



UNIVERSIDADE CATÓLICA DOM BOSCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

# **DESENHO DE PEPTÍDEOS ANTIBACTERIANOS UTILIZANDO *DEEP REINFORCEMENT LEARNING***

Autor: Abel Gil Ley

Orientador. Dr. Octávio L. Franco

Co-orientador: Dr. Fabiano C. Fernandes

Co-orientador: Dr. Anselmo J. Otero González

Campo Grande

Mato Grosso do Sul

Setembro – 2024

UNIVERSIDADE CATÓLICA DOM BOSCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**DESENHO DE PEPTÍDEOS ANTIBACTERIANOS UTILIZANDO  
*DEEP REINFORCEMENT LEARNING***

"Dissertação apresentada, como parte das exigências para obtenção do título de MESTRE EM BIOTECNOLOGIA, no Programa de Pós-Graduação em Biotecnologia da Universidade Católica Dom Bosco - Área de concentração: Biotecnologia"

Autor: Abel Gil Ley

Orientador. Dr. Octávio L. Franco

Co-orientador: Dr. Fabiano C. Fernandes

Co-orientador: Dr. Anselmo J. Otero González

Campo Grande

Mato Grosso do Sul

Setembro – 2024



**“DESENHO DE PEPTÍDEOS ANTIBACTERIANOS UTILIZANDO DEEP REINFORCEMENT  
LEARNING”**

Autor: ABEL GIL LEY

Orientador: Prof. Dr. Octávio Luiz Franco

Coorientador: Prof. Dr. Anselmo de Jesús Otero-González

Coorientador: Prof. Dr. Fabiano Cavalcanti Fernandes

TITULAÇÃO: Mestre em Biotecnologia  
Área de concentração: Biotecnologia.

APROVADA em 13 de setembro de 2024.

---

Dr. Octávio Luiz Franco (Orientador) – UCDB  
Dr. Anselmo de Jesús Otero-González (Coorientador) - Universidad de la Habana-Cuba  
Dr. Fabiano Cavalcanti Fernandes (Coorientador) - IFB  
Dr. Diego Oliveira Nolasco da Silva – ISMEP  
Dr. Marlon Henrique e Silva Cardoso - UCDB

L681d Ley, Abel Gil

Desenho de peptídeos antibacterianos utilizando deep reinforcement learning/ Abel Gil Ley sob orientação do Prof. Dr. Octávio L. Franco; Prof. Dr. Fabiano C. Fernandes e Prof. Dr. Anselmo J. Otero González.-- Campo Grande, MS : 2024.

73 p.: il.

Dissertação (Mestrado em Biotecnologia) - Universidade Católica Dom Bosco, Campo Grande - MS, 2024

Bibliografia: p. 57-69

1. Peptídeos antimicrobianos. 2. Inteligência artificial. 3. Aprendizado profundo por reforço. 4. Projeto de peptídeos. 5. Dinâmica molecular I. Franco, Octávio L. II. Fernandes, Fabiano C. III. González, Anselmo J. Otero. IV. Título.

CDD: 660.6

## **Biografia**

Abel Gil Ley, filho de Barbara Ley Calas e Pedro V. Gil Gonzales. Nasceu em Habana del Este, la Habana, Cuba, no dia 12 de setembro de 1995. Graduado em biologia pela Universidad de la Habana.

Trabalhou como professor assistente da assinatura Genética, no grupo de Biologia de Plantas da Universidad de la Habana.



## Sumário

Lista de Figuras .....	ix
Lista de Tabelas.....	xi
Resumo .....	i
Abstract .....	i
1. Revisão de Literatura .....	1
1.1. Resistência bacteriana .....	1
1.2. Peptídeos Antimicrobianos .....	2
1.2.1. Propriedades Comuns .....	2
1.2.2. Peptídeos Antibacterianos .....	3
1.3. Mecanismo de Ação dos AMPs .....	6
1.4. Avanços no design de AMPs usando métodos de AI .....	9
1.5. Inteligência Artificial .....	9
1.5.1. <i>Machine Learning e Deep Learning</i> .....	10
1.5.2. <i>Recurrent Neural Network (RNN)</i> .....	12
1.5.3. <i>Protein Language Model (pLM)</i> .....	14
1.5.4. <i>Deep Reinforcement Learning</i> .....	16
2. Justificativa .....	20
3. Objetivos .....	22
3.1. Objetivo geral .....	22
3.2. Objetivos específicos .....	22
4. Material e métodos .....	23
4.1. Criação das bases de dados .....	23
4.2. Extração de características de peptídeos .....	24
4.3. Espaço conformacional de AMP .....	25
4.4. Aprendizado por reforço (MRL) .....	25
4.5. Ajustes do modelo AMP-Bert .....	26

4.6.	Seleção das melhores sequências geradas	26
4.7.	Testes <i>in silico</i> do modelo de predição de AMP	27
4.8.	Síntese dos peptídeos e determinação de massa moleculares por espectrometria de massa MALDI-ToF.	28
4.9.	Cepas bacterianas e meios de crescimento:	28
4.10.	Determinação da concentração inibitória mínima (MIC)	29
4.11.	Avaliação de atividade hemolítica	29
4.12.	Simulações de dinâmica molecular	30
5.	Resultados.....	31
5.1.	Seleção das melhores sequências	31
5.2.	Caracterização por MALDI-ToF dos peptídeos antimicrobianos obtidos comercialmente.	34
5.3.	Concentração Inibitória Mínima (MIC)	36
5.4.	Atividade Hemolítica	38
5.5.	Dinâmica molecular do Pep 10 na membrana lipídica	39
6.	Discussão .....	46
7.	Considerações Finais .....	52
8.	Perspectivas .....	53
9.	Referências.....	55
10.	Apêndice A - Banco de dados e modelos utilizados neste estudo. ....	68
11.	Apêndice B - Espectros de massa dos peptídeos antimicrobianos. ....	69



