



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS
QUÍMICAS**

Estudio de la participación de interacciones débiles en
la estructuración del sitio donde la nucleoproteína de
influenza A reconoce ligantes orgánicos derivados de
nucleozina

TESIS

PARA OPTAR POR EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

M. en C. Erick Alberto Correa Padilla

Dra. Karina Martínez Mayorga
Instituto de Química – UNAM
Ciudad Universitaria,
Ciudad de México, 13 de abril del 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS
QUÍMICAS**

Estudio de la participación de interacciones débiles en
la estructuración del sitio donde la nucleoproteína de
influenza A reconoce ligantes orgánicos derivados de
nucleozina

**T E S I S
PARA OPTAR POR EL GRADO DE**

DOCTOR EN CIENCIAS

PRESENTA

M. en C. Erick Alberto Correa Padilla

Dra. Karina Martínez Mayorga
Instituto de Química - UNAM



Ciudad de México, 13 de abril del 2023.

“A.I. won't replace medicinal chemists, but medicinal chemists who use A.I. will replace those who don't”

Drug Discovery Today (2018), 23(7),1373-1384.

DEDICATORIAS

A mi madre, a mi padre y a mis hermanas.

*A mi familia y amigos, quienes han sentido alegría
y nostalgia en este proceso.*

A la memoria de Gerardo, Ofelia y María.

Gracias por toda su paciencia, apoyo y amor.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para la realización de mis estudios de doctorado (CVU: 687699)

Al Dr. Gabriel Cuevas Gonzales Bravo (Instituto de Química, UNAM), quién fue mi primer tutor, por todo su apoyo durante la realización de mis estudios. Su apoyo y consejo fue imprescindible.

A I. Chávez Uribe, E. Huerta Salazar, M.A. Peña González, H. Ríos, Ma. C. García-González y J. Pérez Flores, técnicos académicos del Instituto de Química de la UNAM, por su ayuda para la obtención de la Espectroscopía de Resonancia Magnética Nuclear y Espectrometría de Masas.

Al Dr. Abraham Madariaga (IIMAS-UNAM) por su apoyo durante los estudios de acoplamiento molecular automatizado y a Alejandro Hernández por su apoyo en el manejo del código de Python para la generación de modelos de predicción.

Al Dr. Ignacio Regla, al Dr. Manuel López Ortiz (FES-Zaragoza, UNAM) y al M. en C. Alejandro Aguilera por su apoyo y asesoría en la parte sintética.

Al Dr. Fernando Esquivel Guadarrama (Facultad de Medicina, UAEMor) y a la Maestra en Medicina Molecular Yunuen Acevedo Betancur por su apoyo en la evaluación biológica de los compuestos.

Al Dr. Jaime Pérez Villanueva (UAM-Xochimilco), al Dr. José Alberto Riviera Chávez (Instituto de Química, UNAM), a la Dra. Martha Escárcega (Facultad de Química, UNAM) y al Dr. Joaquín Tamariz (ENCB-IPN), quienes formaron parte de mi comité tutelar durante mis estudios de doctorado. Sus comentarios fueron imprescindibles para la culminación de este trabajo.

A los integrantes de mi jurado por dedicar atención a la revisión del presente trabajo. El jurado está integrado por los profesores: Dr. Fernando Cortés Guzmán (Instituto de Química, UNAM), Dr. Alfonso Sebastián Lira Rocha (Facultad de Química, UNAM), Dr. Jaime Pérez Villanueva (UAM-Xochimilco), Dr. Antonio Romo Mancillas (Facultad de Química, UAQ) y el Dr. Joaquín Barroso Flores (Instituto de Química, UNAM).

A mi tutora la Dra. Karina Martínez Mayorga (Instituto de Química e IIMAS, UNAM) por todo su apoyo y paciencia en este proyecto.

Este trabajo se realizó en el Instituto de Química de la Universidad Nacional Autónoma de México en colaboración con la Facultad de Medicina de la Universidad Autónoma del Estado de Morelos. Corresponde a la continuación de mi proyecto de maestría con la tesis:

Correa Padilla, E. A. Síntesis de análogos de nucleozina con potencial actividad antiviral que incorporan el sistema 2,5 diazabícclo[2.2.1]heptano [Tesis de maestría], F.E.S. Zaragoza, UNAM, 2018 ([liga](#))

El proyecto se presentó, en modalidad de cartel, en el Congreso Internacional de la Sociedad Química de México "*Una Química, Muchas Voces*", organizado en colaboración con la Universidad Autónoma de Yucatán, en la ciudad de Mérida, Yucatán, del 28 de agosto al 2 de septiembre del 2022.

Como resultado de la investigación realizada fue publicado un artículo científico en la revista PLOS-ONE: <https://doi.org/10.1371/journal.pone.0277073>

CONTENIDO

INTRODUCCION	2
MARCO TEÓRICO	4
Química medicinal, receptores y ligandos	5
Relaciones estructura actividad cuantitativa (QSAR)	7
Aprendizaje automatizado para la construcción de modelos predictivos QSAR.....	9
Algoritmos usados en aprendizaje automatizado.....	10
Validación del modelo.....	16
Dominio de Aplicabilidad	17
Acoplamiento molecular	18
Influenza	21
Nucleoproteína de influenza A	27
Nucleozina (nlz)	30
Estructura de la nlz y sus derivados	32
Mecanismo de acción de la nlz	36
Síntesis de análogos de nucleozina.....	39
PLANTEAMIENTO DEL PROBLEMA	42
HIPÓTESIS	42
OBJETIVOS	42
Objetivo general.....	42
Objetivos particulares	42
RESULTADOS	43
Fundamento del proyecto	43
Resultados de acoplamiento molecular para análogos de nlz	45
Resultados de modelos QSAR de predicción de actividad inhibitoria con análogos de nlz.....	57
Síntesis de análogos de nlz.....	60
Caracterización espectroscópica de los compuestos sintetizados.	61
Análisis de descriptores.....	65
Evaluación biológica de análogos de nlz.....	70
DISCUSIÓN DE RESULTADOS	73
CONCLUSIONES	75
METODOLOGÍA	76
Metodología de acoplamiento molecular para análogos de nlz	76
Metodología de modelos QSAR de predicción de actividad inhibitoria con análogos de nlz.....	76
Metodología de síntesis de análogos de nlz.....	83
Metodología de evaluación biológica de análogos de nlz	84
ANEXOS	85
REFERENCIAS	86

INTRODUCCION

La influenza es una infección viral aguda de las vías respiratorias, Los virus de influenza A se presentan a nivel mundial y causan brotes anuales de intensidad variable de comportamiento endémico, epidémico o pandémico. La mutabilidad y la gran frecuencia de reordenamiento genético y los cambios antigénicos resultantes en las glucoproteínas de la superficie viral dificultan el control de virus, evidenciando la necesidad de buscar nuevas moléculas con actividad antiviral. La molécula llamada nucleozina (nlz), es un prometedor agente antiviral capaz de inactivar a la nucleoproteína (NP) del virus mediante la formación de complejos ligando-proteína.

En este trabajo estudiamos el efecto de la modificación del sistema de piperazina sobre la actividad inhibitoria de nucleozina. Mediante técnicas computacionales se estudiaron las nuevas moléculas propuestas de acuerdo con sus propiedades fisicoquímicas y de interacciones intermoleculares en el sitio de reconocimiento y, por técnicas de síntesis orgánica, se obtuvieron algunos compuestos para su posterior evaluación biológica. Se realizó un modelo QSAR con Python como lenguaje de programación para la predicción de actividad biológica de nuevos análogos de nlz; en este trabajo nos enfocaremos en aquellos con modificaciones en el sistema central de piperazina de la molécula de nlz. El modelo fue validado mediante validación interna con el método de *Leave-Many-Out* (CV-LMO) y mediante validación externa. Para medir la bondad de ajuste del conjunto de datos no validados se utilizó el coeficiente de determinación (r^2) de valores predichos contra experimentales con un valor de $r^2 = 0.90$. Los coeficientes de determinación para las validaciones interna y externa son, respectivamente, $q^2_{CV-LMO} = 0.60$, y, $q^2_{ext} = 0.60$.

El cálculo de descriptores moleculares indica la importancia de los átomos de nitrógeno y el número de átomos de carbono en el sistema de piperazina, principalmente sus efectos sobre la lipofilidad y la solubilidad, siendo los átomos de nitrógeno esenciales para la biodisponibilidad. Se utilizó el programa MOE para estudiar el modo de unión para el complejo ligando-proteína, encontrando que la piperazina tiene un efecto conformacional esencial para interactuar con el sitio activo, por lo que modificaciones en este anillo tienen

implicaciones importantes en la estabilización dentro del sitio de reconocimiento. Se realizó una nueva estrategia de síntesis para la obtención de nucleozina y dicha estrategia se utilizó para la síntesis de tres análogos estructurales de nlz. Hasta el momento, se han realizado evaluaciones biológicas de sólo dos de éstos, observando que los cambios de rigidez en el sistema de nucleozina disminuyen la actividad inhibitoria. Los resultados de la evaluación biológica son concordantes con el modelo de predicción QSAR construido en este trabajo de investigación.

MARCO TEÓRICO

El virus de la influenza A es un virus que está en constantes cambios y al que se le debe tener extremas precauciones ya que es capaz de causar epidemias a nivel global. Este virus es capaz de transmitirse entre diferentes especies animales y adaptarse para causar enfermedades a los seres humanos. La influenza aviar e influenza porcina han sido casos históricamente reportados de epidemias a nivel global con diferentes grados de impacto, tanto a nivel salud, económico y social, siendo los años de la primera guerra mundial cuando más impactó al mundo. El 31 de marzo del 2023 se reportó el primer caso de gripe aviar en humanos en el norte de Chile con la cepa H5N1, en un hombre de 53 años, tras detectarse en diferentes especies animales locales, entre ellas lobos marinos y aves de corral desde diciembre de 2022, sin embargo, no hay evidencia de que se presente transmisión de humano a humano¹, y, a la misma fecha, en China se reportó un caso de infección de influenza aviar en humanos con la cepa H3N8.² Los principales antivirales para el tratamiento de la influenza A son Oseltamivir³ y Zanamivir⁴, Peramivir⁵ y Laninamivir⁶, que son inhibidores de la Neuraminidasa⁷ y Baloxavir Marboxil⁸ que es inhibidor de la endonucleasa cap-dependiente, o terapias combinatorias de estos.⁹ Sin embargo, debido a la alta frecuencia de mutaciones el virus de influenza A es capaz de generar resistencia a estos tratamientos,¹⁰ por lo que, es necesario tener presente un “arsenal químico” en caso de que en algún momento se requiera desarrollar, con urgencia, un medicamento contra cepas de Influenza A resistentes.¹¹

Entre los principales enfoques que se proponen para generar terapias contra infecciones de influenza A, destaca la inhibición de Neuraminidasa (NA) y de endonucleasa cap-dependiente, sin embargo, también se propone el uso de inhibidores de Hemaglutinina (HA),¹² así como el bloqueo de canales iónicos con nuevos compuestos derivados de amantano¹³. También se ha reportado el uso de sialidasas (como DAS 181)¹⁴ que evitan la infección del virus removiendo a los receptores, de los tejidos epiteliales del hospedero, necesarios para que se dé la unión del virus de la influenza A y comience el ciclo viral. Otro de los blancos terapéuticos propuestos para encontrar nuevas moléculas con propiedades

antivirales es la nucleoproteína (NP) de influenza A¹⁵; en 2010 se reportó la nucleozina (nlz), una molécula capaz de formar un complejo con esta proteína y evitar la formación de ribonucleoproteína (RNP), inhibiendo así el ciclo del virus.¹⁶

La exploración y el entendimiento de las relaciones estructura actividad de los compuestos con actividad antiviral, como nucleozina, es fundamental para encontrar alguna molécula líder que pueda ser un potencial candidato en la obtención de algún medicamento. En el presente marco teórico haremos una reseña de las técnicas usadas en la ciencia y tecnología para poder encontrar este tipo de relaciones.

Química medicinal, receptores y ligandos

La *Química Medicinal* es una de las disciplinas de la Química que se encarga del descubrimiento, desarrollo, identificación y la interpretación del modo de acción de compuestos biológicamente activos a nivel molecular.¹⁷ El énfasis está puesto en el *desarrollo de fármacos*, pero el interés de la Química Medicinal no se restringe sólo a ese rubro, sino que incluye todos los compuestos bioactivos de manera general. La Química Medicinal está también relacionada al estudio, identificación y síntesis química de compuestos con actividad biológica, fármacos, metabolitos, precursores y compuestos relacionados.

En los sistemas biológicos existen macromoléculas que tienen un papel indispensable en el funcionamiento y la señalización química intercelular e intracelular, éstas pueden encontrarse en la superficie de la célula o en el citoplasma. A las moléculas pequeñas capaces de unirse a estas macromoléculas se les denomina *ligandos*; la principal función de esta unión consiste en activar o desactivar a los receptores, activando o desactivando funciones biológicas en los organismos vivos.

Algunos ejemplos de estos ligandos son:

- Hormonas como la testosterona o los estrógenos involucrados en la diferenciación sexual en los mamíferos.¹⁸

- Neurotransmisores: compuestos liberados por las neuronas para cumplir sus funciones de transmisión de impulso eléctrico, como la acetilcolina.¹⁹
- Antígenos de superficie bacteriana, moléculas simples como la glucosa.²⁰

La función de los ligandos es diversa y generalmente cumplen funciones específicas de señalización celular en los organismos vivos, actúan como verdaderos mensajeros químicos, siendo indispensables para la identificación del medio, la creación de la conciencia, la reproducción, la diferenciación sexual, mecanismos de respuesta a situaciones de alerta, mecanismos de crecimiento y muerte celular, señalización del dolor, funcionamiento del sistema inmune, metabolismo, la obtención de energía, y el mecanismo de acción de fármacos entre muchos otros.

La función biológica específica de estos ligandos está dada por la alta *selectividad* de los receptores por los ligandos, es decir los receptores son capaces de reconocer la estructura de los ligandos en un medio fisiológico y unirse de manera exclusiva a alguno de ellos, efectuando o inhibiendo la función biológica a la que están asociados. La interacción que ocurre entre estos dos compuestos es conocida como interacción ligando-receptor. En el caso de los fármacos estas interacciones ayudan a tratar enfermedades o sus síntomas asociados, y de manera específica se conocen como interacciones fármaco-receptor y son muy estudiadas por la comunidad científica ya que tienen consecuencias interesantes en los organismos vivos.

En la [figura 1](#) se muestra como se ha ido ampliando a través del tiempo el campo de estudio de la Química Medicinal. Los enfoques computacionales hacia el reconocimiento molecular están avanzando de manera veloz debido a la conjunción de datos experimentales de buena calidad, el advenimiento de la Química Combinatoria, el cribado de alto rendimiento, recursos informáticos y avances teóricos.²¹

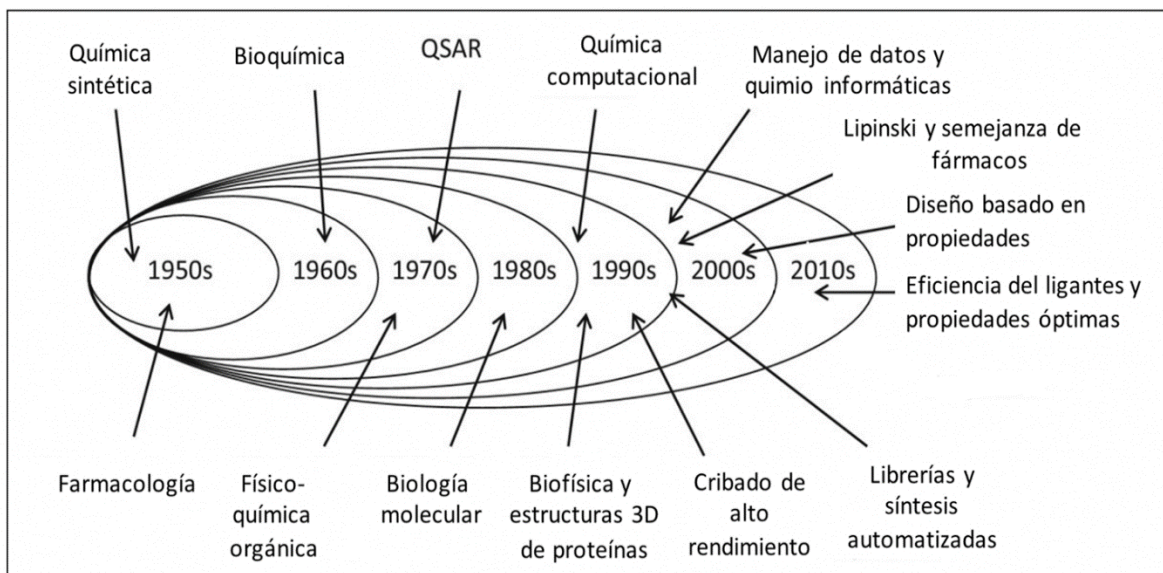


Figura 1. Evolución de las tecnologías en química medicinal. Imagen tomada y adaptada de Meanwell, N. A. Tactics in contemporary drug design; 2014; Vol. 9. <https://doi.org/10.1007/978-3-642-55041-6>.²²

Relaciones estructura actividad cuantitativa (QSAR)

Las relaciones estructura-propiedad cuantitativas QSPR, por sus siglas en inglés *Quantitative Structure Property Relationship*, son herramientas utilizadas en diversas áreas de la Química para la predicción de propiedades moleculares por correlación estructural. El principal objetivo es encontrar **modelos matemáticos** (funciones) que describan propiedades experimentales interesantes de un grupo de moléculas en función de otras propiedades conocidas.

Un **modelo** es una función matemática con diferentes **variables independientes**, capaz de describir alguna propiedad atribuida a la estructura, como la conductividad eléctrica y la actividad biológica; se entiende que estas propiedades serán las **variables dependientes**. Con esta herramienta se puede obtener información útil con fines analíticos y predictivos.

La importancia del modelado de relaciones QSPR en Química Medicinal recae en la predicción de actividades biológicas con potencial efecto terapéutico. Esto permite también optimizar la actividad biológica de fármacos ya existentes por cambios estructurales que modifican la afinidad de los ligandos por los blancos biológicos.²³ Para este caso en

específico, en la Química Medicinal, las relaciones QSPR se denominan relaciones estructura actividad cuantitativa, QSAR por sus siglas en inglés *Quantitative Structure Activity Relationship* y son muy útiles en el desarrollo de fármacos. El uso y aplicación de modelos QSAR *in silico** para predecir la actividad de fármacos ha aumentado recientemente en los últimos años.²⁴ Los modelos QSAR han sido usados intensamente en la Química Medicinal²⁵ para predecir, o descartar, actividades de compuestos para un blanco específico y para estudiar propiedades farmacológicas como el proceso ADME (Absorción, Distribución, Metabolismo y Eliminación),²⁶ así como la predicción de toxicidad²⁷ de los compuestos con aplicaciones en la agricultura, el desarrollo de biocidas²⁸ y búsqueda de nuevas moléculas, ayuda a realizar cribado virtual,²⁹ así como la identificación de alertas estructurales,³⁰ optimización de moléculas líder y es una potencial herramienta para disminuir el uso de animales en experimentación clínica (Tox21)³¹

Algunas de las actividades biológicas que comúnmente se predicen son: la concentración efectiva media (CE_{50}) de un fármaco, su concentración letal media (CL_{50}) y citotóxica media (CC_{50}), o la concentración inhibitoria media (CI_{50}) para antibióticos, antivirales o antifúngicos. En estos casos, la actividad biológica se considera la variable dependiente y está en función de una o más variables independientes que se obtienen a partir de las propiedades fisicoquímicas.

Las variables independientes del modelo son conocidas como *descriptores moleculares*, los cuales son representaciones numéricas de propiedades moleculares calculadas a partir de estructuras moleculares. Estos valores calculados con distintos programas como DRAGON,³² representan propiedades de la molécula como peso molecular, volumen de Van der Waals, momento dipolo, número de átomos de carbono, refractividad molecular, número de anillos aromáticos, coeficientes de partición, área polar superficial, área de superficie hidrofóbica, potencial de ionización, energías de HOMO y/o LUMO, etc.

Si se conocen las actividades biológicas experimentales de un grupo de moléculas con estructuras conocidas, se pueden observar ciertas correlaciones con algunas propiedades

* *in silico*: generado por una computadora.

moleculares integrantes y comienza la parametrización del modelo. Este tipo de **correlaciones** pueden ser **lineales**, **no lineales** o **discontinuas**.

Aprendizaje automatizado para la construcción de modelos predictivos QSAR

Las metodologías actuales para la generación de modelos QSAR involucran el uso de minería de datos e inteligencia artificial (IA, en inglés *Artificial Intelligence*), específicamente, el llamado Aprendizaje Automatizado (ML, por sus siglas en inglés *Machine Learning*). Se han recibido diversas definiciones de aprendizaje automatizado.³³ Arthur Samuel lo definió como “un campo de estudio que brinda a las computadoras la habilidad de aprender sin ser explícitamente programadas”.³⁴ Los algoritmos se entrenan para realizar tareas y hacer predicciones aprendiendo directamente de los datos.³⁵ De acuerdo con Ethem Alpaydin, el aprendizaje automatizado es “programar computadoras para optimizar un criterio de rendimiento usando datos de ejemplo o experiencia pasada”³⁶

Actualmente, este campo tiene un gran número de aplicaciones en el desarrollo de la ciencia y la tecnología, y específicamente dentro del campo de la Química ha sido de utilidad, para el cálculo de energías libres de interacción proteína-ligando, vía simulaciones de Dinámica Molecular, la simulación de dinámicas de deformación en Ciencias de Materiales y en la predicción de reacciones químicas.³⁷

El aprendizaje automatizado se clasifica en dos según su enfoque de estudio (figura 2).³⁸ El *aprendizaje no supervisado* utiliza algoritmos para analizar y agrupar conjuntos de datos no etiquetados. Estos algoritmos descubren patrones ocultos en los datos sin necesidad de intervención humana (no están supervisados). Los modelos de aprendizaje no supervisados se utilizan para tres tareas principales: agrupación, asociación y reducción de la dimensionalidad. Estos métodos moderan la similitud molecular desde los descriptores, pero no usa los valores de actividad en la derivación de un modelo.

El *Aprendizaje Supervisado* se define por el uso de conjuntos de datos etiquetados para

entrenar o "supervisar" algoritmos que clasifiquen datos (clasificación) o predigan resultados de forma precisa (regresión). La propiedad[†] Y se predice con base en ejemplos conocidos denominados elementos, donde cada elemento i tiene asociada la propiedad $Y(i)$ y un grupo de k características $X(i,j)$, $j=1$ hasta k . En términos de QSAR, la variable dependiente Y es una actividad biológica (CE_{50} , CL_{50} o CC_{50}), los elementos son los compuestos químicos y las características son los descriptores moleculares; en este enfoque se utiliza la información de los descriptores para generar el modelo predictivo.

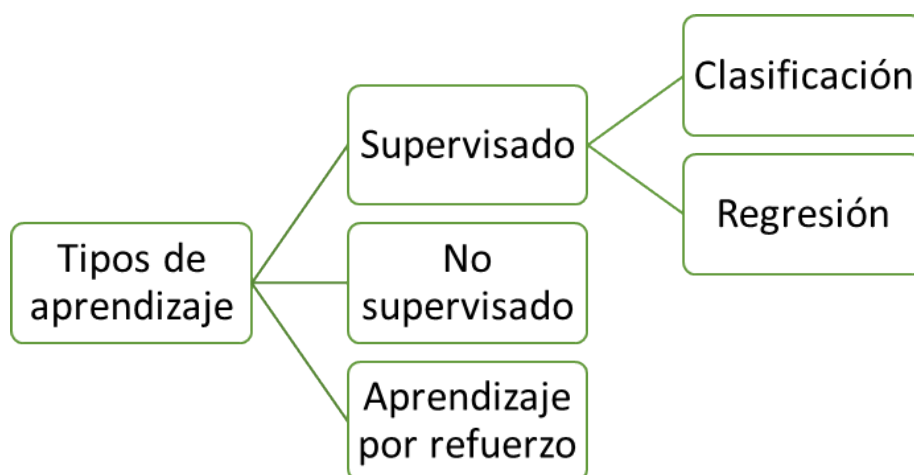


Figura 2. Tipos de aprendizaje automatizado.

Algoritmos usados en aprendizaje automatizado

Existen diferentes algoritmos de clasificación y regresión utilizados en el aprendizaje automatizado,³⁹ en la [figura 3](#) se muestra una clasificación de estos algoritmos de acuerdo con el tipo de enfoque: lineal o enfoque no lineal.⁴⁰

[†] En inglés se usa el término "endpoint" para referirse a las propiedades que se desean predecir o calcular.

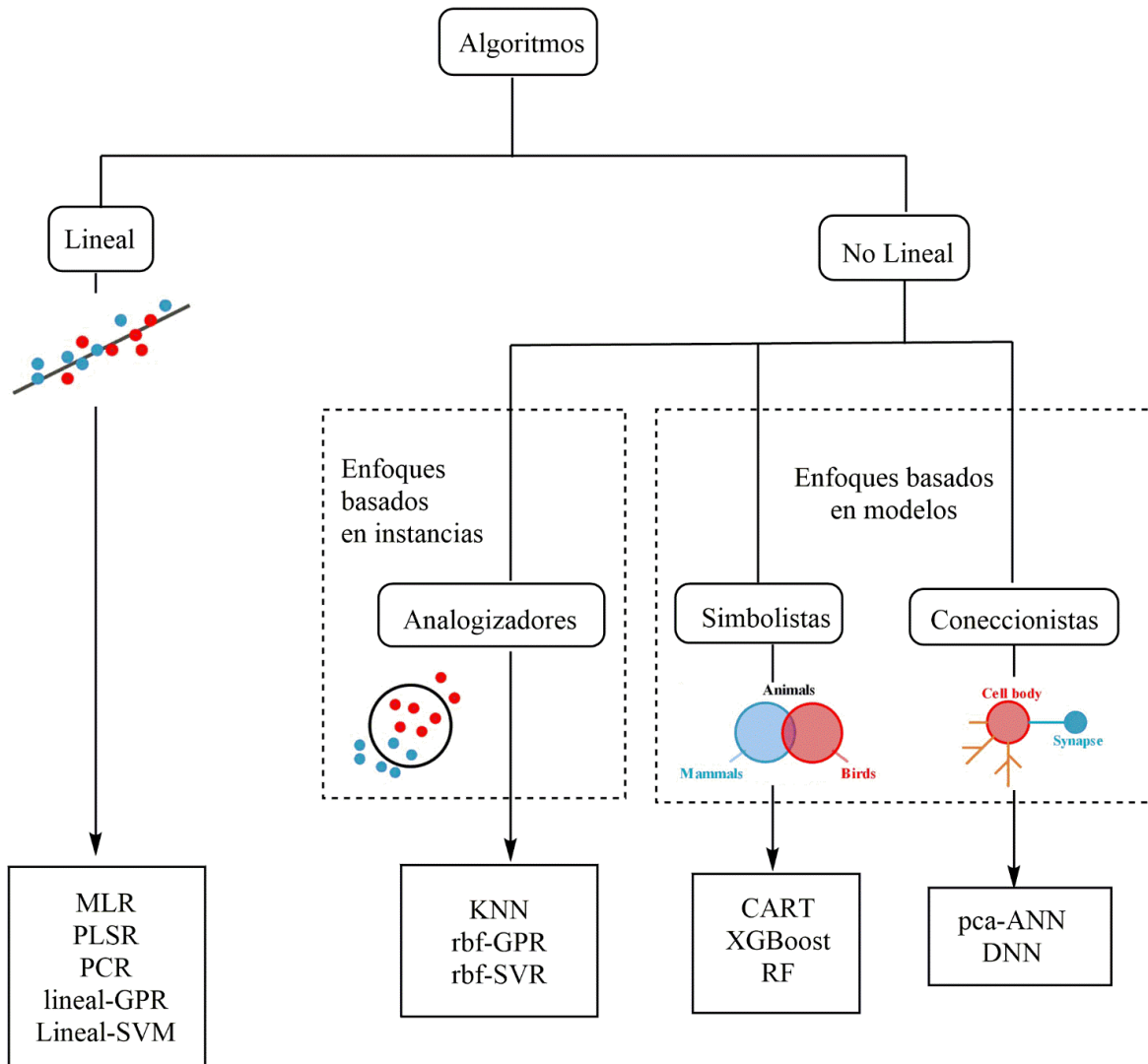


Figura 3. Clasificación de algoritmos utilizados en aprendizaje automatizado. Adaptado de: Wu, Z.; Zhu, M.; Kang, Y.; Leung, E. L.-H.; Lei, T.; Shen, C.; Jiang, D.; Wang, Z.; Cao, D.; Hou, T. Do We Need Different Machine Learning Algorithms for QSAR Modeling? A Comprehensive Assessment of 16 Machine Learning Algorithms on 14 QSAR Data Sets. *Brief Bioinform* **2021**, 22 (4), 1–17. <https://doi.org/10.1093/bib/bbaa321>

Las aproximaciones más simples corresponden a los enfoques lineales donde existe una relación lineal entre las características de los elementos y la propiedad a determinar. La técnica de Regresión Lineal Múltiple (MLR, por sus siglas en inglés, *Multiple Linear Regression*) es el enfoque más común en la construcción de modelos QSAR, genera ecuaciones predictivas de buena interpretabilidad, que tienen la forma:

$$\text{Actividad} = \alpha(\text{descriptor}_1) + \beta(\text{descriptor}_2) + \dots\gamma$$

Para evitar relaciones de insignificancia estadística uno debería aplicar las siguientes dos

reglas.⁴¹

- Por cada descriptor se utilizan 5 compuestos.
- Los descriptores no deberían estar inter-correlacionados: el coeficiente de correlación entre los descriptores no debería ser > 0.5 .

Los algoritmos de aprendizaje de máquina con enfoque lineal como MLR y mínimos cuadrados parciales (PLS, por sus siglas en inglés, *Partial Least Square*) han sido usados para el desarrollo de modelos con grupos de moléculas relativamente pequeños. Con la aplicación de los ensayos HTS (*High Throughput Screening*, cribado de alto rendimiento en castellano) en el desarrollo de fármacos, grandes cantidades de datos de actividad han sido generados, por lo que fue necesario el desarrollo de algoritmos de ML no lineales para generar nuevos modelos predictivos ADMET y QSAR.⁴² Ejemplos de estos algoritmos son: *k*-vecinos más cercanos (kNN, *k-Nearest Neighbors* en inglés), Bosque Aleatorio (RF, *Random Forest*), Máquinas de Soporte Vectorial (SVM, *Support Vector Machine*) y Redes Neuronales Artificiales (ANN, *Artificial Neural Network*).

Entre otros algoritmos lineales utilizados en la generación de modelos QSAR se encuentran Regresor con Mínimos Cuadrados Parciales (PLSR, *Partial Least Squared Regressor*, en inglés), Regresor con Componentes Principales (PCR, *Principal Component Regressor*, en inglés) y algunos más recientes como Regresor Lineal de Procesos Gaussianos (Linear-GPR, *Linear Gaussian Process Regressor*, en inglés) y Máquinas de Soporte Vectorial Lineales (Linear-SVM, *Linear Support Vector Machine*, por sus siglas en inglés).⁴³ En la práctica son más utilizados los enfoques no lineales que pueden ser divididos en dos:⁴⁴

- Enfoques basados en instancias: realizan predicciones por comparación de nuevos puntos de datos con aquellos que están incluidos en el grupo de datos usados para entrenar al modelo (grupo de entrenamiento).
- Enfoques basados en modelos: realizan las predicciones confiando en un modelo definido por un conjunto de parámetros que se infirieron durante el entrenamiento.

En los enfoques basados en instancias la clave del aprendizaje es reconocer la similitud

entre las moléculas y con ello sugerir propiedades de moléculas similares. Si dos moléculas tienen características moleculares similares, aquellas pueden tener similares perfiles fisicoquímicos y biológicos similares. Este grupo incluye algoritmos como KNN y aquellos basados en el uso de kernel como rbf-GPR (*radial basis function - Gaussian Process Regressor*, por sus siglas en inglés) and rbf-SVM (*radial basis function - Support Vector Machine*).

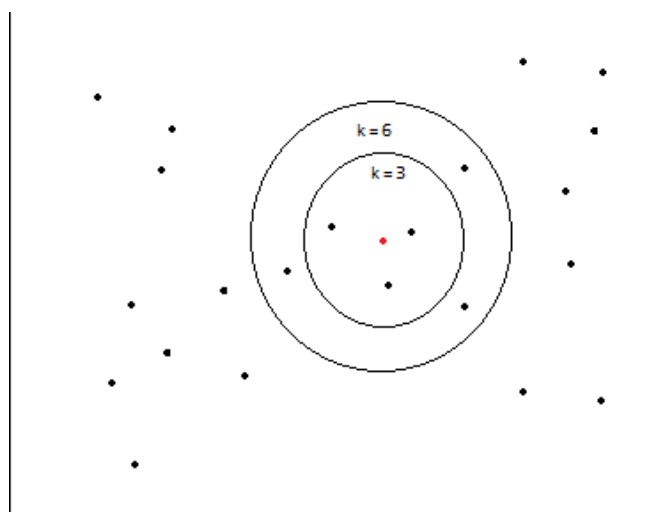


Figura 4. Representación del algoritmo KNN.

El algoritmo KNN basa su predicción de actividad de un compuesto de prueba sobre su vecindad, a un cierto número (k) de compuestos vecinos más cercanos, en un espacio descriptor multivariado. En una serie de compuestos con actividad biológica conocida, se puede utilizar el algoritmo KNN para realizar predicción de valores de actividad biológica mediante regresión; en este caso si $k = 3$, el valor de actividad biológica del compuesto desconocido será predicho como el promedio de la actividad biológica de los tres elementos conocidos, más cercanos a un compuesto desconocido del que se desea predecir sus propiedades en el espacio descriptor como se muestra en la [figura 4](#). Las Máquinas de Soporte Vectorial (SVMs, *Support Vectorial Machines*) fueron originalmente diseñadas como un método de clasificación utilizando avances matemáticos para posicionar un hiperplano para definir y separar dos o más clases, Vapnik⁴⁵ extendió el algoritmo para regresión por el uso de una función de pérdida insensible (ϵ). Este enfoque

se ha convertido muy popular en estudios QSAR.

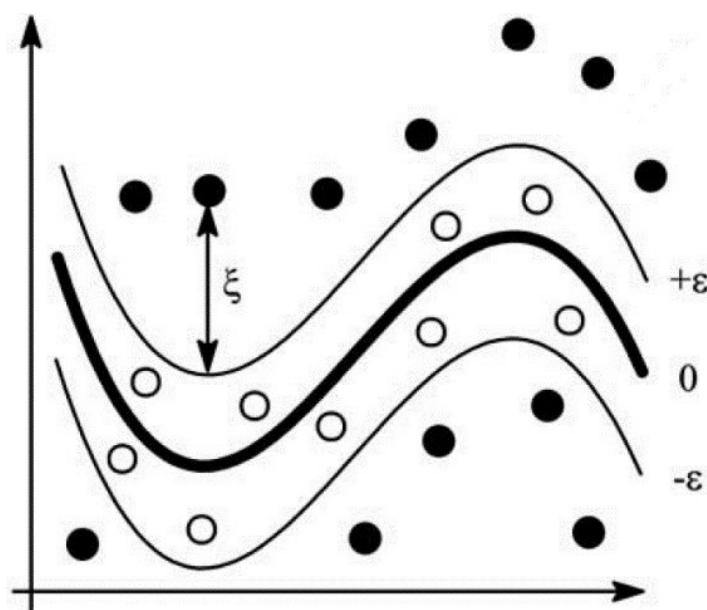


Figura 5. Representación del algoritmo SVR.

Utilizando los patrones de entrenamiento, el regresor SVM genera un modelo representando como un tubo con radio ε ajustado a los datos. El tubo de regresión está bordeado por los hiperplanos denotados con $+\varepsilon$ y $-\varepsilon$ como se muestra en la [figura 5](#). Los patrones situados entre esos hiperplanos tienen una residual (valor absoluto para la diferencia entre valores calculados y experimentales) menor a ε , y en la regresión por SVM el error de esos patrones es considerado cero, y ellos no contribuyen a la penalización general. Los valores que se encuentran fuera del tubo de regresión tienen una residual más grande que ε y por lo tanto ellos tienen un valor de penalización diferente de cero; la penalización para cada punto es proporcional a la distancia (ξ) al borde del tubo de regresión.

Los algoritmos de árbol de decisión se utilizan en clasificación y regresión a través de la construcción de muñones de decisión; cada nodo interno del árbol representa una condición. Los árboles de decisión dividen recursivamente a los elementos del grupo de entrenamiento con particiones binarias utilizando un grupo de reglas que maximizan la separación de elementos, como se muestra en la [figura 6](#), y cuentan con un mecanismo de

poda para mantener la altura mínima del árbol evitando el sobreajuste. Cada hoja denota la función de regresión a utilizar, los coeficientes de esta función de regresión serán las características que guiaron el camino hacia esa hoja.

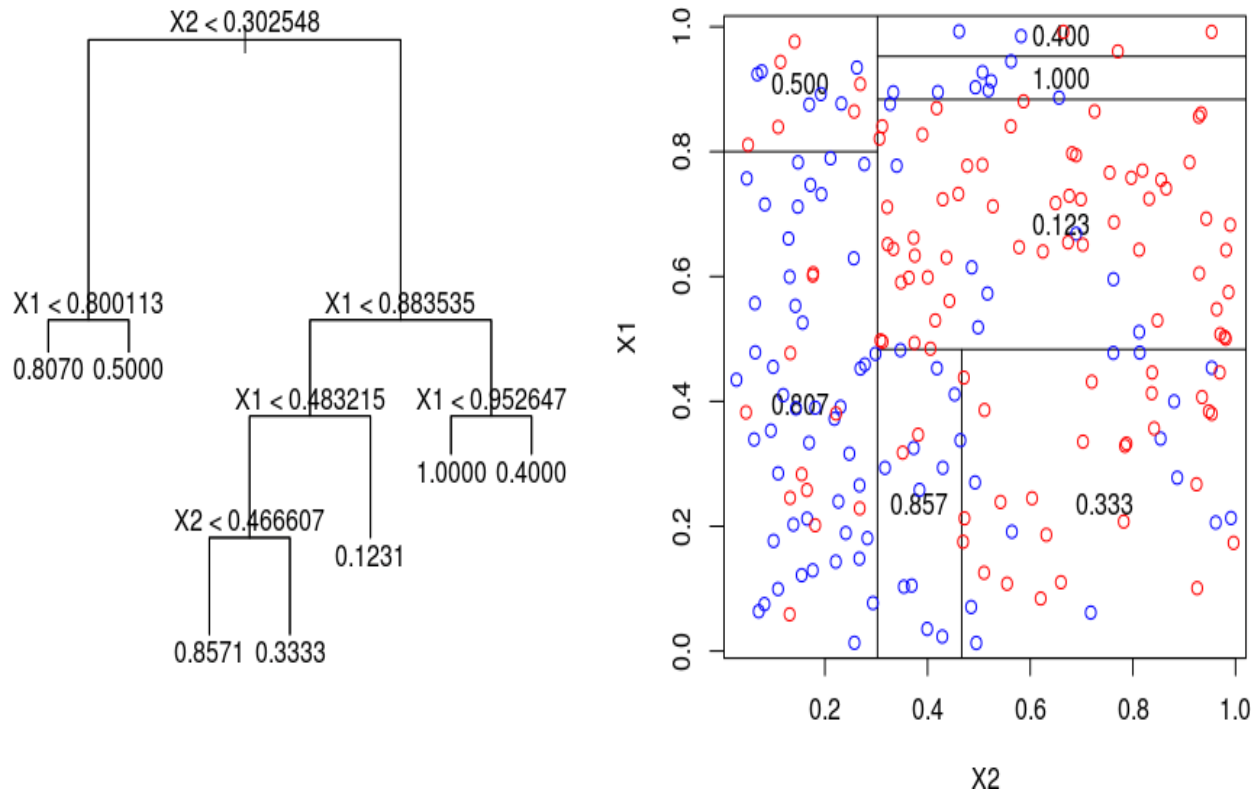


Figura 6. Representación del algoritmo de árbol de decisión.

En cada iteración del método, se evalúan diferentes particiones (evaluando todo el conjunto de datos) y se elige la mejor. La división de los datos genera como salida del método una estructura de árbol, donde cada nodo representa una de las variables de entrada (descriptor). Cada hoja del árbol representa un valor de la variable de destino. Es decir, el valor pronosticado de la variable de destino se obtiene por el camino recorrido desde la raíz hasta una hoja del árbol.⁴⁶

El algoritmo de bosque aleatorio (RF, *Random Forest* en inglés) operan mediante la construcción de una multitud de árboles de decisión en el momento del entrenamiento.⁴⁷

El "bosque" hace referencia a una colección de árboles de decisión no correlacionados, que luego se fusionan para reducir la varianza y crear predicciones de datos más precisas.

Este conjunto arbitrario de árboles de decisión independientes se prueba en conjuntos de datos aleatorios que tienen el mismo número de variables seleccionadas al azar, sin realizar ninguna poda. Para las tareas de regresión se devuelve la predicción promedio de los árboles individuales. Los bosques de decisión aleatorios corrigen el hábito de los árboles de decisión de sobreajustarse a su conjunto de entrenamiento.

Validación del modelo

Una vez que el modelo se desarrolla es necesario saber si éste es estadísticamente relevante, la guía de la OCDE enfatiza la validación estadística de modelos en términos de bondad de ajuste, robustez y predicción.⁴⁸

La calidad de un modelo QSAR puede ser caracterizado estadísticamente por el cálculo de diversos parámetros de calidad y de validación. El propósito de tener diferentes parámetros es juzgar la precisión de la predicción, es decir, la determinación de valores cercanos entre valores experimentales y valores predichos por el modelo. La bondad de ajuste se mide por el cálculo de métricas como el coeficiente de determinación (r^2) o de correlación (r) usando el grupo de entrenamiento, es decir, el grupo de datos que se usó para desarrollar el modelo. La robustez se refiere a la estabilidad cuando se realiza una perturbación al modelo, es decir, la eliminación de datos del grupo de entrenamiento y el redesarrollo del modelo. La predicción externa se refiere a la calidad predictiva obtenida por usar un grupo de compuestos químicos que no se utilizaron durante el desarrollo del modelo. Las métricas de validación interna incluyen coeficientes de determinación (r^2), de técnicas *Leave-One-Out* (q^2_{CV-LOO}) y *Leave-Many-Out* (q^2_{CV-LMO}), así como la prueba de Y-Aleatorización (*Y-scrambling*, en inglés), mientras que la validación externa incluye coeficientes (q^2_{ext}). Se pueden usar otras métricas como el error absoluto promedio (MAE, *Mean Absolute Error*) para checar la calidad y eficiencia de los modelos desarrollados.⁴⁹

Dominio de Aplicabilidad

La predictibilidad de un modelo QSAR es enmarcada por la naturaleza de las moléculas en el grupo de entrenamiento. El dominio de aplicabilidad es la delimitación cuantitativa del espacio químico donde las predicciones son confiables. En este trabajo el dominio de aplicabilidad es definido usando el método del apalancamiento. Los valores del apalancamiento, h_i , son calculados usando la siguiente relación donde X es la matriz de descriptores del modelo y X^T es la matriz transpuesta, del grupo de entrenamiento y x_i es el vector de descriptores para una molécula de consulta y x_i^T es el vector transpuesto.⁵⁰

$$h_i = x_i^T (X^T X)^{-1} x_i$$

Básicamente, los valores de apalancamiento para cada compuesto son proporcionales a la distancia desde el espacio experimental del modelo (que es el centroide del grupo de entrenamiento). Es adecuado para evaluar explícitamente el grado de influencia que tiene un compuesto en particular sobre el grupo de entrenamiento. Los compuestos de bajo apalancamiento tienen menor influencia a la hora de construir el modelo, mientras que los compuestos de alto apalancamiento reforzarían el modelo, aumentando el espacio químico cubierto por el modelo. Los compuestos por encima del umbral[‡] h_{max} están lejos del espacio químico explorado y, por la tanto, la actividad biológica calculada no será confiable.⁵¹

Los resultados son representados en una gráfica de Williams (apalancamiento contra residuales estandarizadas), donde los puntos que salen del dominio de aplicabilidad pueden ser detectados visualmente. La gráfica de Williams es una representación del espacio químico abarcado por el modelo. El diagrama de Williams de Gissi *et al.*⁵² muestra que los dos compuestos del grupo de prueba caracterizados por los más largos residuales, también tienen valores de h mayores a h_{max} , es decir, los puntos con las etiquetas 496 y 472 en la [figura 7](#).⁵³

[‡] $h_{max} = 3p/n$; donde p número de características y n el número de molécula del grupo de entrenamiento.

