



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Maestría y Doctorado en Ciencias Bioquímicas**

**USO Y VALIDACIÓN DE UN SOFTWARE QUE PREDICE Y DISEÑA PÉPTIDOS  
ANTIMICROBIANOS Y PENETRADORES DE CÉLULA**

**TESIS**

**QUE PARA OPTAR POR EL GRADO DE:  
Maestro en Ciencias**

**PRESENTA:**

**DAVID ALEJANDRO CASTILLO GONZÁLEZ**

**GABRIEL DEL RÍO GUERRA  
INSTITUTO DE FISIOLÓGÍA CELULAR**

**GERARDO ALFONSO CORZO BURGUETE  
INSTITUTO DE BIOTECNOLOGÍA**

**LOURIVAL DOMINGOS POSSANI POSTAY  
INSTITUTO DE BIOTECNOLOGÍA**

**Ciudad de México. Septiembre, 2021**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## **Dedicatoria**

Quisiera dedicar este trabajo a mis padres, Alejandra Idalia González Villa y José Luis Castillo González, por brindarme siempre su inmensurable apoyo y amor desde mi nacimiento.

También a Laura Marina Robles Reyes, a quien amo y se ha convertido en mi fuente de motivación.

## **Agradecimientos**

Al Dr. Gabriel Del Río, por haberme permitido realizar mi proyecto en su laboratorio.

A la Dra. María Teresa Lara, por su apoyo técnico durante mi proyecto y por brindarme su ayuda y su experiencia cuando la necesité.

A mis compañeros de laboratorio: Erika, Vladimir, John y Rafael, por siempre aconsejarme en situaciones que se me presentaban, por hacer más amena mi estancia en el laboratorio, por compartir todas sus experiencias.

A los Dres. Chris Wood, Soledad Funes y Andrés Saralegui, por su asistencia técnica para la obtención de los resultados presentados en este trabajo.

Al Consejo Nacional de Ciencia y Tecnología (número de becario 416264, CB-252316) y al Posgrado de Maestría y Doctorado en Ciencias Bioquímicas, por su apoyo para poder realizar mis estudios de maestría.

Al Programa de Apoyo a los Estudios del Posgrado (PAEP) por apoyarme con fondos para presentar mi proyecto en el XXXI Congreso Nacional de Bioquímica.

# Contenido

<b>Resumen</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>Introducción</b>	<b>3</b>
<b>Antecedentes</b>	<b>4</b>
Péptidos antimicrobianos	4
Péptidos penetradores de células	6
Aprendizaje automatizado	8
Árboles de decisión	8
Evaluación de modelos	10
Predicción de péptidos antimicrobianos con aprendizaje automatizado	10
Diseño de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera	11
<b>Hipótesis</b>	<b>15</b>
Objetivos generales	15
Objetivos particulares	15
<b>Materiales</b>	<b>16</b>
Cepas utilizadas en el proyecto	16
Péptidos antimicrobianos	17
Medios y condiciones de cultivo	17
Equipo	18
Software	18
Datos	18
<b>Métodos</b>	<b>20</b>
Predicción de péptidos antimicrobianos	20
Ensayos de crecimiento bacteriano	21
Evaluación de la habilidad de penetración celular de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera. Preparación del inóculo	21
Medición de la internalización péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera por microscopía confocal de fluorescencia	22
Medición de la internalización del péptido $\alpha$ -NLS-C en MAT $\alpha$ y MAT $\alpha$ por microscopía confocal de fluorescencia	22
Medición de la cinética de internalización de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera por citometría de flujo	22
<b>Resultados y discusión</b>	<b>24</b>
Predicción de péptidos antimicrobianos	24
Crecimiento bacteriano en presencia de RPL29	26

Evaluación de la habilidad de penetración celular de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera	28
Evaluación de la selectividad celular de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera	29
<b>Conclusiones</b>	<b>34</b>
<b>Referencias</b>	<b>35</b>

# Lista de figuras

Figura 1.1	Construcción de un árbol de decisión	9
Figura 1.2	Representación esquemática de las feromonas (feromona a y feromona $\alpha$ ) mediante la conjugación sexual de las células haploides de <i>Saccharomyces cerevisiae</i> .	12
Figura 5.1	Curva ROC de evaluación de modelo de predicción de AMPs.	30
Figura 5.2	Secuencia de aminoácidos de RPL29 y b) Proyección helicoidal de aminoácidos 38-55 de RPL29	26
Figura 5.3	Áreas bajo de la curva obtenidas de experimentos de crecimiento bacteriano de <i>Escherichia coli</i> , <i>Pseudomonas aeruginosa</i> y <i>Staphylococcus aureus</i> en presencia de diferentes agentes: RPL29, ampicilina y agua destilada	27
Figura 5.4	Micrografías de la internalización de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera en células de levadura con haplotipo MATa.	29
Figura 5.5	Internalización del péptido $\alpha$ -NLS-C en <i>S. cerevisiae</i> .	30
Figura 5.6	Niveles de fluorescencia celular de inóculos que contienen uno de los haplotipos dobles de célula en un lapso de 60 minutos	32
Figura 5.7	Niveles de fluorescencia celular de inóculos que contienen ambos haplotipos de <i>S. cerevisiae</i> , para medir la tasa de internalización de los péptidos penetradores de célula en un lapso de 60 minutos.	33

# Lista de tablas

Tabla 1.1	Ejemplos de péptidos antimicrobianos que se encuentran en la naturaleza.	5
Tabla 1.2	Secuencias de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera.	19
Tabla 3.1	Cepas utilizadas en el proyecto.	16
Tabla 3.2	Péptidos antimicrobianos usados en este trabajo	17

## Resumen

Este trabajo consistió en identificar nuevos péptidos antimicrobianos mediante el uso de un software de aprendizaje automatizado (*machine learning*), desarrollado previamente por nuestro grupo de trabajo. Para entrenar el clasificador, se usaron dos conjuntos de entrenamiento: uno con secuencias peptídicas reportadas como antibacterianas y otro con secuencias sin actividad antimicrobiana reportada, o que no tuvieran propiedades de péptidos antimicrobianos. Los parámetros empleados para la clasificación fueron composición de aminoácidos, carga neta, ventana móvil de carga, hidrofobicidad, ventana móvil de momento hidrofóbico, punto isoelectrico, coeficiente de partición de agua octanol y aproximación del contenido de alfa-hélices. Una vez entrenado el clasificador, se tomaron los péptidos que serían clasificados, que provenían de regiones conservadas de familias de proteínas, de la base de datos Pfam. Usando este algoritmo, el péptido RPL29, componente de la subunidad 60S del ribosoma de *Saccharomyces cerevisiae*, fue identificado como un potencial péptido antimicrobiano. Sin embargo, los ensayos de crecimiento bacteriano realizados en este trabajo muestran que este péptido no tiene actividad antimicrobiana contra bacterias gram negativas (*Escherichia coli*, *Pseudomonas aeruginosa*) o gram positivas (*Staphylococcus aureus*).

Por otro lado, en este trabajo se evaluó la selectividad celular de los péptidos penetradores de células,  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera, diseñados previamente por nuestro grupo, en células de *Saccharomyces cerevisiae* MAT $\alpha$  y MATa. Se realizaron experimentos de citometría de flujo acoplada a fluorescencia para comparar la velocidad de internalización. Nuestros resultados muestran que la tasa de internalización fue ligeramente mayor en el haplotipo de levadura MATa.

## Abstract

For this work, in order to identify new antimicrobial peptides, a machine learning-based software developed at our laboratory was used. This software, which implements a random forest, was trained with two datasets: the positive set with reported AMPs, and the negative set containing peptides neither reported as AMPs nor desirable properties on AMPs. The features selected for the prediction were amino acid composition, net charge, hydrophobicity, hydrophobic moment window, isoelectric point, water-octanol partition coefficient and alpha-helix content approximation. With the trained predictor, it was used to perform AMP prediction for all the conserved regions contained on the Pfam database, grouped on protein families. After performing the prediction, peptide RPL29, a component of the 60S ribosomal subunit of *Saccharomyces cerevisiae*, was identified as one of the most promising predicted AMPs. Nevertheless, after performing bacterial growth assays for both gram-negative (*Escherichia coli* and *Pseudomonas aeruginosa*) and gram-positive (*Staphylococcus aureus*) bacteria, in presence of the peptide, antimicrobial activity could not be observed.

Additionally, in this work cell penetrating peptides previously synthesized in our laboratory ( $\alpha$ -NLS-C, NLS- $\alpha$ -CE and Chimera) were tested for their selectivity in both haplotypes of yeast, MATa and MAT $\alpha$ , by means of microscopy-coupled flow cytometry assays. As observed in our results, the internalization rate was somewhat higher on one of the yeast haplotypes, MATa.

## Introducción

Dos tipos importantes de péptidos con una aplicación terapéutica son los péptidos antimicrobianos y los péptidos penetradores de células. Los péptidos antimicrobianos (AMPs por sus siglas del inglés *antimicrobial peptides*) son considerados una alternativa a los antibióticos tradicionales debido a que exhiben un amplio espectro antimicrobiano y no inducen resistencia fácilmente (Magana et al, 2020), mientras que los péptidos penetradores de célula (CPPs por sus siglas del inglés *cell penetrating peptides*) tienen una prometedora aplicación en el transporte de fármacos al interior de las células (Xie et al, 2020).

No obstante, las metodologías experimentales para la identificación y diseño de AMPs y CPPs son costosas y consumen mucho tiempo. A fin de reducir los costos de investigación experimental, es importante desarrollar herramientas computacionales para la predicción de péptidos, que identifiquen a los mejores candidatos y ayuden a limitar el gran número de secuencias que se tienen que evaluar en el laboratorio. Una de estas herramientas es el aprendizaje automatizado (*machine learning*). Este enfoque se basa en entrenar y validar modelos sobre bases de datos caracterizadas experimentalmente, para así identificar nuevos péptidos candidatos con la actividad biológica que se está buscando. Utilizando este enfoque, nuestro grupo desarrolló un software para predecir péptidos con alguna función específica, tales como AMPs o CPPs, con el cual posteriormente se diseñaron tres péptidos quiméricos para ser penetradores de células de levadura (Diener et al, 2016).

El presente trabajo consta de dos objetivos principales. Por una parte, entrenar un modelo para identificar nuevas secuencias peptídicas con posible función antimicrobiana, y llevar a cabo la validación experimental del mejor candidato utilizando cepas de *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus*. Por otro lado, comprobar experimentalmente la capacidad de internalización de los péptidos quiméricos ( $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera) diseñados previamente en nuestro grupo en células levadura.

## Antecedentes

### Péptidos antimicrobianos

Los AMPs son secuencias peptídicas menores a 100 residuos que son generadas de manera natural por algunos organismos, y que tienen la habilidad de eliminar un amplio espectro de microorganismos (Maróti et al, 2011).

Los péptidos antimicrobianos están presentes en todos los reinos biológicos. En los organismos multicelulares, los AMPs participan como la primera línea de defensa contra bacterias, hongos, protozoarios y virus, e incluso contra células cancerosas (Ahmed et al, 2019; Mookherjee et al, 2020). En los microorganismos, como eucariotas unicelulares y arqueas, estos péptidos son usados para la autoprotección y competición microbiana (Raffatellu et al, 2018).

A pesar de la enorme diversidad de secuencias que presentan los distintos péptidos antimicrobianos, estos tienen ciertas propiedades fisicoquímicas en común. Generalmente presentan una carga positiva de entre +2 y +9 (debido a la presencia de aminoácidos como lisina y arginina), y una proporción sustancial de residuos hidrofóbicos (>30%). Debido a su carga positiva, estos perturban preferencialmente membranas cargadas negativamente, mientras que la característica de ser anfipáticos les permite interactuar con membranas celulares (Hof et al, 2001).

Los AMPs pueden asumir una variedad de estructuras secundarias, incluyendo  $\alpha$ -hélices, láminas  $\beta$  paralelas, asas y estructuras lineales extendidas (Koehbach y Craik, 2019). Particularmente, en los AMPs  $\alpha$ -helicoidales, agrupaciones de aminoácidos hidrofóbicos y catiónicos se distribuyen espacialmente en regiones discretas. Esta organización resulta en un diseño anfipático que consiste en una separación longitudinal de la hélice en dos partes: en un lado de la hélice se encuentran los aminoácidos polares y catiónicos, mientras que en la otra cara se localizan los aminoácidos hidrofóbicos (Hof, 2001).