## Lista de Figuras

Figura 1: Criação do banco de dados para treinamento de modelos preditivos e generativos.	32
Figura 2: Estrutura tridimensional, projeções de roda helicoidal e sequências dos 15 peptídeos.	42
Figura 3: Mapa de calor de previsão da atividade antimicrobiana dos quinze peptídeos (Pep 1-15) usando 5 modelos preditivos de última geração.	44
Figura 4: Mapa de calor de avaliação da atividade antimicrobiana <i>in vitro</i> dos quinze peptídeos (Pep 1-15) frente as cepas bacterianas suscetíveis (ATCC).	47
Figura 5: Mapa de calor de avaliação da atividade antimicrobiana <i>in vitro</i> do peptídeo 10 frente a cepa suscetível de <i>A. baumannii</i> (ATCC) e as cepas resistentes de <i>A. baumannii</i> (5015 e 8892).	48
Figura 6: Mapa de calor de avaliação da atividade hemolítica dos quinze peptídeos (Pep 1-15) frente a eritrócitos de camundongos.	49
Figura 7: Simulações de MD de Pep 10 em um sistema de bicamadas lipídicas, POPE/POPG (3:1).	50
Figura 8: RMSD do <i>backbone</i> do Pep 10 durante o tempo de simulação de 200ns.	51
Figura 9: RMSF de todos os resíduos de Pep 10 durante o tempo de simulação de 200ns.	51
Figura 10: Número de interações de hidrogênio entre Pep 10 e a bicamada lipídica durante o tempo de simulação de 200ns.	52
Figura 11: Interações entre Pep 10 e a membrana lipídica em 200 ns de simulação.	53
Figura 12: Projeções de roda helicoidal do peptídeo projetado Pep 10 e os peptídeos RR e RR4.	60

Figura S1: Espectros MALDI-TOF dos compostos A (Pep 1), B (Pep 2), C (Pep 3), e D (Pep 4). 77

Figura S2: Espectros MALDI-TOF dos compostos E (Pep 5), F (Pep 6), G (Pep 7), e H (Pep 8). 78

Figura S3: Espectros MALDI-TOF dos compostos I (Pep 10), J (Pep 11), K (Pep 12), e L (Pep 9). 78

Figura S4: Espectros MALDI-TOF dos compostos M (Pep 13), N (Pep 14), e O (Pep 15). 79

## Lista de Tabelas

Tabela 1. Propriedades físico-químicas dos peptídeos utilizados neste estudo.	43
Tabela 2. Dados correspondente a sequência, a fórmula molecular, e os íons $[M+H]^+$ encontrado e calculado para os AMPs.	45
Tabela 3. Interações entre Pep 10 e a membrana lipídica, em 200ns de simulação.	54
Tabela 4. Interações entre Pep 10 e a membrana lipídica, em 200ns de simulação.	55

## Resumo

Com o surgimento de bactérias multirresistentes, os peptídeos antimicrobianos (AMPs) oferecem opções promissoras para o controle desses patógenos. No entanto, a descoberta e o desenvolvimento de AMPs usando métodos tradicionais são processos demorados e caros. O *deep learning* tem sido aplicado ao design de novos AMPs e à classificação de AMPs com alta eficiência. Este estudo desenvolveu uma estrutura computacional que integra *deep reinforcement learning* para o design *in silico* de AMPs. De todas as sequências geradas pelo modelo, foram selecionadas apenas aquelas que atendiam a diversos critérios de seleção, como carga líquida, atividade antibacteriana, hemólise e citotoxicidade, calculados usando diferentes modelos preditivos. As quinze sequências que cumpriram esses critérios foram então sintetizadas e testadas contra vários patógenos, incluindo cepas multirresistentes. Após os testes iniciais de bioatividade, um dos peptídeos, o Pep 10, mostrou inibição a 64  $\mu\text{M}$  contra cepas suscetíveis e 32  $\mu\text{M}$  contra cepas resistentes da bactéria Gram-negativa *Acinetobacter baumannii*. Para entender como esse peptídeo interage com a membrana de *A. baumannii*, foram realizadas duas dinâmicas moleculares (como o peptídeo fora e dentro da membrana) com duração de 200 ns. A dinâmica mostrou que o Pep 10 mantém fortes interações de ponte de hidrogênio com a membrana ao longo da simulação. O *deep learning* mostrou grande potencial em acelerar a descoberta de novos AMPs, sendo uma ferramenta importante para o desenvolvimento de novos fármacos.

**Palavras-chave:** peptídeos antimicrobianos, inteligência artificial, aprendizado profundo por reforço, projeto de peptídeos, dinâmica molecular.

## Abstract

With the emergence of multi-resistant bacteria, antimicrobial peptides (AMPs) offer promising options for controlling these resistant strains. However, discovering and designing AMPs using traditional methods is a time-consuming and expensive process. Deep learning has been applied to the design of new AMPs and the classification of AMPs with high efficiency. This study developed a computational framework that integrates deep reinforcement learning for the *in silico* design of AMPs. From all the sequences generated by the model, only those that met various selection criteria, such as net charge, antibacterial activity, hemolysis, and cytotoxicity (calculated using different predictive models), were chosen. The fifteen sequences that met these conditions were then synthesized and tested against several pathogens, including multi-resistant strains. After initial bioactivity tests, one of the peptides, Pep 10, showed inhibition at 64  $\mu\text{M}$  against susceptible strains and 32  $\mu\text{M}$  against resistant strains of the Gram-negative bacterium *Acinetobacter baumannii*. To understand how this peptide interacts with the membrane of *A. baumannii*, two molecular dynamics simulations (with the peptide outside and inside the membrane) lasting 200 ns were performed. The dynamics showed that Pep 10 maintains strong hydrogen bond interactions with the membrane throughout the simulation. Deep learning has shown great potential in accelerating the discovery of new AMPs, making it an important tool for the development of new drugs.

**keywords:** antimicrobial peptides, artificial intelligence, deep reinforcement learning, peptide design, molecular dynamics.

# 1. Revisão de Literatura

## 1.1. Resistência bacteriana

A preocupação com a resistência aos antimicrobianos (RAM) é considerada uma das 10 principais ameaças à saúde global (*World Health Organization*, 2022). Embora os antibióticos tenham proporcionado inúmeros benefícios na saúde, agricultura, zootecnia e outras áreas, seu uso ilimitado e inadequado levou ao desenvolvimento de micróbios resistentes. Dos mais de 4 milhões de mortes globalmente em 2019 devido a infecções resistentes a medicamentos (Cesaro et al., 2023), a RAM causou diretamente 1,27 milhões dessas mortes (Murray et al., 2022) e estima-se que aumente para 10 milhões de casos anuais até 2050 (Wang et al., 2021). Isso significa que os medicamentos usados para tratar infecções não são mais tão eficazes de maneira confiável (Ji et al., 2024). A resistência aos antibióticos aumentou coletivamente os custos de saúde nos EUA em mais de 55 bilhões de dólares anualmente, quando a perda de produtividade é incluída na estimativa. Os custos totais hospitalares são 29% maiores ao tratar pacientes com patógenos multirresistentes, dado o maior tempo de internação e o uso aumentado de recursos (Morris e Cerceo, 2020). De acordo com relatos registrados, países de baixa e média renda possuem maiores taxas de RAM em comparação com países de alta renda (Algammal et al., 2023).