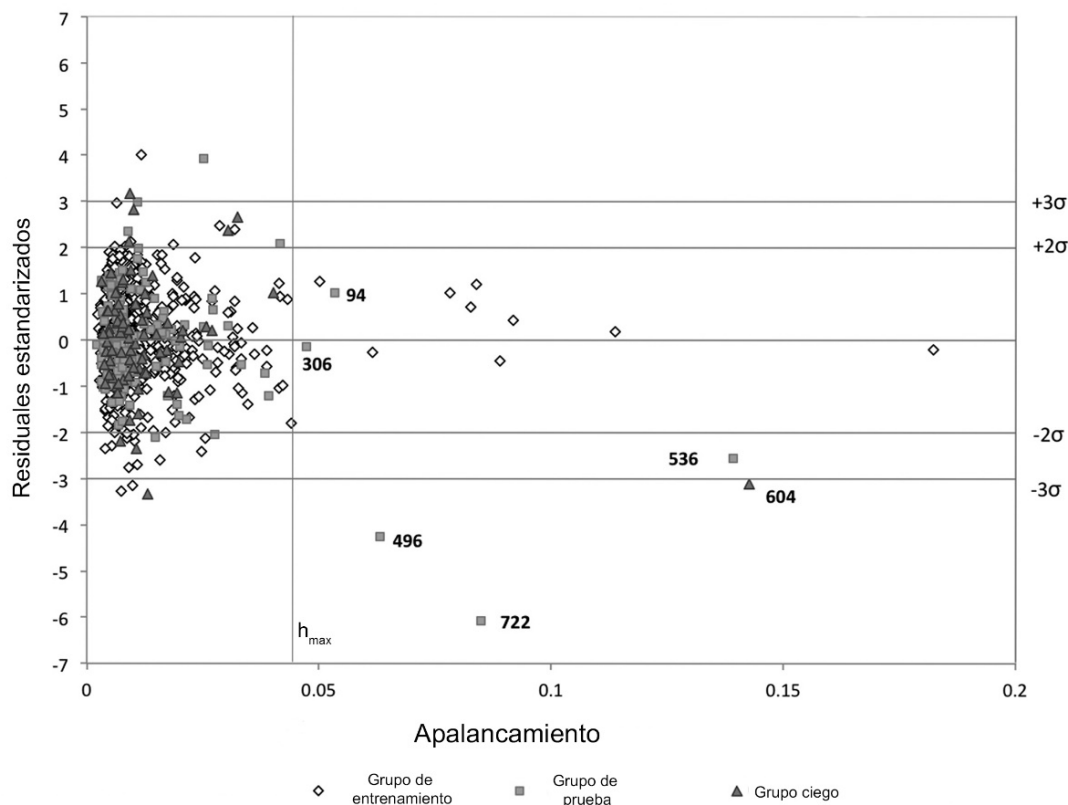


Figura 7. Ejemplo de DA basado en apalancamiento para un modelo de predicción del factor de bioconcentración. Los valores de apalancamiento calculados a partir de nueve descriptores del modelo para los compuestos del grupo de entrenamiento (rombos blancos) se trazan hacia los residuos estandarizados correspondientes. Los productos químicos incluidos en dos conjuntos de validación diferentes (cuadrados y triángulos grises) que violan el umbral h_{max} se marcan como fuera del dominio de aplicación.⁴⁰ Tomada y adaptada de Gissi, A.; Gadaleta, D.; Floris, M.; Olla, S., Carotti, A., Novellino, E.; Benfenati, E.; Nicolotti, O. An alternative QSAR-based approach for predicting the bioconcentration factor for regulatory purposes. *ALTEX - Alternatives to animal experimentation* **2014**, 31(1), 23–36. <https://doi.org/10.14573/altex.1305221>.

Acoplamiento molecular

La técnica de acoplamiento molecular (*docking* en inglés) es una simulación computacional útil para estudiar el modo de unión, y sus energías asociadas[§], entre un ligando y un sitio activo de algún receptor biológico. La técnica consiste en colocar el ligando en diferentes orientaciones y conformaciones dentro del sitio de unión para calcular geometrías y energías de unión óptimas ayudando a predecir la posición y orientación ideales, y la

[§] Energía que no tiene significado fisicoquímico, salvo para asignar una jerarquía.

afinidad dentro del sitio de reconocimiento.⁵⁴ Las poses finales de mayor puntaje (*S*, *Score* en inglés), junto con sus puntajes y energías de conformación, se escriben en una base de datos donde están listas para un análisis posterior.⁵⁵

Es una herramienta importante en el diseño de fármacos debido a que permite predecir sitios de enlace de ligandos en receptores, elucidar el mecanismo de reconocimiento molecular y realizar cribado de grupos de ligandos por simulación virtual de procesos espontáneos,⁵⁶ lo que contribuye a elucidar mecanismos de acción de fármacos y disminuir costos en la reducción de experimentos.

La información del receptor, que suelen ser proteínas, se suele obtener por combinación de técnicas de biología molecular, de difracción de rayos X y de resonancia magnética nuclear (RMN) y se suele depositar en el Protein Data Bank (PDB, *Banco de Datos de Proteínas* en castellano).

Existen tres tipos de acoplamiento molecular: rígido, semiflexible y flexible⁵⁷, en el acoplamiento molecular rígido la conformación del sistema de acoplamiento molecular, es decir receptor y ligando, permanecen inalterados, es adecuado para sistemas grandes como interacciones proteína-proteína y proteína ácido nucleico. Además, el acoplamiento molecular rígido no requiere de mucho calculo, haciéndolo el método más simple de realizar.

Durante el proceso de acoplamiento molecular semiflexible, la conformación del receptor permanece sin cambios, mientras que la conformación del ligando es alterada. Este sistema toma en cuenta la influencia de cambios estructurales en el ligando y es adecuado para el acoplamiento molecular de pequeñas moléculas y macromoléculas, como proteínas o ácidos nucleicos.

Durante el proceso de acoplamiento molecular flexible, el procedimiento permite a las conformaciones del sistema, receptor y ligando, ser modificados a lo largo de la simulación. Esto complica en gran medida el cálculo computacional, pero ayuda a realizar cálculos más precisos.⁵⁸

La técnica de acoplamiento molecular es una técnica que se ha implementado desde los años 70's con el grupo de investigación de Levinthal,⁵⁹ quien usó esta técnica computacional para predecir las posibles conformaciones de las fibras de hemoglobina. En 1982 surgió DOCK, el primer programa para realizar simulaciones y hasta 2016 había más de 60 diferentes programas computacionales para realizar este tipo de estudios entre los que destacan AutoDock, Autodock Vina, ICM, GOLD, MOE, Schrodinger/Maestro, FlexX, Surflex y Affintiy.⁶⁰ A pesar de la gran variedad de programas especiales para realizar estudios de acoplamiento molecular, aún quedan muchos problemas sin resolver como el tratamiento adecuado para la flexibilidad de una proteína, la solvatación y la predicción precisa de la predicción de las afinidades de enlace.⁶¹

La aplicación Dock de MOE⁶² busca modos de unión favorables entre ligandos pequeños a medianos y un objetivo macromolecular no demasiado flexible, generalmente una proteína. Para cada ligando se generan y puntúan varias configuraciones llamadas **poses**. El flujo de trabajo de Dock comienza con un receptor y ligando preparados cargados en MOE. El algoritmo se divide en etapas, como se ilustra en la [figura 8](#).

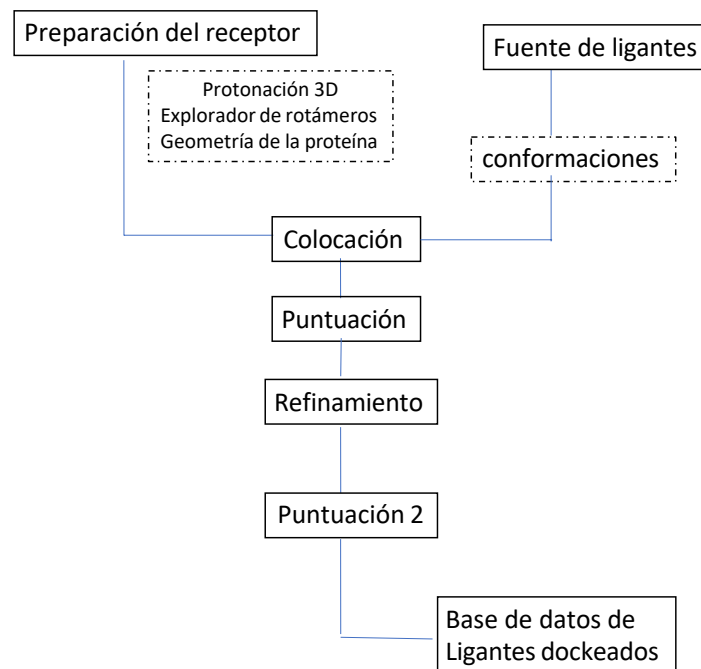


Figura 8. Flujo de trabajo en MOE 2019.

Influenza

Se conocen tres tipos inmunológicos de virus designados A, B y C. La influenza tipo A posee una gran variabilidad antigénica y es la causa de la mayor parte de los casos de epidemias de influenza. La influenza tipo B puede mostrar cambios antigénicos y a veces es causa de epidemias. La influenza tipo C es antigénicamente más estable y en individuos inmunocompetentes sólo es causa de enfermedad leve. Las partículas virales de la influenza, por lo general, son esféricas y de 100nm de diámetro aproximadamente. Los genomas del RNA de cadena sencilla, de sentido negativo de los virus de la influenza A y B poseen ocho segmentos separados. Las partículas del virus de la influenza contienen nueve proteínas estructurales diferentes como se muestra en la [figura 9](#).

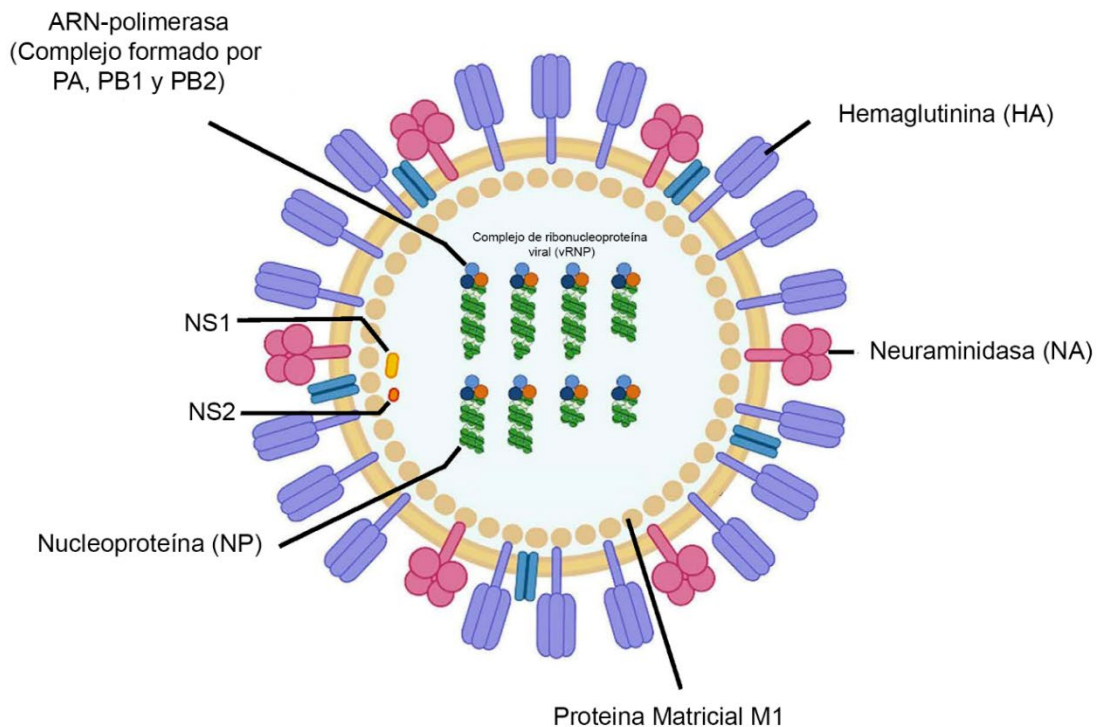


Figura 9. Estructura del virus de la influenza. Tomada de internet y adaptada al castellano; búsqueda en Google (acceso 06-04-2023. <https://www.virology.ws/2009/04/30/structure-of-influenza-virus/>)

La Nucleoproteína (NP) se reúne con el ARN viral para formar una estructura de

ribonucleoproteína (RNP) de 9 nm de diámetro que asume una configuración helicoidal y forma la nucleocápside viral. Tres grandes proteínas (PB1, PB2 y PA) se unen a la RNP viral y se encargan de la transcripción y replicación del ARN. La proteína de la matriz (M₁), que forma una capa por debajo de la cubierta lipídica del virus, es importante en la morfogénesis de la partícula y es un componente principal del virión (casi 40 % de la proteína viral).

La partícula viral está rodeada por una cubierta lipídica derivada de las células. Dos glucoproteínas codificadas por el virus, la Hemaglutinina (HA) y la Neuraminidasa (NA), se introducen en la cubierta y quedan expuestas como espigas de casi 10 nm de longitud sobre la superficie de la partícula. Estas dos glucoproteínas de superficie son los antígenos importantes que determinan la variación antigénica de los virus de la influenza y la inmunidad del huésped. La HA representa casi 25 % de la proteína viral y la NA cerca del 5 %. La proteína del canal iónico M2 y la Proteína NS2 también se encuentran en la cubierta, pero solo en unas pocas copias por partícula.⁶³

La infección viral se inicia con el enlace de la HA a un receptor de membrana que contienen residuos terminales de ácido neuramínico. El virus se internaliza en un endosoma y posteriormente, las ribonucleoproteínas se liberan en el citoplasma. Subsecuentemente, estas pasan por los poros nucleares al núcleo para la transcripción (síntesis de ARN mensajero viral) y la replicación que son mediadas por la ARN polimerasa viral. La síntesis de las proteínas virales la lleva a cabo la célula. Los componentes proteicos necesarios para la formación de las ribonucleoproteínas y la nucleocápside se exportan al núcleo celular. Las proteínas virales de la envoltura se transportan y modifican en el aparato de Golgi para finalmente ser insertadas en la membrana celular. El ensamblaje final de la partícula viral es un proceso no bien comprendido. La liberación de la partícula viral produce lisis celular.

La influenza se transmite de persona a persona mediante gotas de saliva producidas al toser o estornudar, al ser inhaladas depositan un inóculo infeccioso en el epitelio de las vías respiratorias, o bien por contacto con manos o superficies contaminadas. Una vez que los virus se han implantado en el epitelio de las vías respiratorias comienzan a replicarse y diseminarse en el tracto respiratorio, causando la descamación de las células ciliadas y de

las células secretoras de moco. La multiplicación viral lleva a la lisis de estas células con la liberación de antígenos virales que atraen a macrófagos y a linfocitos.

El virus de la influenza tipo A también se puede encontrar en cerdos, caballos, aves silvestres y aves de corral.⁶⁴ Es posible que el virus de la influenza se pueda transmitir entre diferentes especies, dando lugar a recombinaciones genéticas. Por la alta demanda de estos animales en la alimentación, grandes poblaciones se juntan en pequeños espacios dando lugar a transmisiones en las que el hombre puede estar involucrado dando origen a híbridos virales con potencial de diseminación global y, probablemente, con elevada patogenicidad por la ausencia de inmunidad previa en las poblaciones humanas (figura 10).⁶⁵

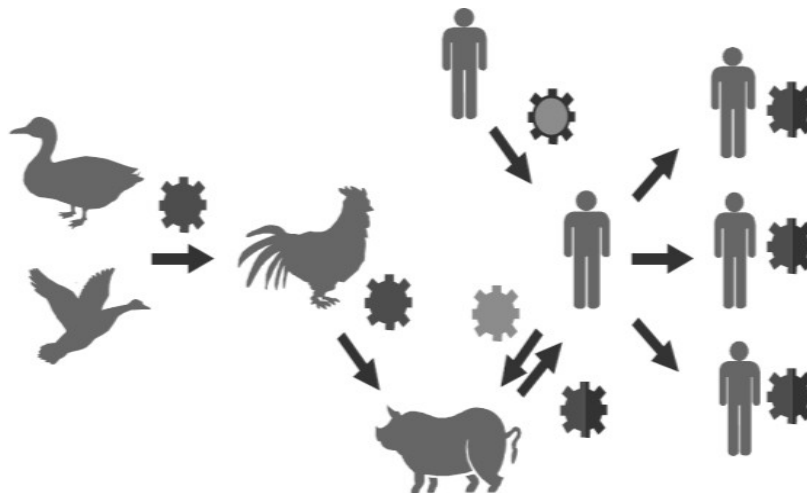


Figura 10. Eventos de recombinación genética entre aves cerdos y humanos, probable mecanismo de génesis del virus AH1N1 de la influenza en 2009. Tomado de: Franco-Paredes C, Río C, Peter Carrasco P, et al. Respuesta en México al actual brote de influenza AH1N1. *salud pública mex.* **2009**, 51(3), 183-186.

Las pandemias más importantes ocasionadas por el virus de la influenza A en los últimos dos siglos han sido la influenza española (1918-1919)** , la influenza asiática (1957-1958), la influenza de Hong Kong (1968-1969) y México en 2009, ocasionada por AH1N1.⁶⁶ (figura

** Cabe mencionar que la pandemia de 1918, que se presentó al final de la primera guerra mundial, ocasionó más muertes que la guerra misma (más de 20 millones). Sin embargo, no fue sino 15 años después cuando se logró el primer aislamiento de virus de influenza tipo A, en 1936 el del tipo B y hasta 1950 el aislamiento del tipo C.

11). En México se registran casos de influenza estacional cada año, causados generalmente por el virus de la influenza AH3N2 e influenza B, pero en 2009 se presentó en México una pandemia provocada por el virus de la influenza porcina AH1N1.⁶⁷

El 21 de abril de 2009 una publicación dio a conocer dos casos de infección respiratoria aguda, de pacientes provenientes de California, ocasionados por un nuevo tipo de influenza A negativo a los subtipos conocidos H1N1 (estacional), H3N2 y H5N1. El 17 de abril de ese año en Atlanta, E.U.A. se determinó que se trataba de un nuevo virus de origen porcino, resistente a amantadina y rimantadina y con dos genes de linaje euroasiático. Se trataba con una recombinación entre cepas de influenza aviar y cepas de influenza humana, teniendo al cerdo como huésped intermediario.

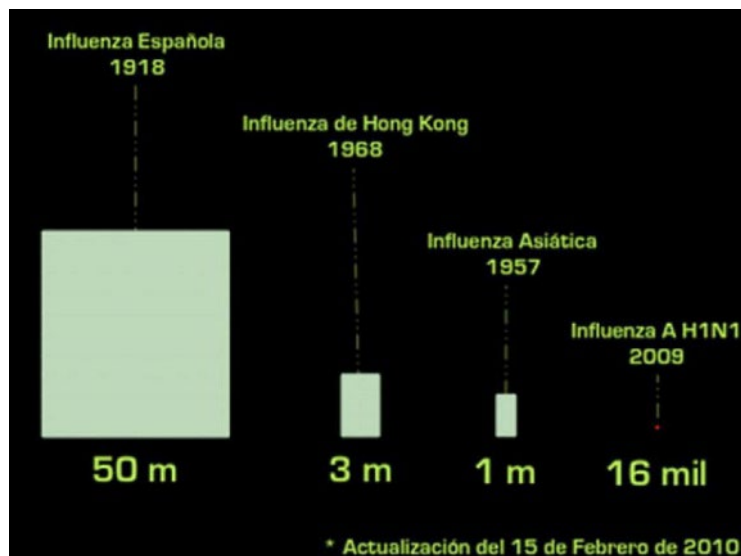


Figura 11. Comparación de la mortalidad de algunas pandemias de influenza (m representa millones). Tomado de: López Cervantes M, García Moreno J, Pacheco Domínguez RL, Escamilla Santiago RA, Villanueva Lozano M. La influenza A/H1N1 2009. Una crónica de la primera pandemia del siglo XXI. *Revista Digital Universitaria* 2010, 11(4).

El 23 de abril de ese año se da a conocer la noticia de que el mismo virus, había sido encontrado en muestras de pacientes mexicanos, procesadas en los laboratorios de Winnipeg en Canadá. Además, se habían registrado al menos 20 muertes en pacientes atendidos en hospitales mexicanos con cuadros clínicos de neumonía viral. Con base en esta información se decidió el cierre inmediato de escuelas públicas y privadas en todos los

niveles educativos, y se implementaron diversas medidas para propiciar el aislamiento social. El virus se diseminó rápidamente en E.U.A., Canadá y cruzó los océanos Atlántico y Pacífico para llegar a España, Inglaterra y Nueva Zelanda.⁶⁸

La letalidad de la enfermedad, tan temida en un inicio, resultó ser selectiva (sólo afecta personas con enfermedades subyacentes) y tan pequeña como la influenza estacional. Gran parte de la población mundial fue infectada, pero desarrolló inmunidad. El 3 de junio de 2009 se informa por la OMS la confirmación de 19 273 casos de influenza A humana H1N1 2009 en 66 países.

El 9 de enero del 2023, se notificó a la OMS una infección humana por el virus de la gripe aviar A(H5). Se trata del primer caso notificado de infección humana causada por el virus de la gripe aviar A(H5) en la región de América Latina y el Caribe. La persona afectada es una niña de nueve años residente en una zona rural de la provincia de Bolívar (Ecuador), quien estuvo en contacto con aves de corral domésticas que habían sido adquiridas una semana antes de que le aparecieran los síntomas.⁶⁹

Existen dos medidas disponibles que pueden reducir el impacto de influenza. La quimioprofilaxis que depende de fármacos y la inmunoprofilaxis que emplea vacunas.⁷⁰ Actualmente el tratamiento incluye medicamentos antivirales, que reducen la intensidad de los síntomas, si son administradas dentro de las primeras 48 horas de iniciada la infección. Para contrarrestar las molestias de la influenza existen diversos medicamentos paliativos como el ibuprofeno para el dolor muscular, el acetaminofén para la fiebre y la aspirina que mejora el estado general de los pacientes.

Hasta 2011 había solo dos tipos de antivirales específicos que eran utilizados en la práctica médica contra la influenza. Ellos eran los derivados de Amantadina que obstruyen el conducto de la proteína M2 y los inhibidores de Neuramidasa: Zanamivir (Renerva) y Oseltamivir (Tamiflu). Existe alta preocupación por la resistencia a estos tratamientos. Por ello es la búsqueda de nuevos tratamientos.⁷¹ Ante casos de influenza de origen AH1N1, se ofrecen a nivel institucional Oseltamivir (Tamiflu) o Zanamivir (Renerva), sus estructuras se muestran en la [figura 12](#).

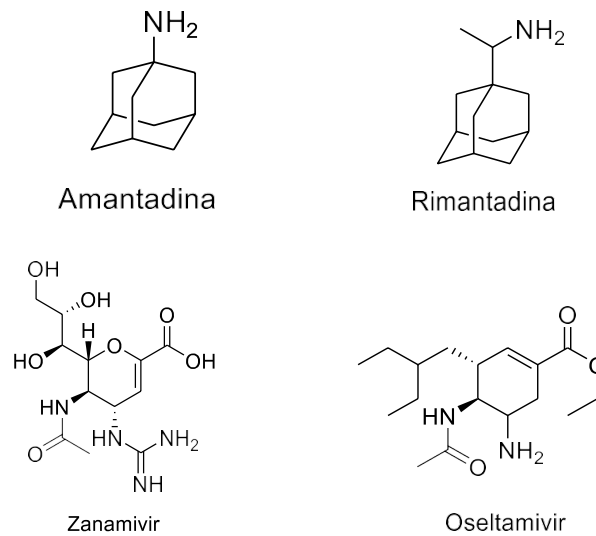


Figura 12. Estructuras de los principales antivirales contra influenza.

Las futuras investigaciones en antivirales son necesarias para minimizar emergencias resultantes por cepas de virus resistentes a fármacos, reducir toxicidad de los fármacos existentes, diseñar antivirales con mayor especificidad basados en conocimientos moleculares dentro de los blancos virales, y el descubrimiento de nuevos antivirales contra virus ante los que no existe algún tratamiento.⁷² Los estudios de virología molecular están avanzando en la identificación de funciones virales específicas que pueden surgir como blancos para terapias antivirales. Las fases de replicación viral en las que se suelen enfocar para el desarrollo de fármacos y terapias incluyen la unión entre el virus y la célula huésped, liberación del genoma viral, síntesis de ácidos nucleicos virales, traducción de proteínas virales, ensamblaje y liberación de partículas virales.⁷³

Entre los principales enfoques para el desarrollo de nuevos antivirales contra influenza A se proponen la inhibición, ya mencionada, de Neuraminidasa (NA) y de endonucleasa cap-dependiente, sin embargo, también se propone el uso de inhibidores de Hemaglutinina (HA),⁷⁴ así como el bloqueo de canales iónicos con nuevos compuestos derivados de amantadano⁷⁵. También se ha reportado el uso de sialidasas (como DAS 181)⁷⁶ que evitan la infección del virus removiendo, de los tejidos epiteliales del hospedero, a los receptores necesarios para que se dé la unión del virus de la influenza A y comience el ciclo viral.

Nucleoproteína de influenza A

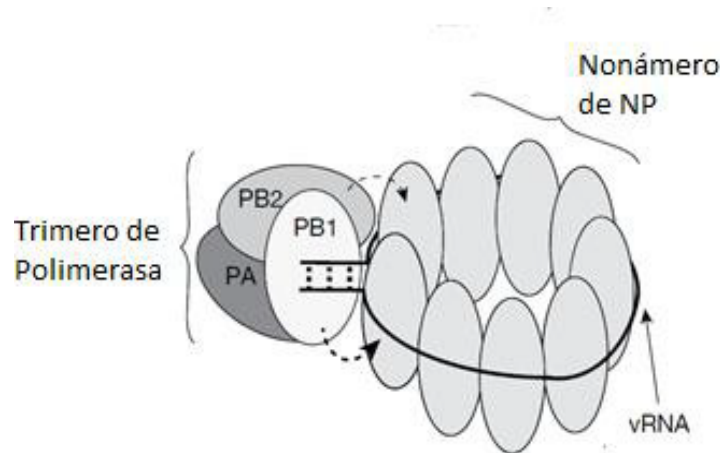


Figura 13. Estructura de la ribonucleoproteína (RNP). Tomado y adaptado de: Cianci, C.; Gerritz, S. W.; Deminie, C.; Krystal, M. Review Influenza Nucleoprotein : Promising Target for Antiviral Chemotherapy. *Antiviral Chemistry & Chemotherapy* 2013, 91, 77–91. <https://doi.org/doi:10.3851/IMP2235>.

La nucleoproteína (NP) es un objetivo atractivo para el tratamiento de la influenza debido a que es la proteína expresada más abundante durante el ciclo de la infección del virus de la influenza con múltiples funcionalidades.⁷⁷ Es fundamental para la síntesis del vARN, la exportación nuclear de las ribonucleoproteínas (RNP), tráfico citoplasmático y el ensamblaje viral. La NP está presente en la forma activa del virus en forma de RNP: nueve unidades de NP se unen al ARN viral junto a las proteínas PA, PB1 y PB2 para formar la estructura de RNP como se muestra en la [figura 13](#).⁷⁸

En la [figura 14](#) se muestra la estructura de Rayos X de la NP, la cual fue caracterizada por Ye *et al.*⁷⁹ Se pueden observar tres dominios principales: cabeza, cuerpo y cola con sus residuos participantes y un surco de unión del ARN viral (ARNv) entre los dominios cabeza y el bucle de cola. Este último surco se muestra en la [figura 15](#) y en la [figura 16](#), y está caracterizado por un residuo Y148, rodeado de los residuos básicos de arginina, el cual se cree que tiene interacción con las nucleobases del RNA viral y puede ser usado como plataforma para el diseño de inhibidores de NP. El bucle de cola tiene una conformación extendida e interactúa con las moléculas vecinas, y es fundamental para la auto oligomerización de NP, a través de diversos tipos de interacciones intermoleculares ([figura 16](#)) incluyendo las interacciones intermoleculares de las hojas β (por ejemplo, $\beta 7$ con $\beta 8$;

$\beta 6$ con $\beta 4$), interacciones hidrofóbicas (por ejemplo, P453 (A) con F420 (B)) y formación de puentes salinos (E 339 (A) con R416 (B)).

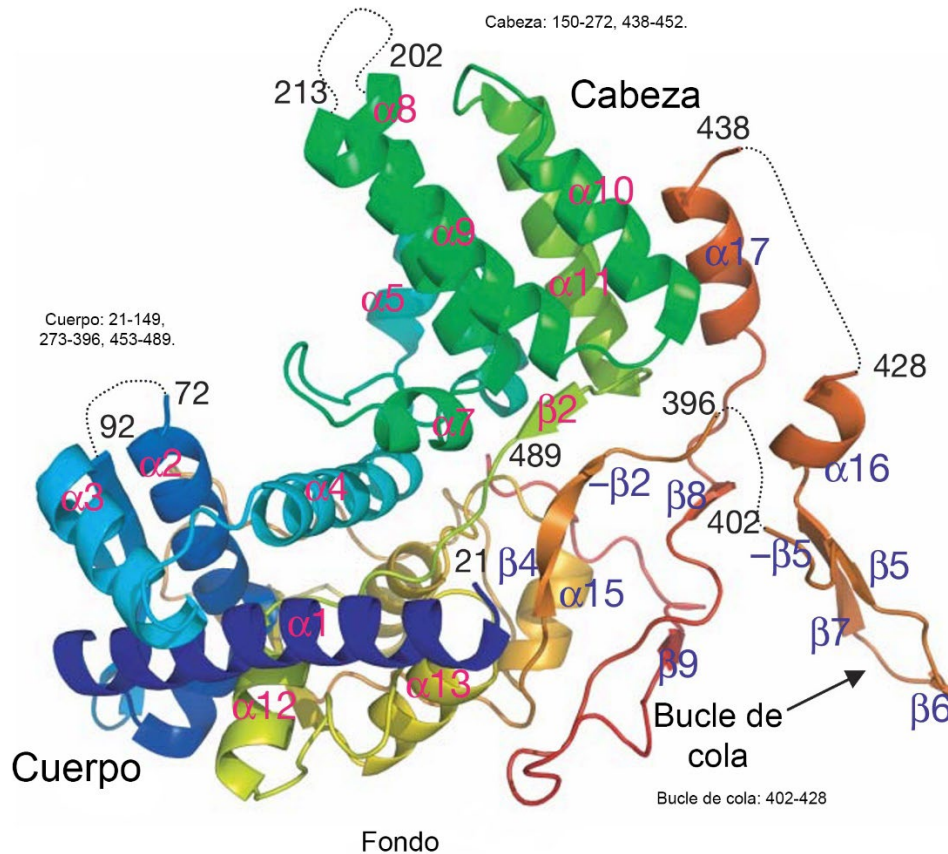


Figura 14. Estructura de NP (código PDB: 2IQH). Tomado de Ye, Q.; Krug, R. M.; Tao, Y. J. The Mechanism by Which Influenza A Virus Nucleoprotein Forms Oligomers and Binds RNA. *Nature* **2006**, *444* (7122), 1078–1082. <https://doi.org/10.1038/nature05379>.

Actualmente existen diversos ligandos dirigidos a la nucleoproteína. Los más estudiados son nucleozina, los derivados de naproxeno⁸⁰, RK424⁸¹, algunos tiazoles 2,4-disustituidos⁸², PPQ-581^{16-b} y S119⁸³ por mencionar algunos. La NP no tiene ninguna contraparte en las células, por lo que se puede usar a esta proteína como objetivo terapéutico con alta especificidad.

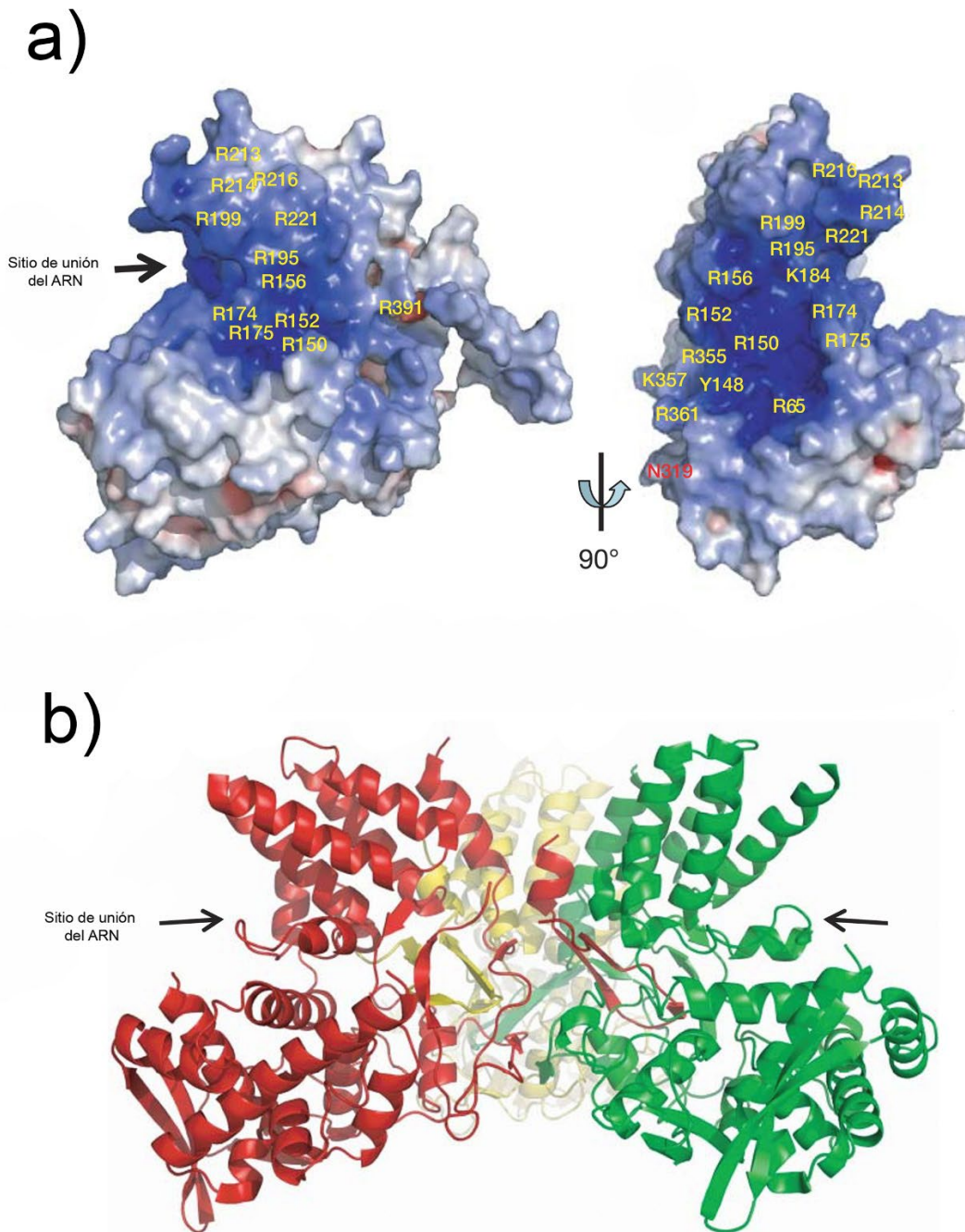


Figura 15. Surco de unión del ARN a NP. **a)** un monómero de NP y sus correspondientes residuos Y148 principalmente. **b)** se muestra el surco de unión del ARN en el trímero de NP. Tomado de Ye, Q.; Krug, R. M.; Tao, Y. J. The Mechanism by Which Influenza A Virus Nucleoprotein Forms Oligomers and Binds RNA. *Nature* **2006**, 444 (7122), 1078–1082. <https://doi.org/10.1038/nature05379>.

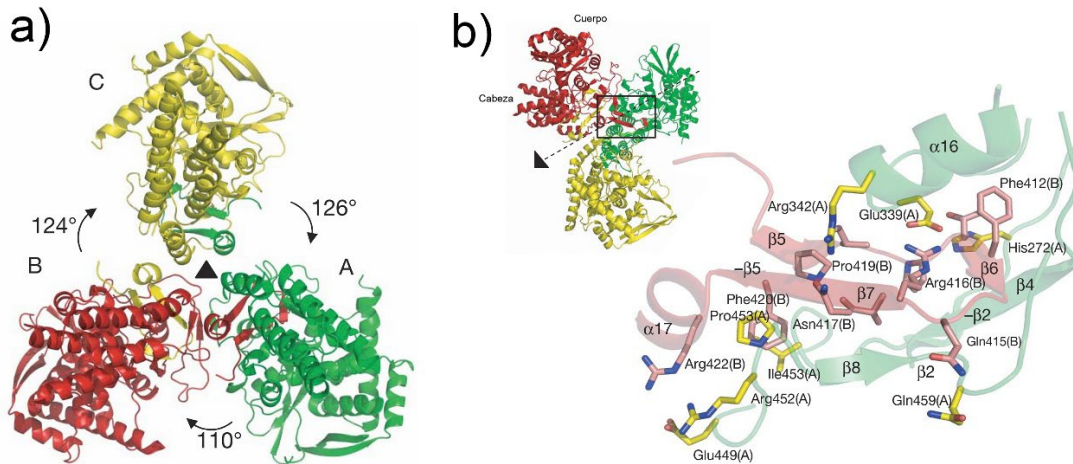
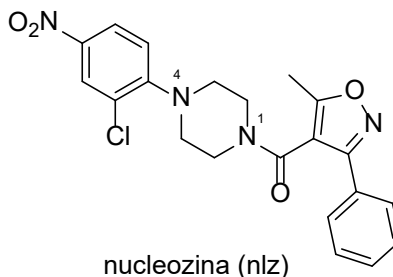


Figura 16. a) el trímero de NP formado por los tres monómeros A, B y C. A en verde, B en rojo y C en amarillo. b) El trímero se genera por las interacciones entre los bucles de cola de los monómeros de NP. Tomado de Ye, Q.; Krug, R. M.; Tao, Y. J. The Mechanism by Which Influenza A Virus Nucleoprotein Forms Oligomers and Binds RNA. *Nature* **2006**, *444* (7122), 1078–1082. <https://doi.org/10.1038/nature05379>.

Nucleozina (nlz)

Desde 2010 se han publicado una serie de amidas derivadas de piperazina con actividad sobre NP, el compuesto hit⁺⁺ fue denominado nucleozina (nlz) [4-(2-cloro-4-nitrofenil)piperazin-1-il]-(5-metil-3-fenil-1,2-oxazol-4-il)metanona y fue reportado por R. Kao y col. ^{16a,c}



En la [tabla 1](#) se muestran la concentración efectiva media (CE₅₀) de la nucleozina fue

⁺⁺ Referente a la frase "Hit to Lead" un enfoque de la química medicinal que se refiere al proceso de llegar a un compuesto líder (con propiedades prometedoras para ser fármaco) partiendo de un conjunto de hits.

evaluada para distintos virus de Influenza mediante PRA (*Plaque Reduction Assay*), mientras que en la [figura 17](#) se muestra la eficacia, en comparación con el Oseltamivir. El compuesto tiene una concentración tóxica media $> 250 \mu\text{M}$ lo que sugiere una amplia ventana terapéutica.

Tabla 1. Concentración Efectiva Media (CE_{50}) de la nucleozina evaluada para diferentes virus de influenza. Tomada y adaptada de Kao, R. Y.; Yang, D.; Lau, L. S.; Tsui, W. H. W.; Hu, L.; Dai, J.; Chan, M. P.; Chan, C. M.; Wang, P.; Zheng, B. J.; Sun, J.; Huang, J. D.; Madar, J.; Chen, G.; Chen, H.; Guan, Y.; Yuen, K. Y. Identification of Influenza A Nucleoprotein as an Antiviral Target. *Nat Biotechnol* 2010, 28 (6), 600–605. <https://doi.org/10.1038/nbt.1638>.

Virus	CI_{50} (μM)
A/WSN/33	0.069 ± 0.003
H3N2	0.16 ± 0.01
Vietnam/1194/04 (H5N1)	0.33 ± 0.04

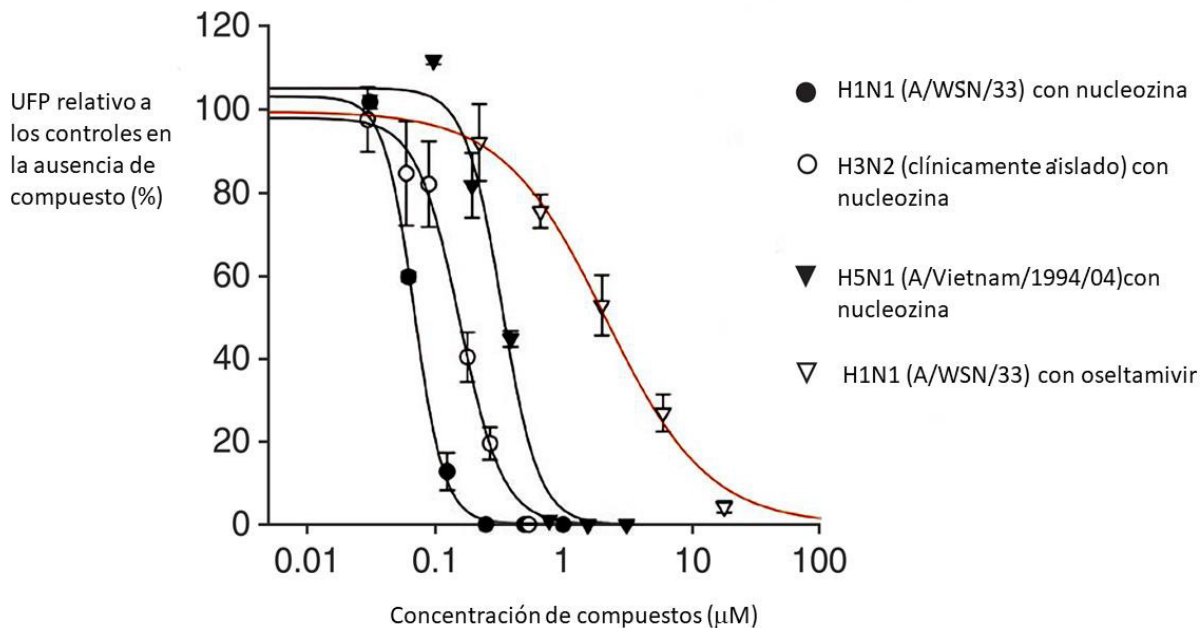


Figura 17. Nucleozina es efectiva contra los virus de influenza humana H1N1, H3N2, y H5N1. Las células MDCK fueron infectadas con diferentes cepas de virus y las actividades antivirales fueron determinados por PRA. Tomada y adaptada de Kao, R. Y.; Yang, D.; Lau, L. S.; Tsui, W. H. W.; Hu, L.; Dai, J.; Chan, M. P.; Chan, C. M.; Wang, P.; Zheng, B. J.; Sun, J.; Huang, J. D.; Madar, J.; Chen, G.; Chen, H.; Guan, Y.; Yuen, K. Y. Identification of Influenza A Nucleoprotein as an Antiviral Target. *Nat Biotechnol* 2010, 28 (6), 600–605. <https://doi.org/10.1038/nbt.1638>.

Estructura de la nlz y sus derivados

En la siguiente sección se describirán las principales modificaciones realizadas a la molécula de nucleozina. Las estructuras son reportadas en código SMILES junto con las actividades biológicas (IC_{50}) la sección de metodología (en la [tabla 12](#)); las IC_{50} están reportadas en unidades de concentración micromolar (μM) y transformadas al $-\log(IC_{50}) = pIC_{50}$. Algunos de los inhibidores derivados de nucleozina corresponden a un esqueleto base que se muestra en la [figura 18](#). Para el caso de la nucleozina (nlz) $R_1 = -Cl$, $R_2 = -NO_2$, $R_3 = -CH_3$, $R_4 = R_5 = R_6 = R_7 = -H$, $Z = N$, $Y = N$ y $X = C$. Para el caso del compuesto 3 reportado por Gerritz *et al.*⁸⁴ (en este trabajo nos referiremos a este compuesto con el nombre Gerritz 3) $R_1 = -Cl$, $R_2 = -NO_2$, $R_3 = -CH_3$, $R_4 = -OCH_3$, $R_5 = R_6 = R_7 = -H$, $Z = N$, $Y = N$ y $X = C$.

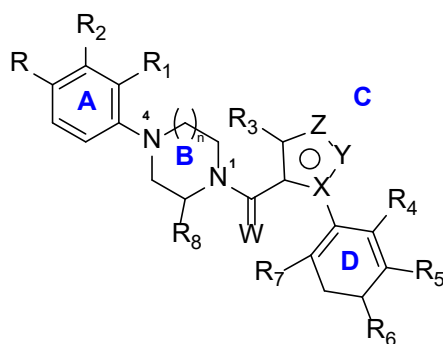
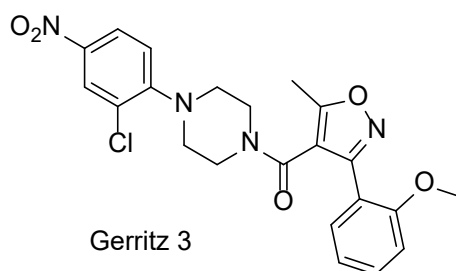


Figura 18. Esqueleto principal de los antivirales derivados de nucleozina.