Los mecanismos de actividad antimicrobiana de los AMPs varían desde la permeabilización en la membrana, despolarización o lisis (Bradshaw, 2003; Izadpanah y Gallo, 2005; Matsuzaki, 2009), la creación de agujeros físicos que causan que el contenido celular se escape (Yang et al, 2000), la activación de procesos de muerte, tal como la inducción de hidrolasas que

degradan la pared celular (Bierbaum y Sahl, 1985), la alteración de la distribución usual de lípidos resultando en la disrupción de las funciones de la membrana (Matsuzaki, 1999).

A la fecha, se han encontrado más de 3000 AMPs provenientes de los seis reinos biológicos (Wang et al, 2016). Algunas de las secuencias peptídicas antimicrobianas más estudiadas se encuentran en la tabla 1.1.

Tabla 1.1 Ejemplos de péptidos antimicrobianos que se encuentran en la naturaleza (Zasloff, 2002). Los subíndices en algunas secuencias indican las cisteínas que forman los puentes disulfuro (las cisteínas con el mismo número forman el puente).

<b>Péptido antimicrobiano</b>	<b>Secuencia</b>	<b>Propiedades</b>	<b>Origen</b>
Cecropina A	KWKLFKKIEKVGQNIRDGIIKAGPAVAVVGQATQIAK	Catiónico, hélice alfa	Polilla
Magainina 2	GIGKFLHSAKKFGKAFVGEIMNS	Catiónico, hélice alfa	Anfibios
LL37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLPRTES	Catiónico, hélice alfa	Humanos
Melitina	GIGAVLKVLTGGLPALISWIKRKRQQ	Catiónico, hélice alfa	Insectos
Buforina II	TRSSRAGLQFPVGRVHRLLRK	Catiónico, hélice alfa	Vertebrados
Taquiplesina	RWC <sub>1</sub> FRVC <sub>2</sub> YRGIC <sub>2</sub> YRKC <sub>1</sub> R	Puentes disulfuro	Cangrejo
Defensina $\alpha$	DC <sub>1</sub> YC <sub>2</sub> RIPAC <sub>3</sub> IAGERRYGTC <sub>2</sub> IYQGRLWAFCC <sub>3</sub> C <sub>1</sub>	Puentes disulfuro	Humano
Defensina $\beta$	NPVSC <sub>1</sub> VRNKGIC <sub>2</sub> VPIRC <sub>3</sub> PGSMKQIGTC <sub>2</sub> VGRAVKC <sub>1</sub> C <sub>3</sub> RKK	Puentes disulfuro	Mamíferos
Indolicidina	ILPWKWPWWPWRR	Lineal	Mamíferos

## **Péptidos penetradores de células**

Los péptidos penetradores de células (CPPs) se definen como secuencias peptídicas menores a 30 aminoácidos que tienen la habilidad intrínseca de atravesar membranas biológicas (Milletti, 2012).

Muchos CPPs se derivan de proteínas naturales, pero otros son quiméricos o completamente sintéticos. Los primeros CPPs fueron identificados en 1988 y 1994, estos se derivan de la proteína transactivadora (Tat) del Virus de Inmunodeficiencia Humana tipo 1, y de la proteína homeótica antennapedia de *Drosophila* (pAntp), respectivamente (Frankel y Pabo, 1988; Derossi et al, 1994). En 1997, un péptido acarreador corto (MPG) que contiene dominios hidrofílicos e hidrofóbicos fue sintetizado para formar complejos no covalentes con cargos (Morris et al, 1997). En 1998, Langel et al diseñaron el primer péptido acarreador quimérico que incluía el fragmento N-terminal del neuropéptido galanina, ligado a un péptido extraído del veneno de avispa (mastosparan), nombrado como el péptido Transportan (Langel et al, 1996; Pooga et al, 1998).

Aunque los CPPs poseen una extensa variedad de secuencias, se pueden clasificar en tres grandes grupos, con base en sus propiedades fisicoquímicas: catiónicos, anfipáticos e hidrofóbicos. Los péptidos catiónicos se caracterizan por estar cargados positivamente debido a la presencia de lisinas y argininas en su secuencia, mientras que los péptidos anfipáticos se distinguen por contener una parte hidrofóbica, así como otra hidrofílica en su secuencia, además de que ésta adopta una estructura helicoidal. En cuanto a los péptidos hidrofóbicos, estos son ricos en aminoácidos apolares y tienen una carga neta baja, o tienen un motivo o grupo químico hidrofóbico crucial para la internalización (Pooga y Langel, 2015).

Un CPP puede contener una secuencia de localización nuclear (NLS por sus siglas del inglés *Nuclear Localization Sequence*). Las NLS consisten en pequeñas secuencias peptídicas que contienen motivos ricos en lisina, arginina y prolina. La presencia de dichos motivos les permiten translocar a través de la membrana nuclear por medio de los poros nucleares, dentro del núcleo (Ragin, 2002). Aunque las NLSs no son capaces de actuar por sí mismas como CPPs, debido a que poseen una carga neta baja, pueden ser usadas en CPPs anfipáticos, al unirse covalentemente a una secuencia hidrofóbica, para así facilitar la introducción del péptido

en el núcleo (Zhang et al, 2016; Morris et al, 2001).

Desde el descubrimiento de los péptidos penetradores celulares ha habido un creciente interés en el estudio de su internalización, no obstante, este mecanismo sigue sin ser comprendido en su totalidad. La internalización de los CPPs puede ocurrir por dos mecanismos en general: endocitosis y translocación directa (Patel et al, 2007). Se piensa que estas vías pueden ocurrir simultáneamente, o que péptidos diferentes pueden utilizar distintos mecanismos dependiendo de sus propiedades biofísicas y carga (Fischer et al, 2002). Aún permanece ambiguo si los CPPs son capaces de interactuar con receptores de membrana.

Los CPPs han sido usados como acarreadores de fármacos y como fármacos por sí mismos, en el tratamiento de diversas enfermedades (Xie et al, 2020). Se les ha utilizado para traslocar un amplio rango de moléculas tales como proteínas, ácidos nucleicos, ARN de interferencia (siRNA), liposomas y nanopartículas, a través de las membranas celulares (Torchilin, 2007; Endoh et al, 2009; Järver y Langel, 2006). Sin embargo, la principal limitante de los CPPs es que son no específicos, es decir, se internalizan de manera no selectiva en distintos tipos celulares (Reissmann, 2014), lo cual restringe su aplicación médica en el transporte de fármacos al interior de la célula.

## Aprendizaje automatizado

Se conoce como aprendizaje automatizado al estudio, desarrollo y aplicación de algoritmos cuyo objetivo es realizar modelos predictivos de una o más variables, a partir de un conjunto de datos, llamados de entrenamiento.

Con base en un conjunto de reglas matemáticas, los algoritmos de aprendizaje automatizado reconocen patrones en un conjunto de datos, los cuales pueden ser utilizados para hacer predicciones sobre datos no caracterizados. Para que el algoritmo sea capaz de clasificar, es necesario alimentarlo con diferentes conjuntos de datos, para que de esa forma, pueda categorizar correctamente (Raschka et al, 2018).

Los algoritmos de aprendizaje se dividen en dos tipos. El primero, llamado aprendizaje supervisado, se caracteriza por tener conocimiento de las diferentes categorías en las que un objeto puede ser clasificado. Por lo tanto, en cada uno de los ejemplos que constituyen el conjunto de entrenamiento, se especifica la categoría a la que pertenecen. Dentro de las metodologías más utilizadas para aprendizaje supervisado, se incluye la máquina de vectores de soporte (SVM, del inglés *support vector machine*), bosque aleatorio, k vecinos más próximos, regresión lineal, regresión, regresión logística, entre otros (Géron, 2019). El segundo se conoce como aprendizaje no supervisado, en donde las observaciones utilizadas para el aprendizaje carecen de etiquetas; es decir, que no han sido categorizadas, por lo que son utilizados para determinar si existe alguna forma de distinguir objetos, con base en los atributos utilizados. Los algoritmos de *clustering* y de redes neuronales son los más utilizados para esta forma de clasificación (Cielen et al, 2016).

### Árboles de decisión

Para construir un árbol de decisión, se requieren datos de observaciones (valores de los atributos o propiedades), con sus respectivas etiquetas. Para la separación de los objetos en diferentes clases, el algoritmo se encarga de definir algunos valores límite de sus atributos, hasta el punto en que solamente se tenga un conjunto de objetos en el que todos pertenezcan a una misma clase. Se utilizan diferentes métricas para medir esta separación, por ejemplo, la impureza Gini. La ecuación que describe esta medición es:

$$\text{Impureza Gini: } G = \sum_{k=1}^c P(i) * (1 - P(i))$$

En donde  $C$  corresponde al número total de clases que se desean predecir, y  $P(i)$  significa la probabilidad de elegir un dato etiquetado con la clase "i".

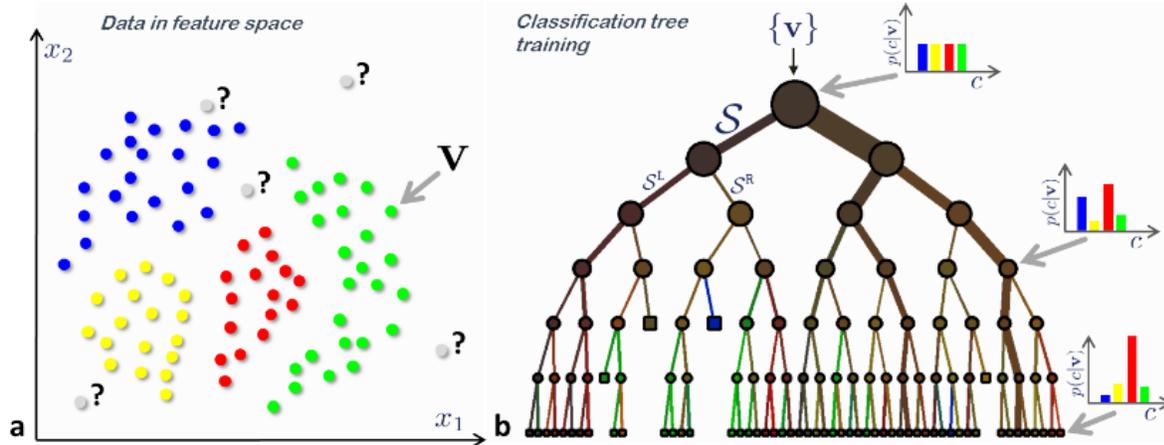


Figura 1.1 Construcción de un árbol de decisión. En **a** se tienen los datos a estudiar; cada dato ( $v$ ) del conjunto ( $V$ ) consta de dos atributos ( $x_1$  y  $x_2$ ) además de su clase (color). En **b** se tiene el árbol de decisión generado de los datos en **a**. En cada nodo del árbol, se crea una ramificación, la cual surge de la necesidad de separar lo mejor posible los puntos en **a**. En el nodo del conjunto  $S$ , al aplicar la regla para separar los datos, se generan dos nuevos subconjuntos,  $S^L$  y  $S^R$ . Cada dato en gris (?) de **a** se introduce en el árbol de decisión generado, y es conducido desde el nodo raíz, haciéndose camino por las ramificaciones (condiciones de separación), hasta llegar a alguno de los nodos finales, o nodos hoja. Los histogramas que se encuentran en algunos nodos indican la probabilidad de contar con un punto de cierto color en un nodo en particular. Imagen tomada de Criminisi et al, 2011.

En vista de que los árboles de decisión deben generar reglas o condiciones de separación para poder realizar la predicción de la manera más precisa, dichas reglas llegan a ser muy estrictas para los datos de entrenamiento, por lo que al momento de usarlas en algún conjunto de datos nuevos, el algoritmo falla con su objetivo. A este comportamiento se le llama sobreentrenamiento (Raschka et al, 2018). Una de las alternativas para contrarrestar este efecto es la creación de un bosque aleatorio (*random forest*). Un bosque aleatorio es un conjunto de árboles de decisión, ensamblado para aumentar la capacidad de generalización, es decir, para mejorar la predicción o clasificación. Se utiliza un método de ensamblaje de estos árboles de decisión llamado bagging (Bootstrap AGGregatING), en donde se construyen diversos árboles con muestreo por reemplazo, por lo que ya no se alcanza a obtener el 100 % de las observaciones iniciales. Una vez generados estos árboles, se realiza una votación entre

ellos, para decidir cuál es la mejor forma de clasificación.

El bosque aleatorio se considera un algoritmo de conjunto, pues toma un grupo de clasificadores débiles (árboles de decisión), pero al trabajar colectivamente, forman un clasificador fuerte (Criminisi et al, 2011).