Enquanto bactérias Gram-positivas, como *Staphylococcus aureus* resistente à meticilina e *Clostridium difficile*, eram anteriormente as mais preocupantes em relação à resistência aos antibióticos, atualmente, as bactérias Gram-negativas resistentes a medicamentos estão em ascensão ao redor do mundo (Morris e Cerceo, 2020). Os seis principais patógenos responsáveis por mortes associadas à resistência (*Escherichia coli*, seguida por *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Acinetobacter baumannii* e *Pseudomonas aeruginosa*) foram responsáveis por 929.000 mortes associadas à RAM em 2019 (Murray et al., 2022).

No Brasil, em 2019, houveram 33.200 mortes atribuíveis à RAM e 137.900 mortes associadas à RAM. O número de mortes por RAM no Brasil é maior do que o número de mortes por diabetes e doenças renais, infecções respiratórias e tuberculose, autoagressão e violência interpessoal, doenças digestivas e doenças respiratórias crônicas (Murray et al., 2022).

Além disso, o uso inadequado de antibióticos causou danos ambientais ecológicos significativos (Ji et al., 2024). Esse fenômeno atenua a eficácia dos antibióticos comumente disponíveis e complica o manejo das infecções. Esses desenvolvimentos podem prolongar o curso da doença, aumentar o ônus econômico nos sistemas de saúde e elevar as taxas de mortalidade (Antimicrobial Resistance, 2023). Consequentemente, é necessário explorar novas abordagens para a terapia antimicrobiana para garantir uma terapia eficaz e conter a propagação de infecções resistentes (Prestinaci et al., 2015).

## **1.2. Peptídeos Antimicrobianos**

### **1.2.1. Propriedades Comuns**

Os peptídeos antimicrobianos (AMPs pela sigla em inglês) são caracterizados pelo seu pequeno tamanho, variando entre 3 e 50 aminoácidos. Eles possuem uma carga positiva (geralmente entre +2 e +13) devido a numerosos aminoácidos carregados positivamente (Kumar et al., 2018) e uma estrutura anfipática com  $\geq 30\%$  de resíduos hidrofóbicos, permitindo-lhes interagir com os lipídios da membrana dos microrganismos (Brandenburg et al., 2012; Huerta-Cantillo e Navarro-García, 2016). Inicialmente, pensava-se que o fato de os AMPs selecionarem múltiplos alvos em sua ação antimicrobiana limitava o desenvolvimento de cepas resistentes a eles (Marr et al., 2006). No entanto, nos últimos anos, foram relatados perfis de resistência induzida por AMPs em várias cepas bacterianas e classes de AMPs (Fleitas et al., 2023). Eles são encontrados em quase todos os organismos vivos, incluindo fungos, bactérias, insetos, vertebrados, plantas (Gutiérrez e Orduz, 2003) e até vírus (Huerta-Cantillo e Navarro-García, 2016). Além de sua atividade antimicrobiana, eles podem exercer atividade anticancerígena (Kumar et al., 2018) e estimular o sistema imunológico (Otero-Gonzalez et al., 2010).

Esses peptídeos possuem um mecanismo de ação complexo porque podem interagir diretamente com o patógeno, mas também podem modular a resposta imune a eles (Kumar et al., 2018). A interação direta com o patógeno pode ocorrer principalmente através da membrana lipídica, mas pode também afetar alvos intracelulares. Por outro lado, eles interagem com o sistema imunológico adaptativo por diferentes vias, como a regulação do processo inflamatório (Kumar et al., 2018; Tellez e Castaño, 2010). Eles são produzidos por diferentes tipos de células de forma constitutiva ou induzida, dependendo do organismo e do tipo de tecido onde a infecção está localizada (Tellez e Castaño, 2010). Entre os diferentes tipos de células que os sintetizam, podemos encontrar queratinócitos, células epiteliais do trato respiratório ou urogenital, células de Paneth do intestino delgado, neutrófilos, células *Natural Killer*, mastócitos e glândulas endócrinas (Castañeda-Casimiro et al., 2009).

Os AMPs apresentam grande diversidade estrutural e em termos de organismo produtor, célula alvo ou mecanismos de ação (Alba et al., 2012), sendo classificados de acordo com essas características. Uma das classificações estruturais mais recentes agrupa-os em três grupos principais: aqueles que adotam uma estrutura de  $\alpha$ -hélice, aqueles que adotam uma estrutura folha  $\beta$  e aqueles que têm uma estrutura estendida (Brogden, 2005).

### **1.2.2. Peptídeos Antibacterianos**

Os AMPs mais bem estudados são aqueles que apresentam atividade antibacteriana, pois as bactérias são o patógeno que mais frequentemente afeta a saúde humana (Jenssen et al., 2006). Eles também podem ter tanto atividade antifúngica quanto antiviral (Boman, 2003). Os peptídeos antibacterianos podem ser bactericidas (se matam as bactérias) ou bacteriostáticos (se inibem o seu crescimento). Geralmente, as concentrações inibitórias mínimas (MIC) e as concentrações bactericidas mínimas (MBC) dos peptídeos antibacterianos coincidem, indicando que esses peptídeos são principalmente bactericidas (Marr et al., 2006).

Uma vantagem dos peptídeos sobre os antibióticos convencionais, é sua capacidade de neutralizar toxinas bacterianas e prevenir sepse, que são complicações



comuns e perigosas durante a infecção (Marr et al., 2006). Além disso, eles não são afetados pelos mesmos mecanismos de resistência que afetam os antibióticos atuais porque sua interação com a membrana não requer um alvo específico (Mookherjee et al., 2020). É por isso que muitos AMPs têm excelente atividade contra cepas multirresistentes, como *Staphylococcus aureus* resistente à meticilina (a meticilina é um antibiótico betalactâmico do grupo das penicilinas) e *Pseudomonas aeruginosa* multirresistente (Marr et al., 2006).

As cecropinas foram os primeiros peptídeos antibacterianos isolados da mariposa gigante da seda, *Hyalophora cecropia* (Chen e Jiang, 2023). Os principais participantes do sistema imunológico dos insetos apresentam hélices anfipáticas e potente atividade antibacteriana Gram-positiva e Gram-negativa (Ciulla e Gelain, 2023). A cecropina P1 é ativa contra várias bactérias clinicamente relevantes, incluindo *Pseudomonas aeruginosa* e *Acinetobacter baumannii* (Chen e Jiang, 2023).