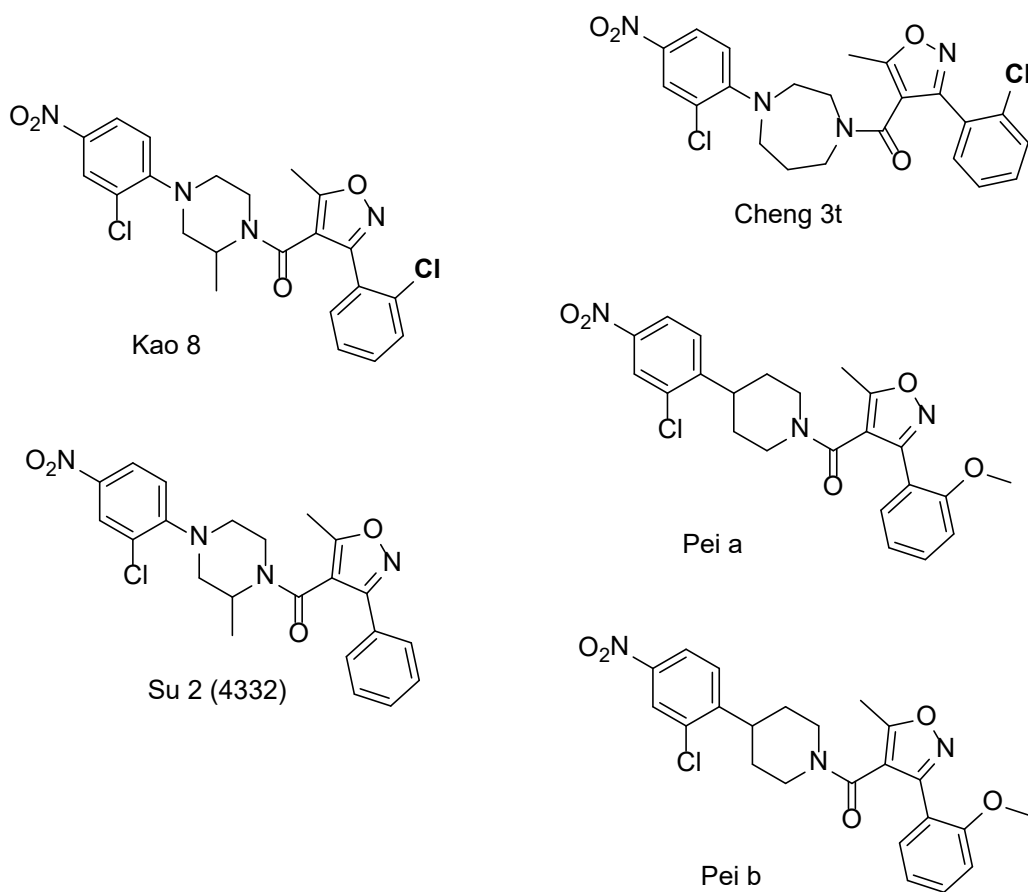


Se han sintetizado análogos donde $W = S$, esa modificación mejora la biodisponibilidad del antiviral en ensayos realizados con Influenza A/Texas/91 (H1N1).⁸⁵

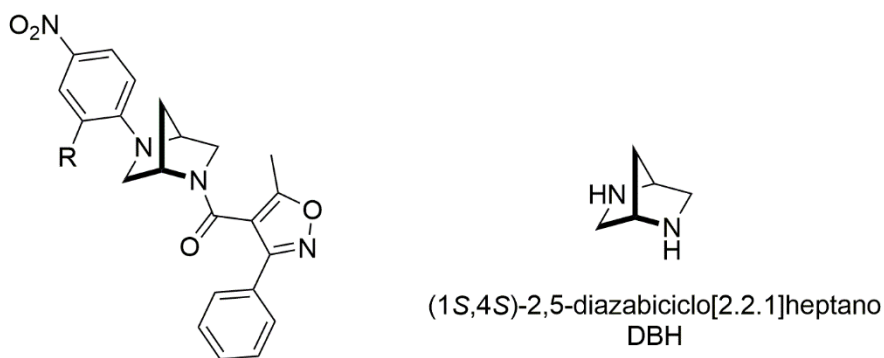
La mayor actividad se ha mostrado en aquellos inhibidores donde A tiene los sustituyentes $R_1 = -Cl$ y $R = -NO_2$. La sustitución de R_1 por sustituyentes con grupos electrodonadores como el $-OCH_3$ disminuye significativamente la actividad biológica, así como la ausencia de

sustituyentes en R.^{16c, 86} La presencia de un grupo pirrolidina en R₂ (Su 7, 0131) da buenos resultados contra la cepa A/Cal/07/2009 H1N1.^{16b}

El anillo B se representa como un anillo de piperazina donde al átomo de nitrógeno de la amida le corresponde la posición 1 (figura 18). Los casos reportados con modificaciones en éste sistema central incluyen análogos con la presencia de grupos metilo en R₈ (Kao 8 y Su 2)^{16b,c} y sustituciones de la piperazina por anillos de 1,4-diazepane, como el compuesto 3t reportado por Cheng *et al.*⁸⁶, y por anillos de piperidina como los compuestos a y b reportados por Pei *et al.*⁸⁷



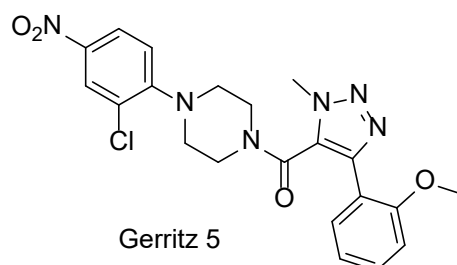
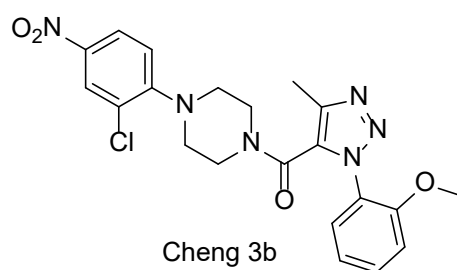
En mis estudios de maestría realicé la síntesis de dos análogos de nucleosina que incorporan el sistema (1S,4S)-2,5-diazabíciclo[2.2.1]heptano (DBH)⁸⁸ como se muestra en la figura 19. No se cuenta con valor de Cl₅₀ para estos análogos.



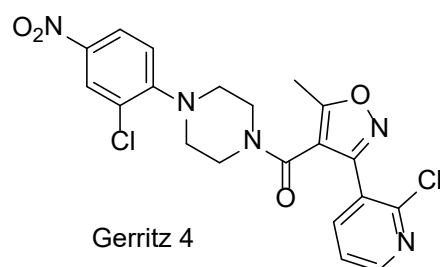
R = -Cl, -H

Figura 19. Análogo de nIz con sustitución en el anillo de piperazina.

El anillo C es un anillo heteroaromático, al cual se le han realizado diversas sustituciones tanto en las respectivas posiciones como en los átomos que conforman el anillo. En cuanto a los sustituyentes, son dos principales, uno en R_3 y otro que suele ser un anillo fenilo (**D**). El sustituyente más común en R_3 suele ser $-CH_3$, el cual, al remplazarlo por $-H$ o $-Et$ se observa una disminución en la actividad. Se ha sustituido el anillo C por diferentes bioisómeros, encontrando, que los análogo donde $X = Y = Z = N$ son buenos candidatos para mejorar la actividad antiviral, el ejemplo más activo es el compuesto 3b reportado por Cheng *et al.*⁸⁶ Lamentablemente estos análogos presentaron una farmacocinética pobre en ratones debido a la inestabilidad metabólica; el análogo 5 reportado por Gerritz *et al.*⁸⁴ mostró ser más estable.



En el anillo **D** también se han realizado sustituciones, encontrando que las mejores sustituyentes son los grupos metoxilo (-OCH₃) en las posiciones R₄, R₆ y R₇. La sustitución del metoxilo en R₄, está involucrada en mejorar la solubilidad. Los ligandos en la posición 5, presentaron poco aporte en la actividad. Otro sustituyente que presentó actividad es el -Cl principalmente en R₄ y R₇. La presencia de un grupo -Cl en R₄ da buenos resultados contra la cepa A/Cal/07/2009 H1N1.^{16b} La sustitución del fenilo D por uno de 2-cloro-3-piridinilo (Gerritz 4) también está relacionado con mejorar la solubilidad acuosa.



También se ha remplazado el anillo C por fenilos sustituidos obteniendo un esqueleto como el que se muestra en la [figura 20](#). Se demostró que las sustituciones en las posiciones *orto* y *para* mejoran la actividad antiviral mientras que las sustituciones en *meta* la disminuyen.⁸⁹

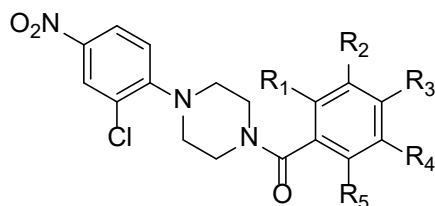
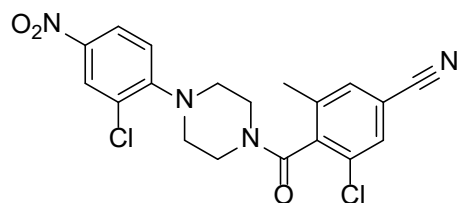


Figura 20. Esqueleto de algunos antivirales derivados de nucleozina.

En la posición *orto* los sustituyentes que mejor actividad presentaron fueron -Br y -CH₃ en comparación con grupos -NH₂, -OH y -OCH₃. Al cambiar -CH₃ por -Et o -Cpr^{††} se observó una disminución en la actividad. Para los sustituyentes en la posición *para* se presentó mejor actividad con grupos electroattractores -CN y -NO₂, sobre -Br, -Cl o -CO₂Me. Una vez establecido -CH₃ como sustituyente en R₁ en la sustitución de los grupos R₅ se encontró que el mejor sustituyente es el -Cl sobre grupos -Br, -NO₂ y -NH₂. El compuesto con mayor

^{††} Cpr = ciclopropilo

actividad fue el compuesto **39** reportado por Liao *et al.*⁸⁹



Liao 39

Mecanismo de acción de la nIz

Cuando nucleozina se enlaza a la NP se forman oligómeros de la proteína, los cuales están indispuestos para migrar al interior del núcleo, lo que a su vez resulta en el cese de la replicación del virus. En la [figura 21](#) se muestra un esquema del oligómero de NP inducido por la amida Gerritz 3. En la [figura 22](#) se observa una imagen de cristalografía de Rayos X del compuesto Gerritz 3; se observa que el reconocimiento molecular se da principalmente por un puente de hidrógeno y dos interacciones π .⁸⁴ La evidencia experimental indicaba que algunas cepas de influenza presentaban mecanismos de resistencia en donde se presentaban algunas de las siguientes mutaciones Y52H, Y289H y N309K o N309T, siendo esta observación cristalográfica lo que evidenciaba que los dos sitios de unión involucrados en la unión con las amidas son los surcos Y289/N309 y Y52.⁸⁴ En 2016 se reportó una nueva interacción CH/ π donde uno de los hidrógenos de la piperazina interactúa con la nube π de arginina 382 ([figura 23](#)).⁹⁰

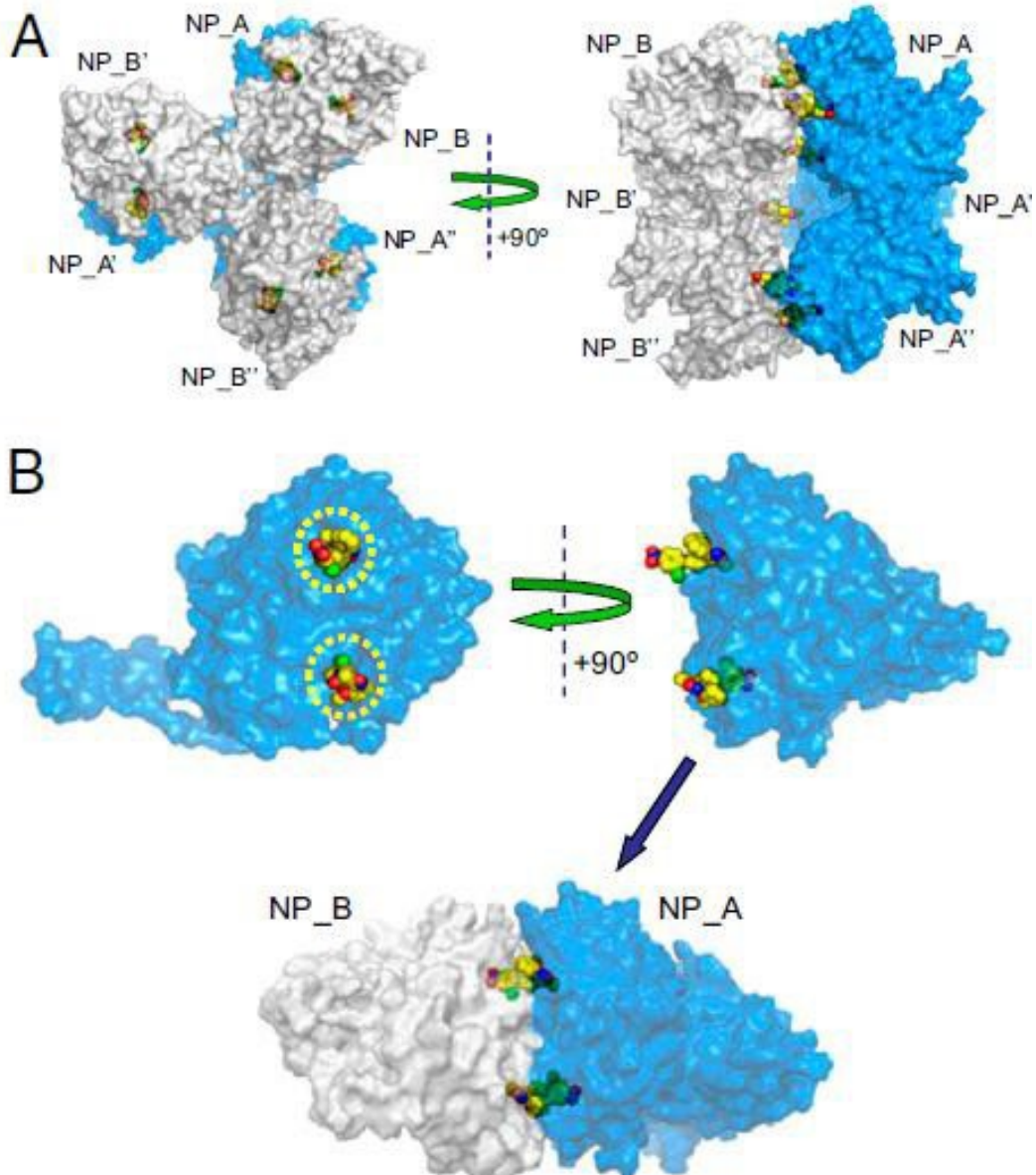


Figura 21. Estructura de Rayos X del complejo Gerritz 3: NP (código PDB: 3RO5). **a)** seis moléculas de Gerritz 3 (amarillo) puentean dos trímeros de NP (NP_A, N_PA', N_PA'' en azul y NP_B, N_PB', N_PB'' en gris) y forman un dímero de trímeros (o hexámero). **b)** Dos moléculas del compuesto 1 puentean una unidad dimérica de NP_A:NP_B (NP_A en azul y NP_B gris) en una orientación antiparalela. Tomado de Gerritz, S. W.; Cianci, C.; Kim, S.; Pearce, B. C.; Deminie, C.; Discotto, L. Inhibition of influenza Virus Replication via Small Molecules That Induce the Formation of Higher-Order Nucleoprotein Oligomers. *PNAS* **2011**, *108* (37), 15366–15371. <https://doi.org/10.1073/pnas.1107906108>.

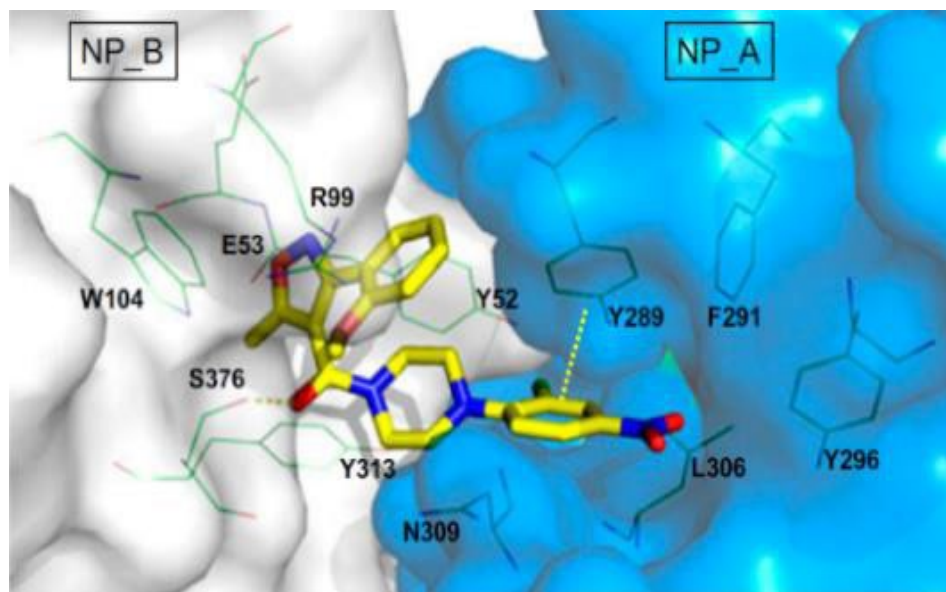


Figura 22. Vista detallada del compuesto 2 enlazado a la subunidad dimérica NP_A:NP_B. Las interacciones clave incluyen un puente de hidrógeno del carbonilo de la amida al residuo S376 de NP_B y dos interacciones π - π entre el 2-cloro-4-nitrofenilo y Y289, y entre, Y52 y el segmento de fenilo. Tomado de Gerritz, S. W.; Cianci, C.; Kim, S.; Pearce, B. C.; Deminie, C.; Discotto, L. Inhibition of influenza Virus Replication via Small Molecules That Induce the Formation of Higher-Order Nucleoprotein Oligomers. *PNAS* **2011**, *108* (37), 15366–15371. <https://doi.org/10.1073/pnas.1107906108>.

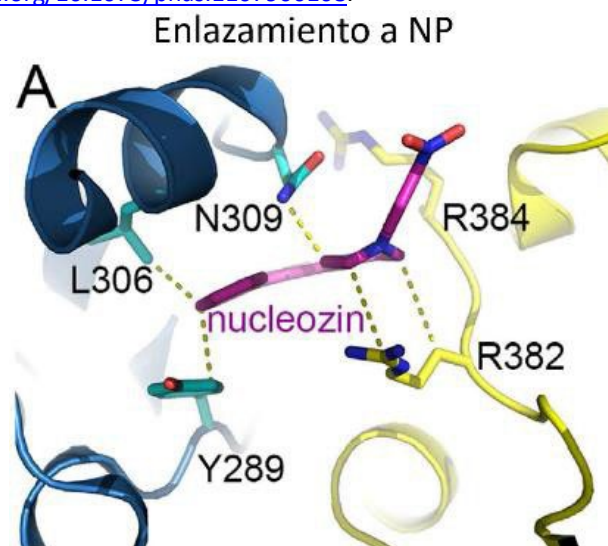


Figura 23. Representación cristalográfica del complejo NP-nucleosina (Código PDB: 5B7B). Tomado de: Pang, B.; Cheung, N. N.; Zhang, W.; Dai, J.; Kao, R. Y.; Zhang, H.; Hao, Q. Structural Characterization of H1N1 Nucleoprotein-Nucleosin Binding Sites. *Sci Rep* **2016**, *6*, 1–9. <https://doi.org/10.1038/srep29684>.

Síntesis de análogos de nucleozina

Las propuestas sintéticas de nucleozina y sus análogos reportados en la literatura consisten en la formación de la amida mediante una reacción de acoplamiento peptídico entre el ácido 5-metil-3-fenil-4-isoxazólico correspondiente y la amina secundaria *N'*-(2-cloro-4-nitrofenil)-*N*-piperazina (I) mostrado en la [figura 24](#).

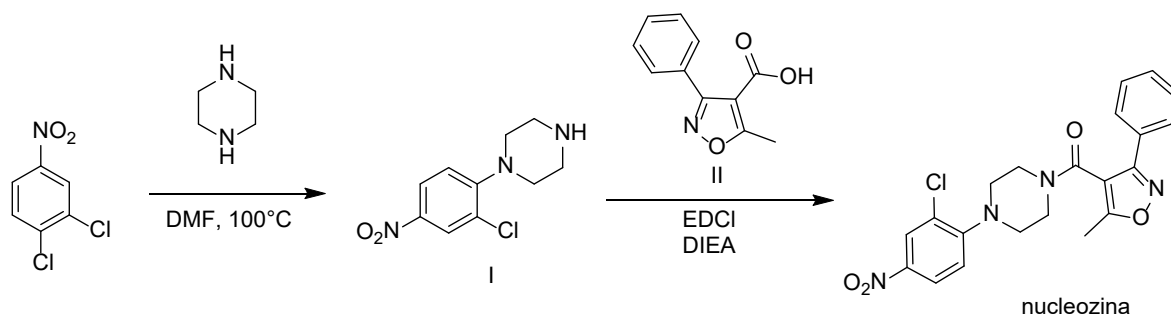


Figura 24. Síntesis de nucleozina según Kao y col.^{16c}

Durante mis estudios de maestría se reporté una nueva estrategia de síntesis⁸⁸ que parte del compuesto 3-cloro-4-fluor-1-nitrobenzénico que es comercialmente disponible, el cual reaccionó con la *N*-(Boc)piperazina mediante una reacción de tipo Sustitución Nucleofílica Aromática (S_NAr) para generar el compuesto III; la posterior remoción del grupo protector *N*-(Boc) nos dió lugar al grupo de la amina secundaria necesario para el acoplamiento peptídico con el anhidrido mixto derivado del ácido carboxílico II ([figura 25](#)).

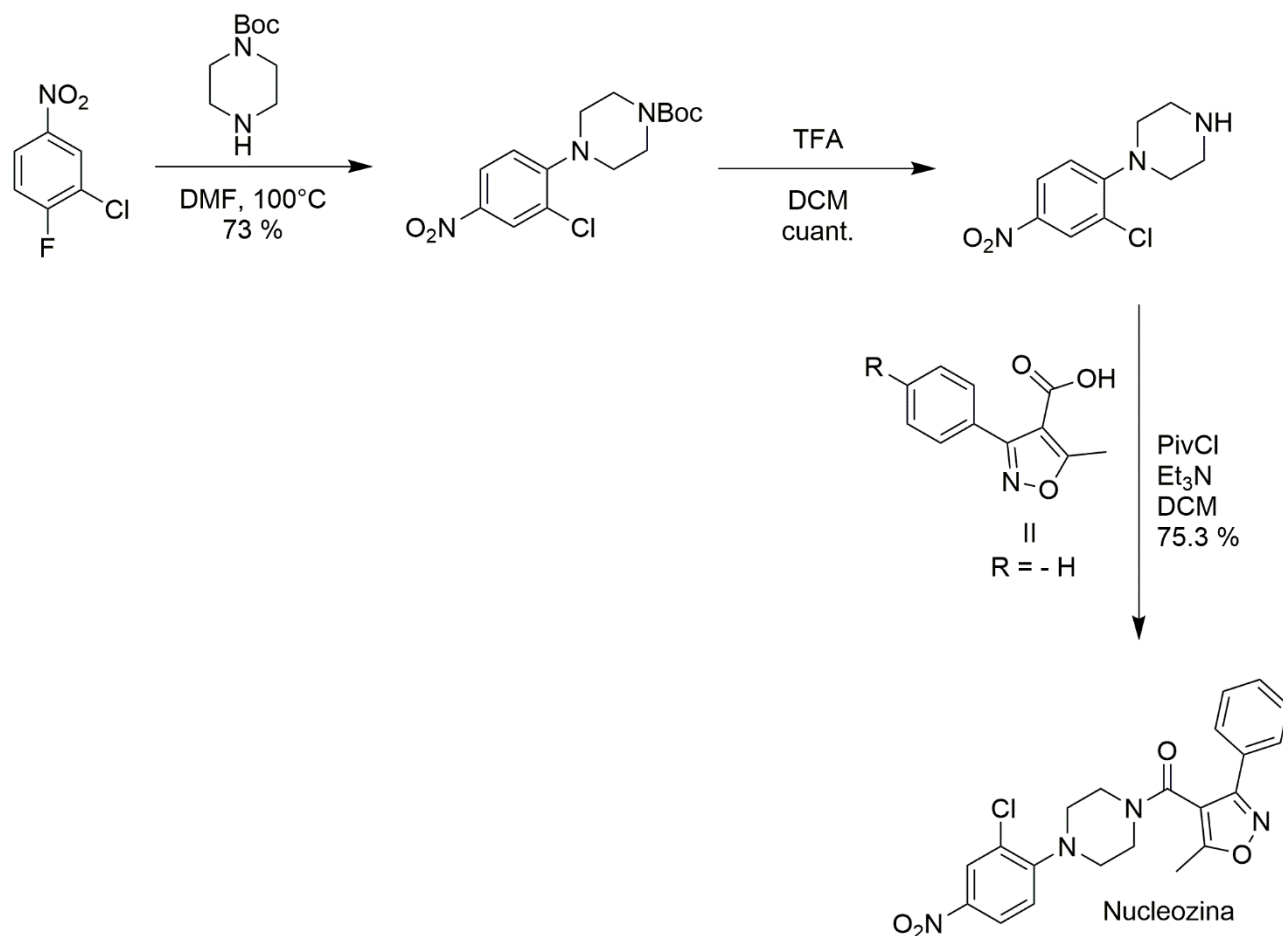


Figura 25. Síntesis de la nucleozina (nlz) propuesta en 2018.

La síntesis del análogo de nucleozina que incorpora el sistema (1*S*,4*S*)-2,5-diazabicyclo[2.2.1]heptano consistió en la homologación de la síntesis de nucleozina para el sistema bicíclico correspondiente, es decir partiendo del (1*S*,4*S*)-*N*-Boc-2,5-diazabicyclo[2.2.1]heptano, como se muestra en la [figura 26](#).

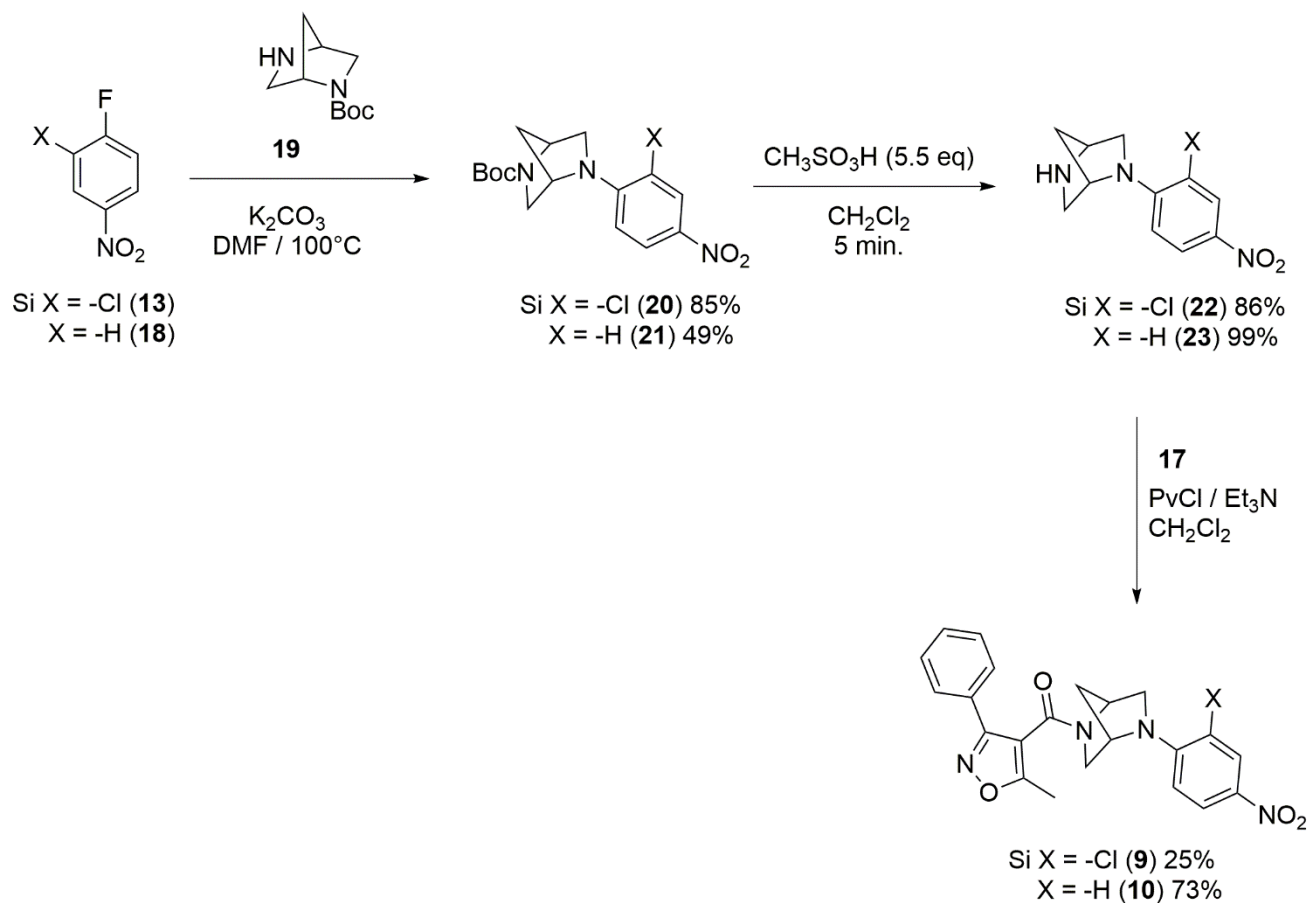


Figura 26. Síntesis de análogos diazabíclicos reportadas en 2018.

La síntesis del ácido isoxazólico también se encuentra reportada. La estrategia utilizada por Gerritz *et al.*⁸⁴ consiste en partir de benzaldehídos sustituidos en el anillo aromático; a partir de estos aldehídos se generará la oxima correspondiente con clorhidrato de hidroxilamina y la posterior halogenación de la posición vinílica de la oxima con *N*-clorosuccinimida (NCS). Posteriormente se realiza la ciclación para formar el anillo de isoxazol mediante una adición tipo Michael y la posterior sustitución del halógeno. También se ha reportado una variante a esta estrategia en la cual el grupo aldehído sustituye a un grupo nitrilo el cual condensa con acetoacetato de etilo con TiCl₄ como catalizador en tolueno. El posterior tratamiento del este producto con el clorhidrato de hidroxilamina da lugar a una especie de transaminación y una posterior reacción de condensación da lugar al ácido isoxazólico.

PLANTEAMIENTO DEL PROBLEMA

Existen pocos análogos de nucleozina con modificaciones en el segmento de piperazina, por lo que se tiene poca evidencia de la función de este ciclo sobre la actividad de la actividad biológica. Esta modificación estructural se aborda en este trabajo.

HIPÓTESIS

Estudios computacionales y experimentales con análogos de nucleozina con modificaciones en la piperazina evidenciarán el papel de este segmento en el reconocimiento con NP. Dado que la piperazina es la parte central en la molécula de nucleozina, las modificaciones estructurales en este sitio afectaran (modificarán) la topología y con ello las interacciones involucradas en el reconocimiento con NP.

OBJETIVOS

Objetivo general

Evidenciar y comprender la importancia de la piperazina en la molécula de nucleozina sobre la actividad antiviral (influenza A).

Objetivos particulares

- Analizar el modo de unión entre nucleozina y sus análogos con la nucleoproteína (NP), así como sus características estructurales de la actividad biológica, mediante acoplamiento molecular automatizado.
- Generar modelos de predicción de actividad antiviral para los análogos de la nucleozina, mediante modelos QSAR.
- Sintetizar análogos de nucleozina.
- Evaluar la actividad biológica de los análogos de nucleozina (Facultad de Medicina, Universidad Autónoma del Estado de Morelos).

RESULTADOS

Fundamento del proyecto

Para conocer la influencia del anillo central de nucleozina en la actividad anti-influenza abordaremos el estudio de una serie de análogos de nucleozina con modificaciones en el anillo de piperazina. Una de las principales modificaciones fue la sustitución de la piperazina por el sistema (1S,4S)-2,5-diazabicyclo[2.2.1]heptano (DBH) formando el compuesto **1** que se muestra en la [figura 27](#), también se tomó en cuenta su imagen especular, es decir su diasterómero.

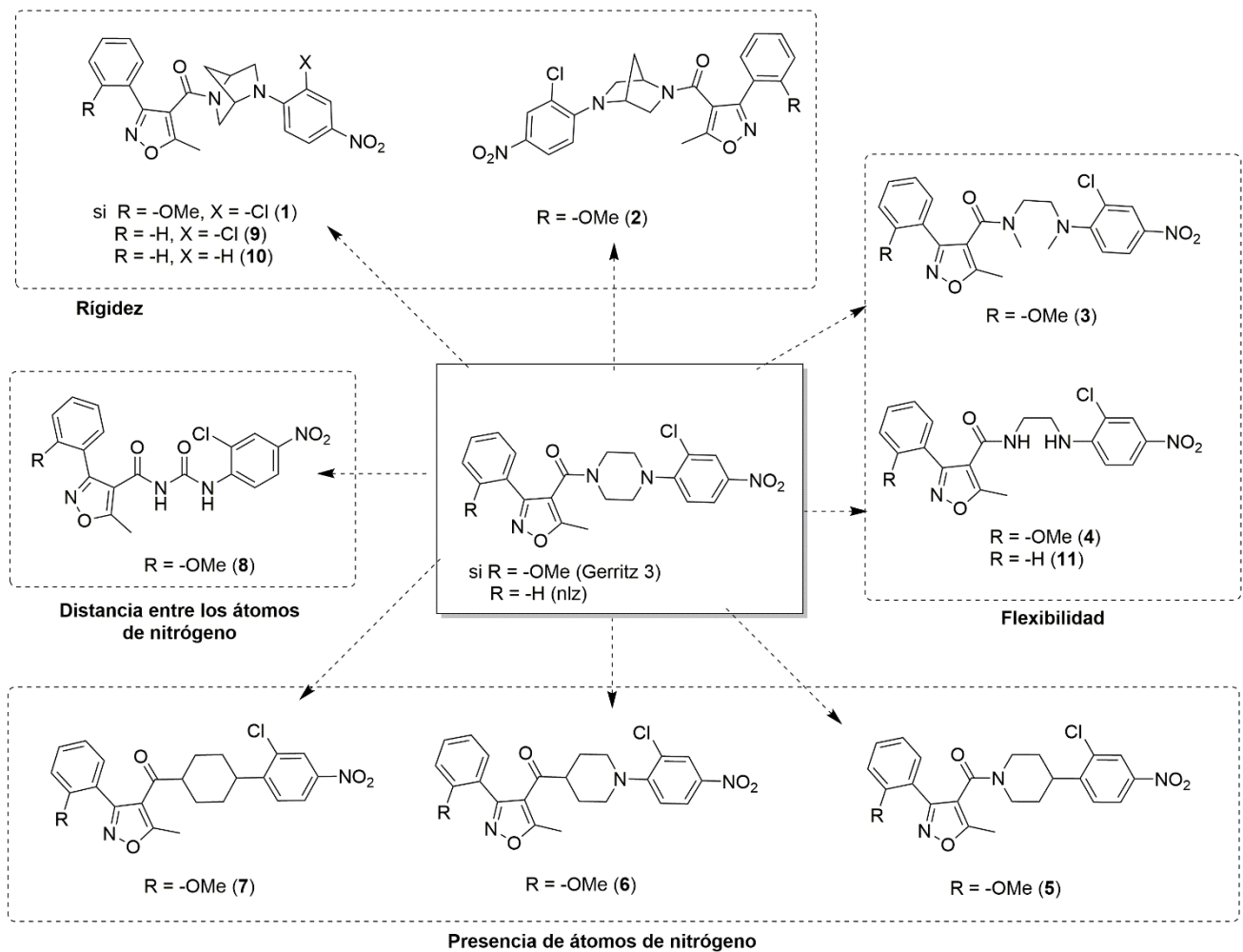


Figura 27. Análogos propuestos en este trabajo.

El sistema DBH es estructuralmente equivalente a la piperazina, pero al ser un anillo conformacionalmente restringido puede modificar la interacción con la nucleoproteína NP en comparación con su análogo de nucleozina. En la [figura 28](#) esquema se muestran los cambios conformacionales y configuracionales de la piperazina para los isómeros en su conformación de silla, comparados al del DBH se muestra que el DBH solo genera tres configuraciones posibles que se muestran en la [figura 29](#).

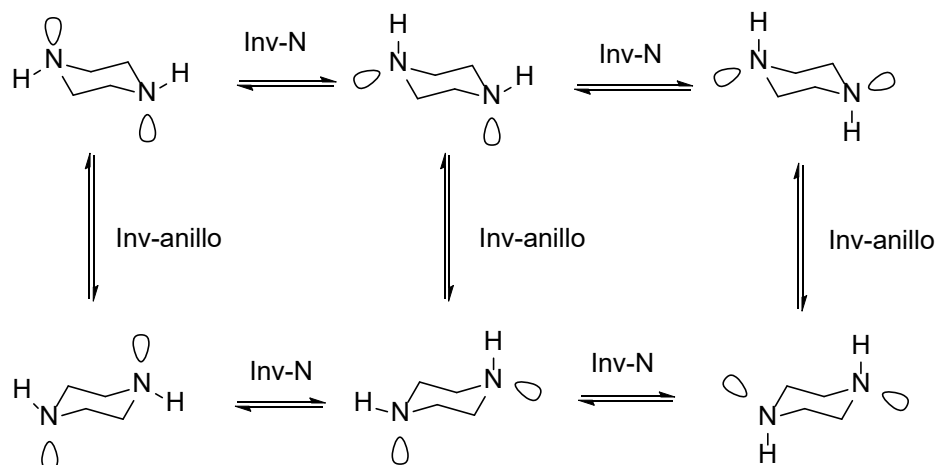


Figura 28. Equilibrio conformacional y configuracional de la piperazina.

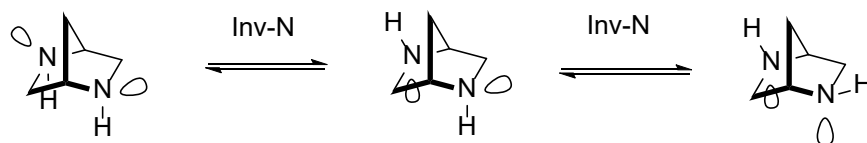


Figura 29. Equilibrio configuracional del sistema (1S,4S)-2,5-diazabíciclo[2.2.1]heptano (DBH).

Así mismo, se propuso el análogo de etilendiamina **3** con la capacidad de ser más flexible que el análogo de nucleozina debido a la libre rotación de la cadena de etileno. Así mismo, la eliminación de dos carbonos del ciclo nos permite hacer incorporación de grupos NH polares con la finalidad de formar interacciones intramoleculares de tipo puente de hidrógeno dentro del sitio de reconocimiento, lo que da lugar al compuesto **4**.

Una parte importante de esta molécula que se ha estudiado muy poco es el papel de los átomos de nitrógeno. Con perspectivas a entender su función, se propone la sustitución

de átomos de N por grupos metinos, lo que da lugar a los compuestos **5**, **6** y **7**.

También se busca disminuir el tamaño de la cadena espaciadora ente los átomos de nitrógeno del ciclo a un solo átomo de C espaciador. Por ello se propone el derivado de urea **8**, la cual sería una molécula estable que no sería susceptible de hidrólisis, así mismo se busca incorporar grupos polares NH para mejorar el reconocimiento

Se busca estudiar las consecuencias biológicas de cambiar la estructura molecular de los ligandos mencionados utilizando cuatro enfoques de investigación propios de la Química Medicinal que se enlistan a continuación:

- Acoplamiento molecular.
Ayuda a indagar el modo de reconocimiento del ligando por parte de nucleoproteína NP de influenza A, y entender la influencia de la estructura química de los ligandos.
- Modelos de predicción de CI_{50} de inhibidores de la NP (modelos QSAR):
Ayuda a establecer de manera cuantitativa relaciones numéricas entre la estructura química de los ligandos y la actividad biológica.
- Síntesis de análogos de nucleozina:
Obtención de los análogos de nucleozina planteados en el estudio y algunos intermediarios.
- Evaluación biológica *in vitro* de análogos de nucleozina: los resultados se obtuvieron en colaboración con la Facultad de Medicina del Estado de Morelos.

Resultados de acoplamiento molecular para análogos de nlz

En las siguientes figuras se muestran los resultados del acoplamiento molecular de los ligandos propuestos. En color amarillo se muestra el ligando cocrystalizado o nativo, Gerritz 3 (código PDB: 3RO5). Este complejo proteína-ligando es un dímero del tipo [2(Gerritz 3):2NP] donde las dos moléculas del ligando interactúan con dos monómeros de NP (A y B). En estas estructuras el monómero A se muestra en color verde y está

conformado por un bucle de residuos entre A_Y289 y A_N309, mientras que el monómero B está representado en color azul, y está conformado por una serie de residuos, entre los que se encuentran, B_Y289, B_S376, B_Y313, B_W104 y B_R99.

En la [figura 30](#) se muestra un mapa de interacciones en 2D para el compuesto 1. En la [figura 31](#) muestra la posición de acoplamiento de los compuestos 1 y 2 dentro del sitio de unión ortostérico, los átomos de hidrógeno del puente de metileno están en la vecindad de una pared polar, formada por los residuos B_Y313 y A_N309. Estos residuos (B_Y313 y A_N309) tienen una interacción de enlace de hidrógeno que une ambos monómeros (A y B). Existe un contacto débil entre el puente de metileno y el residuo de A_N309, Los valores de puntaje y afinidad asociados se muestran en la [tabla 2](#). Para el caso del compuesto 2 se encontró la pose que se muestra en rojo en la [figura 32](#). Se observa que en esta pose en oxígeno del carbonilo se aleja del residuo S376.

Tabla 2. Valores de puntaje y de afinidad calculados

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 1 (púrpura)	-26.15	-13.75	145.47
Compuesto 2 (rojo)	-29.07	-13.17	149.98
Compuesto 2 (naranja)	-28.37	-13.45	144.61

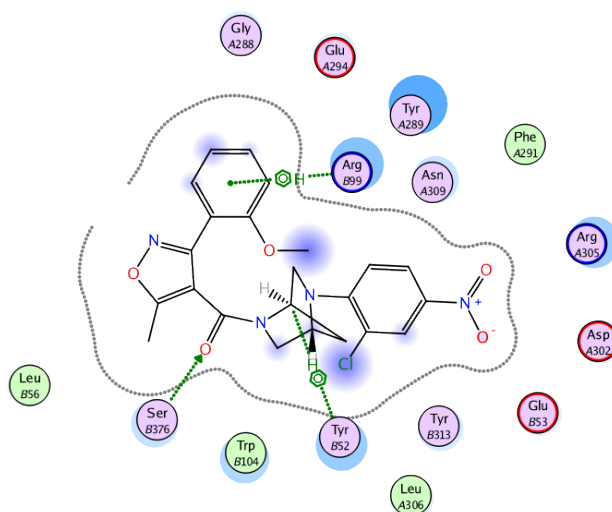


Figura 30. Mapa de interacciones en 2D para el compuesto 1.

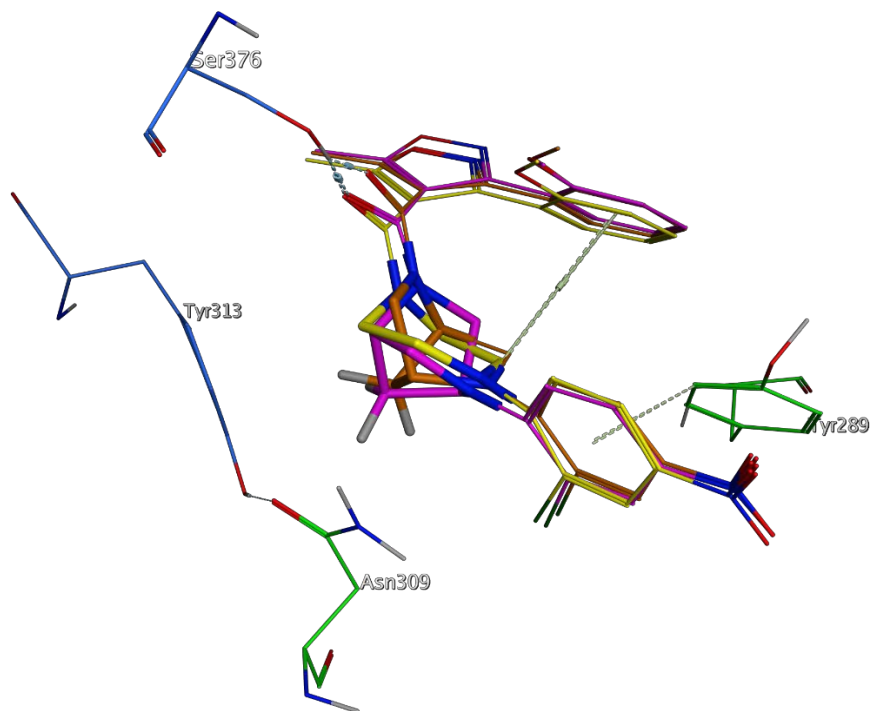


Figura 31. Resultados de la simulación de acoplamiento molecular. En amarillo se representa el ligando cocrystalizado Gerritz 3. **a)** poses seleccionadas para el compuesto 1 en violeta y compuesto 2 en naranja.