### *Evaluación de modelos*

Además de implementar alguna de las técnicas de aprendizaje automatizado para predecir o clasificar, es necesario usar alguna métrica o método que indique qué tan certera o confiable es la predicción. Uno de estos procedimientos es el de validación cruzada. Este consiste en dividir el conjunto de entrenamiento en  $n$  partes iguales; después, realizar el entrenamiento en  $n-1$  partes, y realizar la tarea de predicción en la parte restante. Este mismo procedimiento se ejecuta iterativamente  $n$  veces, de modo que todo el conjunto de datos sea usado tanto para entrenamiento como para evaluación.

### **Predicción de péptidos antimicrobianos con aprendizaje automatizado**

A pesar de poseer propiedades fisicoquímicas en común, los AMPs tienen poca homología de secuencia (Hancock y Chappel, 1999). Por lo tanto, no es tarea sencilla desarrollar un método para predecir péptidos antimicrobianos que se base en similitud. Adicionalmente, la implementación de metodologías experimentales para la identificación y diseño de péptidos antimicrobianos es costosa y consume mucho tiempo. Es por eso que es necesario desarrollar herramientas computacionales para predecir AMPs. Una de estas herramientas es el aprendizaje automatizado (machine learning). Este enfoque se basa en entrenar y validar modelos sobre bases de datos caracterizadas experimentalmente, para así identificar nuevos péptidos candidatos con la actividad biológica que se está buscando.

Entre las primeras aplicaciones del aprendizaje automatizado para encontrar AMPs está el trabajo de Lata et al, quienes desarrollaron una herramienta de clasificación basada en redes neuronales artificiales y SVM (Lata et al, 2007). Posteriormente, otros grupos comenzaron a emplear distintos algoritmos, incluyendo bosques aleatorios, SVM y redes neuronales artificiales, para desarrollar herramientas que logren identificar péptidos con actividad antimicrobiana (Cherkasov et al, 2008; Fjell et al, 2009; Wang et al, 2011; Torrent et al, 2011; Maccari et al, 2013; Xiao et al, 2013; Schneider et al, 2017). Estas estrategias han demostrado su capacidad de predicción de péptidos; sin embargo, es necesario seguir mejorando en la

predicción, ya que es necesario disminuir la cantidad de falsos positivos resultantes e incrementar la precisión de la predicción.

Nuestro grupo desarrolló un software de aprendizaje automatizado llamado Dcf (<https://github.com/cdiener/DCF>), el cual se basa en propiedades fisicoquímicas, estructurales y de composición de aminoácidos para predecir y diseñar péptidos con alguna función específica, tales como AMPs o CPPs (Diener et al, 2016). Específicamente, los descriptores empleados para la clasificación corresponden a 27 valores: 20 corresponden al porcentaje de apariciones de cada aminoácido en la secuencia. Los otros siete son carga neta, ventana móvil de carga, hidrofobicidad, ventana móvil de momento hidrofóbico (Tossi et al, 2003), punto isoeléctrico (Lide, 2004), coeficiente de partición de agua octanol (Tao et al, 1999) y aproximación del contenido de alfa-hélices (Pace y Scholtz, 1998). Estos descriptores se eligieron debido a que se ha demostrado su relevancia en la determinación de péptidos con actividad antimicrobiana (Wang et al, 2011, Torrent et al, 2012, Meher et al, 2017). Dcf implementa un algoritmo de bosque aleatorio. Entre las ventajas de trabajar con bosques aleatorios, es que se obtienen altos porcentajes de precisión (es capaz de generalizar correctamente debido al *bagging*, como se explicó anteriormente), además de mostrar rapidez al momento de entrenar el modelo (Qi, 2012).

### **Diseño de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera**

Los CPPs han sido usados para aplicaciones médicas, particularmente para transportar moléculas (fármacos) dentro de las células (Guidotti et al, 2017). Sin embargo, una limitante de estos péptidos es que se internalizan de manera no específica en distintos tipos celulares (Reissmann, 2014) lo que conduce a que penetren en células no diana, potenciando los efectos secundarios causados con relación al efecto terapéutico.

Una estrategia para superar este inconveniente y asegurar que los CPPs sean capaces de internalizar en células o tejidos específicos, sería agregar una secuencia de aminoácidos al péptido que sea reconocida exclusivamente por estas mismas células. Por ejemplo, un péptido que porte alguna secuencia que funcione como señal y que sea reconocida por algún receptor que sea producido solamente por un tipo celular (Diao et al, 2012).

En *S. cerevisiae* existen dos haplotipos diferentes: MAT $\alpha$  y MAT $\alpha$ . El apareamiento entre dos células de levadura con diferentes haplotipos da lugar a una célula diploide MAT $\alpha$ / $\alpha$ . (Engebrecht, 2003). El mecanismo por el cual se da el proceso de apareamiento es el

siguiente: tanto las células MATa como las MAT $\alpha$  producen feromonas (feromona a y feromona  $\alpha$ , respectivamente) que se liberan en el ambiente extracelular y que son reconocidas por el haplotipo contrario (Bardwell, 2005). Para el haplotipo MATa, el receptor encargado de reconocer a la feromona  $\alpha$  es Ste2p, mientras que para MAT $\alpha$ , la feromona a es reconocida por Ste3p (Burkholder y Hartwell, 1985; Hagen et al, 1986). (Figura 1.2).

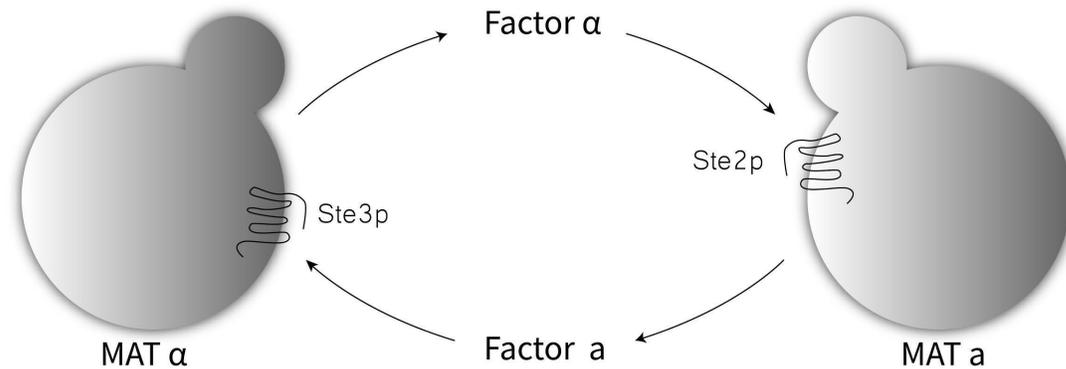


Figura 1.2. Representación esquemática de las feromonas (feromona a y feromona  $\alpha$ ) mediando la conjugación sexual de las células haploides de *Saccharomyces cerevisiae*.

En vista de que las células de levadura MATa reconocen a la feromona  $\alpha$ , debido a la presencia de Ste2p, si un péptido quimérico tiene embebida en su secuencia a la feromona  $\alpha$ , éste posiblemente se internalice preferencialmente en células con haplotipo MATa.

Utilizando Dcf, nuestro grupo diseñó tres péptidos quiméricos llamados  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera, a los cuales se les incorporó la secuencia peptídica de la feromona del haplotipo MAT $\alpha$ ; además de otros aminoácidos que le confieren a la secuencia la capacidad de ser penetradora de células y una secuencia de localización nuclear (NLS) (Diener et al, 2016). Esta última se agregó ya que se deseaba que los péptidos diseñados fueran capaces de introducirse al núcleo.

El procedimiento que utiliza el software Dcf para el diseño de los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera es el siguiente:

1. Primero, un predictor de tipo bosque aleatorio es entrenado sobre un conjunto de secuencias de CPPs y de no CPPs.
2. Luego, el software recibe como entrada una o varias secuencias semilla; en este caso, la secuencia de la feromona  $\alpha$  y una NLS.

3. Los péptidos semilla son unidos mediante unas secuencias peptídicas llamadas "linkers". También, es posible añadir aminoácidos en los extremos N y/o C terminales.
4. Dcf calcula la probabilidad de que la secuencia obtenida sea CPP, usando el modelo entrenado en el paso 1. Esta probabilidad es convertida a una medida de energía ( $E = 1 - \text{PCPP}(s)$ ), para así comenzar con el proceso de optimización mediante el algoritmo de recocido simulado (del inglés *simulated annealing*), hasta obtener un péptido con una alta probabilidad de ser CPP.
5. Para optimizar la secuencia obtenida, ocurren tres procesos diferentes en los linkers y en los aminoácidos añadidos en los flancos: adición, deleción o sustitución de aminoácidos.

El **péptido  $\alpha$ -NLS-C** se nombró de ese modo, pues en el amino terminal se encuentra la secuencia de la feromona  $\alpha$ , seguida de una secuencia de localización nuclear, y por último, se encuentran dos aminoácidos, para que el péptido tuviera capacidad de ser penetrador de células (Tabla 1.2).

Para el caso del **NLS- $\alpha$ -CE**, NLS se encuentra en el extremo amino terminal, y en el extremo carboxilo la feromona  $\alpha$ . Entre estas dos secuencias, se agregaron aminoácidos para que el péptido resultante fuera un buen penetrador de células, es decir, que se internalice de forma eficiente (Tabla 1.2).

Por último, se encuentra el péptido **Chimera**, que también incluye a la feromona  $\alpha$  en el extremo carboxilo, pero ahora la secuencia NLS no se encuentra en el terminal amino del péptido (Tabla 1.2).

Tabla 1.2. Secuencias de los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera. En rojo se indica la secuencia del factor  $\alpha$ , en verde la secuencia de localización nuclear (NLS). P(CPP) indica la probabilidad de ser penetrador de célula (Diener et al, 2016).

Péptido	Secuencia	P(CPP)
$\alpha$ -NLS-C	RL <b>WHWLQLKPGQPMY</b> WR <b>QPKSKRKV</b> RR	0.91
$\alpha$ -NLS-C	FRKWRRK <b>PKKKRKV</b> WWRKVKRR <b>WHWLQLKPGQPMY</b>	0.97
Chimera	KRRWRFVWMN <b>PKKKRKV</b> PPWPYLLW <b>WHWLQLKPGQPMY</b>	0.88

Como se observa en la Tabla 1.2, de acuerdo a los resultados obtenidos de la predicción, los péptidos que se diseñaron tienen alta probabilidad de ser CPPs. Dado que los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera contienen la secuencia de la feromona  $\alpha$  de *S. cerevisiae*, se espera que éstos presenten selectividad por cepas que expresan Ste2p, el receptor de la feromona  $\alpha$ ; y por tanto, que el haplotipo MATa presente una mayor tasa de internalización de estos péptidos.

## Hipótesis

- El empleo del algoritmo de aprendizaje automatizado desarrollado previamente en nuestro grupo, basado en las propiedades fisicoquímicas, estructurales y de composición de aminoácidos de los AMPs y que utiliza el algoritmo de bosques aleatorios, permitirá predecir una secuencia peptídica que será capaz de inhibir el crecimiento bacteriano *in vitro*.
- Los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera, que portan la secuencia de la feromona  $\alpha$  de *Saccharomyces cerevisiae*, serán selectivos para células con haplotipo MAT $\alpha$ .

### Objetivos generales

- Emplear el algoritmo de aprendizaje automatizado desarrollado previamente en nuestro grupo para identificar una secuencia peptídica con posible función antimicrobiana.
- Validar experimentalmente la capacidad antimicrobiana del potencial AMP.
- Demostrar la selectividad de los péptidos penetradores de células,  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera, en *Saccharomyces cerevisiae*.

### Objetivos particulares

- Entrenar un algoritmo de aprendizaje automatizado para que sea capaz de predecir péptidos antimicrobianos.
- Identificar péptidos, a partir de la base de datos Pfam, con la característica de ser antimicrobianos, mediante el uso del algoritmo de aprendizaje automatizado.
- Evaluar la capacidad antimicrobiana del péptido seleccionado, previamente predicho como potencial AMP por el algoritmo, mediante ensayos de crecimiento bacteriano en *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus*.
- Comparar la velocidad de internalización de los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera, en los haplotipos MAT $\alpha$  y MAT $\alpha$  de *S. cerevisiae* mediante citometría de flujo acoplada a fluorescencia.

## Materiales

### Cepas utilizadas en el proyecto

Tabla 3.1 Cepas utilizadas en el proyecto.