Vários peptídeos anfipáticos catiônicos foram identificados em anfíbios, entre os quais a magainina pode ser considerada um dos principais participantes. A maioria dos peptídeos anfíbios exibe atividade antimicrobiana de amplo espectro contra bactérias, fungos e protozoários. Curiosamente, D-aminoácidos foram encontrados em alguns AMPs de anfíbios. Essa modificação reduz a proteólise e o espectro de antibióticos (Ciulla e Gelain, 2023).

Peptídeos cíclicos são estruturas promissoras para biofármacos e agentes poderosos contra a resistência a antibióticos. Vários estudos sobre peptídeos antibacterianos cíclicos foram relatados nas últimas duas décadas, provavelmente porque suas propriedades farmacológicas favoráveis os tornam compostos terapêuticos muito atraentes (Ciulla e Gelain, 2023). De fato, os peptídeos cíclicos desempenham um papel crucial no desenvolvimento de novos antibióticos devido à sua maior estabilidade em condições fisiológicas, maior permeabilidade da membrana e maior biodisponibilidade oral do que análogos lineares (Yan et al., 2021). Além disso, seu custo reduzido de síntese, baixa toxicidade e, especialmente, sua melhor atividade os tornam candidatos promissores como novos agentes antimicrobianos (Ciulla e Gelain, 2023).

A vancomicina é um peptídeo cíclico natural atualmente usado para tratamento oral contra bactérias Gram-positivas (Yan et al., 2021). Recentemente, o peptídeo cíclico daptomicina foi aprovado pelo FDA para o tratamento tópico de infecções causadas por bactérias Gram-positivas. Ele causa dissipação do potencial de membrana das membranas bacterianas devido às suas mudanças conformacionais (induzidas pela ligação a íons de cálcio extracelulares) e aumento da anfipaticidade capaz de formar estruturas transmembrana com subsequente ruptura da integridade da membrana (Ciulla e Gelain, 2023).

A maioria dos AMPs (> 86%) tem 11–50 aminoácidos com uma carga líquida de 0 a +7 e uma porcentagem hidrofóbica de 31–70%. Essas características são consistentes com o mecanismo de ação dos AMPs (Chen e Jiang, 2023). Embora existam muitos tipos de AMPs naturais, a maioria deles mostra sequências longas, o processo de extração é complicado e o custo de síntese é relativamente alto. Devido ao seu baixo custo de produção e potencial de otimização, os AMPs curtos (menos de 15 aminoácidos) são superiores a alguns AMPs naturais. As características comuns desses AMPs curtos são suas cargas positivas e hidrofobicidade, embora as sequências principais sejam diversas. As cargas positivas fornecem ao peptídeo uma afinidade eletrostática pela membrana celular carregada negativamente da bactéria. Em contraste, a hidrofobicidade e a anfifilicidade geral ancoram o peptídeo na membrana celular e podem se autoassociar à membrana. A destruição da membrana do peptídeo faz com que o conteúdo da célula flua para fora e, finalmente, leve à morte bacteriana (Yan et al., 2021).

Além disso, a abundância e a repetição de alguns AMPs de aminoácidos influenciam sua atividade biológica. Os antibióticos naturais, especialmente os AMPs, mostram sequências repetidas de poliaminoácidos devido às suas excelentes propriedades antibacterianas (por exemplo, propriedades catiônicas, interações aprimoradas da membrana celular, etc.) (Chen e Jiang, 2023). A lisina e a arginina são responsáveis pela cationicidade dos AMPs em pH fisiológico e fornecem fortes interações eletrostáticas com lipídios aniônicos em membranas bacterianas Gram-negativas e com paredes celulares de bactérias Gram-positivas. Agrupamentos de

resíduos carregados conferem capacidade bactericida ao se ligarem ao núcleo hidrofóbico da bicamada lipídica, induzindo assim uma partição. No entanto, o comprimento da cadeia peptídica e a anfipaticidade também influenciam a propensão à permeabilidade da membrana bacteriana. Assim, peptídeos ricos em lisina podem despolarizar membranas celulares bacterianas por meio de interações eletrostáticas, causando ruptura da membrana e morte bacteriana. No entanto, foi amplamente demonstrado que, apesar de sua excelente atividade antibacteriana, altas quantidades, disposição e o padrão de lisina e argininas aumentam a citotoxicidade (Ciulla e Gelain, 2023).

### **1.3. Mecanismo de ação dos AMPs**

O modo de ação dos AMPs geralmente envolve a desestabilização da integridade da membrana citoplasmática bacteriana de várias maneiras. No entanto, outros mecanismos antimicrobianos também foram caracterizados, visando processos celulares essenciais, como a síntese de DNA e proteínas, o dobramento de proteínas, a atividade enzimática e a síntese da parede celular. As interações com a membrana permanecem cruciais, mesmo para peptídeos que atuam no interior da célula, pois eles precisam de um meio para translocação. Além disso, para alguns dos AMPs mais estudados, como magainina e indolicidina, foi demonstrado que eles podem ter múltiplos modos de ação (Nguyen et al., 2011).

A interação inicial dos AMPs com a membrana do patógeno ocorre por meio de interações eletrostáticas (AMPs carregados positivamente e membrana lipídica carregada negativamente) e hidrofóbicas. Ao interagir com a membrana, eles alteram sua permeabilidade ou causam lise celular através da formação de poros (Huerta-Cantillo e Navarro-García, 2016). Estudos com magaininas mostraram que as mudanças de permeabilidade da membrana induzidas por peptídeos estão relacionadas à natureza aniônica do lipídio. Foi demonstrado que em lipossomos de fosfatidilglicerol, um lipídio abundantemente observado em membranas bacterianas, o peptídeo induz permeabilização eficaz. Por outro lado, em lipossomos de fosfatidilserina, um lipídio característico de membranas de células de mamíferos, o peptídeo é menos eficaz. No

entanto, outros peptídeos, como melitina, paracina, dermaseptinas e cecropinas, são principalmente líticos em células bacterianas e eritrócitos de mamíferos (Reddy et al., 2004).

O mecanismo de permeabilização da membrana e formação de poros pelos AMPs tem sido descrito por vários modelos. Entre os modelos mais estudados estão: o modelo de carpete, o modelo de barril e o modelo de poro toroidal (Tellez e Castaño, 2010). Outros mecanismos de formação de poros também foram descritos, como o modelo de Shai-Huang-Matsuzaki e o modelo de eletroporação (Nguyen et al., 2011). Todos esses modelos têm em comum a interação inicial dos peptídeos com os grupos negativos nas regiões polares dos lipídios na superfície da membrana, adotando uma orientação paralela a esta (Jenssen et al., 2006).

