



UNIVERSIDAD REGIONAL AMAZÓNICA IKIAM

FACULTAD CIENCIAS DE LA VIDA

CARRERA DE BIOTECNOLOGÍA

**Predicción *in silico* de la bioactividad antimicrobiana de una Librería
virtual de péptidos usando modelos de Aprendizaje Automatizado**

Proyecto de Investigación previo a la obtención del título de:

INGENIERO EN BIOTECNOLOGÍA

AUTOR: ARNOLDD RODRIGO HERNÁNDEZ HURTADO

TUTOR: MSC. Marco Andrés Viteri Yáñez

Napo, Ecuador

2024

Carrera de Biotecnología

Declaración de derecho de autor, autenticidad y responsabilidad

Tena, 29 de agosto de 2024

De mi consideración:

Yo, Hernández Hurtado Arnoldo Rodrigo con documento de identidad N.º 2200280564, declaro que los resultados obtenidos en la investigación que presento en este documento final, previo a la obtención del título Ingeniero en Biotecnología, son absolutamente inéditos, originales, auténticos y personales.

En virtud de lo cual, el contenido, criterios, opiniones, resultados, análisis, interpretaciones, conclusiones, recomendaciones y todos los demás aspectos vertidos en el presente documento, son de mi autoría y de mi absoluta responsabilidad.

Por la favorable atención a la presente, suscribo de usted,

A handwritten signature in blue ink, consisting of several overlapping loops and strokes, positioned above the printed name.

Hernández Hurtado Arnoldo Rodrigo

CI: 2200280564

Carrera de Biotecnología

Autorización de Publicación en el Repositorio Institucional

Tena, 29 de agosto de 2024

Yo, Hernández Hurtado Arnoldo Rodrigo, con documento de identidad N.º 2200280564 en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulación “Predicción *in silico* de la bioactividad antimicrobiana de una Librería virtual de péptidos usando modelos de Aprendizaje Automatizado” de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN, reconozco a favor de la Universidad Regional Amazónica Ikiám una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Así mismo autorizo a la Universidad Regional Amazónica Ikiám para que realice la publicación de este trabajo de titulación en el Repositorio Institucional de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.



Hernández Hurtado Arnoldo Rodrigo

CI: 2200280564

Carrera de Biotecnología

Certificado de Dirección de Trabajo de Titulación

Tena, 29 de agosto de 2024

Certifico que el trabajo de titulación: “Predicción *in silico* de la bioactividad antimicrobiana de una Librería virtual de péptidos usando modelos de Aprendizaje Automatizado” aprobado bajo el mecanismo de titulación de: Tesis, fue realizado Hernández Hurtado Arnoldo Rodrigo, bajo mi dirección.

El mismo ha sido revisado en su totalidad y analizado por la herramienta de verificación de similitud de contenido; por lo tanto, cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad Regional Amazónica Ikiám, para su entrega y defensa.



MSc. Marco Andrés Viteri Yáñez
C.I: 1718163759

ÍNDICE DE CONTENIDOS

DECLARACIÓN DE DERECHO DE AUTOR, AUTENTICIDAD Y RESPONSABILIDAD.....	ii
AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL	iii
CERTIFICADO DE DIRECCIÓN DE TRABAJO DE TITULACIÓN.....	iv
ÍNDICE DE CONTENIDOS	v
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS.....	viii
RESUMEN	ix
ABSTRACT.....	x
1. CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN	1
1.1 Antecedentes	1
1.1.1 Definiciones y clasificación.....	1
1.1.2 Propiedades Físicoquímicas	3
1.1.3 Inteligencia artificial (IA)	6
1.2 Planteamiento del problema.....	11
1.3 Justificación de la Investigación.....	12
1.4 Objetivos de la investigación	13
1.4.1 Objetivo general	13
1.4.2 Objetivos específicos.....	13
2. CAPITULO II. MARCO METODOLÓGICO.....	14
2.1 Construcción de la librería (base de datos)	14
2.2 Determinación de propiedades físicoquímicas.	15
2.3 Construcción de los modelos	16
2.4 Evaluación de métricas de rendimiento	17
2.5 Construcción del algoritmo de recomendación	19
3 CAPÍTULO III: PRESENTACIÓN DE DATOS Y RESULTADOS.....	20
3.1 Exploración inicial de la Librería virtual (base de datos)	20
3.2 Comparativa de la librería propuesta para el cálculo de propiedades físicoquímicas. ...	21
3.3 Determinación de propiedades físicoquímicas de la Librería virtual	23
3.4 Entrenamiento de los modelos	24
3.5 Validación de los modelos con secuencias peptídicas externas a la Librería virtual.....	30

3.6	Validación de un algoritmo de recomendación	33
CAPITULO IV: INTERPRETACIÓN Y DISCUSIÓN.....		35
3.7	Distribución de aminoácidos de la Librería virtual.....	35
3.8	Comparativa de la librería propuesta para el cálculo de propiedades fisicoquímicas ...	35
3.9	En cuanto a la distribución de frecuencias de propiedades fisicoquímicas de la Librería virtual	36
3.10	Entrenamiento de los modelos	37
3.11	Validación de los modelos con secuencias peptídicas externas a la Librería virtual.....	37
3.12	Validación de un algoritmo de recomendación	38
CAPITULO V: CONCLUSIONES Y RECOMENDACIONES.....		40
3.13	Conclusiones.....	40
3.14	Recomendaciones	41
REFERENCIAS BIBLIOGRÁFICAS.....		
ANEXOS.....		

ÍNDICE DE TABLAS

Tabla 1. Métodos de cálculo de las propiedades fisicoquímicas.	15
Tabla 2. Proceso de filtrado de secuencias desde el minado inicial hasta el etiquetado por actividad.	20
Tabla 3. Comparación de las propiedades fisicoquímicas entre el algoritmo propuesto (my_properties) con herramientas disponibles en la web.	22
Tabla 4. Resumen métricas de precisión para el modelo de Red Neuronal propuesto (my_ANN).	26
Tabla 5 . Resumen métricas de precisión para el modelo Máquinas de Vectores de Soporte (my_SVM).	28
Tabla 6. Comparación de métricas de rendimiento de los modelos propuestos.	30
Tabla 7. Comparativa entre el modelo propuesto, con métodos de última generación, en la predicción de 8 AMPs conocidos.	31
Tabla 8. Comparativa entre el modelo propuesto, con métodos de última generación, en la predicción de 8 No-AMPs conocidos.	32
Tabla 9. Funcionalidad de un algoritmo de recomendación.	34

ÍNDICE DE FIGURAS

Figura 1. Flujo de trabajo básico de la aplicación de modelos predictivos de Aprendizaje Automatizado.	7
Figura 2. Esquematización de una Neurona y Red Neuronal Artificial. Fuente: (Prince, 2023; Tablada - Germán & Torres, 2009)	9
Figura 3. Arquitectura básica de una Red Neuronal Artificial. Fuente: (Prince, 2023)	10
Figura 4. Arquitectura de la Red Neuronal utilizada.....	17
Figura 5. Distribución de aminoácidos de los péptidos en base a la CMI por actividad de la Librería virtual.	21
Figura 6. Distribución y tendencias de propiedades fisicoquímicas de Secuencias de péptidos de la Librería virtual.	24
Figura 7. Matrices de confusión del modelo basado en redes neuronales propuesto (my_ANN).....	25
Figura 8. Curva ROC del modelo basado en redes neuronales propuesto (my_ANN) ...	26
Figura 9. Matriz de confusión del modelo Máquinas de Vectores de Soporte propuesto (my_SVM).....	27
Figura 10. Curva ROC del modelo Máquinas de Vectores de Soporte (my_SVM).....	28
Figura 11. Matriz de confusión del modelo Árbol de Decisión propuesto (my_DTree).	29

RESUMEN

La resistencia a los antibióticos ha constituido un problema en la salud pública debido a la rápida propagación de cepas bacterianas resistentes, es por ello que se investigan metodologías que optimicen el proceso de la búsqueda e identificación de nuevos compuestos antimicrobianos que puedan combatirlos eficazmente. Los modelos basados en Inteligencia Artificial representan una metodología de optimización de búsqueda e identificación a gran escala de estos compuestos bioactivos. En este estudio se construyó una librería virtual (base de datos propia) de péptidos a partir de bases de datos públicas, filtrando las secuencias en función de la longitud y la concentración inhibitoria mínima (CIM) frente a *Escherichia coli* y *Staphylococcus aureus*. Las propiedades fisicoquímicas de los péptidos se determinaron utilizando una paquetería Python personalizada, *my_properties*, que mostró una elevada concordancia con herramientas ya establecidas. Se desarrollaron y evaluaron tres modelos de aprendizaje automático: máquina de vectores de apoyo (SVM), árbol de decisión y red neuronal profunda (DNN). Respecto a los modelos SVM y de árbol de decisión, el modelo DNN, denominado *my_ANN*, demostró un rendimiento superior tanto en precisión (84.91%), como en sensibilidad (84,71% y especificidad (85.12%), validándose mediante el uso de secuencias externas de péptidos, mostrando una capacidad de predicción robusta comparable a los métodos más avanzados. Se desarrolló una herramienta de recomendación que utiliza el modelo *my_ANN* para predecir la actividad antimicrobiana de nuevas secuencias peptídicas y generar variaciones con una bioactividad potencialmente mejorada. Los resultados sugieren que los modelos de aprendizaje automático, en particular el aprendizaje profundo, ofrecen un enfoque prometedor para el descubrimiento rápido y rentable de nuevos agentes antimicrobianos.

Palabras clave: Aprendizaje automatizado, máquinas de vectores de soporte, aprendizaje profundo, predicción, árboles de decisión.

ABSTRACT

The Antibiotic resistance has become a public health problem due to the rapid spread of resistant bacterial strains. Therefore, methodologies that optimize the search and identification process of new antimicrobial compounds that can effectively combat them are being investigated. Models based on Artificial Intelligence represent a methodology for optimizing the search and identification of these bioactive compounds on a large scale. In this study, a virtual library (proprietary database) of peptides was constructed from public databases, filtering the sequences according to length and minimum inhibitory concentration (MIC) against *Escherichia coli* and *Staphylococcus aureus*. The physicochemical properties of the peptides were determined using a custom Python package, *my_properties*, which showed high agreement with established tools. Three machine learning models were developed and evaluated: support vector machine (SVM), decision tree and deep neural network (DNN). Regarding the SVM and decision tree models, the DNN model, named *my_ANN*, demonstrated superior performance in both accuracy (84.91%), sensitivity (84.71% and specificity (85.12%), validated using external peptide sequences, showing robust prediction capability comparable to state-of-the-art methods. A recommendation tool was developed using *my_ANN* model to predict the antimicrobial activity of new peptide sequences and generate variations with potentially improved bioactivity. The results suggest that machine learning models, in particular deep learning, offer a promising approach for rapid and cost-effective discovery of new antimicrobial agents.

Keywords: *machine learning, support vector machines, deep learning, prediction, decision trees.*

1. CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

1.1 Antecedentes

1.1.1 Definiciones y clasificación

1.1.1.1 Péptidos

Los péptidos se definen como cadenas cortas de aminoácidos, que suelen contener de 12 a 50 residuos de aa, unidos entre sí por enlaces peptídicos entre el grupo carboxilo de un aminoácido y el grupo amino del aminoácido adyacente (Cabello Ruiz et al., 2016). La secuencia y el número de aminoácidos (aa) de un péptido determinan su conformación estructural, propiedades y actividades funcionales o actividad biológica. Se distinguen de las proteínas por su menor longitud (Doonan, 2002). Aunque puede llegar a haber cierto solapamiento de longitud (tamaño) entre las proteínas más pequeñas y los péptidos. Las diversas funciones de los péptidos incluyen la señalización (hormonas, neurotransmisores), la defensa (péptidos antimicrobianos), regulación enzimática y aplicaciones farmacéuticas (Mookherjee et al., 2020; Pablo Ramírez-Galvis, 2023). Las conformaciones de las cadenas peptídicas dependen de las interacciones entre los residuos de aminoácidos y el entorno del disolvente, aunque la comprensión de las relaciones estructura-función de los péptidos sigue siendo un área de estudio activa en todos los campos biomédicos (David L. Nelson et al., 2023; Mookherjee et al., 2020).

1.1.1.2 Péptidos antimicrobianos

Los péptidos antimicrobianos (AMPs) son una clase de péptidos cortos, que forman parte integral de las defensas inmunitarias innatas en organismo, ya que exhiben una potencia antimicrobiana de amplio espectro contra bacterias, hongos, virus y parásitos a través de mecanismos que implican la permeabilización y disrupción de las membranas

celulares microbianas (Benfield & Henriques, 2020; David L. Nelson et al., 2023). Su naturaleza catiónica y anfipática facilita las interacciones electrostáticas e hidrofóbicas con componentes de la membrana celular microbiana cargados negativamente, como los fosfolípidos (Strömstedt et al., 2010). Los AMPs, se distinguen de los antibióticos convencionales por su amplia gama de objetivos microbianos, y por su menor propensión a inducir resistencia microbiana, aunque esta gradualmente pueda ocurrir (Baindara et al., 2020). Además de la eliminación microbiana, algunos AMPs presentan actividad inmunomoduladora a través de propiedades quimiotácticas y de señalización que median en las respuestas inflamatorias del huésped (Téllez & Castaño, 2010). La función crucial de los AMPs en la inmunidad innata y sus prometedoras aplicaciones médicas han impulsado investigaciones hacia el desarrollo de variantes más estables y eficaces, con miras a crear novedosas terapias contra infecciones (da Cunha et al., 2017).

1.1.1.3 Mecanismos de acción

Aunque cada AMP puede funcionar a través de múltiples mecanismos, los efectos microbicidas se centran en la alteración de la integridad de la membrana de las células microbianas, al interactuar con estas (Seyfi et al., 2020a). Muchos AMP catiónicos se agrupan inicialmente en las superficies lipídicas aniónicas y luego se insertan en las membranas o se translocan a través de ellas de manera dependiente de la concentración (Benfield & Henriques, 2020). Las propiedades catiónicas y anfipáticas de los AMPs impulsan estas interacciones iniciales de unión electrostática e hidrofóbica. Una vez incrustados en las membranas microbianas, los AMPs pueden inducir la permeabilización y muerte celular a través de diversas vías disruptivas de membrana y no disruptivas de membrana (Benfield & Henriques, 2020). La formación de poros constituye una vía clásica de disrupción de membrana, manifestándose como poros en forma de barril, poros toroidales o poros lipídico-péptidos que comprometen la permeabilidad de la membrana, disipan gradientes electroquímicos, alteran el equilibrio osmótico celular y causan eflujo de contenidos intracelulares (David L. Nelson et al., 2023).

Alternativamente, ciertos AMPs llamados detergentes peptídicos pueden cubrir las superficies de las membranas a altas concentraciones y luego solubilizar las membranas de manera similar a un detergente (Seyfi et al., 2020a). Otros mecanismos implican la translocación a través de las membranas permitiendo interacciones con objetivos dentro de la célula bacteriana afectando funciones microbianas como la síntesis de la pared celular, síntesis de proteínas, síntesis de ácidos nucleicos y actividad enzimática, culminando en la inhibición del crecimiento o la apoptosis (Benfield & Henriques, 2020; Seyfi et al., 2020).