<i>Escherichia coli</i>	DH5 $\alpha$ , genotipo: supE44 $\Delta$ lacU169 (o80 lacZ $\Delta$ M15) HSDr17 recA1 endA1 GYRa96 thi-relA1
<i>Pseudomonas aeruginosa</i>	PAO1 Cepa silvestre
<i>Saccharomyces cerevisiae</i>	<p>BY4741, genotipo: MATa his3<math>\Delta</math>1 leu2<math>\Delta</math>0 met15<math>\Delta</math>0 ura3<math>\Delta</math>0</p> <p>BY4742, genotipo: MAT his3<math>\Delta</math>1 leu2<math>\Delta</math> lys2<math>\Delta</math>0 ura3<math>\Delta</math>0</p> <p>MATa<math>\Delta</math>bar1, genotipo: MATa, his3<math>\Delta</math>1 leu2<math>\Delta</math>0 met15<math>\Delta</math>0 ura3<math>\Delta</math>0, gen,nulo:bar1</p> <p>MATa RPL9A-GFP, genotipo: MATa his3<math>\Delta</math>1 leu2<math>\Delta</math>0 met15<math>\Delta</math>0 ura3<math>\Delta</math>0 RPL9A-GFP</p>
<i>Staphylococcus aureus</i>	Staphylococcus aureus ATCC6538

## Péptidos antimicrobianos

Tabla 3.2 Péptidos antimicrobianos usados en este trabajo.

RPL29	MAKSKNHTAHNQTRKAHRNGIKKPKTYKYPSLKGVDPKFRRNHHKHALHGTA KALAAKK
$\alpha$ -NLS- C	RLWHWLQLKPGQPMYWRQPFSKRKVR
NLS- $\alpha$ - CE	FRKWRRKPKKKRKYVWRKVKRRWHWLQLKPGQPMY
Chimer a	KRRWRFVWMNPKKKRKYVWPYLLWWHWLQLKPGQPMY
TAMRA - $\alpha$ -NLS -C	TAMRA-RLWHWLQLKPGQPMYWRQPFSKRKVR
TAMRA -NLS- $\alpha$ -CE	TAMRA-FRWRRKPKKKRKYVWRKVKRRWHWLQLKPGQPMY
TAMRA -Chime ra	TAMRA-KRRWRFVWMNPKKKRKYVWPYLLWWHWLQLKPGQPMY

### Medios y condiciones de cultivo

Para las cepas de bacteria se empleó medio Luria Bertani (LB) con NaCl 5g/L, triptona 10g/L, extracto de levadura 5 g/L, ajustando además el pH a 7.5.

Las cepas de levadura se cultivaron en medio YPD (Yeast Extract-Peptone-Dextrose) líquido y sólido (agar 1.5%). El medio YPD se preparó bajo las siguientes especificaciones: extracto de

levadura 1%, peptona 2%, glucosa 2%. Los medios de cultivo fueron esterilizados a 120°C por 30 minutos.

## Equipo

- Nanodrop Thermo Scientific.
- Synergy Mx, BioTek.
- Thermomixer Comfort, Eppendorf.
- Microscopio confocal Zeiss LSM510 META
- Citómetro de flujo-Microscopio Amnis ImageStream Mark II

## Software

- Dcf, utilizado para la predicción de secuencias peptídicas con actividad antimicrobiana. Disponible en: <https://github.com/cdiener/dcf>
- Amnis IDEAS: preparación de datos de experimentos de internalización de péptidos penetradores de célula en haplotipos de *Saccharomyces cerevisiae*.
- Matplotlib (Hunter, 2007): generación de gráficos para visualización de experimentos de crecimiento bacteriano e internalización de péptidos penetradores de célula en haplotipos de *Saccharomyces cerevisiae*.

## Datos

- Se utilizaron dos conjuntos de entrenamiento (positivo y negativo) para la predicción de péptidos antimicrobianos. El conjunto positivo consistió de 843 péptidos reportados para actividad antimicrobiana (<http://aps.unmc.edu/AP/main.php>), de la base de datos APD, la cual fue curada manualmente por Wang et al (Wang et al, 2016). El conjunto negativo se conforma de 843 péptidos sin anotaciones (experimental o predicha) de ser antimicrobianos (<https://github.com/cdiener/dcf/tree/master/examples>). Estas secuencias se obtuvieron aleatoriamente de la base de datos Uniprot (Magrane y consorcio Uniprot, 2011), procurando que no tuvieran anotaciones de antimicrobiana, antibacteriana, antifungal o efector, de acuerdo con la metodología seguida en trabajos anteriores (Torrent et al., 2011; Xiao et al., 2013). Ambos conjuntos contienen secuencias peptídicas en formato de texto.
- El conjunto de validación consiste en 712 secuencias peptídicas reportadas como antimicrobianas, y 712 secuencias que no han sido reportado como antimicrobianas

obtenidas del sitio de la herramienta Antimicrobial Peptide Scanner, de Veltri et al (Veltri et al, 2018) ([https://www.dveltri.com/ascan/v2/data/AMP\\_Scan2\\_OrigPaper\\_Dataset.zip](https://www.dveltri.com/ascan/v2/data/AMP_Scan2_OrigPaper_Dataset.zip))

- Secuencias peptídicas de la base de datos Pfam sobre las que se realizó la predicción de función antimicrobiana. <https://pfam.xfam.org/>.

## Métodos

### Predicción de péptidos antimicrobianos

Para la predicción de péptidos antimicrobianos, se utilizó el software desarrollado por nuestro grupo de trabajo (Diener et al, 2016), el cual implementa el algoritmo de bosque aleatorio (*random forest*), con una validación cruzada tipo k-fold, con un valor de k igual a 4. El código original fue modificado, para que presentara no solo la probabilidad promedio por familia de proteínas, sino también para cada proteína. Originalmente el predictor arrojaba la probabilidad de cada familia de ser AMP. Para esto, se sumaba la probabilidad de ser AMP de cada péptido de la familia, y el valor obtenido se dividía entre la cantidad de péptidos que conformaban a la familia. Sin embargo, para este trabajo, se modificó para que el predictor también arrojara la probabilidad de ser AMP de cada péptido.

Se entrenó a un clasificador, para que lograra identificar secuencias peptídicas capaces de ser antimicrobianas. Para entrenar el clasificador, se usaron dos conjuntos de entrenamiento: uno con 843 secuencias peptídicas reportadas como antibacterianas en la base de datos APD (*antimicrobial peptide database*, <https://wangapd3.com/>) (Wang et al, 2016) y otro con 843 secuencias para las cuales no se ha reportado actividad antimicrobiana, o no poseen propiedades fisicoquímicas de péptidos antimicrobianos.

Después de haber generado el modelo de clasificación, se decidió correrlo contra un conjunto de datos de evaluación que consiste en 712 secuencias peptídicas reportadas como antimicrobianas, y 712 secuencias que no han sido reportado como antimicrobianas obtenidas del sitio de la herramienta Antimicrobial Peptide Scanner, de Veltri et al (Veltri et al, 2018).

Con el modelo entrenado, se tomaron los péptidos que serían clasificados, que provenían de regiones conservadas de las 16035 familias de proteínas, de la base de datos Pfam (Finn et al, 2016). Pfam contiene una gran colección de familias de proteínas que están representadas por alineamientos de secuencia múltiples.

Los resultados arrojados por este algoritmo consisten en valores de probabilidad, con un rango

de 0 a 1: un valor de 0 corresponde a un no-AMP, y un valor de 1 a un AMP. Por lo tanto, una probabilidad cercana a 1 indica que el péptido tiene potencial de ser antimicrobiano, mientras que un valor cercano a 0 dice que el péptido no tiene capacidad de ser AMP. La secuencia seleccionada, además de tener una alta probabilidad de ser antimicrobiana, también debía ser de *Saccharomyces cerevisiae*, ya que en el grupo del laboratorio se trabaja con este microorganismo, y se tenía planeado extraer el péptido a partir de esta levadura.

### **Ensayos de crecimiento bacteriano**

Para comprobar la capacidad antimicrobiana del péptido seleccionado, se realizaron experimentos de crecimiento bacteriano. Para cada bacteria se siguió el siguiente procedimiento. Se tomó una colonia de la cepa y se dejó crecer en medio LB, por un periodo de 12 horas. Transcurrido este tiempo, se tomó una alícuota del inóculo y se diluyó hasta obtener una concentración de  $8.0 \times 10^5$  células/mL. En una placa transparente, en cada pozo se colocaron 100 $\mu$ L de inóculo diluido y 50 $\mu$ L de solución de péptido, a diferentes concentraciones. Como control positivo se utilizó ampicilina. Para cada ensayo, se realizaron mediciones de OD 600nm por cada hora, durante 24 horas, en el equipo Synergy, a una temperatura de 37 °C. Se realizaron dos experimentos independientes por triplicado.

Las bacterias utilizadas para estos ensayos fueron: *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus*. Se escogieron estas cepas debido a que son bacterias que han desarrollado altos niveles de resistencia a antibióticos en los últimos años (Pourmand et al 2017; Caruso, 2018).

### **Evaluación de la habilidad de penetración celular de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera. Preparación del inóculo**

Para cada haplotipo, en un tubo de ensayo se añadieron 5 mL de medio y una colonia de levadura. El inóculo se dejó crecer durante 14 horas. Transcurrido este tiempo, se realizaron las diluciones necesarias para que el inóculo alcanzara una densidad óptica de 0.4 para experimentos en microscopía y 1.0 para mediciones en el AMNIS).

### **Medición de la internalización péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera por microscopía confocal de fluorescencia**

Se agregaron 13.5  $\mu$ L de inóculo de levadura y 1.5  $\mu$ L de solución de péptido en un portaobjetos de 26.4x76.2x1.2 mm, el cual se cubrió con un cubreobjetos de 22x22x1.5 mm. La muestra se observó bajo un microscopio confocal multifotón Olympus FV100, y se tomaron imágenes a 0 y 10 minutos después de haber colocado el portaobjetos. Todas las imágenes se tomaron con un apilamiento de enfoque (z-stack) de 10 niveles. Las imágenes se analizaron mediante el uso de Fiji.

### **Medición de la internalización del péptido $\alpha$ -NLS-C en MAT $\alpha$ y MAT $\alpha$ por microscopía confocal de fluorescencia**

En un portaobjetos, se colocaron 12.5  $\mu$ L de inóculo de levadura, 1.5  $\mu$ L de solución de péptido 50  $\mu$ M y 1  $\mu$ L de solución de agarosa. Inmediatamente después de haber agregado solución de péptido, se realizaron las tomas de imágenes mediante un microscopio confocal (Zeiss LSM510 META).

Para cada muestra colocada en el microscopio, se realizó una exploración visual, hasta localizar un conjunto de células que fueran favorables para su observación (células sin movimiento que comenzaran a ser internalizadas por el péptido). Después, se tomaron imágenes de las células localizadas, por intervalos de 3 minutos, durante una hora, con un apilamiento de enfoque de 10 niveles. Para realizar el análisis de las imágenes obtenidas y el cálculo de los valores de fluorescencia de cada célula, se utilizó el paquete computacional Fiji (Schindelin et al, 2012). Usando este software, se determinó la intensidad de fluorescencia de los pixeles de la célula. Para obtener valores más precisos, la fluorescencia calculada se dividió entre el área de la célula. Todos estos experimentos fueron realizados en el Laboratorio Nacional de Microscopía Avanzada localizado en el Instituto de Biotecnología de la UNAM.

### **Medición de la cinética de internalización de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera por citometría de flujo**

En estos experimentos, se utilizó un equipo de citometría de flujo acoplado a microscopio (Citómetro de flujo-Microscopio Amnis ImageStream Mark II). Para las mediciones, en un tubo eppendorf de 1.5  $\mu$ L se añadieron 116  $\mu$ L del inóculo de células de levadura y 4  $\mu$ L de péptido a concentración de 200 M. A continuación, se colocó el tubo en el Amnis y se realizaron las mediciones de fluorescencia por célula.

Los parámetros usados para los experimentos fueron las siguientes: contabilización de 10,000 eventos(células) por experimento por intervalos de cinco minutos, en un lapso de 60 minutos; enfoque de células con valor superior al 50%; un área entre 20 y 50  $\mu\text{m}$  y una razón de aspecto (*aspect ratio*) mayor a 0.7. Esta razón de aspecto corresponde a la razón entre el largo y el ancho de la célula. Entonces, una razón de aspecto cercana a 1.0 indicará que la célula es redonda y que se trata de una sola célula; por otra parte, valores bajos indicarían posiblemente aglomerado de células, o en estado de gemación.