Além dos modelos já mencionados, outros não envolvem a formação de poros na membrana. A presença de peptídeos na bicamada lipídica pode causar sua curvatura (modelo de afinamento/engrossamento da membrana), bem como seu remodelamento, formando domínios ricos em lipídios aniônicos ao redor do peptídeo (modelo de aglomeração de lipídios aniônicos). Em alguns casos, a ligação do peptídeo à membrana pode aumentar se eles tiverem como alvo lipídios oxidados (modelo de alvo de lipídios oxidados). Além disso, o acúmulo de AMPs pode atrair pequenos ânions para o outro lado da membrana e causar seu efluxo (modelo de transportador de ânions) ou dissipar o potencial da membrana sem causar nenhum outro dano aparente à membrana (modelo de despolarização não lítica da membrana) (Nguyen et al., 2011).

Recentemente, a influência dos AMPs na curvatura da membrana bacteriana foi investigada. A curvatura intrínseca da membrana pode interferir nos mecanismos de ação de alguns AMPs. Ao mesmo tempo, alguns AMPs podem induzir curvaturas positivas e negativas da membrana em bactérias, levando à formação de poros toroidais e micelização. Os AMPs membranalíticos também são conhecidos por interagir primeiro com as superfícies bacterianas por interações eletrostáticas, interferindo subsequentemente na permeabilidade, fluidez e curvatura da membrana. Peptídeos, como magainina 2, induzem curvaturas positivas da membrana para a formação de poros

toroidais. Por outro lado, lipídios indutores de curvatura negativa inibem a formação de poros (Cardoso et al., 2024).

Os AMPs também podem ter como alvo a formação de componentes estruturais, como a parede celular. A mersacidina, produzido por bactérias, interfere na transglicosilação do lipídio II, uma etapa necessária na síntese de peptidoglicano. A nisina, também pode se ligar ao lipídio II, inibindo assim a síntese da parede celular, além de sua atividade de formação de poros. Curiosamente, este é o mesmo processo biossintético que é alvo do antibiótico vancomicina; no entanto, acredita-se que a mersacidina e a nisina atuem interagindo com diferentes fragmentos moleculares dentro do lipídio II, explicando por que esses peptídeos ainda são ativos contra bactérias resistentes à vancomicina (Jenssen et al., 2006).

Uma quantidade substancial de AMPs, incluindo a polifemusina, pode se translocar através da membrana e induzir a morte do patógeno afetando processos citoplasmáticos, como a síntese de proteínas (Fjell et al., 2012). Esses AMPs podem ter vários alvos intracelulares, como ácidos nucleicos, aos quais se ligam e interferem na sua síntese. Eles também podem inibir a atividade de algumas enzimas e a síntese da parede celular, interromper a divisão celular, estimular a autólise ou causar danos à célula devido ao acúmulo no interior, etc. (Huerta-Cantillo e Navarro-García, 2016).

Um dos AMPs que segue esse mecanismo de ação é a buforina II. Em vários estudos foi mostrado que este AMP é translocado através das membranas bacterianas sem causar permeabilização das mesmas e se liga tanto ao DNA quanto ao RNA. De maneira semelhante, peptídeos em hélice- $\alpha$  derivados de pleurocidina e dermaseptina inibem a síntese de DNA e RNA sem desestabilizar a membrana de *E. coli* (Jenssen et al., 2006). Outro peptídeo que também inibe a síntese de DNA e RNA é a indolicidina, enquanto o PR-39 interrompe a síntese de proteínas e induz a degradação de algumas proteínas necessárias para a replicação do DNA (Brogden, 2005). A pirrolozina penetra na célula alvo e pode se ligar à DnaK, uma proteína de choque térmico envolvida no dobramento de proteínas. Especificamente, este peptídeo inibe a atividade ATPase exibida pela DnaK, impedindo o dobramento de proteínas e causando um acúmulo de

polipeptídeos não funcionais, o que leva à morte celular (Castañeda-Casimiro et al., 2009).

#### **1.4. Inteligência Artificial**

Existem muitas definições de inteligência artificial (*artificial intelligence* - AI). No teste de Turing, a AI é definida como a capacidade das máquinas de se comunicar com humanos (utilizando dispositivos eletrônicos de saída) sem revelar que não são humanos, onde o critério essencial de julgamento é binário. Marvin Minsky, um dos pioneiros da AI, definiu a AI como a habilitação de máquinas para fazer coisas que requerem inteligência humana. A escola simbólica acredita que a AI é a operação de símbolos, e os símbolos mais primitivos correspondem a entidades físicas. Embora as descrições de AI sejam variadas, acredita-se amplamente que o núcleo da AI são as teorias de pesquisa, métodos, tecnologias e aplicações para simular, estender e expandir a inteligência humana (Kaynak, 2021).

A pesquisa em AI tem continuado por mais de 65 anos e alcançou impressionantes avanços tanto no estudo teórico quanto em aplicações do mundo real. A AI está sendo utilizada em quase todos os lugares e é considerada uma habilidade fundamental para o futuro. Estima-se que o mercado de AI atinja 190 bilhões de dólares até 2025, com uma taxa de crescimento anual de mais de 36% entre 2018 e 2025 (Kaynak, 2021).

Os campos de pesquisa da AI incluem sistemas e engenharia, medicina, psicologia, ciências cognitivas, matemática, ciência da computação e muitos outros. Os campos de aplicação da AI são extensos, abrangendo (mas não se limitando a) reconhecimento de fala, processamento de imagens, processamento de linguagem natural, robôs inteligentes, veículos autônomos, sistemas de energia, cuidados de saúde, etc. Em áreas limitadas, a AI superou os humanos. O fenômeno da AI cada vez mais avançada desencadeou uma nova onda de discussões sobre como a AI pode mudar a sociedade humana (Jiang et al., 2022).

### **1.5. *Machine Learning e Deep Learning***

*Machine Learning* (ML) ou aprendizado de máquina, refere-se amplamente ao processo de ajuste de modelos preditivos a dados ou à identificação de agrupamentos informativos dentro dos dados. O campo do aprendizado de máquina essencialmente tenta aproximar ou imitar a capacidade humana de reconhecer padrões, embora de maneira objetiva, utilizando computação (Sharifani e Amini, 2023).