1.1.2 Propiedades Fisicoquímicas

Los AMPs representan una parte vital del sistema de la inmunidad innata, que exhibe una amplia gama de actividad en contra de patógenos (David L. Nelson et al., 2023). Su eficacia depende en gran manera de las propiedades fisicoquímicas de los residuos que componen la secuencia, en particular propiedades de hidrofobicidad, carga neta, momento hidrofóbico y punto isoeléctrico. Normalmente este tipo de compuestos bioactivos tienen pesos moleculares que suelen oscilar umbrales entre 1 a 5 kDa (David L. Nelson et al., 2023; Gagnon et al., 2017a). Sus momentos hidrofóbicos suelen oscilar valores entre 0.5 en escala de Eisenberg, lo que indica un fuerte carácter anfipático para la interacción con membranas bacterianas (Porto et al., 2022). La mayoría de los AMPs suelen ser catiónicos, una carga neta positiva entre rangos entre +2 a +9, principalmente debido al contenido de argininas y lisinas, que facilita las interacciones electrostáticas con membranas bacterianas cargadas negativamente (Gagnon et al., 2017b). Estas propiedades colectivamente contribuyen a la capacidad de estos compuestos bioactivos para alterar las membranas bacterianas.

1.1.2.1 Longitud de secuencia

La longitud de la cadena se refiere al número de residuos de aminoácidos que componen la secuencia primaria de un péptido (David L. Nelson et al., 2023). La longitud de la cadena puede influir en las propiedades anfipáticas, capacidad de formar estructuras

secundarias y profundidad de inserción en la membrana, modulando así la bioactividad y selectividad del péptido antimicrobiano (Doonan, 2002).

1.1.2.2 Peso molecular

Se define como la suma de las masas atómicas de todos los átomos que constituyen la secuencia y se calcula en base a las masas atómicas de cada residuo. Este parámetro influye en aspectos de comportamiento y función biológica, siendo los péptidos de más bajo peso molecular, los que exhiben una mayor permeabilidad de la membrana, lo que llega a facilitar una difusión pasiva a través de esta (Al Saiqali et al., 2018; Seyfi et al., 2020b).

1.1.2.3 Hidrofobicidad

La hidrofobicidad describe la naturaleza apolar a excluirse del contacto con el agua y la asociación con entornos hidrofóbicos (David L. Nelson et al., 2023). Está determinada por la cantidad de residuos no polares presentes en la secuencia, de esta manera mientras más residuos hidrofóbicos (no polares), mayor será la hidrofobicidad global de la secuencia (David L. Nelson et al., 2023).

Esta propiedad desempeña un papel fundamental en la actividad biológica del péptido, debido a que interactúa directamente con el entorno. Los péptidos más hidrofóbicos tienden a interactuar mejor con las membranas celulares, facilitando su inserción en las bicapas lipídicas, característica que llega a ser importante en el caso de péptidos con funciones antimicrobianas o de señalización (David L. Nelson et al., 2023).

La hidrofobicidad también influye en el plegamiento y la estructura tridimensional del péptido, los residuos hidrofóbicos tienden a agruparse en el interior de la estructura, estabilizando su conformación, permitiendo determinar en como el péptido interactúa con otras moléculas (Acquah et al., 2018).

1.1.2.4 Momento hidrofóbico

El momento hidrofóbico determina la periodicidad o segmentación de las regiones polares y no polares a lo largo de la secuencia peptídica (Doonan, 2002). Es una medida de anfipaticidad que indica la segregación de dominios hidrófobos e hidrófilos dentro de la secuencia, aplicada especialmente a péptidos alfa-helicoidales en los que una cara interactúa con membranas (Cabello Ruiz et al., 2016). Esto influye significativamente en la forma en como el péptido va a interactuar con las membranas biológicas. Péptidos con umbrales altos de momento hidrofóbico tienden a conformaciones en donde los residuos hidrofóbicos se orientan hacia la fase lipídica de la membrana, mientras que los hidrofílicos hacia el medio acuoso. Además, dependiendo de la distribución específica de los residuos hidrofóbicos e hidrofílicos, determina la preferencia del péptido a ciertos tipos de membrana, de esta manera influyendo en su selectividad y actividad biológica (Andreev et al., 2018).

1.1.2.5 Carga neta

La carga neta del péptido se refiere a la cationicidad global conferida por el número y los tipos de residuos de aminoácidos cargados, principalmente Arginina, lisina e histidina (Doonan, 2002). La cationicidad favorece la atracción electrostática inicial entre un AMP y los componentes cargados negativamente de las membranas bacterianas (Kumar et al., 2018). Sin embargo, una cationicidad demasiado alta, puede llegar a inducir toxicidad (Duque et al., 2023).

1.1.2.6 Punto isoeléctrico (pI)

El punto isoeléctrico (pI) es el valor del pH en el que el péptido no tiene carga neta o es eléctricamente neutro. De esta manera, a un pH específico la secuencia adopta un estado zwitterionico (una molécula que tiene tanto cargas positivas como negativas). El punto isoeléctrico, está determinado por los valores de pKa de los grupos ionizables,

tanto del grupo amino terminal como de los radicales de las cadenas laterales y el grupo carboxilo terminal (David L. Nelson et al., 2023).

1.1.3 Inteligencia artificial (IA)

Se define como la capacidad de las máquinas para emplear algoritmos, reconocer patrones, aprender de datos y utilizar este aprendizaje en procesos de toma de decisiones y tareas que tradicionalmente requerían de inteligencia humana (Rouhiainen, 2018).

Según Wang (2019) la IA abarca el diseño de sistemas computacionales inteligentes que perciben su entorno y toman acciones para maximizar su precisión en algún objetivo o tarea en específico.

Actualmente se aplican modelos basados en IA en aplicaciones epidemiológicas, modelado de propagación de enfermedades y factores de riesgo, microscopía, reconocimiento de patógenos a partir de imágenes, optimización en análisis de series temporales, predicción de la actividad biológica de compuestos bioactivos, predicción de estructuras de proteínas, diseño y edición de genes y aplicaciones en el diseño de fármacos (Bhardwaj et al., 2022).

Las técnicas de IA, como el aprendizaje automático (ML) y el aprendizaje profundo (DL), se utilizan ampliamente para analizar grandes conjuntos de datos. Esto es crucial en campos como la genómica, donde la IA ayuda a identificar patrones no optimizados con métodos de análisis tradicionales (Russell, 2018).

1.1.3.1 Aprendizaje Automático – Machine Learning (ML)

El aprendizaje automatizado (ML) se refiere al uso de algoritmos informáticos que tienen la capacidad de "aprender" progresivamente patrones complejos a partir de grandes

cantidades de datos, sin estar explícitamente programados para tareas específicas (Shinde & Shah, 2018).

Una de las características clave que distingue al aprendizaje automatizado de la programación tradicional basada en código, es que con esta última el desarrollador o programador define las reglas, procesos, y lógica que llegan a un resultado específico, mientras que en contraste, los modelos de ML se alimentan a partir de datos de entrada y salida, para dar resultados nuevos, pero el proceso intermedio al cual se llega a un resultado en específico no es programable, o es de difícil comprensión (Russell, 2018). Este proceso de desarrollo de modelos basado en datos, junto con la adaptación de algoritmos a lo largo del tiempo a medida que se dispone de más datos, son aspectos fundamentales del aprendizaje automático (Mahesh, 2018).

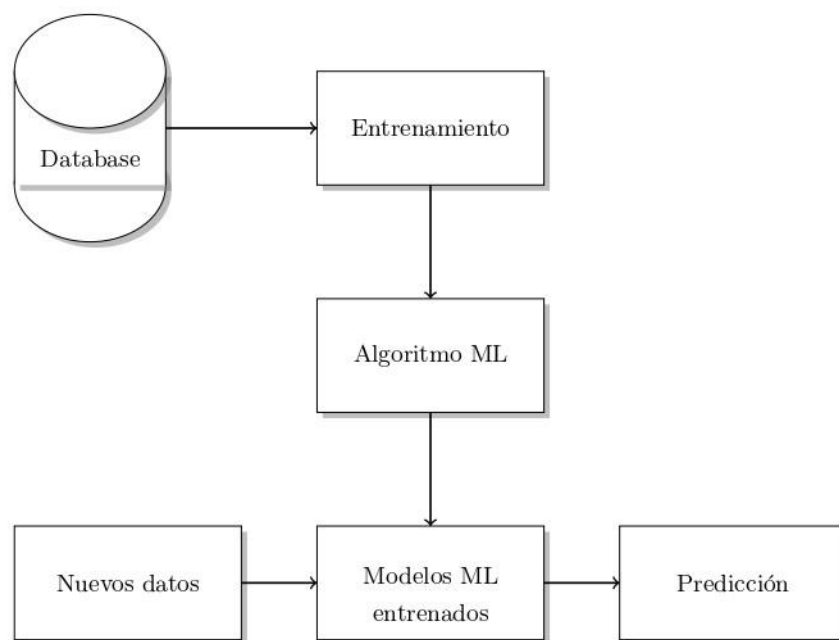


Figura 1. Flujo de trabajo básico de la aplicación de modelos predictivos de Aprendizaje Automatizado.

Realizado por: Hernandez, A, 2024.

El aprendizaje automático implica el desarrollo de representaciones de modelos matemáticos y arquitectura computacional para la minería de grandes conjuntos de

datos complejos con el fin de descubrir relaciones y patrones, permitiendo realizar inferencias sobre nuevas situaciones basadas en el aprendizaje de experiencias pasadas integradas en conjuntos de datos de entrenamiento. Los algoritmos utilizan diversas técnicas estadísticas, probabilísticas y de optimización para mejorar iterativamente el rendimiento del modelo en tareas específicas como la clasificación, agrupación, predicción, traducción y detección de anomalías (Bzdok et al., 2018; Russell, 2018).

1.1.3.1.1 Modelos tradicionales

1.1.3.1.2 Aprendizaje supervisado

El modelo Máquina de Vectores de Soporte (SVM, por sus siglas en inglés) es un modelo de aprendizaje supervisado, que busca construir un hiperplano para separar las observaciones de un conjunto de datos, de acuerdo con las etiquetas de sus clases. De esta manera, se puede usar a modo de clasificador binario, en donde para nuevos datos, las observaciones que pertenecen a una clase quedan por encima del hiperplano, y las que pertenecen a la otra clase quedan por debajo (Mammone et al., 2009).

1.1.3.1.3 Árboles de decisión

Los modelos de árboles de decisión son un modelo de aprendizaje supervisado, usado para tareas tanto de regresión como de clasificación, tiene una estructura de árbol jerárquico, que consta de un nodo raíz, ramas y nodos internos. En donde cada nodo corresponde a una prueba sobre un atributo (o variable de entrada), y cada rama representa el resultado de la prueba, que corresponde a una etiqueta que puede ser de clase o un valor continuo (Kotsiantis, 2013).

1.1.3.2 Aprendizaje profundo - Deep Learning (DL)

El Deep Learning o aprendizaje profundo, se define como el subconjunto de los modelos de Aprendizaje Automatizado que utiliza redes neuronales. Las redes neuronales, se

basan en modelos que se encuentran inspirados en el funcionamiento del cerebro humano, en específico en el comportamiento biológico de las neuronas (Gurney, 2018; Yuste, 2015a).

Según Tablada -Germán & Torres, (2009), las neuronas tienen como función principal transmitir impulsos nerviosos. Este proceso inicia en las dendritas de la neurona, continúa a lo largo del axón, y finaliza en las terminaciones axónicas, donde el impulso se transfiere a otra neurona a través de la sinapsis. De esta manera, computacionalmente estas neuronas son modeladas como unidades de procesamiento, en donde reciben múltiples datos de entrada (equivalentes a las señales que llegan a las dendritas), procesan esta información, producen una única salida y transmiten esta salida a otras neuronas (simulando la sinapsis) (Yuste, 2015b).

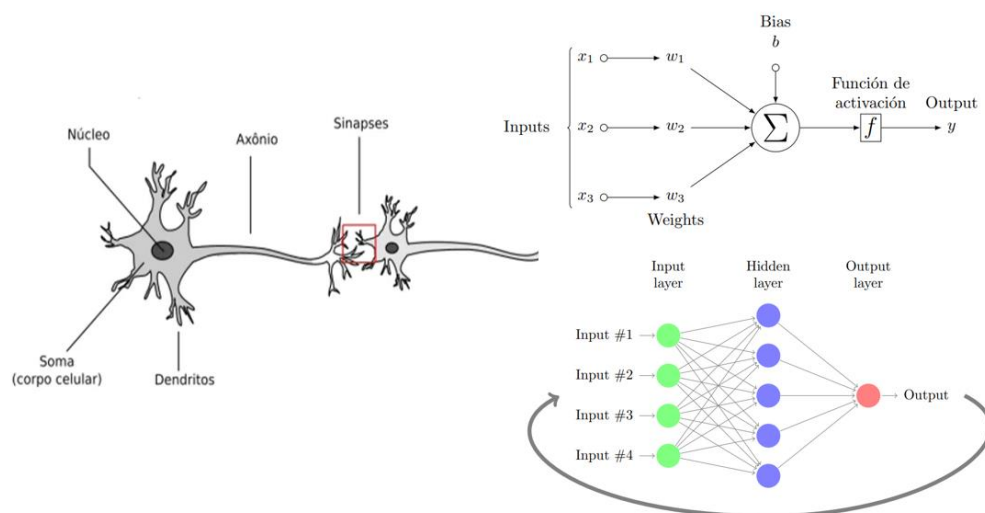


Figura 2. Esquematización de una Neurona y Red Neuronal Artificial. **Fuente:** (Prince, 2023; Tablada - Germán & Torres, 2009)

Realizado por: Hernández, A, 2024.

Por lo general, una red neuronal se llega a componer de la siguiente manera:

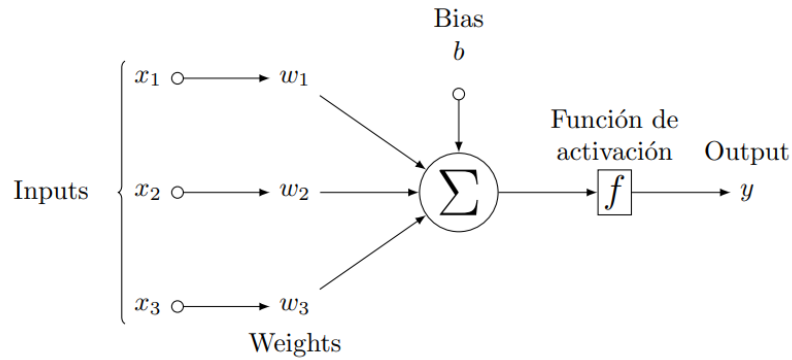


Figura 3. Arquitectura básica de una Red Neuronal Artificial.