Se realizaron dos tipos de experimentos: para cepas individuales y cepas mezcladas. Para este último caso, se debe tener alguna clase de marcador en alguna de las cepas, pues de lo contrario, no se podría determinar si las mediciones corresponden a una cepa o a la otra. Con base en esto, para estos experimentos se utilizó una cepa de MATa, en la cual, la proteína ribosomal RPL9A está unida a una proteína verde fluorescente (GFP). Para los experimentos de cepas individuales, la fuente de excitación fue un láser a una longitud de onda de 561 nm, que corresponde a fluorescencia en color rojo (por el fluoróforo TAMRA) a 100% de intensidad; y para los de cepas mezcladas, además de la fuente mencionada, se utilizó otro láser a 488 nm, ambos a 100% de intensidad.

El análisis de los datos se llevó a cabo usando el software IDEAS ( Image Data Exploration and Analysis Software, Amnis Imaging Flow Cytometry). Lo primero que se hizo fue seleccionar el área de la fluorescencia en las células, que en el software se conoce como máscara. La máscara seleccionada por defecto, corresponde a la de células en campo claro; por lo tanto, no cubre correctamente el área que corresponde a la fluorescencia. Para obtener una mayor cuantificación de la fluorescencia, se cambió a una máscara para el canal que detecta fluorescencia en rojo. Luego se realizó un filtrado de las mediciones y se recolectaron solamente aquellas que presentaran un cierto valor de intensidad de fluorescencia.

Después, las mediciones de las células fluorescentes se graficaron para cada cepa, a los diferentes intervalos, usando el paquete computacional Matplotlib.

## Resultados y discusión

### Predicción de péptidos antimicrobianos

Para la clasificación de péptidos, se utilizó el algoritmo de bosque aleatorio y se entrenó a un clasificador, a fin de que lograra identificar secuencias peptídicas capaces de ser antimicrobianas. Los resultados obtenidos por el algoritmo consisten en valores de probabilidad para ser péptido antimicrobiano, para cada una de las familias, por lo que se obtuvieron 16035 probabilidades. De los resultados arrojados por el algoritmo, se obtuvo un puntaje de precisión del 90.4%.

Para evaluar que el predictor efectivamente sea capaz de identificar AMPs con una precisión del 90%, se decidió correr el algoritmo contra un conjunto de datos de evaluación que consiste en 712 secuencias peptídicas reportadas como antimicrobianas, y 712 secuencias que no han sido reportado como antimicrobianas obtenidas del sitio de la herramienta Antimicrobial Peptide Scanner, de Veltri et al (Veltri et al, 2018). En la figura 5.1, se muestra la curva Característica Operativa del Receptor (curva ROC, Receiver Operating Characteristic) de la evaluación. El área bajo la curva es de 0.94, lo que nos indica que el modelo de predicción de este trabajo sí predice con el porcentaje de certeza del 90%.

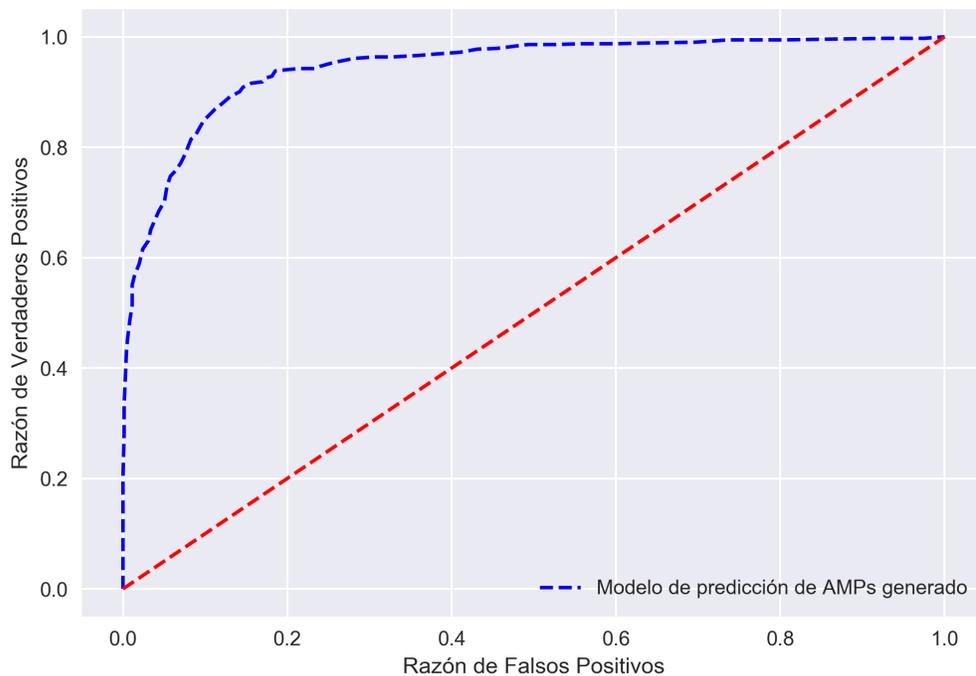


Figura 5.1 Curva Característica Operativa del Receptor (ROC en inglés) del modelo de predicción de péptidos antimicrobianos generado para este trabajo, usando un conjunto de datos de evaluación.

De las secuencias peptídicas que fueron predichas como AMPs, se seleccionó a RPL29 de *Saccharomyces cerevisiae*, debido a que obtuvo una probabilidad de ser AMP de 0.8, y se sintetizó químicamente. RPL29 consta de 59 aminoácidos y se encuentra formando parte de la subunidad 60S del ribosoma de *Saccharomyces cerevisiae*. En la figura 5.2 se muestra la secuencia y proyección helicoidal de aminoácidos de este péptido. De acuerdo con la figura 5.2b, en la hélice que se forma entre los aminoácidos 38-55, se observa una zona polar cargada positivamente en la sección norte (dos residuos de lisina, dos de arginina). También, se encuentran en la otra cara de la hélice aminoácidos hidrofóbicos (dos alaninas, dos leucinas y una fenilalanina), lo cual corresponde a características que poseen los péptidos antimicrobianos.

Adicionalmente, la secuencia de RPL29 se validó usando otros servidores que también predicen péptidos antimicrobianos. Entre estos, iAMPpred indicó una probabilidad de 0.94 de RPL29 de ser antibacteriana. También, se utilizó el predictor AmpGram, el cual predijo a RPL29 como AMP con una probabilidad de 0.84.

## RPL29

a) MAKSKNHTAHNQTRKAHRNGIKKPKTYKYPSLKGVDPKFRRNHHKHALHGTAALAAAKK

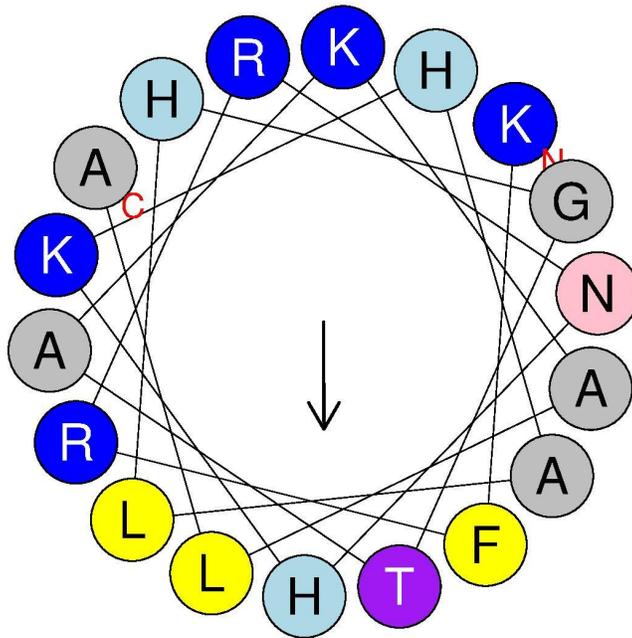


Figura 5.2. a) Secuencia de aminoácidos de RPL29 y b) Proyección helicoidal de aminoácidos 38-55 de RPL29 obtenida de HELIQUEST (Gauntier et al, 2008).

### Crecimiento bacteriano en presencia de RPL29

Para comprobar la capacidad antimicrobiana del péptido RPL29, se realizaron experimentos de crecimiento bacteriano. Se utilizaron cepas de *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus*. Para estos ensayos, se utilizaron concentraciones de péptido de 4, 8, 12 y 16  $\mu\text{M}$ .

En las figura 5.3 se muestran las áreas bajo la curva para los experimentos de crecimiento de *E.coli*, *P. aeruginosa* y *S. aureus* respectivamente, en presencia de RPL29. Se observa que, para todas las bacterias, las curvas de crecimiento para los ensayos con RPL29 son similares a las del grupo control (agua destilada). Estos resultados indican que el péptido no presenta actividad antimicrobiana, tanto para bacterias gram positiva, como para gram negativa.

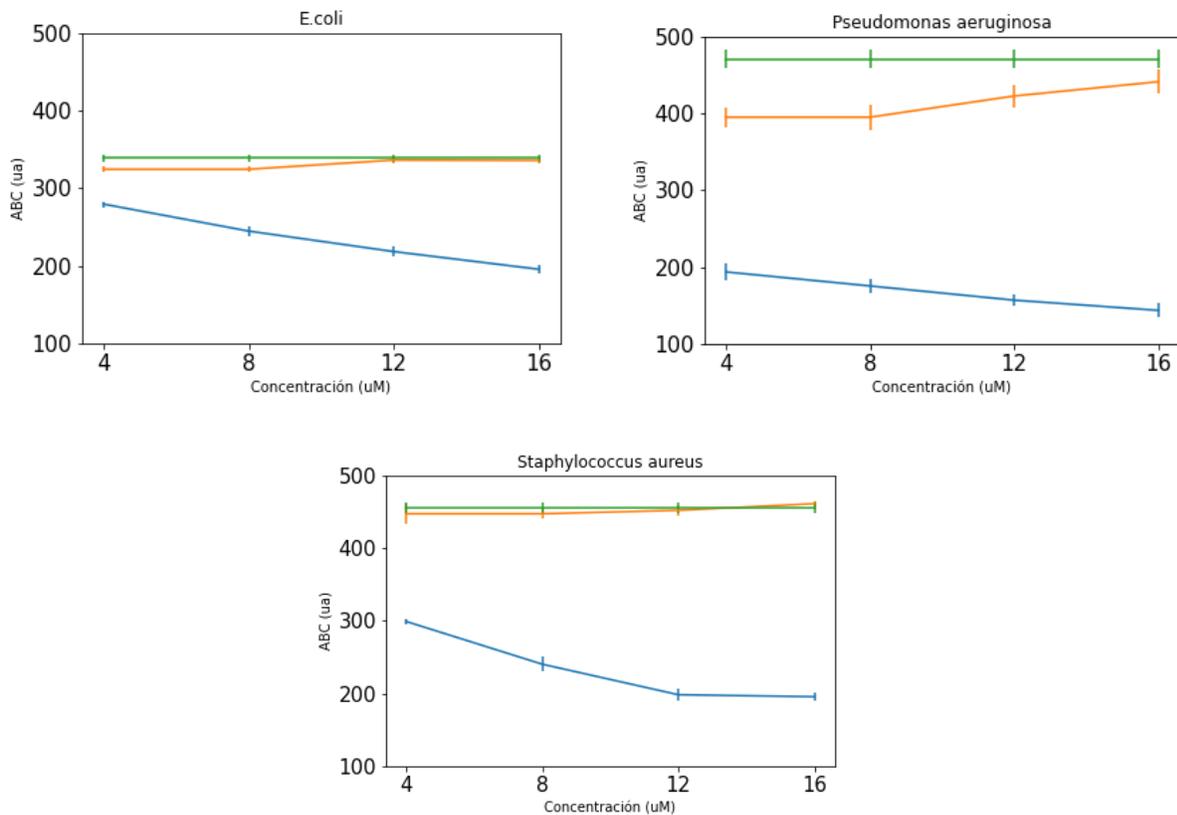


Figura 5.3. Areas bajo de la curva obtenidas de experimentos de crecimiento bacteriano de *Escherichia coli*, *Pseudomonas aeruginosa* y *Staphylococcus aureus* en presencia de diferentes agentes: RPL29 (amarillo), ampicilina (azul), agua destilada (verde).

