1, siendo este último valor el que indica un modelo de mejor calidad. Para el caso del Molprobability score, valores cercanos a 0 indican un mejor modelo.

4.3.3.2. Desarrollo y validación de un modelo de docking

Selección de la mejor estructura

Para cada programa de docking, se evaluó la *performance* del modelo que surge de dockear los compuestos en la estructura nativa, sin realizar ningún pre-procesamiento de la misma o post-procesamiento de los complejos ligando-receptor obtenidos. De esta manera se estableció una línea de base, que representa un procedimiento de docking clásico donde los ligandos son dockeados en una estructura obtenida experimentalmente.

A la hora de evaluar el desempeño individual de las estructuras del *ensemble*, así como también de las distintas funciones de puntuación, se utilizó este modelo base como referencia y se expresó la *performance* de cada estructura de manera relativa. Es decir, un valor de 10 % significa que esta permite una mejora del 10 % en la métrica TauB con respecto a la línea de base, mientras que valores negativos indican que el modelo funciona peor.

Para QVina2 se consideró solo la pose de menor energía (LE), ya que el programa no realiza el clustering de las poses como si lo hace Autodock4, mientras que para AD4-GPU se analizaron la pose de menor energía (LE) y la pose de menor energía del cluster más numeroso (LC_LE). Para cada conformación del receptor y cada función de puntuación, se tomaron 1000 muestras con reemplazo (*bootstrapping*) estratificadas, siendo el valor informado la media de este muestreo.

Para el docking con QVina2, considerando la métrica TauB, existen estructuras cuya *performance* es mayor que la nativa, aunque el porcentaje de mejora es inferior al 20 % en todos los casos. Incluso se observan estructuras que funcionan peor que la nativa, cuyos porcentajes son negativos (Figura 4.11). Minimizar los complejos con la misma función de puntuación (vina) no mejora sustancialmente los resultados. Sin embargo, el *re-scoring* con ad4 mejora, en casi todos los casos, la correlación entre los rankings. El *re-scoring* con vinardo, permite un incremento aún mayor en los valores de TauB, incluso con respecto a ad4.

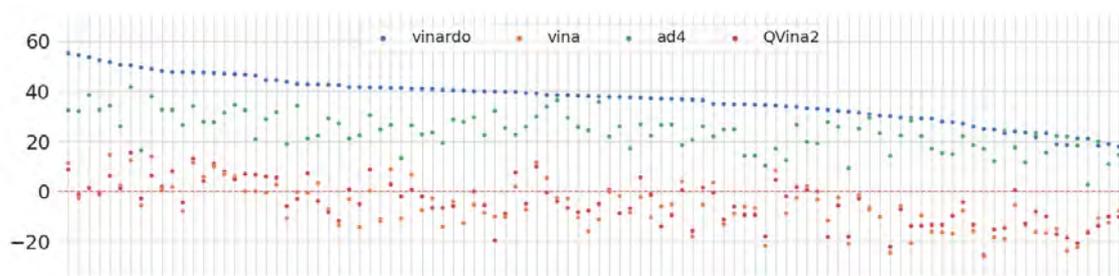


Figura 4.11: Selección de la mejor estructura. Las 102 conformaciones del canal NaV1.2 se encuentran representadas en el eje X, mientras que en el eje Y se muestra la *performance* relativa al modelo de base, utilizando el programa QVina2 y considerando la métrica TauB. Las estructuras se encuentran ordenadas en orden decreciente según su porcentaje de mejora, considerando la función de puntuación Vinardo.

El mismo análisis comparativo de todas las estructuras del *ensemble* se realizó también para AD4-GPU, considerando las poses LE y LC_LE, y la misma métrica (TauB) (Figura B.3). En este caso ad4 supera ligeramente a vinardo como función de puntuación para el *re-scoring*, pero la *performance* general de los modelos obtenidos es inferior que para QVina2.

Para cuantificar este efecto observado, se evaluó el desempeño de cada programa de docking y cada función de puntuación utilizada para re-scorear los complejos, considerando los datos de todas las estructuras del *ensemble*. Para ello, los datos provenientes de todas las estructuras se agruparon por programa de docking y función de puntuación. A su vez, para cada estructura se usaron las 1000 muestras con reemplazo generadas previamente, por lo que cada uno de los valores informados es la media de 102.000 datos (102 x 1000) (Figura 4.12).

Docking	Pose	Re-score	R	Rho	TauB
QVina2	LE	-	0.399	0.443	0.301
		vina	0.394	0.441	0.298
		vinardo	0.612	0.617	0.433
		ad4	0.520	0.564	0.392
Autodock4 GPU	LE	-	0.418	0.462	0.316
		vina	0.358	0.375	0.253
		vinardo	0.524	0.529	0.366
		ad4	0.517	0.547	0.376
	LC_LE	-	0.339	0.379	0.254
		vina	0.334	0.356	0.239
		vinardo	0.510	0.504	0.349
		ad4	0.499	0.535	0.368

Figura 4.12: *Performance* de cada programa de docking (QVina2 o Autodock4-GPU) y función de puntuación (vina, vinardo, ad4) utilizada para re-scorear los complejos obtenidos. Los datos se encuentran agrupados según la pose que se esté considerando en cada caso (LE o LC_LE).

Efectivamente se observa que, a nivel global, la estrategia de docking con QVina2 y *re-scoring* con vinardo es la mejor para este sistema en particular y por lo tanto esta será la elegida para realizar el VS prospectivo. El mejor modelo de docking obtenido consiste en dockear los compuestos en la estructura número 53 con QVina2 y luego re-scorear los complejos obtenidos con la función de puntuación vinardo. Para esta combinación, se obtiene un $R=0.681\pm 0.027$, $Rho=0.678\pm 0.028$ y $TauB=0.488\pm 0.024$, lo que representa una mejora del 63.6 %, 48.5 % y 54.4 % con respecto al modelo base, para cada métrica respectivamente. Adicionalmente, se evaluó su desempeño sobre el **Dataset B** y la **base simulada DUD-E**, para los cuales se obtuvieron AUC-ROCs de 0.790 ± 0.032 y 0.762 ± 0.039 , respectivamente.

Para darle un contexto a estos resultados, podemos mencionar el caso del *D3R grand challenge*, una competencia organizada por el *Drug Design Data Resource* (D3R), donde se publican desafíos relacionados con el desarrollo y aplicación de métodos computacionales para la predicción de afinidad y poses de unión para moléculas pequeñas. Durante el desafío de predicción de afinidad, se utiliza el co-

eficiente TauB para comparar los rankings de afinidad generados por los distintos métodos. En su última edición (año 2020), la competencia incluyó dos blancos moleculares: Cathepsin S (CatS) y beta secretase 1 (BACE). El mejor método alcanzó un TauB de 0.38 ± 0.05 para el *dataset* BACE1 (154 moléculas) y un 0.54 ± 0.02 para CatS (459 moléculas) (**Parks** et al. 2020). Nuestro modelo, por su parte, logró un $\text{TauB} = 0.488 \pm 0.024$ para el **Dataset A** (323 moléculas).

Referencias Capítulo 4

- Daly, J. y B. Witkop** (1971). “Batrachotoxin, an extremely active cardio- and neurotoxin from the Colombian arrow poison frog *Phylllobates aurotaenia*”. *Clinical Toxicology* 4.3, págs. 331-342. DOI: [10.3109/15563657108990484](https://doi.org/10.3109/15563657108990484).
- Eswar, N. et al.** (2006). “Comparative Protein Structure Modeling Using Modeller”. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]* 0 5, Unit-5.6. DOI: [10.1002/0471250953.bi0506s15](https://doi.org/10.1002/0471250953.bi0506s15).
- Gantner, M. E. et al.** (2021). “Computational and Synthetic Target-based Approaches to the Discovery of Novel Anticonvulsant Compounds”. *Current Medicinal Chemistry* 28.33, págs. 6866-6894. DOI: [10.2174/0929867328666210506161234](https://doi.org/10.2174/0929867328666210506161234).
- Kiralj, R. y M. M. C. Ferreira** (2009). “Basic validation procedures for regression models in QSAR and QSPR studies: theory and application”. *Journal of the Brazilian Chemical Society* 20.4, págs. 770-787. DOI: [10.1590/S0103-50532009000400021](https://doi.org/10.1590/S0103-50532009000400021).
- Koehler Leman, J., B. K. Mueller y J. J. Gray** (2017). “Expanding the toolkit for membrane protein modeling in Rosetta”. *Bioinformatics* 33.5, págs. 754-756. DOI: [10.1093/bioinformatics/btw716](https://doi.org/10.1093/bioinformatics/btw716).
- Linford, N. J. et al.** (1998). “Interaction of batrachotoxin with the local anesthetic receptor site in transmembrane segment IVS6 of the voltage-gated sodium channel”. *Proceedings of the National Academy of Sciences* 95.23, págs. 13947-13952. DOI: [10.1073/pnas.95.23.13947](https://doi.org/10.1073/pnas.95.23.13947).
- Lomize, M. A. et al.** (2012). “OPM database and PPM web server: resources for positioning of proteins in membranes”. *Nucleic Acids Research* 40.Database issue, págs. D370-D376. DOI: [10.1093/nar/gkr703](https://doi.org/10.1093/nar/gkr703).

- Oelstrom**, K. et al. (2014). “Evolutionarily conserved intracellular gate of voltage-dependent sodium channels”. *Nature Communications* 5.1, pág. 3420. DOI: [10.1038/ncomms4420](https://doi.org/10.1038/ncomms4420).
- Palestro**, P. H. et al. (2018). “Searching for New Leads To Treat Epilepsy: Target-Based Virtual Screening for the Discovery of Anticonvulsant Agents”. *Journal of Chemical Information and Modeling* 58.7, págs. 1331-1342. DOI: [10.1021/acs.jcim.7b00721](https://doi.org/10.1021/acs.jcim.7b00721).
- Pan**, X. et al. (2019). “Molecular basis for pore blockade of human Na⁺ channel NaV 1.2 by the α -conotoxin KIIIA”. *Science* 363.6433, págs. 1309-1313. DOI: [10.1126/science.aaw2999](https://doi.org/10.1126/science.aaw2999).
- Parks**, C. D. et al. (2020). “D3R grand challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies”. *Journal of Computer-Aided Molecular Design* 34.2, págs. 99-119. DOI: [10.1007/s10822-020-00289-y](https://doi.org/10.1007/s10822-020-00289-y).
- Wang**, S.-Y. et al. (2006). “How Batrachotoxin Modifies the Sodium Channel Permeation Pathway: Computer Modeling and Site-Directed Mutagenesis”. *Molecular Pharmacology* 69.3, págs. 788-795. DOI: [10.1124/mol.105.018200](https://doi.org/10.1124/mol.105.018200).

Capítulo 5

Cribado virtual y
ensayos *in vitro*



5.1. Campaña de cribado virtual

Con el objetivo de reposicionar fármacos ya aprobados como ACDs, se recurrió a la base de datos DrugBank5.1, la cual comprende más de 10.000 fármacos aprobadas, nutracéuticos y fármacos en fase experimental. Las moléculas se procesaron de la misma forma en que se prepararon los *datasets* para las validaciones. De esta manera, se obtuvieron 10.046 estructuras listas para dockear.

El VS se realizó en paralelo, utilizando los 4 modelos desarrollados previamente para ambos blancos moleculares.

5.1.1. Canal TRPV1

Para el caso del modelo QSAR clasificador, desarrollado para el canal TRPV1, los descriptores moleculares requeridos por el modelo no pudieron ser calculados para unas 183 moléculas, por lo tanto fueron 9.863 las que pudieron ser cribadas. Dentro de las 6.796 que se encuentran dentro del DA del modelo QSAR, 113 fueron predichas como activas por el mismo, considerando el *cutoff* por defecto de 0.5. Para el caso del modelo de docking, 769 moléculas obtuvieron un *score* ≥ 46 , por lo que fueron clasificadas como activas. De esta manera, un total de 43 compuestos fueron predichos como activos por ambos tipos de modelos (Figura 5.1).

En este caso, la correlación entre los rankings generados por los modelos no fue calculada, debido a que la probabilidad de activo que predice un modelo QSAR no guarda una relación directa con su potencia, si no con su similitud a los compuestos activos/inactivos con los que se entrenó el algoritmo. A su vez, este modelo QSAR fue desarrollado para distinguir entre compuestos activos (antagonistas) e inactivos, sin considerar como activos a los agonistas del TRPV1. Por lo tanto, es probable que un agonista actuando en el sitio de los capsaicinoides presente una alta energía de interacción predicha mediante el docking, pero que sea clasificado como inactivo por el modelo basado en la estructura de los ligandos.

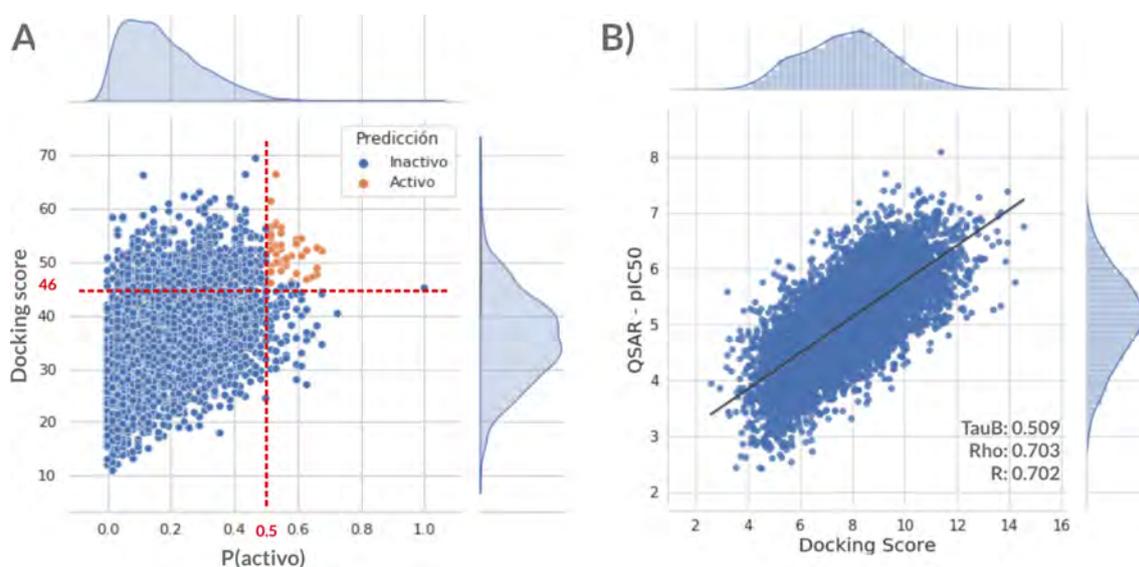


Figura 5.1: Predicciones de los modelos QSAR y docking sobre la base DrugBank 5.1. A) Canal TRPV1. En el eje X se representa la probabilidad de activo $P(\text{activo})$ calculada por el modelo QSAR, mientras que en el eje Y se muestra el *score* de docking. Solo los compuestos dentro del DA del modelo QSAR fueron considerados en el análisis (6.796). En los márgenes del gráfico se observa la distribución de las predicciones de cada modelo, mientras que en el centro se observa la distribución conjunta de las predicciones. Las líneas rojas punteadas delimitan los valores de corte utilizados para cada modelo: 46 para el docking y 0.5 para el QSAR. Las moléculas clasificadas como activas por ambos modelos se muestran en naranja mientras que las inactivas se muestran en azul. B) Canal NaV1.2. En el eje X se representan los *scores* de docking, mientras que en el eje Y se muestra el valor de pIC50 predicho por el modelo de regresión. Solo los compuestos dentro del DA del modelo QSAR fueron considerados en el análisis (6.602). En los márgenes del gráfico se observa la distribución de las predicciones de cada modelo, mientras que en el centro se observa la distribución conjunta. La recta negra representa una regresión lineal entre los valores de x e y. En ambos casos, las métricas TauB, Rho y R muestran la correlación entre ambos rankings.

5.1.2. Canal NaV1.2

Luego de calcular los descriptores moleculares requeridos por el modelo QSAR del canal NaV1.2, un total de 9.290 estructuras pudieron ser procesadas sin errores. Dentro de estas, 6.602 moléculas se encuentran dentro del DA del modelo.

Por tratarse de un modelo de regresión, no es posible contabilizar el número de *hits* predichos como en el caso del canal TRPV1. Sin embargo, es posible estimar la correlación entre los rankings generados por ambos tipos de modelos. En la imagen de la Figura 5.1, donde se muestra un gráfico de distribuciones conjuntas, puede observarse que existe una buena correlación entre las predicciones de ambos modelos (TauB=0.509, Rho=0.703 y R=0.702).

5.1.3. Selección de los candidatos

La selección de los candidatos se basó, principalmente, en el análisis conjunto de las predicciones provenientes de los 4 modelos. Sin embargo, también se tuvieron en cuenta otros factores durante este proceso, como por ejemplo la novedad del *scaffold* o andamiaje químico, el precio y la accesibilidad, los datos de farmacocinética y toxicidad disponibles, y la inspección visual de las poses de docking. Finalmente, 3 compuestos fueron seleccionados para ser evaluados experimentalmente: MTK, NVB y CNZ. Las estructuras químicas de estas moléculas junto con las predicciones de todos los modelos se muestran en la Figura 5.2.

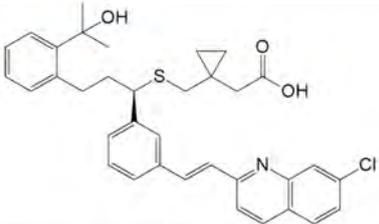
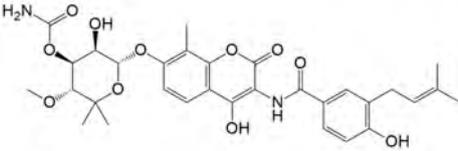
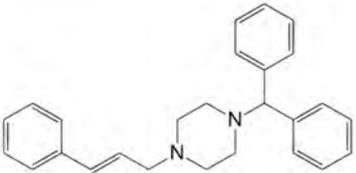
Droga	Estructura	TRPV1		NaV1.2	
		QSAR	Docking	QSAR	Docking
MTK		0.56	-58.33	5.75	-11.63
NVB		0.63	-53.24	5.38	-12.68
CNZ		0.52	-47.73	6.44	-9.12

Figura 5.2: Predicciones de los 4 modelos para los *hits* que fueron seleccionados para ser evaluados experimentalmente. Para el QSAR del canal TRPV1 se muestra la probabilidad de activo, mientras que para el canal NaV1.2 la predicción corresponde al pIC₅₀ predicho en el ensayo de desplazamiento de BTX. En cuanto al docking, se muestra el menor *score* obtenido luego del *re-scoring*, por lo que las diferencias de magnitud entre *targets* se deben a la función de puntuación utilizada: ad4 para el TRPV1 y vinardo para el canal NaV1.2.

Para el caso del canal TRPV1, la interacción de los *hits* seleccionados con el blanco molecular se estudió mediante simulaciones de MD, cuyos resultados se describen en el Capítulo 3, Sección 3.3.3.3.

En la Figura 5.3 se muestran los resultados del docking para los *hits* seleccionados sobre el canal NaV1.2. Aún cuando la región de docking se definió como una caja rectangular que incluye todo el poro del canal, desde la base hasta el filtro de selectividad (anillo DEKA), las predicciones del modelo ubican a los 3 compuestos en la misma región del poro, la cual se corresponde con el sitio de unión de ACDs y anestésicos locales, en la proximidad del dominio 3. El análisis de las interacciones revela algunos patrones comunes, como por ejemplo interacciones de tipo π -stacking

con el residuo Phe1764. A su vez, MTK y NVB forman puentes de hidrógeno con la Tyr1771. Para el caso de CNZ, se observan varias interacciones hidrofóbicas con los residuos Leu1465, Ile1469, Ile1473, Phe1764, Val1767, Val1768, Tyr1771, Ile1772 y Ile1775.

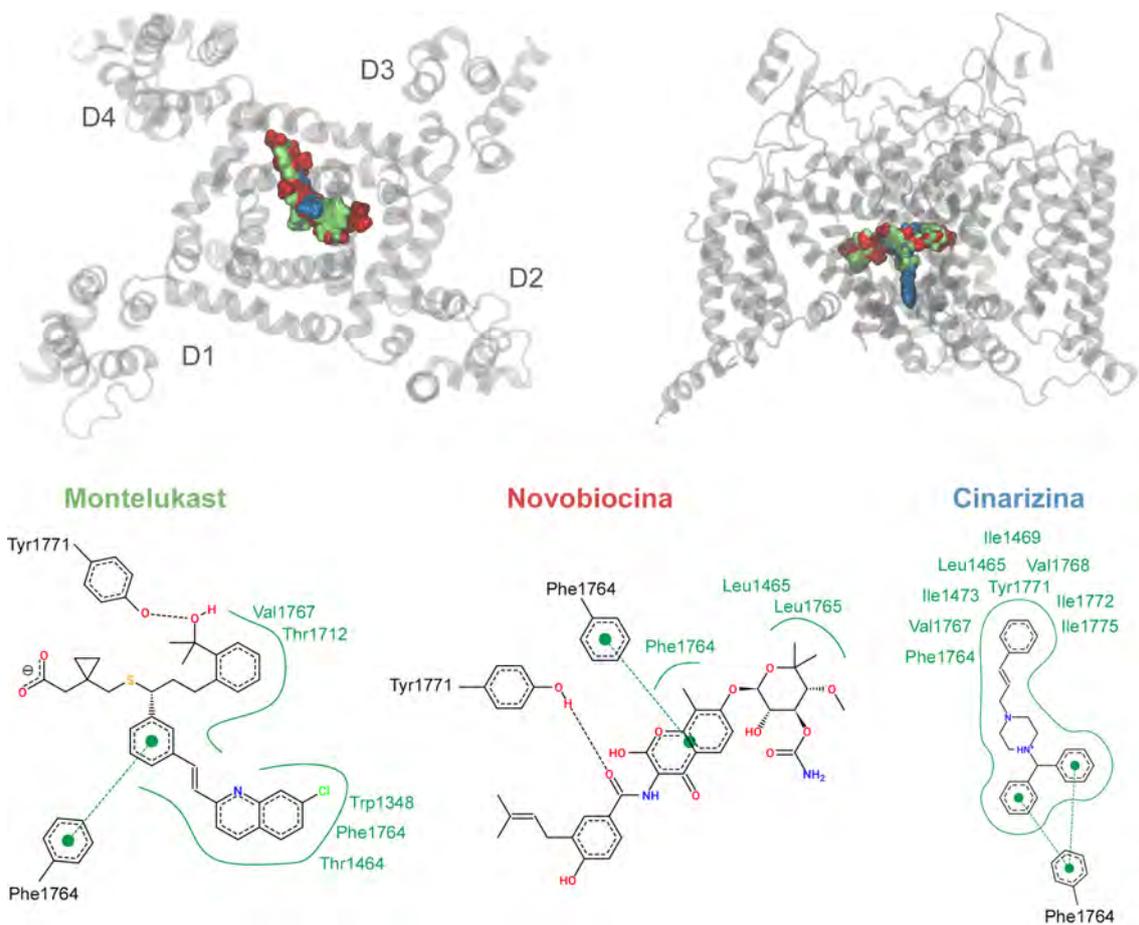


Figura 5.3: Docking de los candidatos seleccionados sobre el modelo 53 del canal NaV1.2. En el panel superior se muestra una vista superior (izquierda) y lateral (derecha) del canal, donde se puede observar la ubicación de las 3 moléculas, representadas como superficies coloreadas (MTK: verde, NVB: rojo y CNZ: azul). Para permitir una mejor visualización se utilizó un plano de corte sobre la estructura de la proteína. En el panel inferior se muestran los diagramas de interacciones 2D generados para cada ligando. Las líneas verdes sólidas muestran las interacciones de tipo hidrofóbicas, mientras que las de tipo π -stacking se representan como círculos unidos por líneas punteadas.

Montelukast

El Montelukast (MTK) es un fármaco antagonista del receptor de Cisteín Leucotrieno (CysLT), ampliamente utilizado para el tratamiento del asma bronquial y las alergias (Markham y Faulds 1998). Luego de la administración repetida de una dosis de 10 mg por vía oral en sujetos sanos, MTK alcanza una concentración plasmática máxima (CP_{max}) de aproximadamente 600 ng/mL (1 μ M) (Zhao et al.

1997).

Novobiocina

La Novobiocina (NVB), también conocida como Albamicina o Catomicina, es un antibiótico aminocumarínico producido por la bacteria *Streptomyces niveus*. Este fármaco se une a la subunidad GyrB de la ADN girasa bacteriana inhibiendo la actividad ATPasa de la misma, lo que le impide a la bacteria sintetizar nuevo ADN (Maxwell 1993).

En Estados Unidos, la formulación oral de NVB fue retirada del mercado por falta de efectividad (como antibiótico). Sin embargo, estudios de fase I y II realizados en pacientes con cáncer recibiendo distintas quimioterapias, demostraron que este fármaco alcanza concentraciones plasmáticas de al menos 150 μM durante 24 h luego de administrar una dosis oral de 5.5 g, sin presentar reacciones de toxicidad serias (Hahm et al. 2000; Murren et al. 2000).

Este antibiótico natural no es sustrato de la PgP y es capaz de inhibir transportadores de eflujo como la Breast Cancer-Resistance Protein (BCRP). De hecho, es utilizado como *tool compound* en ensayos de transporte que involucran a la BCRP (Su et al. 2007; Suzuki et al. 2019).

Cinarizina

La Cinarizina (CNZ) pertenece a la familia de las piperazinas y es comúnmente utilizado en el tratamiento de las náuseas y vómitos asociados con el movimiento, el vértigo o la quimioterapia. Posee propiedades antihistamínicas y bloqueante de canales de calcio, por lo que puede producir sedación (Kirtane et al. 2019). Luego de su administración oral (75 mg), CNZ alcanza una CP_{max} de 275 ± 36 ng/mL (0.746 μM) a un tiempo estimado de 3.0 ± 0.5 h y su tiempo de vida media es de 23.6 ± 3.2 h (Castañeda-Hernández et al. 1993).

5.2. La técnica de patch clamp

5.2.1. Generalidades

La membrana celular, por su capacidad de separar y acumular cargas eléctricas de signo contrario a cada uno de sus lados, presenta un comportamiento capacitivo. A su vez, la actividad de los canales iónicos que se encuentran embebidos en ella y que permiten el paso de iones le confieren a la membrana un componente resistivo/conductivo. La capacitancia y la conductividad, denominadas características eléctricas pasivas, tienen un papel esencial en la generación y regulación del potencial de transmembrana que participa en la regulación de la actividad celular, tanto de células excitables (que disparan potenciales de acción) como de células no excitables (**Kornreich 2007**).

La técnica de patch clamp, introducida por Neher y Sakmann en 1976 (**Sakmann y Neher 1984**), permite estudiar estas características y registrar tanto corrientes a través de los canales iónicos insertos en la membrana celular, como cambios en el potencial de membrana. Existen dos grandes variantes de esta técnica: el clampeo de voltaje o *voltage-clamp*, donde se mantiene fijo el potencial de membrana y se miden las variaciones en la corriente; y el clampeo de corriente o *current-clamp*, donde se fija la corriente y registran las variaciones de potencial de la membrana. La primera de estas variantes es la más frecuentemente utilizada para el estudio del comportamiento de los canales iónicos, y la que fue empleada en esta tesis.

La técnica consiste en el establecimiento de un sello de alta resistencia o *giga-seal* (de aproximadamente $10\text{ G}\Omega$), entre el vidrio de una micropipeta y la membrana celular, denominado *patch* o parche. En el modo de *voltage-clamp*, el amplificador operacional de patch clamp permite fijar sobre ese parche una diferencia de potencial a ambos lados de la membrana, midiendo en simultáneo la corriente que atraviesa la misma con una resolución de hasta $4\text{ }\mu\text{s}$.

Las células de interés se observan con un microscopio estabilizado mecánicamente sobre una mesa antivibratoria y recubierto con una caja de Faraday, para filtrar los ruidos eléctricos del ambiente. La micropipeta de vidrio con una punta cercana a $1\text{ }\mu\text{m}$ se aproxima a la superficie de la membrana de la célula elegida

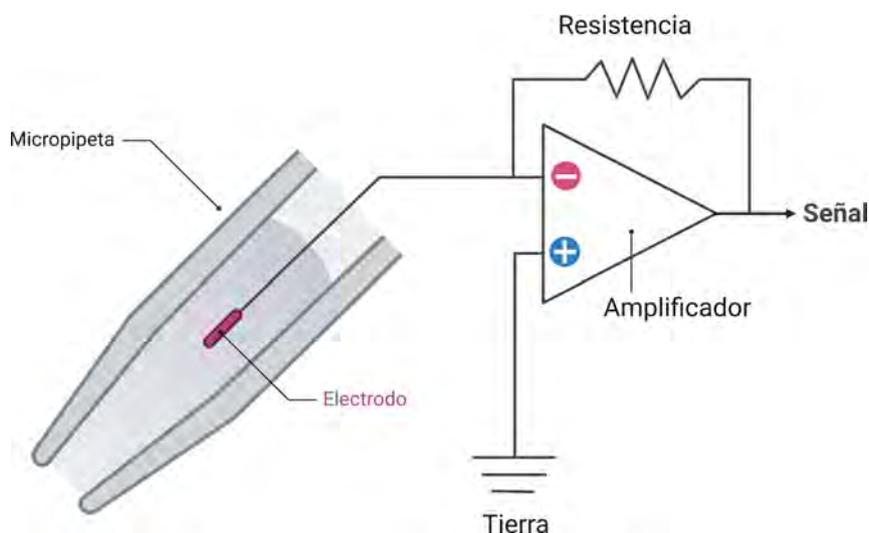


Figura 5.4: Esquema básico de un circuito de patch clamp.

para el establecimiento del sello por medio de un micromanipulador. Para poder fijar o clampar el potencial del parche de membrana y medir simultáneamente la corriente que lo atraviesa, se utilizan dos electrodos de plata recubiertos en cloruro de plata: uno se encuentra dentro de la micropipeta de vidrio (Figura 5.4) y el otro, denominado “tierra”, en la solución del baño que recubre a las células. Inicialmente la corriente fluye directamente entre ambos electrodos, pero cuando la micropipeta toca la membrana celular la resistencia a este flujo se ve incrementada. Luego, ejerciendo una suave succión desde la pipeta, se logra establecer un sello de mayor resistencia o *giga-seal*, lo que evita el flujo de corriente directo entre la pipeta y el baño, permitiendo así fijar el potencial de ese parche y medir las corrientes que lo atraviesan, reduciendo significativamente el ruido en el registro (Sakmann y Neher 1984). Esta configuración se conoce como *cell-attached* (CA), y a partir de ella es posible derivar en otras configuraciones conocidas como *whole-cell* (WC), *inside-out* (IO) u *outside-out* (OO) (Kornreich 2007), las cuales se encuentran esquematizadas en la Figura 5.5.

La diferencia principal entre las configuraciones radica en la persistencia o no de la integridad celular y qué cara de la membrana está en contacto con la solución de la micropipeta o el baño (esta última puede a su vez ser modificada durante el experimento). Además, en la configuración WC se registran corrientes macroscópicas representativas de la totalidad de canales iónicos presentes en la membrana celular, mientras que las configuraciones CA, IO y OO permiten registrar la corriente de un único canal iónico.

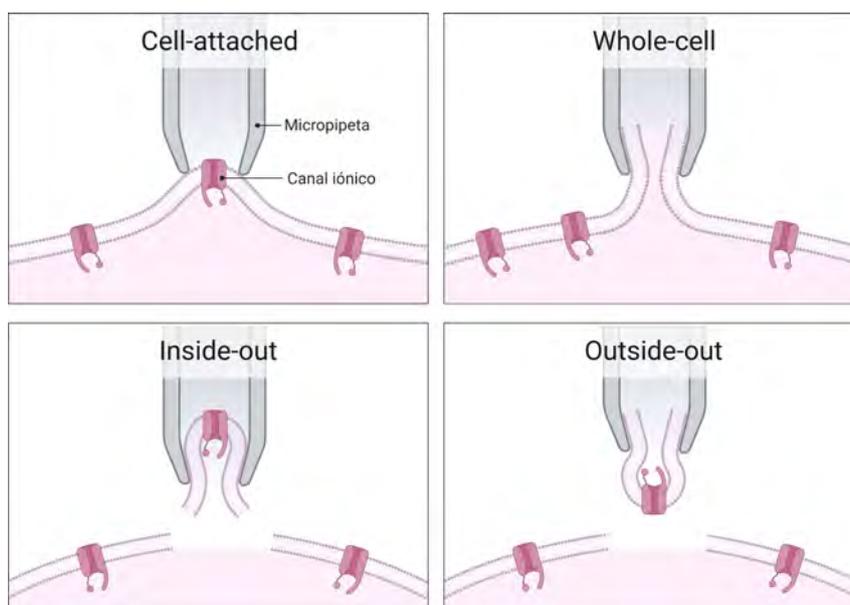


Figura 5.5: Esquema de las configuraciones de patch clamp comúnmente utilizadas.

En esta tesis doctoral se utilizó únicamente la configuración de WC. Para obtener la misma, una vez obtenido el CA, se realiza una segunda succión que remueve la porción de membrana celular del área bajo la punta de la pipeta, de modo de generar una continuidad entre el citosol y la solución de la pipeta (Figura 5.5). En esta configuración los componentes celulares de mayor tamaño permanecen en la célula, pero se dializa la misma con la solución de la pipeta de manera que hay un completo control de las condiciones iónicas del medio intracelular.

5.2.2. Condiciones experimentales de registro

Para ambos canales iónicos las corrientes se registraron en células HEK293 como sistema de expresión heterólogo. Las mismas se cultivaron en un medio esencial mínimo (DMEM) (GIBCO) con alta glucosa y 10% de suero fetal bovino (SFB) (Internegocios), en atmósfera de CO_2 al 5% a 37 °C y con una humedad del 95%. Los repicados se realizaron cuando las mismas alcanzaron un 70-80% de confluencia.