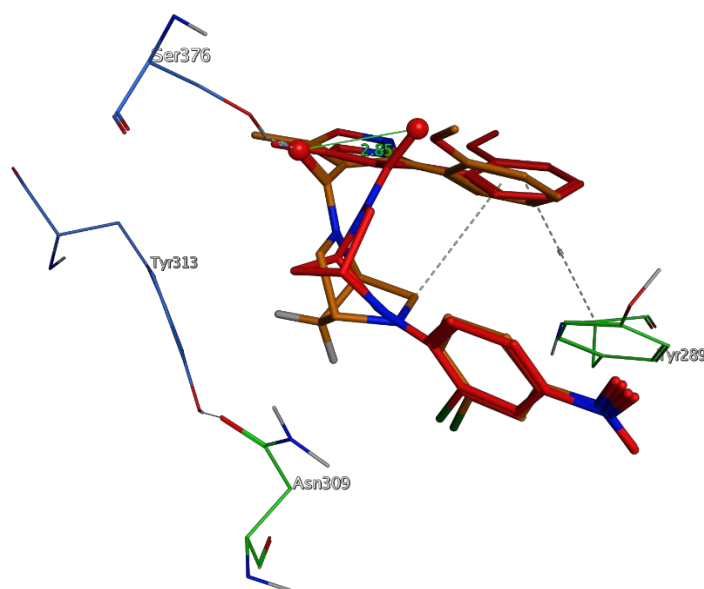


Figura 32. Comparación de poses del compuesto 2.

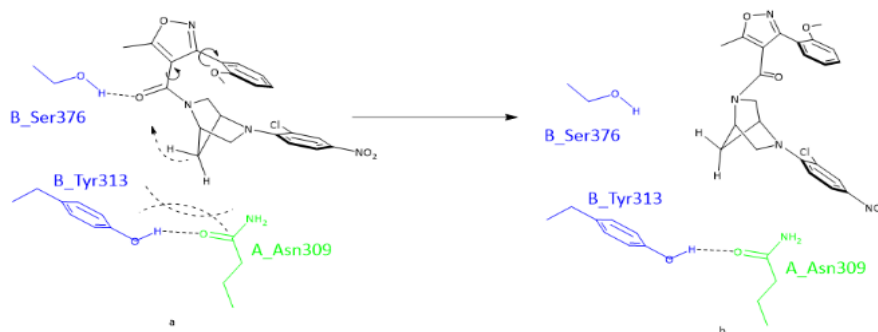


Figura 32 (continuación). Giro de enlaces del compuesto 2.

Dado que los análogos de etilendiamina **3** y **4** son una cadena abierta con rotación libre, hay nueve enlaces rotables en estos compuestos, en comparación con seis enlaces rotables del compuesto Gerritz 3, por lo que estos análogos muestran un espacio conformacional más grande. En la figura 33 se muestra un ejemplo donde la conformación extendida del análogo de etilendiamina **3**, de color cian oscuro, libera energía interna a expensas de perder la interacción con el residuo A_Y289. En color cian se muestra la pose preferida del acoplamiento molecular. Para el caso del compuesto **4**, que se muestra en la figura 34, debido a la sustitución de los metilos (CH_3) unidos a nitrógeno, por átomos de H, hace que se pierda el aporte estérico, la capacidad de rotar los enlaces del ligando aumenta en comparación con el compuesto **3** y disminuye la probabilidad de obtener la conformación adecuada para interactuar con la proteína, además de que los grupos NH no forman enlaces de hidrógeno como se esperaba. Los valores de puntaje y afinidad asociados para las poses de los compuestos **3** y **4** se muestran en la tabla 3.

Tabla 3. Valores de puntaje y de afinidad calculados para los compuestos **3** y **4**.

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 3 Cian	-27.37	-12.99	129.99
Compuesto 3 Cian oscuro	-23.71	-11.59	115.93
Compuesto 4	-24.98	-13.03	88.77

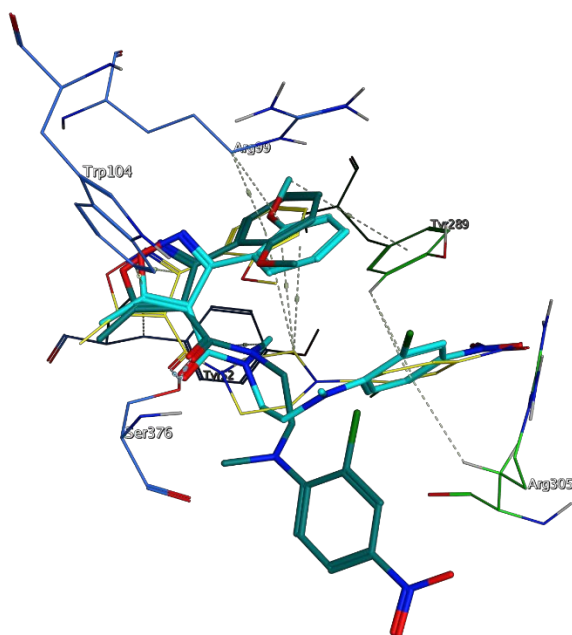


Figura 33. Pose preferida para el ligando **3**, en cian, en cian oscuro se muestra una pose donde el resto 2-cloro-4-nitrofenilo no tiene contacto con el monómero A.

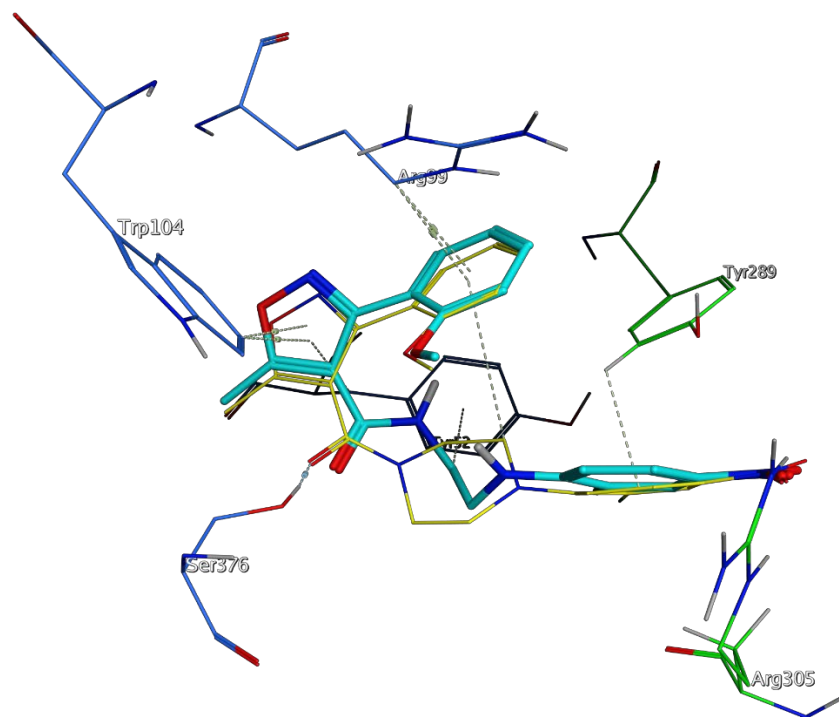


Figura 34. Pose preferida del compuesto **4**.

El papel de los átomos de nitrógeno de la piperazina: acoplamiento de los ligandos 5, 6 y 7.

Los ligandos **5-7** se propusieron para comprender el papel de los átomos de nitrógeno de la piperazina. En el caso del ligando **5** el cambio de átomo de nitrógeno por el grupo CH afecta principalmente a la geometría en esa región de la molécula.

En la [figura 35](#), el ligando nativo, que se muestra en amarillo, tiene un átomo de nitrógeno en la piperazina, de geometría trigonal plana con carácter sp^2 por la conjugación con el segmento de 2-cloro-4-nitrofenilo; para el ligando **5**, la conjugación se pierde debido a la sustitución del átomo de nitrógeno por el átomo de carbono, siendo tetraédrica su topología, y ahora hay un átomo de hidrógeno que se debe posicionar en algún lado, como se muestra en las estructuras de color cian y morado en la [figura 35](#). Esto tendría consecuencias en la disposición espacial del segmento de 2-cloro-4-nitrofenilo. Esta región del compuesto Gerritz 3 interactúa con el monómero A de NP, a través de introducirse a una especie de caja hidrofóbica que incluye a los residuos R305, L306, D302, Y289 y E294; y para el caso del compuesto **5**, este cambio en la geometría del átomo puede provocar que el segmento 2-cloro-4-nitrofenilo quede apretado dentro de la caja hidrofóbica; para evitarlo el ligando **5** mueve la parte central, alejando el carbonilo del residuo B_S376 ([figura 36](#)), por lo que la geometría del este átomo tiene un papel importante. Los valores de puntaje y afinidad asociados se muestran en la [tabla 4](#).

Tabla 4. Valores de puntaje y de afinidad calculados para el compuesto **5**.

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 5 naranja	-29.79	-13.32	102.07
Compuesto 5 (cian)	-26.38	-13.35	94.06
Compuesto 5 (púrpura)	-28.17	-13.34	86.75

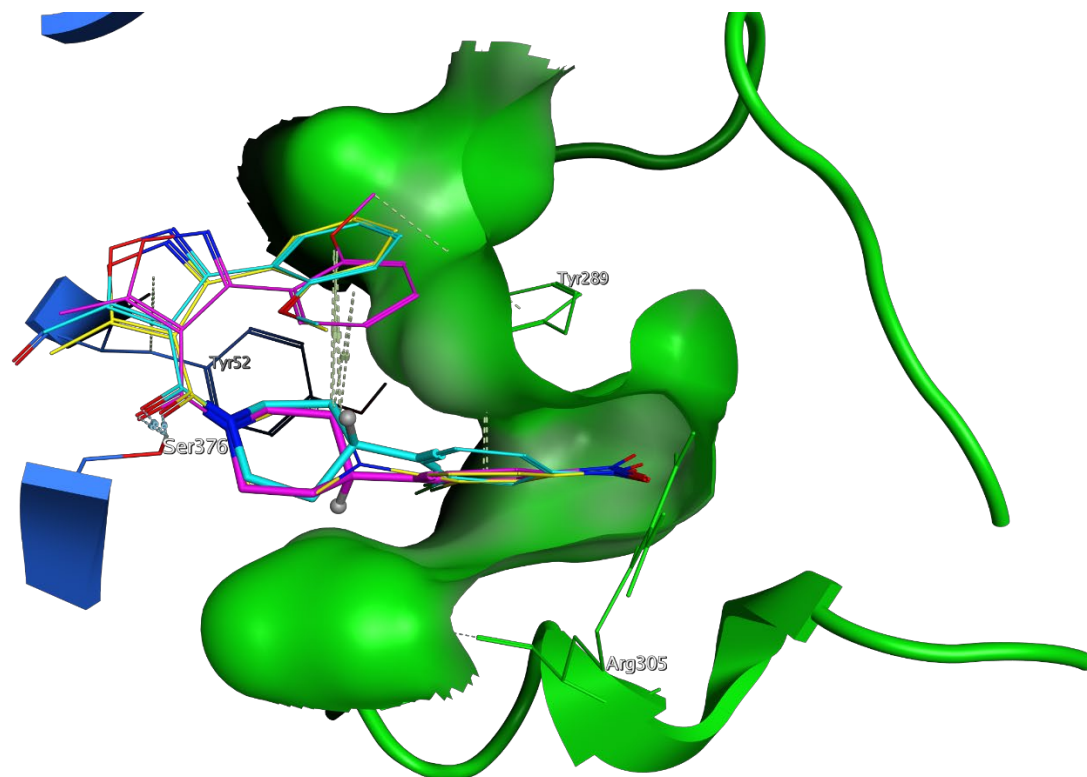


Figura 35. Se muestran dos poses para el ligando **5** comparadas con el ligando nativo: la sustitución del átomo de nitrógeno por un grupo metino genera un cambio de geometría y en la disposición espacial del segmento 2-cloro-4-nitrofenilo. la pose que se muestra en color rosa queda en tercer lugar del ranking, la pose que se muestra en color cian quedó en cuarto lugar.

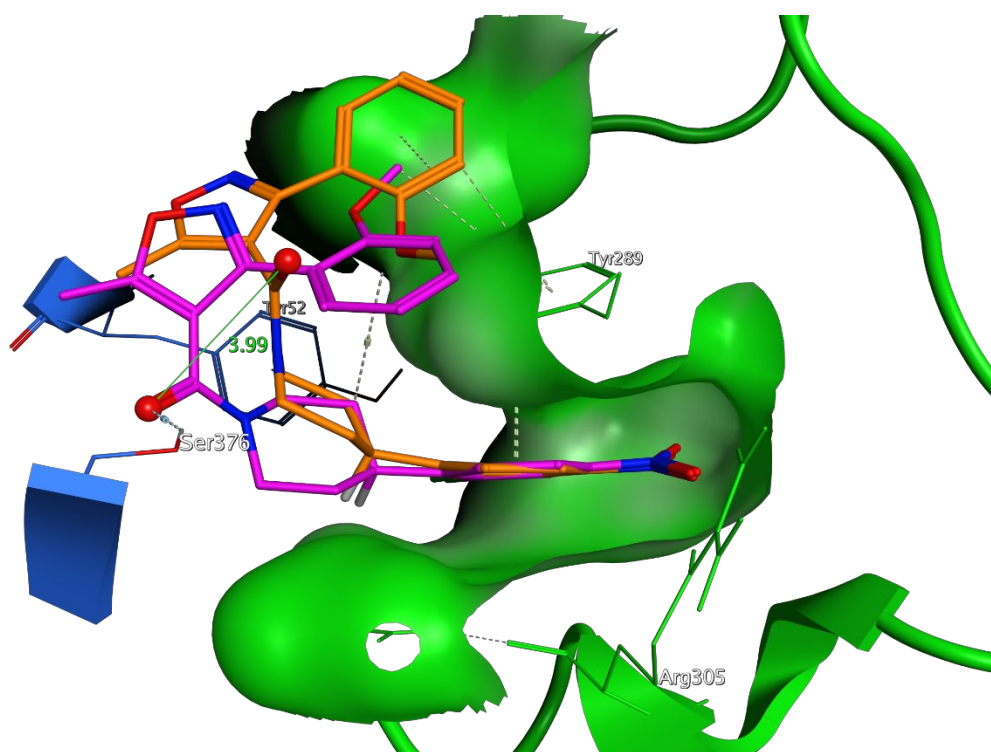


Figura 36. Comparación de poses del compuesto **5**; para la pose de color rojo se observa que el átomo de oxígeno está a 3.9 Å de distancia del mismo átomo en la pose púrpura.

Para el ligando **6** (figura 37) el cambio del átomo de nitrógeno en la amida, por un grupo metino (C-H), forma una cetona que, posiblemente, se tautomeriza, formando dos posibles formas ceto y pasando por un enol intermedio en la interconversión, además, este cambio incide en un cambio en la geometría del átomo de trigonal plano a tetraédrico. El tautómero del enol se superpone de manera casi total a la pose nativa debido a que el carbono del ligando **6** y el nitrógeno de Gerritz 3 tienen la misma geometría trigonal plana, en comparación con las formas ceto donde se observan algunos cambios y el carbono es tetraédrico. Parece que la geometría de la posición estudiada tiene una gran función para estabilizar el ligando dentro del sitio de reconocimiento, aquí el átomo de nitrógeno juega un papel importante por que aporta la geometría trigonal plana. Los valores de puntaje y afinidad asociados se muestran en la tabla 5.

Tabla 5. Valores de puntaje y de afinidad calculados para el compuesto 6.

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 6	-28.78	-13.53	113.38

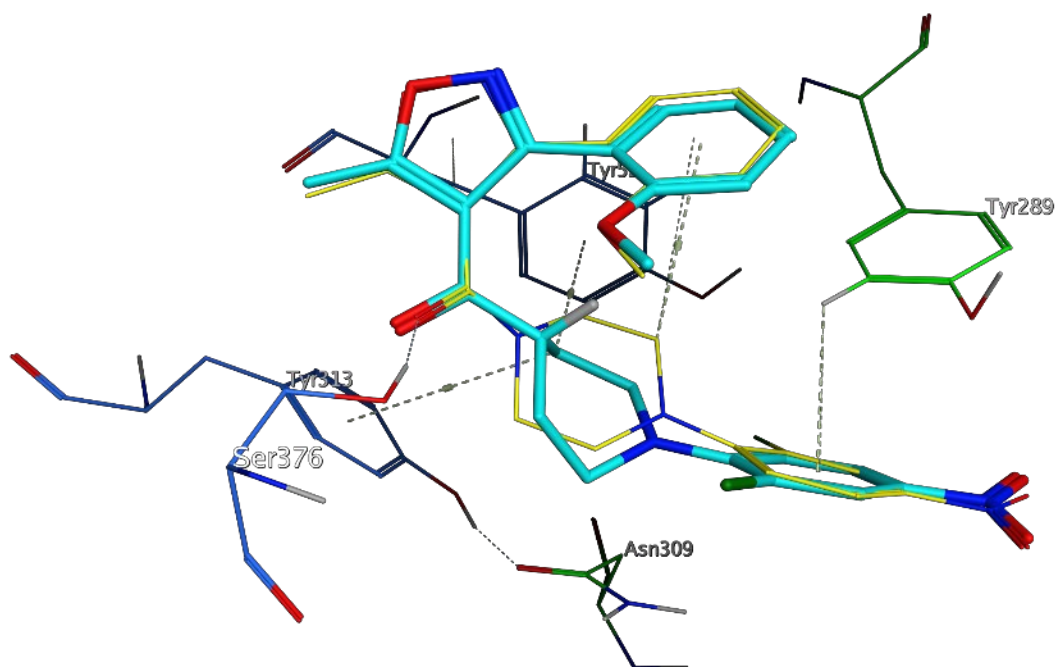


Figura 37. Se muestra la pose del compuesto 6 más similar al ligando nativo, en cian. Esta pose quedó en el primer lugar del ranking.

Dado que el compuesto 7 contiene un grupo cetona es necesario considerar teóricamente dos tautómeros de forma ceto que corresponden a la forma *syn* periplanar y forma *anti* periplanar, y el tercero es la forma enol, las afinidades calculadas indican que la forma ceto *anti* tiene mayor afinidad (-13.63 kcal/mol) por el sitio de reconocimiento que la forma ceto *syn* (-12.59 kcal/mol) y entre estos, la forma *anti* es la que más se parece al ligando cocrystalizado, esto se muestra la [figura 38](#), donde dicho compuesto se representa en color cian. Otras poses encontradas indican que el el ligando *syn*, representado de color morado en la [figura 38](#), hay rotación de los enlaces (similar al caso del ligando 5) que provoca un alejamiento del residuo B_S376 dificultando el reconocimiento del ligando, dicho efecto se observa en el tautómero *anti* con menor frecuencia. Nuevamente vuelve a resaltar la importancia de la geometría de los átomos de nitrógeno en la estabilización del complejo

[2(Gerritz 3):2NP]. Los valores de puntaje y afinidad asociados se muestran en la [tabla 6](#). En la [figura 39](#) se representa otro conformero de forma *syn* periplanar.

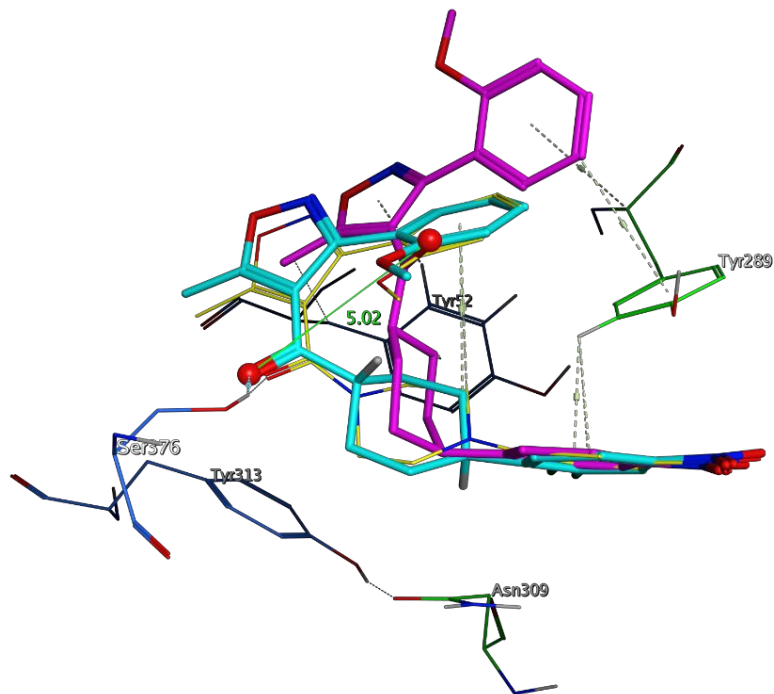


Figura 38. Comparación entre dos poses del compuesto 7. En amarillo se muestra el ligando nativo, en morado se muestra el isómero ceto *syn* y en cian se muestra el isómero ceto *anti* que es más parecido al nativo.

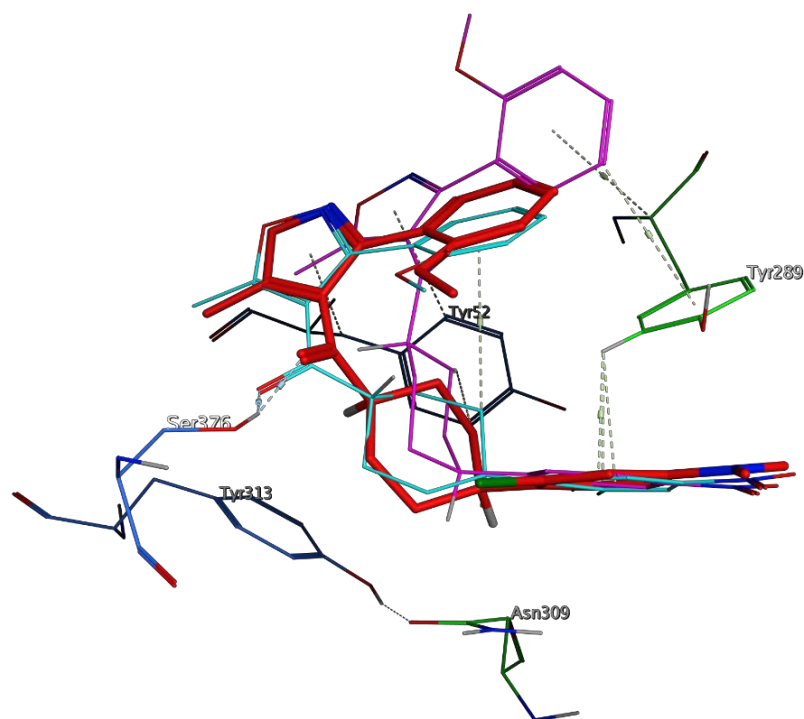


Figura 39. Se observa que el compuesto **7** en su forma *syn periplanar* formando un puente de hidrógeno con S376.

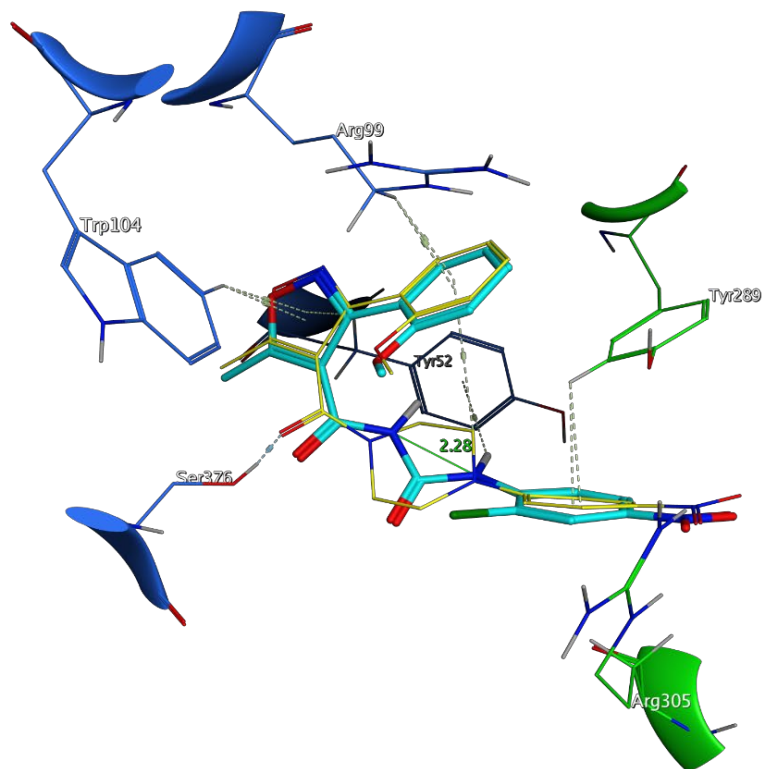
Tabla 6. Valores de puntaje y de afinidad calculados para el compuesto **7**.

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 7 trans cis	-29.18	-13.63	103.73
Compuesto 7 púrpura cis	-30.41	-12.59	107.64
Compuesto 7 cis rojo	-29.71	-12.73	109.06

Para el compuesto **8**, en la [figura 40](#) se muestra la pose seleccionada que es la más similar al ligando amarillo, en esta pose se observa que los carbonilos del compuesto **8** están en posición *syn periplanar*, la distancia N-N en el sitio de reconocimiento es de 2.83 Å para nIz, y de 2.28 Å para el compuesto **8**. No se observa la formación del puente del hidrógeno con algún grupo NH. Los valores de puntaje y afinidad asociados se muestran en la [tabla 7](#).

Tabla 7. Valores de puntaje y de afinidad calculados para el compuesto **8**.

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 8	-24.46	-13	71.12

**Figura 40.** Pose del compuesto **8** en el sitio de reconocimiento.

En la [tabla 8](#) se muestra un resumen de las energías de afinidad y potencial, junto con el puntaje asociadas a las poses del acoplamiento molecular.

Tabla 8. Resumen de valores de puntaje y de afinidad calculados.

Pose	Puntaje (kcal/mol)	Afinidad (kcal/mol)	E (kcal/mol)
Compuesto 1 (púrpura)	-26.15	-13.75	145.47
Compuesto 1 (rojo)	-28.98	-13.47	144.42
Compuesto 2 (rojo)	-29.07	-13.17	149.98
Compuesto 2 (naranja)	-28.37	-13.45	144.61
Compuesto 3 Cian	-27.37	-12.99	129.99
Compuesto 3 Cian oscuro	-23.71	-11.59	115.93
Compuesto 4	-24.98	-13.03	88.77
Compuesto 5 Rojo	-29.79	-13.32	102.07
Compuesto 5 cian	-26.38	-13.35	94.06
Compuesto 5 (púrpura)	-28.17	-13.34	86.75
Compuesto 6	-28.78	-13.53	113.38
Compuesto 7 trans cis	-29.18	-13.63	103.73
Compuesto 7 púrpura cis	-30.41	-12.59	107.64
Compuesto 7 cis rojo	-29.71	-12.73	109.06
Compuesto 8	-24.46	-13	71.12
Gerritz 3 (RMSD = 0.33)	-27.97	-13.50	131.00
Gerritz 3		-13.58	131.63

Resultados de modelos QSAR de predicción de actividad inhibitoria con análogos de nlz

En la [tabla 9](#) se muestran los estadísticos del mejor modelo obtenido. Se realizó el análisis del dominio de aplicabilidad por el método del apalancamiento.

Tabla 9. Estadísticos para el mejor modelo usando el algoritmo de SVR.

	r^2	q^2_{CV-LMO}	q^2_{ext}
	0.94	0.60	0.60
MSE	0.03	0.25	0.21

En la [figura 40a](#) se muestra la distribución de pCl_{50} para el grupo de prueba en gris y el de entrenamiento en naranja. La [figura 40b](#) muestra los valores experimentales de pCl_{50} de los inhibidores comparados contra el valor calculado por el mejor modelo usando SVR (*Support Vector Regressor*) como algoritmo de entrenamiento.

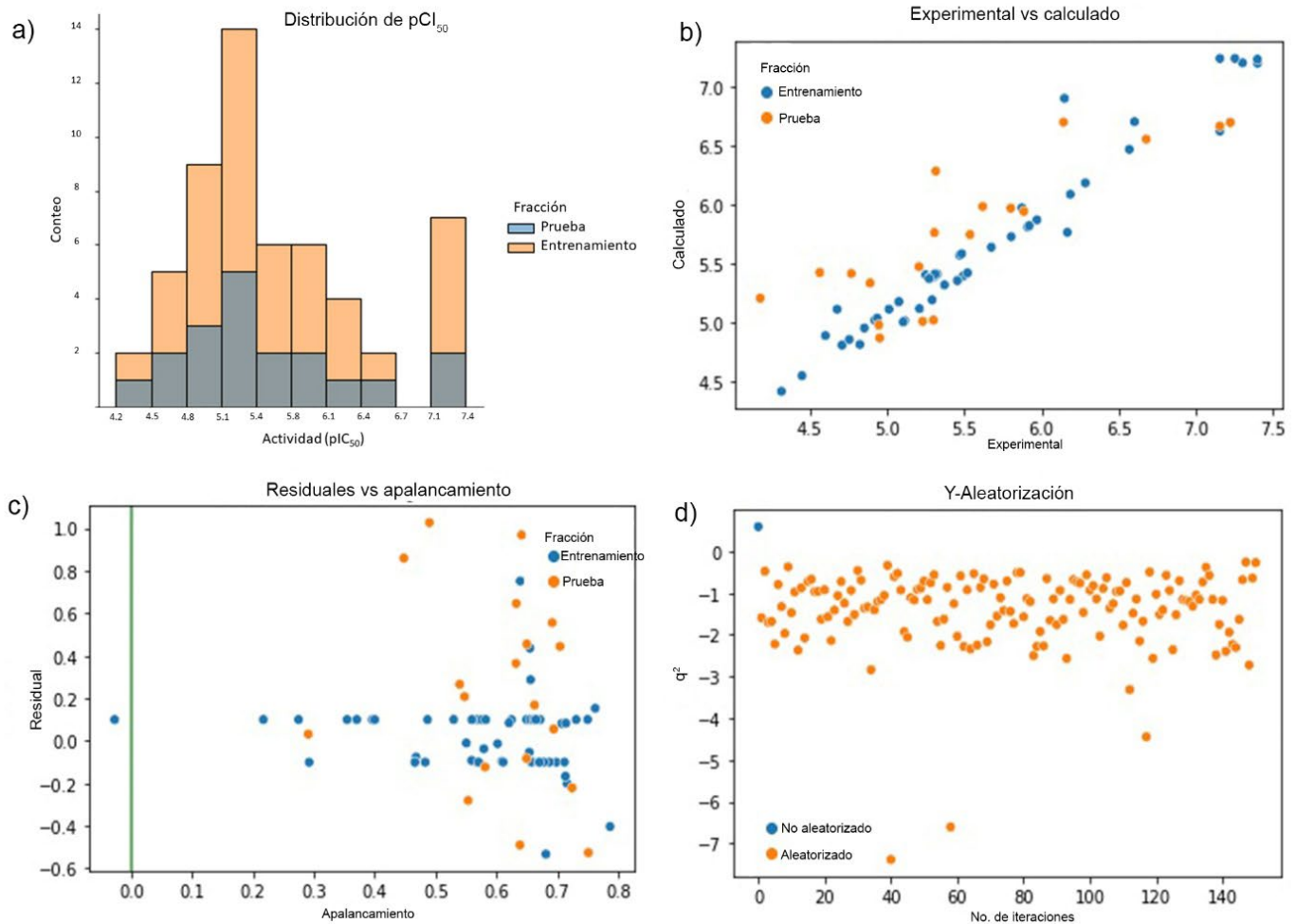


Figura 40. **a)** Distribución de pCl_{50} para el grupo de entrenamiento (naranja) y el grupo de prueba (gris) **b)** Gráfica de regresión del modelo SVR. El gráfico muestra los valores calculados por el modelo de regresión contra los valores experimentales para el conjunto de entrenamiento (en azul) y el conjunto de prueba (en naranja). **c)** Diagrama de Williams para el modelo calculado: el Dominio de Aplicabilidad se calculó por el método de apalancamiento. **d)** Prueba Y-Scrambling para el modelo de predicción: en el gráfico se muestran, en naranja, los resultados de los valores de q^2_{CV-LMO} para modelos aleatorizados y en azul el valor de q^2_{CV-LMO} para el modelo verdadero (no aleatorizado).

En la [figura 40c](#) se muestra el diagrama de Williams para el modelo de predicción: el dominio de aplicabilidad se calculó por el método de apalancamiento; el punto azul que sale del dominio de aplicabilidad corresponde a la molécula Cheng 3t, la cual posee

modificaciones en el anillo de piperazina. Esto indica que es necesario incluir más ejemplos de compuestos con modificaciones en el anillo de piperazina para reforzar el modelo. En la [figura 40d](#) se muestra el resultado de la prueba de Y-Scrambling para el modelo de predicción: en el gráfico se muestran, en naranja, los resultados de los valores de q^2_{CV-LMO} para modelos aleatorizados y en azul el valor de q^2_{CV-LMO} para el modelo verdadero (no aleatorizado). Se observa que el mejor valor se obtuvo para el modelo no aleatorizado, lo que indica que el modelo es robusto. Este modelo nos permite calcular el valor de pCl_{50} para el grupo de compuestos propuesto. Los valores de predicción se muestran en la [tabla 10](#).

Tabla 10. Actividad calculada para los análogos de la nucleozina (nlz) propuestos en este trabajo.

Compuesto	Experimental	Calculado
Nlz	7.22	6.69
Gerritz 3	7.39	7.19
compuesto 1		5.87
compuesto 2		5.86
compuesto 3		6.03
compuesto 4		5.50
compuesto 5		6.60
compuesto 6		6.59
compuesto 6 enol		6.59
compuesto 7 anti		6.04
compuesto 7 enol		6.25
compuesto 7 syn		5.86
compuesto 8		5.34
compuesto 9		5.54
compuesto 10		5.29
compuesto 11		4.90
compuesto 12		5.06

Síntesis de análogos de nIz

Durante la realización de este trabajo doctoral se sintetizó el compuesto **11**.^{§§} La ruta de síntesis se muestra en la [figura 41](#). La síntesis consiste en la formación del *N*-Bocetilendiamina (**13**) a partir de etilendiamina; posteriormente, a partir del ácido carboxílico (**II**) se formó la amida (**14**) mediante la formación inicial de un halogenuro de acilo con cloruro de oxalilo y DMF y la posterior adición de la amina monoprottegida. Después se removió el grupo protector *N*-Boc con 5.5 equivalentes de $\text{CH}_3\text{SO}_3\text{H}$, para obtener el compuesto **15**, para después realizar una $\text{S}_{\text{N}}\text{Ar}$ sobre el átomo de flúor del compuesto 3-cloro-4-nitro-1-fluorobenceno. Los compuestos **14** y **11** fueron caracterizados por RMN- ^1H , RMN- ^{13}C y Espectrometría de Masas (EM) por la técnica de ionización DART+ ([tabla 11](#)). Las señales se encuentran asignadas en la sección de caracterización espectroscópica de los compuestos sintetizados. Aún no se ha realizado la evaluación biológica del compuesto **11**, pero se tiene contemplado realizarla.

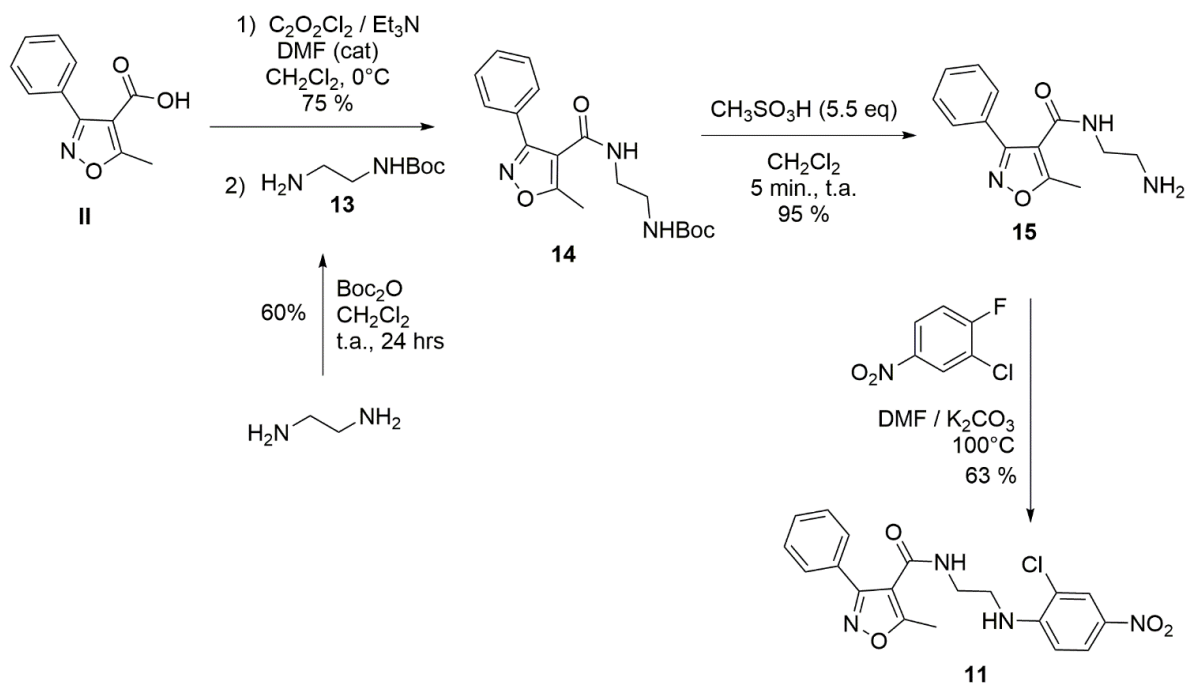


Figura 41. Síntesis del análogo **11** de nucleosina.

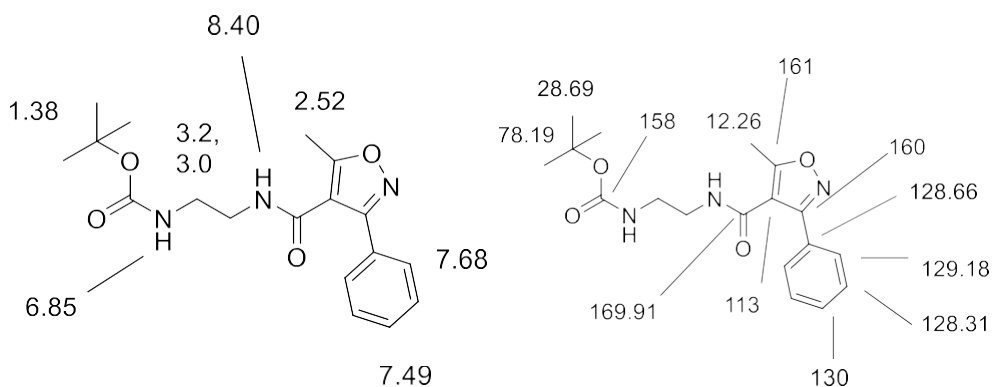
^{§§} También se intentó sintetizar, sin éxito, otros análogos de nIz como el compuesto **12**, el compuesto **6** y el compuesto **7**, pero sólo se logró sintetizar algunos intermediarios propuestos. Esos resultados no se publican en este trabajo.

Tabla 11. Resultados de síntesis.

Compuesto	Rendimiento	Punto de fusión	Espectroscopía
13	60 %	No disponible	No disponible
14	75 %	160-161°C	RMN- ¹ H, RMN- ¹³ C, EM (DART+)
15	95 %	No disponible	No disponible
11	63 %	161-162 °C	RMN- ¹ H, RMN- ¹³ C, EM (DART+)

Caracterización espectroscópica de los compuestos sintetizados.

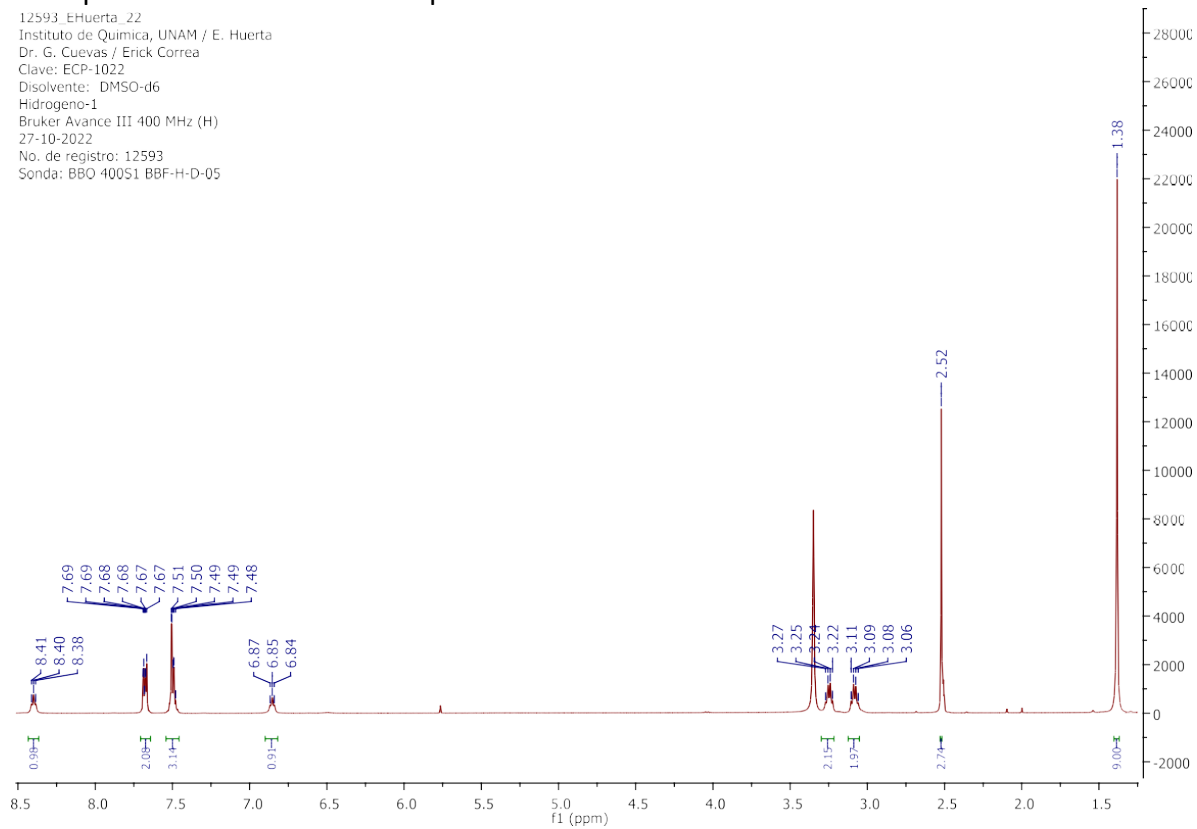
Caracterización del compuesto **14**.



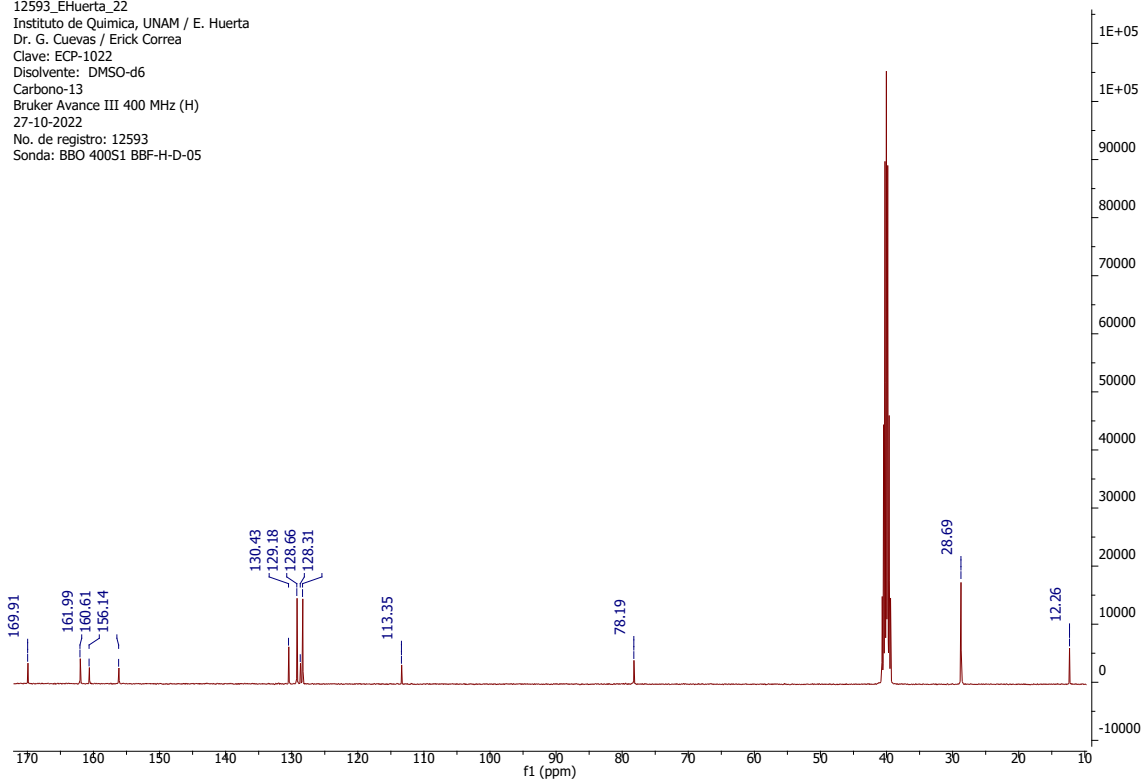
PM = 345.40 g/mol

Espectro de RMN-¹H del compuesto **14**.

12593_EHuerta_22
 Instituto de Química, UNAM / E. Huerta
 Dr. G. Cuevas / Erick Correa
 Clave: ECP-1022
 Disolvente: DMSO-d₆
 Hidrogeno-1
 Bruker Avance III 400 MHz (H)
 27-10-2022
 No. de registro: 12593
 Sonda: BBO 400S1 BBF-H-D-05

Espectro de RMN-¹³C del compuesto **14**.