Se ha reportado que algunas proteínas ribosomales son capaces de ser antimicrobianos. Fernandes y Smith encontraron actividad antibacteriana en la ribonucleoproteína 40S RpS30 de células eucariotas (Fernandes, J y Smith, V, 2002). Ma y colaboradores encontraron que la proteína ribosomal S23 es capaz de matar bacterias tanto gram-positivas como gram-negativas (Ma, Z. et al, 2020). Por último Pütsep y colaboradores, reportaron que en el extremo N-terminal de la proteína ribosomal L1 de *Helicobacter pylori* posee actividad antibacteriana (Pütsep et al, 1999). Por lo tanto, se pensó que RPL29 podría tener actividad antimicrobiana. Sin embargo, RPL29 no mostró capacidad para inhibir el crecimiento de *Escherichia coli*, ni el de *Pseudomonas aeruginosa* o *Staphylococcus aureus*.

Es posible que el conjunto de datos de entrenamiento negativo, es decir, los datos de péptidos no antimicrobianos, no haya sido eficaz. Los conjuntos de entrenamiento negativos usados para la generación de modelos de predicción de péptidos antimicrobianos, usualmente se

obtienen de la base de datos Uniprot, extrayendo secuencias para las cuales no se ha reportado actividad antimicrobiana, o que no poseen propiedades fisicoquímicas de péptidos antimicrobianos (Veltri et al, 2018; Burdukiewicz et al, 2020). Sin embargo, es posible que estas secuencias usadas como conjunto de datos de entrenamiento negativo no aporten suficiente información de categorización al modelo.

### **Evaluación de la habilidad de penetración celular de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera**

Para comprobar que los péptidos diseñados poseen la capacidad de introducirse en células, se observaron células MATa de *Saccharomyces cerevisiae* en presencia de  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera en un microscopio confocal de fluorescencia. Para poder visualizar el proceso de internalización, se agregó una molécula de TAMRA covalentemente, en el extremo amino de cada péptido. Además, para cada péptido, se tomaron micrografías: una inmediatamente después de haber colocado el inóculo de levadura junto con el péptido, y otra 10 minutos después. En estos experimentos, se usaron como control estos tres mismos péptidos, pero con los aminoácidos arginina y lisina sustituidos por glutamato. A estos péptidos se les llamó woRK (WithOut aRginine or Lysine).

Tal como se observa en la figura 5.4, los péptidos diseñados como CPPs logran introducirse en las células de levadura desde el inicio de la toma de micrografías. Este fenómeno se observa de manera más clara para  $\alpha$ -NLS-C y NLS- $\alpha$ -CE. Transcurridos diez minutos después de la primera toma de imágenes, algunas células presentan mayor fluorescencia. En cambio, para los péptidos woRK, no se observa este mismo fenómeno; al inicio del experimento, y a los 10 minutos transcurridos, no se detecta fluorescencia en las células MATa de levadura.

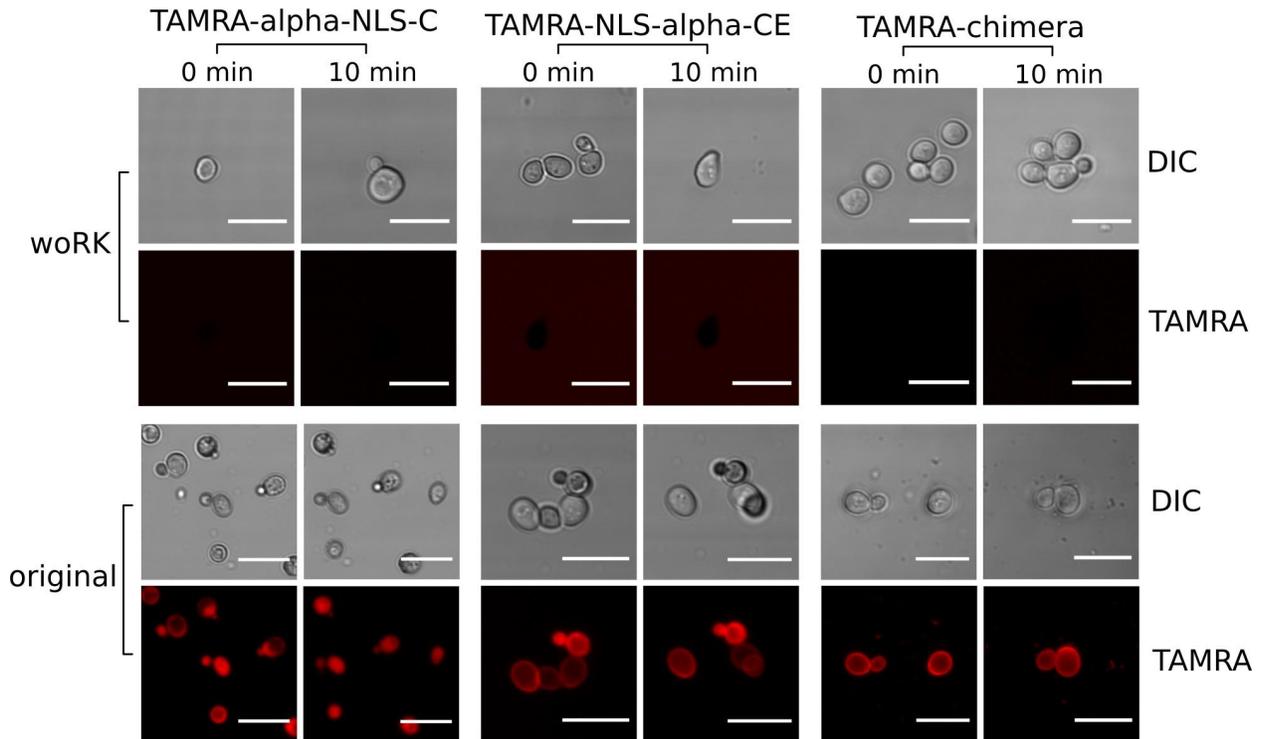


Figura 5.4. Micrografías de la internalización de los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera en células de levadura con haplotipo MAT $\alpha$ . En la sección inferior (original), se encuentran imágenes de microscopio confocal que indican que los 3 péptidos se internalizan dentro de las células de levadura. Para comprobar que los péptidos no se introducían de forma aleatoria, se utilizaron los péptidos woRK; en esta sección, no se observa penetración por parte de los péptidos.

### Evaluación de la selectividad celular de los péptidos $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera

En la figura 5.5 se muestran las imágenes obtenidas por microscopía confocal de fluorescencia, de los experimentos de internalización del péptido  $\alpha$ -NLS-C a 5  $\mu$ M, en los haplotipos MAT $\alpha$  y MAT $\alpha$ . En los paneles a) y b), se muestran las micrografías de MAT $\alpha$  y MAT $\alpha$ , respectivamente, al inicio del experimento. En los paneles c) y d), se muestran gráficas que indican la fluorescencia de cada célula observada en las imágenes del microscopio, a intervalos de 3 minutos, por un lapso de 60 minutos. Desde el inicio, de acuerdo con las gráficas de fluorescencia de MAT $\alpha$  y MAT $\alpha$ , es notoria la diferencia de valores entre ambos haplotipos. Además, el valor máximo de fluorescencia de estas últimas es hasta un orden de magnitud menor que el de las células MAT $\alpha$ . Las imágenes corresponden a un conjunto de células que se observó que presentaban fluorescencia.

A pesar de haber observado que en células MAT $\alpha$  existe una mayor intensidad de fluorescencia, estos resultados no pueden ser considerados como evidencia definitiva para concluir que los péptidos diseñados como CPPs son selectivos para MAT $\alpha$ , pues solamente se analizó un grupo muy reducido de células (7 de MAT $\alpha$  y 9 de MAT $\alpha$ ). Entonces, para lograr una medición cuantitativa de este fenómeno, se requiere de otra técnica que determine la fluorescencia de una población de células expuestas al péptido, sin perder la información detallada de cada célula que proporciona la microscopía. Por estas razones, se decidió utilizar un equipo de citometría de flujo acoplado a microscopía de fluorescencia.

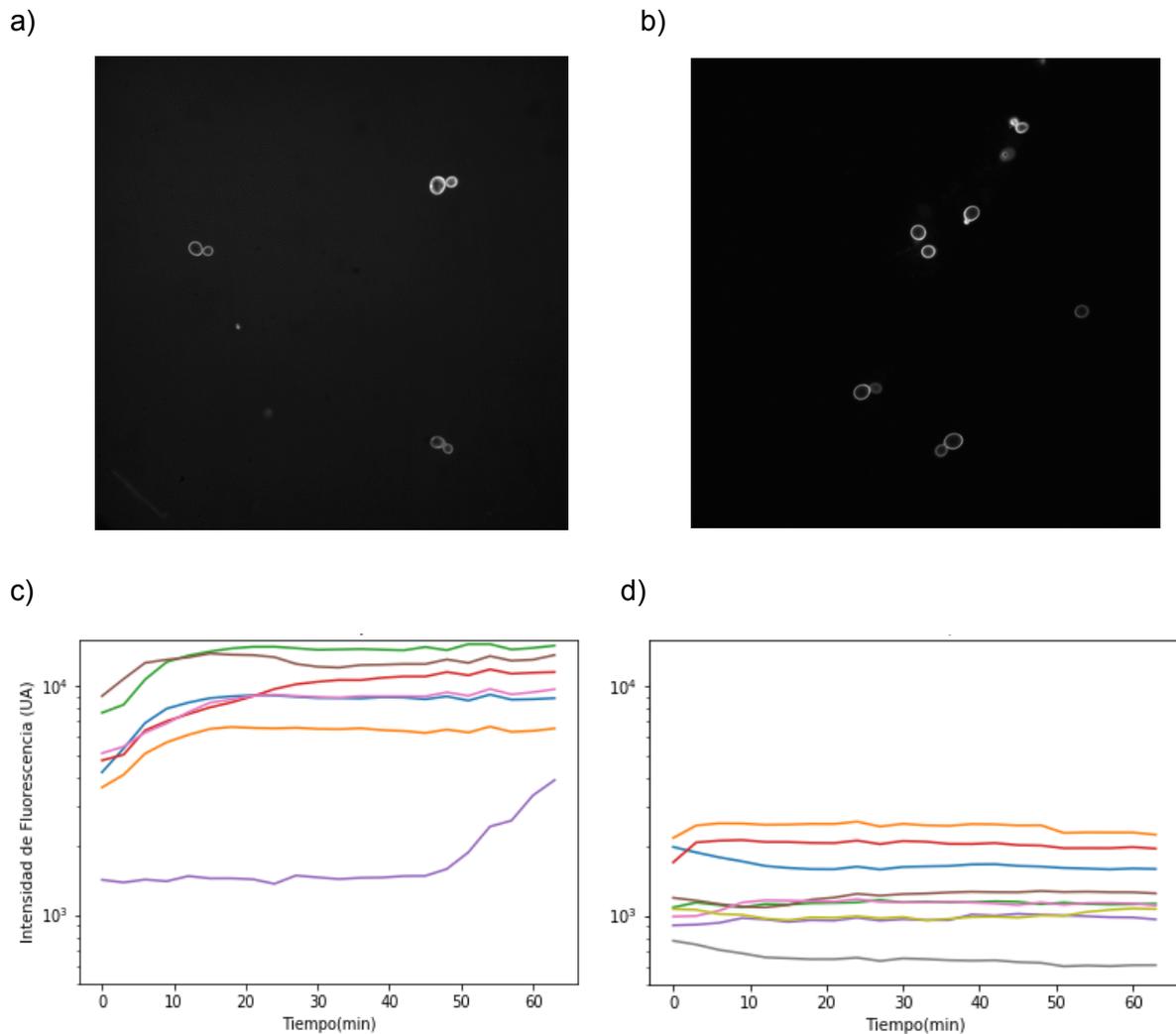


Figura 5.5: Internalización del péptido  $\alpha$ -NLS-C en *S. cerevisiae*. En a) y b) se presentan células MAT $\alpha$  y MAT $\alpha$ , respectivamente, al inicio del experimento. En c) y d) se observan los valores de intensidad de cada una de las células captadas con el microscopio, durante 60 minutos para MAT $\alpha$  y MAT $\alpha$ .

En el caso de mediciones en cepas expuestas a cada péptido de forma independiente, que se

muestran en la figura 5.6, se observa que para el péptido  $\alpha$ -NLS-C y NLS- $\alpha$ -CE, las tres cepas siguen el mismo comportamiento a lo largo del experimento, pues a los primeros minutos, gran cantidad de células se encuentran en niveles bajos de intensidad de fluorescencia. Conforme avanza el tiempo, el tamaño de la caja comienza a incrementar. Este incremento en la amplitud a valores altos de fluorescencia, indica que una mayor cantidad de péptido se encuentra internalizada dentro de las células.