Las soluciones ensayadas fueron perfundidas con una pipeta multivías colocada en la proximidad de la célula investigada. Luego de cada experimento en una sola célula, la placa fue reemplazada por otra que contenía una nueva muestra. Todos los experimentos se realizaron a temperatura ambiente (22°C).

Las pipetas de vidrio WPI PG52165-4 se obtuvieron mediante un *puller* vertical

(Pul-100, World Precision Instruments), con resistencias en el rango 2-4 M Ω . Las corrientes se filtraron mediante un filtro de tipo Bessel de 8 polos (Multiclamp 700A amplifier, Molecular Devices) a una frecuencia de 2 kHz y se digitalizaron a una frecuencia de muestreo de 20 kHz para TRPV1 y de 200 kHz para NaV1.2 (Digidata 1322A, Axon Instruments).

5.3. Registros en canales TRPV1

5.3.1. Métodos

La actividad *in vitro* de los compuestos seleccionados se evaluó a través su capacidad de inhibir las corrientes del canal TRPV1 activadas por CAP en un sistema de expresión heterólogo, mediante la técnica de patch clamp.

Como sistema de expresión se utilizaron células HEK293 transfectadas transientemente con el canal TRPV1. Los cADN de la subunidad α de la isoforma rTRPV1, junto a la proteína verde fluorescente (eGFP, del inglés enhanced Green Fluorescent Protein) -utilizada como reportero de la transfección-, se co-transfectaron utilizando el método de co-precipitación con fosfato de calcio (**Chen y Okayama 1987; Martín et al. 2021**). Las medidas electrofisiológicas se realizaron 18 h después de la transfección.

Las células se observaron empleando un microscopio invertido de epifluorescencia (Arcano) mecánicamente estabilizado, equipado con una lente objetivo de 40X. Las corrientes macroscópicas se registraron en la configuración de WC en células que evidenciaban la expresión de eGFP (**Hamill et al. 1981**).

La solución de registro extracelular del baño y la intracelular de la pipeta fue la misma conteniendo (mM): 130 NaCl, 5.4 KCl, 1.2 MgCl₂, 5 HEPES, 6 glucosa, 1 EGTA; pH 7.4 ajustado con NaOH. La CAP y los compuestos a ensayar se adicionaron en la solución extracelular.

El protocolo de voltaje utilizado consistió en un pulso de 500 ms desde un potencial de mantenimiento de 0 mV hasta +100 mV, para lograr una activación

sostenida, seguido de una rampa de voltaje hiperpolarizante de 0.1 mV/ms hasta -100 mV, buscando registrar la actividad de los canales a distintos potenciales de membrana (Figura 5.6). Para evaluar la capacidad de los compuestos de inhibir la activación mediada por CAP, la corriente se activó utilizando CAP 250 nM. Luego de que la corriente se estabilizó, se perfundió la solución extracelular conteniendo CAP 250 nM y el compuesto a ensayar a una concentración de 100 nM.

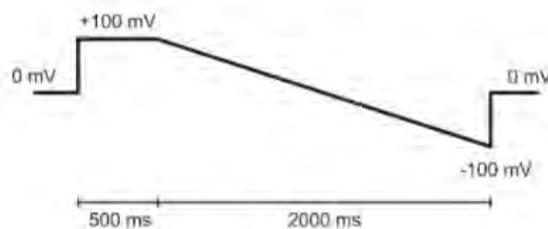


Figura 5.6: Protocolo de voltaje utilizado para registrar, en la configuración de WC, las corrientes activadas por CAP en los canales TRPV1.

Los resultados se expresaron como la media \pm SEM (del inglés, Standard Error of the Mean). Para evaluar la significancia estadística de la inhibición se utilizó un test t de Student frente a un valor hipotético del 0% de inhibición. Para comparar la inhibición a diferentes voltajes se utilizó la prueba t de Student para datos apareados (Graphpad Prism v8). En todos los casos se consideró un valor de p inferior a 0,05 para establecer diferencias estadísticamente significativas.

5.3.2. Resultados

Todos los compuestos evaluados demostraron actividad inhibitoria nanomolar sobre las corrientes activadas por CAP en canales TRPV1. Para los 3 compuestos, el% de inhibición a una concentración de 100 nM y un potencial de -100 mV fue mayor al 50%. Además, para el caso de MTK y NVB esta inhibición presenta una marcada voltaje-dependencia, siendo más pronunciada a potenciales más hiperpolarizados (Figura 5.7-D).

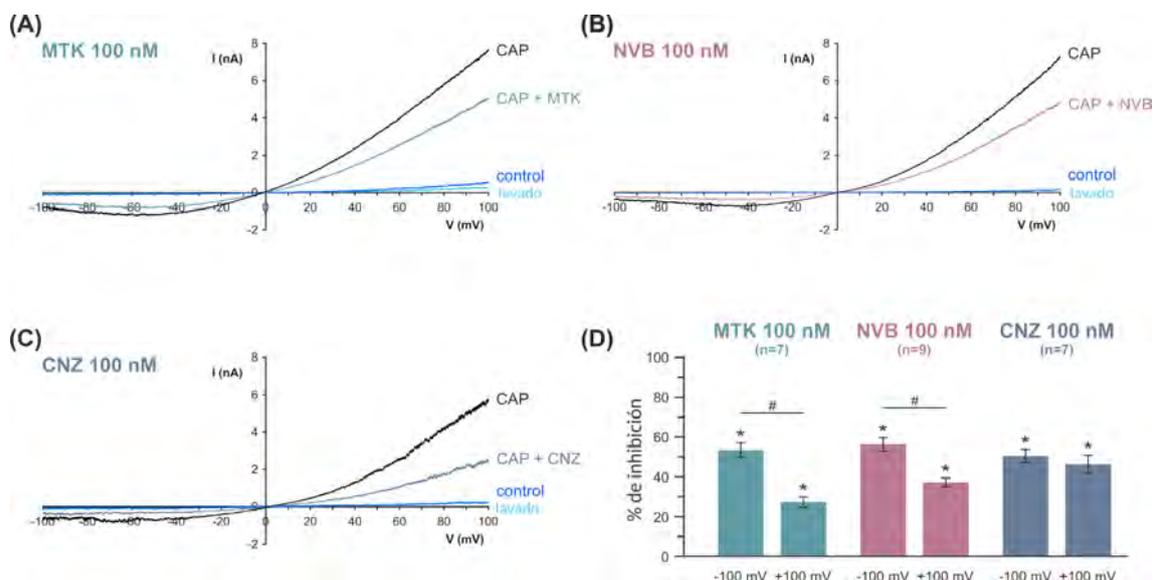


Figura 5.7: Inhibición de las corrientes activadas por CAP en canales TRPV1. A-C) Curvas de relación corriente-voltaje (IV) obtenidas en células HEK293 expresando transientemente canales TRPV1, en la configuración WC, en condiciones control, en presencia de CAP 250 nM, y luego de perfundir CAP junto a 100 nM de MTK (A), NVB (B) o CNZ (C). D) Valores medios \pm SEM para la inhibición de las corrientes activadas por CAP: MTK a -100 mV 53.4 ± 7.1 %, a +100 mV 27.3 ± 5.3 % (n=7); NBV a -100 mV 56.3 ± 6.7 %, a +100 mV 37.3 ± 4.2 % (n=9); CNZ a -100 mV 50.4 ± 9.2 %, a +100 mV 46.3 ± 6.8 % (n=7). * indica una diferencia estadísticamente significativa con una inhibición del 0 % ($p < 0.05$), mientras que # indica diferencias significativas entre la inhibición a -100 mV vs. +100 mV ($p < 0.05$).

Durante el transcurso de esta tesis, Liang y colaboradores reportaron que NVB inhibe el transporte de CAP a través de la membrana intestinal, mediado por el canal TRPV1, de manera dosis-dependiente. Mediante experimentos de qRT-PCR y *Western Blot*, encontraron que NVB es capaz de disminuir la expresión de este canal a nivel intestinal. Los mismos efectos se observaron para SB-705498, un inhibidor de referencia para los canales TRPV1. Debido a que NVB no fue capaz de permear a través de la pared intestinal, como si lo hace CAP, los autores sugieren que NVB no es un inhibidor competitivo sino que simplemente inhibe la función de transporte del TRPV1, aun cuando el efecto de NVB sobre la funcionalidad del canal no fue medido en forma directa (Liang et al. 2018).

Por su parte, nuestros modelos predijeron que NVB es capaz de unirse al sitio

de los capsaicinoides el canal TRPV1 y los experimentos patch clamp confirmaron que NVB antagoniza directamente la activación del canal mediada por CAP.

5.4. Registros en canales NaV1.2

5.4.1. Métodos

Los *hits* seleccionados fueron evaluados por su capacidad de modular a los canales NaV1.2, en un sistema de expresión heterólogo, mediante la técnica de patch clamp.

Se utilizaron células HEK293 con expresión estable de la subunidad del canal NaV1.2, generosamente provistas por GlaxoSmithKline, Stevenage, Reino Unido. Las condiciones de cultivo fueron similares a las descritas en la sección anterior, pero suplementando el medio de cultivo con 0,5 % de sulfato de Geneticina G418 (Aldrich), de modo de ejercer una presión de selección sobre las células. Dada la expresión estable, en este caso no se usó gen reportero sino que se seleccionó la célula a ensayar en base al aspecto de su membrana plasmática.

Las corrientes macroscópicas se registraron en la configuración de WC, utilizando como soluciones de registro:

- Solución extracelular del baño (SE) en mM: NaCl 140, $CaCl_2$ 2, $MgCl_2$ 1, HEPES 10 y glucosa 11; El pH se ajustó a 7.4 con NaOH.
- Solución intracelular de la pipeta (SI) en mM: CsF 100, CsCl 40, EGTA 10, HEPES 10, NaCl 5, $MgCl_2$ 2, Na_2 .ATP 4; El pH se ajustó a 7.3 con CsOH.

La actividad de los compuestos sobre los canales NaV1.2, se evaluó a través de los siguientes protocolos de voltaje, validados por Qiao y colaboradores en el mismo sistema de expresión (Qiao et al. 2014):

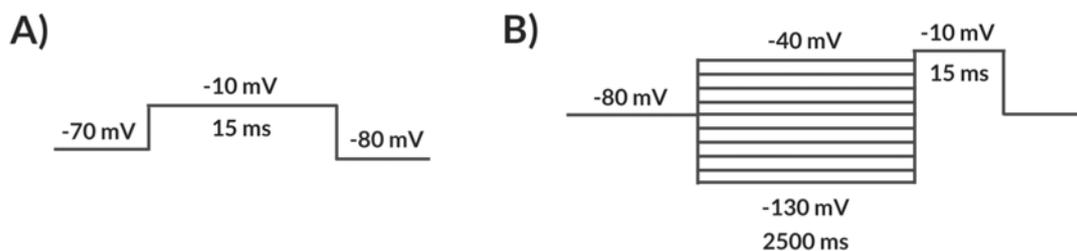


Figura 5.8: Protocolos de voltaje utilizados durante los registros en canales NaV1.2. A) Inhibición de la corriente al pico B) Inactivación de estado estacionario (curva h).

Inhibición de la corriente al pico

Luego de obtener la configuración de WC, el potencial de membrana se fija en -70 mV. Desde este potencial de mantenimiento o *holding*, se generan pulsos rectangulares (15 ms) despolarizantes hasta -10 mV, a intervalos de 10 s (Figura 5.8-A). Una vez que la corriente se estabiliza, este protocolo se aplica en presencia del compuesto a evaluar. Idealmente se espera que, un compuesto con actividad inhibitoria sobre los canales NaV1.2, disminuya el valor de la corriente al pico. Este protocolo se utilizó como pulso de seguimiento para evaluar la estabilidad de la corriente en todo momento, monitoreando la amplitud del pico de corriente generado por el pulso despolarizante. Además, permite evaluar la actividad de los compuestos sobre la corriente neta global, sin importar si los mismos presentan mayor afinidad por los estados cerrados o inactivados de canal.

Curvas de inactivación de estado estacionario (curva h)

Este registro consiste en un protocolo de doble pulso de voltaje. Desde un potencial de mantenimiento de -80 mV, se aplica una serie de pulsos pre-condicionantes de 2500 ms desde -130 mV hasta -40 mV, cada uno seguido de un pulso test de 15 ms a -10 mV (valor de voltaje donde la corriente de sodio tiene su máxima activación en estas células) (Figura 5.8-B). Las curvas h control y en presencia de los compuestos estudiados se construyen graficando el valor del pico de corriente registrado en respuesta al pulso test (normalizado por el valor máximo de corriente) en función del voltaje del pulso pre-condicionante. Esta curva representa la fracción de canales

disponibles a cada valor de potencial estudiado y puede ajustarse con una ecuación de Boltzmann (Ecuación 5.1). En este caso el $V_{1/2}$ representa el valor de potencial en el cual el 50% de los canales de sodio están disponibles para ser activados (en estado cerrado o de *resting*). La reducción en la fracción disponible se debe a que la recuperación del estado inactivado es tiempo y voltaje-dependiente, por lo que la curva h refleja esta última dependencia.

$$\frac{I}{I_{max}} = \frac{1}{1 + \exp\left(-\frac{V_{1/2} - V}{k}\right)} \quad (5.1)$$

Ecuación 5.1. Cálculo de la fracción disponible. En donde I es la corriente registrada al potencial de clampeo V , I_{max} es la corriente máxima registrada en cada condición experimental (control o tratado), $V_{1/2}$ es el valor de voltaje en donde el 50% de los canales está disponible y k refleja la sensibilidad al voltaje del proceso de inactivación.

Los resultados se expresaron como la media \pm SEM (del inglés, Standard Error of the Mean). Para evaluar la significancia estadística de la inhibición neta se utilizó una prueba t de Student vs. una inhibición del 0%, mientras que para los cambios en el $V_{1/2}$ la misma se realizó vs. un corrimiento de 0 mV. Por su parte, para evaluar las diferencias entre concentraciones se utilizó la prueba t de Student para datos apareados (Graphpad Prism v8). En todos los casos se consideró un valor de p inferior a 0,05 para establecer diferencias estadísticamente significativas.

5.4.2. Resultados

Los *hits* seleccionados fueron evaluados a concentraciones de 1 y 10 μ M, utilizando los protocolos de voltaje descritos anteriormente.

Para evaluar la inhibición neta de la corriente del canal NaV1.2, se calculó el porcentaje de inhibición cuando la misma es evocada por un pulso despolarizante desde un potencial de membrana de mantenimiento en valores cercanos a una condición fisiopatológica de mayor excitabilidad (-70 mV). El % de inhibición a cada concentración se calculó midiendo el valor de la corriente al pico luego de perfundir el compuesto y comparándolo con la situación control.

De los tres compuestos ensayados, MTK (1 μM : $73.1 \pm 11.6\%$, $n=6$; 10 μM : $95.3 \pm 4.2\%$, $n=6$) y CNZ (1 μM : $71.6 \pm 3.0\%$, $n=7$; 10 μM : $98.1 \pm 1.5\%$, $n=6$) presentaron una potente actividad inhibitoria sobre los canales NaV1.2 (Figura 5.9). Para el caso de NVB, se observó una tendencia a la inhibición de la corriente, si bien no se alcanzaron diferencias significativas con una inhibición del 0% a las dosis ensayadas (1 μM : $9.7 \pm 8.3\%$, $n=5$; 10 μM : $16.1 \pm 7.8\%$, $n=5$).

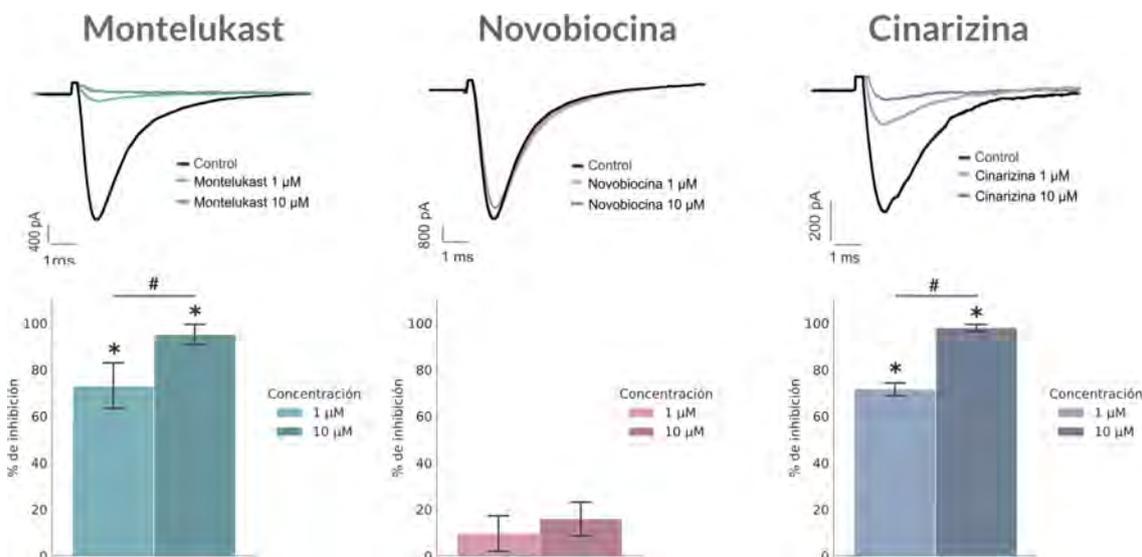


Figura 5.9: Inhibición de las corrientes de sodio en canales NaV1.2. En el panel superior se muestran los registros típicos de las corrientes registradas para la isoforma hNaV1.2 expresada en células HEK293, en la configuración de WC, en condiciones control y a las concentraciones 1 y 10 μM de cada droga. La corriente fue evocada mediante un pulso de 15 ms a -10 mV desde un potencial de mantenimiento de -70 mV. El trazo coloreado en cada caso fue el que se utilizó para calcular el % de inhibición a cada concentración, cuyos valores medios \pm SEM se muestran en los gráficos de barra del panel inferior. Los valores de inhibición de las corrientes de sodio fueron: MTK a 1 μM $73.1 \pm 11.6\%$ ($n=6$), a 10 μM $95.3 \pm 4.2\%$ ($n=6$); CNZ a 1 μM $71.6 \pm 3.0\%$ ($n=7$), a 10 μM $98.1 \pm 1.5\%$ ($n=6$); NBV a 1 μM $9.7 \pm 8.3\%$ ($n=5$), a 10 μM $16.1 \pm 7.8\%$ ($n=5$). * indica una diferencia estadísticamente significativa con una inhibición del 0% ($p<0.05$), mientras que # indica diferencias significativas entre la inhibición a 1 μM vs. 10 μM ($p<0.05$).

En las mismas células se registró además la dependencia del voltaje para la inactivación del estado estacionario, el cuál permite evaluar la habilidad de generar un bloqueo dependiente de estado, característico de muchas ACDs. A partir de la

corrientes evocadas por el protocolo de doble pulso descrito en la sección anterior, se construyó la curva h de fracción disponible vs. voltaje. El $V_{1/2}$ de cada curva h se estimó a partir del ajuste de la función de Boltzmann, y el $\Delta V_{1/2}$ se calculó como la diferencia entre la condición control y luego de perfundir la droga en estudio.

Como puede observarse en la Figura 5.10, todos los *hits* presentaron una inhibición dependiente del estado debido a la estabilización del estado inactivado, la cual se evidencia como un desplazamiento hacia la izquierda en las curvas h : MTK a 1 μM -9.55 ± 1.88 mV (n=7), a 10 μM -25.66 ± 3.21 mV (n=6); CNZ a 1 μM -10.30 ± 2.19 mV (n=7), a 10 μM -30.75 ± 3.00 mV (n=5); NBV a 1 μM -0.85 ± 0.39 mV (n=5), a 10 μM -3.32 ± 0.86 (n=5). Sin embargo, para NVB este corrimiento solo presentó diferencias significativas con un corrimiento de 0 mV a la mayor concentración ($p < 0.05$). La voltaje-dependencia de este bloqueo se evidencia en los gráficos de barra del panel inferior en la Figura 5.10, donde puede observarse que el % de inhibición de los compuestos es mayor a potenciales más despolarizados.

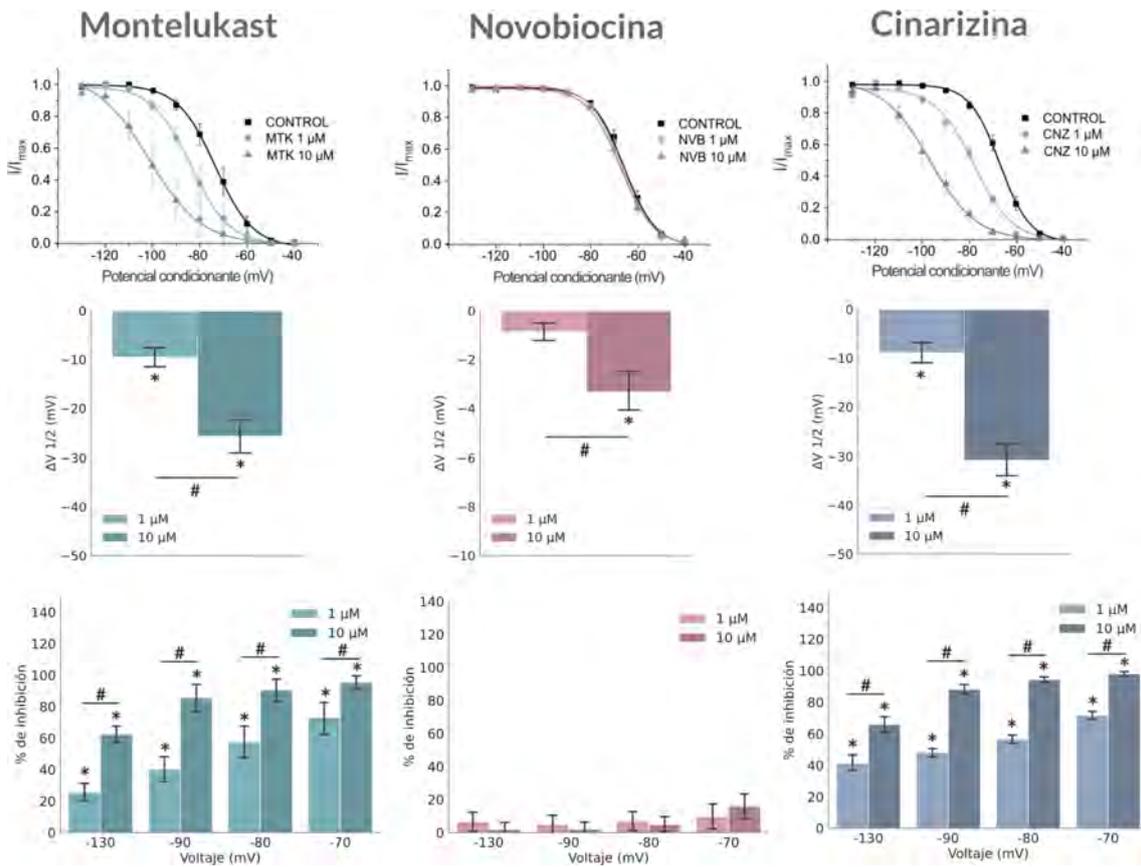


Figura 5.10: Inactivación del estado estacionario en canales NaV1.2. En el panel superior se muestran, para cada compuesto, las curvas h obtenidas para las condiciones control y luego de perfundir la droga a las concentraciones 1 y 10 μ M. El panel medio muestra los valores de $\Delta V_{1/2}$, estimados para cada droga y concentración luego de ajustar los datos con una función de Boltzmann. En el panel inferior los gráficos de barras muestran el bloqueo dependiente del voltaje, para cada candidato, a las dos concentraciones evaluadas. * indica una diferencia estadísticamente significativa con respecto a un corrimiento de 0 mV ($p < 0.05$), mientras que el # indica una diferencia significativa entre las concentraciones ($p < 0.05$).

Referencias Capítulo 5

- Castañeda-Hernández, G.** et al. (1993). “Pharmacokinetics of cinnarizine after single and multiple dosing in healthy volunteers”. *Arzneimittel-Forschung* 43.5, págs. 539-542.
- Chen, C.** y **H. Okayama** (1987). “High-efficiency transformation of mammalian cells by plasmid DNA”. *Molecular and Cellular Biology* 7.8, págs. 2745-2752. DOI: [10.1128/mcb.7.8.2745-2752.1987](https://doi.org/10.1128/mcb.7.8.2745-2752.1987).
- Hahm, H. A.** et al. (2000). “Novobiocin in combination with high-dose chemotherapy for the treatment of advanced breast cancer: a phase 2 study”. *Biology of Blood and Marrow Transplantation: Journal of the American Society for Blood and Marrow Transplantation* 6.3A, págs. 335-343. DOI: [10.1016/s1083-8791\(00\)70059-0](https://doi.org/10.1016/s1083-8791(00)70059-0).
- Hamill, O. P.** et al. (1981). “Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches”. *Pflügers Archiv* 391.2, págs. 85-100. DOI: [10.1007/BF00656997](https://doi.org/10.1007/BF00656997).
- Kirtane, M. V.** et al. (2019). “Cinnarizine: A Contemporary Review”. *Indian Journal of Otolaryngology and Head & Neck Surgery* 71.Suppl 2, págs. 1060-1068. DOI: [10.1007/s12070-017-1120-7](https://doi.org/10.1007/s12070-017-1120-7).
- Kornreich, B. G.** (2007). “The patch clamp technique: principles and technical considerations”. *Journal of Veterinary Cardiology: The Official Journal of the European Society of Veterinary Cardiology* 9.1, págs. 25-37. DOI: [10.1016/j.jvc.2007.02.001](https://doi.org/10.1016/j.jvc.2007.02.001).
- Liang, Q.** et al. (2018). “Novobiocin, a Newly Found TRPV1 Inhibitor, Attenuates the Expression of TRPV1 in Rat Intestine and Intestinal Epithelial Cell Line IEC-6”. *Frontiers in Pharmacology* 9. DOI: [10.3389/fphar.2018.01171](https://doi.org/10.3389/fphar.2018.01171).

- Markham, A. y D. Faulds** (1998). “Montelukast”. *Drugs* 56.2, 251-6, discussion 257. DOI: [10.2165/00003495-199856020-00010](https://doi.org/10.2165/00003495-199856020-00010).
- Martín, P. et al.** (2021). “Arachidonic acid effect on the allosteric gating mechanism of BK (Slo1) channels associated with the 1 subunit”. *Biochimica Et Biophysica Acta. Biomembranes* 1863.9, pág. 183550. DOI: [10.1016/j.bbamem.2021.183550](https://doi.org/10.1016/j.bbamem.2021.183550).
- Maxwell, A.** (1993). “The interaction between coumarin drugs and DNA gyrase”. *Molecular Microbiology* 9.4, págs. 681-686. DOI: [10.1111/j.1365-2958.1993.tb01728.x](https://doi.org/10.1111/j.1365-2958.1993.tb01728.x).
- Murren, J. R. et al.** (2000). “Phase I and pharmacokinetic study of novobiocin in combination with VP-16 in patients with refractory malignancies”. *Cancer Journal (Sudbury, Mass.)* 6.4, págs. 256-265.
- Qiao, X. et al.** (2014). “Properties of human brain sodium channel -subunits expressed in HEK293 cells and their modulation by carbamazepine, phenytoin and lamotrigine”. *British Journal of Pharmacology* 171.4, págs. 1054-1067. DOI: [10.1111/bph.12534](https://doi.org/10.1111/bph.12534).
- Sakmann, B. y E. Neher** (1984). “Patch clamp techniques for studying ionic channels in excitable membranes”. *Annual Review of Physiology* 46, págs. 455-472. DOI: [10.1146/annurev.ph.46.030184.002323](https://doi.org/10.1146/annurev.ph.46.030184.002323).
- Su, Y. et al.** (2007). “Using novobiocin as a specific inhibitor of breast cancer resistant protein to assess the role of transporter in the absorption and disposition of topotecan”. *Journal of Pharmacy & Pharmaceutical Sciences: A Publication of the Canadian Society for Pharmaceutical Sciences, Societe Canadienne Des Sciences Pharmaceutiques* 10.4, págs. 519-536. DOI: [10.18433/j3qp4w](https://doi.org/10.18433/j3qp4w).
- Suzuki, K. et al.** (2019). “Usefulness of novobiocin as a selective inhibitor of intestinal breast cancer resistance protein (Bcrp) in rats”. *Xenobiotica* 0.0, págs. 1-7. DOI: [10.1080/00498254.2019.1708514](https://doi.org/10.1080/00498254.2019.1708514).
- Zhao, J. J. et al.** (1997). “Pharmacokinetics and bioavailability of montelukast sodium (MK-0476) in healthy young and elderly volunteers”. *Biopharmaceutics & Drug Disposition* 18.9, págs. 769-777. DOI: [10.1002/\(sici\)1099-081x\(199712\)18:9<769::aid-bdd60>3.0.co;2-k](https://doi.org/10.1002/(sici)1099-081x(199712)18:9<769::aid-bdd60>3.0.co;2-k).

Capítulo 6

Actividad anticonvulsiva.

Ensayos *in vivo*



6.1. Introducción

Durante las etapas preclínicas del desarrollo de nuevas ACDs se utilizan modelos animales, los cuales se seleccionan, diseñan y conducen cuidadosamente, de manera de reproducir total o parcialmente las características clínicas de la enfermedad, para luego trasladar los resultados y conclusiones obtenidos a sistemas más complejos, como el ser humano u otras especies animales (**Löscher 2017**).

Desde hace más de 40 años, el Instituto Nacional de Trastornos Neurológicos y Accidentes Cerebrovasculares de Estados Unidos, dependiente del National Institute of Health (NIH), ha fijado los lineamientos para la identificación y desarrollo de nuevas ACDs a través de diferentes programas. El programa vigente hoy en día, renombrado como Programa de Detección de Terapias para la Epilepsia (ETSP) en 2016, distingue dos etapas principales en la búsqueda de tratamientos para la Epilepsia farmacorresistente a nivel preclínico. Durante la primera fase, denominada de identificación, los candidatos son sometidos a una serie de diversos ensayos, de modo de identificar compuestos promisorios con potencial de superar a las ACDs conocidas. Posteriormente, durante la fase de diferenciación, los *hits* son evaluados en modelos crónicos de Epilepsia, cuyo costo y complejidad son considerablemente mayores, pero permiten reflejar varios aspectos clínicos de la patología (**Kehne et al. 2017**). La fase de identificación propone dos modelos de crisis agudas en ratones sanos: el MES test y 6Hz test. A su vez, se recomienda también evaluar el potencial neurotóxico de los compuestos mediante el Rotorod test (Löscher, 2011). Existen, sin embargo, una gran variedad de modelos animales que pueden utilizarse para caracterizar el perfil anticonvulsivo de un compuesto (**Löscher 2017**).

Con el objetivo de evaluar la acción anticonvulsiva *in vivo* de los compuestos seleccionados, en este trabajo de tesis se realizaron diversos estudios en animales, los cuales estuvieron a cargo de investigadores especialistas en la materia. Los ensayos pertenecientes a la fase de identificación del protocolo: MES, 6 Hz, y Rotorod fueron llevados a cabo en nuestro laboratorio (LIDeB). Con el fin de aumentar el conocimiento sobre el perfil anticonvulsivo de los compuestos, se incluyó el ensayo PTZ test, también realizado en nuestro laboratorio. Adicionalmente, los compuestos se ensayaron en un modelo de Epilepsia genética audiogénica en hámsters (GASH:Sal).

Estos ensayos se realizaron en el Laboratorio de Trastornos Audiomotores de la Universidad de Salamanca - España, como producto de una colaboración establecida con el grupo que lidera la Dra. Dolores López García. A continuación se describirán las características de los modelos utilizados.

6.2. Condiciones experimentales de los ensayos *in vivo*

Los ensayos llevados a cabo en el LIDeB fueron realizados bajo la aprobación de los protocolos correspondientes por parte del Comité Institucional de Cuidado y Uso de Animales de Laboratorio (CICUAL), de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata (FCE-UNLP), a través del cual se garantiza que las evaluaciones se desarrollen de acuerdo con los estándares internacionales de bienestar animal.

Se utilizaron ratones macho de la cepa NIH:Swiss, con pesos entre 20 y 25 gramos, provistos por el Laboratorio de Animales de Experimentación (LAE) de la Facultad de Ciencias Veterinarias de la UNLP (FCV-UNLP), con certificación SPF (libre de patógenos específicos). Los animales fueron alojados y mantenidos en el bioterio del LIDeB, bajo condiciones estandarizadas de temperatura, luz y humedad, dentro de cajas de policarbonato de 300 mm x 200 mm x 190 mm, con tapa rejilla de acero inoxidable. Se dispusieron en grupos de 6 ratones por caja, con agua y alimento balanceado *ad libitum*. Durante el periodo de adaptación (una semana), los mismos fueron administrados con un volumen de 0,1 ml de solución fisiológica (SF) de forma intraperitoneal (i.p.). Para el caso de los ensayos de MES y 6 Hz los animales fueron restringidos manualmente con acercamiento de los electrodos, de modo de acostumarlos y minimizar así la influencia de la manipulación en los resultados observados el día del ensayo.

Con respecto a los ensayos realizados en la Universidad de Salamanca, todos los procedimientos y protocolos experimentales se realizaron de acuerdo a los lineamientos de la European Communities Council Directive (2010/63/UE) para el cuidado y uso de animales de laboratorio, y fueron aprobados por el Comité de Bioética de

la Universidad de Salamanca.

Se utilizaron hámsters de la cepa GASH:Sal con 4 meses de edad, criados y mantenidos en el bioterio de la Universidad de Salamanca. Los animales se mantuvieron en ciclos de 14/10 h de luz/oscuridad, a una temperatura ambiente controlada (22-24°C) y con agua y alimento balanceado *ad libitum*. Para evitar la influencia del ritmo circadiano, todos los experimentos se realizaron en la misma hora del día.