O aprendizado de máquina é particularmente útil quando o conjunto de dados que se deseja analisar é muito grande (muitos pontos de dados individuais) ou muito complexo (contém muitas características) para análise humana, ou quando se deseja automatizar o processo de análise de dados para estabelecer um fluxo de trabalho reprodutível e eficiente em termos de tempo. Os dados de experimentos biológicos frequentemente possuem essas propriedades; os conjuntos de dados biológicos cresceram enormemente tanto em tamanho quanto em complexidade nas últimas décadas, e está se tornando cada vez mais importante não apenas ter algum meio prático de dar sentido a essa abundância de dados, mas também ter um entendimento sólido das técnicas que são usadas (Greener et al., 2021).

As redes neurais artificiais ou *artificial neural networks* (ANNs) consistem em representações matemáticas de unidades de processamento conectadas chamadas neurônios artificiais. Como as sinapses no cérebro, cada conexão entre os neurônios transmite sinais cuja força pode ser amplificada ou atenuada por um peso, continuamente ajustado durante o processo de aprendizado. Os sinais são processados por neurônios subsequentes apenas se um certo limiar for excedido, conforme determinado por uma função de ativação. Tipicamente, os neurônios são organizados em redes com diferentes camadas. Uma camada de entrada recebe geralmente os dados de entrada (por exemplo, imagens de produtos de uma loja online), e uma camada de saída produz o resultado (por exemplo, categorização de produtos). No meio, há zero ou mais camadas ocultas responsáveis por aprender um mapeamento não linear entre a entrada e a saída. O número de camadas e neurônios, entre outras escolhas de propriedades, como a taxa de aprendizado ou função de ativação, não pode ser aprendido pelo algoritmo de aprendizado. Eles constituem os hiper parâmetros de um

modelo e devem ser definidos manualmente ou determinados por uma rotina de otimização (Janiesch et al., 2021). Uma propriedade chave das redes neurais é que, com muitas poucas suposições, uma rede neural corretamente configurada pode aproximar qualquer função matemática a um nível arbitrário de precisão. Em outras palavras, se qualquer processo (biológico ou outro) puder ser pensado como uma função de um conjunto de variáveis, então esse processo pode ser modelado a qualquer grau arbitrário de precisão, governado apenas pelo tamanho ou complexidade do modelo (Greener et al., 2021).

*Deep Learning* (DL) ou redes neurais profundas consistem tipicamente em mais de uma camada oculta, organizadas em arquiteturas de rede profundamente aninhadas. Além disso, geralmente contêm neurônios avançados em contraste com as ANNs simples. Isto é, podem usar operações avançadas (por exemplo, convoluções) ou múltiplas ativações em um neurônio, em vez de usar uma simples função de ativação. Essas características permitem que redes neurais profundas sejam alimentadas com dados de entrada brutos e descubram automaticamente uma representação necessária para a tarefa de aprendizado correspondente. Esta é a capacidade central das redes, comumente conhecida como *deep learning*. ANNs simples (por exemplo, *autoencoders* rasos) e outros algoritmos de ML (por exemplo, árvores de decisão) podem ser classificados sob o termo aprendizado de máquina raso, uma vez que não fornecem tais funcionalidades. Arquiteturas de *deep learning* incluem camadas convolucionais, recorrentes e de atenção (Ofer et al., 2021). Enquanto alguns algoritmos de ML rasos são considerados inerentemente interpretáveis por humanos e, portanto, transparentes, a tomada de decisão da maioria dos algoritmos avançados de ML é por si ininteligível, a menos que explicada de outra forma e, portanto, constitui uma caixa preta (Janiesch et al., 2021).

Muitos algoritmos de *deep learning* requerem inúmeros conjuntos de treinamento. Embora os dados disponíveis em sistemas biológicos tenham crescido tremendamente, a ordem de magnitude ainda é pequena para muitos frameworks de *deep learning*, e as vantagens do treinamento em *deep learning* não podem ser exploradas. Além de selecionar ordens de magnitude mais precisas de informações, também precisamos



melhorar os modelos de *deep learning* para resolver alguns problemas pequenos e esparsos no conjunto de dados biológicos. Métodos de *deep learning* também têm dificuldade em explicar a saída de seus modelos do ponto de vista biológico. Mas muitos cálculos para redes biológicas não só querem obter o resultado, mas também esperam encontrar potenciais mecanismos biológicos e sistemas de redes funcionais que formam esse resultado. Portanto, como transformar esses métodos em uma interpretação e análise biológica também é um desafio importante (Jin et al., 2020).

### **1.5.1. Recurrent Neural Network (RNN)**

As redes neurais recorrentes ou *recurrent neural network* (RNN) são mais adequadas para dados na forma de sequências ordenadas, de modo que existe alguma dependência ou correlação entre um ponto na sequência e o próximo. Sua principal aplicação é no processamento de linguagem natural, onde o texto é tratado como uma sequência de palavras ou caracteres (Wang et al., 2020). As RNNs podem ser pensadas como um bloco de camadas de rede neural que recebem como entrada os dados correspondentes a cada entrada (ou passo de tempo) em uma sequência e produzem uma saída para cada entrada, que depende das entradas processadas anteriormente. Elas também podem ser usadas para gerar uma representação de toda a sequência passada para camadas posteriores da rede para gerar a saída. Isso é útil, pois sequências de qualquer comprimento podem ser convertidas em uma representação de tamanho fixo e inseridas em um perceptron multicamada (Greener et al., 2021). Exemplos óbvios do uso de RNNs na biologia incluem a análise de sequências de genes ou proteínas, com tarefas incluindo a identificação de regiões promotoras em sequências de genes, a previsão da estrutura secundária de proteínas ou a modelagem dos níveis de expressão gênica ao longo do tempo. As variantes mais avançadas, como a memória de longo prazo (*Long Short-Term Memory*, LSTM) ou a unidade recorrente com porta das RNNs, têm muitos usos na biologia, incluindo a previsão da estrutura de proteínas, o design de peptídeos e a previsão de diagnósticos clínicos a partir de registros de saúde. Esses métodos mais avançados são frequentemente usados em combinação com redes neurais convolucionais (*Convolutional Neural Networks*, CNNs), o que pode aumentar a precisão (Wang et al., 2020). As RNNs podem ser muito robustas na análise de dados

baseados em sequências. Por exemplo, RNNs treinadas em milhões de sequências de proteínas demonstraram a capacidade de capturar informações evolutivas e estruturais e podem ser aplicadas a uma variedade de tarefas supervisionadas, incluindo tarefas relacionadas ao design de novas sequências de proteínas (Greener et al., 2021).