Otro resultado que se observa, es que las cepas con la delección del gen BAR1 registraron una mayor intensidad de fluorescencia que las demás, para  $\alpha$ -NLS-C y Chimera. En el genotipo de MATa, se encuentra un gen que codifica para una proteína llamada BAR1, una proteasa que se encarga de degradar al factor  $\alpha$ . Para investigar el efecto que BAR1 tendría sobre la internalización de los péptidos, se realizaron experimentos de citometría de flujo acoplada a microscopía de fluorescencia, durante un lapso de una hora, en células MATa $\Delta$ BAR1. Por lo tanto, si no se encuentra la proteasa presente, se esperaría que no haya una regulación sobre la internalización de secuencias peptídicas que porten el factor  $\alpha$ . Por esta razón, la fluorescencia es mayor en las células MATa $\Delta$ BAR1.

En experimentos en los que las cepas MATa y MAT $\alpha$  se presentaron de forma conjunta a cada péptido, la distinción entre distribución de intensidades de fluorescencia es totalmente diferente. Como se observa en la figura 5.7, los tres péptidos se internalizan preferentemente en las células MATa conforme avanza el tiempo. En cambio, durante los 60 minutos, las células MAT $\alpha$  se mantienen que se basa en en el mismo rango de intensidad de fluorescencia.

Tal como se esperaba, en los experimentos de citometría de flujo acoplada a microscopía de fluorescencia (AMNIS) se observó que la población de células MATa tiende a niveles de intensidad de fluorescencia mayores que el conjunto de células MAT $\alpha$ , lo que podría indicar mayor internalización de péptido. Estos resultados sugieren que el receptor Ste2p del haplotipo MATa posiblemente reconoce la secuencia del factor  $\alpha$  que se encuentra en el péptido.

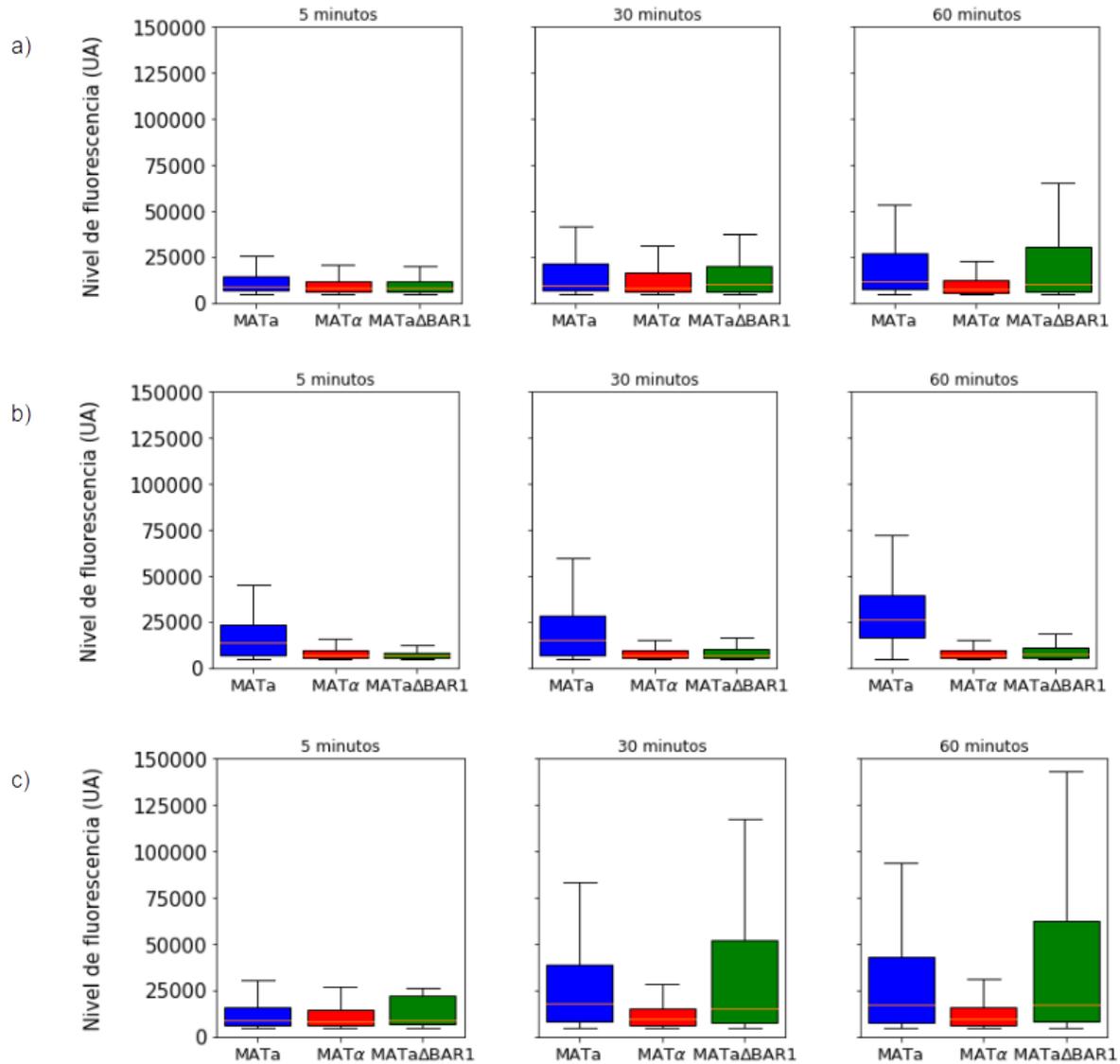


Figura 5.6: Niveles de fluorescencia celular de inóculos que contienen uno de los haplotipos de célula en un lapso de 60 minutos. Se utilizó un péptido diferente para cada experimento: a) TAMRA-α-NLS-C, b) TAMRA-NLS-α-CE y c) TAMRA-Chimera. Las gráficas color azul representan a resultados de MATa, las rojas a MATα y las verdes a MATaΔBAR1.

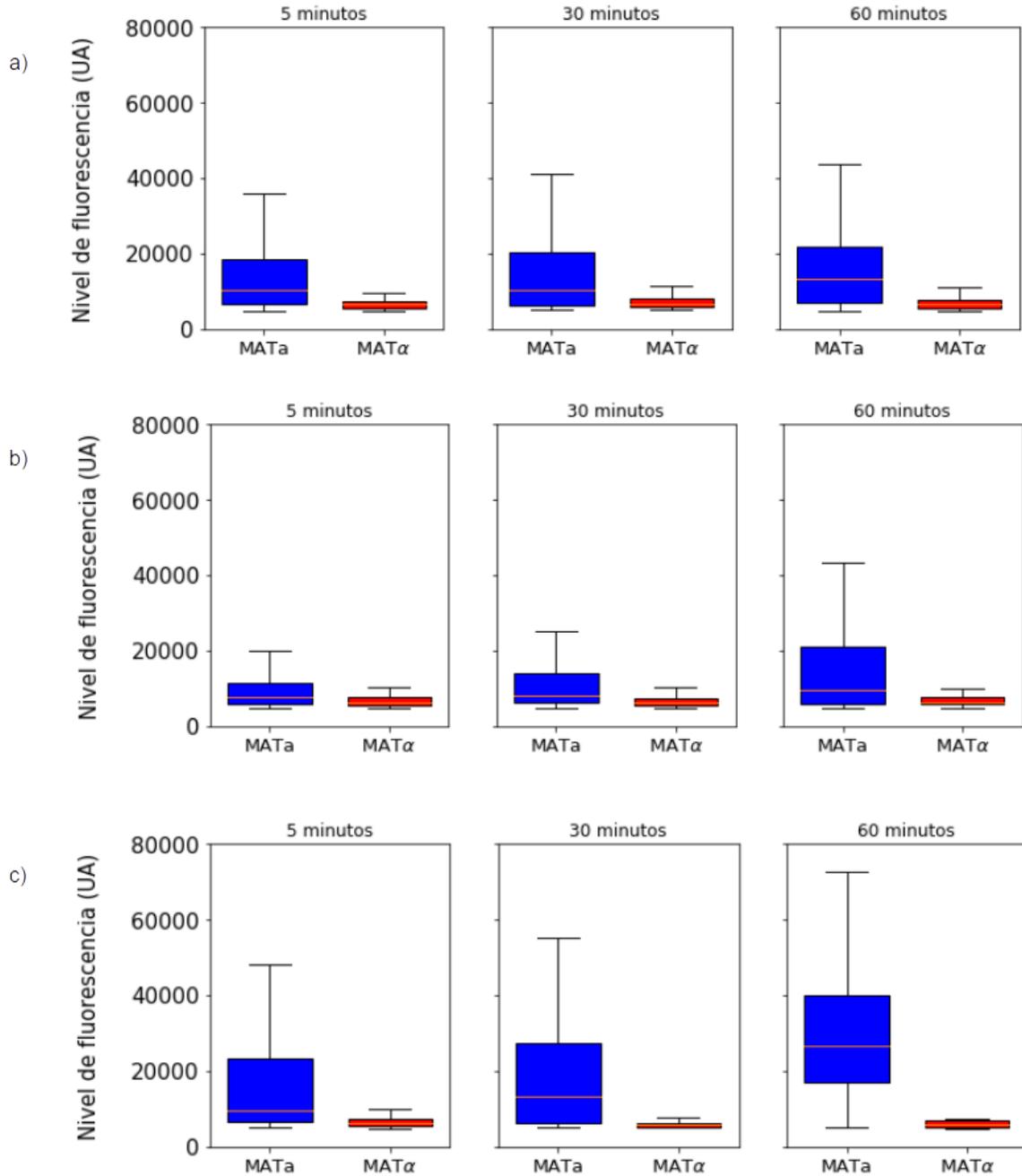


Figura 5.7: Niveles de fluorescencia celular de inóculos que contienen ambos haplotipos de *S. cerevisiae*, para medir la tasa de internalización de los péptidos penetradores de célula en un lapso de 60 minutos. Se utilizó un péptido diferente para cada experimento: a) TAMRA- $\alpha$ -NLS-C, b) TAMRA-NLS- $\alpha$ -CE y c) TAMRA-Chimera. Las gráficas en rojo representan las fluorescencias de MATa y las azules a MAT $\alpha$ .

## Conclusiones

RPL29, una secuencia de 59 aminoácidos, de la levadura *Saccharomyces cerevisiae*, fue predicha como potencial AMP por el software de aprendizaje automatizado Dcf. A pesar de que el software de predicción de péptidos antimicrobianos alcanzó una precisión del 90%, en los ensayos de actividad antimicrobiana, la presencia de RPL29 no tuvo efecto alguno en el crecimiento de bacterias gram negativas (*Escherichia coli*, *Pseudomonas aeruginosa*) o gram positivas (*Staphylococcus aureus*).

Adicionalmente, nuestros resultados muestran que los péptidos  $\alpha$ -NLS-C, NLS- $\alpha$ -CE y Chimera, diseñados previamente en el laboratorio para ser CPPs en *Saccharomyces cerevisiae*, efectivamente logran introducirse en las células de la levadura y presentan una tendencia a internalizarse un poco más en el haplotipo MATa, lo que sugiere que el receptor Ste2p del haplotipo MATa posiblemente reconoce la secuencia del factor  $\alpha$  que se encuentra en el péptido.

# Referencias

Ahmed, A., Siman-Tov, G., Hall, G., Bhalla, N., & Narayanan, A. (2019). Human antimicrobial peptides as therapeutics for viral infections. *Viruses*, 11(8), 704.

Bardwell L (2005). A walk-through of the yeast mating pheromone response pathway. *Peptides* 26: 339-350.

Bhadra, P., Yan, J., Li, J., Fong, S., & Siu, S. (2018). AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Scientific reports*, 8(1), 1697.

Bierbaum, G., & Sahl, H. G. (1985). Induction of autolysis of staphylococci by the basic peptide antibiotics Pep 5 and nisin and their influence on the activity of autolytic enzymes. *Archives of microbiology*, 141(3), 249-254.

Bradshaw, J.P., (2003). Cationic Antimicrobial Peptides- Issues for Potential Clinical Use. *Biodrugs* 17, 233-240.

Burdukiewicz, M., Sidorcuk, K., Rafacz, D., Pietluch, F., Chilimoniuk, J., Rödiger, S., & Gagat, P. (2020). Proteomic screening for prediction and design of antimicrobial peptides with AmpGram. *International journal of molecular sciences*, 21(12), 4310.

Burkholder AC y Hartwell LH (1985). The yeast alpha-factor receptor: structural properties deduced from the sequence of the STE2 gene. *Nucleic Acids Res* 13(23):8463-75.