Para la realización de los perfiles, las drogas se administraron a dosis de 30 y 100 mg/kg y se evaluaron a las 0,5 y 4 h.

Soluciones

Todas las soluciones de los compuestos a evaluar se prepararon en el momento previo al ensayo y fueron administradas por vía i.p. Para los ensayos realizados en el LIDeB, NVB se preparó en SF (volumen máximo 10 ml/kg), CNZ en DMSO (volumen máximo 2 ml/kg) y MTK se preparó en carboximetilcelulosa (CMC) al 0.5%. Con respecto a los ensayos llevados a cabo en la Universidad de Salamanca, MTK y NVB se disolvieron en SF, mientras que CNZ se preparó en una mezcla Etanol:Cremophor:SF (2:1:17).

6.2.1. Ensayo MES

Desarrollado por Toman y colaboradores hace más de 70 años, este fue el primer modelo animal validado para crisis convulsivas, permitiendo la identificación de la Fenitoína por Merrit y Putnam en 1937 (**Putnam y Merritt 1937**). Comúnmente se asocian los compuestos activos en este ensayo con aquellos capaces de modular a los canales NaV. Sin embargo, muchos de estos compuestos interaccionan también con otros blancos moleculares implicados en la fisiopatología de la Epilepsia, como por ejemplo los receptores AMPA y NMDA, o los canales de potasio voltaje-operados (**Bialer y White 2010**).

Utilizando un generador de pulsos, se aplica un estímulo eléctrico corto (0.2 s) de intensidad 50 mA (ratones) o 150 mA (ratas), con una frecuencia de 60 Hz y un ancho de pulso de 0.6 ms, mediante electrodos corneales o auriculares. Estas condi-

ciones corresponden a un estímulo eléctrico 5-10 veces superior al estímulo necesario para alcanzar el umbral convulsivo, de manera de evitar variaciones en la inducción de convulsiones tónico-clónicas debido a las fluctuaciones intra e interindividuales en el umbral convulsivo. Por lo tanto, el 100 % de los animales sin tratamiento anticonvulsivante exhibirán convulsiones tras la estimulación, observadas como extensión tónica de las extremidades traseras. La prueba se considera positiva (esto es, que el candidato administrado no protege la crisis) si dentro de los 10 s siguientes a la estimulación, el animal presenta una extensión de los miembros posteriores que muestra un ángulo de más de 90° y se mantiene durante más de 3 segundos (**Castel-Branco et al. 2009**). Este modelo ofrece un punto final claro, y existe una alta correlación entre la capacidad de un fármaco de proteger a los roedores en el MES test y su efectividad respecto de las crisis tónico-clónicas generalizadas en humanos (**Castel-Branco et al. 2009**; **Kandratavicius et al. 2014**).

El estímulo eléctrico se aplicó mediante electrodos auriculares, utilizando un generador de pulsos UGO Basile ECT- 57800 (Ugo Basile SRL, Italia). Como control positivo se utilizó Fenitoína a una dosis de 15 mg/kg, mientras que como control negativo se inyectó el vehículo utilizado para disolver la droga. Los resultados se reportaron como número de animales protegidos vs. número de animales evaluados, a cada dosis y a cada tiempo.

6.2.2. Ensayo de 6 Hz

El modelo de 6 Hz se describe como un modelo agudo en el que se aplica, a través de electrodos corneales, una estimulación a baja frecuencia (6-Hz/44-mA), larga duración (3 s) y un ancho de pulso de 1 ms, con el fin de inducir en los animales crisis parciales, similares a las crisis psicomotoras observadas en pacientes con Epilepsia parcial, como la Epilepsia del lóbulo temporal o Epilepsia límbica. Las crisis se caracterizan mediante la presencia de diferentes signos en el animal, como por ejemplo inmovilidad, clonías de miembros anteriores y cola erecta, más conocido como *straub tail*. Este estado, por lo general, tiene una duración mayor a 10 s, y los animales que no presenten ninguna de estas manifestaciones, o que lo hagan por un tiempo menor a 10 s, son considerados protegidos por el candidato administrado (**Barton et al. 2001**).

Como control positivo se utilizó Levetiracetam a una dosis de 30 mg/kg, que equivale aproximadamente a dos veces la dosis efectiva 50 (14,9 mg/kg). El control negativo se realizó inyectando el vehículo utilizado para disolver la droga en cada caso. Los resultados se reportaron como número de animales protegidos vs. número de animales evaluados, a cada dosis y tiempo.

6.2.3. Ensayo de PTZ_{sc}

El Pentilentetrazol (PTZ) o Metrazol es un antagonista no competitivo de los receptores GABA-A, que actúa en el sitio de unión de la Picrotoxina (**Jung, Lal y Gatch 2002**). Existen varios modelos animales de crisis convulsivas que emplean esta sustancia, utilizando distintas vías de administración y monitoreando diversos parámetros como criterio de punto final (**Velíšková, Shakarjian y Velíšek 2017**).

En el ensayo PTZ_{sc} realizado en nuestro laboratorio, a cada animal se le administró por vía s.c una solución de PTZ de 80 mg/kg (a razón de 5 ml/kg), suficiente para inducir crisis clónicas con una duración de al menos 5 s en el 97% de los animales evaluados (**Löscher 2011**). Los animales se observaron durante los 30 minutos posteriores a la inyección, y como criterio de protección se utilizó la ausencia de crisis clónica, parcial o generalizada (**Mandhane, Aavula y Rajamannar 2007; Yuen y Trocóniz 2015**).

Como control negativo se administró Fenobarbital a una dosis de 25 mg/kg y se evaluó a los 30 minutos post inyección, tiempo que coincide con su tiempo de efecto máximo, mientras que como control positivo se administró solamente el vehículo. Los resultados se reportaron como número de animales protegidos vs. número de animales evaluados, a cada dosis y tiempo.

6.2.4. Ensayo audiológico GASH:Sal

El modelo GASH:Sal es desarrollado y mantenido por el Laboratorio de Trastornos Audiomotores de la Universidad de Salamanca - España. A diferencia de los modelos descritos previamente, en este caso las convulsiones no son inducidas mediante estímulos proconvulsivantes como impulsos eléctricos o sustancias químicas,

sino que ocurren espontáneamente gracias a un desencadenante sonoro (**Carballosa González 2008**). Este modelo se asocia principalmente a las convulsiones de tipo generalizadas (**Carballosa González 2008**).

Se trata de hámsters de la especie *Mesocricetus auratus*, también conocidos como hámsters sirios o dorados, con una predisposición genética a sufrir convulsiones inducidas por un estímulo sonoro. Estas se caracterizan por un período corto de latencia luego de la estimulación auditiva, seguido por una carrera salvaje o *wild running*, una fase convulsiva y finalmente una fase de estupor (**Muñoz et al. 2017**). Las convulsiones se analizan a través de un índice de severidad categorizado (cSI), el cual varía entre 0 y 8 (Tabla 6.1).

Tabla 6.1: Índices de severidad de convulsiones. El índice de severidad según **Garcia-Cairasco et al. 1996** (SI), con sus respectivas descripciones conductuales, es transformado en variables discretas con fines estadísticos: Índice de gravedad categorizado (cSI). Tomado de **Rossetti et al. 2006**. *Categorías que generalmente son seguidas por convulsiones clónicas de las extremidades posteriores (CCV2).

SI	Comportamientos convulsivos	cSI
0.00	Sin convulsiones	0
0.11	Una carrera salvaje	1
0.23	Una carrera salvaje (carrera, saltos y caída atónica)	2
0.38	Dos carreras salvajes	3
0.61	Convulsión tónica (Opistótonos)	4
0.85	Convulsiones tónicas más convulsiones clónicas generalizadas	5
0.90	Flexión ventral de la cabeza más cSI5	6
0.95	Extensión de extremidades anteriores más cSI6*	7
1.00	Extensión de miembros posteriores más cSI6*	8

Todos los animales fueron sometidos a un estímulo sonoro de pre-tratamiento (frecuencia 0–18 kHz e intensidad 115-120 dB). Los hámsters GASH:Sal presentan típicamente una severidad $cSI \geq 5$ (**Rossetti et al. 2006**), por lo tanto aquellos que no alcancen este valor de cSI inicial fueron excluidos del análisis. Luego de adminis-

trar la droga, los mismos fueron estimulados nuevamente a los 30 min y 4 h, para poder observar posibles efectos en la severidad de las convulsiones. Conjuntamente, se tomaron muestras de sangre a los tiempos 0 min, 15 min, 60 min, 2 h, 4 h y 24 h post administración, con el fin de realizar futuras determinaciones farmacocinéticas (Figura 6.1). Un animal se considera protegido si la droga logra disminuir el cSI por debajo de 5. Se utilizaron 3 animales para cada droga y dosis, y los resultados se expresaron como número de animales protegidos vs. número de animales evaluados, a cada dosis y tiempo.

Al finalizar el ensayo los animales fueron eutanasiados bajo anestesia profunda (mediante inhalación de CO_2) y sus cerebros fueron congelados a $-70^{\circ}C$ para ser utilizados en futuros experimentos a realizarse en el Laboratorio de Trastornos Audiomotores de la Universidad de Salamanca.

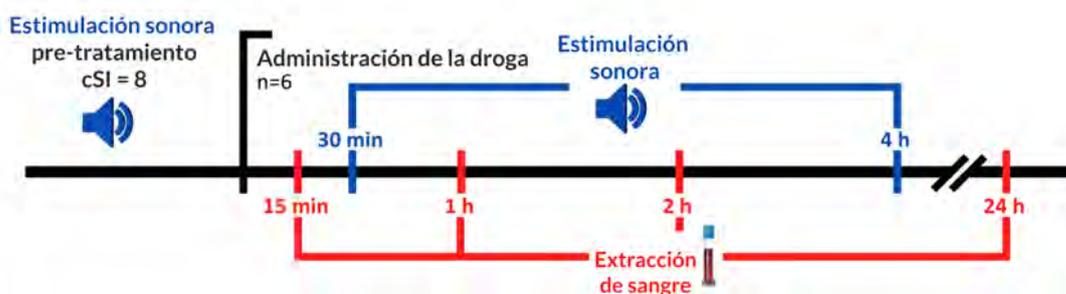


Figura 6.1: Esquema del diseño experimental para el ensayo audiogénico en hámsters GASH:Sal.

6.2.5. Ensayo Rotorod

El ensayo de Rotorod permite evaluar si el compuesto en cuestión afecta la función motora del animal, lo cual es un indicio de neurotoxicidad y se correlaciona con signos clínicos como sedación, somnolencia, ataxia, entre otros (Dunham y Miya 1957).

La prueba consta de un cilindro estriado rotatorio con compartimentos separados, donde se ubica el animal mientras éste rota a una velocidad de 6 revoluciones por minuto (rpm) (Figura 6.2). El animal debe permanecer sobre el mismo sin caer durante 1 minuto para superar la prueba. Si este no puede mantenerse girando sobre el rodillo, al cabo de 3 caídas se considera que el compuesto ensayado presenta

rasgos de neurotoxicidad (Deacon 2013).

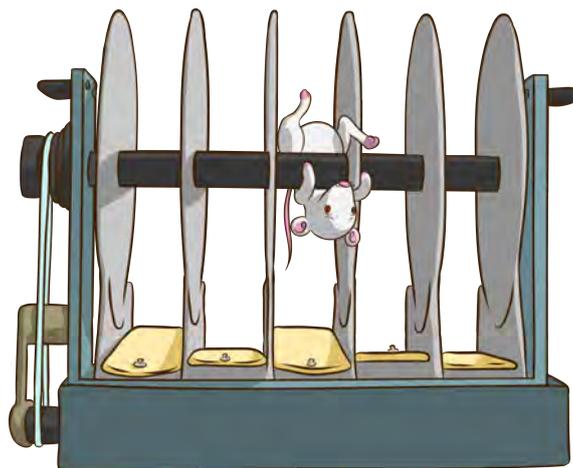


Figura 6.2: Ilustración de la prueba de Rotorod. Tomado de <https://imgur.com/gallery/44vStxD>

Luego de administrar los compuestos a evaluar y antes de la inducción de la convulsión, los animales se colocaron sobre el cilindro del equipo y se observaron durante 1 minuto, registrando las caídas desde el rodillo si estas existieran. Los resultados se reportaron como número de animales que no superaron la prueba vs. número de animales evaluados, a cada dosis y tiempo.

6.3. Resultados

A continuación se describen los resultados obtenidos en los ensayos *in vivo* (Tabla 6.2). En algunos casos se encontraron resultados previos en literatura, los que fueron adicionados al análisis sin repetir el ensayo, priorizando utilizar animales de experimentación solo para complementar el conocimiento científico existente, aportando información novedosa que permita completar la caracterización del perfil anticonvulsivo a nivel preclínico.

Tabla 6.2: Actividad anticonvulsiva y neurotóxica *in vivo*. * Número de animales protegidos sobre número total de animales ensayados. ** Número de animales que presentan signos de neurotoxicidad sobre número total de animales ensayados. NT: no testeado. a: **Lenz** et al. 2014, b: **Cevik** et al. 2015, c: **Fleck** et al. 2015, d: **Fleck** et al. 2016, e: **Abdel-kader** et al. 2017, f: **Desmedt, Niemegeers y Janssen** 1975, g: **Brahmane** et al. 2010, h: **Abdelsameea** 2015.

Droga	Dosis (mg/kg)	Tiempo (h)	MES test*	PTZ test*	6 Hz test*	GASH:Sal*	Rotorod test**
MTK	30	0.5	0/2		0/2	3/3	0/2
		4	0/2	Activo	0/2	3/3	0/2
	100	0.5	0/2	a,b,c,d,e	0/2	3/3	0/2
		4	1/4		0/2	2/3	0/4
NVB	30	0.5	0/2	0/2	0/2	3/3	0/6
		4	3/4	0/2	0/2	2/3	0/8
	100	0.5	0/2	0/2	0/2	3/3	0/6
		4	3/4	0/2	0/2	1/3	0/8
CNZ	30	0.5	Activo	Activo	0/2	1/2	0/2
		4	f,g,h	g	0/2	1/2	0/2
	100	0.5	NT	NT	NT	1/2	NT
		4				2/2	

Montelukast

La actividad anticonvulsiva de MTK en ensayos de PTZ fue previamente reportada en ratones (**Abdel-kader** et al. 2017; **Fleck** et al. 2016; **Lenz** et al. 2014; **Rehni y Singh** 2011) y ratas (**Cevik** et al. 2015). A su vez, la combinación de MTK con otras ACDs como Valproato o Fenobarbital ha demostrado potenciar la acción de las mismas en este tipo de ensayo (**Abdel-kader** et al. 2017; **Fleck** et al. 2015). En el ensayo MES, MTK protegió 1/4 ratones a la dosis de 100 mg/kg y un tiempo de 4 h, mientras que en el ensayo de 6 Hz no se observó actividad. Con respecto al ensayo audiogénico en hámsters, los resultados de MTK fueron prometedores, protegiendo al menos 2/3 animales a todas las dosis y tiempos (Tabla 6.2). En particular, para ambas dosis y un tiempo de 0.5 h, MTK logró abolir completamente las crisis en todos los animales tratados (cSI 0 ± 0).

Novobiocina

En el ensayo MES, NVB demostró una excelente actividad protegiendo 3/4

ratones para ambas dosis (30 y 100 mg/kg) y un tiempo de 4 h. En los ensayos de PTZ y 6 Hz, NVB no mostró actividad a las dosis ensayadas. Por su parte, en el ensayo GASH:Sal NVB demostró actividad a todas las dosis y tiempos ensayados, alcanzando un máximo a las 0.5 h, donde protegió el 100% de los hámsters a las dosis de 30 y 100 mg/kg (Tabla 6.2).

Cinarizina

La actividad de CNZ, junto con otros miembros de la familia de bloqueantes de canales de calcio, ha sido previamente reportada en el ensayo MES (**Abdelsameea 2015; Brahmane et al. 2010; Desmedt, Niemegeers y Janssen 1975**) y también en un ensayo de PTZ agudo (**Brahmane et al. 2010**). En el ensayo de 6 Hz realizado en nuestro laboratorio, CNZ no demostró actividad alguna, aunque solo pudo ser evaluada a la dosis mas baja por motivos de solubilidad. Con respecto al ensayo audiogénico en hámsters GASH:Sal, CNZ demostró actividad a todas las dosis y tiempos ensayados (Tabla 6.2). En particular, a la dosis de 100 mg/kg y un tiempo de 4 h, logró abolir completamente las convulsiones inducidas por sonido (cSI 0 ± 0). Cabe destacar que en esta determinación 2 animales fueron excluidos del análisis, ya que durante el estímulo sonoro de pre-tratamiento no alcanzaron un cSI ≥ 5 . La actividad anticonvulsiva de CNZ en modelos similares de crisis inducidas por sonido fue previamente reportada en ratones (G. B. **De Sarro, Nistico y Meldrum 1986**) y ratas (G. **De Sarro et al. 1990**).

Ninguno de los candidatos evaluados presentó signos de neurotoxicidad evidenciados mediante el test de Rotorod, a ninguna de las dosis evaluadas (Tabla 6.2).

Referencias Capítulo 6

- Abdel-kader**, Z. et al. (2017). “Protective Effect Of Montelukast Against Pentylene-tetrazole-Induced Acute Seizures And Kindling In Mice”. *Zagazig University Medical Journal* 21.6, págs. 1-10. DOI: [10.21608/zumj.2017.4577](https://doi.org/10.21608/zumj.2017.4577).
- Abdelsameea**, A. A. (2015). “Effect of Verapamil, Cinnarizine and Memantine on Maximal Electroshock, Picrotoxin, and Pilocarpine-Induced Seizure Models in Albino Mice”. *Egyptian Journal of Basic and Clinical Pharmacology* 5. DOI: [10.11131/2015/101356](https://doi.org/10.11131/2015/101356).
- Barton**, M. E. et al. (2001). “Pharmacological characterization of the 6 Hz psychomotor seizure model of partial epilepsy”. *Epilepsy Research* 47.3, págs. 217-227. DOI: [10.1016/S0920-1211\(01\)00302-3](https://doi.org/10.1016/S0920-1211(01)00302-3).
- Bialer**, M. y H. S. **White** (2010). “Key factors in the discovery and development of new antiepileptic drugs”. *Nature Reviews. Drug Discovery* 9.1, págs. 68-82. DOI: [10.1038/nrd2997](https://doi.org/10.1038/nrd2997).
- Brahmane**, R. I. et al. (2010). “Role of cinnarizine and nifedipine on anticonvulsant effect of sodium valproate and carbamazepine in maximal electroshock and pentylene-tetrazole model of seizures in mice”. *Journal of Pharmacology & Pharmacotherapeutics* 1.2, págs. 78-81. DOI: [10.4103/0976-500X.72348](https://doi.org/10.4103/0976-500X.72348).
- Carballosa González**, M. (2008). “Hacia un nuevo modelo de epilepsia: el hámster GASH:sal”. DOI: [10.14201/gredos.55970](https://doi.org/10.14201/gredos.55970).
- Castel-Branco**, M. M. et al. (2009). “The maximal electroshock seizure (MES) model in the preclinical assessment of potential new antiepileptic drugs”. *Methods and Findings in Experimental and Clinical Pharmacology* 31.2, págs. 101-106. DOI: [10.1358/mf.2009.31.2.1338414](https://doi.org/10.1358/mf.2009.31.2.1338414).

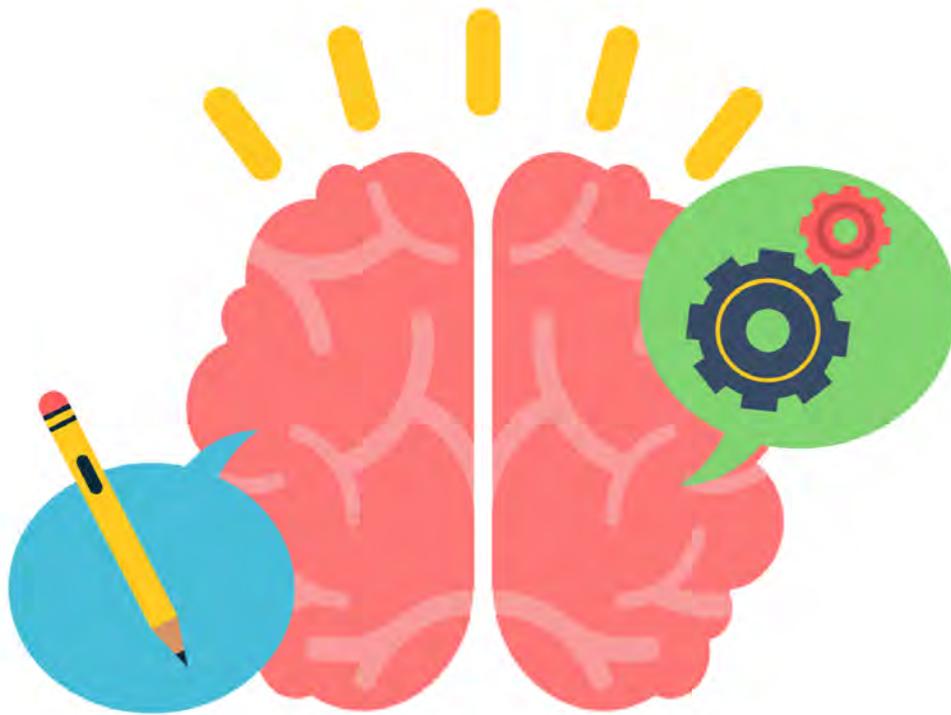
- Cevik, B. et al.** (2015). "Montelukast Inhibits Pentylentetrazol-Induced Seizures in Rats". *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research* 21, págs. 869-874. DOI: [10.12659/MSM.892932](https://doi.org/10.12659/MSM.892932).
- De Sarro, G. B., G. Nistico y B. S. Meldrum** (1986). "Anticonvulsant properties of flunarizine on reflex and generalized models of epilepsy". *Neuropharmacology* 25.7, págs. 695-701. DOI: [10.1016/0028-3908\(86\)90084-5](https://doi.org/10.1016/0028-3908(86)90084-5).
- De Sarro, G. et al.** (1990). "Anticonvulsant properties of some calcium antagonists on sound-induced seizures in genetically epilepsy prone rats". *General Pharmacology: The Vascular System* 21.5, págs. 769-778. DOI: [10.1016/0306-3623\(90\)91032-M](https://doi.org/10.1016/0306-3623(90)91032-M).
- Deacon, R. M. J.** (2013). "Measuring motor coordination in mice". *Journal of Visualized Experiments: JoVE* 75, e2609. DOI: [10.3791/2609](https://doi.org/10.3791/2609).
- Desmedt, L. K., C. J. Niemegeers y P. A. Janssen** (1975). "Anticonvulsive properties of cinnarizine and flunarizine in rats and mice". *Arzneimittel-Forschung* 25.9, págs. 1408-1413.
- Dunham, N. W. y T. S. Miya** (1957). "A Note on a Simple Apparatus for Detecting Neurological Deficit in Rats and Mice**College of Pharmacy, University of Nebraska, Lincoln 8." *Journal of the American Pharmaceutical Association (Scientific ed.)* 46.3, págs. 208-209. DOI: [10.1002/jps.3030460322](https://doi.org/10.1002/jps.3030460322).
- Fleck, J. et al.** (2016). "Montelukast reduces seizures in pentylentetrazol-kindled mice". *Brazilian Journal of Medical and Biological Research* 49.4. DOI: [10.1590/1414-431X20155031](https://doi.org/10.1590/1414-431X20155031).
- Fleck, J. et al.** (2015). "Montelukast potentiates the anticonvulsant effect of phenobarbital in mice: An isobolographic analysis". *Pharmacological Research* 94, págs. 34-41. DOI: [10.1016/j.phrs.2015.02.001](https://doi.org/10.1016/j.phrs.2015.02.001).
- Garcia-Cairasco, N. et al.** (1996). "Audiogenic and audiogenic-like seizures: Locus of induction and seizure severity determine postictal prolactin patterns". *Pharmacology Biochemistry and Behavior* 53.3, págs. 503-510. DOI: [https://doi.org/10.1016/0091-3057\(95\)02040-3](https://doi.org/10.1016/0091-3057(95)02040-3).
- Jung, M. E., H. Lal y M. B. Gatch** (2002). "The discriminative stimulus effects of pentylentetrazol as a model of anxiety: recent developments". *Neuroscience & Biobehavioral Reviews* 26.4, págs. 429-439. DOI: [10.1016/S0149-7634\(02\)00010-6](https://doi.org/10.1016/S0149-7634(02)00010-6).

- Kandratavicius, L. et al.** (2014). “Animal models of epilepsy: use and limitations”. *Neuropsychiatric Disease and Treatment* 10, págs. 1693-1705. DOI: [10.2147/NDT.S50371](https://doi.org/10.2147/NDT.S50371).
- Kehne, J. H. et al.** (2017). “The National Institute of Neurological Disorders and Stroke (NINDS) Epilepsy Therapy Screening Program (ETSP)”. *Neurochemical Research* 42.7, págs. 1894-1903. DOI: [10.1007/s11064-017-2275-z](https://doi.org/10.1007/s11064-017-2275-z).
- Lenz, Q. F. et al.** (2014). “Cysteinyl leukotriene receptor (CysLT) antagonists decrease pentylenetetrazol-induced seizures and blood-brain barrier dysfunction”. *Neuroscience* 277, págs. 859-871. DOI: [10.1016/j.neuroscience.2014.07.058](https://doi.org/10.1016/j.neuroscience.2014.07.058).
- Löscher, W.** (2011). “Critical review of current animal models of seizures and epilepsy used in the discovery and development of new antiepileptic drugs”. *Seizure* 20.5, págs. 359-368. DOI: [10.1016/j.seizure.2011.01.003](https://doi.org/10.1016/j.seizure.2011.01.003).
- (2017). “Animal Models of Seizures and Epilepsy: Past, Present, and Future Role for the Discovery of Antiseizure Drugs”. *Neurochemical Research* 42.7, págs. 1873-1888. DOI: [10.1007/s11064-017-2222-z](https://doi.org/10.1007/s11064-017-2222-z).
- Mandhane, S. N., K. Aavula y T. Rajamannar** (2007). “Timed pentylenetetrazol infusion test: A comparative analysis with s.c.PTZ and MES models of anticonvulsant screening in mice”. *Seizure - European Journal of Epilepsy* 16.7, págs. 636-644. DOI: [10.1016/j.seizure.2007.05.005](https://doi.org/10.1016/j.seizure.2007.05.005).
- Muñoz, L. J. et al.** (2017). “The genetic audiogenic seizure hamster from Salamanca: The GASH:Sal”. *Epilepsy & Behavior: E&B* 71.Pt B, págs. 181-192. DOI: [10.1016/j.yebeh.2016.03.002](https://doi.org/10.1016/j.yebeh.2016.03.002).
- Putnam, T. J. y H. H. Merritt** (1937). “Experimental Determination of the Anticonvulsant Properties of Some Phenyl Derivatives”. *Science* 85.2213, págs. 525-526. DOI: [10.1126/science.85.2213.525](https://doi.org/10.1126/science.85.2213.525).
- Rehni, A. K. y T. G. Singh** (2011). “Modulation of leukotriene D4 attenuates the development of seizures in mice”. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 85.2, págs. 97-106. DOI: [10.1016/j.plefa.2011.04.003](https://doi.org/10.1016/j.plefa.2011.04.003).
- Rossetti, F. et al.** (2006). “EEG wavelet analyses of the striatum-substantia nigra pars reticulata-superior colliculus circuitry: audiogenic seizures and anticonvulsant drug administration in Wistar audiogenic rats (War strain)”. *Epilepsy Research* 72.2-3, págs. 192-208. DOI: [10.1016/j.eplepsyres.2006.08.001](https://doi.org/10.1016/j.eplepsyres.2006.08.001).

- Velíšková, J., M. P. Shakarjian y L. Velíšek** (2017). “Chapter 34 - Systemic Chemoconvulsants Producing Acute Seizures in Adult Rodents”. *Models of Seizures and Epilepsy (Second Edition)*. Ed. por A. **Pitkänen** et al. Academic Press, págs. 491-512. DOI: [10.1016/B978-0-12-804066-9.00035-3](https://doi.org/10.1016/B978-0-12-804066-9.00035-3).
- Yuen, E. S. M. e I. F. Trocóniz** (2015). “Can pentylenetetrazole and maximal electroshock rodent seizure models quantitatively predict antiepileptic efficacy in humans?” *Seizure* 24, págs. 21-27. DOI: [10.1016/j.seizure.2014.11.006](https://doi.org/10.1016/j.seizure.2014.11.006).

Capítulo 7

Conclusiones



7.1. Conclusiones generales

Durante el presente trabajo de tesis doctoral se llevaron a cabo dos campañas de VS con el objetivo de identificar compuestos *multitarget*, con potencial actividad anticonvulsiva, que actúen simultáneamente sobre los canales TRPV1 y NaV1.2. Ambas campañas se realizaron en paralelo, combinando modelos predictivos basados en la estructura de los ligandos (QSAR) y del receptor (docking), desarrollados para cada uno de los blancos moleculares en cuestión.

El *pipeline* de modelado QSAR, llevado a cabo durante esta tesis, fue aplicado en el desarrollo y validación de modelos clasificadores binarios (TRPV1) y de regresión (NaV1.2). Como resultado se obtuvieron modelos robustos y con una alta capacidad predictiva, lo que constituye una validación de la metodología propuesta al aplicarla con resultados satisfactorios en dos escenarios diferentes (clasificación binaria y regresión), pero que apuntan a un objetivo en común, la identificación de compuestos activos.

Con respecto a los modelos basados en la estructura del receptor, se evaluaron diferentes modelos estructurales de la proteína sobre la cual se realiza la simulación, distintos programas de docking y también distintas estrategias de post-procesamiento de los complejos ligando-receptor obtenidos. Para cada una de estas combinaciones se validaron distintos parámetros como el *docking power*, *scoring power* o *screening power*, con el objetivo de evaluarlos comparativamente, lo que finalmente nos permitió desarrollar modelos con una buena capacidad predictiva. Existen una gran cantidad de parámetros que pueden explorarse a la hora de optimizar la *performance* de un modelo de docking para una tarea en particular (por ejemplo VS). Sin embargo, para ello es necesario contar con protocolos de validación bien diseñados, que permitan comparar los distintos modelos a través de diferentes aspectos.

Para ambos blancos moleculares, la estrategia de post-procesamiento consistió en re-scorear los complejos ligando-receptor generados, utilizando diferentes funciones de puntuación. Este *re-scoring* es precedido por una minimización de los complejos, considerando como flexibles las cadenas laterales del sitio activo que se encuentran a 3 Å o menos del ligando en cuestión. En todos los casos se encontró

que este post-procesamiento mejora el *scoring power* de los modelos, algo que ya hemos observado en investigaciones previas (Castro et al. 2020; Llanos et al. 2021, 2022a,b). Para el canal TRPV1, el *re-scoring* con la función de puntuación ad4 resultó ser el más apropiado, mientras que para el canal NaV1.2 fue el *re-scoring* con vinardo.

Con el objetivo de reposicionar fármacos ya aprobados como ACDs, se eligió la base de datos DrugBank5.1 para realizar los VS. El reposicionamiento de fármacos presenta varias ventajas con respecto a las técnicas convencionales de descubrimiento de fármacos, descritas en profundidad en el Capítulo 2, Sección 2.4. La selección de los *hits* para ser evaluados experimentalmente incluyó varios criterios como la novedad del *scaffold*, el precio y la accesibilidad, datos de farmacocinética y toxicidad disponibles, y las predicciones conjuntas sobre los canales TRPV1 y NaV1.2. Finalmente, 3 fármacos fueron seleccionados para su evaluación experimental: CNZ, MTK y NBV.

Todos los *hits* evaluados demostraron actividad *in vitro* sobre ambos blancos moleculares. En este sentido, la combinación de aproximaciones ortogonales, como lo son aquellas basadas en la estructura de los ligandos y del receptor, demostró ser una estrategia prometedora a la hora de identificar nuevos compuestos *multitarget*. Sin embargo, cabe destacar que para poder validar experimentalmente estas metodologías o incluso calcular un *hit rate*, sería necesario evaluar experimentalmente un número mayor de compuestos. A su vez, las predicciones aisladas de los modelos no fueron el único criterio que se consideró para la selección de los candidatos, como se discutió previamente.

Un subproducto de este trabajo de tesis doctoral fue el desarrollo de una metodología propia de agrupamiento, denominada Silhouette Optimized Molecular Clustering (SOMoC), la cual fue aplicada para agrupar en *clusters* los *datasets* utilizados en el desarrollo de los modelos. La validación de esta metodología consistió en un *benchmarking*, donde se la evaluó junto a otros 3 métodos de *clustering* (iRaPCA, Ward y Butina) contra una serie de 29 *datasets* diversos. SOMoC e iRaPCA (una metodología alternativa desarrollada también en el LIDeB), demostraron ser consistentemente superiores a los demás métodos, generando en todos los casos *clusters* compactos y bien separados, a juzgar por las diferentes métricas de validación in-

terna calculadas (**Prada Gori et al. 2022**). Esta metodología fue implementada en forma de [Web App](#), a través de la plataforma Streamlit, lo que permite a cualquier usuario utilizarla de manera libre, gratuita y sencilla. Alternativamente, es posible descargar una versión *standalone* para correr de manera local si el usuario así lo requiere.

7.2. Búsqueda de nuevos inhibidores del canal TRPV1

Con respecto al canal TRPV1, se logró compilar una base de datos de compuestos evaluados frente al blanco molecular, con y sin actividad inhibitoria. La misma se utilizó para desarrollar y validar un modelo QSAR del tipo clasificador (que permite predecir compuestos que antagonizan la activación del canal por CAP), así como también un modelo de docking capaz de identificar compuestos que interaccionan con el sitio de unión de CAP.