Fuente: (Prince, 2023)

Realizado por: Hernández, A, 2024.

En donde:

- Inputs: representan los datos de entrada etiquetados como x_1, x_2, x_3
- Weights: representan los pesos, en donde cada entrada está asociada con un peso correspondiente w_1, w_2, w_3
- Σ : representa la suma ponderada, en donde la neurona suma todas las entradas multiplicadas por sus pesos respectivos.
- Bias: representa el sesgo, denotado como b , se añade este sesgo a la suma ponderada, para permitir un umbral de activación.
- f : representa una función de activación, después de la suma ponderada + el sesgo, esta función introduce no linealidad al proceso, permitiendo a la red aprender patrones más complejos.
- Output: Valor de salida

$$y = f\left(\sum (w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n) + b\right)$$

1.2 Planteamiento del problema

La resistencia a los antibióticos es una crisis mundial creciente, con la rápida propagación de cepas bacterianas resistentes a los fármacos antibacterianos existentes (Dña et al., 2021; Murray et al., 2022). Esta resistencia se produce cuando las bacterias desarrollan mecanismos de defensa para eludir los efectos de los antibióticos, lo que dificulta el tratamiento y aumenta el riesgo de desenlaces graves y potencialmente mortales (Frieri et al., 2017).

El impacto económico de la resistencia antimicrobiana también es alarmante. En los Estados Unidos, por ejemplo, las infecciones resistentes a los antibióticos generan un costo adicional estimado de \$20 mil millones en gastos médicos directos y \$35 mil millones en pérdidas de productividad anuales (Centers for Disease Control, 2019). Mientras que, en Europa, el costo relacionado con la resistencia a los antimicrobianos se estima en al menos €1.5 mil millones anuales en costos de atención médica y pérdida de productividad (Friedrich, 2019).

La resistencia a los antibióticos no solo incrementa la mortalidad y los costos económicos, sino que también prolonga las estancias hospitalarias y requiere tratamientos más intensivos y caros, incluyendo el uso de medicamentos de segunda o tercera línea que pueden ser menos efectivos (Centers for Disease Control, 2019). Esto pone una carga adicional sobre los sistemas de salud, especialmente en países en desarrollo donde los recursos son limitados.

Para hacer frente a este problema, es fundamental identificar nuevos compuestos antimicrobianos que puedan combatir eficazmente estas cepas resistentes. Los péptidos naturales representan una fuente prometedora debido a su amplia variedad estructural y capacidad para interactuar con diferentes sitios en las células bacterianas (la Fuente-Salcido et al., 2015). Sin embargo, la identificación de péptidos antimicrobianos presenta desafíos significativos. La abundancia de péptidos en la naturaleza, combinada con el

tiempo y los costos elevados de los análisis de laboratorio, complica el proceso de descubrimiento de compuestos efectivos.

1.3 Justificación de la Investigación

La crisis de resistencia a antibióticos exige soluciones urgentes e innovadoras para poder combatir contra diferentes cepas bacterianas de rápida evolución, y que representan una amenaza a la salud pública. La identificación de nuevos compuestos antimicrobianos es crucial para combatir las cepas bacterianas resistentes. Se busca proponer una metodología innovadora para la predicción de probabilidad de que compuestos biológicos tengan actividad biológica, utilizando técnicas de aprendizaje automático (ML, por sus siglas en inglés) y aprendizaje profundo (DL, por sus siglas en inglés) (Liu et al., 2020). La aplicación de estos modelos busca optimizar la identificación de péptidos con propiedades antimicrobianas, reduciendo el tiempo y los costos asociados con los métodos tradicionales de laboratorio.

Estos modelos pueden entrenarse a partir de grandes conjuntos de datos existentes que contengan información sobre péptidos antimicrobianos conocidos, sus estructuras moleculares y su actividad frente a distintas cepas bacterianas. Mediante el análisis de patrones y correlaciones en estos datos, los algoritmos de ML pueden predecir la probabilidad de que un nuevo péptido tenga propiedades antimicrobianas, lo que reduce el número de compuestos que deben evaluarse experimentalmente (Lee et al., 2023).

Además, la inclusión temprana de herramientas computacionales en el proceso de descubrimiento de nuevos compuestos antimicrobianos ayudaría a reducir costos y tiempos de desarrollo de nuevos fármacos, lo que podría contribuir a mejorar la disponibilidad de tratamientos innovadores para enfermedades infecciosas.

Pregunta de Investigación

¿Cuál es el nivel de precisión en la predicción de la bioactividad antimicrobiana de péptidos mediante el uso de modelos de aprendizaje automatizado?

Hipótesis

Los modelos de Aprendizaje Automatizado basados en redes neuronales tienen una mayor precisión en la predicción de la bioactividad antimicrobiana de péptidos que los modelos basados en algoritmos de Aprendizaje Automatizado tradicionales.

1.4 Objetivos de la investigación

1.4.1 Objetivo general

Evaluar la bioactividad antimicrobiana *in silico* de una librería virtual de péptidos aplicando modelos de Aprendizaje Automatizado para la identificación de nuevos agentes antimicrobianos.

1.4.2 Objetivos específicos

- Construir una librería virtual de péptidos a partir de bases de datos.
- Determinar propiedades fisicoquímicas de los péptidos de la librería virtual.
- Comparar modelos de Aprendizaje Automatizado (basados en redes neuronales y tradicionales) para la predicción de la bioactividad antimicrobiana de la librería virtual.
- Desarrollar un algoritmo de recomendación a partir del mejor modelo de Aprendizaje Automatizado.

2. CAPITULO II. MARCO METODOLÓGICO

2.1 Construcción de la librería (base de datos)

Para la construcción de la librería virtual (base de datos), se realizó una minería de secuencias de péptidos a partir de bases de datos públicas del NCBI (National Center for Biotechnology Information), PDB (Protein Data Bank) y Uniprot (Universal Protein Resource). Se hizo uso de la Biopython v1.83 una colección de módulos de acceso libre para el análisis de datos de biología molecular, haciendo uso del lenguaje Python v3.11.9, utilizando los módulos Bio.Entrez Bio.PDB y Bio.Uniprot.

En donde se usaron las palabras clave “peptide”, “antimicrobial peptide”, “antimicrobial”, “AMP”, “Bioactive peptide”.

Se aplicaron filtros para seleccionar secuencias cuya actividad biológica contra *Escherichia coli* y *Staphylococcus aureus*, hayan sido validadas experimentalmente mediante la Concentración Mínima Inhibitoria (CMI). Posteriormente, se seleccionaron únicamente aquellas secuencias que tengan una longitud en el umbral de 6 a 30 aa, y las que estén por fuera de este umbral se eliminaron. Esto debido a que se ha reportado que la actividad antimicrobiana de péptidos esta influenciada por la longitud, péptidos con secuencias cortas no llegan a tener buena actividad, mientras que péptidos con secuencias largas pueden incrementar el riesgo de toxicidad.

Adicionalmente, para evitar redundancias y sesgos en el conjunto de datos, se realizó un proceso de eliminación de secuencias repetidas. Con el objetivo de comprender mejor la actividad antimicrobiana, se estandarizaron los valores reportados para la actividad biológica (CMI) a [μM] *Ec* (1).

$$\mu M = \frac{c(\mu g \times mL^{-1})}{mw(Da)} \times 1000 \quad Ec (1)$$

Posteriormente, se etiquetó como AMPs a secuencias cuya actividad biológica, basada en CMI, fueron menores a 15 μM y como No-AMPs a secuencias con CMI mayores a 25 μM . Las secuencias que se encontraban en umbrales intermedios a estos fueron descartadas con el fin de tener una clara distinción entre AMPs y No-AMPs (Bournez et al., 2023; De Breij et al., 2018). Esto debido a que rangos bajos de CMI se considera como indicativo de buena actividad antimicrobiana, mientras que los de altas CMI no se consideran buenos candidatos para aplicaciones antimicrobianas, ya que podrían necesitar dosis demasiados altas para ser prácticas o seguras (Hancock & Sahl, 2006).

2.2 Determinación de propiedades fisicoquímicas.

Se desarrolló una paquetería en lenguaje Python v3.11.9, basada en definiciones para calcular propiedades fisicoquímicas de la Librería virtual, en donde recibe como entrada la secuencia del péptido, y devuelve diferentes descriptores especificados (Tabla 1).

Tabla 1. Métodos de cálculo de las propiedades fisicoquímicas.

Propiedad	Fórmula planteada	Referencia
Peso molecular (PM)	$PM(seq) = \sum_{i=1}^n m(x_i)$ $PM(seq) = m(x_1) + m(x_2) + \dots + m(x_n)$ En donde: $m(x)$: masas moleculares de cada aminoácido	(Voet et al., 2013)
Punto isoeléctrico (pI)	$pI(seq) = \frac{(pK_1 + \sum_{i=1}^n pK_2)}{n}$ En donde: pK_1: punto isoeléctrico del grupo amino terminal. $\sum_{i=1}^n pK_2$: sumatoria de los puntos isoeléctricos de los grupos R de las cadenas laterales. n: número total de residuos de la secuencia.	(Cargile et al., 2008)

Índice de inestabilidad (Ind)	$Ind(seq) = \sum_{i=1}^n f(x_i)$	(Guruprasad et al., 1990)
-------------------------------	----------------------------------	---------------------------

$$Ind(seq) = f(x_1) + f(x_2) + \dots + f(x_n)$$

En donde:

$f(x)$: factor de inestabilidad para cada aminoácido

Hidrofobicidad (Hi)	$Hi(seq) = \frac{\sum_{i=1}^n h(x_i)}{n}$	(Eisenberg, 1984)
---------------------	---	-------------------

$$Ind(seq) = h(x_1) + h(x_2) + \dots + h(x_n)$$

En donde:

$h(x)$: valores de hidrofobicidad para cada aminoácido, basados en la escala de Kyte & Doolittle en 1982.

Carga Neta (Z)	$Z(seq) = Z_{N-terminal} + \sum_{i=1}^n Z_{aa_i}$	(Cargile et al., 2008)
----------------	---	------------------------

En donde:

$Z_{N-terminal}$: Carga del grupo amino terminal.

Z_{aa} : Carga de cada residuo de aminoácido.

Momento hidrofóbico (μH)	$\mu H = \sqrt{\left\{ \left[\sum_{n=1}^n H_n \sin(\delta n) \right]^2 + \left[\sum_{n=1}^n H_n \cos(\delta n) \right]^2 \right\}}$	(Eisenberg, 1984)
---------------------------------	---	-------------------

En donde:

H_n : Hidrofobicidad residual según la escala propuesta por Eisenberg.

δ : distancia angular

n : número de aminoácidos de la secuencia

Realizado por: Hernández, A., 2024.

2.3 Construcción de los modelos

Se tomó el 80% de las secuencias de los péptidos de la librería virtual para entrenar los modelos a implementar, mientras que el 20% restante se usó como validación de estos.

Se entrenaron dos modelos, un modelo basado en Máquinas de Vectores de Soporte (SVM) y un modelo basado en Árboles de decisión para predecir la bioactividad antimicrobiana de los AMPs. Adicionalmente, se desarrolló una Red Neuronal Profunda (DNN), en donde esta tomó como entrada la secuencia del péptido, codificado como vectores numéricos, bajo la metodología one hot encoding. La salida fue un score o una puntuación de probabilidad para cada secuencia recibida (Figura 4)

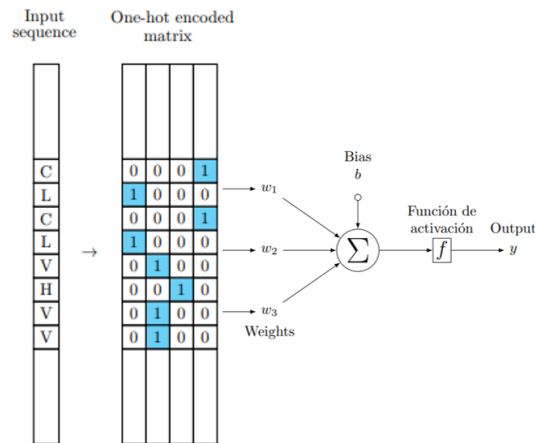


Figura 4.Arquitectura de la Red Neuronal utilizada.

Realizado por: Hernández, A.,2024.

2.4 Evaluación de métricas de rendimiento

Para evaluar el rendimiento de los modelos, se utilizaron las métricas de precisión, Recall, Exactitud, Especificidad, error cuadrático medio (MSE) y F1-score (James et al., 2013).

- **Precisión**

Mide la proporción de predicciones de verdaderos positivos identificados correctamente.

$$Precisión = \frac{TP}{(TP + FP)}$$

- Sensibilidad (Recall)

Mide proporción de verdaderos positivos detectados.

$$\text{Sensibilidad (TPR)} = \frac{TP}{(TP + FN)}$$

- Exactitud (Accuracy)

Mide la proporción de predicciones correctas, tanto positivas como negativas.

$$\text{Exactitud} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

- Especificidad

Mide la proporción de verdaderos negativos identificados correctamente.

$$\text{Especificidad} = \frac{TN}{(TN + FP)}$$

- Error cuadrático medio (MSE)

Mide el promedio de los errores al cuadrado entre los valores predichos y los valores reales.

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

- F1 – score

Es una media armónica entre precisión y sensibilidad, proporciona un valor que balancea estas métricas.

$$F1 - score = \frac{2}{\left(\frac{1}{\text{Precisión}} + \frac{1}{\text{Sensibilidad}} \right)}$$

- ROC (AUC)

La curva ROC es una representación gráfica de la sensibilidad frente a la tasa de falsos positivo.

El AUC proporciona una medida agregada del rendimiento en todos los umbrales de clasificación posibles.

En donde: $TP = True\ Positive (Verdadero\ Positivo)$

$TN = True\ Negative (Verdadero\ Negativo)$

$FP = False\ Positive (Falso\ Positivo)$

$FN = False\ Negative (Falso\ Negativo)$

2.5 Construcción del algoritmo de recomendación

Se seleccionó el modelo que mostró las mejores métricas de rendimiento para la predicción de probabilidad de la actividad antimicrobiana. En este sentido, el algoritmo de recomendación está constituida por la paquetería diseñada para el cálculo de propiedades fisicoquímicas, y el modelo que mostró las mejores métricas de rendimiento, así el algoritmo de recomendación recibirá como entrada una nueva secuencia peptídica, y dará como resultado el cálculo de sus respectivas propiedades fisicoquímicas ver Tabla 1, y un score de probabilidad propio del modelo implementado, que predice la probabilidad de tener actividad antimicrobiana. Mientras que, para la generación de las variaciones de secuencia, dada una secuencia de entrada, se realiza sustituciones en la secuencia por residuos catiónicos que confieren actividad antimicrobiana a AMPs, esto con el objetivo de aumentar la probabilidad de ser antimicrobiano, y que las variaciones salientes no lleguen a ser aleatorias.