A LSTM é uma arquitetura especial de RNN, oferecendo um desempenho superior às RNNs tradicionais, especialmente projetada para superar os problemas de explosão/desaparecimento de gradiente que geralmente surgem ao aprender dependências de longo prazo (Wang et al., 2020). A arquitetura do LSTM consiste em um conjunto de sub-redes conectadas de forma recorrente, conhecidas como blocos de memória (Hochreiter e Schmidhuber, 1997). O propósito do bloco de memória é manter seu estado ao longo do tempo e regular o fluxo de informações por meio de unidades de controle não lineares. A saída do bloco é conectada de volta à entrada do bloco e a todas as entradas. Esse sistema recorrente pode lidar com uma ampla variedade de problemas, incluindo previsão de séries temporais, reconhecimento de texto, processamento de linguagem natural, legendas de imagens e vídeos, análise de sentimentos e visão computacional. É uma prática comum hibridizar CNNs com LSTM para obter um desempenho ideal. Em tais modelos híbridos, camadas de convolução e *pooling* são usadas para reduzir a dimensionalidade do problema enquanto suprimem grandemente a redundância nas representações (Van Houdt et al., 2020).

Outro problema identificado com as RNNs é a dificuldade que elas têm em examinar partes específicas de uma sequência de entrada, o que é importante para gerar uma saída altamente precisa. A adição de um mecanismo de atenção às RNNs, que permite ao modelo acessar todas as partes da sequência de entrada ao calcular cada saída, foi introduzida para aliviar esse problema (Greener et al., 2021). Recentemente, foi demonstrado que a RNN nem mesmo é necessária, e que a atenção por si só pode ser usada; os modelos resultantes, chamados de "*transformers*", obtiveram resultados de ponta em muitos *benchmarks* de processamento de linguagem natural. Modelos de *transformers*, frequentemente treinados em bilhões de sequências usando milhares de unidades de processamento gráfico, recentemente mostraram maior precisão do que as RNNs para tarefas em sequências biológicas (Van Houdt et al., 2020). O sucesso notável

do AlphaFold2 no 14<sup>o</sup> Experimento de Avaliação Crítica da Predição da Estrutura de Proteínas (CASP14), uma avaliação cega de abordagens computacionais para prever a estrutura de proteínas a partir de sequências, sugere que os modelos que usam atenção também têm potencial para tarefas em biologia estrutural (Greener et al., 2021). Recentemente, o modelo AlphaFold3 com uma arquitetura baseada em difusão substancialmente atualizada foi desenvolvido, capaz de prever a estrutura conjunta de complexos, incluindo proteínas, ácidos nucleicos, pequenas moléculas, íons e resíduos modificados. O novo modelo AlphaFold demonstrou precisão substancialmente melhorada em relação a muitas ferramentas especializadas anteriores: precisão muito maior para interações proteína-ligante em comparação com ferramentas de encaixe de última geração, precisão muito maior para interações proteína-ácido nucleico em comparação com preditores específicos de ácido nucleico e precisão de previsão anticorpo-antígeno substancialmente maior em comparação com o modelo anterior (Abramson et al., 2024).

### **1.5.2. Protein Language Model (pLM)**

O processamento de linguagem natural ou *natural language processing* (NLP) é um ramo da AI que permite que as máquinas compreendam a linguagem natural. A representação vetorial de palavras (*word embeddings*) é o ponto crítico que afeta o desempenho final do modelo, pois torna os dados de texto compreensíveis para os modelos de ML. Usando milhões de sequências de proteínas disponíveis, pesquisadores de bioinformática empregaram abordagens de aprendizado auto-supervisionado para o desenvolvimento de modelos de linguagem de proteínas (pLM). Vários pLMs foram lançados e alcançaram ótimo desempenho em tarefas de previsão subsequentes (por exemplo, estrutura de proteínas, funcionalidade, localização subcelular, etc.) (Du et al., 2024).

Assim como a linguagem humana, as sequências de proteínas podem ser naturalmente representadas como cadeias de letras. O alfabeto de proteínas consiste em 20 aminoácidos comuns (AAs) (excluindo aminoácidos não convencionais e raros). Além disso, como a linguagem natural, as proteínas naturalmente evoluídas são

tipicamente compostas por elementos modulares reutilizados que exibem pequenas variações que podem ser rearranjadas e montadas hierarquicamente. Por essa analogia, motivos (*motifs*) e domínios proteicos comuns, sendo os blocos funcionais básicos das proteínas, são semelhantes a palavras, frases e sentenças na linguagem humana. Dadas essas semelhanças em forma e substância, parece natural aplicar métodos de processamento de linguagem natural às sequências de proteínas. Nas últimas décadas, houve um fluxo contínuo de algoritmos estatísticos e de aprendizado de máquina do campo do NLP para a bioinformática (Ofer et al., 2021).

O modelo de representação de linguagem BERT (*Bidirectional Encoder Representations from Transformers*) (Devlin et al., 2018) foi projetado para pré-treinar representações bidirecionais profundas a partir de texto não rotulado, condicionando conjuntamente no contexto esquerdo e direito em todas as camadas. Como resultado, o modelo BERT pré-treinado pode ser ajustado com apenas uma camada de saída adicional para criar modelos de ponta para uma ampla gama de tarefas, como resposta a perguntas e inferência de linguagem, sem modificações substanciais na arquitetura específica da tarefa.

O AMP-BERT (Lee et al., 2022) foi desenvolvido utilizando um modelo ajustado do ProtBert-BFD (Elnaggar et al., 2020), um modelo baseado em BERT que foi pré-treinado com aproximadamente 2,1 bilhões de sequências de proteínas do Big Fantastic Database (BFD). Este modelo ajustado pode usar o mecanismo de atenção do BERT para fazer previsões melhores e implementar uma análise de auto-atenção subsequente que captura características estruturais importantes em AMPs. Isso melhora os modelos existentes de classificação de AMP, faz previsões de classe mais precisas com dados externos e sugere resíduos importantes que contribuem para o processo de classificação, além de meras métricas de avaliação de previsão (por exemplo, precisão, sensibilidade, especificidade) (Lee et al., 2022).