Todos los compuestos evaluados demostraron actividad inhibitoria nanomolar sobre las corrientes activadas por CAP en canales TRPV1, evaluadas mediante la técnica de patch clamp en un sistema de expresión heterólogo en células HEK293. Los porcentajes de inhibición (calculados a -100 mV y una concentración de 100 nM) oscilaron alrededor del 50% en todos los casos: MTK = $53.4 \pm 7.1\%$ (n=7), NBV = $56.3 \pm 6.7\%$ (n=9), CNZ = $50.4 \pm 9.2\%$ (n=7). Para el caso de MTK y NVB, se observó una marcada voltaje-dependencia en la inhibición, siendo esta más pronunciada a potenciales más hiperpolarizados.

En el contexto de una patología como la Epilepsia, la inhibición a potenciales más negativos, cercanos al potencial de membrana en reposo, es una característica deseable para una ACD que actúe sobre los canales TRPV1. Varios estudios sugieren que, el aumento del Ca^{2+} intracelular mediado por la activación de estos canales, genera una despolarización del potencial de membrana en reposo (**Naziroglu 2015**; **Storozhuk, Moroz y Zholos 2019**), lo que conlleva a un aumento del *firing* neuronal y la liberación de neurotransmisores, entre ellos se ha descrito un aumento en la descarga de las sinapsis glutamatérgicas excitatorias (**Fawley, Hofmann y Andresen 2014**). Por otro lado, la contribución de las corrientes mediadas por el TRPV1 a la forma y duración de los potenciales de acción es despreciable. Por lo tanto, la

inhibición de los canales TRPV1 presentaría un mayor impacto cuando esta ocurre a potenciales de membrana más negativos, como se observó para MTK y NBV.

Mediante simulaciones de MD, se estudió la interacción de los candidatos seleccionados con el sitio de los capsaicinoides en el modelo del canal hTRPV1. La estrategia planteada de incluir 4 copias de cada ligando, ocupando simultáneamente los 4 sitios de unión de CAP, nos permitió incrementar considerablemente el muestreo o *sampling* de la simulación, sin aumentar el costo computacional. Para los 3 compuestos, se observaron interacciones favorables con el blanco molecular, las cuales se mantuvieron estables durante toda la simulación.

Desde la perspectiva de la química medicinal, los resultados de esta tesis doctoral suponen un valioso aporte al conocimiento científico actual sobre los antagonistas TRPV1 y la hipertermia en humanos. Se lograron identificar 3 nuevos *scaffolds* de antagonistas con potencia nanomolar, los cuales son fármacos que se utilizan hoy día en la clínica, y ninguno ha demostrado efectos adversos relacionados con la modulación de la temperatura corporal. Esto nos permite validar la hipótesis de trabajo inicial, la cual plantea que “a través de una estrategia de reposicionamiento de fármacos asistida por computadora, es posible identificar nuevos antagonistas TRPV1 sin efectos adversos relacionados con la modulación de la temperatura corporal”.

A partir de la correlación que existe entre el efecto de los antagonistas sobre los distintos modos de activación y el resultado que estos producen sobre la temperatura corporal, es posible diseñar antagonistas modo-selectivos que sean térmicamente neutros (**Garami et al. 2018**; **Garami et al. 2020**). Sin embargo, esto requiere necesariamente conocer el perfil del antagonista frente a los distintos modos de activación, para poder identificar ciertas características electrónico-estructurales en estas moléculas que puedan relacionarse con el modo de activación que afectan. En este sentido, la perspectiva a futuro es caracterizar cada uno de los antagonistas identificados —idealmente utilizando el clon humano hTRPV1— con respecto a los distintos modos de activación, para corroborar si se ajustan o no al modelo propuesto por Garami y colaboradores (**Garami et al. 2020**).

7.3. Búsqueda de nuevos inhibidores del canal NaV1.2

Para el caso del canal NaV1.2, se compiló inicialmente una base de datos de moléculas evaluadas en el ensayo de desplazamiento de H-BTX. Este *dataset* nos permitió desarrollar un modelo QSAR de regresión, capaz de predecir el valor de pIC₅₀ para un compuesto en este ensayo, lo cual representa una medida de su afinidad por el sitio de unión de moléculas pequeñas, en la región del poro del canal. Debido a que los ensayos de desplazamiento de H-BTX se realizaron en sinaptosomas, y que la secuencia aminoacídica en la región del poro es prácticamente invariable entre las isoformas NaV (Oelstrom et al. 2014), este modelo permite identificar moléculas activas sobre todas las isoformas cerebrales de estos canales, no solo sobre la isoforma 1.2.

En este caso, la estructura tridimensional del blanco molecular fue determinada experimentalmente mediante Cryo-EM. La estructura nativa fue refinada mediante el software Rosetta, generando así un *ensemble* conformacional alrededor de la misma. Luego, el *dataset* compilado previamente fue dockeado en todos los modelos tridimensionales (102), utilizando diferentes condiciones de docking. Esta estrategia nos permitió mejorar considerablemente el *scoring power* del modelo, en comparación con un protocolo de docking clásico utilizando la estructura experimental, tal y como se obtiene del Protein Data Bank.

Con respecto a los ensayos *in vitro*, se evaluó la capacidad de los *hits* seleccionados de inhibir las corrientes de sodio, registradas mediante la técnica de patch clamp, en canales hNaV1.2 expresados en células HEK293. En particular, MTK (1 μ M: 73.1 \pm 11.6 %, n=6; 10 μ M: 95.3 \pm 4.2 %, n=6) y CNZ (1 μ M: 71.6 \pm 3.0 %, n=7; 10 μ M: 98.1 \pm 1.5 %, n=6) demostraron una alta potencia inhibitoria, mientras que el bloqueo tónico de NVB no fue estadísticamente significativo a las concentraciones ensayadas (p<0.05) (1 μ M: 9.7 \pm 8.3 %, n=5; 10 μ M: 16.1 \pm 7.8 %, n=5).

Como se mencionó en el Capítulo 4, nuestro grupo de investigación desarrolló previamente un modelo por homología del canal hNaV1.2, el cual se utilizó en una campaña de VS para identificar nuevos inhibidores. A una concentración de 300 μ M se obtuvieron valores de inhibición cercanos al 20 % para Ciprofloxacina, Losartan y Valsartan. El compuesto más potente fue la N,N-Difenetilsulfamida, logrando una

inhibición del $36.9 \pm 8.2\%$ (n=3) a una concentración de $100 \mu\text{M}$, un valor similar al obtenido para Fenitoína $37.4 \pm 3.4\%$ (n=6), una ACDs clásica cuyo mecanismo de acción se asocia a la inhibición de los canales NaV (**Paestro et al. 2018**). Aun cuando todos los compuestos ensayados demostraron actividad inhibitoria *in vitro*, su potencia fue considerablemente menor en comparación a los reportados en esta tesis. Esto puede atribuirse, en parte, a la combinación de modelos basados en la estructura de los ligandos y del receptor para la identificación de nuevos *hits*. A su vez, este modelo de docking se construyó en base a una estructura experimental del canal, y no un modelo por homología.

Con respecto a la inhibición dependiente de estado, los 3 compuestos demostraron un potente bloqueo fásico, debido a la estabilización del estado inactivado. A una concentración de $10 \mu\text{M}$, se obtuvieron valores de $\Delta V_{1/2}$ estadísticamente significativos ($p < 0.05$) para MTK: $-25.66 \pm 3.21 \text{ mV}$ (n=6), CNZ: $-30.75 \pm 3.00 \text{ mV}$ (n=5) y NVB: -3.32 ± 0.86 (n=5). A modo de referencia, podemos mencionar los $\Delta V_{1/2}$ que se obtienen, a la misma concentración, para ACDs clásicas como Fenitoína: -5 mV , Carbamazepina: -1 mV o Lamotrigina: -2 mV (**Qiao et al. 2014**)

El bloqueo dependiente de estado o *state-dependent*, donde los bloqueantes estabilizan el estado inactivado de los canales NaV, es una característica común a la mayoría de las ACDs que actúan sobre estos canales. Los fármacos que presentan este tipo de bloqueo actúan preferencialmente sobre las neuronas más despolarizadas, y por lo tanto hiperexcitables, que participan en el foco epiléptico. A su vez, por retrasar la recuperación del estado inactivado, estos compuestos disminuyen la fracción disponible de canales para pasar del estado cerrado o *resting* al abierto. Esta disminución es acumulativa y aumenta con la frecuencia de estimulación (*use-dependent*), por lo que aquellas neuronas involucradas en la propagación de la actividad epileptiforme, las cuales presentan mayores frecuencias de disparo, serán más susceptibles al bloqueo por parte de estos compuestos (**Nardi et al. 2012**; **Qiao et al. 2014**).

7.4. Actividad anticonvulsiva *in vivo*

La actividad anticonvulsiva de los compuestos fue evaluada en cuatro modelos animales de crisis agudas. Todos los *hits* presentaron actividad en al menos uno de estos ensayos, y ninguno presentó signos de neurotoxicidad evidenciada mediante el test de Rotorod (Tabla 6.2).

En particular, para NVB se observó una buena actividad en el ensayo MES, donde protegió a 3/4 ratones a la dosis más baja (30 mg/kg) y un tiempo de 4 h. Con respecto a los ensayos en el modelo audiogénico GASH:Sal, realizados en la Universidad de Salamanca, las 3 drogas evaluadas (MTK, NVB y CNZ) mostraron resultados prometedores, disminuyendo la severidad de las convulsiones a todas las dosis y tiempos ensayados. El más potente fue MTK, quien logró abolir completamente las convulsiones a dosis de 30 y 100 mg/kg y un tiempo de 0.5 h. Por su parte, CNZ demostró un comportamiento similar a la dosis mas alta y un tiempo de 4 h.

A la luz de los resultados obtenidos en el modelo audiogénico GASH:Sal, se plantea como perspectiva a futuro evaluar Capsaicina (CAP) y Capsazepina (CPZ), un agonista y un antagonista TRPV1 de referencia, respectivamente. A su vez, se medirá el grado de expresión de este canal en varias regiones del cerebro de los animales tratados, para evaluar si las drogas afectan su expresión. De esta manera, se intentará validar al TRPV1 como un blanco molecular implicado en este modelo animal particular. Cabe destacar que existen reportes previos que vinculan al canal TRPV1 con modelos de Epilepsia audiogénica en ratas GEPR-3 (Cho et al. 2018) y WAR (Lazarini-Lopes et al. 2022). Estos experimentos se llevarán a cabo en el Laboratorio de Trastornos Audiomotores de la Universidad de Salamanca, en el marco de una nueva tesis doctoral.

Como se discutió en el Capítulo 1, Sección 1.7, los compuestos *multitarget* representan una aproximación prometedora en el tratamiento de enfermedades complejas, como la Epilepsia. De hecho, varias ACDs de uso clínico actúan simultáneamente sobre más de un blanco molecular. En este sentido, aun cuando puedan existir otros mecanismos de acción asociados a la actividad anticonvulsiva reportada para estas moléculas, nuestros resultados validan la hipótesis de trabajo inicial, la cual pos-

tula que “es posible encontrar compuestos *multitarget* que actúen simultáneamente sobre los canales TRPV1 y NaV1.2, con actividad anticonvulsiva *in vivo*, a través de una estrategia de reposicionamiento de fármacos asistida por computadora”. La modulación conjunta de los canales TRPV1 y NaV1.2 emerge como una estrategia promisoriosa para el desarrollo de nuevas ACDs, más eficaces, seguras y mejor toleradas.

Referencias Capítulo 7

- Castro**, M. A. et al. (2020). “Citrus reticulata peel oil as an antiatherogenic agent: Hypolipogenic effect in hepatic cells, lipid storage decrease in foam cells, and prevention of LDL oxidation”. *Nutrition, Metabolism and Cardiovascular Diseases* 30.9, págs. 1590-1599. DOI: [10.1016/j.numecd.2020.04.033](https://doi.org/10.1016/j.numecd.2020.04.033).
- Cho**, S. J. et al. (2018). “Inhibition of transient potential receptor vanilloid type 1 suppresses seizure susceptibility in the genetically epilepsy-prone rat”. *CNS Neuroscience & Therapeutics* 24.1, págs. 18-28. DOI: [10.1111/cns.12770](https://doi.org/10.1111/cns.12770).
- Fawley**, J. A., M. E. **Hofmann** y M. C. **Andresen** (2014). “Cannabinoid 1 and Transient Receptor Potential Vanilloid 1 Receptors Discretely Modulate Evoked Glutamate Separately from Spontaneous Glutamate Transmission”. *The Journal of Neuroscience* 34.24, págs. 8324-8332. DOI: [10.1523/JNEUROSCI.0315-14.2014](https://doi.org/10.1523/JNEUROSCI.0315-14.2014).
- Garami**, A. et al. (2018). “TRPV1 antagonists that cause hypothermia, instead of hyperthermia, in rodents: Compounds’ pharmacological profiles, in vivo targets, thermoeffectors recruited and implications for drug development”. *Acta Physiologica (Oxford, England)* 223.3. DOI: [10.1111/apha.13038](https://doi.org/10.1111/apha.13038).
- Garami**, A. et al. (2020). “Hyperthermia induced by transient receptor potential vanilloid-1 (TRPV1) antagonists in human clinical trials: Insights from mathematical modeling and meta-analysis”. *Pharmacology & Therapeutics* 208, pág. 107474. DOI: [10.1016/j.pharmthera.2020.107474](https://doi.org/10.1016/j.pharmthera.2020.107474).
- Lazarini-Lopes**, W. et al. (2022). “Increased TRPV1 Channels and FosB Protein Expression Are Associated with Chronic Epileptic Seizures and Anxiogenic-like Behaviors in a Preclinical Model of Temporal Lobe Epilepsy”. *Biomedicines* 10.2, pág. 416. DOI: [10.3390/biomedicines10020416](https://doi.org/10.3390/biomedicines10020416).

- Llanos, M. A. et al.** (2021). “Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study”. *Journal of Chemical Information and Modeling* 61.8, págs. 3758-3770. DOI: [10.1021/acs.jcim.1c00404](https://doi.org/10.1021/acs.jcim.1c00404).
- Llanos, M. A. et al.** (2022a). “Homology Modeling and Molecular Dynamics Simulations of Trypanosoma cruzi Phosphodiesterase b1”. *Chemistry & Biodiversity* 19.1, e202100712. DOI: <https://doi.org/10.1002/cbdv.202100712>.
- Llanos, M. A. et al.** (2022b). “Structure-Based Virtual Screening Identifies Novobiocin, Montelukast, and Cinnarizine as TRPV1 Modulators with Anticonvulsant Activity In Vivo”. *Journal of Chemical Information and Modeling*. DOI: [10.1021/acs.jcim.2c00312](https://doi.org/10.1021/acs.jcim.2c00312).
- Nardi, A. et al.** (2012). “Advances in Targeting Voltage-Gated Sodium Channels with Small Molecules”. *ChemMedChem* 7.10, págs. 1712-1740. DOI: [10.1002/cmdc.201200298](https://doi.org/10.1002/cmdc.201200298).
- Naziroglu, M.** (2015). “TRPV1 Channel: A Potential Drug Target for Treating Epilepsy”. *Current Neuropharmacology* 13.2, págs. 239-247. DOI: [10.2174/1570159X13666150216222543](https://doi.org/10.2174/1570159X13666150216222543).
- Oelstrom, K. et al.** (2014). “Evolutionarily conserved intracellular gate of voltage-dependent sodium channels”. *Nature Communications* 5.1, págs. 3420. DOI: [10.1038/ncomms4420](https://doi.org/10.1038/ncomms4420).
- Palestro, P. H. et al.** (2018). “Searching for New Leads To Treat Epilepsy: Target-Based Virtual Screening for the Discovery of Anticonvulsant Agents”. *Journal of Chemical Information and Modeling* 58.7, págs. 1331-1342. DOI: [10.1021/acs.jcim.7b00721](https://doi.org/10.1021/acs.jcim.7b00721).
- Prada Gori, D. N. et al.** (2022). “iRaPCA and SOMoC: Development and Validation of Web Applications for New Approaches for the Clustering of Small Molecules”. *Journal of Chemical Information and Modeling*. DOI: [10.1021/acs.jcim.2c00265](https://doi.org/10.1021/acs.jcim.2c00265).
- Qiao, X. et al.** (2014). “Properties of human brain sodium channel α -subunits expressed in HEK293 cells and their modulation by carbamazepine, phenytoin and lamotrigine”. *British Journal of Pharmacology* 171.4, págs. 1054-1067. DOI: [10.1111/bph.12534](https://doi.org/10.1111/bph.12534).
- Storozhuk, M. V., O. F. Moroz y A. V. Zholos** (2019). “Multifunctional TRPV1 Ion Channels in Physiology and Pathology with Focus on the Brain, Vas-

culature, and Some Visceral Systems". *BioMed Research International* 2019, pág. 5806321. DOI: [10.1155/2019/5806321](https://doi.org/10.1155/2019/5806321).

Detalle de publicaciones, congresos y becas

Durante este trabajo de tesis doctoral se realizaron 19 presentaciones a congresos (6 nacionales y 13 internacionales) y 10 publicaciones en revistas científicas con referato. A su vez, 4 publicaciones fueron submitidas y se encuentran en etapa de revisión.

En el año 2017, recibimos una beca de pasaje y estadía para asistir al *workshop* “SIRAH OpenLab: performing MD simulations with the SIRAH force field”, Institut Pasteur, Montevideo, Uruguay.

La Society for Laboratory Automation and Screening (SLAS) otorgó, en tres oportunidades, el premio Tony B. Academic Travel Award, que consistió de una beca de viaje, estadía e inscripción para participar de las ediciones:

- SLAS2018 International Conference and Exhibition - San Diego, CA, USA.
- SLAS2019 International Conference and Exhibition - Washington DC, USA.
- SLAS2019 Europe International Conference and Exhibition - Barcelona, España.

En el marco del proyecto REpeat protein Function Refinement, Annotation and Classification of Topologies (REFRACT) - Marie Skłodowska Curie Research and Innovation Staff Exchange (RISE)-COST Action, un consorcio internacional de investigación en proteínas repetitivas, en 2019 se realizó una estancia de 6 meses en el Laboratoire de Bioinformatique Structurale et Modélisation Moléculaire (BiSMM) - Centre National de la Recherche Scientifique (CNRS) - Montpellier, Francia.

En 2019 Capacity Building for Bioinformatics in Latin America (CABANA) otorgó una beca de viaje, estadía y viáticos para asistir al workshop “Resources and tools for target identification and drug discovery”, Mendoza, Argentina. Ese mismo año, A Global Network for Neglected Tropical Diseases otorgó una beca de viaje, estadía y viáticos para asistir al “Workshop in Techniques and Technologies in Drug Discovery”, en la misma ciudad de Mendoza.

En el año 2020 The World Academy of Sciences for the advancement of science in developing countries (TWAS) junto a la Bibliotheca Alexandrina otorgaron una beca de viaje e inscripción para asistir a BiovisionAlexandria 2020 conference en Alexandria, Egipto (Evento pospuesto debido a la pandemia de COVID19).

En el año 2021 se obtuvo por concurso un subsidio de jóvenes investigadores de la Universidad Nacional de La Plata.

A continuación, se detallan las publicaciones realizadas durante el transcurso de esta tesis:

10. **Novel Dimeric hHv1 Model and Structural Bioinformatic Analysis Reveal an ATP Binding Site Resulting in a Channel Activating Effect.** Llanos et al., Journal of Chemical Information and Modeling, 2022, [10.1021/acs.jcim.1c01396](https://doi.org/10.1021/acs.jcim.1c01396).
9. **Structure-based virtual screening identifies Novobiocin, Montelukast and Cinnarizine as TRPV1 modulators with anticonvulsant activity in vivo.** Llanos et al., Journal of Chemical Information and Modeling, 2022, [10.1021/acs.jcim.2c00312](https://doi.org/10.1021/acs.jcim.2c00312).
8. **iRaPCA and SOMoC: Development and validation of Web Applications for new approaches for the clustering of small molecules.** Prada Gori DN., Llanos MA., Bellera CL., Talevi A., and Alberca LN., Journal of Chemical Information and Modeling, 2022, [10.1021/acs.jcim.2c00265](https://doi.org/10.1021/acs.jcim.2c00265)
7. **Homology Modeling and Molecular Dynamics Simulations of Trypanosoma cruzi Phosphodiesterase b1.** Llanos et al., Chemistry and Biodiversity, 2021. [10.1002/cbdv.202100712](https://doi.org/10.1002/cbdv.202100712)

6. **Strengths and weaknesses of docking simulations in the SARS-CoV-2 era: the main protease (MPro) case study.** Llanos et al., Journal of Chemical Information and Modeling, 2021. [10.1021/acs.jcim.1c00404](https://doi.org/10.1021/acs.jcim.1c00404)
5. **Computational and synthetic target-based approaches to the discovery of novel anticonvulsant compounds.** Gantner ME, Llanos MA, Garofalo FM, Villalba ML, and Gavernet L., Current Medicinal Chemistry, 2021. [10.2174/0929867328666210506161234](https://doi.org/10.2174/0929867328666210506161234)
4. **Can drug repurposing strategies be the solution to the COVID-19 crisis?** Bellera CL., Llanos MA., Gantner ME., Rodriguez S., Gavernet L., Comini M., and Talevi A. Expert Opinion on Drug Discovery, 2020. [10.1080/17460441.2021.1863943](https://doi.org/10.1080/17460441.2021.1863943)
3. **A Structure-Based Approach towards the Identification of Novel Antichagasic Compounds: Trypanosoma Cruzi Carbonic Anhydrase Inhibitors.** Llanos et al., Journal of Enzyme Inhibition and Medicinal Chemistry, 2020. [10.1080/14756366.2019.1677638](https://doi.org/10.1080/14756366.2019.1677638)
2. **Point mutations affecting yeast prion propagation change the structure of its amyloid fibrils.** Anna I. Sulatskaya, Stanislav A. Bondarev, Maksim I. Sulatsky, Nina P. Trubitsina, Mikhail V. Belousov, Galina A. Zhouravleva, Manuel A. Llanos, Andrey V. Kajava, Irina M. Kuznetsova, Konstantin K. Turoverov; Journal of Molecular Liquids, 2020. [10.1016/j.molliq.2020.113618](https://doi.org/10.1016/j.molliq.2020.113618)
1. **Citrus reticulata peel oil as an antiatherogenic agent: hypolipogenic effect in hepatic cells, lipid storage decrease in foam cells, and prevention of LDL oxidation.** Castro MA., Llanos MA., Rodenak-Kladniew BE., Gavernet L., Galle ME., and Crespo R., Metabolism and Cardiovascular Diseases, 2020. [10.1016/j.numecd.2020.04.033](https://doi.org/10.1016/j.numecd.2020.04.033)

Material suplementario

Material suplementario A

Material suplementario Capítulo 3

Tabla A.1: Validación interna del *clustering*. Distintos CVIs calculados por SOMoC para el *clustering* final de las moléculas con un K=9. A fines comparativos se incluye la media y el desvío estándar (SD) de 500 iteraciones de *random clustering*, donde las moléculas son asignadas a cada clúster de manera aleatoria.

	Random		
	Valor	Media	SD
Silhouette score	0.978	-0.143	0.033
Davies Bouldin score	0.170	66.548	18.436
Calinski Harabasz score	203188.197	1.060	0.525
Dunn Index	1.101	0.000	0.000

Tabla A.2: Selección del mejor algoritmo. Métricas de clasificación binaria calculadas mediante 100 x 10-fold CV en el conjunto de entrenamiento, para todos los algoritmos evaluados utilizando los hiperparámetros por defecto en cada caso.

Modelo	ROC-AUC		Accuracy		Kappa		Precision		Recall		F1	
	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
AdaBoost	0.932	0.033	0.896	0.037	0.792	0.073	0.891	0.047	0.914	0.051	0.901	0.035
DT	0.902	0.038	0.902	0.038	0.805	0.075	0.902	0.048	0.913	0.052	0.906	0.036
ExtraTrees	0.980	0.017	0.943	0.029	0.887	0.058	0.933	0.040	0.961	0.034	0.946	0.027
KNN	0.954	0.027	0.924	0.032	0.852	0.063	0.891	0.044	0.979	0.026	0.932	0.028
LGBM	0.965	0.024	0.932	0.030	0.864	0.060	0.919	0.043	0.955	0.037	0.936	0.028
LR	0.950	0.029	0.913	0.034	0.827	0.067	0.898	0.047	0.943	0.041	0.919	0.031
MLP	0.961	0.027	0.930	0.030	0.861	0.060	0.918	0.043	0.953	0.037	0.934	0.028
RF	0.975	0.020	0.942	0.029	0.885	0.057	0.926	0.041	0.968	0.032	0.946	0.026
SGD	0.928	0.038	0.876	0.045	0.752	0.090	0.882	0.053	0.880	0.069	0.879	0.046
SVC	0.968	0.022	0.942	0.028	0.886	0.056	0.923	0.040	0.972	0.029	0.947	0.026

Tabla A.3: Selección del mejor algoritmo. Aleatorización de Fisher para todos los algoritmos evaluados. Para cada métrica se expresa la media y su desvío estándar (SD), calculados mediante 1000 iteraciones de randomización de la variable dependiente.

Modelo	Set	ROC-AUC		Accuracy		Kappa		Precision		Recall		F1	
		Media	SD	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
AdaBoost	Test	0.532	0.061	0.018	0.110	0.521	0.052	0.509	0.055	0.546	0.079	0.509	0.055
DT	Test	0.515	0.047	0.010	0.084	0.519	0.042	0.505	0.042	0.513	0.060	0.505	0.042
ExtraTrees	Test	0.513	0.059	-0.011	0.111	0.508	0.054	0.494	0.055	0.520	0.070	0.494	0.055
KNN	Test	0.515	0.063	-0.011	0.112	0.508	0.055	0.495	0.056	0.526	0.081	0.495	0.056
LGBM	Test	0.526	0.060	0.008	0.114	0.517	0.054	0.504	0.057	0.537	0.074	0.504	0.057
LogisticRegression	Test	0.536	0.069	0.016	0.133	0.520	0.061	0.508	0.066	0.555	0.084	0.508	0.066
MLP	Test	0.522	0.057	0.005	0.101	0.516	0.049	0.503	0.051	0.531	0.071	0.503	0.051
RF	Test	0.508	0.060	-0.015	0.117	0.507	0.057	0.492	0.058	0.510	0.069	0.492	0.058
SGD	Test	0.561	0.117	0.103	0.215	0.562	0.104	0.551	0.107	0.568	0.146	0.551	0.107
SVC	Test	0.556	0.133	0.032	0.248	0.521	0.110	0.516	0.124	0.602	0.169	0.516	0.124

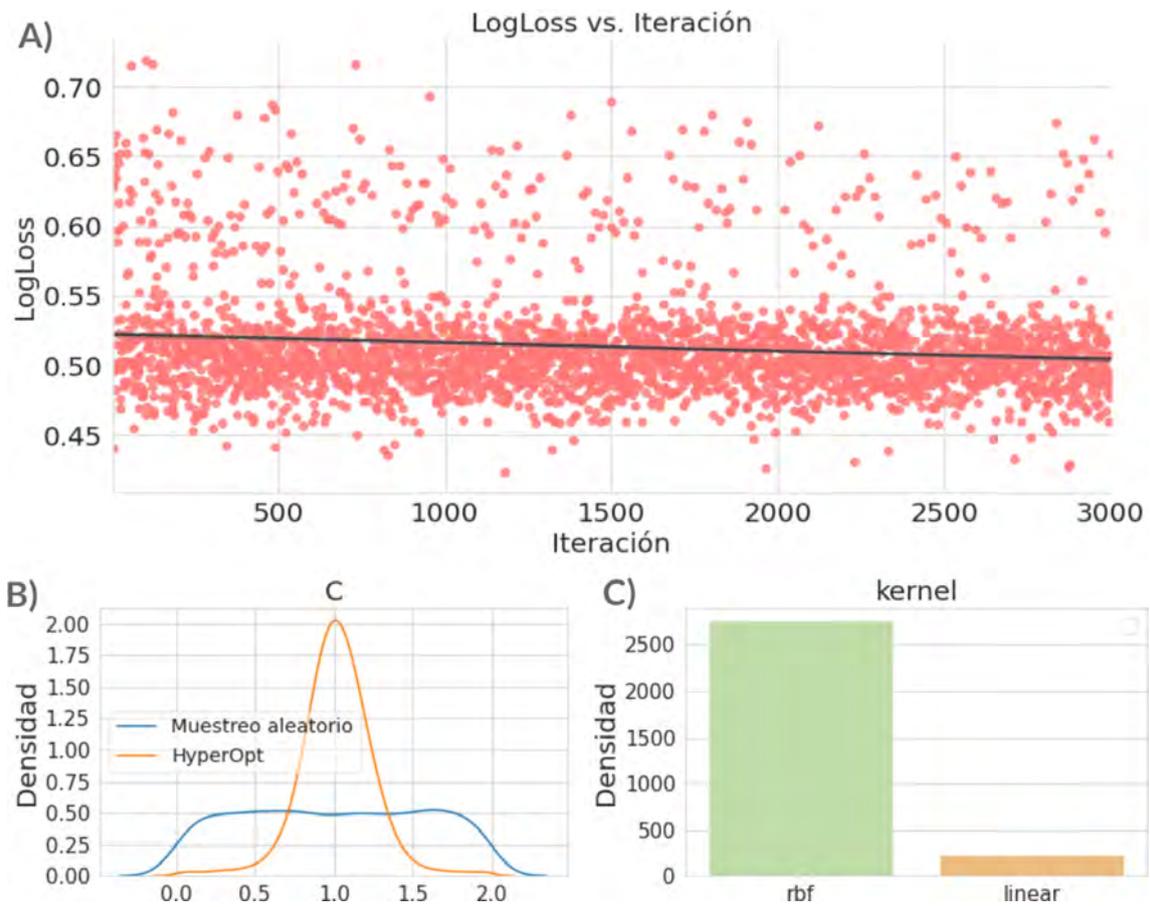


Figura A.1: Optimización del algoritmo SVC. En el panel superior (A) se muestra como varía la métrica de pérdida (LogLoss) a medida que avanza la búsqueda. La línea negra representa una regresión lineal entre los valores de x e y, mientras que el sombreado representa un intervalo de confianza del 95%. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: C (B) y kernel (C). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

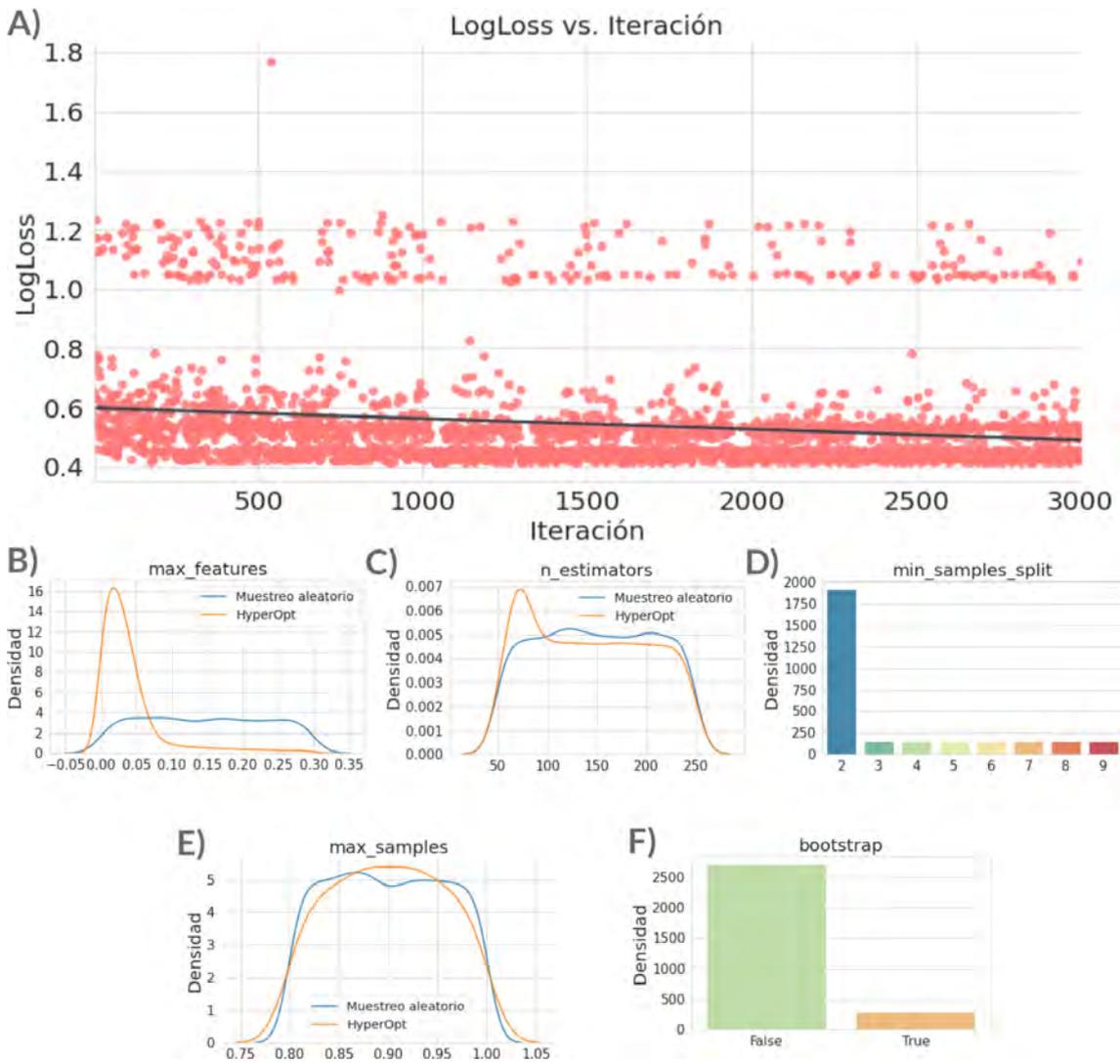


Figura A.2: Optimización del algoritmo RF. En el panel superior (A) se muestra como varía la métrica de pérdida (LogLoss) a medida que avanza la búsqueda. La línea negra representa una regresión lineal entre los valores de x e y, mientras que el sombreado representa un intervalo de confianza del 95 %. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: max_features (B), n_estimators (C), min_samples_split (D), max_samples (E) y bootstrap (F). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

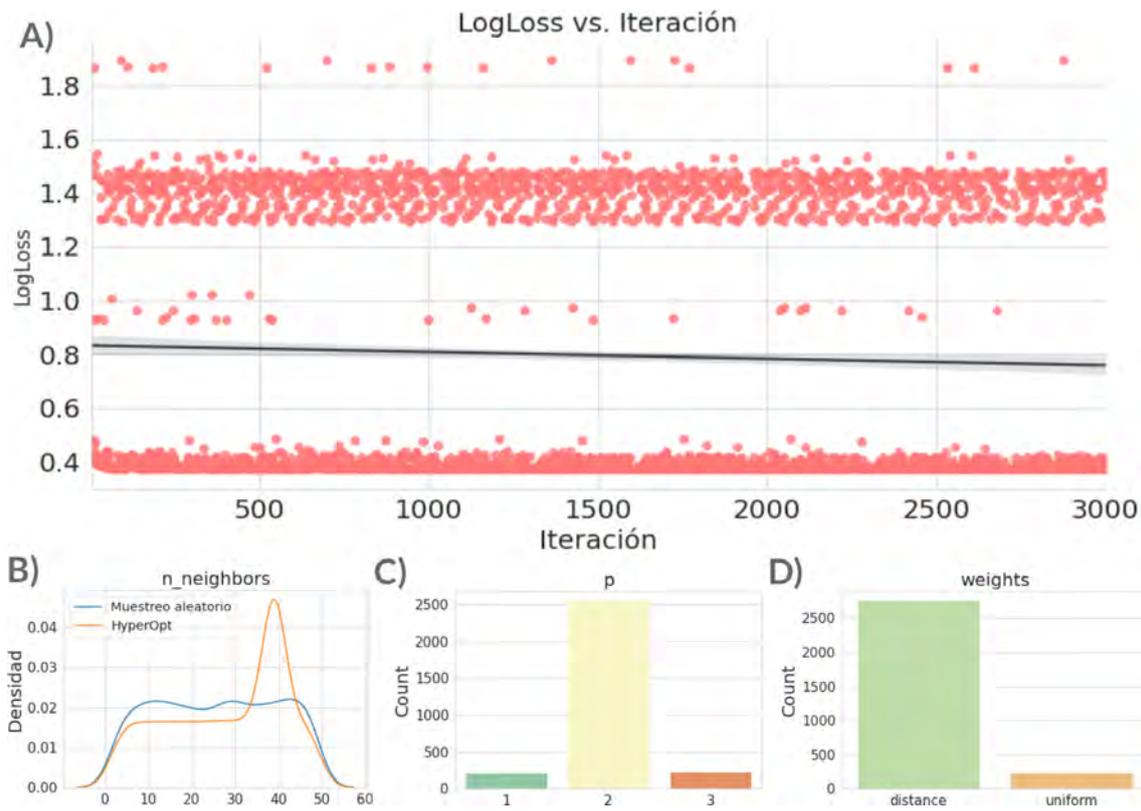


Figura A.3: Optimización del algoritmo KNN. En el panel superior (A) se muestra como varía la métrica de pérdida (LogLoss) a medida que avanza la búsqueda. La línea negra representa una regresión lineal entre los valores de x e y, mientras que el sombreado representa un intervalo de confianza del 95 %. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: n_neighbors (B), p (C) y weights (D). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

Tabla A.4: Aleatorización de Fisher para los algoritmos optimizados. Para cada métrica se expresa la media y su desvío estándar (SD), calculados mediante 100 iteraciones de randomización de la variable dependiente.