3 CAPITULO III: PRESENTACIÓN DE DATOS Y RESULTADOS

El minado inicial resultó en un total de 82,400 secuencias distribuidas entre las bases de datos NCBI, Uniprot y PDB. Después de la aplicación de filtros de actividad biológica y longitud de secuencia, el conjunto de datos se redujo a 2,350 secuencias (Tabla 2).

Tabla 2. Proceso de filtrado de secuencias desde el minado inicial hasta el etiquetado por actividad.

Base de datos	Minado Inicial	Filtros de Actividad y longitud (de 6 a 30 aa)	AMPs (CMI < 15 μ M)	No-AMPs (CMI > 25 μ M)
NCBI	27.467	850	320	230
Uniprot	27.460	780	270	270
PDB	27.466	720	110	200
Total	82.393	2.350	700	700

Elaborado por: Hernández, A., 2024.

Posterior al etiquetado, se identificaron 701 AMPs (secuencias con CMI menor a 15 μ M) y 700 de No-AMPs (secuencias con CMI mayor a 25 μ M) dando un total de 1401 secuencias.

3.1 Exploración inicial de la Librería virtual (base de datos)

Se observaron qué aminoácidos como la Leucina (L), Alanina (A), Glicina (G) y Serina (S) son los más abundantes tanto para AMPs como para No-AMPs, pero su frecuencia se nota con mayor prevalencia en AMPs (Figura 5).

Asimismo, se puede evidenciar que la distribución de aminoácidos catiónicos como la Lisina (K), Arginina (R) e Histidina (H) muestran frecuencias mayores en AMPs, con excepción de la Histidina (H) la cual muestra una frecuencia relativamente menor en ambas categorías. Por otro lado, aminoácidos como la Cisteína (C), Ácido aspártico (D), Ácido Glutámico (E), Metionina (M), Prolina (P), Triptófano (W), Tirosina (Y) y otros

muestran frecuencias bajas tanto en AMPs como en No-AMPs, implicando su menor ocurrencia en las secuencias de péptidos analizadas.

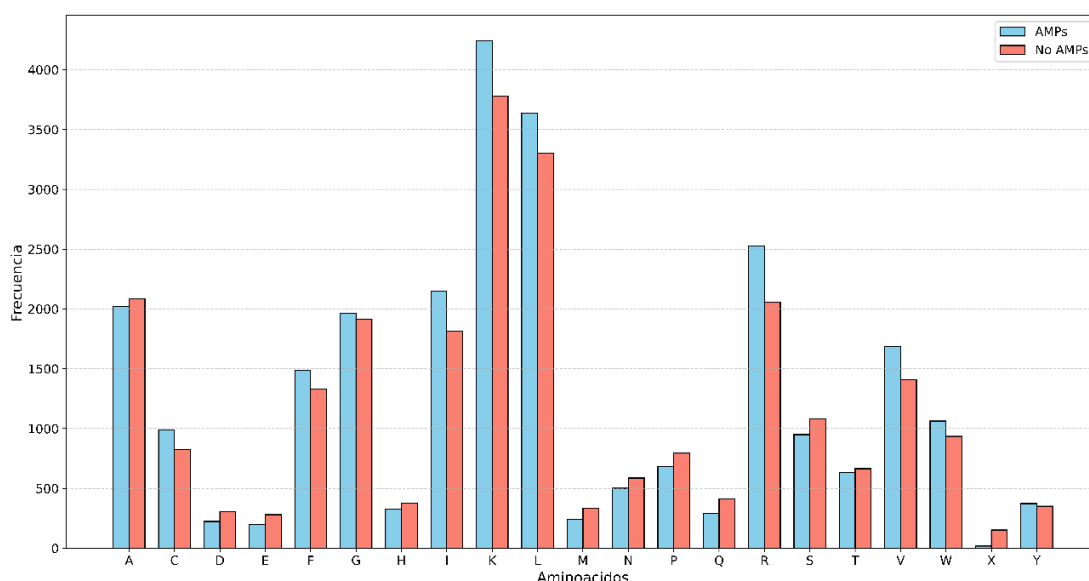


Figura 5. Distribución de aminoácidos de los péptidos en base a la CMI por actividad de la Librería virtual.

Realizado por: Hernández, A., 2024

3.2 Comparativa de la librería propuesta para el cálculo de propiedades fisicoquímicas.

Al comparar el algoritmo propuesto para el cálculo de propiedades fisicoquímicas, el cual está implementado en lenguaje Python “*my_properties*” con diferentes herramientas y algoritmos de referencia (Tabla 3), se observa un alto nivel de concordancia, en donde la mayoría de las diferencias son inferiores a 1%.

Se utilizó un péptido con actividad antimicrobiana conocida, en este caso el péptido Dermaseptina-S1 << ALWKTMLKKLGTMALHAGKAALGAAADTISQGTQ >>, un péptido aislado de las secreciones de la piel de la rana sudamericana *Phyllomedusa sauvagii*, el cual mediante ensayos *in vivo* ha demostrado ser eficaz contra un amplio espectro de microorganismos incluidos bacterias, hongos y protozoos. Los valores reportados de actividad biológica para Dermaseptina-S1 incluyen una concentración mínima

inhibitoria (MIC) de 5 µg/ml contra *Escherichia coli*, 15 µg/ml contra *Cryptococcus neoformans*, y 100 µg/ml contra *Aspergillus fumigatus* (Rudi et al., 2010).

Para el peso molecular, se observa una diferencia mínima del 0,424% entre los 3455,056 reportados por "my_properties" y los 3455,06 de Peptide Calculator. La carga neta presenta una discrepancia del 0,270%, con valores de 3,087 y 3,09 respectivamente reportados por APD3. El porcentaje de residuos hidrofóbicos es idéntico en ambos casos, fijándose en el 50%, al igual que la hidrofobicidad con un valor de 0,185 en ambas herramientas, siendo comparada con ProtParam de ExPASy.

El punto isoeléctrico muestra la mayor diferencia de la tabla con un 1,35%, donde "my_properties" reporta 10,537 frente a los 10,55 de Peptide Calculator. En cuanto al índice alifático, hay una variación del 0,294%, con "my_properties" indicando 92,353 y un algoritmo ProtParam de ExPASy 92,35. El índice de Boman también presenta una pequeña diferencia del 0,324%, siendo los valores 0,163 reportado por "my_properties" y 0,16 por APD3. Ambos métodos concuerdan perfectamente en el índice de inestabilidad con un valor de 1,85 reportado por ProtParam de ExPASy. Mientras que, el momento hidrofóbico muestra una variación del 1,10%, con "my_properties" reportando 0,509 y EMBOSS explorer reportando 0,52.

Tabla 3. Comparación de las propiedades fisicoquímicas entre el algoritmo propuesto (my_properties) con herramientas disponibles en la web.

Propiedades fisicoquímicas	my_properties	Otras calculadoras/algoritmos		%Diferencia
		Valor	Referencia	
Longitud	34	34	<u>APD3</u>	0%
Peso molecular	3455,056	3455,06	<u>Peptide Calculator</u>	0,424%
Carga neta	3,087	3,09	<u>APD3</u>	0,270%
% Residuos Hidrofóbicos	50	50	<u>APD3</u>	0%
Hidrofobicidad	0,185	0,185	<u>ProtParam - Expasy</u>	0 %
Punto isoeléctrico	10,537	10,55	<u>Peptide Calculator</u>	1,35%
Índice alifático	92,353	92,35	<u>ProtParam - Expasy</u>	0,294%
Índice de Boman	0,163	0,16	<u>APD3</u>	0,324%

Índice de inestabilidad	1,85	1,85	<u>ProtParam - Expasy</u>	0%
Momento hidrofóbico	0,509	0,52	<u>Hmoment – EMBOSS explorer</u>	1,10%

Elaborado por: Hernández, A., 2024.

3.3 Determinación de propiedades fisicoquímicas de la Librería virtual

Los valores de los parámetros o propiedades fisicoquímicas de la librería virtual (base de datos) (Figura 6) presentan rangos de hidrofobicidad entre -4,5 a 2,12 mostrando una ligera inclinación positiva, lo que indica un predominio de péptidos con hidrofobicidad baja a moderada. El momento hidrofóbico con rangos entre 0 y 1,233, muestra una distribución casi simétrica con un sutil sesgo a la derecha, lo que sugiere un equilibrio entre regiones hidrofóbicas e hidrofílicas. El índice alifático muestra rangos de 0 a 292,5 y aproximándose a una distribución normal, denotando un contenido alifático de moderado a alto en la mayoría de los péptidos. Por el contrario, el índice de inestabilidad con un rango de -67,41 a 509,95 muestra una marcada asimetría positiva, lo que revela que, si bien la mayoría de los péptidos son estables, algunos presentan una inestabilidad.

Asimismo, la carga neta con rangos de -8,143 a 17,982 muestra una asimetría positiva, lo que indica que predominan secuencias con cargas positivas. El punto isoeléctrico muestra rangos entre 3,545 y 13,646, presentando una distribución bimodal, lo que sugiere que las secuencias predominan con carga positiva. Finalmente, el índice de Boman presenta un rango de -3,127 a 14,92 mostrando una asimetría positiva, lo que indica un potencial de unión predominantemente de bajo a moderado.

Se puede evidenciar que la mayoría de las propiedades fisicoquímicas siguen una distribución Gaussiana no normal, corroborada por la prueba de normalidad Shapiro-wilk.

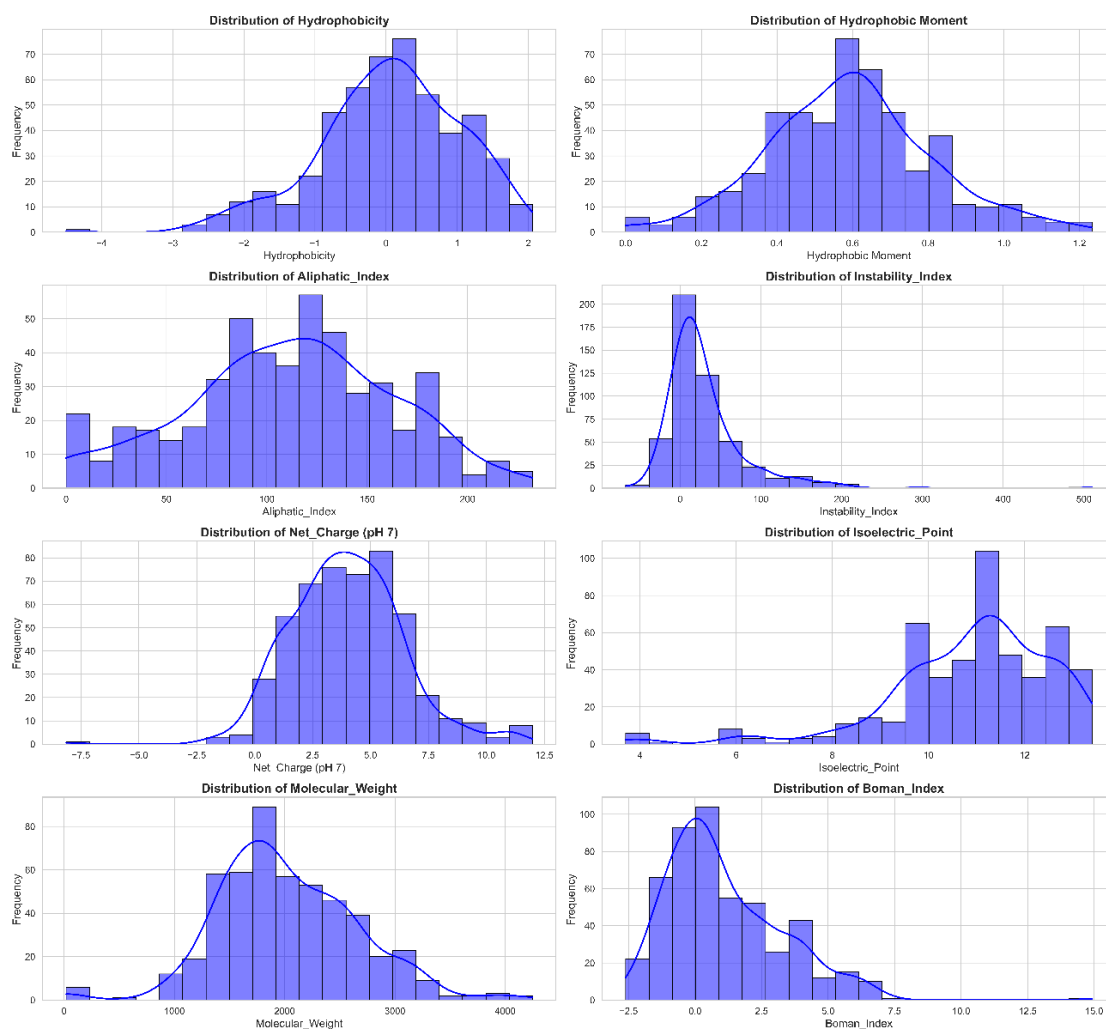


Figura 6. Distribución y tendencias de propiedades fisicoquímicas de Secuencias de péptidos de la Librería virtual.

Elaborado por: Hernández, A., 2024.

3.4 Entrenamiento de los modelos

- Modelo basado en Redes Neuronales propuesto (*my_ANN*)

En cuanto al modelo basado en redes neuronales propuesto, según datos obtenidos posterior al entrenamiento (Figura 7) se muestra las matrices de confusión obtenidas, tanto para el conjunto de datos que se usó para el entrenamiento como para el conjunto de validación.

En los datos de entrenamiento, el modelo identificó correctamente 935 casos, distribuidos en 450 verdaderos negativos (TN) y 485 verdaderos positivos (TP), mientras que clasificó erróneamente 145 casos, con 75 falsos positivos (FP) y 70 falsos negativos (FN). Para los datos de prueba, el modelo acertó en 215 casos, con 115 TN y 100 TP, y erró en 65 casos, con 30 FP y 35 FN.

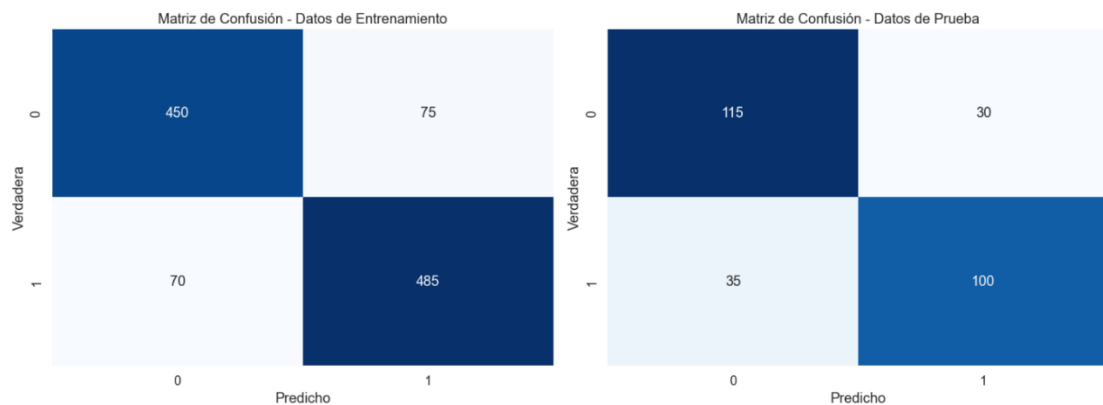


Figura 7. Matrices de confusión del modelo basado en redes neuronales propuesto (*my_ANN*).
Realizado por: Hernández, A. 2024.