12593_EHuerta_22
 Instituto de Química, UNAM / E. Huerta
 Dr. G. Cuevas / Erick Correa
 Clave: ECP-1022
 Disolvente: DMSO-d₆
 Carbono-13
 Bruker Avance III 400 MHz (H)
 27-10-2022
 No. de registro: 12593
 Sonda: BBO 400S1 BBF-H-D-05

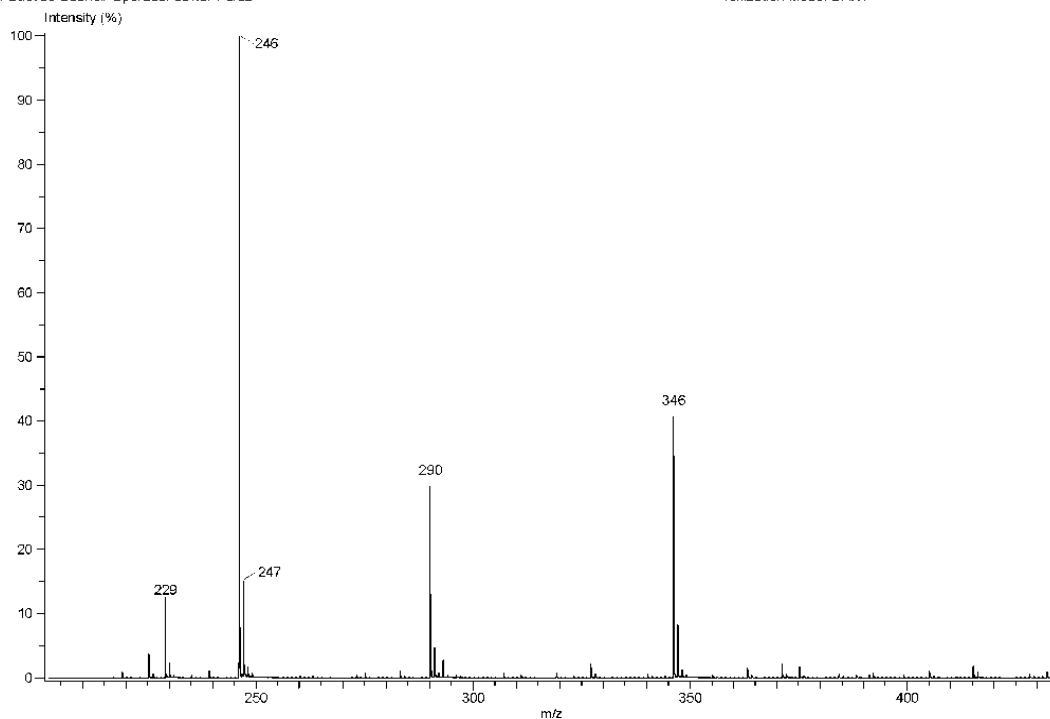
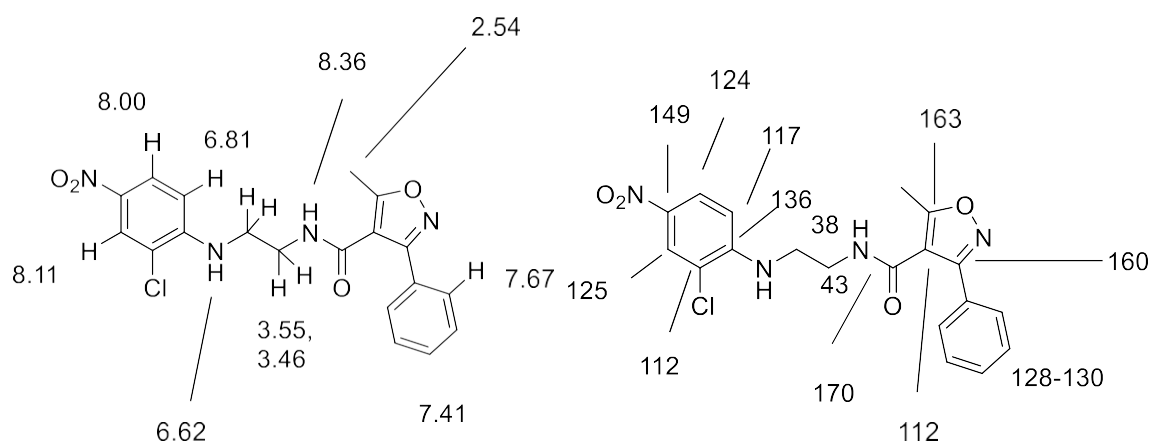


Espectro de Masas (DART+) del compuesto **14**.

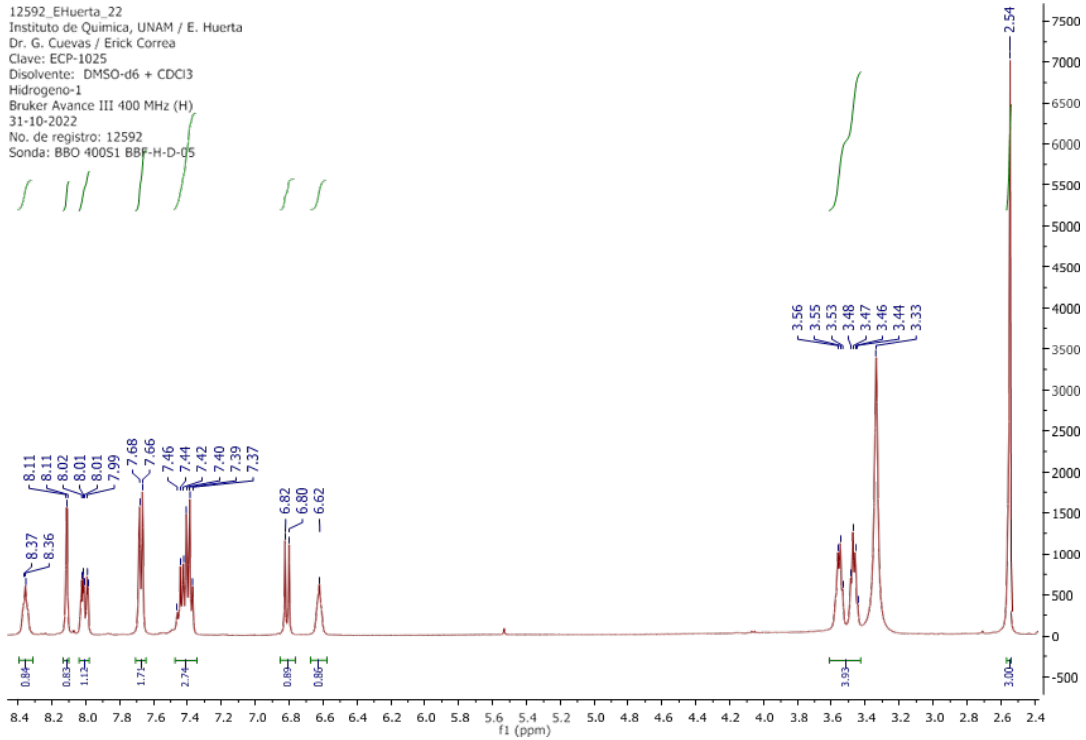
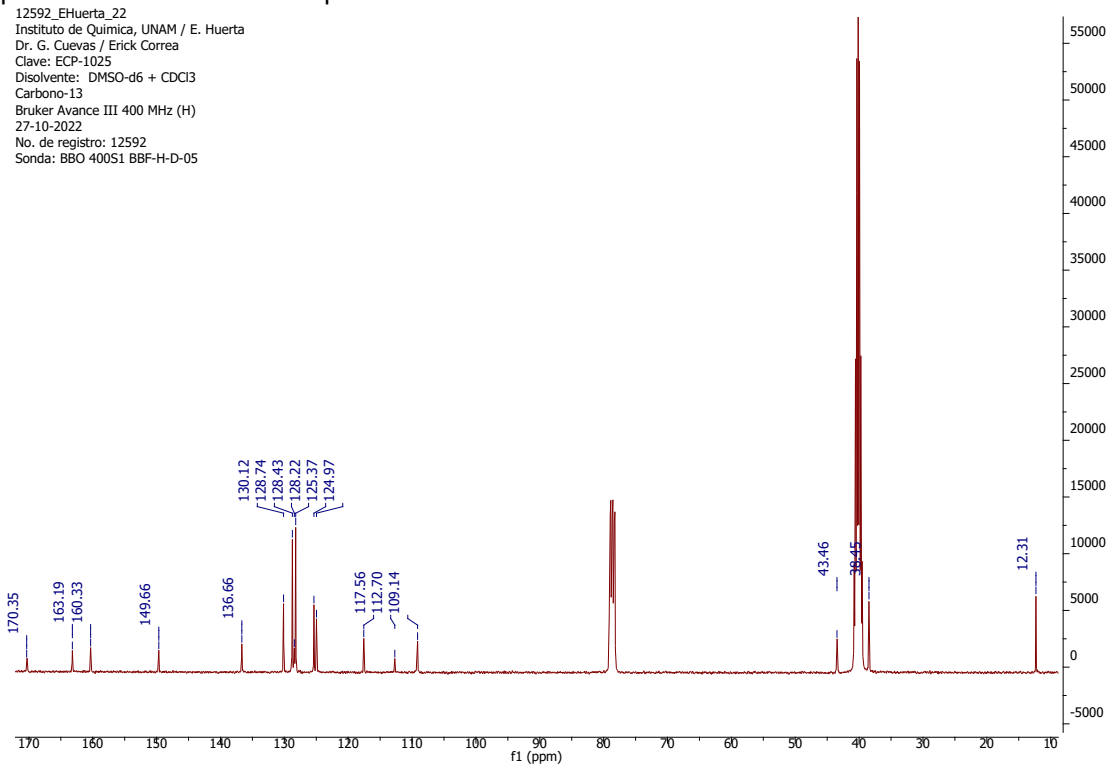
INSTITUTO DE QUIMICA, UNAM
LABORATORIO DE ESPECTROMETRIA DE MASAS

Acq. Data Name: 3630 ECP-1022
Creation Parameters: Average(MS[1] Time:0.1)
Dr. Cuevas Gabriel/ Operador Javier Perez

Experiment Date/Time: 10/25/2022 4:11:01 PM
Instrument: JEOL The AccuTOF : JMS-T100LC
Ionization Mode: DART+

Caracterización del compuesto **11**.

PM = 400.82 g/mol

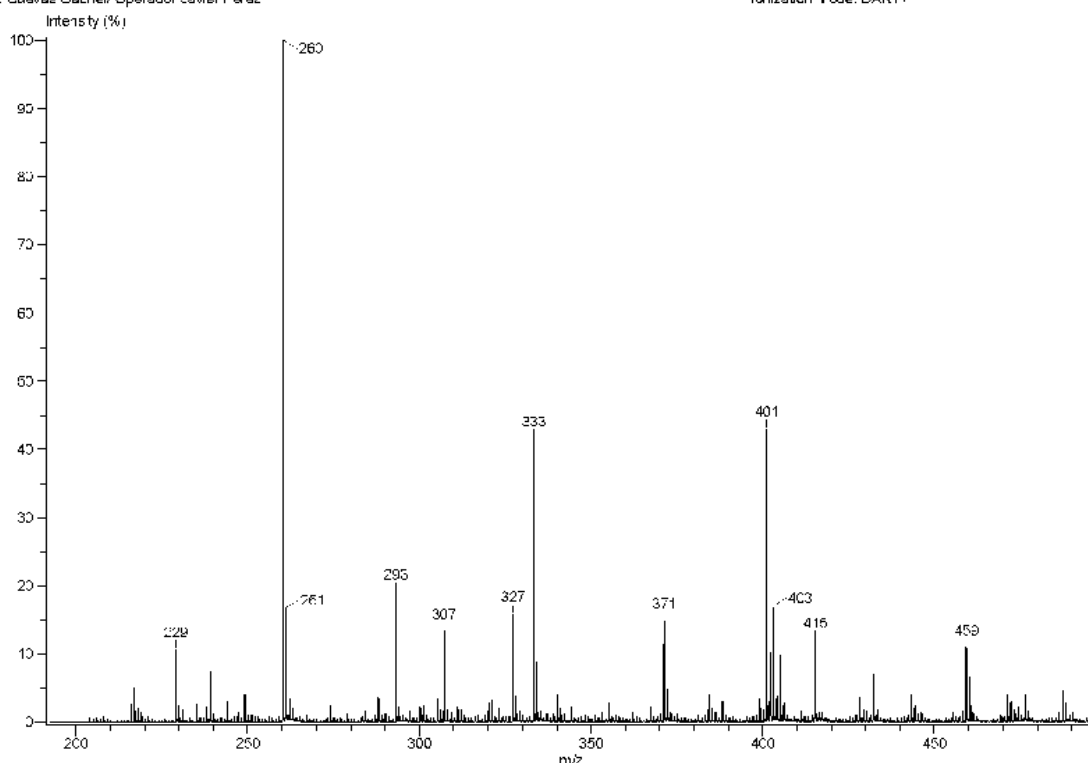
Espectro de RMN-¹H del compuesto **11**.Espectro de RMN-¹³C del compuesto **11**.

Espectro de Masas (DART+) del compuesto **11**.

INSTITUTO DE QUIMICA, UNAM
LABORATORIO DE ESPECTROMETRIA DE MASAS

Acq. Data Name: 3531 FC2-1025
Creation Parameters: Average(MS)1 Time: 2.3s
Dr. Cuevas Gabriel/ Operador: Javier Perez

Experiment Date/Time: 10/25/2022 4:13:50 PM
Instrument: JEOL The AccuTOF JMS-T100CLC
Ionization Mode: DART+



Análisis de descriptores

Los descriptores calculados con el software MOE 2019 para nucleozina y sus análogos nos ayudan a comprender algunas propiedades fisicoquímicas de los ligandos involucrados en la actividad antiviral. Es importante en el desarrollo de fármacos no solo ver a IC_{50} como el principal objetivo, sino que la biodisponibilidad también será importante para una formulación posterior. Por ello, en la [figura 42](#) se comparan los valores de los descriptores *** $SlogP$ y $logS$ de MOE para nIz y otros antivirales comerciales, utilizados

*** Descriptor $SlogP$: Logaritmo del coeficiente de partición octanol/agua ($\log P$). El conjunto de entrenamiento para $SlogP$ fue de aproximadamente 7000 estructuras. Descriptor $Log S$: Logaritmo de la solubilidad acuosa (mol/L). Esta propiedad se calcula a partir de un modelo de tipo de átomo lineal de contribución atómica con $r^2 = 0.90$ de aproximadamente 1200 moléculas.

para el tratamiento de las infecciones gripales como Oseltamivir, Zanamivir, Amantadina y Rimantidina.

También se compararon con otros antivirales de amplio espectro como Arbidol y Aciclovir, y otros como Foscarnet, Peramivir, Remdesivir, Nirmaltrevir o Ritonavir. El Naproxeno, el Acetaminofeno y la Clorfeniramina son otros fármacos que se usan en combinación con antivirales para el tratamiento de infecciones por influenza.

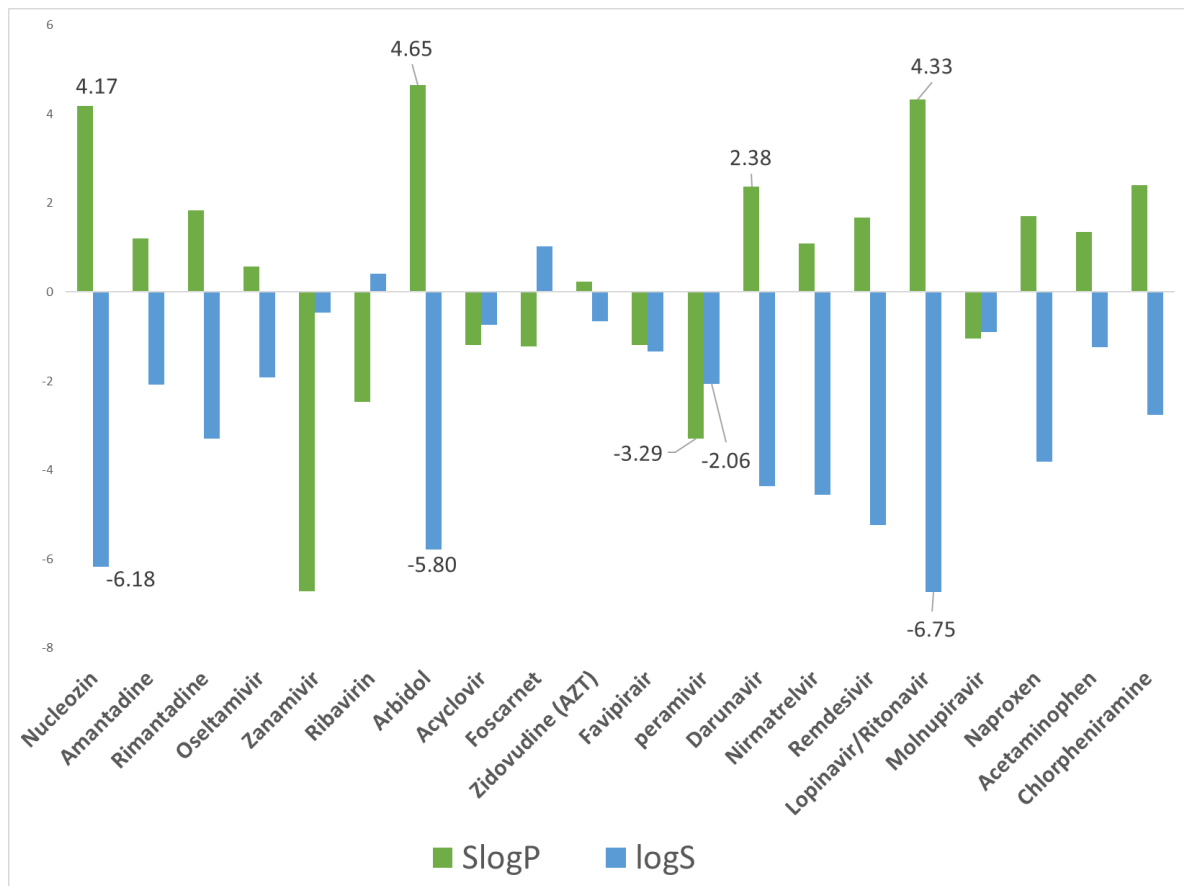


Figura 42. Valores de los descriptores SlogP y logS para nIz, algunos antivirales comerciales y algunas otras sustancias usadas para tratar los síntomas de influenza.

La lipofilidad y la solubilidad tienen un profundo impacto en el proceso ADME de un fármaco porque está estrechamente relacionado con la biodisponibilidad del fármaco, la unión a proteínas plasmáticas (PPB), la depuración metabólica, el volumen de distribución y la unión enzima/receptor. De manera general, a mayor lipofilidad menor solubilidad acuosa e inversamente, también cuanto más positivo es el valor LogS es más soluble el

compuesto y valores negativos de LogS indican una baja solubilidad. Los valores positivos de SlogP indican mayor lipofilidad.

Los valores calculados de SlogP para nlz y Gerritz 3 son 4,17 y 4,18 respectivamente. Podemos notar que son valores muy diferentes en comparación con los medicamentos antigripales comerciales, Amantadina, Rimantadina y Oseltamivir, con Zanamivir la diferencia es aún mucho mayor. También el valor de SlogP para la nlz es mayor en comparación con los valores del Naproxeno, del Paracetamol y de la Clorfeniramina. Entonces, con estas observaciones es normal que nlz y Gerritz 3 tengan una solubilidad acuosa baja en comparación con Rimantadina, Amantadina y Oseltamivir, el Zanamivir presentó la mayor solubilidad; la nlz también es menos soluble que el Naproxeno, el Paracetamol y la Clorfeniramina. Así, con estos descriptores es posible observar que, si bien la alta lipofilidad es la característica molecular que permitió este tipo de compuestos interactuar con el sitio de reconocimiento, esta misma propiedad podría ser un problema en la farmacocinética y el ADME debido a la baja solubilidad de los compuestos, en comparación con otros fármacos comerciales. Sin embargo, el Arbidol y el Aciclovir, que son antivirales de amplio espectro clínicamente aprobados, tienen una mayor lipofilidad que nlz y Gerritz 3 y también presentan valores bajos de solubilidad.⁸⁴

Para los compuestos reportados en la literatura las moléculas menos lipofílicas y más solubles son los compuestos Cheng 3h, que contiene un grupo fenol terminal en el segmento D, y el compuesto Cheng 3q que carece del átomo de cloro del segmento A. En el caso del compuesto Liao 10 la eliminación del anillo C hace menos lipofílico al compuesto ([figura 43](#)). Comparando Gerritz 3 con nlz, podemos observar que el grupo metoxilo del anillo D disminuye la solubilidad.

El nitrógeno como heteroátomo tiene un papel bastante importante en la solubilidad. Podemos observar que los compuestos Pei a y Pei b, que carecen de un átomo de nitrógeno en la piperazina, tienen menor solubilidad que la nlz. Podemos observar que la introducción de átomos nitrógenos en los sistemas aromáticos aumenta la solubilidad como en el caso de Cheng 3a y Cheng 3b. Mientras que retirar heteroátomos de los anillos

aromáticos disminuye la solubilidad como en el caso de los compuestos Cheng 5a y Cheng 5b. La introducción del átomo de nitrógeno en el sistema D mejora la solubilidad como en el caso de Gerritz 4 y Gerritz 5, sin embargo, la introducción del anillo de morfolina hace al compuesto más soluble (el valor logS para nlz es de -6.17 y para el ligando Su 11 es de -5.66) pero no presentó actividad biológica contra el virus. Algunas de estas estructuras se muestran en la [figura 44](#) y [figura 45](#).

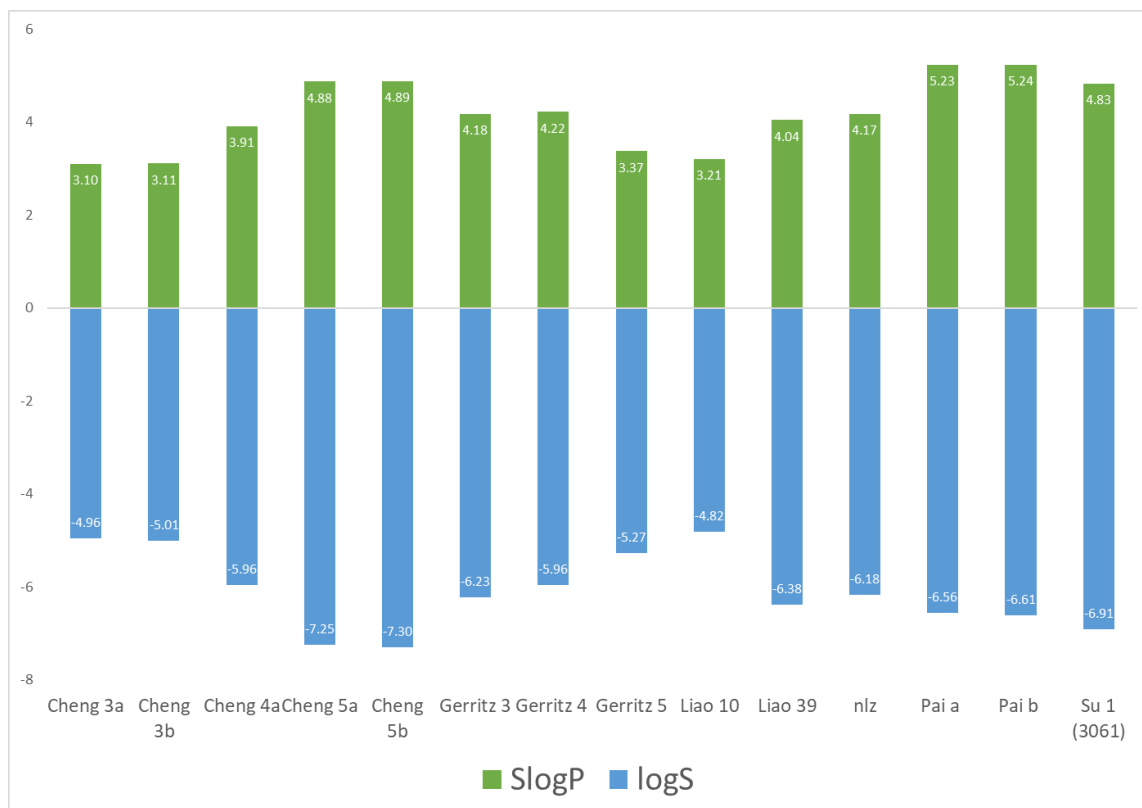


Figura 43. Valores de los descriptores SlogP y logS para nlz, Gerritz 3 y los análogos propuestos.

En la [figura 46](#) se muestra que los compuestos **1** y **2** son menos solubles ($\log S = -6.57$) que nlz ($\log S = -6.17$) debido a la introducción del puente metileno, los compuestos **3** y **4** fueron ligeramente más solubles ($\log S = -6.07$ y -6.05 , respectivamente). Cuando se eliminan los átomos de nitrógeno en el resto de piperazina, el valor de logS disminuye ($\log S = -6.61$ para **5** y -6.65 para **6**), la forma tautomérica enólica del compuesto **6** es más soluble que nlz ($\log S = -6.11$) y para el caso del compuesto **7** son aún menos solubles ($\log S = -7.66$, para la forma ceto y $\log S = -7.1$ para la forma enol). Incluso con grupos NH, el compuesto **8** es menos soluble ($\log S = -6.51$) que nlz. En este caso se puede notar que los átomos de

nitrógeno tienen un aporte importante para la solubilidad.

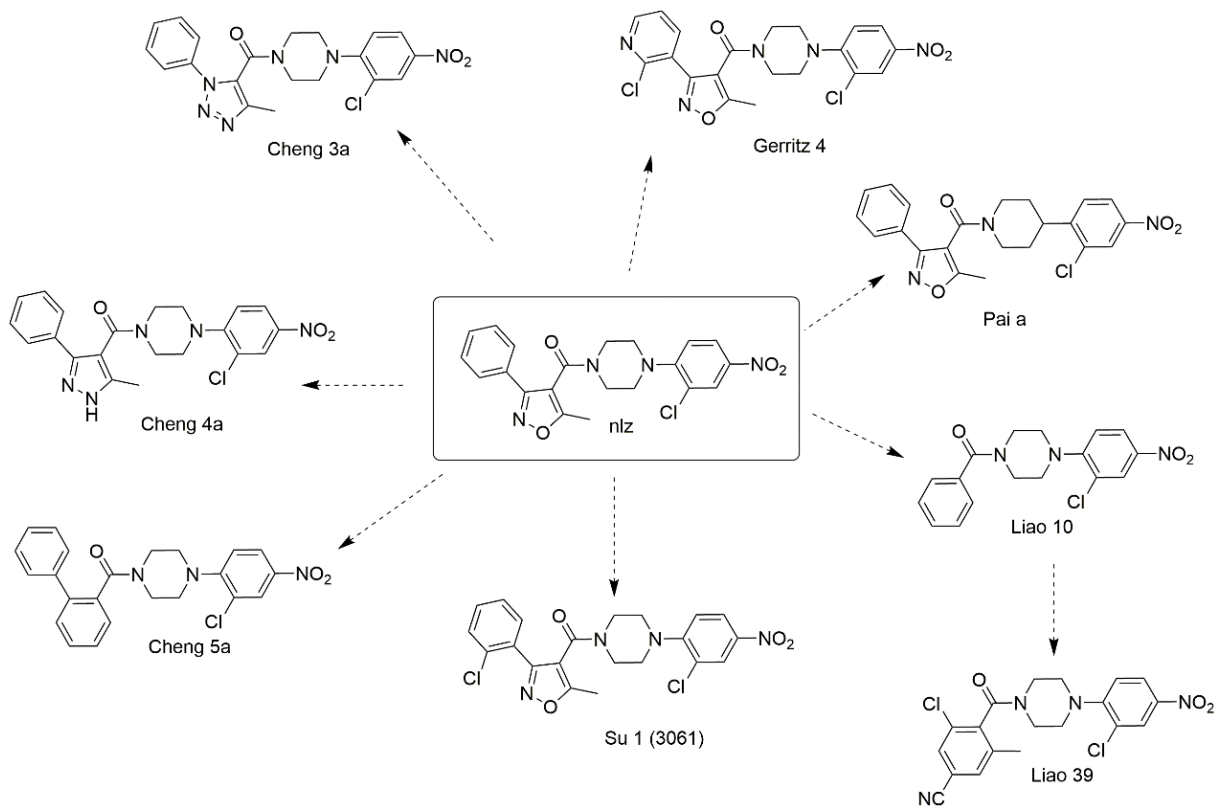


Figura 44. Estructuras químicas para algunos análogos de nucleosina.

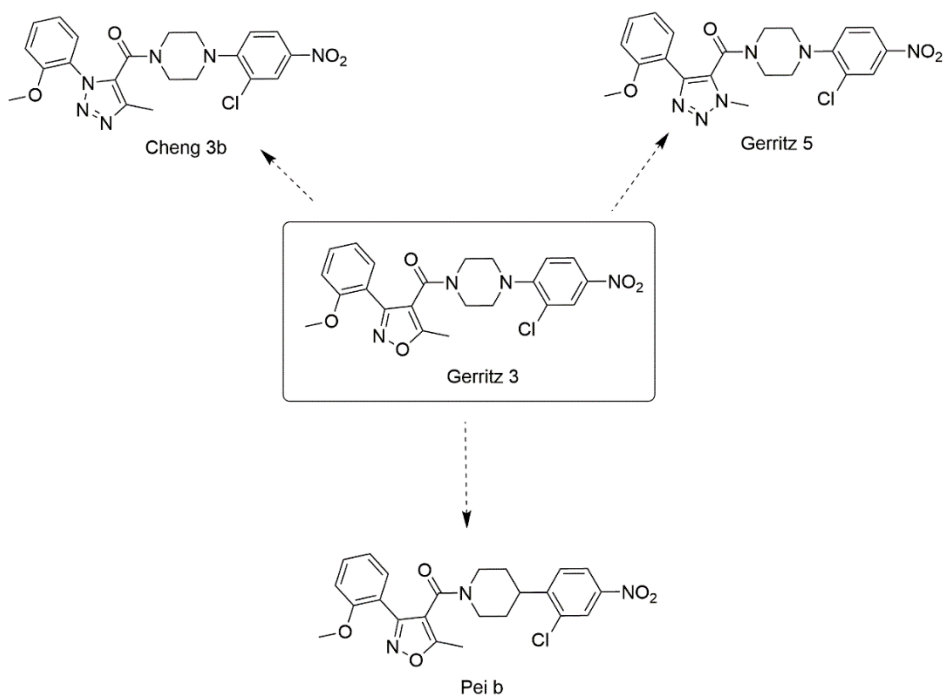


Figura 45. Estructuras químicas para algunos análogos de nucleosina.

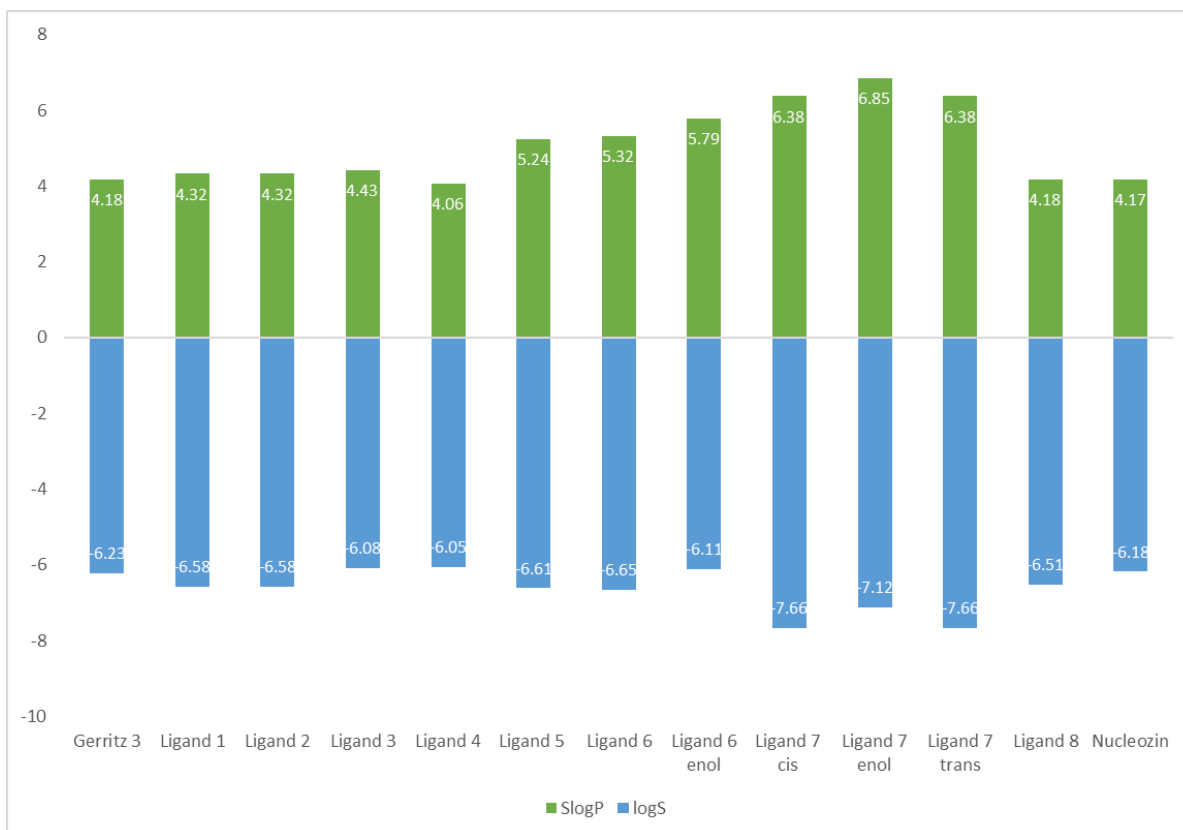


Figure 46. Valores de los descriptores de MOE llamados SlogP y LogS calculados para los análogos propuestos en este trabajo.

Evaluación biológica de análogos de nlz

En la [figura 37](#) se muestran las curvas de dilución, realizadas a muestras de cultivos celulares infectados con el virus de influenza A/Caledonia H1N1, a diferentes concentraciones del inhibidor. Cuando se utilizó nlz el resultado fue que, bajo este sistema, a una concentración de 1 μM se observó un efecto inhibitorio moderado, mientras que a 10 μM , el compuesto inhibía totalmente la propagación del virus como se muestra en la [figura 37a](#). Cuando se probaron los análogos **9** y **10**, no se observó ningún efecto en la reducción de la propagación del virus a 10 μM ([figura 37b](#) y [figura 37c](#)).

Una observación importante es que estos compuestos y nlz presentan baja solubilidad acuosa, lo que pudiera tener problemas con la biodisponibilidad y la formulación si se pensará en proponerlos como candidatos a fármacos. Cuando se realizaron estos experimentos no se contaba con el compuesto **11**, por lo que no fue incluido en la evaluación, pero se tiene como planeado realizarla en un futuro.

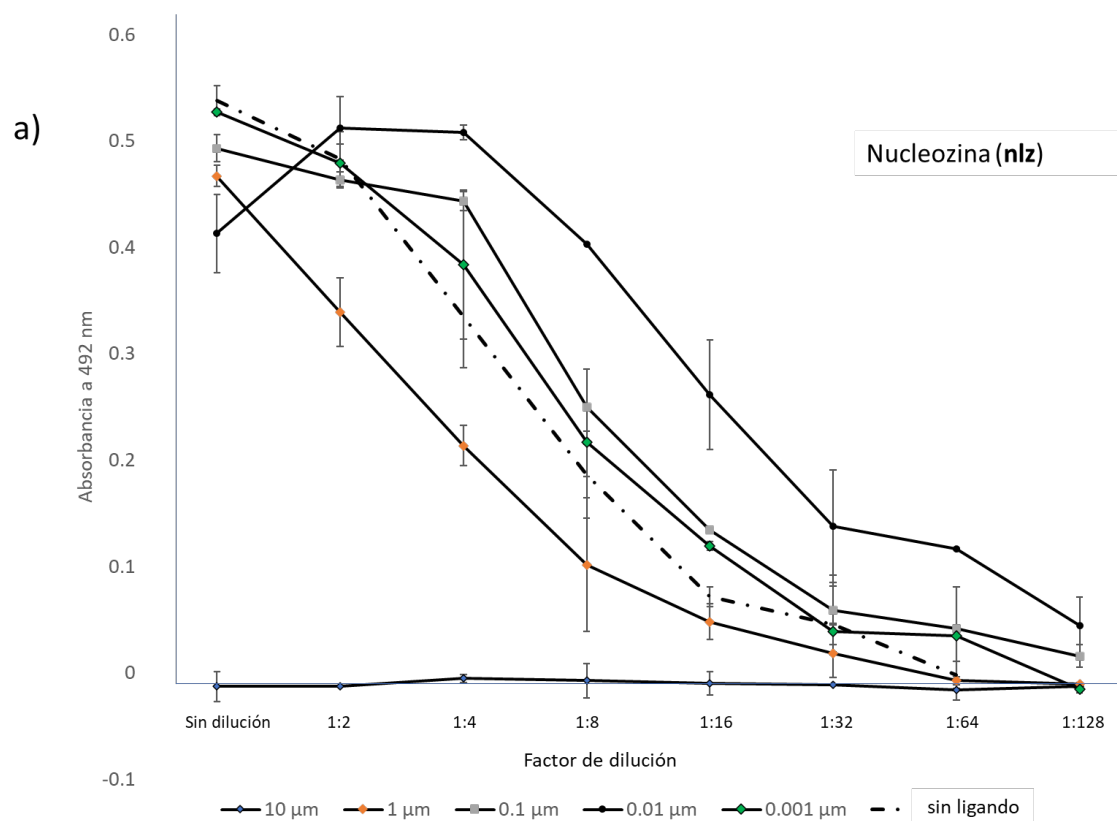


Figura 47. Evaluación biológica de los compuestos en cultivos celulares de células MDCK, se muestran las curvas de dilución, realizadas a muestras de cultivos celulares infectados con el virus de influenza A/Caledonia H1N1, a diferentes concentraciones del inhibidor. **a)** Evaluación biológica de nlz. Se observa a concentración de 1 µM un efecto inhibitorio moderado sobre la replicación del virus, y, a concentración de 10 µM, fue observada la inhibición total del crecimiento viral.

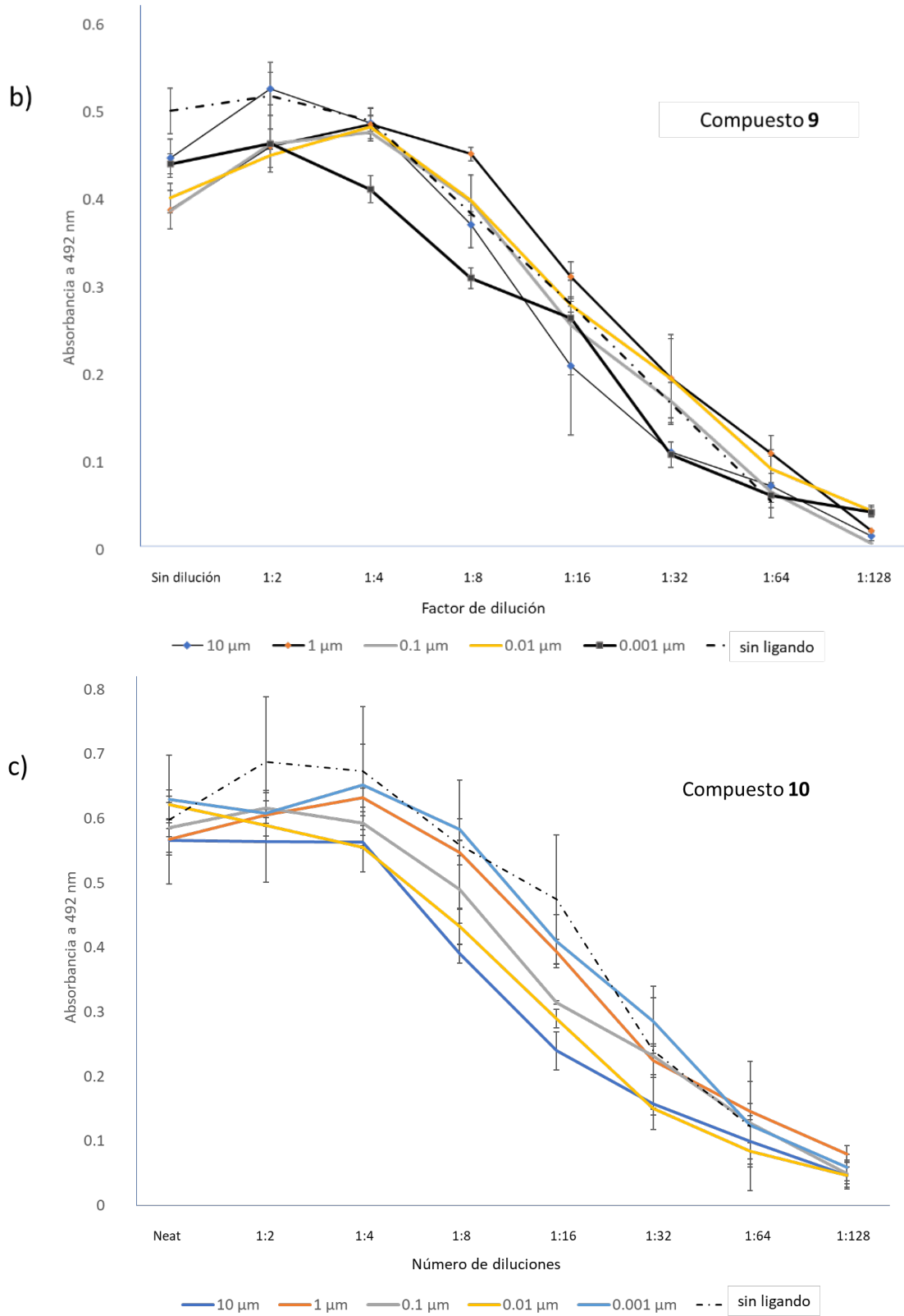


Figura 47 (continuación). Evaluación biológica de los compuestos en cultivos celulares de células MDCK, se muestran las curvas de dilución, realizadas a muestras de cultivos celulares infectados con el virus de influenza A/Caledonia H1N1, a diferentes concentraciones del inhibidor. **b)** Evaluación biológica del compuesto **9**: no se observa inhibición del virus a 10 µM. **c)** Evaluación biológica del compuesto **10**: no se observa inhibición del virus a 10 µM.

DISCUSIÓN DE RESULTADOS

La conformación del anillo de piperazina es importante en la estabilización del ligando en el sitio de unión. En el aspecto conformacional, el estudio de acoplamiento indica que la geometría debida a los átomos de nitrógeno de piperazina es crucial para mantener la conformación adecuada, la sustitución de estos átomos por átomos de carbono genera cambios de geometría afectando el modo de unión con la proteína. El cálculo de descriptores indica que la ausencia de nitrógenos disminuye la solubilidad de estos análogos ya que tienden a ser más lipofílicos que nlz o Gerritz 3, sugiriendo que estos átomos tienen un papel importante en la biodisponibilidad.

Para el reconocimiento de nlz, debe haber cierto nivel de rigidez en la molécula, brindar a la molécula de flexibilidad podría no ser conveniente debido a que se pueden generar o evitar interacciones en el sitio de reconocimiento. La modificación del puente metileno, de los compuestos diazabíclicos, no es adecuada para mejorar la solubilidad o para aumentar la potencia, por las siguientes razones:

- el cambio en la preferencia conformacional.
- el puente de metileno aumenta el impedimento estérico dentro del sitio de reconocimiento
- el puente de metileno aumenta la lipofilidad en una región cercana a los residuos polares B_Y313 y A_N309 en el sitio de reconocimiento.

Con estos resultados se propone que se puede mejorar la afinidad entre los derivados de nlz y NP mediante la introducción de grupos polares en el anillo central, que interaccionen con el residuo Y313 y/o con el residuo N309. En posible que, incluso, se puedan encontrar análogos más solubles siguiendo esta estrategia. Durante la evaluación biológica de los compuestos **9** y **10**, comparando con nlz, se observó que éstos presentan baja solubilidad acuosa, incluso, fue necesario disolverlos en dimetilsulfóxido (DMSO), lo que pudiera ser un problema para desarrollar un fármaco y posteriormente un medicamento.

Futuras modificaciones buscarán mejorar la solubilidad sin perder la lipofilidad necesaria

para interactuar con el sitio de reconocimiento. Los modelos QSAR validados permitirán predecir la actividad antiviral de este tipo de moléculas y guiar esfuerzos de síntesis de otros análogos, sin embargo, la observación de que la molécula Cheng 3t sale del dominio de aplicabilidad indica que es necesario obtener datos de actividad para análogos de nucleozina con modificaciones en el anillo de piperazina para reforzar el modelo, por ello es necesario sintetizar análogos con modificaciones en la piperazina y determinar la CI_{50} con la cepa adecuada para ampliar el dominio de aplicabilidad del modelo. Los resultados biológicos, muestran que los compuestos **9** y **10** son menos activos que nIz y el cálculo de descriptores muestra que su solubilidad es baja. El compuesto **11** que se sintetizó en este trabajo no ha sido evaluado, pero, se tienen perspectivas de realizar su evaluación y obtener el valor CI_{50} para este compuesto y los compuesto **9** y **10**.

CONCLUSIONES

La conformación del anillo de piperazina es importante en la estabilización del ligando en el sitio de unión y los cambios estructurales en esta región de la molécula tendrán consecuencias sobre la actividad antiviral.