		ROC-AUC		Accuracy		Kappa		Precision		Recall		F1	
Modelo	Set	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
ExtraTrees	Test	0.494	0.055	0.494	0.055	-0.013	0.111	0.507	0.054	0.507	0.076	0.506	0.062
KNN	Test	0.487	0.103	0.487	0.103	-0.026	0.207	0.497	0.094	0.553	0.146	0.522	0.114
RF	Test	0.495	0.057	0.495	0.057	-0.010	0.114	0.509	0.054	0.520	0.074	0.513	0.060
SVC	Test	0.507	0.081	0.507	0.081	0.014	0.162	0.517	0.076	0.548	0.113	0.531	0.092

Tabla A.5: Validación del *scoring power* para el modelo TRPV1. Valores de AUC-ROC obtenidos para el conjunto de datos balanceado, para todas las combinaciones de modelo de homología, programa de docking y función de puntuación evaluada para el *re-scoring*. La media y su desvío estándar (SD) se estimaron mediante 1000 muestras de *bootstrapping*.

Modelo	Docking	Re-scoring	AUC-ROC	
			Media	SD
Agonist (5IRX)	QVina2	-	0.504	0.013
Agonist (5IRX)	QVina2	ad4	0.721	0.019
Agonist (5IRX)	QVina2	vina	0.616	0.021
Agonist (5IRX)	QVina2	vinardo	0.683	0.020
APO (5IRZ)	QVina2	-	0.772	0.010
APO (5IRZ)	QVina2	ad4	0.837	0.015
APO (5IRZ)	QVina2	vina	0.794	0.017
APO (5IRZ)	QVina2	vinardo	0.801	0.017
Antagonist (5IS0)	QVina2	-	0.635	0.012
Antagonist (5IS0)	QVina2	ad4	0.785	0.017
Antagonist (5IS0)	QVina2	vina	0.644	0.021
Antagonist (5IS0)	QVina2	vinardo	0.763	0.018
Agonist (5IRX)	AD4-GPU	-	0.684	0.014
Agonist (5IRX)	AD4-GPU	ad4	0.779	0.018
Agonist (5IRX)	AD4-GPU	vina	0.676	0.020
Agonist (5IRX)	AD4-GPU	vinardo	0.722	0.020
APO (5IRZ)	AD4-GPU	-	0.760	0.012
APO (5IRZ)	AD4-GPU	ad4	0.802	0.016
APO (5IRZ)	AD4-GPU	vina	0.712	0.020
APO (5IRZ)	AD4-GPU	vinardo	0.723	0.018
Antagonist (5IS0)	AD4-GPU	-	0.787	0.012
Antagonist (5IS0)	AD4-GPU	ad4	0.836	0.015
Antagonist (5IS0)	AD4-GPU	vina	0.605	0.022
Antagonist (5IS0)	AD4-GPU	vinardo	0.642	0.022

Tabla A.6: Evaluación del *screening power*. Métricas de enriquecimiento calculadas para todas las combinaciones de modelo de homología, programa de docking y función de puntuación evaluada para el *re-scoring*, utilizando la base de datos simulada DUDE. La media y su desvío estándar (SD) se estimaron mediante 1000 muestras de *bootstrapping*.

Modelo	Docking	Re-scoring	AUC-ROC		BEDROC		RIE		EF 1 %		EF 10 %	
			Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
Agonist (5IRX)	AD4-GPU	-	0.452	0.032	0.024	0.017	0.403	0.282	0.000	0.000	0.277	0.255
Agonist (5IRX)	AD4-GPU	ad4	0.655	0.037	0.060	0.018	1.006	0.293	0.000	0.000	1.233	0.474
Agonist (5IRX)	AD4-GPU	vina	0.639	0.033	0.077	0.022	1.279	0.364	0.000	0.000	1.559	0.521
Agonist (5IRX)	AD4-GPU	vinardo	0.696	0.032	0.103	0.025	1.713	0.415	0.000	0.000	2.307	0.587
APO (5IRZ)	AD4-GPU	-	0.375	0.038	0.057	0.029	0.946	0.485	1.515	1.900	0.828	0.405
APO (5IRZ)	AD4-GPU	ad4	0.622	0.044	0.110	0.027	1.841	0.455	0.349	1.043	2.143	0.582
APO (5IRZ)	AD4-GPU	vina	0.678	0.039	0.123	0.033	2.046	0.550	1.786	1.831	2.140	0.597
APO (5IRZ)	AD4-GPU	vinardo	0.656	0.038	0.122	0.032	2.036	0.527	0.057	0.482	2.022	0.553
Antagonist (5IS0)	AD4-GPU	-	0.497	0.034	0.046	0.025	0.772	0.409	0.968	1.612	0.607	0.344
Antagonist (5IS0)	AD4-GPU	ad4	0.670	0.034	0.094	0.031	1.569	0.517	1.935	1.927	1.624	0.537
Antagonist (5IS0)	AD4-GPU	vina	0.576	0.039	0.094	0.034	1.572	0.573	4.084	2.747	1.405	0.539
Antagonist (5IS0)	AD4-GPU	vinardo	0.608	0.035	0.055	0.018	0.920	0.305	0.000	0.000	0.939	0.423
Agonist (5IRX)	QVina2	-	0.584	0.019	0.016	0.003	0.274	0.047	0.000	0.000	0.061	0.071
Agonist (5IRX)	QVina2	ad4	0.727	0.033	0.185	0.037	3.084	0.610	0.075	0.433	3.477	0.670
Agonist (5IRX)	QVina2	vina	0.680	0.033	0.079	0.021	1.318	0.351	0.000	0.000	1.663	0.525
Agonist (5IRX)	QVina2	vinardo	0.665	0.032	0.073	0.020	1.224	0.338	0.000	0.000	1.578	0.528
APO (5IRZ)	QVina2	-	0.706	0.020	0.127	0.018	2.117	0.299	0.007	0.091	2.103	0.334
APO (5IRZ)	QVina2	ad4	0.766	0.038	0.257	0.046	4.278	0.761	3.673	2.835	3.885	0.709
APO (5IRZ)	QVina2	vina	0.698	0.034	0.135	0.031	2.252	0.511	0.000	0.000	2.409	0.590
APO (5IRZ)	QVina2	vinardo	0.694	0.035	0.138	0.033	2.306	0.554	0.033	0.321	2.578	0.613
Antagonist (5IS0)	QVina2	-	0.633	0.019	0.067	0.012	1.124	0.196	0.000	0.000	1.560	0.302
Antagonist (5IS0)	QVina2	ad4	0.763	0.031	0.185	0.037	3.086	0.618	3.820	2.729	3.584	0.644
Antagonist (5IS0)	QVina2	vina	0.714	0.030	0.121	0.030	2.016	0.497	0.155	0.704	2.548	0.636
Antagonist (5IS0)	QVina2	vinardo	0.715	0.030	0.115	0.028	1.914	0.465	0.000	0.000	2.225	0.590

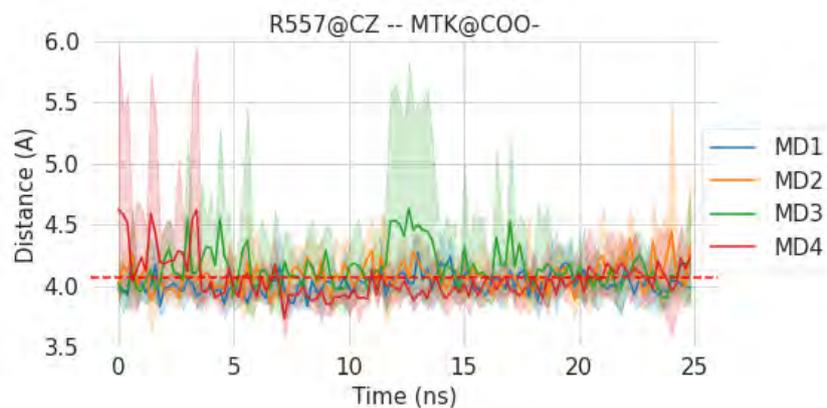


Figura A.4: B) Distancia (Å) entre el carbono Cz de la Arg557 y el carbono carbonílico del MTK, calculada a lo largo de toda la simulación para todas las copias del ligando y agrupada por corrida. La línea roja punteada muestra el valor medio calculado para todos los ligandos y corridas: 4.07 ± 0.21 Å.

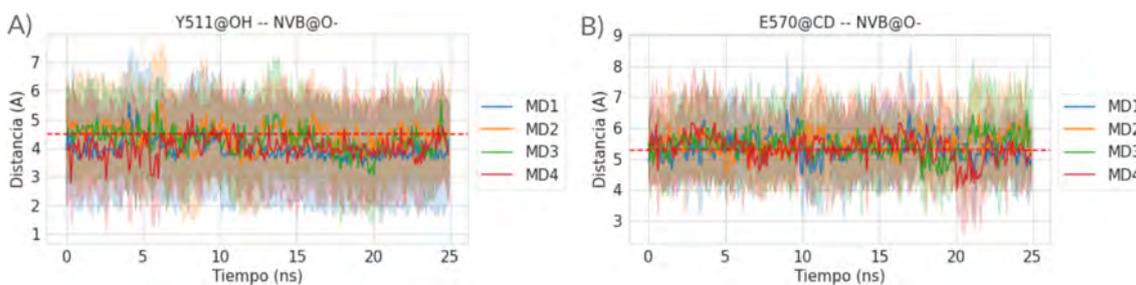


Figura A.5: A) Distancia (Å) entre el oxígeno de la Tyr511 y el oxígeno cumarínico de la NVB, calculada a lo largo de toda la simulación para todas las copias del ligando y agrupada por corrida. La línea roja punteada muestra el valor medio calculado para todos los ligandos y corridas: 4.41 ± 1.48 Å. B) Distancia (Å) entre el carbono Cd del Glu570 y el oxígeno del grupo hidroxilo libre de NVB, calculada a lo largo de toda la simulación para todas las copias del ligando y agrupada por corrida. La línea roja punteada muestra el valor medio calculado para todos los ligandos y corridas: 5.28 ± 1.10 Å.

Material suplementario B

Material suplementario Capítulo 4

Tabla B.1: Validación interna del *clustering* para el **Dataset A**. Distintos CVIs calculados por SOMoC para el *clustering* final de las moléculas con un $K=8$. A fines comparativos se incluye la media y el desvío estándar (SD) de 500 iteraciones de *random clustering*, donde las moléculas son asignadas a cada clúster de manera aleatoria.

	Random		
	Valor	Media	SD
Silhouette score	0.871	-0.113	0.019
Davies Bouldin score	0.273	26.336	9.688
Calinski Harabasz score	1767.053	1.003	0.403
Dunn Index	0.015	0.000	0.000

Tabla B.2: Selección del mejor algoritmo. *Performance* de los algoritmos evaluados, expresada como la media de 100 x 10-fold CV en el conjunto de entrenamiento acompañada de su desvío estándar (SD). Todos los algoritmos se entrenaron utilizando los hiperparámetros por defecto.

Modelo	MAE		R^2		Rho		TauB	
	Media	SD	Media	SD	Media	SD	Media	SD
GBT	0.382	0.065	0.648	0.162	0.805	0.092	0.644	0.094
LGBM	0.385	0.066	0.635	0.175	0.796	0.096	0.635	0.096
LR	0.409	0.067	0.607	0.164	0.795	0.082	0.622	0.086
Ridge	0.407	0.067	0.610	0.162	0.796	0.081	0.623	0.085
SGD	0.400	0.064	0.628	0.149	0.804	0.077	0.633	0.083
SVR	0.352	0.054	0.719	0.108	0.841	0.069	0.673	0.080

Tabla B.3: Selección del mejor algoritmo. Aleatorización de Fisher para todos los algoritmos evaluados. Para cada métrica se expresa la media y su desvío estándar (SD), calculados mediante 1000 iteraciones de aleatorización de la variable dependiente.

Modelo	MAE		R^2		Rho		TauB	
	Media	SD	Media	SD	Media	SD	Media	SD
GBT	0.858	0.062	-0.245	0.169	-0.014	0.163	-0.009	0.112
LGBM	0.877	0.066	-0.289	0.180	-0.003	0.164	-0.003	0.111
SLR	0.819	0.060	-0.133	0.155	-0.014	0.236	-0.008	0.161
Ridge	0.818	0.054	-0.126	0.134	-0.012	0.204	-0.009	0.138
SGD	0.807	0.050	-0.104	0.122	-0.011	0.195	-0.007	0.133
SVR	0.800	0.049	-0.089	0.131	0.023	0.204	0.014	0.138

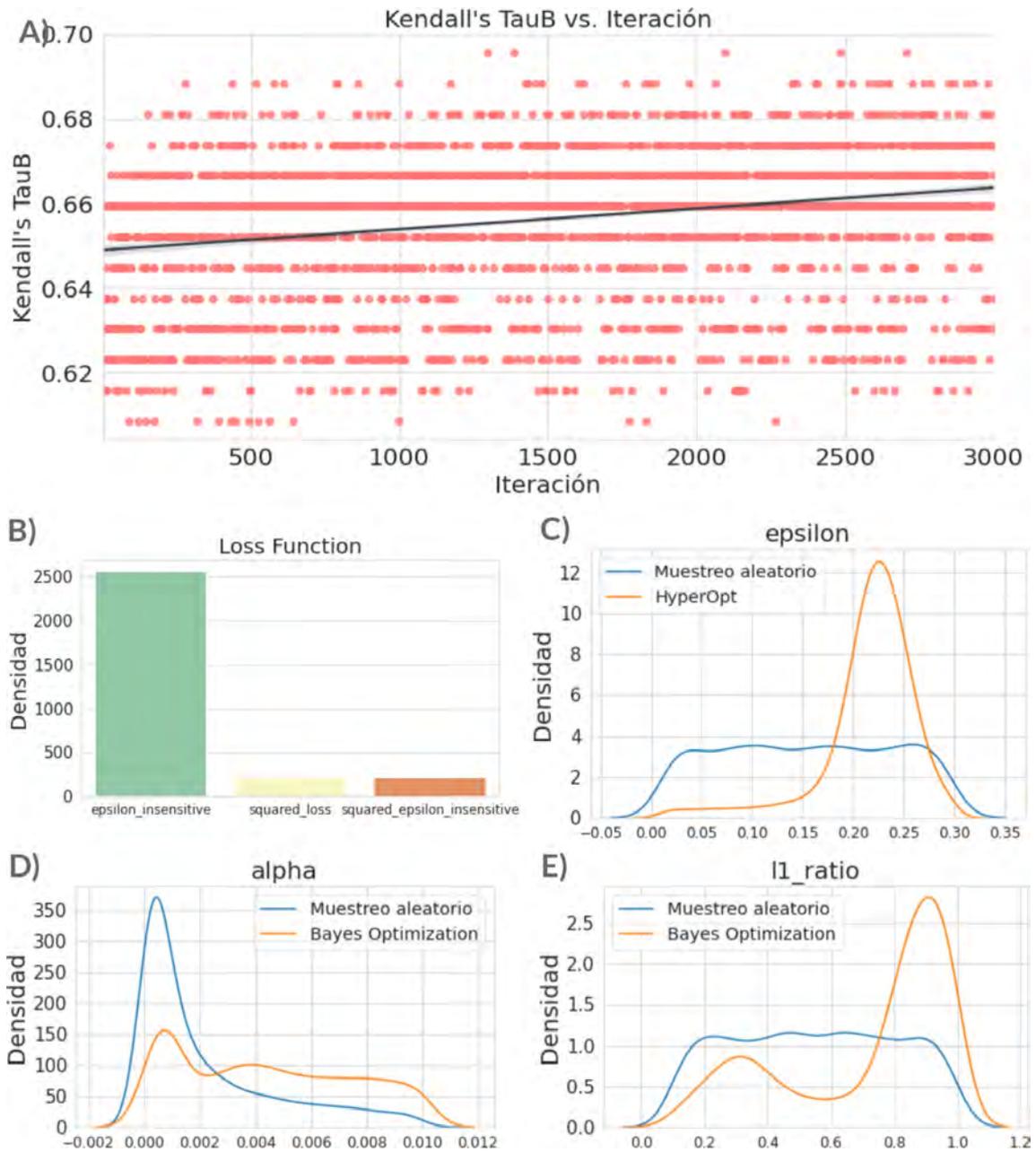


Figura B.1: Optimización del algoritmo SGD. En el panel superior (A) se muestra como varía la métrica de pérdida (TauB) a medida que avanza la búsqueda. La línea negra representa una regresión lineal de los datos y el sombreado un intervalo de confianza del 95%. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: Loss Function (B), epsilon (C), alpha (D), l1_ratio (E). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

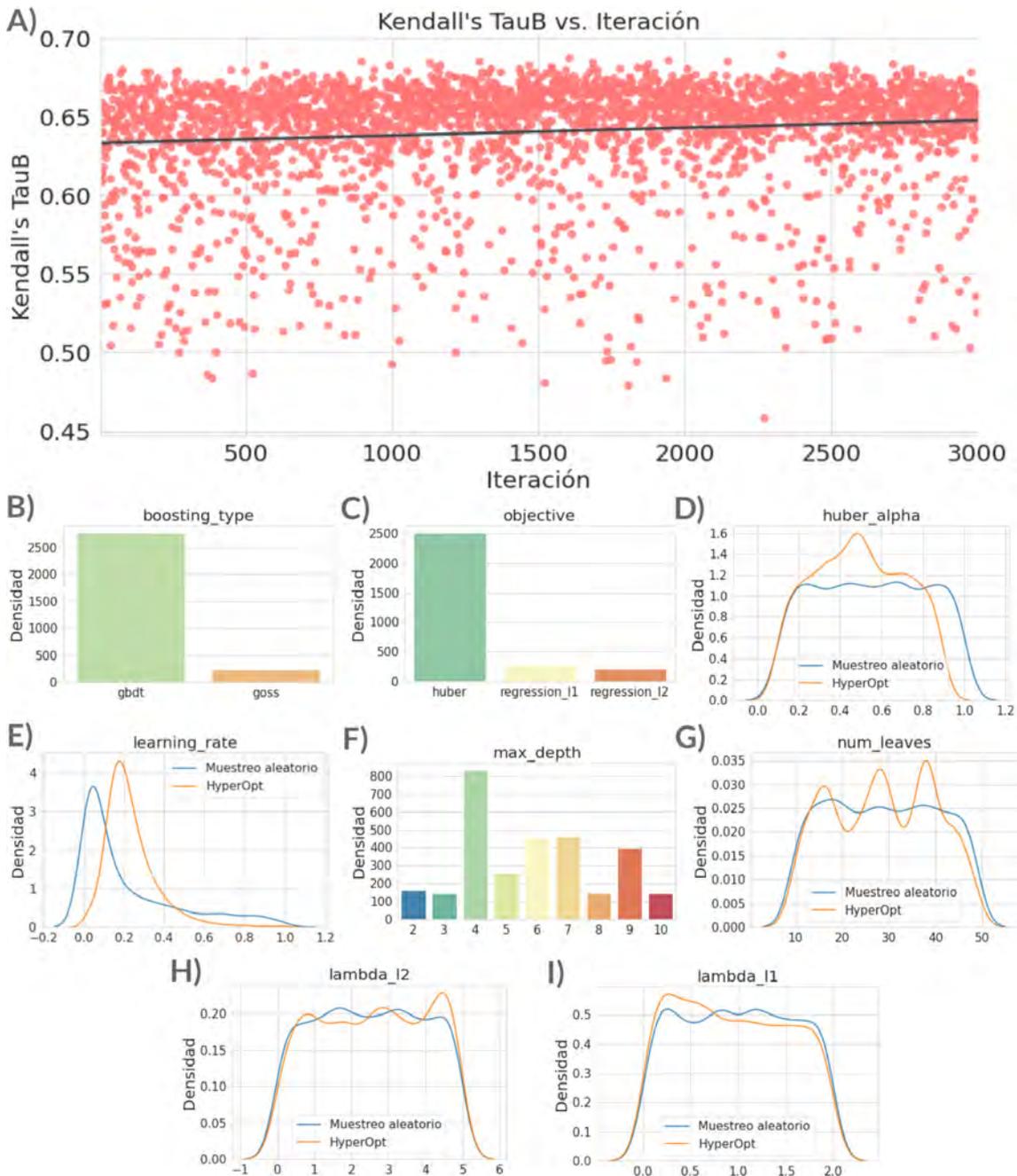


Figura B.2: Optimización del algoritmo LGBM. En el panel superior (A) se muestra como varía la métrica de pérdida (TauB) a medida que avanza la búsqueda. La línea negra representa una regresión lineal de los datos y el sombreado un intervalo de confianza del 95%. En los paneles inferiores se observan las distribuciones de los hiperparámetros optimizados: `boosting_type` (B), `objective` (C), `huber_alpha` (D), `learning_rate` (E), `max_depth` (F), `num_leaves` (G), `lambda_l2` (H) y `lambda_l1` (I). Para el caso de las variables continuas, en azul se representa la distribución de valores que se obtiene con un muestreo aleatorio, mientras que en naranja se representa el muestreo que efectivamente realizó HyperOpt.

Tabla B.4: Aleatorización de Fisher para los algoritmos optimizados. Para cada métrica se expresa la media y su desvío estándar (SD), calculado mediante 1000 iteraciones de aleatorización de la variable dependiente.

Modelo	MAE		R^2		Rho		TauB	
	Media	SD	Media	SD	Media	SD	Media	SD
SGD	0.792	0.068	-0.084	0.172	0.003	0.304	0.002	0.210
SVR	0.784	0.050	-0.056	0.131	0.003	0.250	0.002	0.172
LGBM	0.799	0.059	-0.093	0.154	-0.005	0.241	-0.003	0.164

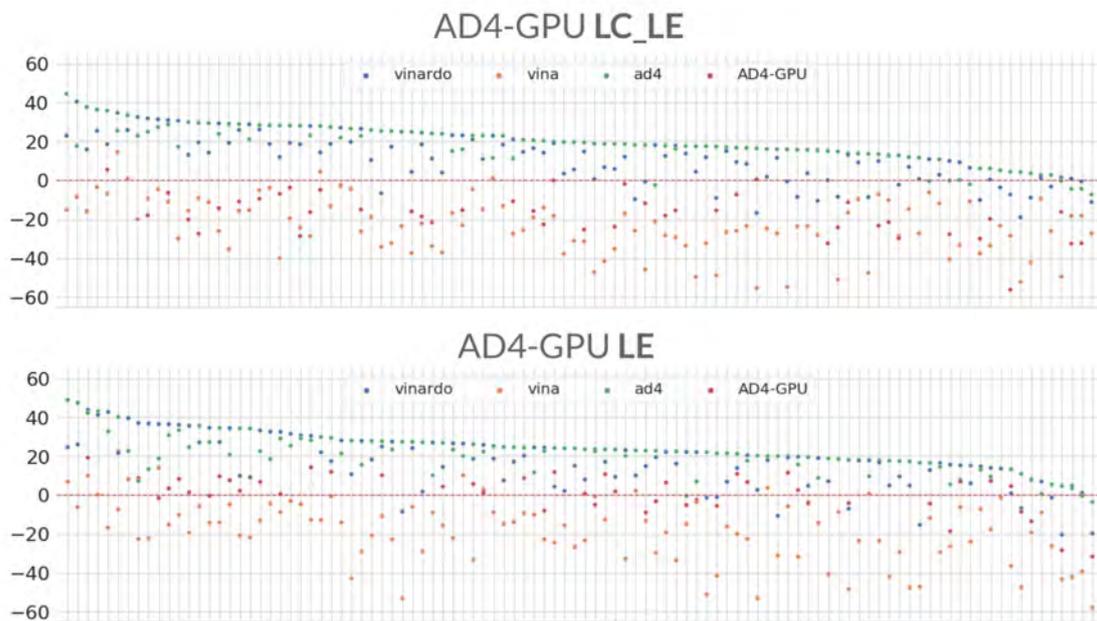


Figura B.3: Selección de la mejor estructura. Las 102 conformaciones del canal NaV1.2 se encuentran representadas en el eje X, mientras que en el eje Y se muestra la *performance* relativa al modelo de base, en términos de la métrica TauB, considerando las poses LE y LC LE generadas por AD4-GPU. Las estructuras se encuentran ordenadas en orden decreciente según su porcentaje de mejora, considerando la función de puntuación ad4 utilizada durante el *-scoring*.

MATERIAL SUPLEMENTARIO B. MATERIAL SUPLEMENTARIO

CAPÍTULO 4

Dataset A. Conjunto de datos que se utilizó para desarrollar y validar los modelos QSAR y de docking para el canal Nav1.2. Las estructuras de los compuestos se muestran en formato SMILES (del inglés, Simplified Molecular Input Line Entry Specification).