En cuanto al rendimiento del modelo *my_ANN*, los resultados de entrenamiento muestran una precisión general de 0.86 86.68%, indicando un alto rendimiento inicial. La sensibilidad en este conjunto es del 87.39%, demostrando una alta eficacia en detectar secuencias realmente AMPs; mientras que, la especificidad alcanza un 85.71%, señalando una buena capacidad para identificar No-AMPs. El F1-Score para los datos de entrenamiento es de 87.03%, reflejando un equilibrio sólido entre precisión y exhaustividad. En contraste, los resultados de validación (o prueba) revelan una ligera disminución en el rendimiento, con una precisión del 76.79%, una sensibilidad del 74.07% y una especificidad del 79.31%. El F1-Score en este conjunto es de 75.41%, indicando un rendimiento algo menor pero aún robusto en datos no vistos previamente. Por otro lado, en las métricas generales del modelo (Tabla 4) se observa un rendimiento general del 84.91%, sensibilidad general de alrededor de 84,71% indicando que el modelo mantiene una alta tasa de detección de AMPs, mientras que la especificidad del 85.12%, demuestra su eficacia en identificar No-AMPs. El F1 Score general de 84.81%

sugiere un buen equilibrio entre precisión y exhaustividad en el rendimiento global del modelo.

Tabla 4. Resumen métricas de precisión para el modelo de Red Neuronal propuesto (*my_ANN*).

	Exactitud	Sensibilidad	Especificidad	F1-score	AUC	MSE
<i>my_ANN</i>	0.8668	0.8739	0.8571	0.8703	0.915	0.093
<i>Train</i>						
<i>my_ANN</i>	0.7679	0.7407	0.7931	0.7541	0.835	
<i>Val</i>						
<i>my_ANN</i>	0.8491	0.8471	0.8512	0.8481	0.890	0.093

Realizado por: Hernández, A., 2024.

Adicionalmente, el área bajo la curva ROC (AUC) general del modelo, se refleja con un umbral de 0.89 indicando una alta capacidad para distinguir entre secuencias de péptidos AMPs y No-AMPs, el pronunciado aumento de la curva (Figura 8) demuestra la elevada especificidad y sensibilidad del modelo propuesto.

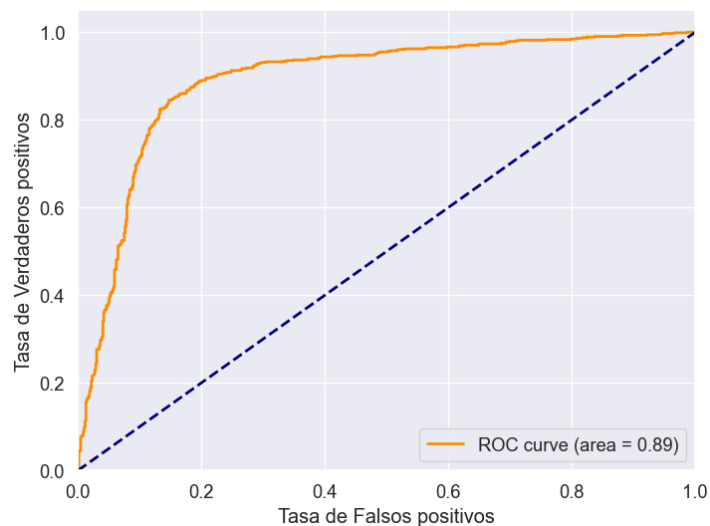


Figura 8. Curva ROC del modelo basado en redes neuronales propuesto (*my_ANN*)

Realizado por: Hernández, A., 2024.

- **Modelo basado en Máquinas de Vectores de Soporte propuesto (*my_SVM*)**

El modelo de Máquinas de Vectores de Soporte propuesto (*my_SVM*) muestra un rendimiento robusto o moderado con una precisión general de alrededor del 80,2% indicando una alta exactitud en sus predicciones generales (Figura 9). Además, muestra una sensibilidad del 85% para identificar correctamente secuencias de péptidos bioactivos. Por otro lado, la especificidad del 76,9%, aunque notable, sugiere un margen de mejora en la detección de péptidos No-AMPs.

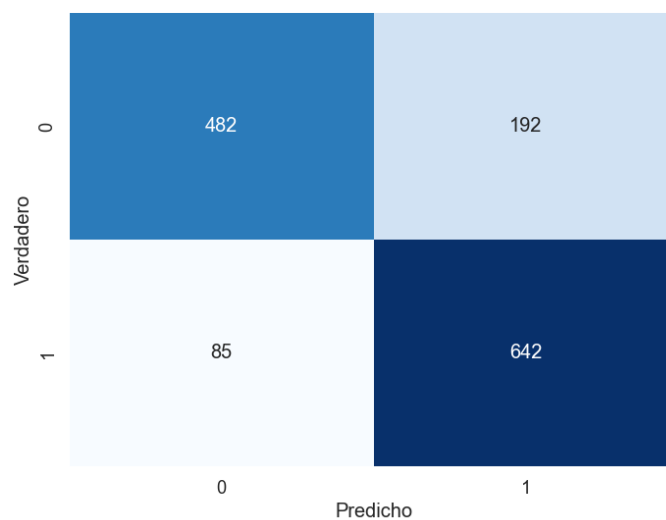


Figura 9. Matriz de confusión del modelo Máquinas de Vectores de Soporte propuesto (*my_SVM*).

Realizado por: Hernández, A. 2024.

El F1-score de 0,776 refleja un equilibrio adecuado entre precisión y exhaustividad, mientras que el área bajo la curva ROC (AUC) de 0,76 (Figura 10) señala una buena capacidad discriminativa del modelo, con potencial de optimización. El error cuadrático medio (MSE) de 0.086, relativamente bajo, indica que las predicciones del modelo se acercan considerablemente a los valores reales (Tabla 5).

Tabla 5 . Resumen métricas de precisión para el modelo Máquinas de Vectores de Soporte (*my_SVM*).

	Exactitud	Sensibilidad	Especificidad	F1-score	AUC	MSE
<i>my_SVM</i>	0.802	0.850	0.769	0.776	0.76	0.086

Realizado por: Hernández, A. 2024.

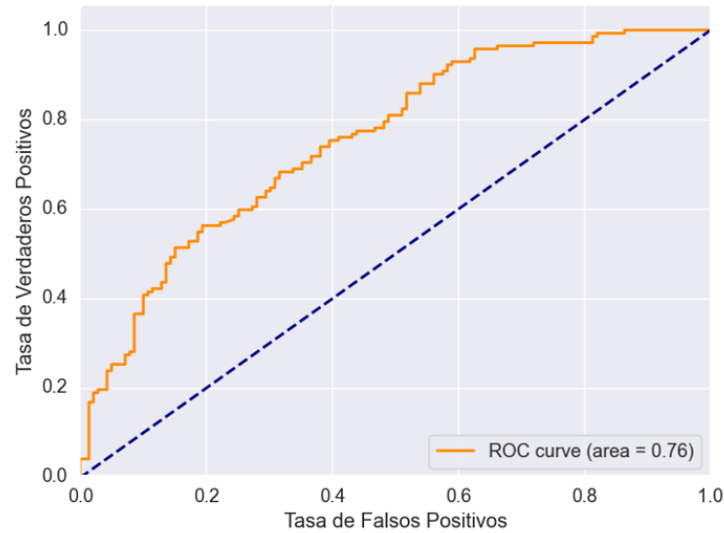


Figura 10. Curva ROC del modelo Máquinas de Vectores de Soporte (*my_SVM*)
Realizado por: Hernández, A., 2024.

El modelo alcanza una precisión global del 76,36%, lo que indica que clasifica correctamente más de tres cuartas partes de los péptidos analizados. Esta métrica refleja la capacidad general del árbol de decisión para realizar predicciones precisas. El modelo muestra una sensibilidad del 83,6% y una excelente capacidad para identificar correctamente los péptidos AMPs (Figura 11). Esto sugiere que el árbol de decisión es moderadamente eficaz en la detección de péptidos con actividad biológica.

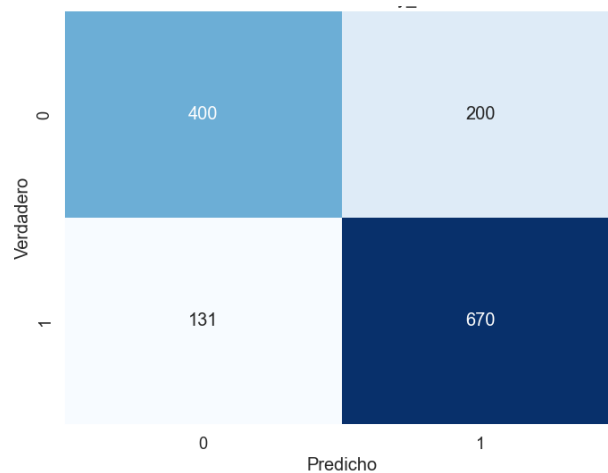


Figura 11. Matriz de confusión del modelo Árbol de Decisión propuesto (my_DTree).
Elaborador por: Hernández, A., 2024.

En cuanto a la Especificidad se reporta alrededor del 66.7% indica que el modelo identifica correctamente dos tercios de los péptidos no bioactivos (Tabla 4). Aunque es un resultado sólido, sugiere que hay margen de mejora en la clasificación de péptidos sin actividad biológica.

Tabla 4. Resumen métricas de precisión para el modelo de Máquinas de Vectores de Soporte (my_SVM)

	Exactitud	Sensibilidad	Especificidad	F1-score	AUC	MSE
<i>my_DTree</i>	0.7636	0.836	0.667	0.707	--	--

Realizado por: Hernández, A. 2024.

Comparación de los modelos

Tal y como se evidencia en la Tabla 5, de los modelos propuestos y trabajados, el modelo que presentó las mejores métricas de rendimiento para la identificación tanto de verdaderos positivos (AMPs), como en verdaderos negativos (No AMPs) fue *my_ANN* con una exactitud de 84,91% superior a los otros modelos propuestos *my_SVM* con 80,02% y *my_DTree* con 76,3%, respectivamente.

Tabla 6. Comparación de métricas de rendimiento de los modelos propuestos.

	Exactitud	Sensibilidad	Especificidad	F1-score	AUC	MSE
<i>my_DTree</i>	0.763	0.836	0.667	0.707	--	0.1091
<i>my_SVM</i>	0.802	0.850	0.749	0.776	0.76	0.086
<i>my_ANN</i>	0.8491	0.8471	0.8512	0.8481	0.890	0.093

Realizado por: Hernández, A., 2024.

3.5 Validación de los modelos con secuencias peptídicas externas a la Librería virtual

Se evaluó y comparó la capacidad de predicción del modelo con las mejores métricas calculadas, con otros modelos de última generación disponibles en la literatura (Tabla 7). Se usaron 8 secuencias de péptidos con actividad antimicrobiana conocida, que ya han sido validadas experimentalmente, las cuales son externas a la librería virtual (base de datos) interna.

El modelo propuesto *my_ANN*, logró predecir correctamente la probabilidad de cada una de las secuencias de los péptidos analizados, exhibiendo todas las predicciones con umbrales superiores al 0.5, marcador que sugiere que el péptido tiene probabilidad de tener actividad antimicrobiana.

Estos resultados son comparables con los demás modelos de referencia, si bien se puede observar que hay margen de mejora para igualar el rendimiento con modelos más nuevos como AMP Scanner vr.2 y AI4AMP, que a menudo muestran un rendimiento muy alto (puntuaciones por encima de 0,99), el rendimiento de *my_ANN* está más alineado con los modelos CAMPR4. Sin embargo, supera a iAMPred en algunos péptidos.

Tabla 7. Comparativa entre el modelo propuesto, con métodos de última generación, en la predicción de 8 AMPs conocidos.

Péptido	Secuencia	my_ANN	CAMPR4			AMP Scanner vr.2	AI4AMP	DBSSA	iAMPred
			SVM	RF	ANN				
Indolicidina	ILPWKWPWW PWRR	0,9464	0,86	0,83	0,51	1,00	1,00	AMP	0.89
Cecropina P1	SWLSKTAKKLE NSAKKRISGIA IAIQGGPR	0,7669	0,83	0,87	0,89	0,9924	0,9947	AMP	0.72
Magainina 2	GIGKFLHSAKKF GKAFFVGEIMNS	0,8860	0,95	0,99	0,94	0,9994	1,0000	AMP	0.99
Temporina A	FLPLIGRVLSGIL	0,7840	0,88	1	0,91	0,9995	0,9981	AMP	0.91
LL-37	LLGDFFRKSKEK IGKEFKRIVQRI KDFLRNLVPRT ES	0,7011	0,64	0,66	0,81	0,9994	1,0000	AMP	0.60
Defensina HNP-1	ACYCRIPACIAG ERRYGTCTYQG RLWAFCC	0,7656	0,94	0,94	0,95	1	0,9967	AMP	0.92
Bactenecin	RLCRIVVIRVCR	0,8532	0,95	0,91	0,98	0,9994	0,9985	AMP	0.94
Dermaseptina-S1	ALWKTMLKKL GTMALHAGKA ALGAAADTISQ GTQ	0,9464	0,89	0,61	0,30	0,999	0,9997	AMP	0.92

Realizado por: Hernández, A., 2024.

Por otro lado, en la evaluación de péptidos No-AMPs, el modelo propuesto my_ANN muestra un rendimiento muy consistente en las predicciones de baja probabilidad (Tabla 8), logrando predecir correctamente todas las secuencias que no tienen actividad antimicrobiana, todas por debajo del umbral de probabilidad de 0.5. Se observa que para el péptido Glucagon, my_ANN predice una probabilidad de 0.28, mientras que para Beta-Endorphin y GLP-1_Humano, las predicciones son 0.3004 y 0.3204 respectivamente, lo que indica una correcta identificación de estos como No-AMPs.

En comparación con los otros modelos, se muestra variabilidad en sus predicciones. Los modelos CAMPR4 (SVM, RF, ANN) presentan puntuaciones que varían considerablemente entre los diferentes péptidos. Siendo para el péptido VIP_Humano, las predicciones de los modelos SVM, RF y ANN son 0.3088, 0.14 y 0.17 respectivamente. Aunque en general predicen correctamente la baja probabilidad de tener actividad

antimicrobiana, esta variabilidad puede ser una desventaja en términos de consistencia. Los modelos más recientes como AMP Scanner vr.2 y AI4AMP tienden a dar puntuaciones muy bajas, como 0.0038 y 0.0009 respectivamente para el péptido Glucagón, lo cual es coherente con la naturaleza no antimicrobiana de estos péptidos. Sin embargo, en algunos casos, como en Neurotensin, AI4AMP muestra una puntuación relativamente alta de 0.61368287, lo que podría indicar un falso positivo.