Caruso G. (2018). Antibiotic Resistance in *Escherichia coli* from Farm Livestock and Related Analytical Methods: A Review. *Journal of AOAC International*, 101(4), 916–922.

Cherkasov, A., Hilpert, K., Jenssen, H., Fjell, C. D., Waldbrook, M., Mullaly, S. C., Volkmer, R., & Hancock, R. E. (2009). Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. *ACS chemical biology*, 4(1), 65–74.

Cielen D, Myesman A, Ali M (2016). *Introducing data science big data, machine learning, and more, using python tools*. Manning Publications Co.

Criminisi A., Shotton J., and Konukoglu E (2011). Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical Report MSR-TR-2011-114, Microsoft Research, Ltd., Cambridge, UK.

Derossi, D., Joliot, A. H., Chassaing, G., & Prochiantz, A. (1994). The third helix of the Antennapedia homeodomain translocates through biological membranes. *Journal of Biological Chemistry*, 269(14), 10444-10450.

Diao, Y., Han, W., Zhao, H., Zhu, S., Liu, X., Feng, X., Gu, J., Yao, C., Liu, S., Sun, C., & Pan, F. (2012). Designed synthetic analogs of the  $\alpha$ -helical peptide temporin-La with improved antitumor efficacies via charge modification and incorporation of the integrin  $\alpha v \beta 3$  homing domain. *Journal of peptide science : an official publication of the European Peptide Society*, 18(7), 476–486.

Diener, C., Garza Ramos Martínez, G., Moreno Blas, D., Castillo Gonzalez, D. A., Corzo, G., Castro-Obregon, S., & Del Rio, G. (2016). Effective design of multifunctional peptides by combining compatible functions. *PLoS computational biology*, 12(4), e1004786.

Endoh, T., & Ohtsuki, T. (2009). Cellular siRNA delivery using cell-penetrating peptides modified for endosomal escape. *Advanced drug delivery reviews*, 61(9), 704-709.

Engbrecht J (2003) Cell signaling in yeast sporulation. *Biochem Biophys Res Commun*. 306:325-328.

Fernandes, J. M., & Smith, V. J. (2002). A novel antimicrobial function for a ribosomal peptide from rainbow trout skin. *Biochemical and biophysical research communications*, 296(1), 167-171.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... & Salazar, G. A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*, 44(D1), D279-D285.

Fischer, R., Waizenegger, T., Köhler, K., & Brock, R. (2002). A quantitative validation of fluorophore-labelled cell-permeable peptide conjugates: fluorophore and cargo dependence of import. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1564(2), 365-374.

Fjell, C. D., Jenssen, H., Hilpert, K., Cheung, W. A., Panté, N., Hancock, R. E., & Cherkasov, A. (2009). Identification of novel antibacterial peptides by chemoinformatics and machine learning. *Journal of medicinal chemistry*, 52(7), 2006–2015.

Frankel, A. D., & Pabo, C. O. (1988). Cellular uptake of the tat protein from human immunodeficiency virus. *Cell*, 55(6), 1189–1193.

Gautier, R., Douguet, D., Antony, B., & Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific  $\alpha$ -helical properties. *Bioinformatics*, 24(18), 2101-2102.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly.

Giguère, S., Laviolette, F., Marchand, M., Tremblay, D., Moineau, S., Liang, X., Biron, É., & Corbeil, J. (2015). Machine learning assisted design of highly active peptides for drug discovery. *PLoS computational biology*, 11(4), e1004074.

Guidotti, G., Brambilla, L., & Rossi, D. (2017). Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends in pharmacological sciences*, 38(4), 406–424.

Hagen DC, et al. (1986) Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone a factor: gene sequence and implications for the structure of the presumed receptor. *Proc Natl Acad Sci U S A* 83(5):1418-22.

Hancock, R. E., & Chapple, D. S. (1999). Peptide antibiotics. *Antimicrobial agents and chemotherapy*, 43(6), 1317-1323.

Hof, W. V. T., Veerman, E. C., Helmerhorst, E. J., & Amerongen, A. V. N. (2001). Antimicrobial peptides: properties and applicability.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.

Izadpanah, A., & Gallo, R. L. (2005). Antimicrobial peptides. *Journal of the American Academy of Dermatology*, 52(3), 381-390.

Järver, P., & Langel, Ü. (2006). Cell-penetrating peptides—a brief introduction. *Bba-Biomembranes*, 3(1758), 260-263.

Koebach, J., & Craik, D. J. (2019). The vast structural diversity of antimicrobial peptides. *Trends in pharmacological sciences*, 40(7), 517-528.

Langel, Ü., Pooga, M., Kairane, C., Zilmer, M., & Bartfai, T. (1996). A galanin-mastoparan chimeric peptide activates the Na<sup>+</sup>, K<sup>+</sup>-ATPase and reverses its inhibition by ouabain. *Regulatory peptides*, 62(1), 47-52.

Lata, S., Sharma, B. K., & Raghava, G. P. (2007). Analysis and prediction of antibacterial peptides. *BMC bioinformatics*, 8, 263.

Lide, D. R. (Ed.). (2004). *CRC Handbook of Chemistry and Physics*, 72nd Edition, CRC Press, 1991.

Maccari, G., Di Luca, M., Nifosí, R., Cardarelli, F., Signore, G., Boccardi, C., & Bifone, A. (2013). Antimicrobial peptides design by evolutionary multiobjective optimization. *PLoS computational biology*, 9(9), e1003212.

Magana, M., Pushpanathan, M., Santos, A. L., Leanse, L., Fernandez, M., Ioannidis, A., & Cherkasov, A. (2020). The value of antimicrobial peptides in the age of resistance. *The Lancet Infectious Diseases*.

Magrane, M., consorcio UniProt (2011). UniProt knowledgebase: a hub of integrated protein data. *Database*, 2011, bar009.

Maróti, G., Kereszt, A., Kondorosi, E., & Mergaert, P. (2011). Natural roles of antimicrobial peptides in microbes, plants and animals. *Research in microbiology*, 162(4), 363-374.

Matsuzaki K (2009) Control of cell selectivity of antimicrobial peptides. *Biochim Biophys Acta* 1788:687-1692.

Matsuzaki, K. (1999). Why and how are peptide–lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1462(1-2), 1-10.

Meher, P. K., Sahu, T. K., Saini, V., & Rao, A. R. (2017). Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Scientific reports*, 7, 42362.

Milletti, F. (2012). Cell-penetrating peptides: classes, origin, and current landscape. *Drug Discovery Today*, 17(15-16), 850–860.

Mookherjee, N., Anderson, M. A., Haagsman, H. P., & Davidson, D. J. (2020). Antimicrobial host defence peptides: Functions and clinical potential. *Nature reviews Drug discovery*, 19(5), 311-332.

Morris, M. C., Vidal, P., Chaloin, L., Heitz, F., & Divita, G. (1997). A new peptide vector for efficient delivery of oligonucleotides into mammalian cells. *Nucleic acids research*, 25(14), 2730–2736.

Morris, M. C., Depollier, J., Mery, J., Heitz, F., & Divita, G. (2001). A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nature biotechnology*, 19(12), 1173–1176.

Pace, C. Nick; Scholtz, J. Martin (1998). A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal*. 75. pp. 422–427.

Patel, L. N., Zaro, J. L., & Shen, W. C. (2007). Cell penetrating peptides: Intracellular pathways and pharmaceutical perspectives. *Pharmaceutical Research*.

Pooga, M., Hällbrink, M., Zorko, M., & Langel, U. (1998). Cell penetration by transportan. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 12(1), 67–77.

Pooga, M., & Langel, Ü. (2015). Classes of Cell-Penetrating Peptides. *Methods in molecular biology (Clifton, N.J.)*, 1324, 3–28.

Pourmand, A., Mazer-Amirshahi, M., Jasani, G., & May, L. (2017). Emerging trends in antibiotic resistance: implications for emergency medicine. *The American journal of emergency medicine*, 35(8), 1172-1176.

Qi, Y. (2012). Random forest for bioinformatics. In *Ensemble machine learning* (pp. 307-323). Springer, Boston, MA.

Raffatellu, M. (2018). Learning from bacterial competition in the host to develop antimicrobials. *Nature medicine*, 24(8), 1097-1103.

Ragin, A. D., Morgan, R. A., & Chmielewski, J. (2002). Cellular import mediated by nuclear localization signal peptide sequences. *Chemistry & biology*, 9(8), 943-948

Raschka, S. (2015). *Python Machine Learning*. Packt Publishing - ebooks Account. ISBN: 1783555130

Reissmann S. (2014). *Journal of peptide science: an official publication of the European Peptide Society*, 20(10), 760–784.

Schindelin, J.; Arganda-Carreras, I. & Frise, E. et al. (2012), "Fiji: an open-source platform for biological-image analysis", *Nature methods* 9(7): 676-682, PMID 22743772.

Schneider, P., Müller, A. T., Gabernet, G., Button, A. L., Posselt, G., Wessler, S., Hiss, J. A., & Schneider, G. (2017). Hybrid Network Model for "Deep Learning" of Chemical Data: Application to Antimicrobial Peptides. *Molecular informatics*, 36(1-2), 10.1002/minf.201600011.

Shi, N. Q., Qi, X. R., Xiang, B., & Zhang, Y. (2014). A survey on "Trojan Horse" peptides: opportunities, issues and controlled entry to "Troy". *Journal of controlled release*, 194, 53-70.

Stalmans, S., Bracke, N., Wynendaele, E., Gevaert, B., Peremans, K., Burvenich, C., Polis, I., & De Spiegeleer, B. (2015). Cell-Penetrating Peptides Selectively Cross the Blood-Brain Barrier In Vivo. *PLoS one*, 10(10), e0139652.

Suckfuell, M., Lisowska, G., Domka, W., Kabacinska, A., Morawski, K., Bodlaj, R., Klimak, P., Kostrica, R., & Meyer, T. (2014). Efficacy and safety of AM-111 in the treatment of acute sensorineural hearing loss: a double-blind, randomized, placebo-controlled phase II study. *Otology & neurotology* : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 35(8), 1317–1326.

Tao, P., Wang, R. & Lai, L (1999). Calculating Partition Coefficients of Peptides by the Addition Method. *Journal of Molecular Modeling*, 5(10):189-195.

Torrent, M., Andreu, D., Nogués, V. M., & Boix, E. (2011). Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PloS one*, 6(2), e16968.

Torrent, M., Di Tommaso, P., Pulido, D., Nogués, M. V., Notredame, C., Boix, E., & Andreu, D. (2012). AMPA: an automated web server for prediction of protein antimicrobial regions. *Bioinformatics (Oxford, England)*, 28(1), 130–131.

Tossi, A., Sandri, L., & Giangaspero, A. (2003). New consensus hydrophobicity scale extended to non-proteinogenic amino acids. In *27th European Peptide Symposium* (pp. 416-417). Edizioni Ziino.

Torchilin, V. P. (2007). Tatp-mediated intracellular delivery of pharmaceutical nanocarriers. *Biochemical Society Transactions*, 35(4), 816-820.

Veltri, D., Kamath, U., & Shehu, A. (2018). Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, 34(16), 2740-2747.

Wang, G., Li, X. and Wang, Z. (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research* 44, D1087-D1093.

Wang, P., Hu, L., Liu, G., Jiang, N., Chen, X., Xu, J., Zheng, W., Li, L., Tan, M., Chen, Z., Song, H., Cai, Y. D., & Chou, K. C. (2011). Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. *PloS one*, 6(4), e18476.

Xiao, X., Wang, P., Lin, W. Z., Jia, J. H., & Chou, K. C. (2013). iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Analytical biochemistry*, 436(2), 168–177.

Xie, J., Bi, Y., Zhang, H., Dong, S., Teng, L., Lee, R. J., & Yang, Z. (2020). Cell-penetrating peptides in diagnosis and treatment of human diseases: From preclinical research to clinical application. *Frontiers in pharmacology*, 11, 697.

Yang, L., Weiss, T. M., Lehrer, R. I., & Huang, H. W. (2000). Crystallization of antimicrobial pores in membranes: magainin and protegrin. *Biophysical Journal*, 79(4), 2002-2009.

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415(6870), 389-395.

Zhang, L., Zhang, Y., Tai, L., Jiang, K., Xie, C., Li, Z., Lin, Y. Z., Wei, G., Lu, W., & Pan, W. (2016). Functionalized cell nucleus-penetrating peptide combined with doxorubicin for synergistic treatment of glioma. *Acta biomaterialia*, 42, 90–101.