ChEMBL_ID	SMILES	pIC50
CHEMBL75880	C(C1CCCC1)C2CCCC2)C3CCCC3	6.29
CHEMBL334104	C(C1CCCC1)c2nc(c[nH]2)c3ccc(cc3)c4ccccc4	7.16
CHEMBL43064	C(C=Cc1ccccc1)N2CCN(CC2)C(c3ccccc3)c4ccccc4	6.37
CHEMBL3098423	C(CC(c1ccccc1)c2ccccc2)NC(c3ccccc3)c4ccccc4	6.12
CHEMBL331513	C(CC1CCCC1)C2nc(c[nH]2)c3ccc(cc3)c4ccccc4	7.72
CHEMBL3098425	C(CCC(c1ccccc1)c2ccccc2)NC(c3ccccc3)c4ccccc4	6.25
CHEMBL3098427	C(CCCC(c1ccccc1)c2ccccc2)C(CCN(c3ccccc3)c4ccccc4	6.25
CHEMBL3098426	C(CCN(c1ccccc1)c2ccccc2)C(Cc3ccccc3)c4ccccc4	6.46
CHEMBL3098420	C(CN1CCN(CC1)C)c2ccccc2)c3ccccc3)C4ccccc4	6.72
CHEMBL3098424	C(CNC(c1ccccc1)c2ccccc2)C(Cc3ccccc3)c4ccccc4	5.88
CHEMBL20166	C(CNCCCc1ccccc1)Cc2ccccc2	6.60
CHEMBL283775	C(CNCCc1ccccc1)Cc2ccccc2	6.47
CHEMBL3098421	C(Cc1ccccc1)N2CCN(CC2)C(c3ccccc3)c4ccccc4	6.59
CHEMBL119139	C(Cc1nc(c[nH]1)c2ccc(cc2)c3ccccc3)C4CCCC4	7.58
CHEMBL3098422	C(NC(c1ccccc1)c2ccccc2)C(c3ccccc3)c4ccccc4	5.50
CHEMBL3037819	C1[C@@H](N[C@@H](c2ccccc2)c3[nH]c4ccccc4c13)c5nc(c[nH]5)c6ccccc6	5.84
CHEMBL1276308	CC#C[C@H](1)O[C@H](C)[C@H]2[C@@H]3COC4=C(O)CCC4=C3[C@H](C)[C@H]12C)c5ccc(cc5)N(C)C	5.23
CHEMBL157101	CC(=O)N1CCN(CC1)c2ccc(OCC3COC(Cn4cnc4)(O3)c5ccc(Cl)cc5Cl)cc2	4.62
CHEMBL1400945	CC(=O)[C@H]1[C@@H](C)[C@H]2[C@@H]3C=C4C[C@@H](O)CC[C@H]4(C)[C@H]3CC(C)[12C]C#N	4.61
CHEMBL221753	CC(C)(C)CC(C)C)c1ccc(OCCOCC[N+](C)(C)Cc2ccccc2)cc1	6.59
CHEMBL239764	CC(C)(C)COC(=O)NCCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	6.43
CHEMBL114	CC(C)(C)NC(=O)[C@H]1[C@@H](C)[C@H]2CC[C@H]2CN1C[C@@H](O)[C@H](Cc3ccccc3)NC(=O)[C@H](CC(=O)N)NC(=O)C4ccc5ccccc5n4	6.12
CHEMBL394196	CC(C)(C)OC(=O)NCCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	6.74
CHEMBL239345	CC(C)(C)OC(=O)NCCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.19
CHEMBL239133	CC(C)(C)OC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	6.84
CHEMBL305660	CC(C)(C)c1ccc(cc1)C(=O)CCN2CC(C)C2)OC(c3ccccc3)c4ccccc4	7.40
CHEMBL3098418	CC(C)(C)c1ccc(cc1)C(=O)CCN2CCN(CC2)C(c3ccccc3)c4ccccc4	6.89
CHEMBL17157	CC(C)(C)c1ccc(cc1)C(O)CCN2CC(C)C2)C(O)c3ccccc3)c4ccccc4	7.11
CHEMBL124211	CC(C)CNCCOC(=O)c1ccc(N)cc1	4.26
CHEMBL394245	CC(C)COC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.58
CHEMBL1008	CC(C)COC(CN(Cc1ccccc1)c2ccccc2)N3CCCC3	6.08
CHEMBL493	CC(C)[C@H]1N2C(=O)[C@H](N1)C(=O)[C@H]3CN(C)[C@H]4C5c(Br)[nH]6c8ccc(C4=C3)c56)(O)[C@H]2(O)[C@H]7CCCN7C1=O)C(C)C	5.17
CHEMBL1306	CC(C)N1CCN(CC1)c2ccc(O[C@H]3COC[C@@H](Cn4cnc4)(O3)c5ccc(Cl)cc5Cl)cc2	5.54
CHEMBL27	CC(C)NCC(O)COc1ccccc2ccccc12	4.81
CHEMBL672	CC(C)OC(=O)C(C)(C)OC1ccc(cc1)C(=O)c2ccc(Cl)cc2	4.71
CHEMBL1255837	CC(C)[C@H]1(NC(=O)[C@H]2[C@H]3[C@@H](C4c[nH]5c6ccc3c45)N(C)2)O[C@H]6(O)[C@H]7CCCN7C1=O)C[C@H](Cc8ccccc8)N6C1=O	5.64
CHEMBL163	CC(C)[C@H](N(C)=O)N(C)C1csc(n1)C(C)C(=O)N[C@H](C)[C@H](O)[C@H](Cc2ccccc2)NC(=O)CC3cnc3)C4ccccc4	5.62
CHEMBL596226	CC(C)c1cc(CCCCC(O)(C=O)N(C)F)F)cc(C(C)C)c1O	4.66
CHEMBL595057	CC(C)c1cc(CCCCC(O)(C=O)N(C)F)F)cc(C(C)C)c1O	5.24
CHEMBL526	CC(C)c1ccccc(C(C)C)c1O	4.59
CHEMBL643	CC(CN1c2ccccc2Sc3ccccc13)N(C)C	5.62
CHEMBL1197	CC(CN1CCCC1)OC(=O)c2ccccc2	5.15
CHEMBL3098428	CC(CNc1ccccc1)OC(=O)c2ccccc2	5.15
CHEMBL753	CC(COc1ccccc1)N(CCC)C2c2ccccc2	4.35
CHEMBL24072	CC(Cc1ccccc1)NCCC(c2ccccc2)c3ccccc3	6.52
CHEMBL1762	CC(N)C(=O)Nc1c(C)ccccc1C	3.70
CHEMBL1255834	CC(N)Cc1c[nH]c2ccc(OCC3ccccc3)cc12	5.66
CHEMBL225197	CC(O)(C=O)N1CCN(CC1)C(c2ccccc2)c3ccccc3)c4ccccc4	5.20
CHEMBL584171	CC1(C)N=C(N)N=C(N)N1C2ccc(Cl)cc1C2	4.72
CHEMBL574661	CC1(C)N=C(N)N=C(N)N1C2ccc(Cl)cc1C2	4.45
CHEMBL116028	CC1(C)N=C(N)N=C(N)N1C2ccccc2	4.27
CHEMBL575493	CC1(C)N=C(N)N(C(=N1)Nc2ccc(Br)cc2	4.74
CHEMBL533117	CC1(C)N(C)=NC(=N1)N)Nc2ccc(Cl)cc1C2	4.61
CHEMBL582877	CC1(C)N(C)=NC(=N1)N)Nc2ccc(Br)cc1C2	4.90
CHEMBL582944	CC1(C)N(C)=NC(=N1)N)Nc2ccc(Cl)cc1C2	4.78
CHEMBL573515	CC1(C)N(C)=NC(=N1)N)Nc2ccccc2Br	4.24
CHEMBL573274	CC1(C)N(C)=NC(=N1)N)Nc2ccccc2Cl	4.07
CHEMBL298513	CC12CC[C@@]3(C)C1CC[C@]2(O)O3)OCc4ccccc(Br)cc4	5.00
CHEMBL290743	CC12N(Cc3ccccc13)C(=O)NC2=O	3.14
CHEMBL127865	CC1CCCCN1CCOC(C=O)c2ccccc2	4.89
CHEMBL83	CCC(=C)/(c1ccccc1)c2ccc(OCCN(C)C)cc2)c3ccccc3	5.90
CHEMBL411	CCC(=C)C(C)c1ccc(O)cc1)c2ccc(O)cc2	5.33
CHEMBL64391	CCC(C)N1N=C(N1=O)c2ccc(cc2)N3CCN(CC3)c4ccc(OCC5COC(Cn6cnc6)(O5)c7ccc(Cl)cc7Cl)cc4	5.59
CHEMBL376872	CCC(O)(C=O)NCCc1ccccc1)c2ccccc2	4.52
CHEMBL3098433	CCC(OCCc1ccccc1)C(=O)CCCc2ccccc2)c3ccccc3	5.37
CHEMBL222105	CCC(OCCc1ccccc1)C(=O)NCCCc2ccccc2)c3ccccc3	5.37
CHEMBL40	CCC1(C=O)N(C)=O)NC1=O)c2ccccc2	5.60
CHEMBL224836	CCC1(NC(=O)N(C1=O)S(=O)(=O)c2ccc(C)cc2)c3ccccc3	4.30
CHEMBL225170	CCC1(NC(=O)N(CCN2CCN(CC2)C(c3ccccc3)c4ccccc4)C1=O)c5ccccc5	5.96
CHEMBL376544	CCC1(NC(=O)N(CN2CCN(CC2)C(c3ccccc3)c4ccccc4)C1=O)c5ccccc5	5.48
CHEMBL553025	CCC1=C[C@@H]2CN(C1)Cc3c([nH]4c5ccccc34)[C@@]2(C2)(C(=O)OC)c5cc6cc(cc5OC)N(C)[C@H]7[C@@](O)([C@H](OC(=O)C)[C@H]8(C)C=CCN9CC[C@]67[C@H]89)C(=O)OC	5.28
CHEMBL119103	CCCC(CCC)c1nc(c[nH]1)c2ccc(F)cc2	6.29
CHEMBL120306	CCCC(CCC)c1nc(c[nH]1)c2ccc(cc2)C(C)C	7.52
CHEMBL118626	CCCC(CCC)c1nc(c[nH]1)c2ccc(cc2)c3ccc(Br)cc3	6.49
CHEMBL118682	CCCC(CCC)c1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.00
CHEMBL390536	CCCC1(NC(=O)N(C1=O)S(=O)(=O)C)c2ccccc2	3.30
CHEMBL288861	CCCC1(NC(=O)NC1=O)c2ccccc2	3.79
CHEMBL222263	CCCC1(OC(=O)NC1=O)c2ccccc2	3.35
CHEMBL39810	CCCCC1(NC(=O)N(C)C1=O)c2ccccc2	3.55
CHEMBL39247	CCCCC1(NC(=O)NC1=O)c2ccc(C)cc2	4.02
CHEMBL39889	CCCCC1(NC(=O)NC1=O)c2ccc(C)cc2	4.24
CHEMBL117331	CCCCC(C)C(C)C)c1nc(c[nH]1)c2ccc(F)cc2	7.60
CHEMBL332257	CCCCC(C)C(C)C)c1nc(c[nH]1)c2ccc(cc2)c3ccccc3	8.10
CHEMBL119494	CCCCC(C)C(C)C)c1nc(c[nH]1)c2ccccc2	7.07
CHEMBL425125	CCCCC(C)C(C)C)c1nc(CN1CCN(CC1)C(c2ccccc2)c3ccccc3)c4ccccc4	5.32
CHEMBL540451	CCCCC(C)C(C)C)c1nc(C)C1=O)c2ccccc2)c3ccccc3	5.44
CHEMBL225408	CCCCC(C)C(C)C)c1nc(C)C1=O)c2ccccc2	5.44
CHEMBL225405	CCCCC(C)C(C)C)c1nc(C)C1=O)c2ccccc2	5.30
CHEMBL594410	CCCCC1(NC(=O)N(C)C1=O)c2ccccc2	4.41
CHEMBL222770	CCCCC1(NC(=O)N(C1=O)S(=O)(=O)C)c2ccccc2	3.95
CHEMBL373877	CCCCC1(NC(=O)N(CN2CCN(CC2)C(c3ccccc3)c4ccccc4)C1=O)c5ccccc5	5.89

MATERIAL SUPLEMENTARIO B. MATERIAL SUPLEMENTARIO CAPÍTULO 4

CHEMBL373380	CCCCC1(NC(=O)N(Cc2ccccc2)C1=O)c3ccccc3	5.16
CHEMBL288505	CCCCC1(NC(=O)NC1=O)c2ccccc2	4.41
CHEMBL374202	CCCCC1(OC(=O)NC1=O)c2ccccc2	3.50
CHEMBL289499	CCCCC1(NC(=O)NC1=O)c2ccccc2	4.89
CHEMBL596058	CCCCCCC(O)(C(=O)N)c1cc(C)c(NC(=O)CN(CC)CC)c(C)c1	5.22
CHEMBL287713	CCCCCCC(O)(C(=O)N)c1ccccc1	5.05
CHEMBL225389	CCCCCCC(O)(C(=O)OCC)c1ccccc1	4.44
CHEMBL376661	CCCCCCC1(NC(=O)N(C1=O)S(=O)(=O)C)c2ccccc2	4.89
CHEMBL225495	CCCCCCC1(NC(=O)N(CN2CCN(CC2)C(c3ccccc3)c4ccccc4)C1=O)c5ccccc5	5.66
CHEMBL593347	CCCCCCC1(NC(=O)NC1=O)c2cc(C)c(NC(=O)CN(CC)CC)c(C)c2	4.87
CHEMBL417155	CCCCCCC1(NC(=O)NC1=O)c2ccccc2	5.30
CHEMBL376560	CCCCCCC1(OC(=O)NC1=O)c2ccccc2	4.52
CHEMBL610750	CCCCCCC1(NC(=O)NC1=O)c2ccccc2	5.30
CHEMBL222208	CCCCCCC(O)(C(=O)N)c1ccccc1	5.19
CHEMBL222910	CCCCCCC(O)(C(=O)O)c1ccccc1	4.08
CHEMBL388171	CCCCCCC1(NC(=O)N(C1=O)S(=O)(=O)C)c2ccccc2	4.93
CHEMBL374113	CCCCCCC1(NC(=O)N(Cc2ccccc2)C1=O)c3ccccc3	5.29
CHEMBL39043	CCCCCCC1(NC(=O)NC1=O)c2ccccc2	5.30
CHEMBL388327	CCCCCCC1(OC(=O)NC1=O)c2ccccc2	4.96
CHEMBL588119	CCCCCCC[C@H](C)[C@H](O)c1ccc(SC(C)C)cc1	6.76
CHEMBL239765	CCCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.82
CHEMBL331942	CCCCC1nc(c[nH]1)c2ccc(C(C)C)cc2	7.55
CHEMBL118603	CCCCC1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.80
CHEMBL129529	CCCN(CCC)CCOC(=O)c1ccc(N)cc1	5.57
CHEMBL1098	CCCN1CCCCC1C(=O)Nc2c(C)cccc2C	5.27
CHEMBL698	CCCNc1ccc(cc1)C(=O)OCCN(C)C	5.47
CHEMBL240191	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(Br)cc2	6.88
CHEMBL240400	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(C(C)C)cc2	7.39
CHEMBL240402	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(C)cc2	7.02
CHEMBL392180	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(CCC)cc2	7.21
CHEMBL392179	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(F)cc2	6.00
CHEMBL240192	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)C(F)F	6.34
CHEMBL239981	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)C3CCCC3	8.15
CHEMBL240193	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)N3CCCC3	7.85
CHEMBL239556	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.50
CHEMBL1200	CCCCOc1ccc(cc1N)C(=O)OCCN(CC)CC	5.54
CHEMBL1198	CCCCOc1ccc(OCCN2CCOCC2)cc1	5.00
CHEMBL127643	CCCCc1cc2ccccc2c(OCCN(C)C)n1	5.47
CHEMBL633	CCCCc1cc2ccccc2c1C(=O)c3ccc(I)c(OCCN(CC)CC)c(I)c3	6.18
CHEMBL492	CCCN(CC)C(C)C(=O)Nc1c(C)cccc1C	5.46
CHEMBL531	CCCN1C[C@H](CS)C[C@H]2[C@H]1Cc3c[nH]c4ccccc2c34	5.76
CHEMBL127810	CCCN(C)C(C)COC(=O)c1ccccc1	5.12
CHEMBL1093	CCCN(C)C(=O)Nc1c(C)ccc1C(=O)O	4.29
CHEMBL1194	CCCN(C)C(=O)Nc1ccccc1C	4.27
CHEMBL631	CCCN(C)C(O)C1ccccc1C(=O)OCCc2ccccc2	5.89
CHEMBL239554	CCCCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.07
CHEMBL1769	CCCCc1ccc(N)ccc1C(=O)OCCN(CC)CC	4.62
CHEMBL1196	CCCCc1ccc(cc1N)C(=O)OCCN(CC)CC	5.72
CHEMBL157138	CCN(CC)C(=O)N[C@H]1CN(C)[C@H]2Cc3c[nH]c4ccccc(C2=C1)c34	5.40
CHEMBL1231	CCN(CC)CC#CCOC(=O)C(O)(C1CCCC1)c2ccccc2	5.95
CHEMBL595810	CCN(C)CC(=O)Nc1c(C)ccc(cc1C)C2(NC(=O)NC2=O)c3ccccc3	4.40
CHEMBL79	CCN(CC)CC(=O)Nc1c(C)cccc1C	4.29
CHEMBL593479	CCN(CC)CC(=O)Nc1ccc(cc1)C2(NC(=O)NC2=O)c3ccccc3	4.40
CHEMBL7568	CCN(CC)CCCC(C)Nc1c2ccc(C)cc2nc3ccc(OCC)cc13	5.89
CHEMBL1123	CCN(CC)CCOC(=O)C1(CCCOC1)C2CCCC2	5.71
CHEMBL569	CCN(CC)CCOC(=O)c1ccc(N)cc1	3.96
CHEMBL954	CCN(CC)CCOc1ccc(cc1)C(=C(C))c2ccccc2c3ccccc3	6.20
CHEMBL38588	CCN1C(=O)NC(=O)C1(C)c2ccccc2	3.60
CHEMBL2103747	CCN1C[C@]2(COC)[C@H](O)C[C@]3[C@]4[C@]5[C@]6(O)[C@]7[C@]8(O)[C@]9(C)[C@]10(C)[C@]11(O)C(C)[C@]12(C)[C@]13(C)[C@]14(C)[C@]15(C)[C@]16(O)C(=O)C17	5.35
CHEMBL1480	CCOC(=O)C=C(C)NC(=C(C1c2ccc(C)cc2)C1)C(=O)O	4.43
CHEMBL475534	CCOC(=O)C1=C(C)NC(=C(C1c2ccc(cc2)N+])=O)]O-]C(=O)O)C	4.44
CHEMBL239347	CCOC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	7.15
CHEMBL278172	CCOC(=O)c1ccc(N)cc1	3.14
CHEMBL127123	CCOc1ccc(NC(=Nc2ccc(OCC)cc2)C)cc1	5.77
CHEMBL421669	CCOc1ccc(cc1)C(=O)OCCN(CC)CC	4.62
CHEMBL1221512	CC[C@H](C)[C@H](O)[C@H](O)C1ccccc1C(=O)N[C@H](CC(=O)O)C(C)N(C)C[C@H](N)CS	6.27
CHEMBL159	CC[C@]1(O)C[C@]2(C)[C@H]2CN(C)C3c1[nH]c4ccccc34[C@]5(C)C(=O)O)C5c6cc(cc5CO)N(C)[C@H]7(C)[C@]8(O)[C@H](O)C(=O)C[C@]9(C)C=CCN9C[C@]10(C)[C@H]8(C)C(=O)O)C1	4.60
CHEMBL2088377	CN(C(=O)CN1C(=O)O)c2ccc(cc12)c3ccc(O)cc3)c4ccccc4	6.72
CHEMBL2442937	CN(C(=O)CN1C(=O)O)c2ccc(cc12)c3ccc(O)cc3)c4ccccc4	6.16
CHEMBL421665	CN(C)C(=O)C(CCN1CC(O)CC1)c2ccc(C)cc2)C(F)F)(c3ccccc3)c4ccccc4	5.45
CHEMBL1193	CN(C)CC(C)c1ccccc1c2ccccc2	4.48
CHEMBL629	CN(C)CCC=C1c2ccccc2CCc3ccccc13	5.64
CHEMBL1628227	CN(C)CCC=C1c2ccccc2COc3ccccc13	5.59
CHEMBL549	CN(C)CCCC1(OCc2cc(ccc12)C#N)c3ccc(F)cc3	5.58
CHEMBL415	CN(C)CCCN1c2ccccc2CCc3ccc(C)cc13	6.13
CHEMBL71	CN(C)CCCN1c2ccccc2Sc3ccc(C)cc13	5.55
CHEMBL564	CN(C)CCCN1c2ccccc2Sc3ccccc13	5.59
CHEMBL55733	CN(C)CC[C@]2(C)[C@H](O)[C@]12CCCC3C[C@]1(CCC13C)OCc4ccccc4)O2	5.35
CHEMBL657	CN(C)CCOC(c1ccccc1)c2ccccc2	5.85
CHEMBL900	CN(C)CCOC(c1ccccc1)c2ccccc2C	5.60
CHEMBL1201353	CN(C)CC[C@]2(C)[C@H](c1ccc(C)cc1)c2ccccc2	4.82
CHEMBL64894	CN(C)c1ccc(cc1)C(=C2C=CC(=[N+](C)C)C=C2)c3ccc(cc3)N(C)C	6.34
CHEMBL822	CN(CC=C#CC(C)C)C)Cc1ccccc2ccccc12	5.85
CHEMBL990	CN(Cc1ccc(cc1)C(C)C)Cc2ccc3ccccc23	5.53
CHEMBL70586	CN(C[C@H]1CC[C@]2(C)C(CO)2)C[C@]3H1N3CCCC3(C(=O)Cc4ccc(C)cc1)c4	5.51
CHEMBL516	CN1CCC(=C2c3ccccc3C=C4c4ccccc24)CC1	5.93
CHEMBL395110	CN1CCCC(CN2c3ccccc3Sc4ccccc24)C1	5.60
CHEMBL1087	CN1CCCC1C(=O)Nc2c(C)cccc2C	3.72
CHEMBL1626	CN1CCC[C@]2H1CCO[C@]1(C)c2ccccc2)c3ccc(C)cc3	6.07
CHEMBL648	CN1CCN(C)C1C(c2ccccc2)c3ccccc3	5.51
CHEMBL728	CN1CCN(CCCN2c3ccccc3Sc4ccc(C)cc24)CC1	5.90
CHEMBL8437	CN1CCN2C(C1)c3ccccc3C4c4ccccc24	5.38
CHEMBL1732	CN1C[C@]2H1[C@]3H2[C@]4H1C3c[nH]c4ccccc2c34)C(=O)N(C)[C@]5(C)O[C@]6(O)C[C@]7H7CCCN7C(=O)C[C@]8(Cc9ccccc9)N6C5=O	5.30
CHEMBL19215	CN1C[C@]2H1(CNC(=O)O)C2c3ccccc3)C[C@]3H3[C@]4H1C4c4cn(C)c5ccccc3c45	6.10

MATERIAL SUPLEMENTARIO B. MATERIAL SUPLEMENTARIO

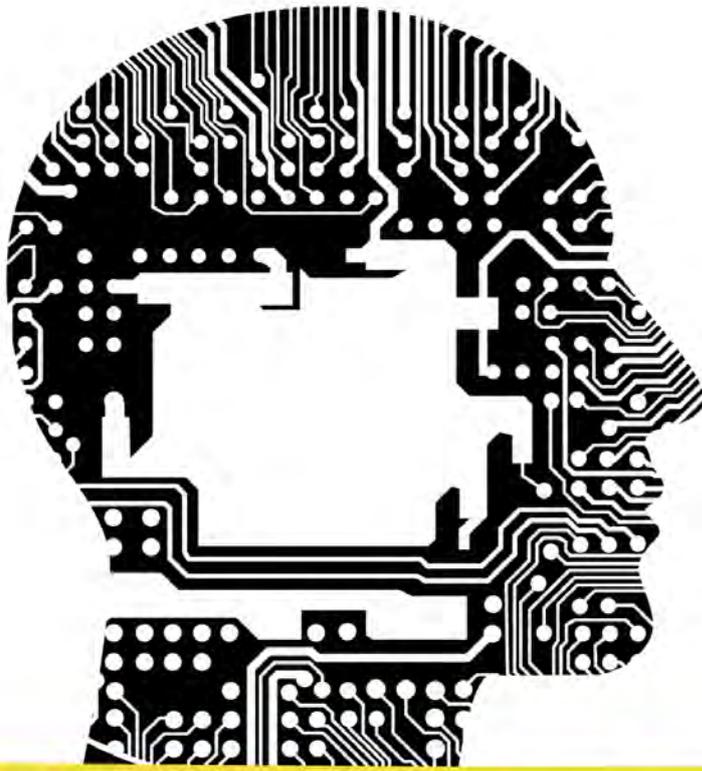
CAPÍTULO 4

CHEMBL1201203	CN1[C@@H]2CC[C@H]1C[C@H](C2)OC(c3ccccc3)c4ccccc4	5.78
CHEMBL52539	CNC(=O)[C@@H](C)N1CCC23CCCC2[C@@H]1C4ccc(OC)c5ccccc5F)c5cc34	5.28
CHEMBL85899	CNC(=O)[C@@H](CO)NC1c1ccc(OC)c2ccccc2F)c2cc1	4.75
CHEMBL311029	CNC(=O)[C@@H](CO)NC1c1ccc(OC)c2ccccc2F)c2cc1	4.58
CHEMBL41	CNCCC(OC)c1ccc(cc1)C(F)F)c2ccccc2	7.05
CHEMBL445	CNCCC=C1c2ccccc2CC3c1ccccc13	5.71
CHEMBL21731	CNCCC12CCC(c3ccccc13)c4ccccc24	5.59
CHEMBL809	CN[C@@H]1CC[C@@H](c2ccc(Cl)c(Cl)c2)c3ccccc13	5.59
CHEMBL239346	COC(=O)NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	6.55
CHEMBL370805	COC(=O)[C@@H]1[C@@H](C)[C@@H]2CC[C@H]1N2C)OC(=O)c3ccccc3	4.31
CHEMBL1729	COC1CN(CCCOC2ccc(F)cc2)CCC1NC(=O)c3ccc(Cl)c(N)cc3OC	5.96
CHEMBL296366	COC[C@@H](COC(=O)c1ccccc1)C[C@@H](O)[C@@H]23CCCC4[C@@H](CCC24C)OC)O3	5.00
CHEMBL772	CO[C@@H]1[C@@H](C)[C@@H]2CN3CCc4c([nH]c5ccc(OC)ccc45)[C@@H]3[C@@H]2[C@@H]1C(=O)OC)OC(=O)c6ccc(OC)c(OC)c(O)c6	6.11
CHEMBL52359	CO[C@@H]12CCC3(C)C(CCC[C@@H]3(O)1)[C@@H](O)C[C@@H](COC(=O)c4ccccc4)CN(C)C)C2	5.44
CHEMBL126075	COc1cc(oc(OC)c1OC)C(=O)OCCN2CCCN(CCCOC(=O)c3ccc(OC)c(OC)c(OC)c3)CC2	6.22
CHEMBL707	COc1cc2nc(nc(N)c2cc1OC)N3CCN(CCC3)C(=O)C4COc5ccccc5O4	5.96
CHEMBL2	COc1cc2nc(nc(N)c2cc1OC)N3CCN(CCC3)C(=O)c4ccccc4	4.67
CHEMBL51149	COc1ccc(CCN(C)CCCC(C#N)(C)(C)c2ccc(OC)c(OC)c2)cc1OC	5.44
CHEMBL6966	COc1ccc(CCN(C)CCCC(C#N)(C)(C)c2ccc(OC)c(OC)c2)cc1OC	5.59
CHEMBL296419	COc1ccc(CCN2CC(C)CC2)Nc3nc4ccccc4n3Cc5ccc(F)cc5cc1	7.48
CHEMBL511	COc1ccc(CN(CCN(C)C)c2ccccc2)cc1	5.28
CHEMBL573962	COc1ccc(CN2C(=NC(=NC2(C)C)N)N)cc1	4.22
CHEMBL64767	COc1ccc(CNC(=O)C(N2CCN(C)C)OCc3c(C)cccc3C)CC2)c4ccc(C)cc4cc1	6.46
CHEMBL64114	COc1ccc(CNC(=O)C(N2CCN(C)C)OCc3c(C)cccc3C)CC2)c4ccc(F)cc4cc1	6.46
CHEMBL61501	COc1ccc(CNC(=O)C(N2CCN(C)C)OCc3c(C)cccc3C)CC2)c4ccc(OC)cc4cc1	6.24
CHEMBL279835	COc1ccc(CNC(=O)C(N2CCN(C)C)OCc3c(C)cccc3C)CC2)c4ccc(cc4)C(F)F)cc1	6.45
CHEMBL64505	COc1ccc(CNC(=O)C(N2CCN(C)C)OCc3ccc(OC)cc3)CC2)c4ccc(F)cc4cc1	6.67
CHEMBL60622	COc1ccc(CNC(=O)C(N2CCN(C)C)OCc3ccc(OC)cc3)CC2)c4ccc(cc4)C(F)F)cc1	6.45
CHEMBL64583	COc1ccc(OC)c(c1)S(=O)(=O)NCC(N2CCN(C)C)OCc3ccc(Br)cc3)CC2)c4ccccc4	6.28
CHEMBL416848	COc1ccc(OC)C(O)CN2CCC(C)N(C(=O)c3ccc(Cl)cc3)c4ccc(F)cc4cc1	6.67
CHEMBL64018	COc1ccc(OC)C(O)CN2CCC(C)N(C(=O)c3ccc(Cl)cc3)c4ccc(F)cc4cc1	6.64
CHEMBL433100	COc1ccc(OC)C(O)CN2CCC(C)N(C(=O)c3ccc(OC)cc3)c4ccc(F)cc4cc1	6.72
CHEMBL59712	COc1ccc(OC)C(O)CN2CCC(C)N(C(=O)c3ccc(OC)cc3)c4ccccc4cc1	6.74
CHEMBL63060	COc1ccc(OC)C(O)CN2CCC(C)N(C(=O)c3ccc(Cl)cc3)c4ccc(F)cc4cc1	6.71
CHEMBL62240	COc1ccc(OC)C(O)CN2CCC(C)N(C(=O)c3ccc(Cl)cc3)c4ccccc4cc1	6.58
CHEMBL291629	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccc(C)cc4cc1	6.24
CHEMBL60847	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccc(F)cc4cc1	6.59
CHEMBL59644	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccc(cc4)C(F)F)cc1	6.37
CHEMBL64491	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccc(F)cc4cc1	6.28
CHEMBL64831	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccc(F)cc4cc1	6.06
CHEMBL61336	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccc(F)cc4cc1	6.60
CHEMBL64177	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3)c4ccccc4cc1	6.13
CHEMBL64688	COc1ccc(OC)C(O)CN2CCN(C)C(=O)NCC3ccccc3OC)c4ccccc4cc1	6.14
CHEMBL61283	COc1ccc(cc1)C(=O)NCC(N2CCN(C)C)OCc3ccc(Cl)cc3)CC2)c4ccccc4	6.27
CHEMBL583757	COc1ccc(cc1)N2C(=NC(=NC2(C)C)N)N	5.11
CHEMBL439496	COc1ccc(cc1)OC)C(=O)C[C@@H]2[C@@H]3[C@@H]4[C@@H]5[C@@H]6(C)C[C@@H](O)C[C@@H]7(O)C[C@@H](CN8C[C@@H](C)CC[C@@H]8)C[C@@H]7(O)C[C@@H](O)C[C@@H]35O[C@@H]24O	4.66
CHEMBL22207	COc1ccc2CC3C4CCCC4(CCN3C)c2c1	5.89
CHEMBL52440	COc1ccc2C[C@@H]3[C@@H]4[C@@H]5CCCC[C@@H]4(CCN3C)c2c1	5.89
CHEMBL1201151	COc1ccc2[C@@H]3C[C@@H]4[C@@H]5[C@@H]6(C)C[C@@H](C)C[C@@H]7(O)C[C@@H]8(O)C[C@@H]9(O)C[C@@H]35O[C@@H]24O	5.25
CHEMBL573445	COc1ccc(NC2=NC(=NC(C)N2)N)c1	4.12
CHEMBL575299	COc1ccc(cc1)N2C(=NC(=NC2(C)C)N)N	5.29
CHEMBL53463	COc1ccc2C(=O)c3c(O)c4C[C@@H](O)C[C@@H](O)C[C@@H]5[C@@H](N)[C@@H](O)C[C@@H](C)O5)c4c(O)c3C(=O)c12)C(=O)CO	4.82
CHEMBL60666	COc1ccc(cc1)C(=O)NCC(N2CCN(C)C)OCc3ccc(Cl)cc3)CC2)c4ccccc4	6.44
CHEMBL573514	COc1ccc(cc1)N2C(=NC(=NC2(C)C)N)N	4.91
CHEMBL723	COc1ccc(cc1)OCCN(C)C)OCc2ccc3[nH]c4ccccc4c23	6.34
CHEMBL479	CSc1ccc2Sc3ccccc3N(CCC4CCCC4)C2c1	5.69
CHEMBL51777	C[C@@H](N)1CCC23CCCC2[C@@H]1C4ccc(OC)c5ccccc5F)c5cc34)C(=O)N	5.58
CHEMBL315738	C[C@@H](N)1C(Nc1ccc(NC2c1ccccc2F)cc1)C(=O)N	4.95
CHEMBL85962	C[C@@H](N)1C(Nc1ccc(OC)C2c1ccccc2)cc1)C(=O)N	5.11
CHEMBL82327	C[C@@H](N)1C(Nc1ccc(OC)c2ccc(F)cc2)cc1)C(=O)N	5.07
CHEMBL1630578	C[C@@H]1[C@@H]2[C@@H]3[C@@H]4[C@@H]5C4=CC(=O)C=C[C@@H]2(C)C[C@@H]1(C)O)N5CCN(C)C5)c6ccc(nc6)n7N8CCCC7N8CCCC8	5.87
CHEMBL83621	C[C@@H](N)1C(Nc1ccc(NC2c1ccccc2F)cc1)C(=O)N	4.86
CHEMBL84167	C[C@@H](N)1C(Nc1ccc(OC)C2c1ccccc2)cc1)C(=O)N	5.85
CHEMBL85054	C[C@@H](N)1C(Nc1ccc(OC)C2c1ccccc2)cc1)C(=O)N	5.96
CHEMBL85371	C[C@@H](N)1C(Nc1ccc(OC)C2c1ccccc2)cc1)C(=O)N	4.88
CHEMBL48582	C[C@@H](N)1C(Nc1ccc(OC)c2ccc(F)cc2)cc1)C(=O)N	5.09
CHEMBL717	C[C@@H]1[C@@H]2[C@@H]3[C@@H]4[C@@H]5C4=CC(=O)C=C[C@@H]2(C)C[C@@H]1(C)O)N5CCN(C)C5)c6ccc(nc6)n7N8CCCC7N8CCCC8	4.83
CHEMBL1162	C[C@@H]12CC[C@@H]3[C@@H]4[C@@H]5CCCC[C@@H]4(C)C[C@@H]3(C)C(=O)C)C(=O)C)C[C@@H]4(C)CCC(=O)C=C14	4.81
CHEMBL691	C[C@@H]12CC[C@@H]3[C@@H]4[C@@H]5CCCC[C@@H]4(C)C[C@@H]3(C)C(=O)C)C(=O)C)C[C@@H]4(C)CCC(=O)C=C14	4.65
CHEMBL282575	C[C@@H]12CC[C@@H]3[C@@H]4[C@@H]5CCCC[C@@H]4(C)C[C@@H]3(C)C(=O)C)C(=O)C)C[C@@H]4(C)CCC(=O)C=C14	5.48
CHEMBL1479	C[C@@H]12CC3c1ccc3C=C1CC[C@@H]4[C@@H]5C4=CC(=O)C=C[C@@H]2(C)C[C@@H]1(C)O)N5CCN(C)C5)c6ccc(nc6)n7N8CCCC7N8CCCC8	5.53
CHEMBL127734	C[N+](C)(OC(=O)N)C1c1ccccc1)CC(=O)N2c1ccccc2	5.41
CHEMBL141343	Cc1c(C)c2OC(C)CN3CCN(C)C4c1cc(nc4)N5CCCC5)N6CCCC6)CCc2c(C)C1O	5.45
CHEMBL584	Cc1c(C)ccc1C(=O)N[C@@H](CS2c1ccccc2)C[C@@H](O)CN3C[C@@H]4CCCC[C@@H]4(C)C[C@@H]3(C)C(=O)N(C)C)C)C	5.73
CHEMBL762	Cc1ccc(O)c(C)c1CC2=NCN2(C)C)C)C	5.46
CHEMBL220709	Cc1ccc(O)c(C)c1CC2=NCN2(C)C)C)C	5.43
CHEMBL357995	Cc1cc2c(OC)C(C)C)OC(=O)c3ccccc3)ccccc2[nH]1	6.22
CHEMBL582969	Cc1ccc(CN2C(=NC(=NC2(C)C)N)N)cc1	4.49
CHEMBL583174	Cc1ccc(NC2=NC(=NC(C)N2)N)cc1	4.15
CHEMBL60568	Cc1ccc(cc1)C(N2CCN(C)C)OCc3c(C)cccc3C)CC2)C(=O)NCC4c1ccccc4	6.49
CHEMBL573513	Cc1ccc(cc1)N2C(=NC(=NC2(C)C)N)N	5.38
CHEMBL118	Cc1ccc(cc1)c2ccc(nc23ccc(cc3)S(=O)(=O)N)C(F)F)F	6.77
CHEMBL3098431	Cc1ccc(cc1)C)C(=O)CCN2CCN(C)C2)C(c3ccccc3)c4ccccc4	4.81
CHEMBL3098432	Cc1ccc2c(ccc2c1)S(=O)(=O)NCCCCCN	4.92
CHEMBL305691	Cc1ccc(O)c1OCC(O)CN2CCN(C)C2)C(=O)NCC3ccccc3)c4ccc(F)cc4	6.83
CHEMBL60843	Cc1ccc(O)c1OCC(O)CN2CCN(C)C2)C(=O)NCC3ccccc3)c4ccc(cc4)C(F)F)F	6.53
CHEMBL575703	Cc1ccc(O)c1N2C(=NC(=NC2(C)C)N)N	5.33
CHEMBL574657	Cc1ccc(O)c1N2C(=NC(=NC2(C)C)N)N	5.04
CHEMBL330800	Cc1nc(c[nH]1)c2ccc(cc2)c3ccccc3	5.84
CHEMBL91	Cc1ccc(COC(Cn2ccn2)c3ccc(Cl)cc3Cl)c(Cl)c1	5.51
CHEMBL808	Cc1ccc(COC(Cn2ccn2)c3ccc(Cl)cc3Cl)c(Cl)c1	5.46
CHEMBL1262	Cc1ccc(CON=C(Cn2ccn2)c3ccc(Cl)cc3Cl)c(Cl)c1	5.83
CHEMBL1221	Cc1ccc(CSC(Cn2ccn2)c3ccc(Cl)cc3Cl)c1	5.70
CHEMBL1208858	Cc1ccc(cc1)C(c2ccc(Cl)cc2)[n+]3ccn(C)OC4ccc(Cl)cc4C)c5ccc(Cl)cc5C)c3	5.15