Tabla 8. Comparativa entre el modelo propuesto, con métodos de última generación, en la predicción de 8 No-AMPs conocidos.

Péptido	Secuencia	my_ANN	CAMPR4			AMP			
			SVM	RF	ANN	Scanner vr.2	AI4AMP	DBSSAP	iAMPred
Glucagón	HSQGTFTSDYSKYL DSRRRAQDFVQWL MNT	0,28	0,05	0,03	0,07	0,0038	0,0009	Non-AMP	0,007
Beta-Endorphin	YGGFMTSEKSQTPL VTLFKNAIKNAYK GE	0,30 04	0,48	0,48	0,42	0,0066	0,018523932	Non-AMP	0,234
GLP-1_Humano	HAEGTFTSDVSSYL EGQAAKEFIWLV KGR	0,32 04	0,04	0,11	0,1	0,0035	0,013697088	Non-AMP	0,023
VIP_Humano	HSDAVFTDNYTRLR KQMAVKKYLNSILN	0,30 88	0,14	0,17	0,24	0,0523	0,025065035	Non-AMP	0,018
Galanin	GWTLSAGYLLGP HAIDNHRSFHDY GLT	0,26	0,3	0,4	0,39	0,0081	0,012679011	Non-AMP	0,111
Neurotensin	ELYENKPRRPYIL	0,28	0,27	0,04	0,1	0,4166	0,61368287	Non-AMP	0,487
Vasoactive_in testinal_pepti de	HSDAVFTDNYTRLR KQMAVKKYLNSILN	0,30 88	0,14	0,17	0,24	0,0523	0,025065035	Non-AMP	0,018
Glucagon- like_peptide- 1	HAEGTFTSDVSSYL EGQAAKEFIWLV KGR	0,32 04	0,04	0,11	0,1	0,0035	0,013697088	Non-AMP	0,023

Realizado por: Hernández, A., 2024.

Se debe tener en cuenta que la puntuación calculada, la cual está entre un umbral entre 0 y 1, predice la probabilidad de que el péptido penetre en las células bacterianas.

- Una puntuación superior a 0,5 sugiere que el péptido tiene probabilidad de penetrar en las células.
- Puntuaciones más cercanas a 1 indican mayor confianza en la penetración celular.
- Puntuaciones más cercanas a 0 sugieren que es muy poco probable que el péptido penetre en las células.

3.6 Validación de un algoritmo de recomendación

El algoritmo de recomendación quedó implementado con el modelo que mostró las mejores métricas de precisión, sensibilidad, especificidad, F1-score y AUC. En este caso, el modelo propuesto, basado en redes neuronales *my_ANN*, junto con la librería *my_properties*, y la implementación de la función de sustitución de aminoácidos en las secuencias, tiene 2 funciones implementadas.

La primera función, calcula las propiedades fisicoquímicas de una secuencia dada, y predice la probabilidad de que sea antimicrobiana. Se usó dos péptidos conocidos para validar la funcionalidad de un algoritmo propuesto, un péptido con actividad antimicrobiana conocida, y uno sin actividad antimicrobiana, siendo estos Dermaseptina-S1 y Glucagon-like_peptide-1, respectivamente.

Se logró calcular con precisión cada una de las propiedades fisicoquímicas y la predicción de probabilidad, de dos péptidos (AMPs y No-AMPs). El péptido Dermaseptina-S1 (AMP) presentó valores de 0,185 de hidrofobicidad, 0,509 de momento hidrofóbico, 92,353 de índice alifático, 1,847 de índice de inestabilidad, 3,030 de carga neta, 10,537 de punto isoeléctrico, 3455,056 de peso molecular, 0,163 de índice Boman y una predicción de 0,946. Mientras que el péptido Glucagon-like_peptide-1 (No-AMP) presentó valores de -0,230 de hidrofobicidad, 0,483 de momento hidrofóbico, 71,667 de índice alifático, 17,693 de índice de inestabilidad, -0,968 de carga neta, 5,617 de punto isoeléctrico,

3298,659 de peso molecular, 1,322 de índice Boman y una predicción de 0,320, lo cual coincide con las predicciones realizadas anteriormente (Tabla 9).

Tabla 9. Funcionalidad de un algoritmo de recomendación.

Secuencia	Hidrofobicidad	Momento Hidrofóbico	Índice Alifático	Índice Inestabilidad	Carga Neta	Punto Isoeléctrico	Peso Molecular	Índice Boman	Predicción
ALWKTMLK									
KLGTMALH									
AGKAALGA	0,185	0,509	92,353	1,847	3,030	10,537	3455,056	0,163	0,946
AADTISQGT									
Q									
HAEGTFTSD									
VSSYLEGQA									
AKEFIAWL	-0.230	0.483	71.667	17.693	-0.968	5.617	3298.659	1.322	0.320
KGR									

Realizado por: Hernández, A., 2024.

La segunda función, genera variaciones dada una secuencia, sustituyendo aminoácidos en diferentes posiciones a lo largo de la secuencia, junto con el cálculo de propiedades fisicoquímicas y la predicción de probabilidad para cada variación. Se uso una secuencia inicial como entrada, el péptido con actividad antimicrobiana conocida Dermaseptina-S1. Un algoritmo de recomendación, logro generar 55 variaciones de secuencia del péptido analizado ver Anexo 2, y logrando predecir con precisión la probabilidad de actividad antimicrobiana, en donde se mantuvo constante, sugiriendo que estas variaciones no afectaron significativamente la eficacia antimicrobiana predicha de las secuencias.

CAPITULO IV: INTERPRETACIÓN Y DISCUSIÓN

3.7 Distribución de aminoácidos de la Librería virtual

La alta prevalencia de aminoácidos como la Leucina, Alanina, Glicina y Serina en secuencias peptídicas activas (AMPs), concuerda con estudios previos de Pasupuleti et al., (2012) en donde mencionan que los AMPs suelen ser ricos en estos aminoácidos, particularmente en Leucina y Alanina, que contribuyen a la formación de estructuras anfipáticas cruciales para la actividad antimicrobiana. La Glicina, por su parte, proporciona flexibilidad estructural, mientras que la serina puede participar en interacciones polares con membranas bacterianas (Hancock & Sahl, 2006). La presencia de aminoácidos catiónicos como la Lisina y la Arginina en AMPs es mencionada por Doonan (2002), que menciona que la carga neta positiva de los AMPs, conferida principalmente por estos residuos, es crucial para la interacción electrostática inicial con membranas bacterianas cargadas negativamente. Asimismo, Duque et al. (2023), sugiere que la escasez de aminoácidos ácidos en AMPs ayuda a mantener la carga neta positiva.

3.8 Comparativa de la librería propuesta para el cálculo de propiedades fisicoquímicas

La alta concordancia observada entre la librería propuesta *my_properties* y las herramientas establecidas de cálculo de propiedades fisicoquímicas de secuencias de péptidos, respalda la viabilidad de desarrollar soluciones computacionales personalizadas en bioinformática. Esta concordancia, con diferencias generalmente inferiores al 1%, no sólo valida la precisión de la librería *my_properties* propuesta, sino que también subraya la robustez de los métodos subyacentes en el cálculo de estas propiedades críticas, alineándose con los estándares de precisión reportados por Porto et al., (2017) para herramientas similares.

La implementación de *my_properties* como una librería en lenguaje Python para el cálculo de propiedades fisicoquímicas de secuencias aminoacídicas ofrece ventajas significativas en términos de accesibilidad y flexibilidad. A diferencia de las herramientas web, un algoritmo local permite la integración seamless (fluidez) en pipelines de análisis más amplios y la posibilidad de procesar grandes conjuntos de datos de manera eficiente tal como lo demuestra Lee et al., (2023).

3.9 En cuanto a la distribución de frecuencias de propiedades fisicoquímicas de la Librería virtual

La ligera inclinación positiva en la hidrofobicidad de la librería virtual sugiere un predominio de péptidos con hidrofobicidad baja a moderada. Este hallazgo es consistente con estudios previos que han demostrado la importancia de un equilibrio hidrofóbico en péptidos bioactivos, particularmente en AMPs como los estudios de Gong et al. (2019), en donde mencionan que, un aumento en la hidrofobicidad puede mejorar la bioactividad de los AMPs, aunque con el riesgo de incrementar también la toxicidad hacia las células huésped.

Asimismo, la distribución casi simétrica del momento hidrofóbico (rango de 0 a 1,233) indica un balance entre regiones hidrofóbicas e hidrofílicas. Este equilibrio se alinea con estudios realizados por Chetia et al., (2024) , en donde mencionan que este equilibrio en AMPs puede ser un parámetro crítico para producir péptidos terapéuticos altamente selectivos contra bacterias.

La asimetría positiva en la carga neta y la distribución bimodal del punto isoeléctrico indican un predominio de secuencias con cargas positivas (Kiraga et al., 2007). La presencia de péptidos con diferentes cargas netas en la librería podría permitir la selección de candidatos para diversas aplicaciones.

El índice de Boman, sugiere un potencial de unión predominantemente de bajo a moderado, esto indica que la mayoría de los péptidos en la librería podrían tener una

interacción selectiva con sus dianas, minimizando potencialmente los efectos secundarios no deseados. Caso contrario, como mencionan Terziyski et al., (2023) a valores altos de Índice de Boman, es probable que el péptido tenga varias actividades biológicas.

3.10 Entrenamiento de los modelos

La alta precisión general (84.91 %) y el AUC (0.89) del modelo propuesto *my_ANN* superan a los modelos tradicionales, como Máquina de Vectores de Soporte (*my_SVM*) y Árbol de Decisión (*my_DTree*). Este desempeño es consistente con los hallazgos de estudios recientes como Li et al., (2019), quienes abogan por el uso de modelos de aprendizaje profundo en bioinformática debido a su capacidad mejorada para reconocer patrones complejos.

El desarrollo de modelos computacionales para la predicción de la probabilidad de que un péptido tenga, actividad biológica representa un área de investigación crucial en la lucha contra la resistencia antimicrobiana. Los resultados obtenidos con el modelo propuesto *my_ANN* aportan nuevas perspectivas a este campo, al tiempo que plantean interrogantes sobre los enfoques actuales en el diseño de estos modelos predictivos.

3.11 Validación de los modelos con secuencias peptídicas externas a la Librería virtual.

La validación de modelos de predicción de péptidos antimicrobianos (AMPs) utilizando secuencias externas a la base de datos de entrenamiento es un paso crucial para evaluar la robustez y aplicabilidad real de estos modelos. La evaluación del modelo *my_ANN* con secuencias de AMPs y No-AMPs conocidos y validados experimentalmente, proporciona insights valiosos sobre su rendimiento y generalización.

Sin embargo, se tiene que considerar lo señalado en investigaciones de Yang et al., (2019), en donde argumentan que la capacidad de un modelo para predecir

correctamente compuestos conocidos (como los datos de entrenamiento) es un paso clave, pero no suficiente para garantizar su utilidad en el descubrimiento de nuevos compuestos o desconocidos. Esta observación subraya la necesidad de considerar estos resultados como un indicador positivo, más no definitivo, del rendimiento del modelo.

El rendimiento de *my_ANN*, aunque comparable con modelos establecidos como CAMPR4 (Gawde et al., 2022), no alcanza los niveles de precisión extremadamente altos reportados por modelos más recientes como AMP Scanner vr.2 y AI4AMP (Lin et al., 2021; Veltri et al., 2018). Esta discrepancia suscita una reflexión crítica sobre el equilibrio entre precisión y capacidad de generalización en los modelos de predicción de AMPs. Sin embargo, se debe tener en cuenta también, que alcanzar niveles altos de precisión no necesariamente implica un buen rendimiento de generalización, sino que puede ser un indicador de sobreajuste (Srivastava et al., 2014).

3.12 Validación de un algoritmo de recomendación

La integración de un algoritmo de recomendación presenta una alternativa óptima que no solo permite la predicción precisa de la actividad antimicrobiana, sino que también facilita la generación y evaluación de múltiples variantes de la secuencia. Las variaciones de secuencia se generaron mediante la sustitución estratégica de aminoácidos por triptófano (W), arginina (R) y lisina (K), con el objetivo de aumentar la probabilidad de que las secuencias resultantes tuvieran actividad antimicrobiana. Este enfoque se alinea con investigaciones previas que han demostrado la importancia de estos residuos en la función de los AMPs. Estudios realizados por Fernández García, (2016) observaron que la introducción de triptófano en posiciones específicas mejoraba significativamente la actividad antimicrobiana de péptidos derivados de la lactoferricina bovina.

Asimismo, Jiang et al., (2008) señalan que el reemplazo estratégico de residuos por lisina (K) en la secuencia peptídica incrementa la carga neta positiva, lo cual puede mejorar la interacción con las membranas bacterianas cargadas negativamente, y aumentar la actividad antimicrobiana. Sin embargo, este aumento en la carga positiva debe ser

cuidadosamente balanceado, ya que un exceso puede llevar a una toxicidad hemolítica no deseada.

CAPITULO V: CONCLUSIONES Y RECOMENDACIONES

3.13 Conclusiones

Las secuencias de péptidos minadas de bases de datos públicas e implementadas en la librería virtual se construyeron de manera proporcional para ambas categorías (tanto AMPs como No-AMPs) para evitar sesgos en el entrenamiento o sobreajustes en la evaluación con datos externos. En total, se incluyeron 1401 secuencias de péptidos, distribuidas equitativamente entre AMPs y No-AMPs.

Se logró desarrollar “*my_properties*”, una librería implementada en lenguaje Python para el cálculo de propiedades fisicoquímicas de secuencias aminoacídicas, mostrando una concordancia significativa con herramientas ya establecidas, facilitó la caracterización de la Librería virtual propuesta.

A partir de los análisis de resultados en la comparación de las métricas de rendimiento de los modelos propuestos tanto para secuencias de la librería de AMPs propuestas como para secuencias peptídicas externas, se puede concluir que el modelo basado en redes neuronales (*my_ANN*) es el más adecuado para predecir la probabilidad de actividad antimicrobiana, con una precisión del 84.91% y un AUC de 0.89. Este modelo superó a los modelos tradicionales como el de Máquina de Vectores de Soporte (*my_SVM*) y Árbol de Decisión (*my_DTree*).

En cuanto a un algoritmo de recomendación, esta representa un enfoque alternativo personalizado, para la predicción de probabilidad de AMPs, así como la generación de péptidos análogos basado en una secuencia dada. La capacidad de un algoritmo para predecir con precisión la probabilidad de que una secuencia tenga actividad antimicrobiana y generar variaciones efectivas constituye un avance significativo en la automatización y aceleración del proceso de diseño de AMPs.

3.14 Recomendaciones

Se recomienda que, si bien la capacidad del modelo DNN (*my_ANN*) para predecir correctamente la actividad antimicrobiana es comparable con otras herramientas predictoras, se necesita complementar con estudios adicionales a nivel de cribado virtual, cribado molecular y modelado estructural.