La sustitución de piperazina por los sistemas diazabíclicos no aumenta la potencia antiviral, debido a:

- el cambio de preferencia conformacional.
- el puente de metileno aumenta el impedimento estérico dentro del sitio de reconocimiento.
- el puente de metileno aumenta la hidrofobicidad en un segmento que está cercano a dos residuos polares (B_Y313 y A_N309).

La átomos de nitrógeno de piperazina son cruciales para mantener la conformación adecuada dentro de la NP y tienen un papel importante en la solubilidad y en la biodisponibilidad.

Es necesario mejorar la solubilidad de nucleozina.

METODOLOGÍA

Metodología de acoplamiento molecular para análogos de nlz

Los estudios de acoplamiento molecular automatizado se realizaron con el programa Molecular Operating Environment 2019 (MOE; Chemical Computing Group, Montreal, QC, Canadá)⁹¹. La estructura tridimensional de la proteína se descargó del PDB con el código 3RO5. Se eliminaron las moléculas de agua, se realizó el ajuste de cargas, se agregaron hidrógenos faltantes, se ajustaron los estados de protonación a pH = 7 y se realizó la minimización de energía utilizando el campo de fuerza MMFF94x. Como referencia, se realizó el acoplamiento molecular del ligando cocrystalizado, se empleó el método de colocación Triangle Matcher y el campo de fuerza para el refinamiento fue FFMM94x, se usó ASE como las opciones 1 y 2 de rescoring. La pose del ligando cocrystalizado se reprodujo con un RMSD de 0.33.

Metodología de modelos QSAR de predicción de actividad inhibitoria con análogos de nlz

Los datos de actividad biológica se buscaron en la literatura.^{16,86,89} Se encontraron valores de CI_{50} reportados con cuatro principales cepas del virus de influenza que incluyen A/WSN/33 H1N1, A/PR/8/34 H1N1, H1N1 Ca/07/09 y A/HK/8/68 y H3N2 evaluados mediante técnicas de Ensayo de Reducción en Placa (PRA, *Plaque Reduction Assay* en inglés). Para este trabajo se tomaron solo aquellos valores de CI_{50} que se obtuvieron con la cepa A/H1N1/WSN/33; se obtuvieron 74 datos, a este conjunto de datos se le nombrará grupo de datos (Data Set). La CI_{50} y su respectivo valor en escala p se muestra en la [tabla 12](#) junto a los códigos SMILES de cada compuesto.

Las 76 estructuras de los inhibidores de NP se dibujaron en el programa ChemDraw⁹², se obtuvo el código SMILES y este se usó para modelar las estructuras en el programa MOE 2019, donde se formó una base de datos en formato .mdb, posteriormente, y de acuerdo con el manual de MOE 2019, se agregaron los hidrógenos faltantes, se realizó el ajuste del estado de protonación a pH = 7, se realizó el ajuste de cargas y la minimización de la energía con el campo de fuerza MMFF94x. Esta base de datos se guardó como un archivo tipo .mol2 y fue utilizado para el cálculo de descriptores con el programa DRAGON versión 7.0 el cual nos ayudó a calcular 4185 descriptores moleculares 2D y 3D para los elementos del grupo de datos. La información de los descriptores se guardó como un archivo .csv. y se migró a la plataforma de Google Colaboratory, también conocida como Colab, la cual trabaja con lenguaje de programación de Python.

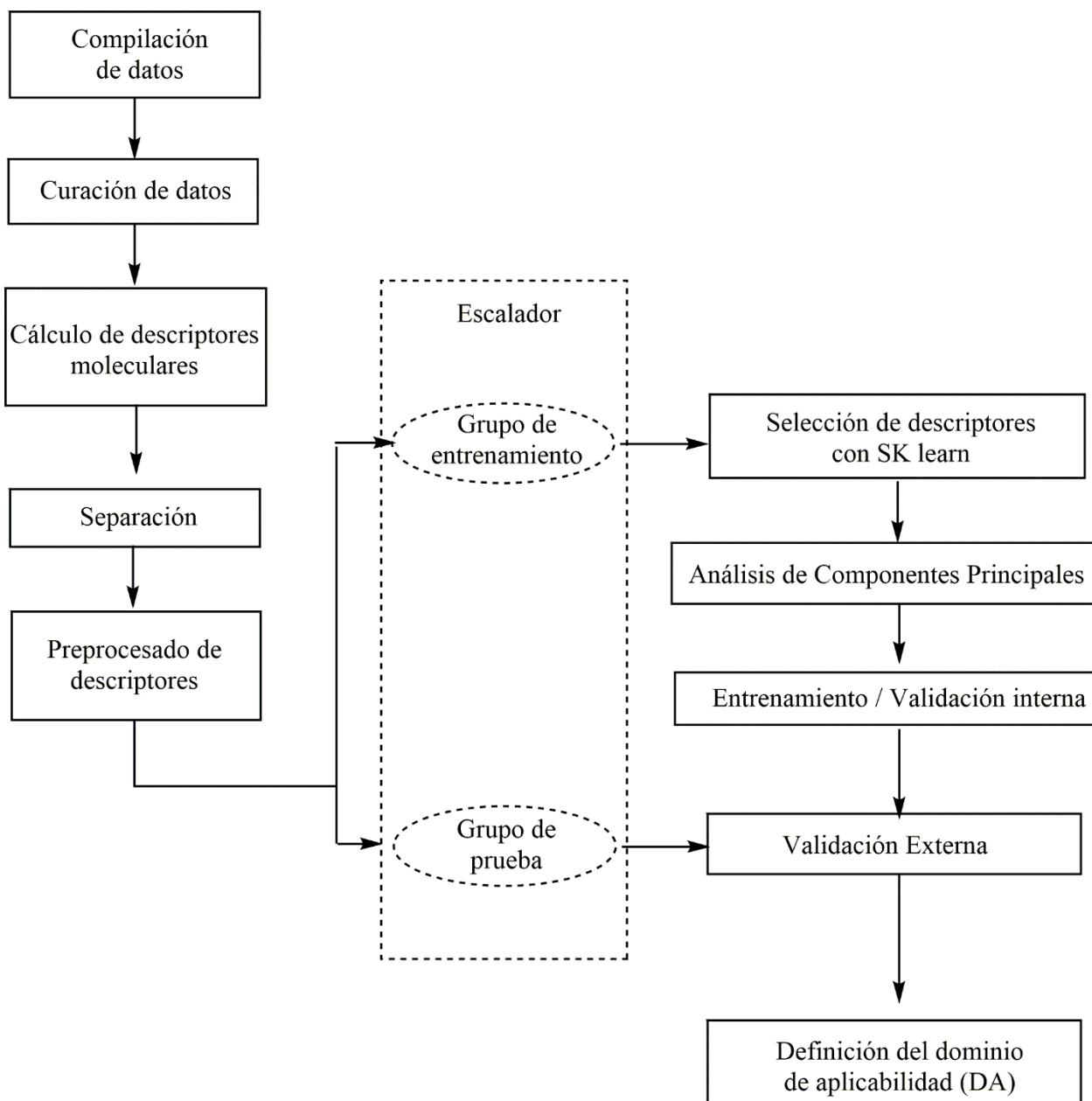


Figure 48. Diagrama de flujo con el procedimiento realizado para obtener el modelo de predicción de actividad biológica.

Utilizando el algoritmo VarianceThreshold se realizó un preprocesamiento de descriptores para eliminar aquellos descriptores constantes y semiconstantes. Posteriormente se realizó la separación de los datos en grupo de entrenamiento (train set), que son los datos que nos ayudarán a construir el modelo y grupo de prueba (test set), que son datos que nos ayudarán a evaluar el poder predictivo del modelo. Posteriormente se realizó un escalamiento tipo estandarización tomando como referencia al grupo de entrenamiento. Se utilizó el algoritmo FeatureSelection de Scikit Learn⁹³ para seleccionar los mejores 65 descriptores moleculares y posteriormente se utilizó la técnica de Análisis de Componentes Principales (PCA, Principal Components Analysis) para realizar la reducción de

dimensionalidad y formar 15 Componentes Principales. Estos se utilizaron para realizar el entrenamiento con el algoritmo de SVR de Sci-kit Learn. Se realizó validación cruzada Leave Many Out (LMO) con 10 folds y se realizó la validación externa con el grupo de prueba, se calcularon los coeficientes de determinación q^2_{CV-LMO} , r^2 , q^2_{ext} , así como los respectivos valores de MAE y MSE para cada uno de los grupos. El dominio de aplicabilidad se definió por el método del apalancamiento. El código se encuentra disponible en la siguiente dirección de git hub: <https://github.com/erickcorrea3092/nlzQSAR>

Tabla 12. Valores de Cl_{50} y su respectivo valor de $-\log(Cl_{50})$ para los compuestos reportados en la literatura.

Entrada	Ligando	Cl_{50} (μM)	pCl_{50}	Código SMILES
1	Cheng 2a	4.83	5.32	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(onc1C)-c1cccc1</chem>
2	Cheng 2b	0.71	6.15	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(onc1C)-c1cccc1OC</chem>
3	Cheng 3a	3.13	5.50	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1</chem>
4	Cheng 3b	0.68	6.17	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1OC</chem>
5	Cheng 3c	5.35	5.27	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cc(OC)ccc1</chem>
6	Cheng 3d	5.61	5.25	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1ccc(OC)cc1</chem>
7	Cheng 3e	1.2	5.92	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1Cl</chem>
8	Cheng 3f	4.89	5.31	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cc(Cl)ccc1</chem>
9	Cheng 3g	4.93	5.31	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1ccc(Cl)cc1</chem>
10	Cheng 3h	1.58	5.80	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1O</chem>
11	Cheng 3i	2.4	5.62	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1C</chem>
12	Cheng 3j	3.22	5.49	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1OCC</chem>
13	Cheng 3k	8.42	5.07	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1OC(C)C</chem>
14	Cheng 3m	1.34	5.87	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1c(OC)cccc1OC</chem>
15	Cheng 3n	8.36	5.08	<chem>Clc1cc([N+](O)=O)ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1OC</chem>
16	Cheng 3p	48.12	4.32	<chem>Clc1cccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1OC</chem>
17	Cheng 3q	16.98	4.77	<chem>O(C)c1cccc1-n1nnc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1</chem>
18	Cheng 3t	35.45	4.45	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1C)-c1cccc1OC</chem>

Tabla 12 (continuación). Valores de Cl_{50} y su respectivo valor de $-\log(Cl_{50})$ para los compuestos reportados en la literatura.

Entrada	Ligando	Cl_{50} (μM)	pCl_{50}	Código SMILES
19	Cheng 4a	3.38	5.47	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(n[nH]c1C)-c1cccc1</chem>
20	Cheng 4b	0.72	6.14	<chem>Clc1cccc1-c1n[nH]c(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
21	Cheng 5a	5.24	5.28	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1-c1cccc1</chem>
22	Cheng 5b	1.57	5.80	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1-c1cccc1OC</chem>
23	Gerritz 3	0.04	7.40	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1cccc1OC</chem>
24	Gerritz 4	0.07	7.15	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1cccnc1Cl</chem>
25	Gerritz 5	0.07	7.15	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1n(nnc1-c1cccc1OC)C</chem>
26	Kao 1	0.05	7.30	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1ccc(O)cc1</chem>
27	Kao 10	25	4.60	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1c([N+](=O)[O-])cccc1Cl</chem>
28	Kao 2	0.06	7.22	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1cccc1</chem>
29	Kao 3	0.056	7.25	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1ccc(N)cc1</chem>
30	Kao 4	0.25	6.60	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1ccc(N=[N+]=[N-])cc1</chem>
31	Kao 5	0.04	7.40	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
32	Kao 6 (R)	0.21	6.68	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
33	Kao 8	5.1	5.29	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1</chem>
34	Kao 9	12	4.92	<chem>Clc1cccc(Cl)c1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1</chem>
35	Liao 10	19.52	4.71	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1</chem>
36	Liao 11	5.2	5.28	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1C</chem>

Tabla 12 (continuación). Valores de Cl_{50} y su respectivo valor de $-\log(Cl_{50})$ para los compuestos reportados en la literatura.

Entrada	Ligando	Cl_{50} (μM)	pCl_{50}	Código SMILES
37	Liao 12	14.96	4.83	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cc(ccc1)C</chem>
38	Liao 13	5.87	5.23	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1ccc(cc1)C</chem>
39	Liao 14	4.93	5.31	<chem>Clc1ccccc1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
40	Liao 15	11.14	4.95	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cc(Cl)ccc1</chem>
41	Liao 16	7.66	5.12	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1ccc(Cl)cc1</chem>
42	Liao 17	4.24	5.37	<chem>Brc1ccccc1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
43	Liao 18	11.28	4.95	<chem>Brc1cc(ccc1)C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
44	Liao 19	7.85	5.11	<chem>Brc1ccc(cc1)C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
45	Liao 20	6.2	5.21	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1[N+](=O)[O-]</chem>
46	Liao 21	27.19	4.57	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cc([N+](=O)[O-])ccc1</chem>
47	Liao 22	3.01	5.52	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1ccc([N+](=O)[O-])cc1</chem>
48	Liao 23	4.72	5.33	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1O</chem>
49	Liao 24	12.84	4.89	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1N</chem>
50	Liao 25	9.69	5.01	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1CC</chem>
51	Liao 26	11.55	4.94	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1C1CCCC1</chem>
52	Liao 27	6.16	5.21	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1cccc1OC</chem>
53	Liao 28	3.5	5.46	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1ccc(cc1)C#N</chem>
54	Liao 29	66.03	4.18	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1ccc(cc1)C(OC)=O</chem>
55	Liao 30	0.65	6.19	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(cccc1N)C</chem>
56	Liao 31	3.28	5.48	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(cccc1[N+](=O)[O-])C</chem>
57	Liao 32	1.07	5.97	<chem>Brc1cccc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
58	Liao 33	0.52	6.28	<chem>Clc1cccc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
59	Liao 34	4.88	5.31	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(cccc1C)C</chem>
60	Liao 35	2.11	5.68	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1ccc([N+](=O)[O-])cc1C</chem>

Tabla 12 (continuación). Valores de Cl_{50} y su respectivo valor de $-\log(Cl_{50})$ para los compuestos reportados en la literatura.

Entrada	Ligando	Cl_{50} (μM)	pCl_{50}	Código SMILES
61	Liao 36	2.9	5.54	<chem>Brc1cc(C)c(cc1)C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
62	Liao 37	1.23	5.91	<chem>Brc1cc([N+](=O)[O-])ccc1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
63	Liao 38	5.33	5.27	<chem>Brc1cc(F)c(cc1)C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
64	Lia0 39	0.27	6.57	<chem>Clc1cc(cc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl)C#N</chem>
65	Nucleozin (nlz)	0.06	7.22	<chem>Clc1cc([N+](=O)[O-])ccc1N1CCN(CC1)C(=O)c1c(noc1C)-c1cccc1</chem>
66	Su 1 (3061)	0.07	7.15	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1Cl</chem>
67	Su 10 (9168)	19.5	4.71	<chem>Clc1cccc([N+](=O)[O-])c1N1CCN(CC1)C(=O)c1c(noc1C)-c1cccc1</chem>
68	Su 2 (4332)	1.3	5.89	<chem>Clc1cc([N+](=O)[O-])ccc1N1CC(N(CC1)C(=O)c1c(noc1C)-c1cccc1)C</chem>
69	Su 3 (2130)	3.5	5.46	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc(Cl)cc1[N+](=O)[O-]</chem>
70	Su 4 (3822)	1.2	5.92	<chem>o1nc(-c2cccc2OC)c(C(=O)N2CCN(CC2)c2ccc([N+](=O)[O-])cc2)c1C</chem>
71	Su 5 (6074)	5	5.30	<chem>Clc1cccc(Cl)c1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1</chem>
72	Su 6 (0927)	14	4.85	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1c([N+](=O)[O-])cccc1Cl</chem>
73	Su 7 (0131)	17.5	4.76	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1cc(N2CCCC2)c([N+](=O)[O-])cc1</chem>
74	Su 9 (4427)	21	4.68	<chem>Clc1cccc1-c1noc(C)c1C(=O)N1CCN(CC1)c1ccc([N+](=O)[O-])cc1</chem>

Metodología de síntesis de análogos de nIz

Síntesis del compuesto **13**

En un matraz de fondo redondo de 250 mL se disolvieron 6 mL de etilendiamina en 60 mL de DCM seco en atmósfera de nitrógeno y se añadió lentamente con una jeringa una solución de Boc₂O en 60 mL de DCM. La reacción se controló por CCF (Cromatografía en Capa Fina). Al final de la reacción, la fase orgánica se lavó con 60 mL de agua dos veces con salmuera, se secó sobre Na₂SO₄ y se filtró, el DCM se evaporó en un evaporador rotatorio y el producto se purificó por cromatografía en columna. Se obtuvo un sólido cremoso blanco. Rendimiento del 60%. No se cuenta con espectroscopía.

Síntesis del compuesto **14**

En un matraz de fondo redondo de 50 mL se colocaron 647 mg del compuesto **II**, se añadieron 2 gotas de DMF seco y se colocaron en atmósfera de nitrógeno, se añadieron 10 mL de DCM seco, la mezcla de reacción se enfrió en baño de hielo y a 350 °C. Se añadieron lentamente 60 µL de cloruro de oxalilo y luego se dejó en agitación durante 20 minutos. Por otra parte, se preparó una solución de 10 mL de DCM con 577 µL de TEA y 510 mg del compuesto **13**, y éste se agregó lentamente a la mezcla de reacción y luego se dejó en agitación durante 30 minutos, luego, 20 mL de agua, se separaron las fases y la fase orgánica se lavó una vez con 20 mL de solución de Na₂CO₃, una vez más con agua y una vez con salmuera. La fase orgánica se secó sobre Na₂SO₄, se filtró y el DCM se evaporó en un evaporador rotatorio. El compuesto se purificó por cromatografía en columna (Hexano:AcOEt, 6:4). Se obtuvo sólido blanco con 75 % de rendimiento. **Compuesto 14**: p.f.: 160-161 °C; **RMN-¹H** (400 MHz, DMSO-d₆) δ 8.40 (t, J = 5.59 Hz, 1H) 7.60 (dd, J = 6.7, 2.9 Hz, 1H), 7.71 – 7.64 (m, 2H), 7.54-7.46(m, 3H), 6.85 (t, J = 5.56 Hz, 1H), 3.25 (dd, J = 12.25, 6.18 Hz, 2H), 3.08 (dd, J = 12.28, 6.19 Hz, 2H), 2.52 (s, 3H), 1.38 (s, 9H). **RMN-¹³C** (400 MHz, DMSO-d₆) δ 169.91, 161.99, 160.61, 156.14, 130.43, 129.18 (2C), 128.66, 128.31(2C), 113.35, 78.19, 28.69, 12.26. EM (DART+) m/z (% intensity): 346 (40, M+), 290 (30), 246(100), 229 (13).

Síntesis del compuesto **15**

En un matraz de fondo redondo de 50 mL se disolvieron en 10 mL de DCM 2.3 mmol de la *N*-Boc-*N'*-(4-nitroaril)amina monoprottegida (compuestos **15**, **20**, **21**) y se agregaron lentamente 840 µL de MeSO₃H. El final de la reacción se determinó por TLC en 5 minutos. Se añadió solución saturada de Na₂CO₃ hasta pH > 8, se extrajo con DCM (2 X 10 mL), se secó sobre Na₂SO₄ anhidro y se evaporó hasta sequedad obteniendo un líquido incoloro con 95 % de rendimiento. No se cuenta con espectroscopía.

Síntesis del compuesto 11

En un matraz de fondo redondo de 10 mL se colocaron 272 mg del compuesto **26**, 5 mL de DMF, 196 mg del compuesto **13** y 308 mg de K_2CO_3 y se dejó en agitación durante 72 hrs. Se evaporó la DMF y luego se disolvió el producto crudo en 10 mL de AcOEt, se lavó dos veces con 10 mL de agua y luego con 10 mL de salmuera. La fase orgánica se secó sobre Na_2SO_4 , se filtró y el solvente se eliminó en un evaporador rotatorio. El producto se purificó por cromatografía en columna de gel de sílice con DCM como fase móvil. Se obtuvo un sólido amarillo brillante. Rendimiento de 63 %. **Compuesto 11**: m.p.161-162 °C; **RMN-¹H** (400 MHz, DMSO-d₆ + CDCl₃) δ 8.36 (br, 1H) 8.11 (d, J = 2.54 Hz, 1H), 8.00 (m, 1H), 7.67 (dd, J = 6.89, 2.54 Hz, 2H), 7.46 - 7.37 (m, 3H), 6.81 (d, J = 9.23 Hz, 1H), 6.62 (br, 1H), 3.55 (dd, J = 11.21, 5.5 Hz, 2H), 3.46 (dd, J = 11.01, 5.45 Hz, 2H), 2.54 (s, 3H). **RMN-¹³C** (400 MHz, DMSO-d₆ + CDCl₃) δ 170.32 (d, J = 6.51 Hz), 163.19, 160.3, 149.66, 136.66, 130.12, 128.74 (2C), 128.43, 128.22 (2C), 125.37, 124.97, 117.56, 112.70, 109.14, 43.46, 38.45, 12.31. **EM** (DART+) m/z (% intensity): 403 (15, M+2), 401 (47, M+), 333(42), 260 (100).

Metodología de evaluación biológica de análogos de nlz

Se sembraron células de riñón canino Madin-Darby (MDCK) a razón de 104 células por pocillo, en placas de 96 pocillos, en medio de Eagle modificado por Dulbecco (DMEM) suplementado con -L-glutamina 4 mM, piruvato de sodio 1 mM, penicilina 50 U/ml, 50 µg/ml de estreptomina y suero de ternera fetal (FCS) al 10 %, y se incubó durante 24 h a 37 °C. El medio se retiró y las células se infectaron durante 1 h a 37 °C con la cepa A/Caledonia H1N1 del virus de la gripe A, a una concentración de 0,001 unidades formadoras de placa (PFU, Plaque-Forming Unit) por célula. Se retiró el inóculo y se añadieron a las células diferentes concentraciones de nlz o sus análogos, previamente disueltos en DMSO, (de 1 µM a 10 µM), diluidas en DMEM sin suero, y se incubaron. Después de 48 horas, se retiraron los sobrenadantes, se congelaron a -20 °C y luego se titularon en una nueva placa de 96 pocillos con MDCK y se incubaron durante 1 hora. a 37 °C. Se retiró el inóculo, se añadió DMEM sin suero a las células y se incubaron durante 72 horas. a 37 °C. Se retiró el medio, las células se fijaron con paraformaldehído al 4% y se permeabilizaron con metanol-acetona 1:1. Luego, se realizó un ensayo ELISA en las células fijadas-permeabilizadas, utilizando un anticuerpo monoclonal de ratón específico para la proteína viral interna M1 (ATCC; HB-64), seguido de un anticuerpo policlonal secundario Ig de cabra anti-ratón conjugado con peroxidasa de rábano picante (HRP, *Horseradish Peroxidase* en inglés) (Jackson). La reacción se desarrolló con el sustrato diclorhidrato de o-fenilendiamina (Sigma) y se leyó en un lector ELISA (Biotech) a 490 nm. La intensidad de la señal se correlaciona con la cantidad de M1 producida durante la infección. En todos los pasos, DMEM sin suero contenía 1 µg de tripsina tratada con L-1-tosilamida-2-feniletilclorometilcetona (TPCK; Sigma) para la activación de la infectividad del virus. Todos los reactivos de cultivo de tejidos eran de GIBCO.⁹⁴

ANEXOS



Sociedad Química de México, A. C.

otorga la presente

CONSTANCIA

a

Erick Correa Padilla, Karina Martínez Mayorga, Abraham Madarriaga Mazon,
Alejandro Hernandez Cano, Eduardo Ruiz Padilla, Ignacio Regla, Gabriel Cuevas,
Yunuen Acevedo Betancur, Fernando Esquivel Guadarrama

por su participación presentando el Trabajo Profesional en Modalidad Cartel Presencial titulado

Efecto de la modificación del anillo de piperazina sobre la actividad anti-influenza de nucleozina. CISQM-QS-CP19

en el Marco de las Actividades del
Congreso Internacional de la Sociedad Química de México 2022,

"Una Química: Muchas Voces"

29 de agosto al 2 de septiembre de 2022

Mérida, Yucatán, México.



SOCIEDAD QUÍMICA
DE MÉXICO, A.C.
"La química nos une"

Dr. David Cuintanar Guerrero

Presidente Nacional Electo
Presidente de Congresos
Sociedad Química de México, A.C.

M. en C. Wendy Fanny Brito Loeza

Presidente del Comité Organizador Local CISQM 2022
Pro-secretaría, Comité Ejecutivo Nacional
Sociedad Química de México, A.C.

Dr. Alfonso Ramón García Márquez

Cordinador del CISQM2022
Tesorero, Comité Ejecutivo Nacional
Sociedad Química de México, A.C.

Effect of piperazine ring modification on the anti-influenza activity of nucleozin

Erick Correa-Padilla^{1,2}, Alejandro Hernandez-Cano^{1,3}, Eduardo Ruiz-Padilla², Ignacio Regla², Gabriel Cuevas¹, Yunuen Acevedo-Betancur⁴, Fernando Esquivel-Guadarrama⁴, Abraham Madariaga-Mazon^{1,5}, Karina Martinez-Mayorga^{1,5}

¹ Instituto de Química Campus Mérida, Universidad Nacional Autónoma de México, Ucu, Yucatán, México.
² Facultad de Estudios Superiores Zaragoza-UNAM, Batalla 5 de mayo s/n esquina fuerte de Loreto, Colonia Ejército de Oriente, Iztapalapa, Ciudad de México, México.
³ Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, 04510, México.
⁴ Laboratorio de Inmunología Viral, Facultad de Medicina, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México.
⁵ Instituto de Química, Universidad Nacional Autónoma de México, Ciudad de México, 04510 México.
⁶ Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Unidad Mérida, Universidad Nacional Autónoma de México, Mérida, Mexico

Summary

The effect of piperazine modification in the nucleozin (nlz) structure on anti-influenza activity was studied. A group of molecules with modifications in the piperazine ring including rigid, flexible changes, with modifications in the nitrogen atoms and in the number of spacer carbons between them was proposed. Docking studies indicate that the piperazine and the carbons of the nitrogen atoms are necessary to stabilize the ligand within the recognition site. QSAR models allowed the prediction of antiviral activity with metrics of $r^2 = 0.93$, $q^2_{CV-LMO} = 0.57$ and $q^2_{ext} = 0.56$, for the best model. The synthesis and in vitro biological evaluation of analogs were also carried out. Finally, it was observed that the absence of nitrogen decreases the solubility and increases the lipophilicity, so they have an important role in bioavailability.

Introduction

Nucleoprotein (NP) is an attractive biological target for antiviral agents against influenza. Currently, there are several nucleoprotein-directed ligands such as nucleozin (nlz), an amide derived from piperazine, from which a series of derivative analogs with anti-influenza activity have been published. The binding sites of the protein with one of the ligands are formed by the groove that forms between residues Y289/N309 and the loop formed by residue Y52 and its neighboring residues. The main interactions are between the terminal aromatic segments and there is little relevant information about the role of piperazine, where it seems that it only plays a spacer role. In this work, the effect of the modification in the piperazine system on the structure of nucleozin (nlz) and its effect on antiviral activity is studied.

Methodology



Figure 1. Methodology proposed in this work

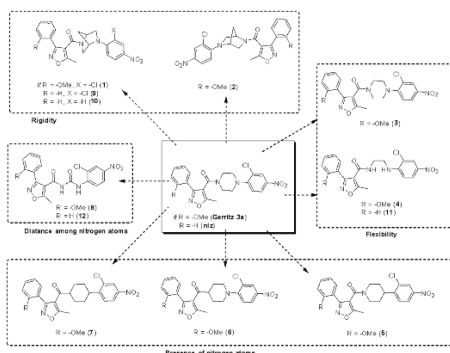


Figure 1. Compounds proposed in this work to study the influence of nucleozin piperazine on antiviral activity.

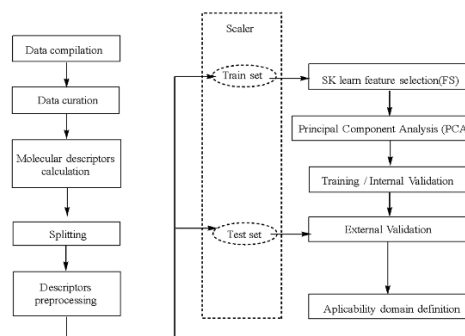
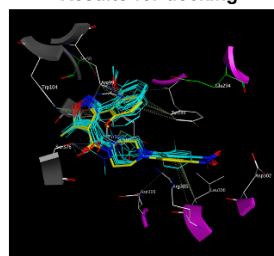


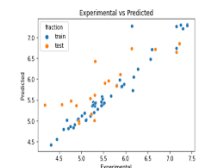
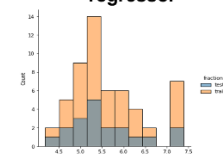
Figure 2. Workflow to perform the QSAR model.

Results for the docking



Compds	Score	RTCNucleos	APF2Gerr1	Affinity (kcal/mol)	E (kcal/mol)
1	14.77	-29.61	103.73	-13.15	147.73
2	15.31	-34.68	95.42	-13.36	147.02
3	-23.74	-24.56	-90.13	-14.76	126.06
4	-17.73	-31.68	-74.81	-13.23	194.80
5	-25.03	-28.43	-110.80	-13.41	86.75
6	-19.85	-30.10	-95.36	-13.15	119.23
6 enol	23.65	-29.16	105.22	-14.33	135.13
7 enol	-22.04	-28.76	-103.18	-13.61	95.45
7 anti	-18.90	-27.51	-110.90	-14.11	102.52
7 syn	-18.81	-27.59	-110.67	-13.51	102.66
8	-23.06	-25.31	-78.53	-13.49	66.65
Gerr1/3	18.22	-28.63	109.59	-14.10	129.68
nlz	-27.05	-28.78	-109.23	-13.68	123.62
Nitrite				-13.88	152.71

Results for the Support Vector regressor



	r^2	q^2_{CV-LMO}	q^2_{ext}
Score	0.93	0.57	0.56
MSE	-0.04	0.22	0.28

Conclusions

Piperazine is essential to maintain the proper conformation of the ligand within the recognition site. In the conformational aspect, the coupling study indicates that the geometry due to the piperazine nitrogen atoms is crucial to maintain the proper conformation, the substitution of these atoms by carbon atoms generates geometry changes, in addition, the calculation of descriptors indicates that the absence of nitrogen decreases solubility and increases lipophilicity, so these atoms have an important role in bioavailability. Modification of the methylene bridge of diazabicyclic compounds is not suitable for improving solubility or for increasing potency. This is because this change changes the conformational preference, increases steric hindrance, and increases lipophilicity. The QSAR models obtained will make it possible to predict the antiviral activity of the proposed molecules and guide efforts to synthesize other analogs. Future modifications will seek to improve solubility without losing the necessary lipophilicity to interact with the recognition site. The biological results show that compounds 9 and 10 are less active than nlz and the calculation of descriptors shows that their solubility is low.

References

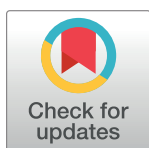
1. Jans, B.; Bernini, L.; Della, G. D.; Dhan, C.; Quilico, S.; Scharf, A. Structural Design of Novel Nucleoside Derivatives Targeting Mammalian Nucleoproteins of Influenza A Virus. *J. Biomol. Struct. Dyn.* 2018, 35 (5), 1889-1912. <https://doi.org/10.1080/07391031.2018.1493230>
2. Wu, Y.; Saqib, M.; Tahir, B.; Wang, Y. Influenza A Virus Nucleoprotein: A Highly Conserved Multi-Functional Viral Protein as a Potential Drug Target. *Chem. Biol. Drug. Des.* 2017, 2017, 2271-2285. <https://doi.org/10.1111/cbdd.12968>
3. Shen, Y. P.; Chen, Y. H.; Chen, S. Y.; Liu, M. E.; Jiao, H. T.; Wu, P. Y.; Li, H. W.; Li, F. Y.; Li, G. H.; Bao, P. H.; Tang, A. S.; Cheng, Y. S.; Wu, Y. T.; Wang, C. H.; Tai, M. D. B39: A Novel Inhibitor of Influenza Virus Nucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* 2011, 108 (26), 10515-10520. <https://doi.org/10.1073/pnas.1113170108>
4. Su, C.; Chang, T. H.; Lin, M.; Wang, K.; Huang, W.; Jia, Z. S. High-Throughput Identification of Compounds Targeting Influenza B Virus-Dependent RNA Polymerase Activity. *2016*, 167 (45), 10151-10156. <https://doi.org/10.1021/acs.jcheminf.6b00226>
5. White, K. M.; Zhang, P.; Wang, J.; Wang, P.; Liu, A.; Antonovics, B.; Chen, A.; Chandra, S.; An, D.; Wu, X.; J. Shen, S. L. Small Molecule Inhibitor of the Influenza A and B Virus Nucleoprotein. *2018*, <https://doi.org/10.1021/acs.jcheminf.8b00226>
6. Kim, Y. S. Antiviral Compounds and Methods of Making and Using Them: From Chemical to Clinical Application. *2015*, 2015 (10), 917-931, 2015.
7. Kim, R.; Yang, D.; Liu, L. S.; Choi, W. H.; Wink, L.; Lim, J.; Cho, M. P.; Cho, C. M.; Wang, P.; Cho, J. J.; Song, J.; Chung, J. D.; Rohde, J.; Cho, G.; Cho, H.; Cho, H.; Yoo, K. Y. Identification of Influenza A Nucleoprotein as an Antiviral Target. *Nat. Chem.* 2010, 2 (6), 490-495. <https://doi.org/10.1038/nchem.1010>
8. Gomez, S. W.; Casici, C.; Kim, S.; Puzos, B. C.; Dunning, C.; Dierker, C. Discovery of Small Molecules That Inhibit the Formation of Influenza Nucleoprotein Oligomers. *PLoS ONE* 2011, 6 (12), e25664. <https://doi.org/10.1371/journal.pone.0025664>
9. Jaramillo, G.; Aguirre, B.; Pineda, G.; Gomez, M. M.; Dima, A.; Pina, A.; Pina, A.; Pina, G. A. Studies of Biologically Active Bridged Diazabicyclic Compounds. *Chem. Biol. Drug. Des.* 2012, 23 (2), 251-256. <https://doi.org/10.1111/j.1177-3245.2012.01611.x>
10. Shen, Y. P.; Wang, P. W.; Thompson, P.; Zhang, Y.; Meng, F. Antiviral Agents: Nucleoprotein Inhibitors and Nucleoside Derivatives as Potential Inhibitors of Influenza Virus. *J. Med. Chem.* 1974, 17 (5), 613-617. <https://doi.org/10.1021/jm00161a012>
11. Ramonon, P.; Bernini, D.; Groll, C.; Neugebauer, J. Homocyclic and Heterocyclic Nucleoside Derivatives as Novel Anti-Influenza A Agents Targeting Virus Nucleoprotein. *J. Med. Chem.* 2012, 55 (2), 2144-2151. <https://doi.org/10.1021/jm201150a>
12. Tashir, K.; Sankar-Sankar, T.; Pineda, G.; Gomez, M. M.; Dima, A.; Pina, A.; Pina, G. A. Synthesis and in Vitro Biological Evaluation of 11-(2,2-Dimethyl-4-Carboxamide Derivatives as New Anti-Influenza A Agents Targeting Virus Nucleoprotein. *J. Med. Chem.* 2012, 55 (2), 2144-2151. <https://doi.org/10.1021/jm201150a>
13. Kim, R.; Yang, D.; Choi, Y.; Wang, M.; Kim, W. J.; Park, Y.; Kim, S.; Cho, J. Design, Synthesis and in Vitro Biological Evaluation of Novel Anti-Influenza A Agents Targeting Virus Nucleoprotein. *BMC Med.* 2020, 18 (1), 444-454. <https://doi.org/10.1186/s12916-020-01621-2>
14. Liu, J.; Cheng, H.; Wang, J.; Chen, F.; Li, Y.; Ding, K.; Jinhua, N.; Du, D.; Li, C.; Cheng, Y. Synthesis of Nucleoside Derivatives as New Influenza A Nucleoprotein Inhibitors. *Open J. Med. Chem.* 2015, 6 (4), 70. <https://doi.org/10.4236/ojmc.2015.64007>

RESEARCH ARTICLE

Modifications in the piperazine ring of nucleozin affect anti-influenza activity

Erick Correa-Padilla^{1,2}, Alejandro Hernández-Cano^{1,3}, Gabriel Cuevas¹, Yunuen Acevedo-Betancur⁴, Fernando Esquivel-Guadarrama⁴, Karina Martínez-Mayorga^{1,5,6*}

1 Institute of Chemistry, National Autonomous University of Mexico, Mexico City, Mexico, **2** Zaragoza School of Higher Education, National Autonomous University of Mexico, Mexico City, Mexico, **3** School of Science, National Autonomous University of Mexico, Mexico City, Mexico, **4** School of Medicine, Autonomous University of the State of Morelos, Cuernavaca, Morelos, México, **5** Institute of Chemistry, Campus Merida, National Autonomous University of Mexico, Merida-Tetiz Highway, Yucatán, México, **6** Institute for Applied Mathematics and Systems, Merida Research Unit, National Autonomous University of Mexico, Sierra Papacal Merida, Yucatan, Mexico

* kmtzm@unam.mx

Abstract

The infection caused by the influenza virus is a latent threat. The limited access to vaccines and approved drugs highlights the need for additional antiviral agents. Nucleozin and its analogs have gained attention for their promising anti-influenza activity. To contribute to the advancement of the discovery and design of nucleozin analogs, we analyzed piperazine-modified nucleozin analogs to increase conformational freedom. Also, we describe a new synthetic strategy to obtain nucleozin and its analogues, three molecules were synthesized and two of them were biologically evaluated *in vitro*. Although the analogues were less active than nucleozin, the loss of activity highlights the need for the piperazine ring to maintain the activity of nucleozin analogs. Interestingly, this result agrees with the prediction of anti-influenza activity made with a QSAR model presented in this work. The proposed model and the synthetic route will be useful for the further development of nucleozin analogs with antiviral activity.

OPEN ACCESS

Citation: Correa-Padilla E, Hernández-Cano A, Cuevas G, Acevedo-Betancur Y, Esquivel-Guadarrama F, Martínez-Mayorga K (2023) Modifications in the piperazine ring of nucleozin affect anti-influenza activity. PLoS ONE 18(2): e0277073. <https://doi.org/10.1371/journal.pone.0277073>

Editor: Matloob Ahmad, Government College University Faisalabad, PAKISTAN

Received: October 18, 2022

Accepted: January 12, 2023

Published: February 10, 2023

Copyright: © 2023 Correa-Padilla et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Influenza is a highly contagious viral infection caused by a single-stranded negative-sense RNA virus of the *Orthomyxoviridae* family. This infection affects the respiratory system of the host, affecting the nasal pharyngeal mucosa, bronchi, and pulmonary alveoli. The symptoms of influenza are similar to the common cold; however, influenza can be deadly, especially in vulnerable groups. The mutation rate of the influenza virus; the high frequency of genetic rearrangement; and the antigenic changes in viral glycoproteins challenge the control of infections, causing even zoonotic interactions, such as avian influenza (H7 and H9) and swine influenza (Cal/09), with high potential for a pandemic threat. Pharmacological therapy is available. Rimantadine is used to treat or prevent the infection of seasonal influenza (Influenza B), and oseltamivir, peramivir, zanamivir are preferred for influenza (influenza A) that has reached epidemic and pandemic levels [1]. Furthermore, emergent strains are potentially life-

threatening illnesses. Therefore, fostering the discovery and development of new therapeutic antiviral agents is paramount. Efforts in that direction led to the discovery of nucleozin (nlz), a potent inhibitor of influenza A virus infections in *in vitro* and *in vivo* assays [2,3]. Nlz, a piperazine amide, and the corresponding analogs induce the aggregation of nucleoprotein (NP), a protein that plays an essential role on the virus replication cycle. Three-dimensional structures, obtained by X-ray crystallography, of nlz and the analog named Gerritz 3 bound to NP, have been reported in the literature, PDB ID: 3RO5 [4] and PDB ID: 5B7B [5], respectively. In the case of Gerritz 3, the three-dimensional structure shows a dimeric complex formed by two NP monomers bridged symmetrically by two molecules of compound Gerritz 3 [2(Gerritz 3):2NP]. In turn, the complex with nlz is belonging formed by six NP monomers and two nlz molecules (2nlz:6NP). These complexes precipitate in the nucleus of the host cell. They do not migrate to the cytoplasm, where it is necessary to form ribonucleoprotein (RNP) and the subsequent assembly of viral structures.

Influenza inhibitors targeting viral RNP constituent proteins have been recently reviewed [6,7]. Among the most promising targets is described NP, being **nlz** the representative inhibitor of this system. Recent strategies involve computational methodologies, such as virtual screening, similarity searches, and pharmacophore modeling.

Notably, **nlz** is an influenza inhibitor ($IC_{50} = 0.06 \mu\text{M}$) [2], more potent than oseltamivir ($IC_{50} = 1\text{--}10 \mu\text{M}$), an approved drug for the treatment of influenza infections [8]. In addition, the median toxic concentration (TC_{50}) of **nlz** and analogs is greater than $250 \mu\text{M}$. Thus, **nlz** and analogs are good drug candidates with a wide therapeutic window. To contribute to this area, here we report a novel synthesis of three **nlz** analogs, and the *in vitro* evaluation of two of those molecules. In addition, considering these molecules and **nlz** analogs reported in the literature, we present the development of predictive models of activity.

Materials and methods

QSAR models

The workflow to develop predictive models of **nlz** analogs with antiviral activity is shown in Fig 1 and consists of the following steps. Associated script can be found in the Supporting Information.

Data compilation

The structures of **nlz** analogs were collected from the literature using the keyword “*nucleozin*” in SciFinder. The noncurated database consisted of 105 molecules; after the removal of datapoints without activity values, a final set contained 74 molecules. The biological activity reported for all these molecules was evaluated in the Plaque Reduction Assay (PRA) with the A/H1N1/WSN/33 strain [2,4,9]. The antiviral IC_{50} values were transformed to molar units and then to $-\log(IC_{50})$; this data is presented in supporting information in S1 Table in S1 File. ChemBioDraw Ultra 13.0 [10] was used to build the structures. Protonation states were assigned at $\text{pH} = 7.0$, and the structures were energy minimized with FFMM94x forcefield in MOE 2022.10 [11,12].

Molecular descriptors. 4185 2D and 3D molecular descriptors were calculated with the software DRAGON (version 7) [13]. Eight additional descriptors (SLogP, Log S, lip_acc, lip_don, TPSA, Weight, b_rotN, b_rotR) were calculated with MOE 2019.10. and analyzed to assess ADME properties.

Feature selection and model training. The dataset (.csv file format) containing the molecular descriptors and activity values was imported into a notebook in Google Colaboratory (Colab) [14,15]. The modules NumPy [16], Matplotlib [17], Pandas [18,19] and Sci-kit Learn [20] were used for data handling, analysis, visualization, and for the generation of the supervised

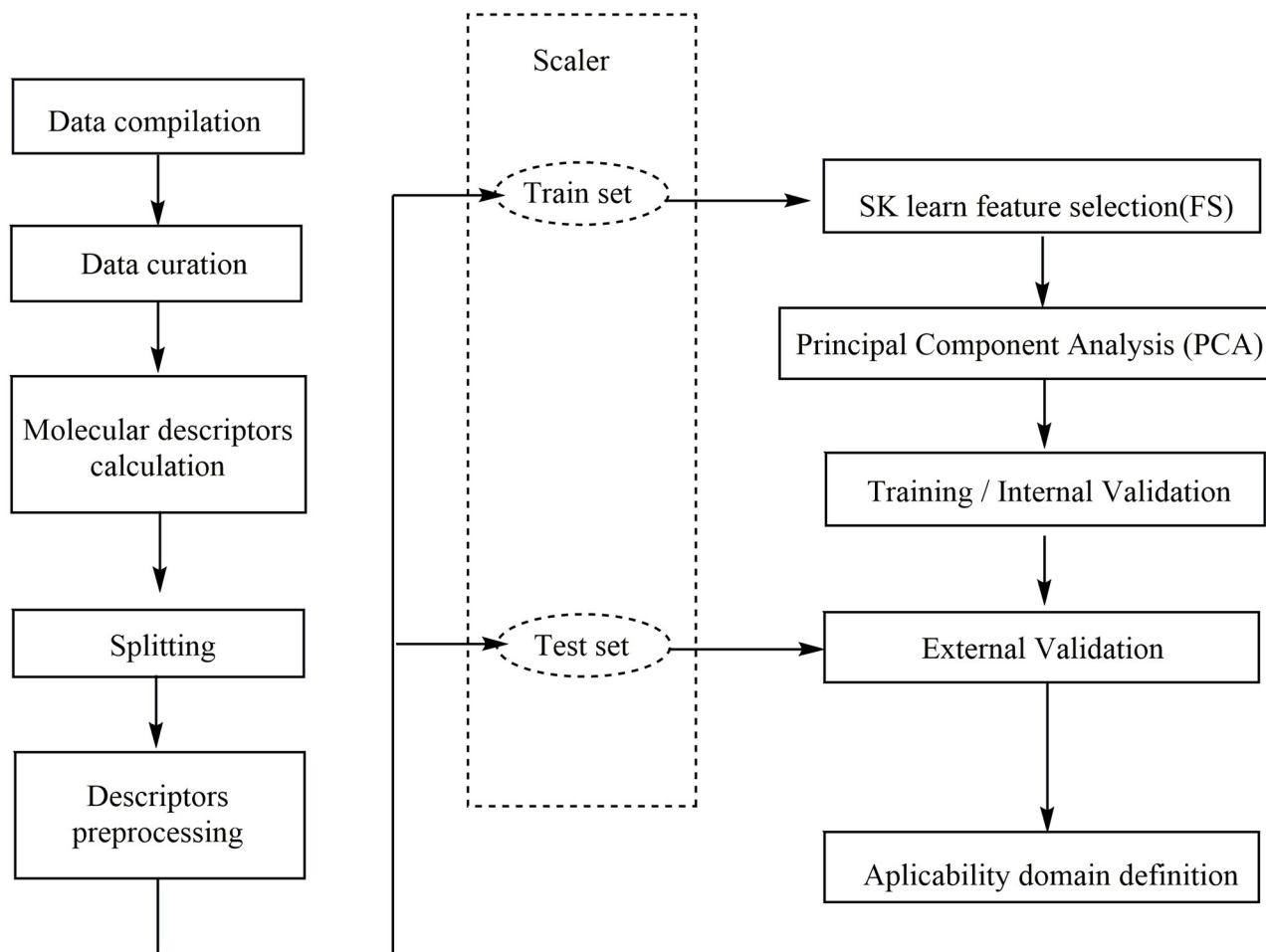


Fig 1. Workflow of the QSAR model.