MATERIAL SUPLEMENTARIO B. MATERIAL SUPLEMENTARIO

CAPÍTULO 4

CHEMBL219916	Clc1ccc2N(C3CCN(CCCN4C(=O)Nc5ccccc45)CC3)C(=O)Nc2c1	5.99
CHEMBL1254129	Clc1ccc(c1)N2CCN(CN3C(=O)CC4(CCc5ccccc5C4)C3=O)CC2	5.61
CHEMBL104	Clc1ccccc1C(c2ccccc2)(c3ccccc3)n4cnc4	5.17
CHEMBL1254220	FC(F)(F)c1ccc(c1)N2CCN(CN3C(=O)CC4(CCc5ccccc5C4)C3=O)CC2	5.03
CHEMBL119385	Fc1ccc(SC2=NN3C=NC(=O)C(=C3C=C2)c4c(C)cccc4Cl)c(F)c1	4.58
CHEMBL267930	Fc1ccc(cc1)C(=O)CCCN2CCCC3(CC2)N(CNC3=O)c4ccccc4	6.22
CHEMBL3098419	Fc1ccc(cc1)C(CCCN2CCN(CC2)C(c3ccccc3)c4ccccc4)c5ccc(F)cc5	6.92
CHEMBL3098429	Fc1ccc(cc1)C(CCCNCCc2ccccc2)c3ccc(F)cc3	6.84
CHEMBL3098430	Fc1ccc(cc1)C(CCCNCCc2ccccc2)c3ccc(F)cc3	5.91
CHEMBL30008	Fc1ccc(cc1)C(N2CCN(CC=Cc3ccccc3)CC2)c4ccc(F)cc4	6.40
CHEMBL490	Fc1ccc(cc1)C@H2CCN(C@H)2COc3ccc4OCOc4c3	5.57
CHEMBL418863	Fc1ccc(cc1)c2c[nH]c(C3CC3CC3)cn2	6.84
CHEMBL77675	N=C(Nc1ccccc1)Nc2ccccc2	5.88
CHEMBL420	NC(=N)NN=Cc1c(Cl)cccc1Cl	5.45
CHEMBL1909072	NC(=O)C1CCN(CCCN2c3ccccc3Sc4ccc(Cl)cc24)CC1	5.57
CHEMBL574656	NC1=NC2(CCCCC2)N(C(=N1)N)c3ccc(Cl)cc3	5.36
CHEMBL582902	NC1=NC2(CCCCC2)N(C(=N1)N)c3ccc(Cl)c3	5.40
CHEMBL574411	NC1=NC2(CCCCC2)N(C(=N1)N)c3ccccc3	4.83
CHEMBL575917	NC1=NC2(CCCCC2)N(C(=N1)N)c3ccccc3Cl	4.89
CHEMBL1188431	NCCCCCNS(=O)(=O)c1ccc2ccccc12	4.66
CHEMBL741	Nc1nnc(c1N)n1c2ccccc(Cl)c2Cl	4.81
CHEMBL3414957	O=C(CN1CCCC1)N2CCN(CC2)C3(C4CC5CC(C3C5)C4)c6ccccc6	5.61
CHEMBL3098134	O=C(NCCN1CCN(CC1)C(c2ccccc2)c3ccccc3)c4ccccc4	6.55
CHEMBL239979	O=C(NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3)OC4CCCC4	7.28
CHEMBL239766	O=C(NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3)OC4CCCC4	7.58
CHEMBL239980	O=C(NCCc1nc(c[nH]1)c2ccc(cc2)c3ccccc3)OC4CCCC4	7.68
CHEMBL39326	O=C1NC(=O)C(N1)(C2CCCC2)c3ccccc3	4.24
CHEMBL290090	O=C1NC(=O)C2(CCCCC2c3ccccc23)N1	3.60
CHEMBL39377	O=C1NC(=O)C2(CCCCC2c3ccccc23)N1	3.60
CHEMBL289046	O=C1NC(=O)C2(CCCCC2c3ccccc23)N1	3.07
CHEMBL289431	O=C1NC(=O)C2(CCCc3ccccc23)N1	2.68
CHEMBL225189	O=C1NC(C(=O)N1CN2CCN(CC2)C(c3ccccc3)c4ccccc4)(c5ccccc5)c6ccccc6	6.66
CHEMBL222210	O=C1NC(C(=O)N1CN2CCN(CC2)C(c3ccccc3)c4ccccc4)(c5ccccc5)c6ccccc6	5.35
CHEMBL1254381	O=C1NC2(CCCc3ccccc3C2)C(=O)N1CN4CCN(CC4)c5ccccc5	5.03
CHEMBL139	OC(=O)Cc1ccccc1Nc2c(Cl)cccc2Cl	4.28
CHEMBL936	OC(CCN1CCCC1)(c2ccccc2)c3ccccc3	5.47
CHEMBL1490	OC(CCN1CCCC1)(C2CCGCC2)c3ccccc3	5.55
CHEMBL59639	OC(COCC=C)CN1CCN(CC1)C(c2ccccc2)c3ccc(Cl)cc3	6.40
CHEMBL62878	OC(COc1ccc(Cl)cc1)CN2CCN(CC2)C(CNC(=O)c3ccc(cc3)C(F)(F)F)c4ccccc4	6.68
CHEMBL60271	OC(COc1ccccc1)CN2CCN(CC2)C(CNC(=O)c3ccc(F)cc3)c4ccccc4	6.44
CHEMBL60256	OC(COc1ccccc1)CN2CCN(CC2)C(c3ccccc3)c4ccc(cc4)C(F)(F)F	6.05
CHEMBL16	OC1=NC(=O)C(N1)(c2ccccc2)c3ccccc3	4.40
CHEMBL726	OCN1CCN(CCCN2c3ccccc3Sc4ccc(cc24)C(F)(F)F)CC1	5.50
CHEMBL81	OC1ccc(cc1)c2sc3cc(O)ccc3c2C(=O)c4ccc(OCCN5CCCC5)cc4	6.28



Material suplementario

Algoritmos

C.1. Silhouette Optimized Molecular Clustering (SOMoC)

Con el objetivo de poder agrupar las moléculas según su similaridad química o estructural, se desarrolló una metodología propia de agrupamiento o *clustering*, denominada Silhouette Optimized Molecular Clustering (SOMoC). La misma fue desarrollada íntegramente en el lenguaje de programación Python y luego implementada en forma de [Web App](#), a través de la plataforma Streamlit, lo que permite a cualquier usuario utilizarla a través de un interfaz sencilla e intuitiva. También se encuentra disponible una versión *standalone*, para aquellos usuarios que requieran correr el algoritmo de manera local. Esta aplicación forma parte del paquete [LIDeBTools](#), desarrollado y mantenido por nuestro laboratorio, el cual es de acceso libre y gratuito.

En una evaluación comparativa o *benchmarking* sobre 29 *datasets* de moléculas pequeñas, SOMoC e iRaPCA (otra metodología de *clustering* desarrollada en nuestro laboratorio), demostraron ser superiores a otros algoritmos clásicos como Ward ([Ward 1963](#)) o Butina ([Butina 1999](#)), generando *clusters* más compactos y distantes entre sí. Los detalles sobre el desarrollo, implementación y validación de estas metodologías pueden encontrarse en la publicación correspondiente ([Prada Gori et al. 2022](#)).

La metodología de agrupamiento SOMoC puede describirse como la combinación secuencial de tres etapas fundamentales: **descripción**, **reducción** y **clustering** (Figura [C.1](#)).

La **descripción** implica representar o codificar la información química/estructural contenida en cada molécula, de forma que pueda ser procesada por un algoritmo de agrupamiento. Para ello, se calculan *fingerprints* moleculares de tipo EState1 ([Hall y Kier 1995](#)), para todas las moléculas del conjunto de datos. Previo al cálculo de los *fingerprints*, el usuario puede optar por estandarizar las moléculas utilizando la librería de Python [MolVS](#).

En líneas generales, los algoritmos de *clustering* no funcionan bien en espacios hiperdimensionales, donde la relación ejemplos/predictores es muy baja. Este

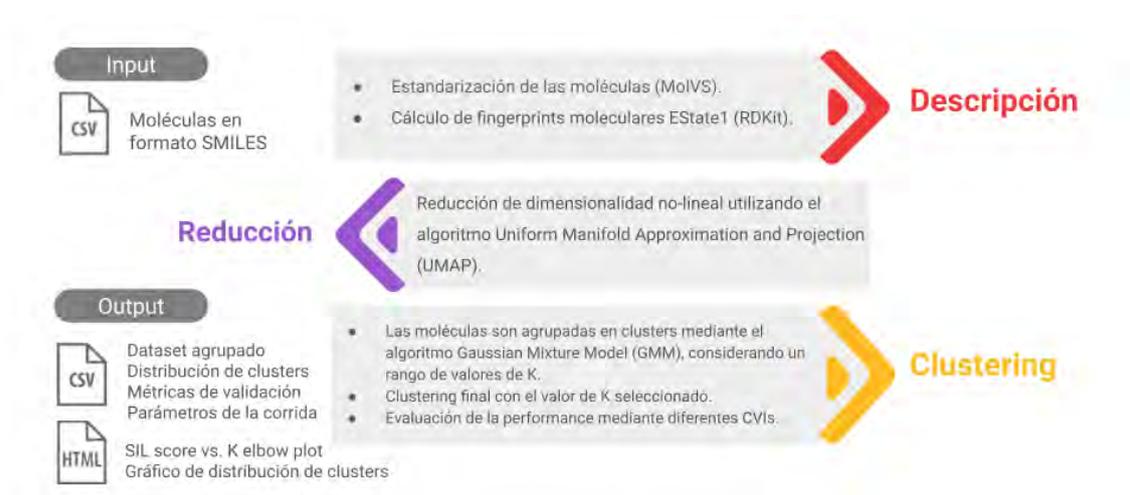


Figura C.1: Esquema general donde se describen las distintas etapas del método SOMoC. Reproducido y adaptado con autorización de **Prada Gori et al. 2022**. Copyright 2022 American Chemical Society.

fenómeno se conoce como *the curse of dimensionality* o “la maldición de la dimensionalidad” (**Bellman 1966**). Por lo tanto, la segunda etapa implica la **reducción** de dimensionalidad del conjunto de datos, previo a la etapa de agrupamiento propiamente dicho, utilizando para ello el algoritmo Uniform Manifold Approximation and Projection (UMAP) (**McInnes, Healy y Melville 2018**).

UMAP es una técnica de reducción de dimensionalidad no lineal, que intenta proyectar los datos desde un espacio hiperdimensional hacia uno más simple, manteniendo la estructura topológica original de los mismos y reteniendo sólo la información más importante en este sentido. Este algoritmo se basa en tres asunciones básicas:

- Los datos se encuentran uniformemente distribuidos sobre una variedad de Riemann o *Riemannian manifold*.
- La métrica de Riemann es localmente constante, o puede ser aproximada como tal.
- La variedad o *manifold* se encuentra localmente conectada.

A partir de estas asunciones, los datos son modelados con una estructura topológica difusa, para luego encontrar una proyección de los mismos en un espacio dimensional más simple, pero manteniendo una estructura topológica equivalente.

Finalmente, durante la etapa final de *clustering* se utilizó el algoritmo Gaussian Mixture Models (GMM) para agrupar las moléculas. Una *gaussian mixture* o mezcla gaussiana es una sumatoria de K distribuciones gaussianas, donde cada una de estas distribuciones se caracteriza por:

- Una media μ que define su centro.
- Una covarianza Z que define el ancho de la distribución.
- Una probabilidad de mezcla π que determina cuán grande será la distribución. La sumatoria de las probabilidades dentro de una mezcla gaussiana debe ser igual a 1.

Cada gaussiana explica los datos contenidos en un *cluster*, o dicho de otra forma, cada *cluster* es representado por una distribución gaussiana particular. A diferencia de otros algoritmos como K-Means, que sólo permite asignar un *cluster* a cada ejemplo, GMM es un método de *soft clustering*, es decir que asigna a cada ejemplo un grado de pertenencia a cada uno de los *clusters* identificados.

Durante la inicialización, los componentes o gaussianas se centran en ciertos puntos dentro del conjunto de datos, generados en nuestro caso utilizando el método de K-Means, aunque también puede inicializarse de manera aleatoria o uniforme. Luego, se calcula para cada punto del conjunto de datos la probabilidad de que ese punto haya sido generado por cada uno de los K componentes del modelo. Con el objetivo de encontrar los valores óptimos para los parámetros μ , Z y π que caracterizan cada distribución, se emplea un algoritmo denominado Expectation Maximization (EM).

El número de componentes gaussianas o *clusters* (K) debe establecerse de antemano. Existen diversos métodos para seleccionar el valor óptimo de K, siendo el criterio del hombro o *elbow criterion* uno de los más utilizados por su simpleza y carácter intuitivo (Bholowalia y Kumar 2014). Evaluando secuencialmente distintos valores de K y calculando para cada valor la proporción de varianza que explica el modelo, es posible encontrar un punto de inflexión o quiebre en la curva, donde la adición de más *clusters* no incrementa sustancialmente la proporción de varianza asociada al modelo. Como medida de varianza comúnmente se considera la relación

entre la varianza entre grupos con respecto a la varianza total, aunque también puede utilizarse el estadístico F, que considera la relación de la varianza entre grupos con respecto a la intra-grupo.

En este caso, se aplicó una variante del método de *elbow* para optimizar K, donde se utilizó el índice de Silhouette (SIL) (**Rousseeuw 1987**) como métrica para evaluar la calidad o *performance* del *clustering*. El coeficiente de Silhouette (S) para cada dato dentro de un *cluster* puede definirse como:

$$\mathbf{S} = \frac{b - a}{\max(a, b)}$$

Donde **a** es la distancia media *intra-cluster*, y **b** es la distancia media *inter-clusters*. **S** puede tomar valores entre 1 y -1, donde 1 implica un *clustering* perfecto, mientras que un valor de -1 implica que ese dato debería ser categorizado dentro de otro *cluster*. Luego, el SIL se calcula como el promedio de **S** para todo el *dataset*.

SOMoC evalúa secuencialmente un rango de valores de K (establecido por el usuario), calculando para cada uno de ellos el SIL. Por defecto, el algoritmo utiliza el K que genera el mayor SIL para el *clustering* final de las moléculas. Sin embargo, también se genera un *elbow plot* de K vs. SIL que permite al usuario identificar gráficamente un punto de inflexión en la curva y elegir así un valor de K diferente, el que puede utilizarse luego para volver a agrupar las moléculas.

Cabe destacar que existen diversas métricas para evaluar la calidad de un agrupamiento cuando no se conoce la verdadera pertenencia de cada ejemplo o *ground truth*, denominadas índices de validez de *cluster* (CVIs, del inglés cluster validity indexes). El concepto subyacente en todas ellas es que la calidad del *clustering* será mayor cuanto menor sea la varianza *intra-cluster* (más compacto) y mayor sea la distancia entre *inter-clusters* (más separados). Algunos ejemplos incluyen el Davies Bouldin score (DB) (**Davies y Bouldin 1979**), el Calinski Harabasz score (CH) (**Caliński y Harabasz 1974**) o el índice Dunn (DUNN) (**Dunn 1973**). En la Tabla **C.1** se muestran los CVIs antes mencionados con sus respectivos rangos y valores óptimos.

Tabla C.1: CVIs calculados por SOMoC durante la validación del *clustering*, acompañados de sus valores mínimos, máximos y óptimos. Reproducido y adaptado con autorización de **Prada Gori et al. 2022**. Copyright 2022 American Chemical Society.

CVI	Mínimo	Máximo	Óptimo
Dunn	0	infinito	Los valores más altos denotan un mejor <i>clustering</i>
DB	0	infinito	Los valores menores denotan un mejor <i>clustering</i>
CH	0	infinito	Los valores más altos denotan un mejor <i>clustering</i>
SIL	-1	1	1 representa un <i>clustering</i> ideal

Comparativamente, el SIL ha demostrado ser altamente efectivo en la selección del número de *clusters*, evaluado sobre una serie de algoritmos de *clustering* diferentes (**Arbelaitz et al. 2013**). Sin embargo, el análisis conjunto de distintos CVIs permite hacer una evaluación más robusta de la *performance* del *clustering* (**Estivill-Castro 2002**). Por lo tanto, una vez determinado el valor óptimo de K, SOMoC calcula los CVIs antes mencionados y devuelve una tabla con los resultados numéricos.

C.2. Métodos de optimización

C.2.1. Optimización bayesiana

La optimización bayesiana (BO, del inglés Bayesian Optimization) es una técnica de optimización, que puede ser entendida como un caso particular de Sequential model-based optimization (SMBO) (**Hutter, Hoos y Leyton-Brown 2011**). En los algoritmos de búsqueda tradicionales, como *random search* o *grid search*, el resultado de las iteraciones pasadas no influye sobre las iteraciones futuras, por lo que se los denomina algoritmos de búsqueda no informados. En los algoritmos de búsqueda informados, como BO, los resultados de las iteraciones pasadas son utilizados para construir, de manera secuencial, un modelo que propone cuáles serán los próximos hiperparámetros a ser evaluados, minimizando alguna función de pérdida.

Esta técnica cuenta con dos componentes principales:

- Un modelo subrogado $\mathbf{F}(\mathbf{x})$ que aproxima la función a optimizar $\mathbf{f}(\mathbf{x})$.
- Una función de adquisición (FA), que será la encargada de proponer los nuevos puntos a evaluar (\mathbf{x}), dada una distribución condicional de la función $\mathbf{f}(\mathbf{x})$

Comúnmente, se utiliza un proceso gaussiano como *prior*, o distribución de probabilidad *a priori*, para aproximar la distribución de probabilidades que tendrá *a posteriori* la función $\mathbf{f}(\mathbf{x})$, y construir así el modelo subrogado. Sin embargo, también existen otras opciones como el Tree-of-Parzen-Estimators (TPE) (J. S. **Bergstra** et al. 2011). Evaluar este modelo subrogado $\mathbf{F}(\mathbf{x})$ es menos costoso que evaluar la función original, por lo que el algoritmo requiere un tiempo extra entre evaluaciones, pero se alcanza el objetivo con un número menor de ellas (J. **Bergstra**, **Yamins** y **Cox** 2013; **Snoek**, **Larochelle** y **Adams** 2012). Es decir, el algoritmo pasa más tiempo explorando las regiones del hiperespacio que son más promisorias.

Con respecto a la FA, esta es calculada a partir del modelo subrogado y se utiliza para decidir cuáles serán los próximos hiperparámetros a muestrear, en base a la probabilidad de que estos generen una solución superadora con respecto a la mejor solución encontrada hasta el momento. Las más utilizadas son el Expected Improvement Criterion (EI) y Upper confidence Bound (UCB) (**Shahriari** et al. 2016). A su vez, esta función será la encargada de balancear la relación de compromiso *exploitation/exploration*, es decir, cuanto se seguirá “explotando” o aprovechando una región del hiperespacio que generó buenos resultados (*exploitation*), en relación a cuanto se exploran otras regiones donde la incertidumbre es mayor (*exploration*). De esta manera, se evita que el algoritmo se estanque en una región del hiperespacio que se corresponde con un mínimo local y no global.

A modo de ejemplo, en la Figura C.2 se esquematiza el proceso de BO para una función simple del tipo $y = f(x)$. Inicialmente el algoritmo evalúa ciertos valores de \mathbf{x} elegidos de modo aleatorio, con los cuales construye el primer modelo subrogado $\mathbf{F}(\mathbf{x})$. Luego, a través de la FA se propone el próximo valor de \mathbf{x} que tendrá la mayor probabilidad de generar un valor de y menor que el valor más pequeño encontrado hasta el momento. El algoritmo evalúa este valor de \mathbf{x} , y luego actualiza el modelo subrogado con esta nueva información. De este modo, con las sucesivas iteraciones

el modelo subrogado $\mathbf{F}(\mathbf{x})$ aproxima cada vez mejor a la función original, generando predicciones cada vez más confiables. De hecho, en las regiones cercanas a donde previamente se evaluaron puntos, la varianza tiende a ser 0 (Figura C.2).

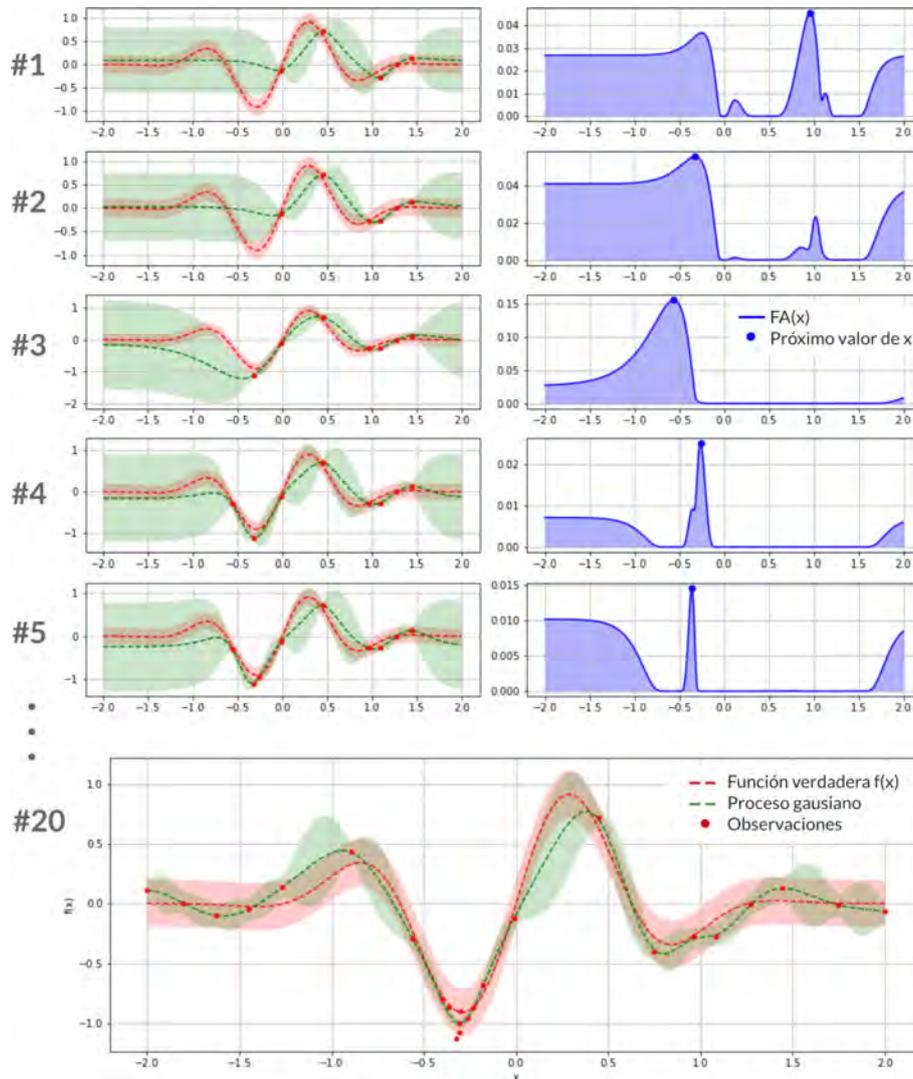


Figura C.2: Esquematización de la BO para una función $y = f(x)$ utilizando un proceso gaussiano. Los primeros 5 valores de \mathbf{x} se muestrean de manera aleatoria. En la figura se muestran las primeras 5 iteraciones del algoritmo, y el resultado final luego de 20 iteraciones. En las gráficas de la primer columna, la línea punteada roja representa el valor verdadero de la función $\mathbf{f}(\mathbf{x})$, el cual en principio se desconoce. La línea punteada verde representa la media estimada por el proceso gaussiano. Los sombreados de las curvas representan la varianza asociada, mientras que los puntos rojos son las observaciones o valores de \mathbf{x} evaluados. En las gráficas de la derecha, se observan los resultados de la FA, donde cada punto azul representa el próximo valor de \mathbf{x} a ser muestreado según esta función. Finalmente, en la última gráfica se muestra el resultado al concluir las 20 iteraciones.

C.2.2. Algoritmos genéticos

Un algoritmo genético es una metaheurística de tipo evolutiva que permite resolver problemas de búsqueda y optimización de manera eficiente, obteniendo una solución satisfactoria a un problema complejo, cuya solución óptima implicaría un alto costo computacional o en muchos casos sería prácticamente inalcanzable. Es decir, nos permite obtener una buena solución a un problema, incluso cuando esta no sea la mejor solución posible (Mirjalili 2019). En el campo del aprendizaje automático, y en particular de la quimioinformática, pueden ser utilizados como método de optimización de los hiperparámetros de un algoritmo (Wicaksono y Afif 2018), o también como método de selección de predictores a partir de un conjunto inicial, con el objetivo de reducir la dimensionalidad de los datos a modelar (Fang et al. 2011).

Estos algoritmos se basan en el principio de evolución natural de las especies, o supervivencia del más apto. En una población determinada, sólo los especímenes más aptos podrán reproducirse y así transferir su información genética a las futuras generaciones. De esta manera, el genotipo de los mejores individuos será cada vez más prevalente en las siguientes generaciones. Este proceso evolutivo es afectado por diversos operadores genéticos, como el *crossing-over* y las mutaciones, que actúan sobre la información genética de los individuos involucrados. El *crossing-over* permite un entrecruzamiento de la información genética contenida en cromosomas homólogos provenientes de ambos progenitores, mientras que las mutaciones son cambios aleatorios en la composición genética de un cromosoma (Sivanandam y Deepa 2008). En el esquema de la Figura C.3 se detallan las etapas principales de un algoritmo genético.



Figura C.3: Esquema de las etapas fundamentales de un algoritmo genético.

La **inicialización** supone la creación de una población inicial de individuos

sobre la cual comenzará a operar el algoritmo. La **evaluación** o cálculo de *fitness*, permite definir que tan buena es cada una de las soluciones o individuos, y de esta manera seleccionar solo a los mejores candidatos para generar descendencia.

El operador genético de ***crossing-over*** permite generar nuevos individuos que combinen la información contenida en cada uno de los progenitores, mientras que las **mutaciones** puntuales operan sobre la descendencia y permiten introducir variabilidad en la misma. El algoritmo finaliza cuando se alcanza el número máximo de generaciones o cuando se cumple el criterio de **finalización**, que se define comúnmente como un cierto número de generaciones sin mejora apreciable en el valor de *fitness*.

C.3. Algoritmos de aprendizaje automático

El aprendizaje automático (ML, del inglés *Machine Learning*) es una rama de la inteligencia artificial (AI, del inglés Artificial Intelligence) que se enfoca en desarrollar técnicas o algoritmos que permiten a una computadora “aprender”, sin haber sido programada explícitamente para ello. Es decir, el programa observa datos tratando de identificar patrones en ellos, construye un modelo basado en estos datos, y luego utiliza este modelo para realizar predicciones sobre nuevos datos (**Russell y Norvig 2010**).

Según el tipo de datos que emplean, los algoritmos de ML pueden ser clasificados como supervisados o no supervisados. Mientras que los métodos supervisados requieren de datos etiquetados (por ejemplo activo/inactivo), los no supervisados son capaces de encontrar patrones en los datos sin etiqueta alguna (un ejemplo clásico son los algoritmos de *clustering*) (**Alloghani et al. 2020**).

A continuación se describen los algoritmos de ML que fueron utilizados para el desarrollo de los modelos QSAR de los canales TRPV1 y NaV1.2.

C.3.1. K-nearest neighbors

El algoritmo de K-vecinos más cercanos (KNN, del inglés K-nearest neighbors) es un algoritmo supervisado no paramétrico, que permite clasificar a los ejemplos

en base a la clase de sus vecinos más cercanos ([Kramer 2013](#)).

A cada compuesto del conjunto de entrenamiento, se le asignará la clase más frecuente entre sus K vecinos más cercanos (Figura C.4). Existen distintas funciones de distancia para medir la cercanía entre los ejemplos en un espacio hiperdimensional, siendo la distancia euclidiana la más común. En este caso, los descriptores deben ser normalizados antes de poder compararlos entre sí. Naturalmente, la dimensionalidad de este hiperespacio vendrá dada por el número de descriptores (p). La contribución de estos vecinos a la predicción final puede ser uniforme (*uniform*) o es posible utilizar una función de pesos, donde los vecinos más cercanos contribuirán a la predicción final más que aquellos lejanos (*distance*).

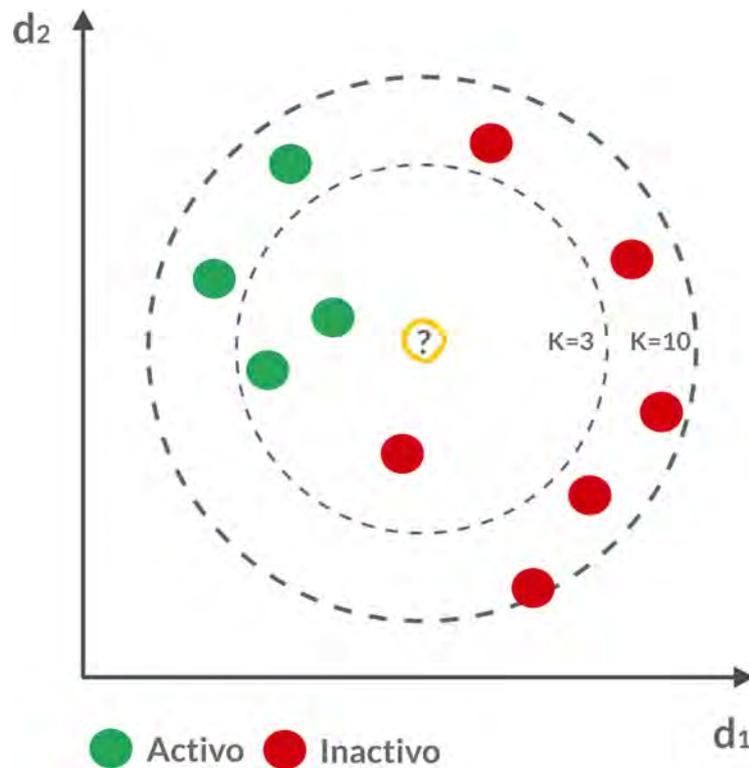


Figura C.4: Representación del funcionamiento del algoritmo KNN. En el diagrama se esquematiza un problema de clasificación binaria, en un espacio bidimensional ($p=2$) delimitado por los descriptores d_1 y d_2 . Para un $K=3$ el ejemplo (?) será clasificado como Activo, mientras que si se considera un $K=10$ el mismo será clasificado como Inactivo.

También es posible aplicar este algoritmo a problemas de regresión, en cuyo caso la predicción para un ejemplo particular será la media de sus K vecinos más cercanos ([Altman 1992](#)).

C.3.2. ExtraTrees

ExtraTrees, también conocido como árboles extremadamente aleatorizados o *Extremely Randomized Trees*, es un algoritmo de aprendizaje automático supervisado. Este meta-predicador, combina las predicciones de distintos árboles de decisión, con el objetivo de mejorar su capacidad predictiva y minimizar el sobreajuste o *overfitting*. Por lo tanto, comenzaremos por explicar estos últimos.

Un árbol de decisión es un algoritmo predictivo que divide el espacio de las variables independientes (predictores), agrupando las observaciones con valores similares para la variable dependiente (respuesta). Aplicando una serie de reglas o decisiones, el algoritmo es capaz de dividir el espacio muestral en subregiones, de modo que cada subregión contenga la mayor proporción posible de ejemplos de una misma clase (**Quinlan 1986**). Cada punto del árbol donde ocurre un *split* o división, basada en una regla, se denomina nodo. Según su altura dentro del árbol estos se clasifican como raíz (*root*), intermedios u hojas (*leaves*) (Figura C.5). De esta manera, la profundidad de un árbol quedará determinada por el número máximo de nodos de una rama.

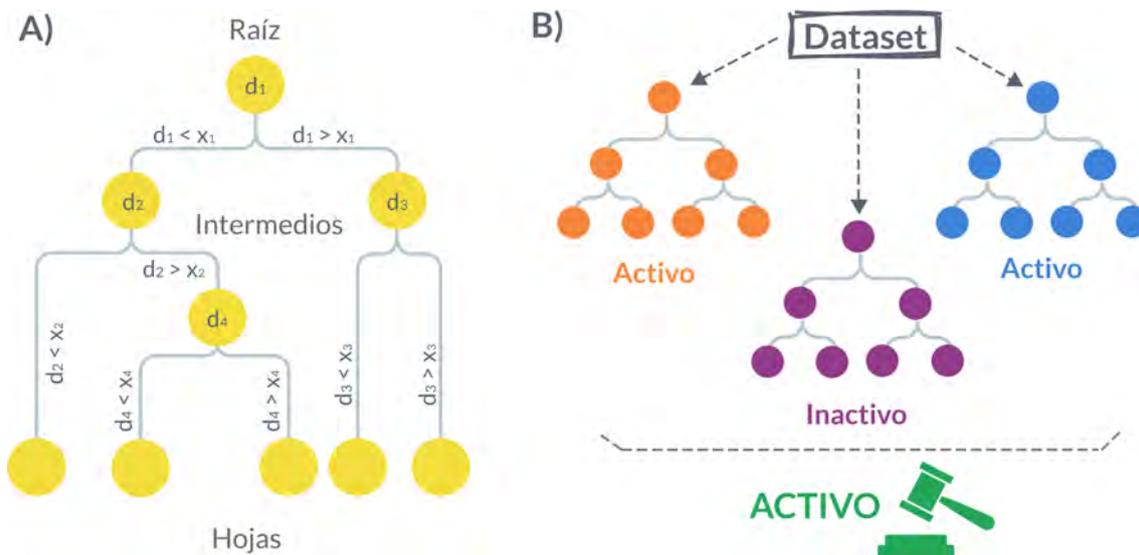


Figura C.5: Funcionamiento del algoritmo ExtraTrees. A) Diagrama general de un árbol de decisión. Los distintos tipos de nodos (raíz, intermedios y hojas) son señalados en el árbol. d_1 , d_2 , d_3 y d_4 se refieren a los descriptores utilizados en los *splits*, mientras que x_1 , x_2 , x_3 y x_4 son los valores de corte para cada uno de ellos. B) Esquemático de un *ensemble* de árboles de decisión donde cada árbol, construido a partir de la totalidad del conjunto de datos (*dataset*), se utiliza para predecir la clase de un compuesto desconocido, y luego la decisión final del *ensemble* es la clase más votada.