Se sugiere que los resultados obtenidos mediante el modelo predictivo propuesto se consideren como un punto de partida para investigaciones más profundas. Ya que es recomendable realizar estudios experimentales *in vitro* e *in vivo* para validar la actividad antimicrobiana de los péptidos seleccionados por el modelo propuesto "*my_ANN*".

REFERENCIAS BIBLIOGRÁFICAS

- Acquah, C., Stefano, E. Di, & Udenigwe, C. C. (2018). Role of hydrophobicity in food peptide functionality and bioactivity. *Journal of Food Bioactives*, 4, 88-98-88-98. <https://doi.org/10.31665/JFB.2018.4164>
- Al Saiqali, M., Tangutur, A. D., Banoth, C., & Bhukya, B. (2018). Antimicrobial and anticancer potential of low molecular weight polypeptides extracted and characterized from leaves of *Azadirachta indica*. *International Journal of Biological Macromolecules*, 114, 906-921. <https://doi.org/10.1016/J.IJBIOMAC.2018.03.169>
- Andreev, K., Martynowycz, M. W., Huang, M. L., Kuzmenko, I., Bu, W., Kirshenbaum, K., & Gidalevitz, D. (2018). Hydrophobic interactions modulate antimicrobial peptoid selectivity towards anionic lipid membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1860(6), 1414-1423. <https://doi.org/10.1016/J.BBAMEM.2018.03.021>
- Baindara, P., Ghosh, A. K., & Mandal, S. M. (2020). Coevolution of Resistance Against Antimicrobial Peptides. *Https://Home.Liebertpub.Com/Mdr*, 26(8), 880-899. <https://doi.org/10.1089/MDR.2019.0291>
- Benfield, A. H., & Henriques, S. T. (2020). Mode-of-Action of Antimicrobial Peptides: Membrane Disruption vs. Intracellular Mechanisms. *Frontiers in Medical Technology*, 2, 610997. <https://doi.org/10.3389/FMEDT.2020.610997/BIBTEX>
- Bhardwaj, A., Kishore, S., & Pandey, D. K. (2022). Artificial Intelligence in Biological Sciences. *Life 2022, Vol. 12, Page 1430*, 12(9), 1430. <https://doi.org/10.3390/LIFE12091430>
- Bournez, C., Riool, M., de Boer, L., Cordfunke, R. A., de Best, L., van Leeuwen, R., Drijfhout, J. W., Zaat, S. A. J., & van Westen, G. J. P. (2023). CalcAMP: A New Machine Learning Model for the Accurate Prediction of Antimicrobial Activity of Peptides. *Antibiotics*, 12(4), 725. <https://doi.org/10.3390/ANTIBIOTICS12040725/S1>
- Bzdok, D., Altman, N., & Krzywinski, M. (2018). THIS MONTH Statistics versus machine learning. *Nature Publishing Group*, 15. <https://doi.org/10.1038/nmeth.4642>
- Cabello Ruiz, E. D., Núñez González, M. A., & Torres de la Cruz, V. M. (2016). Actividad biológica de proteínas y péptidos. *Investigación En Plantas de Importancia Médica*, 2016, ISBN 9788494467370, Págs. 313-350, May 2017, 313-350. <https://dialnet.unirioja.es/servlet/articulo?codigo=8595959&info=resumen&idioma=SPA>
- Centers for Disease Control, U. (2019). *Antibiotic Resistance Threats in the United States, 2019*. <https://doi.org/10.15620/cdc:82532>
- Chetia, M., Sarkar, T., Yadav, M., Dey, C., Sundaravadivelu, P. K., Thummer, R. P., & Chatterjee, S. (2024). Tuning of hydrophobic-hydrophilic balance for the development of a salt-tolerant and protease-resistant lipopeptide AMP. *New Journal of Chemistry*, 48(6), 2764-2777.

- da Cunha, N. B., Cobacho, N. B., Viana, J. F. C., Lima, L. A., Sampaio, K. B. O., Dohms, S. S. M., Ferreira, A. C. R., de la Fuente-Núñez, C., Costa, F. F., Franco, O. L., & Dias, S. C. (2017). The next generation of antimicrobial peptides (AMPs) as molecular therapeutic tools for the treatment of diseases with social and economic impacts. *Drug Discovery Today*, 22(2), 234–248. <https://doi.org/10.1016/J.DRUDIS.2016.10.017>
- David L. Nelson, Michael M. Cox, & Aaron A. Hoskins. (2023). *Lehninger Principles of Biochemistry 8th edition by David L. Nelson, Michael M. Cox, Aaron A. Hoskins*.
- De Breij, A., Riool, M., Cordfunke, R. A., Malanovic, N., De Boer, L., Koning, R. I., Ravensbergen, E., Franken, M., Van Der Heijde, T., Boekema, B. K., Kwakman, P. H. S., Kamp, N., El Ghalbzouri, A., Lohner, K., Zaat, S. A. J., Drijfhout, J. W., & Nibbering, P. H. (2018). The antimicrobial peptide SAAP-148 combats drug-resistant bacteria and biofilms. *Science Translational Medicine*, 10(423). https://doi.org/10.1126/SCITRANSLMED.AAN4044/SUPPL_FILE/AAN4044_TABLE_S5.ZIP
- Dña, :, Ruiz De Galarreta Beristain, M., Jesús, D., Alain, D., & Santander, O. (2021). *La resistencia a los antibióticos, un problema de salud global: las infecciones por microbacterias ultrarresistentes*. <https://repositorio.unican.es/xmlui/handle/10902/23549>
- Doonan, S. (2002). *Péptidos y proteínas (Vol. 15)*. Real Sociedad de Química. (Vol. 15). Royal Society of chemistry.
- Duque, H. M., Rodrigues, G., Santos, L. S., & Franco, O. L. (2023). The biological role of charge distribution in linear antimicrobial peptides. *Expert Opinion on Drug Discovery*, 18(3), 287–302. <https://doi.org/10.1080/17460441.2023.2173736>
- Fernández García, M. (2016). *Estrategias para el desarrollo de un nuevo producto cárnico fermentado de cerdo ibérico con compuestos nitrogenados bioactivos mediante la inoculación de microorganismos seleccionados y la enzima EPg222*.
- Friedrich, A. W. (2019). Control of hospital acquired infections and antimicrobial resistance in Europe: the way to go. *Wiener Medizinische Wochenschrift*, 169(1), 25–30. <https://doi.org/10.1007/S10354-018-0676-5/METRICS>
- Frieri, M., Kumar, K., & Boutin, A. (2017). Antibiotic resistance. *Journal of Infection and Public Health*, 10(4), 369–378.
- Gagnon, M. C., Strandberg, E., Grau-Campistany, A., Wadhwani, P., Reichert, J., Bürck, J., Rabanal, F., Auger, M., Paquin, J. F., & Ulrich, A. S. (2017a). Influence of the Length and Charge on the Activity of α -Helical Amphipathic Antimicrobial Peptides. *Biochemistry*, 56(11), 1680–1695. https://doi.org/10.1021/ACS.BIOCHEM.6B01071/SUPPL_FILE/BI6B01071_SI_001.PDF
- Gagnon, M. C., Strandberg, E., Grau-Campistany, A., Wadhwani, P., Reichert, J., Bürck, J., Rabanal, F., Auger, M., Paquin, J. F., & Ulrich, A. S. (2017b). Influence of the Length and Charge on the Activity of α -Helical Amphipathic Antimicrobial Peptides. *Biochemistry*, 56(11), 1680–1695. https://doi.org/10.1021/ACS.BIOCHEM.6B01071/SUPPL_FILE/BI6B01071_SI_001.PDF

- Gawde, U., Chakraborty, S., Hanif Waghu, F., Shankar Barai, R., Khanderkar, A., Indraguru, R., Shirsat, T., & Idicula-Thomas, S. (2022). CAMP R4 : a database of natural and synthetic antimicrobial peptides. *Nucleic Acids Research*, 51. <https://doi.org/10.1093/nar/gkac933>
- Gong, H., Zhang, J., Hu, X., Li, Z., Fa, K., Liu, H., Waigh, T. A., McBain, A., & Lu, J. R. (2019). Hydrophobic Control of the Bioactivity and Cytotoxicity of de Novo-Designed Antimicrobial Peptides. *ACS Applied Materials and Interfaces*, 11(38), 34609–34620. https://doi.org/10.1021/ACSAMI.9B10028/SUPPL_FILE/AM9B10028_SI_001.PDF
- Gurney, K. (2018). *An Introduction to Neural Networks*. <https://doi.org/10.1201/9781315273570>
- Hancock, R. E. W., & Sahl, H. G. (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nature Biotechnology* 2006 24:12, 24(12), 1551–1557. <https://doi.org/10.1038/nbt1267>
- Jiang, Z., Vasil, A. I., Hale, J. D., Hancock, R. E. W., Vasil, M. L., & Hodges, R. S. (2008). Effects of net charge and the number of positively charged residues on the biological activity of amphipathic α -helical cationic antimicrobial peptides. *Peptide Science*, 90(3), 369–383. <https://doi.org/10.1002/BIP.20911>
- Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczyk, M., Biecek, P., Polak, N., Smolarczyk, K., Dudek, M. R., & Cebrat, S. (2007). The relationships between the isoelectric point and: Length of proteins, taxonomy and ecology of organisms. *BMC Genomics*, 8(1), 1–16. <https://doi.org/10.1186/1471-2164-8-163/FIGURES/8>
- Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261–283. <https://doi.org/10.1007/S10462-011-9272-4/METRICS>
- Kumar, P., Kizhakkedathu, J. N., & Straus, S. K. (2018). Antimicrobial peptides: diversity, mechanism of action and strategies to improve the activity and biocompatibility in vivo. *Biomolecules*, 8(1), 4.
- la Fuente-Salcido, D., Margarita, N., Villarreal-Prieto, J. M., Díaz León, M. Á., & García Pérez, A. P. (2015). Evaluación de la actividad de los agentes antimicrobianos ante el desafío de la resistencia bacteriana. *Revista Mexicana de Ciencias Farmacéuticas*, 46(2), 7–16.
- Lee, H., Lee, S., Lee, I., & Nam, H. (2023). AMP-BERT: Prediction of antimicrobial peptide function based on a BERT model. *Protein Science: A Publication of the Protein Society*, 32(1). <https://doi.org/10.1002/PRO.4529>
- Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y., & Gao, X. (2019). Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods*, 166, 4–21. <https://doi.org/10.1016/J.YMETH.2019.04.008>
- Lin, T.-T., Yang, L.-Y., Lu, I.-H., Cheng, W.-C., Hsu, Z.-R., Chen, S.-H., & Lin, C.-Y. (2021). AI4AMP: an antimicrobial Peptide predictor using physicochemical property-Based encoding method and deep learning. *Msystems*, 6(6), e00299-21.
- Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research*. <https://doi.org/10.21275/ART20203995>
- Mammone, A., Turchi, M., & Cristianini, N. (2009). Support vector machines. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(3), 283–289. <https://doi.org/10.1002/WICS.49>

- Mookherjee, N., Anderson, M. A., Haagsman, H. P., & Davidson, D. J. (2020). Antimicrobial host defence peptides: functions and clinical potential. *Nature Reviews Drug Discovery* 2020 19:5, 19(5), 311–332. <https://doi.org/10.1038/s41573-019-0058-8>
- Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., ... Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Pablo Ramírez-Galvis, J. (2023). Péptidos los mensajeros entre sistemas biológicos. *Revitalia*. <https://doi.org/10.13140/RG.2.2.12772.55683>
- Pasupuleti, M., Schmidtchen, A., & Malmsten, M. (2012). Antimicrobial peptides: key components of the innate immune system. *Critical Reviews in Biotechnology*, 32(2), 143–171. <https://doi.org/10.3109/07388551.2011.594423>
- Porto, W. F., Ferreira, K. C. V., Ribeiro, S. M., & Franco, O. L. (2022). Sense the moment: A highly sensitive antimicrobial activity predictor based on hydrophobic moment. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1866(3), 130070.
- Porto, W. F., Pires, A. S., & Franco, O. L. (2017). Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnology Advances*, 35(3), 337–349. <https://doi.org/10.1016/J.BIOTECHADV.2017.02.001>
- Prince, S. J. D. (2023). *Understanding Deep Learning*. MIT press.
- Rouhiainen, L. (2018). Inteligencia artificial. *Madrid: Alienta Editorial*, 20–21.
- Rudi, J. M., Müller, D. M., Siano, A., Simonetta, A. C., & Tonarelli, G. G. (2010). Péptido antimicrobiano quimérico de dermaseptina-s1 y tigerinina-1: Estructura secundaria y selectividad hacia membranas. *FABICIB: Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional Del Litoral*, 14.
- Russell, R. (2018). *Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python*. www.TechnicalBooksPDF.com
- Seyfi, R., Kahaki, F. A., Ebrahimi, T., Montazersaheb, S., Eyvazi, S., Babaeipour, V., & Tarhriz, V. (2020a). Antimicrobial Peptides (AMPs): Roles, Functions and Mechanism of Action. *International Journal of Peptide Research and Therapeutics*, 26(3), 1451–1463. <https://doi.org/10.1007/S10989-019-09946-9/METRICS>
- Seyfi, R., Kahaki, F. A., Ebrahimi, T., Montazersaheb, S., Eyvazi, S., Babaeipour, V., & Tarhriz, V. (2020b). Antimicrobial Peptides (AMPs): Roles, Functions and Mechanism of Action. *International Journal of Peptide Research and Therapeutics*, 26(3), 1451–1463. <https://doi.org/10.1007/S10989-019-09946-9/METRICS>
- Shinde, P. P., & Shah, S. (2018). A Review of Machine Learning and Deep Learning Applications. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*. <https://doi.org/10.1109/ICCUBEA.2018.8697857>

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Strömstedt, A. A. A., Ringstad, L., Schmidtchen, A., & Malmsten, M. (2010). Interaction between amphiphilic peptides and phospholipid membranes. *Current Opinion in Colloid & Interface Science*, 15(6), 467–478. <https://doi.org/10.1016/J.COCIS.2010.05.006>
- Tablada -Germán, C. J., & Torres, A. (2009). Redes Neuronales Artificiales. *Revista de Educación Matemática*, 24(3). <https://doi.org/10.33044/REVEM.10280>
- Téllez, G. A., & Castaño, J. C. (2010). Péptidos antimicrobianos. *Infectio*, 14(1), 55–67.
- Terziyski, Z., Terziyska, M., Hadzhikoleva, S., & Desseva, I. (2023). A software tool for data mining of physicochemical properties of peptides. *BIO Web of Conferences*, 58, 03007.
- Veltri, D., Kamath, U., & Shehu, A. (2018). *Deep learning improves antimicrobial peptide recognition*. <https://doi.org/10.1093/bioinformatics/bty179>
- Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37.
- Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., & Barzilay, R. (2019). Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model*, 59, 40. <https://doi.org/10.1021/acs.jcim.9b00237>
- Yuste, R. (2015a). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience* 2015 16:8, 16(8), 487–497. <https://doi.org/10.1038/nrn3962>
- Yuste, R. (2015b). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience* 2015 16:8, 16(8), 487–497. <https://doi.org/10.1038/nrn3962>

ANEXOS

Estadísticos descriptivos de la Librería virtual (Base de datos)

Propiedades	count	mean	std	min	25%	50%	75%	max
MIC_μM	1401	3,01E+13	8,3E+14	0	2,929021	8,917759	25	2,72E+16
Hydrophobicity	1401	0,096412	1,025857	-4,5	-0,54211	0,175676	0,884615	2,12
Hydrophobic Moment	1401	0,595707	0,216083	0	0,458334	0,58451	0,719254	1,233191
Aliphatic_Index	1401	112,7303	53,02097	0	76,11111	116,25	150,7692	292,5
Instability_Index	1401	32,12448	50,72138	-67,41	2,52381	19,15455	45,99231	509,95
Net_Charge (pH 7)	1401	3,854351	2,441423	-8,14322	1,99489	3,887158	5,029489	17,98157
Isoelectric_Point	1401	10,92994	1,649209	3,544783	9,834331	11,10307	12,24366	13,64613
Molecular_Weight	1401	2040,815	645,8351	18,01524	1584,963	1937,403	2454,895	4573,485
Boman_Index	1401	1,086403	2,199684	-3,127	-0,538	0,507	2,551176	14,92
Activity_Label	1401	0,505706	0,500146	0	0	1	1	1

Realizado por: Hernández, A., 2024.