<https://doi.org/10.1371/journal.pone.0277073.g001>

machine learning regression models. First, the data was split into training and test sets using a stratified split. The training and test sets contain 55 and 19 molecules, respectively. The descriptors with variance equal to zero, were removed with the algorithm VarianceThreshold, then, the descriptors were scaled with the STDScaler. Then, 70 descriptors relevant to the activity prediction were filtered using the Feature Selection (FS) algorithm from scikit-learn, using the meta transformer SelectFromModel, and the RandomForest Regressor as a base estimator. The feature selection step was next, using Principal Component Analysis (PCA) with 16 Principal Components and SVR (Support Vector Regressor) as the algorithm of regression for the training.

Model validation. Goodness-of-fit was measured with the coefficient of determination r^2 . Cross-Validation leave-many-out with 10 folds (CV-LMO, folds = 10), $q^2_{CV-LMNO}$, and Y-scrambling were performed to evaluate the robustness of the model. Lastly, the external validation, q^2_{ext} , allowed us to assess the predictive power.

Synthesis of nucleozin and analogs

The chemical structures and numbering of the **nlz** analogs studied in this work are shown in Figs 2–5 and described in the Results section and the details of the synthesis can be found in the Supporting information section.

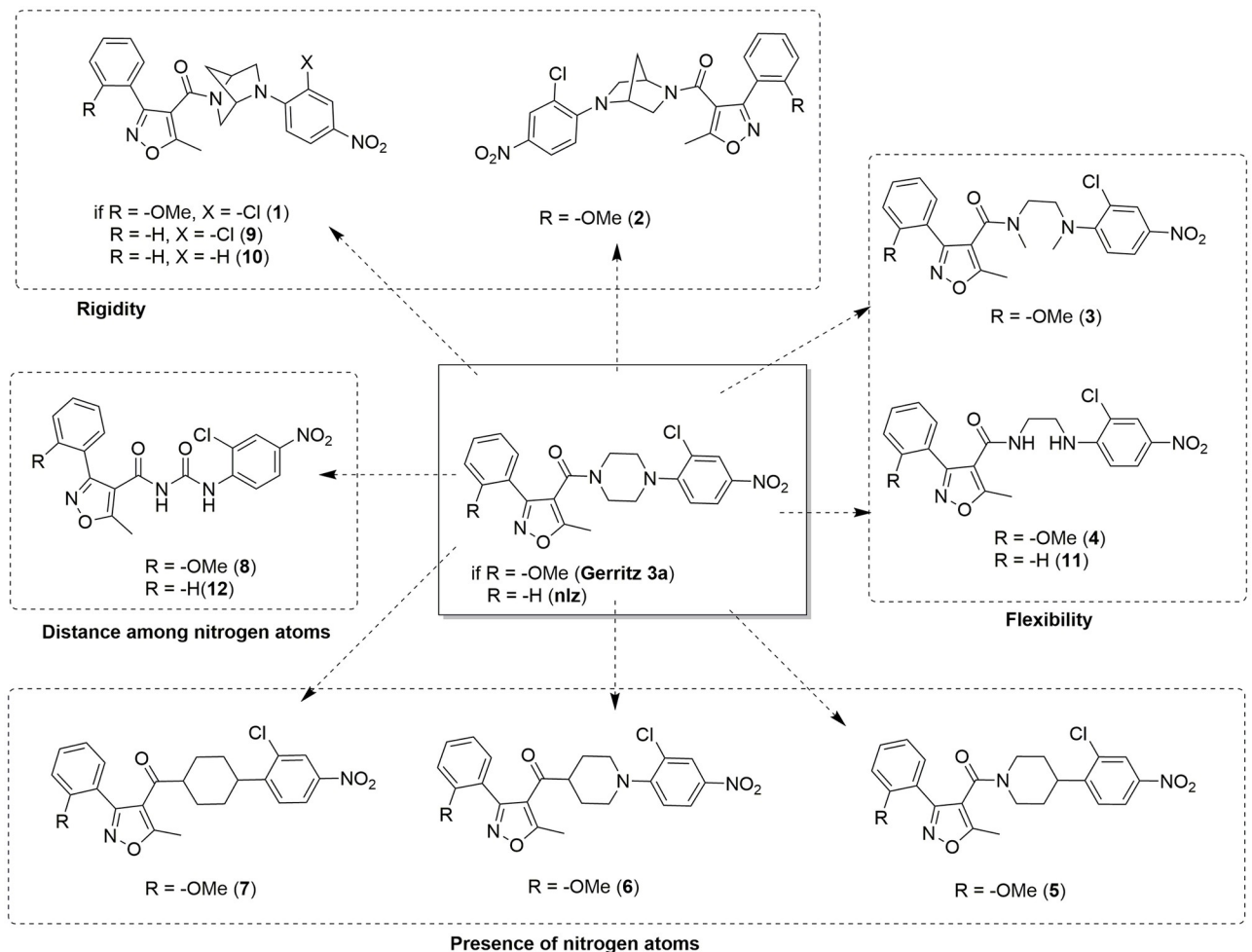


Fig 2. Analogs proposed in this work to study the influence of the piperazine moiety over the antiviral activity within the nucleozin derivatives.

<https://doi.org/10.1371/journal.pone.0277073.g002>

Biological evaluation

A modified antigen-reduction assay was used to determine the activity *nlz* and their analogues [21]. Madin-Darby canine kidney (MDCK) cells were seeded at 104 cells per well, in 96-well plates, in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 4 mM -L-glutamine, 1 mM sodium pyruvate, 50 U/mL penicillin, 50 µg/mL Streptomycin and 10% fetal calf serum (FCS), and incubated for 24 hr at 37 °C. The medium was retired, and cells infected for 1 hr at 37 °C with the influenza A virus strain A/Caledonia H1N1, at a concentration of 0.001 plaque forming units (PFU) per cell. The inoculum was retired, and different concentrations of *nlz* or their analogues (from 1 µM to 10 µM), diluted in DMEM without serum, added to the cells, and incubated. After 48 hrs, the supernatants were removed, frozen at -20 °C and then titrated in a new 96-well plate with MDCK and incubated for 1 hr at 37 °C. The inoculum was retired, DMEM without serum added to the cells and incubated for 72 hrs at 37 °C. The medium was retired, the cells fixed with 4% paraformaldehyde and permeabilized with methanol-acetone 1:1. Then, an ELISA on the fixed-permeabilized cells was performed, using a mouse monoclonal antibody specific for the internal viral protein M1 (ATCC; HB-64), followed by a secondary polyclonal antibody goat anti-mouse Ig's conjugated with horseradish

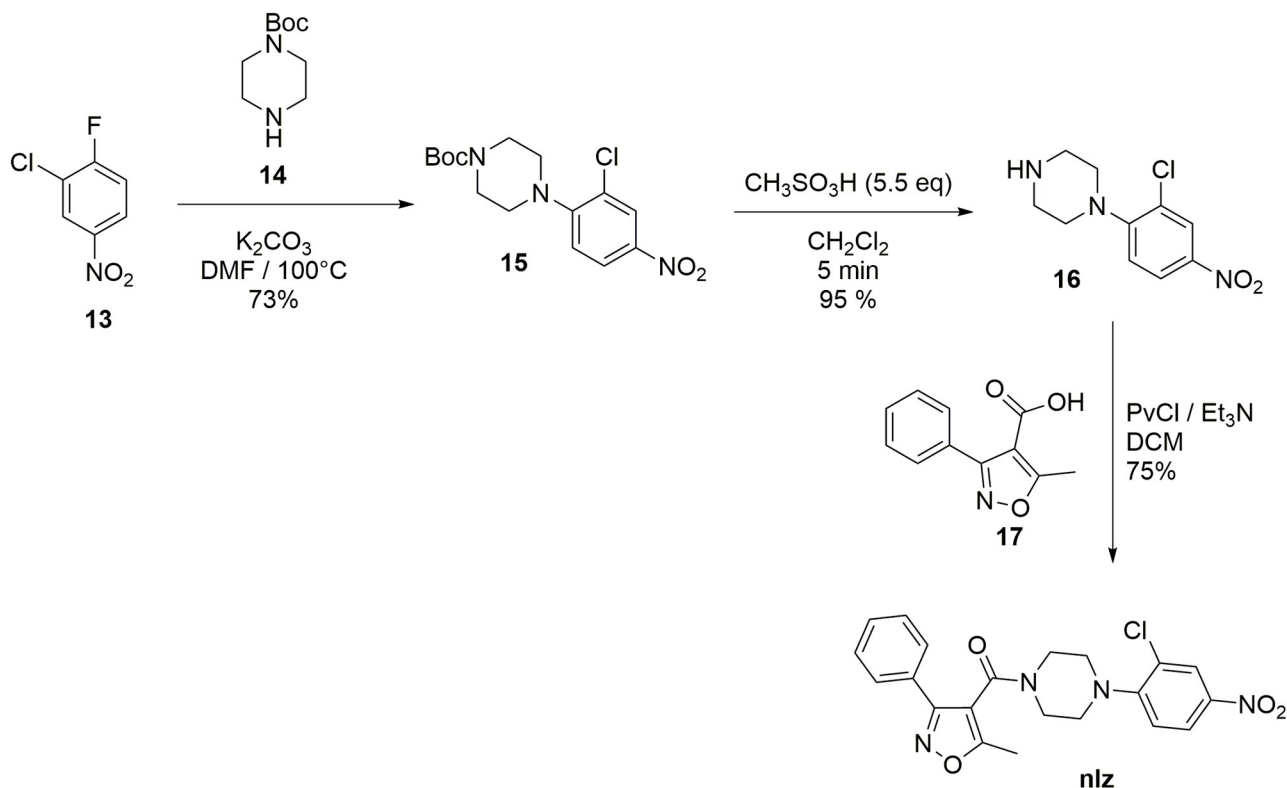


Fig 3. Synthesis of nucleozin.

<https://doi.org/10.1371/journal.pone.0277073.g003>

peroxidase (HRP) (Jackson). The reaction was developed with the substrate *o*-phenylenediamine dihydrochloride (Sigma) and read in an ELISA reader (Biotech) at 490 nm. The intensity of the signal correlates with the amount M1 produced during the infection. In all the steps, DMEM without serum contained 1 μ g of L-1-tosylamide-2-phenylethyl chloromethyl ketone-treated trypsin (TPCK; Sigma) for activation of the virus infectivity. All tissue culture reagents were from GIBCO.

Results and discussion

Nlz analogs, shown in Fig 2, have structural variations in the piperazine ring. Compounds 1 and 2 are analogs with the introduction of the 2,5-diazabicyclo[2.2.1]heptane (DBH) system, where compound 1 corresponds to the (1*S*,4*S*) diastereomer and compound 2 corresponds to the (1*R*,4*R*) diastereomer. This system has been introduced in a variety of compounds [22] with antiparasitic [23], antibiotic [24], and anticarcinogenic activity [25]. It has been suggested that compounds with the DBH system have a better binding ability, compared to the piperazine analogs, due to the rigidity of the bicycle ring. The diamine system of 2,5-diazabicyclo[2.2.1]heptane is traditionally included in screening libraries as a rigid counterpart of the flexible piperazine ring [26–30]. Compounds 3 and 4 exemplify flexible ligands by substituting the piperazine ring for an ethylenediamine functional group. To assess the role of the nitrogen atoms in the piperazine, we analyzed compounds 5, 6, 7, and 8. These molecules allowed us to investigate the relevance of the distance among the nitrogen atoms of the piperazine by diminishing the number of carbon atoms between them.

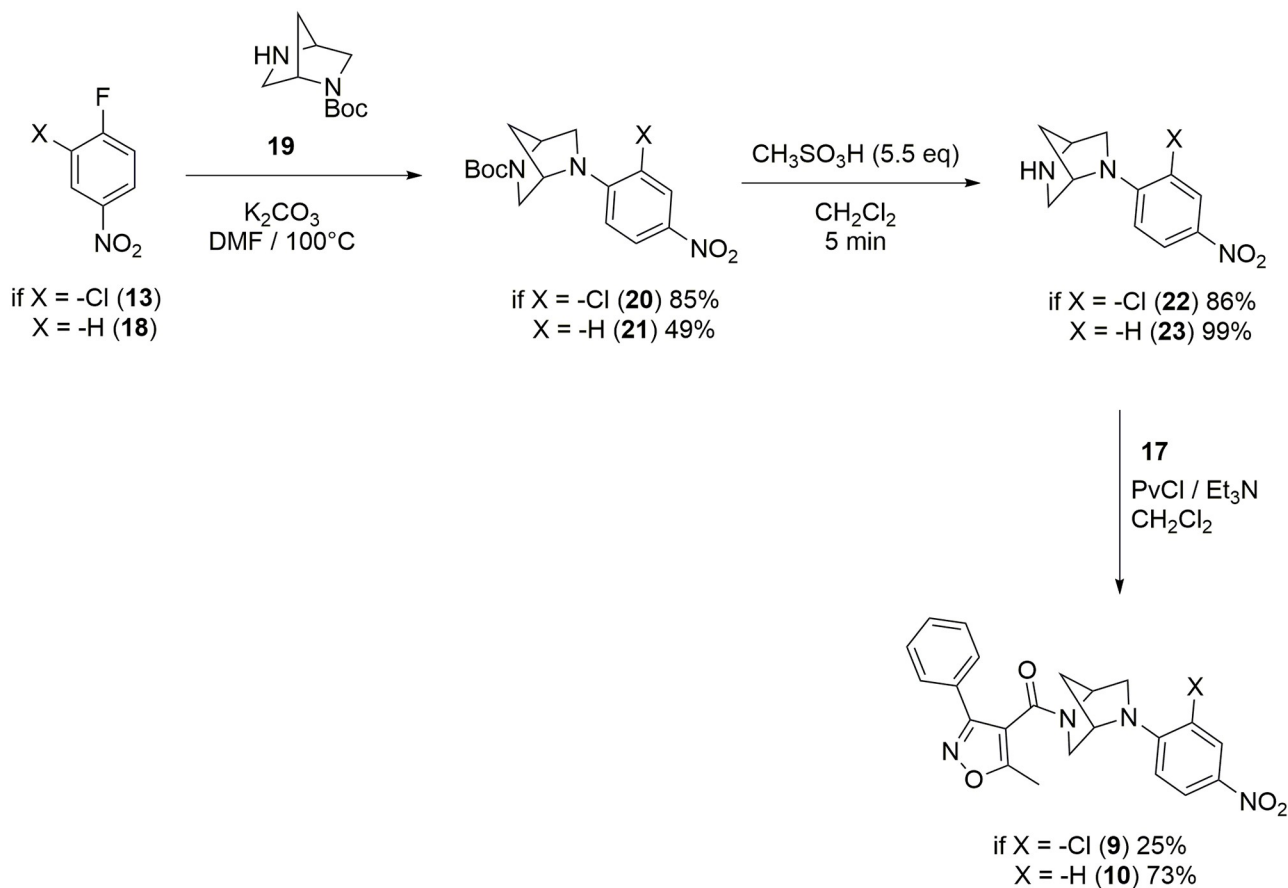


Fig 4. Synthesis of compounds 9 and 10.

<https://doi.org/10.1371/journal.pone.0277073.g004>

Molecular descriptors

Molecular descriptors can be informative of ADME profiles. Here we calculated descriptors with relevance in ADMET properties, with the software MOE 2022.10 [11] (see [Supporting information](#)). S1 Fig in [S1 File](#) shows a comparison of SlogP and logS values. The SlogP corresponds to the octanol/water partition coefficient, calculated for ~7000 structures [31], and is a measure of lipophilicity. logS is the log of the aqueous solubility (mol/L) calculated from a linear model. It is known that lipophilicity greatly impacts ADME properties due to its effect on solubility, plasma protein binding (PPB), metabolic clearance, the volume of distribution, and enzyme/receptor binding, among other pharmacological properties. Thus, as lipophilicity increases, solubility decreases, and the solubility increases as the logS increases (more positive).

S1 Fig in [S1 File](#), shows the SlogP and logS for **nlz** and other commercial antivirals used for the treatment of the influenza infections, such as oseltamivir, zanamivir, amantadine, and rimantadine. Broad spectrum antivirals, such as arbidol and acyclovir, foscarnet, peramivir, remdesivir, nirmaltrevir or ritonavir, are also shown. To note, naproxen, acetaminophen, and chlorpheniramine are other drugs used in combination with antivirals to treat influenza infections.

The calculated values of SlogP for **nlz** and **Gerritz 3** are 4.17 and 4.18, respectively. These are higher values compared to those of commercial anti-flu drugs, amantadine, rimantadine,

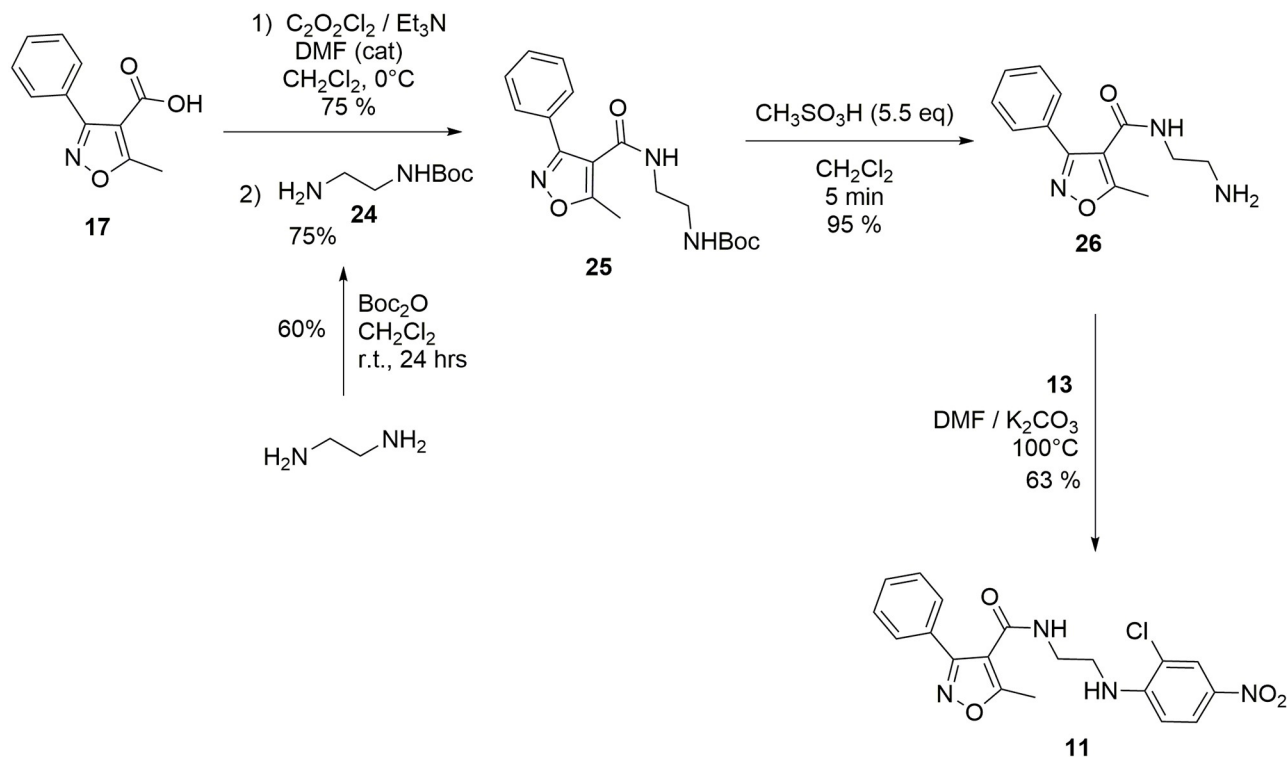


Fig 5. Synthesis of compound 11.

<https://doi.org/10.1371/journal.pone.0277073.g005>

and oseltamivir; the difference between **nlz** and zanamivir is even more significant. The value of **nlz** is also greater than that of naproxen, acetaminophen, and chlorpheniramine. Not surprisingly, **nlz** and **Gerritz 3** have low aqueous solubility compared with rimantadine, amantadine, oseltamivir, and zanamivir. **Nlz** is also less soluble than naproxen, acetaminophen, and chlorpheniramine. Thus, although the high lipophilicity is the molecular characteristic that allowed the ligand to interact with the recognition site, this can also be undesirable in terms of pharmacokinetics and ADME properties due to low solubility. Nonetheless, arbidol and acyclovir, broad-spectrum antivirals, have higher lipophilicity than **nlz** and **Gerritz 3** and low solubility values.

S2 Fig in [S1 File](#) compares SlogP and logS for relevant modifications in **nlz** structure. S3 and S4 Figs in [S1 File](#) show the chemical structures for the analogs mentioned above. Modifications in isoxazole-4-carboxamide moiety have been reported in the literature. Derivatives of 1*H*-1,2,3-triazole-4-carboxamide are more soluble than **nlz**, even more than the 1*H*-pyrazol-4-carboxamides and phenyl-carboxamides derivatives. In an analog reported by Liao *et al.* [32], the absence of the isoxazol ring makes it less lipophilic and more aqueous soluble. In turn, changing the nitrogen atom of the aniline group for a methine group decreases the solubility and increases the lipophilicity [9].

S5 Fig in [S1 File](#) shows that compounds **1** and **2** are less soluble ($\log S = -6.57$) than **nlz** ($\log S = -6.18$) due to the introduction of the methylene bridge, compounds **3** and **4** were slightly more soluble ($\log S = -6.08$ and -6.05 , respectively). When the nitrogen atoms are removed in the piperazine moiety the value of $\log S$ diminished ($\log S = -6.61$ for ligand **5** and -6.65 for compound **6**), the enolic tautomeric form of the compound **6** is more soluble than **nlz** ($\log S = -6.11$) and for the case of the compound **7** are even less soluble with $\log S = -7.66$ for the keto

form and -7.12 for the enol form. Even with NH groups, compound **8** is less soluble ($\log S = -6.51$) than **nlz**. Thus, the nitrogen atoms have an essential contribution to solubility.

QSAR models

The distribution of IC_{50} antiviral activity values of our curated dataset is shown in Fig 6. Activity values range from 4.2 to 7.4. The activity values of training and test sets lie within the same range.

In the preprocessing stage, features with low variance (zero or near to zero) were removed. Then, we transformed all the remaining features to have equal variance. By scaling the features, one gives an equal chance to all the features to influence the model. Thus, the standardized data can be informative. We employed StandardScaler as scaling algorithm. As a next step, we selected the most relevant features by assessing their importance using the meta-transformer SelectFromModel, alongside the estimators listed in S2 Table in S1 File. To limit the number of features, we first explored the best base estimator with the meta-transformer SelectFromModel; then, we fine-tuned this base estimator with the hyperparameters listed in S3 Table in S1 File. Lastly, the number of features was evaluated from 20 to 600 (S4 Table in S1 File), and the best models were obtained with 60 to 70 features.

The models were further refined with hyperparameter tuning [20]. The number of trees in the random forest varied from 50 to 5000. Above 600 trees, the statistics of the models were consistently improved (S3 Table in S1 File).

Max_depth, is the maximum depth of the tree. The default value is None, and then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples (the hyperparameter min_samples_split is the minimum number of samples required to split an internal node, the default value is 2). Max_depth was used in an interval between 1 and 150 (S3 Table in S1 File). Here, the best statistics were reached at max_depth = 50.

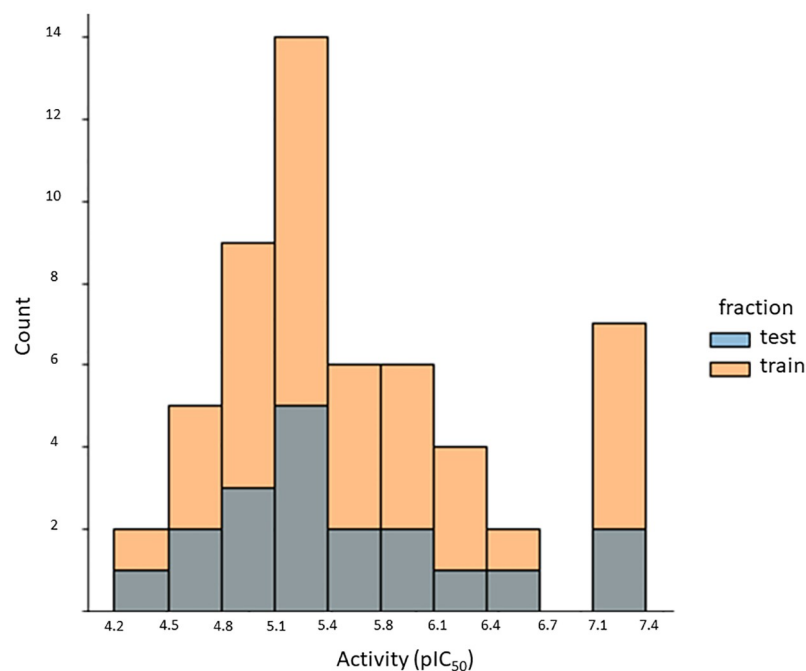


Fig 6. Distribution of pIC₅₀ values for nucleozin derivatives. Molecules in the training set are shown in orange, and molecules in the test set are shown in gray. The inhibitory potency of the test set falls within the interval of pIC₅₀ values of the training set.

<https://doi.org/10.1371/journal.pone.0277073.g006>

Table 1. Search for the best regressor algorithm for training.

Model	Regressor	r^2
1	LinearRegression	0.7706
2	Ridge	0.7823
3	Lasso	0.1625
4	ElasticNet	0.7556
5	BayesianRidge	0.7685
6	SGDRegressor	0.7542
7	KNearestNeighborRegressor	0.5791
8	SVR	0.9141
9	NuSVR	0.9103
10	LinearSVR	0.7275
11	DecisionTreeRegressor	1.0
12	RandomForestRegressor	0.9112
13	MLPRegressor	0.9405

<https://doi.org/10.1371/journal.pone.0277073.t001>

The hyperparameter `Min_samples_leaf` is the minimum number of samples required at a leaf node. A split point at any depth will only be considered if it leaves at least `min_samples_leaf` training samples in each of the left and right branches. The default value is 1. In the model, `Min_samples_leaf` was tested in an interval between 2 and 20 (S3 Table in [S1 File](#)). Here, the best statistics were reached at `Min_samples_leaf = 8`.

The hyperparameter `min_impurity_decrease` refers to a float value, and the default value is equal to 0.0. A node will be split if this split induces a decrease of the impurity greater than or equal to this value. It took values between $1.0E^{-15}$ and 2.0 (S3 Table in [S1 File](#)). Here, the best statistics were reached at $1X10^{-6}$.

Lastly, `warm_start` is a hyperparameter with the Boolean value. The default is False; when set to True, reuse the solution of the previous call to fit and add more estimators to the ensemble; otherwise, just fit a whole new forest. Here, the best statistics were reached with True. All the other parameters were left as default.

The selected 60 or 70 features were further selected using PCA, reasonably good models contained 15 or more PC (S5 Table in [S1 File](#)). After the preprocessing, models were generated with several regression algorithms, listed in [Table 1](#), exploring from linear regression to Random Forest. Models with the best statistics were SVR and the variant NuSVR. Further refinement was performed using hyperparameter tuning.

Using SVR, the following step consisted of hyperparameter tuning ([Table 2](#)). The hyperparameters analyzed were `C` (the regularization parameter) and `max_iter`. The first one, `C` is a parameter that determines the strength of the regularization; it must be positive, and higher values of `C` correspond to less regularization. In these cases, the model tries to fit each data point of the training set as best as possible. At the same time, with low values of the parameter `C`, the algorithms put more emphasis on adjusting to the majority of data points [33]. The default value is 1.0. Here, the best statistics were reached at 2.0 (model 16 in [Table 2](#)). The hyperparameter `max_iter` is a hard limit on iterations within the solver; it is a positive integer value or -1 for no limit (default value). For the best model, `max_iter = 100 000`.

In summary, the best model was developed with SVR, with the hyperparameters `C = 2.0` and `max_iter = 100 000`. The statistics obtained for that model are summarized in [Table 3](#), and the code of Colab is available in the supporting information.

[Fig 7a](#) shows the distribution plot of the values predicted with the SVR model and the experimental values, the train test is shown in blue, and the test set is shown in orange and is

Table 2. Hyperparameter tuning for SVR regressor.

Model	Hyperparameter	r^2
	C	
13	0.1	0.2713
14	0.5	0.7681
15	1.0 (default)	0.9039
16	2.0	0.9270
17	3.0	0.9259
18	4.0	0.9393
19	5.0	0.9458
20	10	0.9594
21	20	0.9698
22	50	0.9878
	<i>max_iter</i>	
31	-1	0.908255
32	100	0.891469
33	1000	0.894005
34	10000	0.883529
35	50000	0.89358
36	100000	0.90203

<https://doi.org/10.1371/journal.pone.0277073.t002>

defined by a linear correlation. Fig 7b shows the results of the Y-scrambling test with 150 iterations (orange dots). As expected, the q^2 obtained for the models generated after aleatorization decreases. They show that the model's performance is not due to chance. Fig 7c shows the applicability domain using the Williams plot. Under that description, the model can define most of the molecules. The statistics of the best model are $r^2 = 0.94$, $q^2 = 0.60$, and $q^2_{\text{ext}} = 0.60$. The activity predicted for analogs 1–12 is shown in Table 4. Interestingly, Geritz 3 and nlz are predicted more actives than the nlz analogs, in agreement with the preliminary experimental data presented here.

Synthesis

The reported synthetic procedure for nlz using 1,2-dichlorobenzene [2] was not reproducible in our lab. Thus, we designed a new method for the synthesis of nlz. The synthetic route is shown in Fig 3. Compound 13 was mixed with the secondary amine 14 in DMF at 120 °C, in the presence of K_2CO_3 , to give compound 15 via a nucleophilic aromatic substitution reaction. The subsequent deprotection of the *N*-Boc group with 5.5 equivalents of methane sulfonic acid in dichloromethane yielded compound 16 quantitatively in five minutes. After this deprotection, a solution of amine 16 was added to a solution of a mixed anhydride, generated *in situ* from the carboxylic acid 17. It was treated with pivaloyl chloride in the presence of triethylamine, as a base, in dry dichloromethane as the solvent; the consequent addition-elimination reaction gave the nlz compound.

Table 3. Results for the support vector regressor.

	r^2	$q^2_{\text{CV-LMO}}$ (Folds = 10)	q^2_{ext}
	0.94	0.60	0.60
MSE	0.03	0.25	0.21

<https://doi.org/10.1371/journal.pone.0277073.t003>

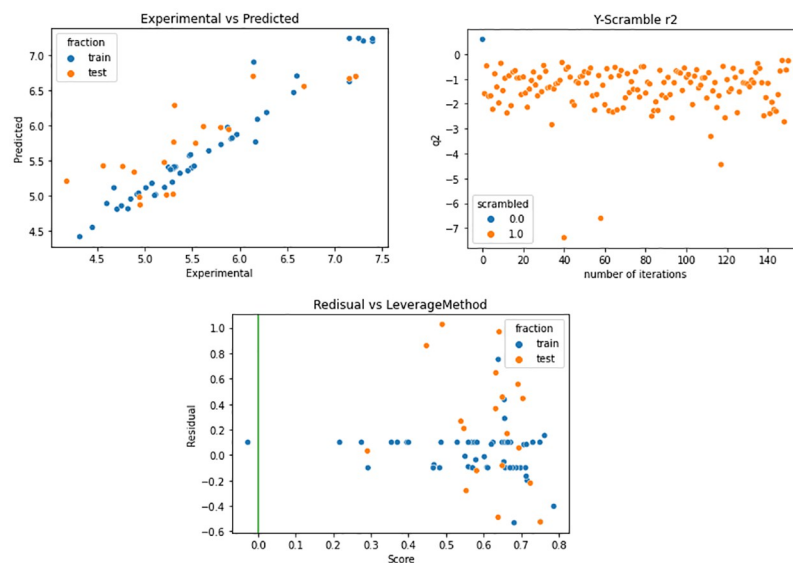


Fig 7. Results of QSAR model. **a)** Regression plot of the SVR model. The plot shows the predicted values by SVR model against the experimental for the train set (in blue) and test set (in orange). **b)** Y-Scramble test for the predicted model: In the plot is shown, in orange, the results of the q^2_{CV-LMO} values for randomized models and in blue the q^2_{CV-LMO} values for the true model (no randomized). **c)** Williams plot for the calculated model: The Applicability Domain was calculated by the leverage method.

<https://doi.org/10.1371/journal.pone.0277073.g007>

The structures of **nlz** and analogues were characterized by $^1\text{H-NMR}$, $^{13}\text{C-NMR}$, DEPT-135, and the corresponding molecular weights were confirmed by DART mass spectrometry. The spectra and assignment of the signals are described in detail as Supporting Information.

Table 4. Predicted activity for nucleozin analogues.

Compound	Experimental	Predicted
Nlz	7.22	6.69
Gerritz 3	7.39	7.19
compound 1		5.87
compound 2		5.86
compound 3		6.03
compound 4		5.50
compound 5		6.60
compound 6		6.59
compound 6 enol		6.59
compound 7 anti		6.04
compound 7 enol		6.25
compound 7 syn		5.86
compound 8		5.34
compound 9		5.54
compound 10		5.29
compound 11		4.90
compound 12		5.06

<https://doi.org/10.1371/journal.pone.0277073.t004>

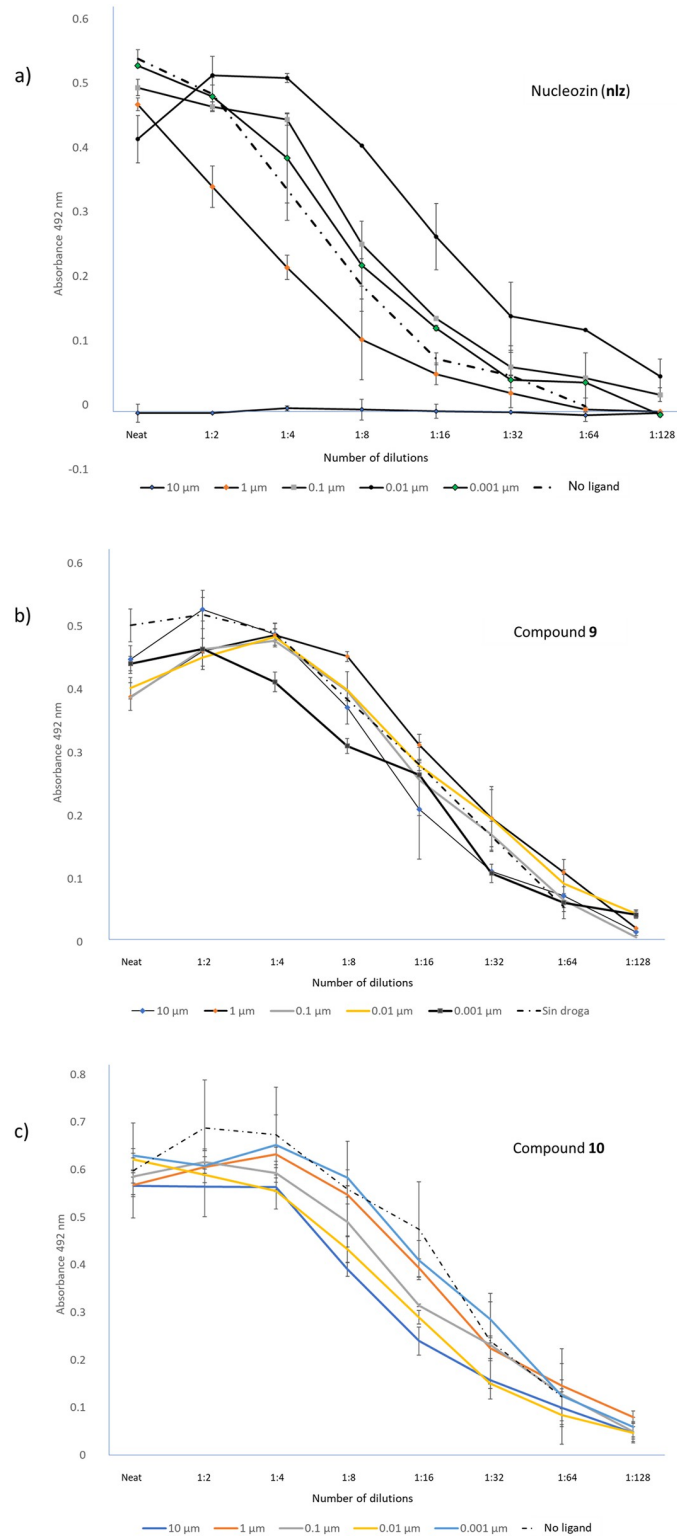


Fig 8. Biological evaluation of nlz and analogs in cell cultures of influenza virus-infected MDCK cells. The experiments were performed by infecting MDCK cells with the Influenza virus A/Caledonia H1N1, at a concentration of 0.001 pfu per cell in presence of various concentrations of nlz or analogs. After 48 hrs of incubation, the supernatants were obtained and titrated in MDCK cells and incubated for 72 hrs. Cells were fixed-permeabilized and an ELISA on the cells performed, using an anti- influenza M1 monoclonal antibody, followed by a goat anti-mouse Ig's

polyclonal antibody conjugated to HRP. After adding the substrate, the color intensity was read at 492 nm in an automated ELISA plate reader. **a)** Evaluation of **nlz**. **b)** biological evaluation of the ligand **9**. **c)** biological evaluation of the ligand **10**.

<https://doi.org/10.1371/journal.pone.0277073.g008>

The same synthetic strategy was employed for the synthesis of compounds **9** and **10** (Fig 4); for these compounds was used the amine (1*S*,4*S*)-*N*-Boc-2,5-diazabicyclo[2.2.1]heptane, *N*-BocDBH, (**19**), reported previously, instead of the amine **14**. Compound **9** was prepared from 2-chloro-4-nitrofluorobenzene (**13**), and compound **10** from the compound 4-nitrofluorobenzene (**18**).

A similar methodology was also used to obtain compound **11**. The ethylenediamine was mono protected using the Boc₂O reagent to obtain compound **24** which was mixed with an acyl chloride generated *in situ* from the carboxylic acid **17** treated with oxalyl chloride in dry dichloromethane in the presence of triethylamine as a base, and drops of dimethylformamide, rendering compound **25**. Then, for the deprotection of the *N*-Boc group, the compound **25** was treated with methane sulfonic acid to form the primary amine **26**, followed by a nucleophilic aromatic substitution, using **13** as an electrophile for obtaining the compound **11** (Fig 5).

Thus, here we employed a new route using 4-nitrofluoroarenes as feedstock, which efficiently led to **nlz** and some derivatives. The aromatic nucleophilic substitution of the fluor atom was performed in dimethylformamide (DMF) as solvent at 120 °C. Some observations in the laboratory indicated that it was not necessary to add DMF to the reaction mixture for the synthesis of compound **20**. For this reason, we carried out the reaction grinding in a mortar with a pestle, for 15 minutes, the monoprotected diamine **19** and the compound **13**, both in solid state, in the presence of five equivalents of K₂CO₃ as a base, without DMF. The reaction was monitored by thin-layer chromatography (TLC), comparing the reaction mixture against the previously characterized compound **20**, showing that most of the raw materials were transformed. Compound **20** was obtained by the addition of 3 mL of hot ethyl acetate (AcOEt), filtered through a small layer of Celite, recrystallized, and isolated from the mother liquor. By thin layer chromatography, a product with a good degree of purity was observed, and the melting point coincided with that reported m.p. = 199.7 °C. Thus, the substitution reaction can be carried out under free-solvent conditions. Future experiments will be carried out to explore this methodology as a potential synthesis route to new **nlz** derivatives.

Biological evaluation. The virus load starts to decline in the presence of 1 μM of **nlz** and is totally inhibited at 10 μM. Under the same conditions but in the presence of analogs **9** and **10**, viral growth was not inhibited, as is shown in Fig 8. It is possible that analogs **9** and **10** will have viral inhibitory effects at larger concentrations. These results show that constraining the core structure of **nlz** (analogs **9** and **10**) results in the loss of activity. Interestingly, the QSAR model developed here predicts the analogs **9** and **10** as less active, consistent with the experimental observation.

Additional molecular modeling studies and experimental evaluation of other analogs will provide information on structural modifications required for the improvement of antiviral activity. QSAR models developed here will aid the synthetic efforts of new **nlz** analogs.

Supporting information

S1 File. Inhibitory activity data of the influenza A/WSN/33 (H1N1) virus, computational details, and synthetic details.

(PDF)

Acknowledgments

Eduardo Ruiz Padilla, Manuel López Ortiz, Ignacio Regla, and Abraham Madariaga-Mazon are gratefully acknowledged for helpful discussions. Erick Correa thanks Conacyt for Ph.D. Scholarship. We appreciate the technical assistance from I. Chávez Uribe, E. Huerta Salazar, M.A. Peña Gonzalez, H. Rios, Ma. C. García-González, J. Pérez Flores (NMR and MS spectroscopy).

Author Contributions

Conceptualization: Gabriel Cuevas, Fernando Esquivel-Guadarrama, Karina Martinez-Mayorga.

Data curation: Erick Correa-Padilla.

Funding acquisition: Karina Martinez-Mayorga.

Investigation: Erick Correa-Padilla, Alejandro Hernández-Cano, Gabriel Cuevas, Yunuen Acevedo-Betancur, Fernando Esquivel-Guadarrama, Karina Martinez-Mayorga.

Project administration: Karina Martinez-Mayorga.

Writing – original draft: Erick Correa-Padilla, Alejandro Hernández-Cano.

Writing – review & editing: Karina Martinez-Mayorga.

References

1. Kumar B, Asha K, Khanna M, Ronsard L, Meseko CA, et al. (2018) The emerging influenza virus threat: status and new prospects for its therapy and control. *Arch Virol* 163: 831–844. <https://doi.org/10.1007/s00705-018-3708-y> PMID: 29322273
2. Kao YRWA, 2011. <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2011005799> (2011) RNA containing modified nucleotides and use thereof in vaccines.
3. Su CY, Cheng TJ, Lin MI, Wang SY, Huang WI, et al. (2010) High-throughput identification of compounds targeting influenza RNA-dependent RNA polymerase activity. *Proc Natl Acad Sci U S A* 107: 19151–19156. <https://doi.org/10.1073/pnas.1013592107> PMID: 20974907
4. Gerritz SW, Cianci C, Kim S, Pearce BC, Deminie C, et al. (2011) Inhibition of influenza virus replication via small molecules that induce the formation of higher-order nucleoprotein oligomers. *Proc Natl Acad Sci U S A* 108: 15366–15371. <https://doi.org/10.1073/pnas.1107906108> PMID: 21896751
5. Pang B, Cheung NN, Zhang W, Dai J, Kao RY, et al. (2016) Structural Characterization of H1N1 Nucleoprotein-Nucleozin Binding Sites. *Sci Rep* 6: 29684. <https://doi.org/10.1038/srep29684> PMID: 27404920
6. Hou L, Zhang Y, Ju H, Cherukupalli S, Jia R, et al. (2022) Contemporary medicinal chemistry strategies for the discovery and optimization of influenza inhibitors targeting vRNP constituent proteins. *Acta Pharm Sin B* 12: 1805–1824. <https://doi.org/10.1016/j.apsb.2021.11.018> PMID: 35847499
7. Nawar N, A review paper on "Developing novel antiviral strategies against pathogenic RNA viruses" 2021, Brac University.
8. Kao RY, Yang D, Lau LS, Tsui WH, Hu L, et al. (2010) Identification of influenza A nucleoprotein as an antiviral target. *Nat Biotechnol* 28: 600–605. <https://doi.org/10.1038/nbt.1638> PMID: 20512121
9. Pei S, Xia S, Yang F, Chen J, Wang M, et al. (2020) Design, synthesis and in vitro biological evaluation of isoxazol-4-carboxa piperidyl derivatives as new anti-influenza A agents targeting virus nucleoprotein. *RSC Adv* 10: 4446–4454. <https://doi.org/10.1039/c9ra10828a> PMID: 35495231
10. ChemDraw. CambridgeSoft Corporation: 875 Massachusetts Ave. C, MA 02139, USA.
11. MOE Molecular Operating Environment (MOE), version 2022.08, Chemical Computing Group Inc., Montreal, Quebec, Canada. <http://www.chemcomp.com>. pp. Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, Quebec, Canada.
12. Gill PEM, W.; Wright, M. H. PRACTICAL OPTIMIZATION; Emerald Group Publishing Limited, 1981.
13. Talete, S. R. L. DRAGON for Windows (Software for Molecular Descriptor Calculations). Kode chemoinformatics 2019.