El algoritmo continua dividiendo los datos hasta que las hojas sean puras (solo contengan datos de la misma clase). Sin embargo, para minimizar el sobreajuste, también es posible finalizar el algoritmo de manera prematura, cuando se alcance algún criterio de finalización como profundidad máxima, número máximo de hojas o número mínimo de ejemplos en una hoja.

Dentro de los árboles de decisión, existen distintas variantes del algoritmo, como por ejemplo ID3, C4.5, CART, CHAID, MARS (Patel y Prajapati 2018). A su vez, existen también varias métricas que pueden utilizarse a la hora de seleccionar que variable o predictor elegir para generar un *split*, dentro de las cuales podemos mencionar al *Information Gain*, *Information Gain Ratio* o *Gini Index* (Berzal et al. 2003).

Los árboles de decisión, construidos sobre la totalidad del conjunto de entrenamiento, a menudo tienden a sobreajustar los datos. Sin embargo, existen metodo-

logías de *ensemble learning*, que combinan las predicciones de árboles construidos a partir de distintos subconjuntos del conjunto de entrenamiento. Quizás el más conocido es *Random Forest* (RF) (Breiman 2001), donde estos subconjuntos son obtenidos mediante *bootstrapping* (tomando una muestra del 100% con reemplazo), para luego combinar las predicciones individuales en una sola, lo que se conoce como *bootstrap aggregating* o *bagging* (Breiman 1996). A su vez, a la hora de elegir el mejor predictor para generar un *split*, RF solo considera un subconjunto del total, muestreado de manera aleatoria. Esta técnica se denomina método de subespacios aleatorios o *feature bagging* (Ho 1998).

ExtraTrees es también un algoritmo de *ensemble learning*, pero a diferencia de RF no utiliza *bootstrapping*, si no que muestrea sin reemplazo el conjunto de entrenamiento (Figura C.5). A su vez, los predictores elegidos para realizar un *split* (y su valor de corte), son seleccionados de manera aleatoria sobre un subconjunto del total, también muestreado de manera aleatoria (Geurts, Ernst y Wehenkel 2006).

Originalmente, los árboles de decisión y sus variantes ensambladas se desarrollaron como algoritmos de clasificación binaria. Sin embargo, estos pueden aplicarse también a problemas multi-clase o de regresión (Loh 2011).

C.3.3. Support Vector Machine

Originalmente desarrollado por Vladimir Vapnik (Cortes y Vapnik 1995), la máquina de vector soporte (SVM, del inglés Support Vector Machine), es un algoritmo de aprendizaje automático supervisado, ampliamente utilizado en diversas aplicaciones científicas y tecnológicas (Byvatov y Schneider 2003).

En su forma mas simple, la clasificación binaria de dos clases que son linealmente separables, SVM intenta encontrar un **hiperplano** que separe ambas clases de la mejor manera posible. Cabe destacar que se los denomina hiperplanos ya que generalmente se trabaja en espacios hiperdimensionales, pero en el caso de espacios bidimensionales serían rectas, o planos si el espacio fuese tridimensional. Al subconjunto de observaciones del conjunto de entrenamiento que determinan la ubicación del hiperplano de separación se las denomina **vectores de soporte** (Figura C.6).

Estos hiperplanos cuentan a su vez con un **margen**, definido como la anchura máxima de una región paralela al hiperplano que no contiene ejemplos en su interior (Figura C.6). El objetivo del algoritmo es maximizar este **margen**, para encontrar así los hiperplanos óptimos (Kowalczyk 2017).

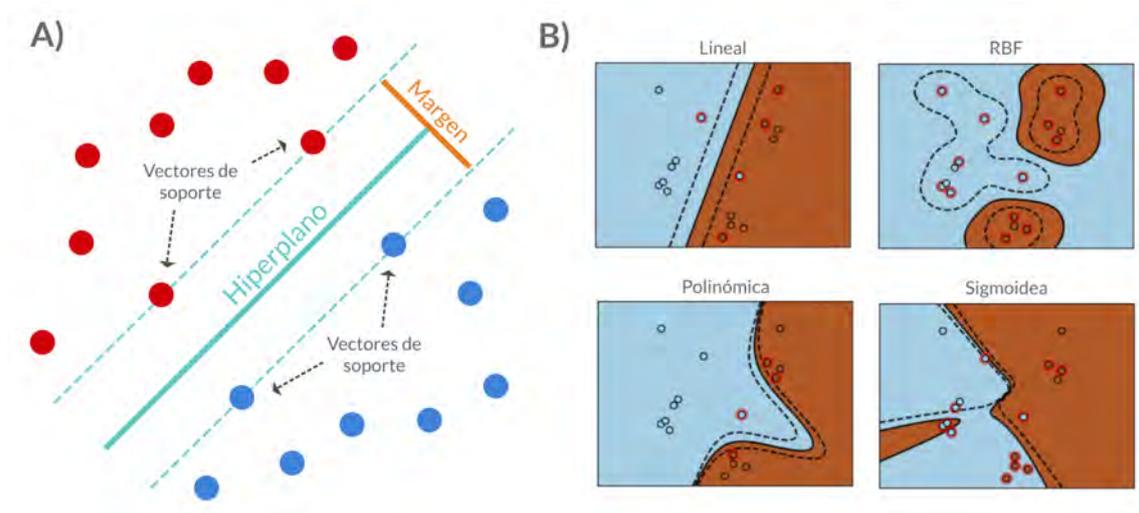


Figura C.6: Funcionamiento del algoritmo SVM. A) Diagrama de los distintos componentes del algoritmo, ejemplificando un caso de clasificación binaria donde ambas clases son linealmente separables. B) Esquematación de los distintos tipos de kernel: Lineal, RBF, Polinómica y Sigmoidea. Los vectores soporte se destacan en color rojo.

Debido a que en los problemas de la vida real, las clases no suelen ser perfectamente separables, es posible suavizar estos márgenes permitiendo al algoritmo cometer algunas clasificaciones erróneas. En la implementación de la librería de Python Skit-learn (Pedregosa et al. 2011), el hiperparámetro **C** permite controlar esta penalidad. Cuando menor sea el valor de **C**, mayor será la penalidad que recibe el algoritmo por cada mala clasificación. Por lo tanto, el margen se volverá mas estrecho y dependerá de menos vectores soporte.

Para los casos donde los datos no sean linealmente separables, es posible utilizar lo que se conoce como truco del **kernel**. Una **kernel** es una función matemática que permite transformar o reasignar los datos a un espacio dimensional diferente, donde se espera que las clases se vuelvan linealmente separables, simplificando así la tarea de encontrar un **hiperplano** o limite de decisión (Hofmann, Schölkopf y Smola 2008). Las **kernel** más utilizadas son la Lineal, RBF (del inglés, Radial

Base Function), Polinómica o Sigmoidea (Figura C.6) (**Ben-Hur** et al. 2008).

Al igual que ExtraTrees, SVM también se desarrolló inicialmente como método de clasificación binaria, pero su aplicación se ha extendido a problemas de clasificación múltiple (**Crammer** y **Singer** 2001) y también regresión. Para este último, la diferencia fundamental es que ahora el hiperplano intentará ajustar la mayor cantidad de datos posibles, similar a una regresión lineal simple (en un espacio bi-dimensional). A su vez, se introduce un hiperparámetro *epsilon* (ϵ) que controla el margen de error aceptado por el modelo. Es decir, las predicciones que se encuentren contenidas entre el intervalo del valor real $\pm \epsilon$ no serán penalizadas por el algoritmo (**Smola** y **Schölkopf** 2004).

Referencias Material suplementario Algoritmos

- Alloghani**, M. et al. (2020). “A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science”. *Supervised and Unsupervised Learning for Data Science*. Ed. por M. W. **Berry**, A. **Mohamed** y B. W. **Yap**. Unsupervised and Semi-Supervised Learning. Cham: Springer International Publishing, págs. 3-21. DOI: [10.1007/978-3-030-22475-2_1](https://doi.org/10.1007/978-3-030-22475-2_1).
- Altman**, N. S. (1992). “An Introduction to Kernel and Nearest-Neighbor Non-parametric Regression”. *The American Statistician* 46.3, págs. 175-185. DOI: [10.1080/00031305.1992.10475879](https://doi.org/10.1080/00031305.1992.10475879).
- Arbelaitz**, O. et al. (2013). “An extensive comparative study of cluster validity indices”. *Pattern Recognition* 46.1, págs. 243-256. DOI: [10.1016/j.patcog.2012.07.021](https://doi.org/10.1016/j.patcog.2012.07.021).
- Bellman**, R. (1966). “Dynamic Programming”. *Science* 153.3731, págs. 34-37. DOI: [10.1126/science.153.3731.34](https://doi.org/10.1126/science.153.3731.34).
- Ben-Hur**, A. et al. (2008). “Support Vector Machines and Kernels for Computational Biology”. *PLOS Computational Biology* 4.10, e1000173. DOI: [10.1371/journal.pcbi.1000173](https://doi.org/10.1371/journal.pcbi.1000173).
- Bergstra**, J., D. **Yamins** y D. **Cox** (2013). “Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms”. Austin, Texas, págs. 13-19. DOI: [10.25080/Majora-8b375195-003](https://doi.org/10.25080/Majora-8b375195-003).
- Bergstra**, J. S. et al. (2011). “Algorithms for Hyper-Parameter Optimization”. *Advances in Neural Information Processing Systems 24*. Ed. por J. **Shawe-Taylor** et al. Curran Associates, Inc., págs. 2546-2554.

- Berzal**, F. et al. (2003). “On the quest for easy-to-understand splitting rules”. *Data & Knowledge Engineering* 44.1, págs. 31-48. DOI: [10.1016/S0169-023X\(02\)00062-9](https://doi.org/10.1016/S0169-023X(02)00062-9).
- Bholowalia**, P. y **A. Kumar** (2014). “EBK-means: A clustering technique based on elbow method and k-means in WSN”. *International Journal of Computer Applications* 105.9.
- Breiman**, L. (1996). “Bagging predictors”. *Machine Learning* 24.2, págs. 123-140. DOI: [10.1007/BF00058655](https://doi.org/10.1007/BF00058655).
- (2001). “Random Forests”. *Machine Learning* 45.1, págs. 5-32. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Butina**, D. (1999). “Unsupervised Data Base Clustering Based on Daylight’s Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets”. *Journal of Chemical Information and Computer Sciences* 39.4, págs. 747-750. DOI: [10.1021/ci9803381](https://doi.org/10.1021/ci9803381).
- Byvatov**, E. y **G. Schneider** (2003). “Support vector machine applications in bioinformatics”. *Applied bioinformatics* 2.2, págs. 67-77.
- Caliński**, T. y **J. Harabasz** (1974). “A dendrite method for cluster analysis”. *Communications in Statistics* 3.1, págs. 1-27. DOI: [10.1080/03610927408827101](https://doi.org/10.1080/03610927408827101).
- Cortes**, C. y **V. Vapnik** (1995). “Support-vector networks”. *Machine Learning* 20.3, págs. 273-297. DOI: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018).
- Crammer**, K. e **Y. Singer** (2001). “On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines”. *Journal of Machine Learning Research* 2.Dec, págs. 265-292.
- Davies**, D. L. y **D. W. Bouldin** (1979). “A Cluster Separation Measure”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1.2, págs. 224-227. DOI: [10.1109/TPAMI.1979.4766909](https://doi.org/10.1109/TPAMI.1979.4766909).
- Dunn**, J. C. (1973). “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters”. *Journal of Cybernetics* 3.3, págs. 32-57. DOI: [10.1080/01969727308546046](https://doi.org/10.1080/01969727308546046).
- Estivill-Castro**, V. (2002). “Why so many clustering algorithms: a position paper”. *ACM SIGKDD Explorations Newsletter* 4.1, págs. 65-75. DOI: [10.1145/568574.568575](https://doi.org/10.1145/568574.568575).

- Fang**, M. et al. (2011). “A new hypothesis of drug refractory epilepsy: neural network hypothesis”. *Medical Hypotheses* 76.6, págs. 871-876. DOI: [10.1016/j.mehy.2011.02.039](https://doi.org/10.1016/j.mehy.2011.02.039).
- Geurts**, P., D. **Ernst** y L. **Wehenkel** (2006). “Extremely randomized trees”. *Machine Learning* 63.1, págs. 3-42. DOI: [10.1007/s10994-006-6226-1](https://doi.org/10.1007/s10994-006-6226-1).
- Hall**, L. H. y L. B. **Kier** (1995). “Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information”. *Journal of Chemical Information and Computer Sciences* 35.6, págs. 1039-1045. DOI: [10.1021/ci00028a014](https://doi.org/10.1021/ci00028a014).
- Ho**, T. K. (1998). “The random subspace method for constructing decision forests”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8, págs. 832-844. DOI: [10.1109/34.709601](https://doi.org/10.1109/34.709601).
- Hofmann**, T., B. **Schölkopf** y A. J. **Smola** (2008). “Kernel methods in machine learning”. *The Annals of Statistics* 36.3, págs. 1171-1220. DOI: [10.1214/009053607000000677](https://doi.org/10.1214/009053607000000677).
- Hutter**, F., H. H. **Hoos** y K. **Leyton-Brown** (2011). “Sequential Model-Based Optimization for General Algorithm Configuration”. *Learning and Intelligent Optimization*. Ed. por C. A. C. **Coello**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, págs. 507-523. DOI: [10.1007/978-3-642-25566-3_40](https://doi.org/10.1007/978-3-642-25566-3_40).
- Kowalczyk**, A. (2017). “Support vector machines succinctly”. *Syncfusion Inc*.
- Kramer**, O. (2013). “K-Nearest Neighbors”. *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Ed. por O. **Kramer**. Intelligent Systems Reference Library. Berlin, Heidelberg: Springer, págs. 13-23. DOI: [10.1007/978-3-642-38652-7_2](https://doi.org/10.1007/978-3-642-38652-7_2).
- Loh**, W.-Y. (2011). “Classification and regression trees”. *WIREs Data Mining and Knowledge Discovery* 1.1, págs. 14-23. DOI: [10.1002/widm.8](https://doi.org/10.1002/widm.8).
- McInnes**, L., J. **Healy** y J. **Melville** (2018). “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. *arXiv:1802.03426 [cs, stat]*.
- Mirjalili**, S. (2019). “Genetic Algorithm”. *Evolutionary Algorithms and Neural Networks: Theory and Applications*. Ed. por S. **Mirjalili**. Studies in Computational Intelligence. Cham: Springer International Publishing, págs. 43-55. DOI: [10.1007/978-3-319-93025-1_4](https://doi.org/10.1007/978-3-319-93025-1_4).

- Patel, H. y P. Prajapati** (2018). “Study and Analysis of Decision Tree Based Classification Algorithms”. *International Journal of Computer Sciences and Engineering* 6, págs. 74-78. DOI: [10.26438/ijcse/v6i10.7478](https://doi.org/10.26438/ijcse/v6i10.7478).
- Pedregosa, F. et al.** (2011). “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research* 12.85, págs. 2825-2830.
- Prada Gori, D. N. et al.** (2022). “iRaPCA and SOMoC: Development and Validation of Web Applications for New Approaches for the Clustering of Small Molecules”. *Journal of Chemical Information and Modeling*. DOI: [10.1021/acs.jcim.2c00265](https://doi.org/10.1021/acs.jcim.2c00265).
- Quinlan, J. R.** (1986). “Induction of decision trees”. *Machine Learning* 1.1, págs. 81-106. DOI: [10.1007/BF00116251](https://doi.org/10.1007/BF00116251).
- Rousseeuw, P. J.** (1987). “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. *Journal of Computational and Applied Mathematics* 20, págs. 53-65. DOI: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Russell, S. y P. Norvig** (2010). *Artificial Intelligence: A Modern Approach*. 3.^a ed. Prentice Hall.
- Shahriari, B. et al.** (2016). “Taking the Human Out of the Loop: A Review of Bayesian Optimization”. *Proceedings of the IEEE* 104.1, págs. 148-175. DOI: [10.1109/JPROC.2015.2494218](https://doi.org/10.1109/JPROC.2015.2494218).
- Sivanandam, S. y S. Deepa** (2008). “Genetic Algorithms”. *Introduction to Genetic Algorithms*. Ed. por S. **Sivanandam** y S. **Deepa**. Berlin, Heidelberg: Springer, págs. 15-37. DOI: [10.1007/978-3-540-73190-0_2](https://doi.org/10.1007/978-3-540-73190-0_2).
- Smola, A. J. y B. Schölkopf** (2004). “A tutorial on support vector regression”. *Statistics and Computing* 14.3, págs. 199-222. DOI: [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).
- Snoek, J., H. Larochelle y R. P. Adams** (2012). “Practical Bayesian Optimization of Machine Learning Algorithms”. *Advances in Neural Information Processing Systems* 25. Ed. por F. **Pereira** et al. Curran Associates, Inc., págs. 2951-2959.
- Ward, J. H.** (1963). “Hierarchical Grouping to Optimize an Objective Function”. *Journal of the American Statistical Association* 58.301, págs. 236-244. DOI: [10.1080/01621459.1963.10500845](https://doi.org/10.1080/01621459.1963.10500845).
- Wicaksono, A. S. y A. Afif** (2018). “Hyper Parameter Optimization using Genetic Algorithm on Machine Learning Methods for Online News Popularity Predic-

tion". *International Journal of Advanced Computer Science and Applications*

9.12. DOI: [10.14569/IJACSA.2018.091238](https://doi.org/10.14569/IJACSA.2018.091238).



Material suplementario

Métricas

Existe una gran variedad de métricas que pueden emplearse para juzgar el desempeño de un modelo predictivo, dependiendo si se trata de un problema de clasificación, regresión o ranking. Este apartado resume algunas de las métricas más comunes para cada uno de los problemas antes mencionados, focalizándose en las que fueron utilizadas en la presente tesis.

D.1. Métricas de regresión

Error Cuadrático Medio (MSE, del inglés Mean Squared Error):

Esta métrica considera la media de los errores elevados al cuadrado. El error para cada predicción es la diferencia entre el valor real y el predicho. Puede tomar valores entre cero e infinito (pero nunca negativos), siendo cero el valor para un modelo perfecto. Debido a que la penalidad del error es cuadrática, esta métrica es sensible a la presencia de valores extremos o *outliers* en el conjunto de datos.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Donde Y_i es el valor predicho para el ejemplo i , \hat{Y}_i es el valor real y n es el número de observaciones.

Error Absoluto Medio (MAE, del inglés Mean Absolute Error):

A diferencia del MSE, esta métrica mide el promedio del valor absoluto de los errores. Al igual que el MSE también toma valores entre cero e infinito, donde cero indica un ajuste perfecto. Esta métrica penaliza los errores de manera lineal, por lo que es menos sensible a la presencia de *outliers* o ruido en el conjunto de datos.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

Donde Y_i es el valor predicho para el ejemplo i , \hat{Y}_i es el valor real y n es el número de observaciones.

Coefficiente de Pearson (R):

El coeficiente R de Pearson es una medida de la correlación lineal entre dos variables X e Y (**Pearson 1895**). Puede tomar valores entre -1 (correlación lineal inversa entre las variables) y 1 (correlación lineal positiva entre las variables). Un valor igual a 0 indica que no existe correlación lineal entre ambas variables. Si la correlación es positiva significa que una variable aumenta si la otra también lo hace. Para dos variables x e y el coeficiente R se calcula como:

$$R = \frac{\sum_{i=1}^n (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^n (x_i - \hat{x})^2} \sqrt{\sum_{i=1}^n (y_i - \hat{y})^2}}$$

Donde x_i es la observación i de la variable x , \hat{x} es la media de todas las observaciones en x , y_i es la observación i de la variable y , \hat{y} es la media de todas las observaciones en y , y n es el número total de observaciones. Esta medida de correlación paramétrica asume ciertas características sobre el conjunto de datos:

1. Ambas variables siguen una distribución normal.
2. No existen *outliers* significativos en el conjunto de datos.
3. Existe una relación lineal entre ambas variables x e y .
4. Homocedasticidad o varianza constante de los residuos.

Coefficiente de determinación (R^2):

El valor del coeficiente de determinación nos habla de la fuerza de la correlación entre dos variables x e y , y se utiliza comúnmente para evaluar la calidad de los modelos predictivos. Es decir, el R^2 representa la proporción de la varianza que puede ser explicada por el modelo. Por lo tanto, se ve fuertemente afectado por la varianza de las variables independientes, y solo nos permite comparar modelos que fueron entrenados con los mismos datos. Puede tomar valores entre 0 (no existe correlación entre los datos) y $+1$ (correlación perfecta). Para el caso de la regresión lineal simple, este coeficiente es simplemente el cuadrado del coeficiente R de Pearson (**Heinisch 1962**).

D.2. Métricas de ranking

Coefficiente Rho de Spearman:

Esta métrica mide el grado de correlación entre dos rankings, y su rango varía entre -1 y 1 según los rankings sean inversos o perfectamente iguales, respectivamente (Zar 2014). El coeficiente de Spearman entre dos variables está definido como coeficiente de Pearson entre los rankings de ambas variables. Mientras que el R de Pearson sólo asume relaciones lineales entre las variables, el coeficiente de Spearman permite evaluar relaciones monotónicas entre ellas, ya sean lineales o no. Esta métrica no realiza ninguna asunción sobre la distribución de los datos, por lo que es considerado un método no paramétrico.

$$\mathbf{Rho} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Donde d_i es la diferencia entre los rankings de x_i e y_i , y n es el número de observaciones.

Coefficiente TauB de Kendall:

Al igual que el Rho de Spearman, esta también es una medida de correlación entre dos rankings (Hoeffding 1957). Su rango varía entre -1 (los ranking son perfectamente inversos) y 1 (ambos rankings son idénticos, es decir que cada elemento ocupa la misma posición en ambos). Un valor igual a 0 indica que no existe asociación alguna entre los rankings. Este método no paramétrico es particularmente apropiado cuando el conjunto de datos es pequeño y existen empates en el ranking. Su fórmula se define como:

$$\mathbf{TauB} = \frac{P - Q}{\sqrt{(P + Q + T) \times (P + Q + U)}}$$

Donde P es el número de pares concordantes, Q el número de pares discrepantes, T es el número de empates solo en x y U el número de empates solo en y . Si ocurre un empate en x e y , no se considera en T ni en U .

D.3. Métricas de clasificación

Para el caso de la clasificación binaria existen cuatro combinaciones posibles de categorías, representadas en lo que se conoce como matriz de confusión (Figura D.1): Verdaderos positivos (VP), Falsos Positivos (FP) o error de tipo I, Verdaderos Negativos (VN) y Falsos Negativos (FN) o error de tipo II.

		Predicción	
		Positivo	Negativo
Observación	Positivo	VP	FN
	Negativo	FP	VN

Figura D.1: Matriz de confusión para un clasificador binario.

De esta matriz, se desprenden varias métricas que comúnmente se utilizan en problemas de clasificación binaria como por ejemplo:

Sensibilidad (Se, del inglés Sensitivity):

También conocida como razón de verdaderos positivos o *recall*, esta métrica considera la proporción de activos correctamente clasificados como tales.

$$Se = \frac{VP}{VP + FN}$$

Especificidad (Sp, del inglés Specificity):

Se refiere a la proporción de inactivos correctamente clasificados como tales.

$$Sp = \frac{VN}{VN + FP}$$

Exactitud (Acc, del inglés Accuracy):

Esta métrica mide la fracción de buenas clasificaciones. Es decir, la cantidad de Verdaderos Positivos y Verdaderos Negativos sobre el total de clasificaciones. La exactitud no permite distinguir si el clasificador comete más errores al clasificar una categoría u otra.

$$\mathbf{Acc} = \frac{VP + VN}{VP + VN + FN + FP}$$

F1-Score:

Esta métrica se define como la media armónica de la precisión y la sensibilidad. De su definición, se desprende que esta métrica no considera a los Verdaderos Negativos.

$$\mathbf{F1} = \frac{2 \times VP}{2 \times VP + FP + FN}$$

Coefficiente de correlación de Matthews (MCC, del inglés Matthews Correlation Coefficient):

Este coeficiente puede considerarse como una medida de la correlación entre las predicciones, por lo que puede tomar valores entre -1 (discrepancia total entre las predicciones) y +1 (clasificador perfecto), mientras que el 0 que representa una clasificación al azar ([Matthews 1975](#)).

$$\mathbf{MCC} = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

Coefficiente K de Cohen (Cohen's Kappa):

Esta métrica permite medir el grado de concordancia entre las evaluaciones cualitativas (categóricas) que realizan distintos observadores, ajustadas por el efecto del azar. Por lo tanto, en el contexto de la clasificación binaria esta métrica puede utilizarse para comparar el desempeño de distintos modelos ([McHugh 2012](#)).

Pérdida logarítmica (LogLoss, del inglés Logarithmic Loss):

La métrica LogLoss se basa en probabilidades, y es ampliamente utilizada en problemas de clasificación. Esta mide cuanto se aleja la probabilidad de una predicción de su valor verdadero. Valores más pequeños de LogLoss indican un mejor modelo (**Gneiting y Raftery 2007**).

$$\mathbf{LogLoss} = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \text{Ln}(p_i) + (1 - y_i) \cdot \text{Ln}(1 - p_i)]$$

donde i es la observación analizada, y_i es su valor real, p_i es la probabilidad de la predicción i y N es el número total de observaciones.

Curvas ROC

Las curvas de Característica Operativa del Receptor (ROC, del inglés Receiver Operating Characteristic) son una representación gráfica de como varia la sensibilidad (Se) frente a la especificidad (Sp) para un clasificador binario, a medida que varia el umbral de discriminación o *cutoff* (valor de corte que se utiliza para clasificar los ejemplos) (**Marzban 2004**). Para el caso de un modelo basado en un algoritmo de aprendizaje automático, el *score* será la probabilidad de clasificación de la clase positiva, mientras que si se trata de un modelo de docking, este será la afinidad predicha por el modelo. En la Figura D.2 se esquematiza el proceso de construcción de una curva ROC.

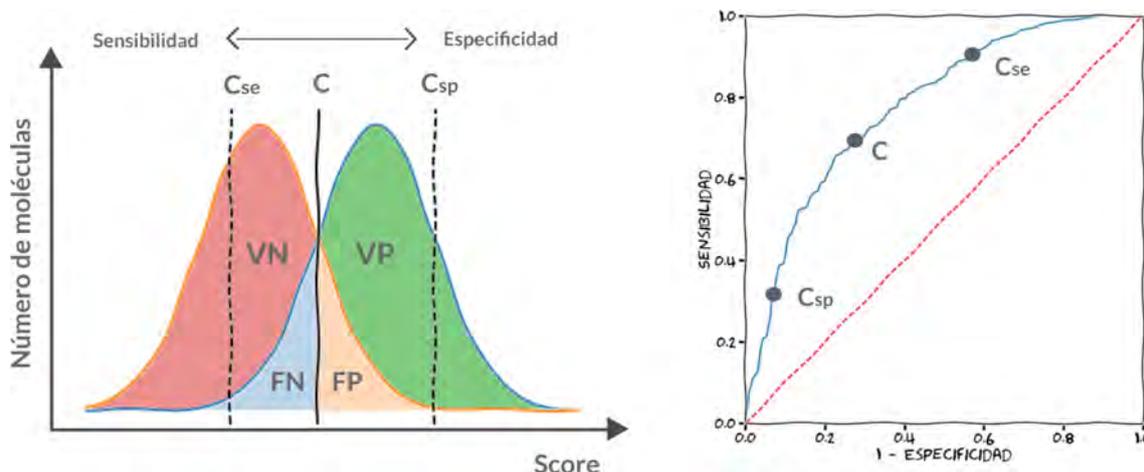


Figura D.2: Representación esquemática de la construcción de una curva ROC. En el esquema de la izquierda se representa la distribución de *score* obtenida para cada clase (positiva/negativa), y diferentes valores de *cutoff* (C, C1, C2). Cada uno de estos valores tiene asociado un valor de Se y Sp, los cuales constituyen el trazo de la curva ROC (derecha). La línea roja punteada sobre la diagonal representa un modelo nulo o clasificador al azar, cuya AUC-ROC es igual a 0.5.

El área bajo la curva ROC (AUC-ROC) puede utilizarse como parámetro para juzgar la capacidad predictiva de un modelo (Triballeau et al. 2005). Un clasificador al azar o modelo nulo, representado por la línea diagonal, tendrá un AUCROC=0.5; mientras que un clasificador perfecto tendrá un área igual a 1, y su curva coincide con los ejes de coordenadas x e y. En líneas generales, un modelo se considera “bueno” si su AUC-ROC es mayor a 0.7.

D.4. Métricas específicas del cribado virtual

Existen ciertas métricas diseñadas específicamente para aplicaciones del VS, donde el *early enrichment* o enriquecimiento temprano es un factor importante. El concepto de enriquecimiento temprano se refiere a la capacidad de un modelo de recuperar a los compuestos activos en los primeros puestos, ya que esta será la zona del ranking donde nos concentramos a la hora de seleccionar los candidatos más promisorios (Truchon y Bayly 2007).

Factor de enriquecimiento:

Esta métrica mide cuantos más activos podemos encontrar en una cierta fracción del ranking, con respecto a una distribución aleatoria de los activos en el mismo. El tamaño de esta fracción es una variable que debemos definir de antemano y generalmente está asociada a nuestra capacidad posterior de análisis y evaluación experimental de los compuestos. Un factor de enriquecimiento al 1 % considera como “zona de enriquecimiento temprano” solo al primer 1 % de nuestra base de datos ordenada (Truchon y Bayly 2007). Los factores de enriquecimiento comúnmente reportados son calculados al 1 % y al 10 %.

Robust Initial Enhancement (RIE):

Esta métrica desarrollada por Sheridan (Sheridan et al. 2001) utiliza un método de pesos decrecientes como función de ranqueo, como estrategia para mitigar uno de los problemas fundamentales asociado a los factores de enriquecimiento cuando la proporción de activos en la base de datos es pequeña, como lo es la varianza. El parámetro α está relacionado con el tamaño de la zona del ranking definida como enriquecimiento temprano. Un valor de RIE=1 significa que el enriquecimiento del modelo es nulo con respecto al azar.

Boltzmann-Enhanced Discrimination of ROC (BEDROC):

Propuesta originalmente por Truchon y Bayly (Truchon y Bayly 2007), esta métrica puede entenderse como una generalización del AUCROC, donde se incorpora una función del tipo exponencial que asigna una mayor importancia a los “activos” recuperados en la zona del ranking que consideramos de “enriquecimiento temprano”. En esencia, esta métrica puede interpretarse como una versión normalizada del RIE, acotada en el rango 0-1.

Al igual que en el RIE, existe un parámetro α relacionado con la zona en ranking a la que daremos más importancia. Comúnmente se utiliza un valor de $\alpha=20$, lo que significa que el primer 8 % del ranking contribuye en un 80 % al valor de BEDROC (Ntie-Kang 2020).

Referencias Material

suplementario Métricas

- Gneiting**, T. y A. E. **Raftery** (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. *Journal of the American Statistical Association* 102.477, págs. 359-378. DOI: [10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437).
- Heinisch**, O. (1962). “Steel, R. G. D., and J. H. Torrie: Principles and Procedures of Statistics. (With special Reference to the Biological Sciences.) McGraw-Hill Book Company, New York, Toronto, London 1960, 481 S., 15 Abb.; 81 s 6 d”. *Biometrische Zeitschrift* 4.3, págs. 207-208. DOI: [10.1002/bimj.19620040313](https://doi.org/10.1002/bimj.19620040313).
- Hoeffding**, W. (1957). “Review of Rank Correlation Methods”. *Econometrica* 25.1, págs. 181-183. DOI: [10.2307/1907752](https://doi.org/10.2307/1907752).
- Marzban**, C. (2004). “The ROC Curve and the Area under It as Performance Measures”. *Weather and Forecasting* 19.6, págs. 1106-1114. DOI: [10.1175/825.1](https://doi.org/10.1175/825.1).
- Matthews**, B. W. (1975). “Comparison of the predicted and observed secondary structure of T4 phage lysozyme”. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405.2, págs. 442-451. DOI: [10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9).
- McHugh**, M. L. (2012). “Interrater reliability: the kappa statistic”. *Biochemia Medica* 22.3, págs. 276-282. DOI: [10.11613/BM.2012.031](https://doi.org/10.11613/BM.2012.031).
- Ntie-Kang**, F. (2020). *Volume 1 Fundamental Concepts*. De Gruyter.
- Pearson**, K. (1895). “Note on Regression and Inheritance in the Case of Two Parents”. *Proceedings of the Royal Society of London* 58, págs. 240-242.
- Sheridan**, R. P. et al. (2001). “Protocols for bridging the peptide to nonpeptide gap in topological similarity searches”. *Journal of Chemical Information and Computer Sciences* 41.5, págs. 1395-1406. DOI: [10.1021/ci0100144](https://doi.org/10.1021/ci0100144).

- Triballeau**, N. et al. (2005). “Virtual screening workflow development guided by the receiver operating characteristic curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4”. *Journal of Medicinal Chemistry* 48.7, págs. 2534-2547. DOI: [10.1021/jm049092j](https://doi.org/10.1021/jm049092j).
- Truchon**, J.-F. y C. I. **Bayly** (2007). “Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem”. *Journal of Chemical Information and Modeling* 47.2, págs. 488-508. DOI: [10.1021/ci600426e](https://doi.org/10.1021/ci600426e).
- Zar**, J. H. (2014). “Spearman Rank Correlation: Overview”. *Wiley StatsRef: Statistics Reference Online*. American Cancer Society. DOI: [10.1002/9781118445112.stat05964](https://doi.org/10.1002/9781118445112.stat05964).