Secuencias análogas generadas, a partir del péptido Dermaseptina S1.

Secuencia	Hidrofobidad	Momento Hidrofóbico	Índice Alifático	Índice Inestabilidad	Carga Neta	Punto Isoeléctrico	Peso Molecular	Índice Boman	Predicción
WLWKTMLKKLGTMAL	0,1058	0,508	89,4	5,476	3,03	10,8	357	0,14	0,78
HAGKAALGAAADTISQ	82	989	117	471	040	0273	0,22	794	3955
GTQ			6		8		9	1	
ALWKTMLKKLGTMALH	0,1852	0,508	92,3	1,847	3,03	10,8	345	0,16	0,78
AGKAALGAAADTISQGT	94	989	529	059	040	0273	5,09	323	3955
Q			4		8		4	5	
ALRKTMLKKLGTMALH	0,0794	0,484	92,3	0,547	4,03	11,2	342	0,67	0,78
AGKAALGAAADTISQGT	12	037	529	059	039	482	5,06	058	3955
Q			4		9		8	8	
ALWRTMLKKLGTMALH	0,1676	0,556	92,3	1,847	3,03	11,0	348	0,43	0,78
AGKAALGAAADTISQGT	47	08	529	059	114	7954	3,10	882	3955
Q			4				8	4	
ALWKTMLKKLGTMALH	0,1852	0,508	92,3	1,847	3,03	10,8	345	0,16	0,78
AGKAALGAAADTISQGT	94	989	529	059	040	0273	5,09	323	3955
Q			4		8		4	5	
ALWKWMLKKLGTMAL	0,1794	0,455	92,3	8,811	3,03	10,8	354	0,01	0,86
HAGKAALGAAADTISQ	12	512	529	765	040	0273	0,20	911	3902
GTQ			4		8		2	8	
ALWKRMLKKLGTMALH	0,0735	0,704	92,3	11,43	4,03	11,2	351	0,52	0,80
AGKAALGAAADTISQGT	29	414	529	529	039	482	0,17	647	096
Q			4		9		7	1	

ALWKKMLKKLGTMALH	0,0911	0,620	92,3	11,43	4,02	10,9	348	0,25	0,76
AGKAALGAAADTISQGT	76	783	529	529	966	7761	2,16	088	8497
Q			4		7		3	2	
ALWKTWLKKLGTMALH	0,1029	0,520	92,3	1,055	3,03	10,8	351	0,16	0,78
AGKAALGAAADTISQGT	41	168	529	882	040	0273	0,11	382	9861
Q			4		8		5	4	
ALWKTMWKKLGTMAL	0,0470	0,496	80,8	4,344	3,03	10,8	352	0,23	0,85
HAGKAALGAAADTISQ	59	548	823	118	040	0273	8,14	941	5371
GTQ			5		8		8	2	
ALWKTMLRKLGTMALH	0,1676	0,595	92,3	10,00	3,03	11,0	348	0,43	0,85
AGKAALGAAADTISQGT	47	597	529	882	114	7954	3,10	882	3297
Q			4				8	4	
ALWKTMLKKLGTMALH	0,1852	0,508	92,3	1,847	3,03	10,8	345	0,16	0,78
AGKAALGAAADTISQGT	94	989	529	059	040	0273	5,09	323	3955
Q			4		8		4	5	
ALWKTMLKRLGTMALH	0,1676	0,539	92,3	13,93	3,03	11,0	348	0,43	0,84
AGKAALGAAADTISQGT	47	747	529	235	114	7954	3,10	882	4656
Q			4				8	4	
ALWKTMLKKLGTMALH	0,1852	0,508	92,3	1,847	3,03	10,8	345	0,16	0,78
AGKAALGAAADTISQGT	94	989	529	059	040	0273	5,09	323	3955
Q			4		8		4	5	
ALWKTMLKKWGTMAL	0,0470	0,486	80,8	1,294	3,03	10,8	352	0,23	0,82
HAGKAALGAAADTISQ	59	328	823	118	040	0273	8,14	941	8647
GTQ			5		8		8	2	
ALWKTMLKKLWTMAL	0,1705	0,506	92,3	6,888	3,03	10,8	358	0,12	0,88
HAGKAALGAAADTISQ	88	871	529	235	040	0273	4,25	235	1764
GTQ			4		8		5	3	
ALWKTMLKKLRTMALH	0,0647	0,608	92,3	10,00	4,03	11,2	355	0,62	0,82
AGKAALGAAADTISQGT	06	944	529	882	039	482	4,23	970	2668
Q			4		9			6	
ALWKTMLKKLKTMALH	0,0823	0,557	92,3	1,847	4,02	10,9	352	0,35	0,68
AGKAALGAAADTISQGT	53	678	529	059	966	7761	6,21	411	068
Q			4		7		6	8	
ALWKTMLKKLGWMAL	0,1794	0,434	92,3	14,93	3,03	10,8	354	0,01	0,92
HAGKAALGAAADTISQ	12	664	529	824	040	0273	0,20	911	5269
GTQ			4		8		2	8	
ALWKTMLKKLGRMALH	0,0735	0,728	92,3	4,344	4,03	11,2	351	0,52	0,84
AGKAALGAAADTISQGT	29	565	529	118	039	482	0,17	647	7801
Q			4		9		7	1	
ALWKTMLKKLGKMALH	0,0911	0,636	92,3	11,43	4,02	10,9	348	0,25	0,78
AGKAALGAAADTISQGT	76	872	529	529	966	7761	2,16	088	1828
Q			4		7		3	2	
ALWKTMLKKLGTMWL	0,1058	0,508	89,4	1,847	3,03	10,8	357	0,14	0,88
HAGKAALGAAADTISQ	82	989	117	059	040	0273	0,22	794	427
GTQ			6		8		9	1	
ALWKTMLKKLGTMAW	0,0470	0,508	80,8	8,811	3,03	10,8	352	0,23	0,82
HAGKAALGAAADTISQ	59	989	823	765	040	0273	8,14	941	1511
GTQ			5		8		8	2	
ALWKTMLKKLGTMAL	0,2529	0,508	92,3	4,391	2,99	10,8	350	-	0,90
WAGKAALGAAADTISQ	41	989	529	176	212	0271	4,16	0,04	9238
GTQ			4		1		6	235	

ALWKTMLKKLGTMALR	0,1470	0,638	92,3	7,511	3,99	11,2	347	0,46	0,87
AGKAALGAAADTISQGT	59	477	529	765	211	4819	4,14	5	75
Q			4		3		1		
ALWKTMLKKLGTMALH	0,0176	0,508	89,4	6,314	4,02	10,9	351	0,37	0,67
KGKAALGAAADTISQGT	47	989	117	706	966	7761	2,18	970	1404
Q			6		7		9	6	
ALWKTMLKKLGTMALH	0,1705	0,508	92,3	4,344	3,03	10,8	358	0,12	0,86
AWKAALGAAADTISQG	88	989	529	118	040	0273	4,25	235	585
TQ			4		8		5	3	
ALWKTMLKKLGTMALH	0,0647	0,508	92,3	4,344	4,03	11,2	355	0,62	0,69
ARKAALGAAADTISQGT	06	989	529	118	039	482	4,23	970	1049
Q			4		9			6	
ALWKTMLKKLGTMALH	0,0823	0,508	92,3	4,344	4,02	10,9	352	0,35	0,63
AKKAALGAAADTISQGT	53	989	529	118	966	7761	6,21	411	4364
Q			4		7		6	8	
ALWKTMLKKLGTMALH	0,1676	0,509	92,3	4,344	3,03	11,0	348	0,43	0,84
AGRAALGAAADTISQGT	47	025	529	118	114	7954	3,10	882	1918
Q			4				8	4	
ALWKTMLKKLGTMALH	0,1852	0,508	92,3	1,847	3,03	10,8	345	0,16	0,78
AGKAALGAAADTISQGT	94	989	529	059	040	0273	5,09	323	3955
Q			4		8		4	5	
ALWKTMLKKLGTMALH	0,0176	0,508	89,4	1,847	4,02	10,9	351	0,37	0,71
AGKKALGAAADTISQGT	47	989	117	059	966	7761	2,18	970	4051
Q			6		7		9	6	
ALWKTMLKKLGTMALH	0,1058	0,508	89,4	5,476	3,03	10,8	357	0,14	0,84
AGKAWLGAAADTISQG	82	989	117	471	040	0273	0,22	794	3897
TQ			6		8		9	1	
ALWKTMLKKLGTMALH	0,1705	0,508	92,3	6,888	3,03	10,8	358	0,12	0,86
AGKAALWAAADTISQG	88	989	529	235	040	0273	4,25	235	9375
TQ			4		8		5	3	
ALWKTMLKKLGTMALH	0,0647	0,585	92,3	10,00	4,03	11,2	355	0,62	0,84
AGKAALRAAADTISQGT	06	928	529	882	039	482	4,23	970	9234
Q			4		9			6	
ALWKTMLKKLGTMALH	0,0823	0,512	92,3	1,847	4,02	10,9	352	0,35	0,79
AGKAALKAAADTISQGT	53	837	529	059	966	7761	6,21	411	8622
Q			4		7		6	8	
ALWKTMLKKLGTMALH	0,1058	0,508	89,4	3,552	3,03	10,8	357	0,14	0,81
AGKAALGWAADTISQG	82	989	117	941	040	0273	0,22	794	5042
TQ			6		8		9	1	
ALWKTMLKKLGTMALH	0,0176	0,558	89,4	1,847	4,02	10,9	351	0,37	0,76
AGKAALGKAADTISQGT	47	009	117	059	966	7761	2,18	970	855
Q			6		7		9	6	
ALWKTMLKKLGTMALH	0,0176	0,508	89,4	1,847	4,02	10,9	351	0,37	0,79
AGKAALGAKADTISQGT	47	989	117	059	966	7761	2,18	970	2758
Q			6		7		9	6	
ALWKTMLKKLGTMALH	0,1058	0,508	89,4	4,344	3,03	10,8	357	0,14	0,82
AGKAALGAAWDTISQG	82	989	117	118	040	0273	0,22	794	6086
TQ			6		8		9	1	
ALWKTMLKKLGTMALH	0,0176	0,508	89,4	4,344	4,02	10,9	351	0,37	0,77
AGKAALGAAKDTISQGT	47	989	117	118	966	7761	2,18	970	863
Q			6		7		9	6	

ALWKTMLKKLGTMALH	0,2617	0,508	92,3	4,344	4,03	11,2	352	-	0,82
AGKAALGAAAWTISQG	65	989	529	118	023	7834	6,21	0,16	4891
TQ			4				9	176	
ALWKTMLKKLGTMALH	0,1735	0,508	92,3	8,764	5,02	11,4	346	0,07	0,78
AGKAALGAAAKTISQGT	29	989	529	706	948	0292	8,18		8956
Q			4		9				
ALWKTMLKKLGTMALH	0,1794	0,508	92,3	6,267	3,03	10,8	354	0,01	0,82
AGKAALGAAADWISQG	12	989	529	647	040	0273	0,20	911	683
TQ			4		8		2	8	
ALWKTMLKKLGTMALH	0,0735	0,508	92,3	4,05	4,03	11,2	351	0,52	0,81
AGKAALGAAADRISQG	29	989	529		039	482	0,17	647	3744
TQ			4		9		7	1	
ALWKTMLKKLGTMALH	0,0911	0,508	92,3	1,273	4,02	10,9	348	0,25	0,78
AGKAALGAAADKISQG	76	989	529	529	966	7761	2,16	088	4163
TQ			4		7		3	2	
ALWKTMLKKLGTMALH	0,0764	0,539	92,3	1,847	4,03	11,2	352	0,50	0,80
AGKAALGAAADTIRQG	71	957	529	059	039	482	4,20	205	1712
TQ			4		9		3	9	
ALWKTMLKKLGTMALH	0,0941	0,508	92,3	0,65	4,02	10,9	349	0,22	0,78
AGKAALGAAADTIKQGT	18	989	529		966	7761	6,19	647	6436
Q			4		7			1	
ALWKTMLKKLGTMALH	0,0647	0,508	92,3	4,344	4,03	11,2	355	0,62	0,78
AGKAALGAAADTISQRT	06	989	529	118	039	482	4,23	970	5704
Q			4		9			6	
ALWKTMLKKLGTMALH	0,0823	0,508	92,3	4,344	4,02	10,9	352	0,35	0,78
AGKAALGAAADTISQKT	53	989	529	118	966	7761	6,21	411	4525
Q			4		7		6	8	
ALWKTMLKKLGTMALH	0,1794	0,508	92,3	10,19	3,03	10,8	354	0,01	0,78
AGKAALGAAADTISQG	12	989	529	118	040	0273	0,20	911	4946
WQ			4		8		2	8	
ALWKTMLKKLGTMALH	0,0735	0,508	92,3	12,22	4,03	11,2	351	0,52	0,78
AGKAALGAAADTISQG	29	989	529	647	039	482	0,17	647	4474
RQ			4		9		7	1	
ALWKTMLKKLGTMALH	0,0911	0,508	92,3	11,02	4,02	10,9	348	0,25	0,78
AGKAALGAAADTISQG	76	989	529	941	966	7761	2,16	088	3754
KQ			4		7		3	2	
ALWKTMLKKLGTMALH	0,1558	0,549	92,3	4,064	4,03	11,2	348	0,43	0,78
AGKAALGAAADTISQGT	82	631	529	706	039	482	3,15	911	4969
R			4		9		1	8	
ALWKTMLKKLGTMALH	0,1735	0,508	92,3	4,064	4,02	10,9	345	0,16	0,78
AGKAALGAAADTISQGT	29	989	529	706	966	7761	5,13	352	411
K			4		7		8	9	

Realizado por: Hernández, A., 2024.