14. Google Colaboratory. https://colab.research.google.com/#scrollTo=GJBs_fiRovLc.
15. Carneiro T, Nóbrega RVMD, Nepomuceno T, Bian GB, Albuquerque VHCD, et al. (2018) Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access* 6: 61677–61685.
16. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, et al. (2020) Array programming with NumPy. *Nature* 585: 357–362. <https://doi.org/10.1038/s41586-020-2649-2> PMID: 32939066
17. Hunter JD (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9: 90–95.
18. McKinney W. Pandas: A Foundational Python Library for Data Analysis and Statistics. In *Python for high performance and scientific computing*; 2011; Vol. 14, pp 1–9.
19. McKinney, W. pandas: a python data analysis library. <http://pandas.sourceforge.net>.
20. Pedregosa FVG.; Gramfort A.; Michel V.; Thirion B.; Grisel O.; Blondel M.; et al. (2012) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825–2830.
21. Gross FL, Bai Y, Jefferson S, Holiday C, Levine MZ (2017) Measuring influenza neutralizing antibody responses to A(H3N2) viruses in human sera by microneutralization assays using MDCK-SIAT1 cells. *J Vis Exp* 129: 56448.
22. Murineddu G, Asproni B, Pinna G, Curzu MM, Dore A, et al. (2012) Synthesis of biologically active bridged diazabicycloheptanes. *Curr Med Chem* 19: 5342–5363. PMID: 22830336
23. Thompson PE, Zeigler JB, McCall W (1974) Antifilarial agents. Diazabicyclooctanes and diazabicycloheptanes as bridged analogs of diethylcarbamazine. *J Med Chem* 17: 481–487. <https://doi.org/10.1021/jm00251a002> PMID: 4830249
24. Remuzon P, Bouzard D, Guiol C, Jacquet JP (1992) Fluoronaphthyridines as antibacterial agents. 6. Synthesis and structure-activity relationships of new chiral 7-(1-, 3-, 4-, and 6-methyl-2,5-diazabicyclo[2.2.1]heptan-2-yl)naphthyridine analogues of 7-[(1R,4R)-2,5-diazabicyclo[2.2.1]heptan-2-yl]-1-(1,1-dimethylethyl)-6-fluoro-1,4-dihydro-4-oxo-1,8-naphthyridine-3-carboxylic acid. Influence of the configuration on blood pressure in dogs. A quinolone-class effect. *J Med Chem* 35: 2898–2909. <https://doi.org/10.1021/jm00093a024> PMID: 1322990
25. Laskar S, Sánchez-Sánchez L, Flores SM, López-Muñoz H, Escobar-Sánchez ML, et al. (2018) Identification of (1S,4S)-2,5-diazabicyclo[2.2.1]heptane-dithiocarbamate-nitrostyrene hybrid as potent antiproliferative and apoptotic inducing agent against cervical cancer cell lines. *Eur J Med Chem* 146: 621–635. <https://doi.org/10.1016/j.ejmech.2018.01.020> PMID: 29407986
26. Britvin SN, Rummyantsev AM (2017) Crystal structure of (1S,4S)-2,5-diazabicyclo[2.2.1]heptane dibromide. *Acta Crystallogr E Crystallogr Commun* 73: 1861–1865. <https://doi.org/10.1107/S2056989017015870> PMID: 29250403
27. Dam JH, Bender D, Peters D, Någren K (2016) [¹¹C]NS9531, [¹¹C]NS9762 and [¹¹C]NS6417, specific SERT tracers: pre-clinical evaluation in pigs and optimization of synthesis conditions using [¹¹C]methyl triflate. *Nucl Med Biol* 43: 42–51.
28. Cernak T, Gesmundo NJ, Dykstra K, Yu Y, Wu Z, et al. (2017) Microscale High-Throughput Experimentation as an Enabling Technology in Drug Discovery: Application in the Discovery of (Piperidinyl)pyridinyl-1H-benzimidazole Diacylglycerol Acyltransferase 1 Inhibitors. *J Med Chem* 60: 3594–3605. <https://doi.org/10.1021/acs.jmedchem.6b01543> PMID: 28252959
29. Llona-Minguez S, Höglund A, Ghassemian A, Desroses M, Calderón-Montaño JM, et al. (2017) Piperazin-1-ylpyridazine Derivatives Are a Novel Class of Human dCTP Pyrophosphatase 1 Inhibitors. *J Med Chem* 60: 4279–4292. <https://doi.org/10.1021/acs.jmedchem.7b00182> PMID: 28508636
30. Wei M, Zhang X, Wang X, Song Z, Ding J, et al. (2017) SAR study of 5-alkynyl substituted quinazolin-4(3H)-ones as phosphoinositide 3-kinase delta (PI3Kδ) inhibitors. *Eur J Med Chem* 125: 1156–1171.
31. Wildman SA, Crippen GM (1999) Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Computer Sciences* 39: 868–873.
32. Liao J C H.; Wan J.; Chen P.; Li Y.; Ding K.; Tortorella M. D.; et al. (2016) Evaluation of Benzamide Derivatives as New Influenza A Nucleoprotein Inhibitors. *Open Journal of Medicinal Chemistry* 6: 43–50.
33. Müller ACG, S. Introduction to Machine Learning with Python; O'Reilly Media, Inc., 2016.

REFERENCIAS

- ¹ a) Hernández Orozco, F. Chile Registra Su Primer Caso de Gripe Aviar En Humanos, Esto Sabemos. *Expansión*. **2023**. <https://expansion.mx/mundo/2023/03/30/primer-caso-gripe-aviar-chile> (acceso 2023-03-30). b) Torres, C. Chile Confirmó El Hallazgo de Más de 70 Lobos Marinos Muertos En Una Isla y No Descartan “Influenza Aviar.” infobae. March 31, 2023. <https://www.infobae.com/america/america-latina/2023/03/31/chile-confirmando-el-hallazgo-de-mas-de-70-lobos-marinos-muertos-en-una-isla-y-no-descartan-influenza-aviar/> (acceso 2023-03-30). c) Montes, R. Un Hombre de 53 Años Con Una Grave Neumonía, El Primer Humano Con Gripe Aviar En Chile. *El País*. Santiago, Chile March 30, **2023**. <https://elpais.com/chile/2023-03-31/un-hombre-de-53-anos-con-una-grave-neumonia-el-primer-humano-con-gripe-aviar-en-chile.html> (acceso 2023-03-30).
- ² China Enciende Alarmas Por Nuevo Caso de Humano Contagiado de Gripe Aviar: Ya Sería El Tercero. *Revista Semana*. **2023**. <https://www.semana.com/mundo/articulo/china-enciende-alarmas-por-nuevo-caso-de-humano-contagiado-de-gripe-aviar-ya-seria-el-tercero/202328/> (acceso 2023-03-30).
- ³ a) Bardsley-Elliott, A.; Noble, S. Oseltamivir. *Drugs* **1999**, *58* (5), 851–860. <https://doi.org/10.2165/00003495-199958050-00007>. b) McClellan, K.; Perry, C. M. Oseltamivir. *Drugs* **2001**, *61* (2), 263–283. <https://doi.org/10.2165/00003495-200161020-00011>.
- ⁴ a) Waghorn, S. L.; Goa, K. L. Zanamivir. *Drugs* **1998**, *55* (5), 721–725. <https://doi.org/10.2165/00003495-199855050-00015>. b) Hayden, F. G.; Osterhaus, A. D. M. E.; Treanor, J. J.; Fleming, D. M.; Aoki, F. Y.; Nicholson, K. G.; Bohnen, A. M.; Hirst, H. M.; Keene, O.; Wightman, K. Efficacy and Safety of the Neuraminidase Inhibitor Zanamivir in the Treatment of Influenzavirus Infections. *New England Journal of Medicine* **1997**, *337* (13), 874–880. <https://doi.org/10.1056/NEJM199709253371302> c) Dunn, C. J.; Goa, K. L. Zanamivir. *Drugs* **1999**, *58* (4), 761–784. <https://doi.org/10.2165/00003495-199958040-00016>.
- ⁵ a) Shetty, A. K.; Peek, L. A. Peramivir for the Treatment of Influenza. *Expert Rev Anti Infect Ther* **2012**, *10* (2), 123–143. <https://doi.org/10.1586/eri.11.174>. b) Scott, L. J. Peramivir: A Review in Uncomplicated Influenza. *Drugs* **2018**, *78* (13), 1363–1370. <https://doi.org/10.1007/s40265-018-0981-8>. c) Shetty, A. K.; Wester, A. Peramivir Injection in the Treatment of Acute Influenza: A Review of the Literature. *Infect Drug Resist* **2016**, *9*, 201–214. <https://doi.org/10.2147/IDR.S86460>.
- ⁶ Ikematsu, H.; Kawai, N. Laninamivir Octanoate: A New Long-Acting Neuraminidase Inhibitor for the Treatment of Influenza. *Expert Rev Anti Infect Ther* **2011**, *9* (10), 851–857. <https://doi.org/10.1586/eri.11.112>.
- ⁷ a) Świerczyńska, M.; Mirowska-Guzel, D. M.; Pindelska, E. Antiviral Drugs in Influenza. *Int J Environ Res Public Health* **2022**, *19* (5), 3018. <https://doi.org/10.3390/ijerph19053018>. b) Shobugawa, Y.; Saito, R.; Dapat, C.; Dapat, I. C.; Kondo, H.; Saito, K.; Sato, I.; Kawashima, T.; Suzuki, Y.; Suzuki, H. Clinical Effectiveness of Neuraminidase Inhibitors—Oseltamivir, Zanamivir, Laninamivir, and Peramivir—for Treatment of Influenza A(H3N2) and A(H1N1)Pdm09 Infection: An Observational Study in the 2010–2011 Influenza Season in Japan. *Journal of Infection and Chemotherapy* **2012**, *18* (6), 858–864. <https://doi.org/10.1007/s10156-012-0428-1>. c) Sarukhanyan, E.; Shanmugam, T. A.; Dandekar, T. In Silico Studies Reveal Peramivir and Zanamivir as an Optimal Drug Treatment Even If H7N9 Avian Type Influenza Virus Acquires Further Resistance. *Molecules* **2022**, *27* (18), 5920. <https://doi.org/10.3390/molecules27185920>.
- ⁸ a) Hayden, F. G.; Asher, J.; Cowling, B. J.; Hurt, A. C.; Ikematsu, H.; Kuhlbusch, K.; Lemuel-Diot, A.; Du, Z.; Meyers, L. A.; Piedra, P. A.; Takazono, T.; Yen, H.-L.; Monto, A. S. Reducing Influenza Virus Transmission: The Potential Value of Antiviral Treatment. *Clinical Infectious Diseases* **2022**, *74* (3), 532–540. <https://doi.org/10.1093/cid/ciab625>. b) Reina, J.; Reina, N. Baloxavir Marboxil: A Potent Cap-Dependent

Endonuclease Inhibitor of Influenza Viruses. *Rev Esp Quimioter* **2019**, 32 (1), 1–5.

- ⁹ Batool, S.; Chokkakula, S.; Song, M.-S. Influenza Treatment: Limitations of Antiviral Therapy and Advantages of Drug Combination Therapy. *Microorganisms* **2023**, 11 (1), 183. <https://doi.org/10.3390/microorganisms11010183>.
- ¹⁰ **a)** Renaud, C.; Kuypers, J.; Englund, J. A. Emerging Oseltamivir Resistance in Seasonal and Pandemic Influenza A/H1N1. *Journal of Clinical Virology* **2011**, 52 (2), 70–78. <https://doi.org/10.1016/j.jcv.2011.05.019>. **b)** Dharan, N. J. Infections With Oseltamivir-Resistant Influenza A(H1N1) Virus in the United States. *JAMA* **2009**, 301 (10), 1034. <https://doi.org/10.1001/jama.2009.294>. **c)** Patel, M. C.; Mishin, V. P.; De La Cruz, J. A.; Chesnokov, A.; Nguyen, H. T.; Wilson, M. M.; Barnes, J.; Kondor, R. J. G.; Wentworth, D. E.; Gubareva, L. V. Detection of Baloxavir Resistant Influenza A Viruses Using next Generation Sequencing and Pyrosequencing Methods. *Antiviral Res* **2020**, 182, 104906. <https://doi.org/10.1016/j.antiviral.2020.104906>. **d)** Harding, J.; Bernasconi, C.; Williams, S.; Wildum, S.; Kinoshita, M.; Uehara, T.; Hurt, A. C. Investigating the Transmission of Baloxavir-resistant Influenza Viruses from Treated Index Patients to Untreated Household Contacts in the BLOCKSTONE Study. *Influenza Other Respir Viruses* **2023**, 17 (1). <https://doi.org/10.1111/irv.13079>. **e)** Hurt, A. C.; Holien, J. K.; Parker, M.; Kelso, A.; Barr, I. G. Zanamivir-Resistant Influenza Viruses with a Novel Neuraminidase Mutation. *J Virol* **2009**, 83 (20), 10366–10373. <https://doi.org/10.1128/JVI.01200-09>.
- ¹¹ **a)** Barik, S. New Treatments for Influenza. *BMC Med* **2012**, 10 (1), 104. <https://doi.org/10.1186/1741-7015-10-104>. **b)** Birnkrant, D.; Cox, E. The Emergency Use Authorization of Peramivir for Treatment of 2009 H1N1 Influenza. *New England Journal of Medicine* **2009**, 361 (23), 2204–2207. <https://doi.org/10.1056/NEJMp0910479>.
- ¹² **a)** Hamilton, B. S.; Whittaker, G. R.; Daniel, S. Influenza Virus-Mediated Membrane Fusion: Determinants of Hemagglutinin Fusogenic Activity and Experimental Approaches for Assessing Virus Fusion. *Viruses* **2012**, 4 (7), 1144–1168. <https://doi.org/10.3390/v4071144>. **b)** Shen, Z.; Lou, K.; Wang, W. New Small-Molecule Drug Design Strategies for Fighting Resistant Influenza A. *Acta Pharm Sin B* **2015**, 5 (5), 419–430. <https://doi.org/10.1016/j.apsb.2015.07.006>.
- ¹³ Vorobjev, Y. N. Design of an Efficient Inhibitor for the Influenza A Virus M2 Ion Channel. *Molekuliarnaia biologii* **2020**, 54 (2), 321–332. <https://doi.org/10.31857/S0026898420020160>.
- ¹⁴ Marjuki, H.; Mishin, V. P.; Chesnokov, A. P.; De La Cruz, J. A.; Fry, A. M.; Villanueva, J.; Gubareva, L. V. An Investigational Antiviral Drug, DAS181, Effectively Inhibits Replication of Zoonotic Influenza A Virus Subtype H7N9 and Protects Mice From Lethality. *J Infect Dis* **2014**, 210 (3), 435–440. <https://doi.org/10.1093/infdis/jiu105>.
- ¹⁵ Ye, Q.; Krug, R. M.; Tao, Y. J. The Mechanism by Which Influenza A Virus Nucleoprotein Forms Oligomers and Binds RNA. *Nature* **2006**, 444 (7122), 1078–1082. <https://doi.org/10.1038/nature05379>.
- ¹⁶ **a)** Kao, R. Y.; Yang, D.; Lau, L. S.; Tsui, W. H. W.; Hu, L.; Dai, J.; Chan, M. P.; Chan, C. M.; Wang, P.; Zheng, B. J.; Sun, J.; Huang, J. D.; Madar, J.; Chen, G.; Chen, H.; Guan, Y.; Yuen, K. Y. Identification of Influenza A Nucleoprotein as an Antiviral Target. *Nat Biotechnol* **2010**, 28 (6), 600–605. <https://doi.org/10.1038/nbt.1638>. **b)** Su, C. Y.; Cheng, T. J. R.; Lin, M. I.; Wang, S. Y.; Huang, W. I.; Lin-Chua, S. Y.; Chen, Y. H.; Wu, C. Y.; Lai, M. M. C.; Cheng, W. C.; Wu, Y. T.; Tsai, M. D.; Cheng, Y. S. E.; Wong, C. H. High-Throughput Identification of Compounds Targeting Influenza RNA-Dependent RNA Polymerase Activity. *Proc Natl Acad Sci U S A* **2010**, 107 (45), 19151–19156. <https://doi.org/10.1073/pnas.10135921097>. **c)** Kao, Y. R. Antiviral compounds and methods of marking and using there of cross reference to related applications. WO 2011/015037 A1, 2011.
- ¹⁷ **a)** Wermuth, C. G. *The Practice of Medicinal Chemistry*, 3rd ed.; Elsevier, 2008. **b)** Fernandes, J. P. S. The Importance of Medicinal Chemistry Knowledge in the Clinical Pharmacist's Education. *Am J Pharm Educ* **2018**, 82 (2), 106–114. <https://doi.org/10.5688/ajpe6083>.

- ¹⁸ Negri-Cesi, P.; Colciago, A.; Celotti, F.; Motta, M. Sexual Differentiation of the Brain: Role of Testosterone and Its Active Metabolites. *J Endocrinol Invest* **2004**, *27* (6 Suppl), 120–127.
- ¹⁹ Blokland, A. Acetylcholine: A Neurotransmitter for Learning and Memory? *Brain Res Rev* **1995**, *21* (3), 285–300. [https://doi.org/10.1016/0165-0173\(95\)00016-X](https://doi.org/10.1016/0165-0173(95)00016-X).
- ²⁰ **a)** del Carmen Fernández-Alonso, M.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. Molecular Recognition of Saccharides by Proteins. Insights on the Origin of the Carbohydrate-Aromatic Interactions. *J Am Chem Soc* **2005**, *127* (20), 7379–7386. <https://doi.org/10.1021/ja051020+>. **b)** Valverde, P.; Martínez, J. D.; Cañada, F. J.; Ardá, A.; Jiménez-Barbero, J. Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* **2020**, *21* (21), 2999–3025. <https://doi.org/10.1002/cbic.202000238>.
- ²¹ Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct* **2003**, *32*, 335–373. <https://doi.org/10.1146/annurev.biophys.32.110601.142532>.
- ²² *Tactics in Contemporary Drug Design*; Meanwell, N. A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2015; Vol. 9. <https://doi.org/10.1007/978-3-642-55041-6>.
- ²³ Souza, S. D.; Prema, K. V. Machine Learning Models for Drug – Target Interactions : Current Knowledge and Future Directions. *Drug Discov Today* **2020**, *25* (4). <https://doi.org/10.1016/j.drudis.2020.03.003>.
- ²⁴ **a)** Carpenter, K. A.; Cohen, D. S.; Jarrell, J. T.; Huang, X. Deep Learning and Virtual Drug Screening. *Future Med Chem* **2018**, *10* (21), 2557–2567. <https://doi.org/10.4155/fmc-2018-0314>. **b)** Zhang, L.; Tan, J.; Han, D.; Zhu, H. From Machine Learning to Deep Learning : Progress in Machine Intelligence for Rational Drug Discovery. *Drug Discov Today* **2017**, *22* (11), 1680–1685. <https://doi.org/10.1016/j.drudis.2017.08.010>.
- ²⁵ Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. v.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem Soc Rev* **2020**, *49* (11), 3525–3564. <https://doi.org/10.1039/D0CS00098A>.
- ²⁶ **a)** Subramanian, G.; Ramsundar, B.; Pande, V.; Denny, R. A. Computational Modeling of β -Secretase 1 (BACE-1) Inhibitors Using Ligand Based Approaches. *J Chem Inf Model* **2016**, *56* (10), 1936–1949. <https://doi.org/10.1021/acs.jcim.6b00290>. **b)** Coi, A.; Fiamingo, F. L.; Livi, O.; Calderone, V.; Martelli, A.; Massarelli, I.; Bianucci, A. M. QSAR Studies on BK Channel Activators. *Bioorg Med Chem* **2009**, *17* (1), 319–325. <https://doi.org/10.1016/j.bmc.2008.10.068>. **c)** Baidya, A. T. K.; Ghosh, K.; Amin, Sk. A.; Adhikari, N.; Nirmal, J.; Jha, T.; Gayen, S. In Silico Modelling, Identification of Crucial Molecular Fingerprints, and Prediction of New Possible Substrates of Human Organic Cationic Transporters 1 and 2. *New Journal of Chemistry* **2020**, *44* (10), 4129–4143. <https://doi.org/10.1039/C9NJ05825G>. **d)** VEERASAMY, R.; RAJAK, H. QSAR Studies on Neuraminidase Inhibitors as Anti-Influenza Agents. *Turk J Pharm Sci* **2021**, *18* (2), 151–156. <https://doi.org/10.4274/tjps.galenos.2020.45556>. **e)** García-Jacas, C. R.; Martínez-Mayorga, K.; Marrero-Ponce, Y.; Medina-Franco, J. L. Conformation-Dependent QSAR Approach for the Prediction of Inhibitory Activity of Bromodomain Modulators. *SAR QSAR Environ Res* **2017**, *28* (1), 41–58. <https://doi.org/10.1080/1062936X.2017.1278616>. **f)** Khaled, D. M.; Elshakre, M. E.; Noamaan, M. A.; Butt, H.; Abdel Fattah, M. M.; Gaber, D. A. A Computational QSAR, Molecular Docking and In Vitro Cytotoxicity Study of Novel Thiouracil-Based Drugs with Anticancer Activity against Human-DNA Topoisomerase II. *Int J Mol Sci* **2022**, *23* (19), 11799. <https://doi.org/10.3390/ijms231911799>. **g)** Naik, P. K.; Dubey, A.; Kumar, R. Development of Predictive Quantitative Structure-Activity Relationship Models of Epipodophyllotoxin Derivatives. *SLAS Discovery* **2010**, *15* (10), 1194–1203. <https://doi.org/10.1177/1087057110380743>. **c)** Zheng, F.; Zhan, M.; Huang, X.; Abdul Hameed, M. D. M.; Zhan, C.-G. Modeling in Vitro Inhibition of Butyrylcholinesterase Using Molecular Docking, Multi-Linear Regression and Artificial Neural Network Approaches. *Bioorg Med Chem* **2014**, *22* (1), 538–549. <https://doi.org/10.1016/j.bmc.2013.10.053>.
- ²⁷ Massarelli, I.; Imbriani, M.; Coi, A.; Saraceno, M.; Carli, N.; Bianucci, A. M. Development of QSAR Models for

- Predicting Hepatocarcinogenic Toxicity of Chemicals. *Eur J Med Chem* **2009**, *44* (9), 3658–3664. <https://doi.org/10.1016/j.ejmech.2009.02.014>.
- ²⁸ Wu, Z.; Park, H.-Y.; Xie, D.; Yang, J.; Hou, S.; Shahzad, N.; Kim, C. K.; Yang, S. Synthesis, Biological Evaluation, and 3D-QSAR Studies of *N*-(Substituted Pyridine-4-Yl)-1-(Substituted Phenyl)-5-Trifluoromethyl-1 *H* -Pyrazole-4-Carboxamide Derivatives as Potential Succinate Dehydrogenase Inhibitors. *J Agric Food Chem* **2021**, *69* (4), 1214–1223. <https://doi.org/10.1021/acs.jafc.0c05702>.
- ²⁹ Pereira, F.; Latino, D.; Gaudêncio, S. QSAR-Assisted Virtual Screening of Lead-Like Molecules from Marine and Microbial Natural Sources for Antitumor and Antibiotic Drug Discovery. *Molecules* **2015**, *20* (3), 4848–4873. <https://doi.org/10.3390/molecules20034848>.
- ³⁰ Low, Y. S.; Alves, V. M.; Fourches, D.; Sedykh, A.; Andrade, C. H.; Muratov, E. N.; Rusyn, I.; Tropsha, A. Chemistry-Wide Association Studies (CWAS): A Novel Framework for Identifying and Interpreting Structure–Activity Relationships. *J Chem Inf Model* **2018**, *58* (11), 2203–2213. <https://doi.org/10.1021/acs.jcim.8b00450>.
- ³¹ Toxicology Testing in the 21st Century (Tox21). <https://www.epa.gov/chemical-research/toxicology-testing-21st-century-tox21> (accesso 09-04-2023).
- ³² Talete, S. R. L. DRAGON for Windows (Software for Molecular Descriptor Calculations). Kode cheminformatics
- ³³ Jakhar, D.; Kaur, I. Artificial Intelligence, Machine Learning and Deep Learning: Definitions and Differences. *Clin Exp Dermatol* **2020**, *45* (1), 131–132. <https://doi.org/10.1111/ced.14029>.
- ³⁴ **a)** Samuel, A. L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev* **1959**, *3* (3), 210–229. <https://doi.org/10.1147/rd.33.0210>. **b)** Machine Learning in Radiation Oncology; el Naqa, I., Li, R., Murphy, M. J., Eds.; Springer International Publishing: Cham, 2015. <https://doi.org/10.1007/978-3-319-18305-3>.
- ³⁵ Basile, A. O.; Yahy, A.; Tatonetti, N. P. Artificial Intelligence for Drug Toxicity and Safety. *Trends Pharmacol Sci* **2019**, *40* (9), 624–635. <https://doi.org/10.1016/j.tips.2019.07.005>.
- ³⁶ Alpaydin, E. Introduction to Machine Learning, 3rd ed.; Cambridge, MA; The MIT Press: London, England, 2014.
- ³⁷ **a)** Madzhidov, T. I.; Rakhimbekova, A.; Afonina, V. A.; Gimadiev, T. R.; Mukhametgaleev, R. N.; Nugmanov, R. I.; Baskin, I. I.; Varnek, A. Machine Learning Modelling of Chemical Reaction Characteristics: Yesterday, Today, Tomorrow. *Mendeleev Communications* **2021**, *31* (6), 769–780. <https://doi.org/10.1016/j.mencom.2021.11.003>. **b)** Jiang, Y.; Yu, Y.; Kong, M.; Mei, Y.; Yuan, L.; Huang, Z.; Kuang, K.; Wang, Z.; Yao, H.; Zou, J.; Coley, C. W.; Wei, Y. Artificial Intelligence for Retrosynthesis Prediction. *Engineering* **2022**. <https://doi.org/10.1016/j.eng.2022.04.021>. **c)** Johansson, S.; Thakkar, A.; Kogej, T.; Bjerrum, E.; Genheden, S.; Bastys, T.; Kannas, C.; Schliep, A.; Chen, H.; Engkvist, O. AI-Assisted Synthesis Prediction. *Drug Discov Today Technol* **2019**, *32–33*, 65–72. <https://doi.org/10.1016/j.ddtec.2020.06.002>.
- ³⁸ **a)** *Supervised and Unsupervised Learning for Data Science*; Berry, M. W., Mohamed, A., Yap, B. W., Eds.; Springer International Publishing: Cham, 2020. <https://doi.org/10.1007/978-3-030-22475-2>. **b)** Delua, J. *Supervised vs. Unsupervised Learning: What's the Difference?* IBM. <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning> (accesso 09-04-2023).
- ³⁹ **a)** Müller, A. C.; Guido, S. Introduction to Machine Learning with Python; O'Reilly Media, Inc., 2016. **b)** Sarkar, D.; Bali, R.; Sharma, T. Practical Machine Learning with Python; Apress: Berkeley, CA, 2018.

<https://doi.org/10.1007/978-1-4842-3207-1>.

- ⁴⁰ Wu, Z.; Zhu, M.; Kang, Y.; Leung, E. L.-H.; Lei, T.; Shen, C.; Jiang, D.; Wang, Z.; Cao, D.; Hou, T. Do We Need Different Machine Learning Algorithms for QSAR Modeling? A Comprehensive Assessment of 16 Machine Learning Algorithms on 14 QSAR Data Sets. *Brief Bioinform* **2021**, *22* (4), 1–17. <https://doi.org/10.1093/bib/bbaa321>.
- ⁴¹ **a)** Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E. How Not to Develop a Quantitative Structure–Activity or Structure–Property Relationship (QSAR/QSPR). *SAR QSAR Environ Res* **2009**, *20* (3–4), 241–266. <https://doi.org/10.1080/10629360902949567>. **b)** Topliss, J. G. Utilization of Operational Schemes for Analog Synthesis in Drug Design. *J Med Chem* **1972**, *15* (10), 1006–1011. <https://doi.org/10.1021/jm00280a002>. **c)** Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J Med Chem* **2014**, *57* (12), 4977–5010. <https://doi.org/10.1021/jm4004285>.
- ⁴² Pantelev, J.; Gao, H.; Jia, L. Recent Applications of Machine Learning in Medicinal Chemistry. *Bioorg Med Chem Lett* **2018**, *28* (17), 2807–2815. <https://doi.org/10.1016/j.bmcl.2018.06.046>.
- ⁴³ Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18* (8), 2674. <https://doi.org/10.3390/s18082674>.
- ⁴⁴ Tao, H.; Wu, T.; Aldeghi, M.; Wu, T. C.; Aspuru-Guzik, A.; Kumacheva, E. Nanoparticle Synthesis Assisted by Machine Learning. *Nat Rev Mater* **2021**, *6* (8), 701–716. <https://doi.org/10.1038/s41578-021-00337-5>.
- ⁴⁵ **a)** Veisi, H. Introduction to SVM; 2023; pp 3–18. https://doi.org/10.1007/978-981-19-6553-1_1. **b)** Vapnik, V. The Support Vector Method of Function Estimation. In *Nonlinear Modeling*; Springer US: Boston, MA, 1998; pp 55–85. https://doi.org/10.1007/978-1-4615-5703-6_3.
- ⁴⁶ Martínez, M. J.; Razuc, M.; Ponzone, I. MoDeSuS: A Machine Learning Tool for Selection of Molecular Descriptors in QSAR Studies Applied to Molecular Informatics. *Biomed Res Int* **2019**, *2019*, 1–12 <https://doi.org/10.1155/2019/2905203>
- ⁴⁷ **a)** Breiman, L. Random Forests. *Mach Learn* **2001**, *45* (1), 5–32. <https://doi.org/10.1023/A:1010933404324>. **b)** Tin Kam Ho. The Random Subspace Method for Constructing Decision Forests. *IEEE Trans Pattern Anal Mach Intell* **1998**, *20* (8), 832–844. <https://doi.org/10.1109/34.709601>.
- ⁴⁸ Gomez-jimenez, G.; Gonzalez-ponce, K.; Castillo-pazos, D. J.; Madariaga-mazon, A.; Barroso-flores, J.; Cortes-Guzman, F.; Martinez-Mayorga, K. The OECD Principles for (Q) SAR Models in the Context of Knowledge Discovery in Databases (KDD), 1st ed.; Elsevier Inc., **2018**. <https://doi.org/10.1016/bs.apcsb.2018.04.001>.
- ⁴⁹ Aher, R. B.; Khan, K.; Roy, K. Ecotoxicological QSARs; Roy, K., Ed.; Methods in Pharmacology and Toxicology; Springer US: New York, NY, 2020. <https://doi.org/10.1007/978-1-0716-0150-1>.
- ⁵⁰ Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules* **2012**, *17* (5), 4791–4810. <https://doi.org/10.3390/molecules17054791>.
- ⁵¹ Rosas-Jimenez, J. G.; Garcia-Revilla, M. A.; Madariaga-Mazon, A.; Martinez-Mayorga, K. Predictive Global Models of Cruzain Inhibitors with Large Chemical Coverage. *ACS Omega* **2021**, *6* (10), 6722–6735. <https://doi.org/10.1021/acsomega.0c05645>.
- ⁵² Gissi, A.; Gadaleta, D.; Floris, M.; Olla, S.; Carotti, A.; Novellino, E.; Benfenati, E.; Nicolotti, O. An alternative QSAR-based approach for predicting the bioconcentration factor for regulatory purposes. *ALTEX* -

- Alternatives to animal experimentation* **2014**, 31(1), 23–36. <https://doi.org/10.14573/altex.1305221>.
- ⁵³ Gadaleta, D.; Mangiatordi, G. F.; Catto, M.; Carotti, A.; Nicolotti, O. Applicability Domain for QSAR Models. International. *Journal of Quantitative Structure-Property Relationships* **2016**, 1 (1), 45–63. <https://doi.org/10.4018/IJQSPR.2016010102>.
- ⁵⁴ Ferreira, L.; dos Santos, R.; Oliva, G.; Andricopulo, A. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, 20 (7), 13384–13421. <https://doi.org/10.3390/molecules200713384>.
- ⁵⁵ Bitencourt-Ferreira, G.; de Azevedo, W. F. Docking Screens for Drug Discovery; de Azevedo, W. F., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2019; Vol. 2053. <https://doi.org/10.1007/978-1-4939-9752-7>.
- ⁵⁶ a) Yuriev, E.; Holien, J.; Ramsland, P. A. Improvements, Trends, and New Ideas in Molecular Docking: 2012–2013 in Review. *Journal of Molecular Recognition* **2015**, 28 (10), 581–604. <https://doi.org/10.1002/jmr.2471>.
b) Gervasio, F. L.; Spiwok, V. Biomolecular Simulations in Structure-Based Drug Discovery; 2018.
- ⁵⁷ Coumar, M. S. Molecular Docking for Computer-Aided Drug Design; Elsevier, 2021. <https://doi.org/10.1016/C2019-0-04960-6>
- ⁵⁸ Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for Molecular Docking: A Review. *Biophys Rev* **2017**, 9 (2), 91–102. <https://doi.org/10.1007/s12551-016-0247-1>.
- ⁵⁹ Levinthal, C.; Wodak, S. J.; Kahn, P.; Dadivanian, A. K. Hemoglobin interaction in sickle cell fibers I: Theoretical approaches to the molecular contacts. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, 72 (4), 1330–1334. <https://doi.org/10.1073/pnas.72.4.1330>.
- ⁶⁰ Gupta, M.; Sharma, R.; Kumar, A. Docking Techniques in Pharmacology: How Much Promising? *Comput Biol Chem* **2018**, 76 (February), 210–217. <https://doi.org/10.1016/j.compbiolchem.2018.06.005>.
- ⁶¹ Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in docking success rates due to dataset preparation. *J. Comput. Aided. Mol. Des.* **2012**, 26 (6), 775–786. <https://doi.org/10.1007/s10822-012-9570-1>.
- ⁶² Molecular Operating Environment (MOE), 2022.02. Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7 2022.
- ⁶³ Jewitz. Microbiología general. Riedel, S.; Hobden, J. A.; Miller, S.; Morse, S. A.; Mietzner, T. A.; Detrick, B.; Mitchell, T. G.; Sakanari, J. A.; Hotez, P.; Mejia, R. *Medical Microbiology*, 28th edition.; Riedel, S., Morse, S. A., Mietzner, T., Miller, S., Eds.; Mc Graw Hill, 2019.
- ⁶⁴ a) Qiu, C.; Lu, P. Avian Influenza in Human; 2021. <https://doi.org/10.1007/978-981-16-1429-3>. b) Yohei Yamauchi. Influenza Virus; Yamauchi, Y., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. 1836. <https://doi.org/10.1007/978-1-4939-8678-1>.
- ⁶⁵ Franco-Paredes C, Río C, Peter Carrasco P, et al. Respuesta en México al actual brote de influenza AH1N1. *salud pública mex.* **2009**, 51(3), 183-186.
- ⁶⁶ Boltz, D. A.; Aldridge, J. R.; Webster, R. G.; Govorkova, E. A. Drugs in Development for Influenza. *Drugs* **2010**, 70 (11), 1349–1362. <https://doi.org/10.2165/11537960-000000000-00000>.
- ⁶⁷ Solórzano SF, Miranda NMG. Influenza. *Bol Med Hosp Infant Mex* **2009**, 66, 461-473.
- ⁶⁸ López Cervantes M, García Moreno J, Pacheco Domínguez RL, Escamilla Santiago RA, Villanueva Lozano M. La influenza A/H1N1 2009. Una crónica de la primera pandemia del siglo XXI. *Revista Digital Universitaria* **2010**, 11(4).
- ⁶⁹ Infección humana por el virus de la gripe aviar A(H5) - Ecuador. OMS.

- [https://www.who.int/es/emergencies/disease-outbreak-news/item/2023-DON434#:~:text=El%209%20de%20enero%20de,provincia%20de%20Bol%C3%ADvar%20\(Ecuador\)](https://www.who.int/es/emergencies/disease-outbreak-news/item/2023-DON434#:~:text=El%209%20de%20enero%20de,provincia%20de%20Bol%C3%ADvar%20(Ecuador).). (acceso 2023-03-30).
- ⁷⁰ Fujita, J. *Influenza*; Fujita, J., Ed.; Respiratory Disease Series: Diagnostic Tools and Disease Managements; Springer Singapore: Singapore, 2021. <https://doi.org/10.1007/978-981-15-9109-9>.
- ⁷¹ Sarker, A.; Gu, Z.; Mao, L.; Ge, Y.; Hou, D.; Fang, J.; Wei, Z.; Wang, Z. Influenza-Existing Drugs and Treatment Prospects. *Eur J Med Chem* **2022**, *232*, 114-189. <https://doi.org/10.1016/j.ejmech.2022.114189>.
- ⁷² Murugan, N. A.; Raja, K. M. P.; Saraswathi, N. T. *Antiviral Drug Discovery and Development*; Liu, X., Zhan, P., Menéndez-Arias, L., Poongavanam, V., Eds.; Advances in Experimental Medicine and Biology; Springer Singapore: Singapore, 2021; Vol. 1322. <https://doi.org/10.1007/978-981-16-0267-2>.
- ⁷³ RübSamen-Schaeff, H.; Buschmann, H. *New Drug Development for Known and Emerging Viruses; Methods and Principles in Medicinal Chemistry*; Wiley, 2022. <https://doi.org/10.1002/9783527810697>.
- ⁷⁴ **a)** Hamilton, B. S.; Whittaker, G. R.; Daniel, S. Influenza Virus-Mediated Membrane Fusion: Determinants of Hemagglutinin Fusogenic Activity and Experimental Approaches for Assessing Virus Fusion. *Viruses* **2012**, *4* (7), 1144–1168. <https://doi.org/10.3390/v4071144>. **b)** Shen, Z.; Lou, K.; Wang, W. New Small-Molecule Drug Design Strategies for Fighting Resistant Influenza A. *Acta Pharm Sin B* **2015**, *5* (5), 419–430. <https://doi.org/10.1016/j.apsb.2015.07.006>.
- ⁷⁵ Vorobjev, Y. N. Design of an Efficient Inhibitor for the Influenza A Virus M2 Ion Channel. *Molekuliarnaia biologiia* **2020**, *54* (2), 321–332. <https://doi.org/10.31857/S0026898420020160>.
- ⁷⁶ Marjuki, H.; Mishin, V. P.; Chesnokov, A. P.; De La Cruz, J. A.; Fry, A. M.; Villanueva, J.; Gubareva, L. V. An Investigational Antiviral Drug, DAS181, Effectively Inhibits Replication of Zoonotic Influenza A Virus Subtype H7N9 and Protects Mice From Lethality. *J Infect Dis* **2014**, *210* (3), 435–440. <https://doi.org/10.1093/infdis/jiu105>.
- ⁷⁷ Amarelle, L.; Lecuona, E.; Sznajder, J. I. Tratamiento Antigripal : Fármacos Actualmente Utilizados y Nuevos Agentes En Desarrollo. *Arch Bronconeumol* **2017**, *53* (1), 19–26. <https://doi.org/http://dx.doi.org/10.1016/j.arbres.2016.07.004>.
- ⁷⁸ Cianci, C.; Gerritz, S. W.; Deminie, C.; Krystal, M. Review Influenza Nucleoprotein : Promising Target for Antiviral Chemotherapy. *Antiviral Chemistry & Chemotherapy* **2013**, *91*, 77–91. <https://doi.org/doi:10.3851/IMP2235>.
- ⁷⁹ Ye, Q.; Krug, R. M.; Tao, Y. J. The Mechanism by Which Influenza A Virus Nucleoprotein Forms Oligomers and Binds RNA. *Nature* **2006**, *444* (7122), 1078–1082. <https://doi.org/10.1038/nature05379>.
- ⁸⁰ Tarus, B.; Bertrand, H.; Zedda, G.; Di Primo, C.; Quideau, S.; Slama-Schwok, A. Structure-Based Design of Novel Naproxen Derivatives Targeting Monomeric Nucleoprotein of Influenza A Virus. *J Biomol Struct Dyn* **2015**, *33* (9), 1899–1912. <https://doi.org/10.1080/07391102.2014.979230>.
- ⁸¹ Hu, Y.; Sneyd, H.; Dekant, R.; Wang, J. Influenza A Virus Nucleoprotein: A Highly Conserved Multi-Functional Viral Protein as a Hot Antiviral Drug Target. *Curr Top Med Chem* **2017**, *17* (20), 2271–2285. <https://doi.org/10.2174/1568026617666170224122508>.
- ⁸² Shen, Y. F.; Chen, Y. H.; Chu, S. Y.; Lin, M. I.; Hsu, H. T.; Wu, P. Y.; Wu, C. J.; Liu, H. W.; Lin, F. Y.; Lin, G.; Hsu, P. H.; Yang, A. S.; Cheng, Y. S. E.; Wu, Y. T.; Wong, C. H.; Tsai, M. D. E339...R416 Salt Bridge of Nucleoprotein as a Feasible Target for Influenza Virus Inhibitors. *Proc Natl Acad Sci U S A* **2011**, *108* (40), 16515–16520. <https://doi.org/10.1073/pnas.1113107108>.

- ⁸³ White, K. M.; Abreu, P.; Wang, H.; De Jesus, P. D.; Manicassamy, B.; García-Sastre, A.; Chanda, S. K.; DeVita, R. J.; Shaw, M. L. Broad Spectrum Inhibitor of Influenza A and B Viruses Targeting the Viral Nucleoprotein. *ACS Infect Dis* **2018**, *4* (2), 146–157. <https://doi.org/10.1021/acinfecdis.7b00120>.
- ⁸⁴ Gerritz, S. W.; Cianci, C.; Kim, S.; Pearce, B. C.; Deminie, C.; Discotto, L. Inhibition of influenza Virus Replication via Small Molecules That Induce the Formation of Higher-Order Nucleoprotein Oligomers. *PNAS* **2011**, *108* (37), 15366–15371. <https://doi.org/10.1073/pnas.1107906108>.
- ⁸⁵ Homman M, Kingi N, Bergman J, Engqvist R. Methanethione compounds having antiviral activity. WO2013171334A1, 2013. p. 45.
- ⁸⁶ Cheng, H.; Wan, J.; Lin, M. I.; Liu, Y.; Lu, X.; Liu, J.; Xu, Y.; Chen, J.; Tu, Z.; Cheng, Y. S. E.; Ding, K. Design, Synthesis, and in Vitro Biological Evaluation of 1 H -1,2,3-Triazole-4-Carboxamide Derivatives as New Anti-Influenza A Agents Targeting Virus Nucleoprotein. *J Med Chem* **2012**, *55* (5), 2144–2153. <https://doi.org/10.1021/jm2013503>.
- ⁸⁷ Pei, S.; Xia, S.; Yang, F.; Chen, J.; Wang, M.; Sun, W.; Li, Z.; Yuan, K.; Chen, J. Design, Synthesis and in Vitro Biological Evaluation of Isoxazol-4-Carboxa Piperidyl Derivatives as New Anti-Influenza A Agents Targeting Virus Nucleoprotein. *RSC Adv* **2020**, *10* (8), 4446–4454. <https://doi.org/10.1039/c9ra10828a>.
- ⁸⁸ **a)** Correa Padilla, E. A. Síntesis de anaálogos de nucleozina con potencial actividad antiviral que incorporan el sistema 2,5 diazabicyclo[2.2.1]heptano [Tesis], F.E.S. Zaragoza, UNAM, 2018. Disponible en: https://repositorio.unam.mx/contenidos/sintesis-de-analogos-de-nucleozina-con-potencial-actividad-antiviral-que-incorporan-el-sistema-25-diazabicyclo221h-102062?c=Bd9V5K&d=false&q=*.:&i=3&v=1&t=search_0&as=0 **b)** Correa-Padilla, E.; Hernández-Cano, A.; Cuevas, G.; Acevedo-Betancur, Y.; Esquivel-Guadarrama, F.; Martínez-Mayorga, K. Modifications in the Piperazine Ring of Nucleozin Affect Anti-Influenza Activity. *PLoS One* **2023**, *18* (2), e0277073. <https://doi.org/10.1371/journal.pone.0277073>.
- ⁸⁹ Liao, J.; Cheng, H.; Wan, J.; Chen, P.; Li, Y.; Ding, K.; Tortorella, M. D.; Tu, Z.; Zhang, Y. Evaluation of Benzamide Derivatives as New Influenza A Nucleoprotein Inhibitors. *Open J Med Chem* **2016**, *6*, 43–50. <https://doi.org/http://dx.doi.org/10.4236/ojmc.2016.63004>.
- ⁹⁰ Pang, B.; Cheung, N. N.; Zhang, W.; Dai, J.; Kao, R. Y.; Zhang, H.; Hao, Q. Structural Characterization of H1N1 Nucleoprotein-Nucleozin Binding Sites. *Sci Rep* **2016**, *6*, 1–9. <https://doi.org/10.1038/srep29684>.
- ⁹¹ Molecular Operating Environment (MOE), 2022.02. Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7 2022.
- ⁹² ChemDraw. CambridgeSoft Corporation: 875 Massachusetts Ave., Cambridge, MA 02139, USA.
- ⁹³ Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. *Journal of Machine Learning Research* **2012**, *12*, 2825–2830.
- ⁹⁴ Gross, F. L.; Bai, Y.; Jefferson, S.; Holiday, C.; Levine, M. Z. Measuring Influenza Neutralizing Antibody Responses to A(H3N2) Viruses in Human Sera by Microneutralization Assays Using MDCK-SIAT1 Cells. *Journal of Visualized Experiments* **2017**, No. 129. <https://doi.org/10.3791/56448>.