Os modelos de linguagem têm o potencial de aprender padrões em sequências de proteínas ao longo da evolução. Essa ideia motiva a pesquisa em modelos de linguagem em escala evolutiva, nos quais modelos básicos aprendem representações que refletem aspectos da biologia subjacente e, com maior capacidade de

representação, capturam a estrutura secundária e terciária em baixa resolução. Apesar da simplicidade de seus objetivos de treinamento, como preencher palavras faltantes ou prever a próxima palavra, os modelos de linguagem de texto demonstraram exibir capacidades emergentes que se desenvolvem como uma função de escala em aumento de poder computacional, dados e vários parâmetros. Modelos de linguagem modernos contendo dezenas a centenas de bilhões de parâmetros mostram habilidades como tradução de linguagem com poucos exemplos, raciocínio de senso comum e resolução de problemas matemáticos, tudo sem supervisão explícita. A estrutura em resolução atômica emerge e continua a melhorar em modelos de linguagem ao longo de quatro ordens de magnitude na escala de parâmetros. Fortes correlações entre a compreensão da sequência de proteínas pelo modelo de linguagem (perplexidade) e a precisão da previsão da estrutura revelam uma ligação estreita entre a modelagem de linguagem e o aprendizado da estrutura (Lin et al., 2022).

A modelagem em escala evolutiva (ESM-2) é o modelo de linguagem mais recente (baseado em modelos de linguagem de proteínas com *Transformers*), treinado em escalas de 8 milhões até 15 bilhões de parâmetros, lançado em 2022 pela *Fundamental AI Research (FAIR) Protein Team* da Meta. Em relação ao modelo de geração anterior, ESM-1b, o ESM-2 introduziu melhorias na arquitetura e nos parâmetros de treinamento e aumenta os recursos computacionais e os dados. A família de modelos resultante, ESM-2, superou o anteriormente estado da arte ESM-1b (um modelo de aproximadamente 650 milhões de parâmetros) com um número comparável de parâmetros, e em *benchmarks* de previsão de estrutura também supera outros modelos de linguagem de proteínas recentes (Lin et al., 2022).

### **1.5.3. Deep Reinforcement Learning**

O aprendizado por reforço ou *reinforcement learning* (RL) é um subcampo da AI usado para resolver problemas de tomada de decisão dinâmica. Isso envolve a análise de ações possíveis e a estimativa da relação estatística entre as ações e seus possíveis resultados, seguida pela determinação de um regime de tratamento que tente encontrar o resultado mais desejável (Popova et al., 2018). No RL, um agente recebe um estado e

uma recompensa do ambiente. A tarefa do agente é determinar a ação apropriada. No aprendizado por reforço, a ação apropriada é maximizar a recompensa (Ladosz et al., 2022). De forma mais simples, os agentes de RL usam aprendizado iterativo para obter com sucesso estratégias de decisão satisfatórias (Li, 2023). O modelo de RL é composto por dois componentes principais: um gerador e um discriminador. A função principal do gerador é criar sequências fictícias de AMP, enquanto o discriminador premia os AMPs gerados (X. Wang et al., 2024).

O aprendizado por reforço profundo (*deep reinforcement learning*, DRL), uma combinação aprofundada de ANN e RL, integra a capacidade de representação de características do *deep learning* com a capacidade de tomada de decisão do aprendizado por reforço, permitindo alcançar poderosas capacidades de controle de aprendizado de ponta a ponta. Na última década, o DRL fez avanços substanciais em muitas tarefas que exigem a percepção de entradas de alta dimensão e a tomada de decisões ótimas ou quase ótimas (X. Wang et al., 2024).

Sua aplicação frequentemente requer abundância de dados interativos e poder computacional significativo. A instabilidade no treinamento representa um comportamento de oscilação no DRL e é frequentemente acompanhada por uma eficiência extremamente baixa. Uma estratégia comum para eliminar a instabilidade no treinamento é evitar mudanças muito grandes na política em cada etapa. É difícil determinar manualmente qual deve ser o tamanho da etapa da política, pois um tamanho de etapa muito pequeno pode levar a uma convergência muito lenta e um tamanho de etapa muito grande pode resultar no risco de instabilidade no treinamento. A otimização de política na região de confiança ou *trust region policy optimization* (TRPO) (Schulman et al., 2015) aborda esse tipo de trade-off formulando uma restrição de região de confiança, que mede o quão distantes estão as novas e antigas políticas uma da outra. O benefício do TRPO é que ele pode garantir uma melhora de forma contínua e consistente em uma direção para a otimização da política em cada iteração. A otimização de política proximal ou *proximal policy optimization* (PPO) (Schulman et al., 2017) é uma versão aprimorada do TRPO que se baseia na técnica de recorte de primeira ordem para evitar a resolução de um problema de otimização de segunda ordem no TRPO. Ao usar

essa técnica, aumenta-se notavelmente a eficiência computacional, mantendo-se atualizações monótonas de política (Li, 2023).

### 1.6. Avanços no desenho de AMPs usando métodos de AI

A descoberta de medicamentos é um procedimento prolongado e desafiador. O custo médio para encontrar novos medicamentos por meios convencionais é estimado em US\$ 2,6 bilhões, com base nos dados disponíveis, e o processo inteiro pode levar mais de 12 anos (Zhang et al., 2023). Um desafio significativo é identificar um composto que atenda a múltiplos critérios, como bioatividade, metabolismo do medicamento, perfil farmacocinético e acessibilidade sintética, entre outros. Infelizmente, a porcentagem do vasto espaço químico que pode ser sintetizado e testado, quanto mais eficientemente, é minúscula. No entanto, o desenho *de novo* utiliza algoritmos para projetar e avaliar virtualmente moléculas, reduzindo assim o espaço químico (Olivecrona et al., 2017).

As metodologias convencionais de desenho de AMP frequentemente se baseiam em investigações de estrutura-atividade, que são frequentemente incorporadas em modelos de relações quantitativas estrutura-atividade. Esses procedimentos envolvem o treinamento de modelos de aprendizado de máquina identificando características pertinentes e, subsequentemente, empregando esses modelos para prever possíveis candidatos a peptídeos promissores (Zhang et al., 2023). Observa-se que essas metodologias são principalmente focadas em avaliar peptídeos existentes e não são projetadas para gerar novos diretamente (L. Wang et al., 2024). Embora o aprendizado de máquina tenha mostrado potencial na identificação e triagem de AMP, ele tem limitações, incluindo a dependência de características extraídas manualmente. No entanto, essas desvantagens foram abordadas pela incorporação de técnicas de *deep learning* (R. Wang et al., 2024).

Estudos recentes demonstraram o potencial de modelos generativos profundos em acelerar a descoberta de materiais e medicamentos. Tais modelos foram utilizados com sucesso para identificar AMPs com propriedades desejadas, alcançando considerável sucesso na descoberta de AMP (Zhang et al., 2023). As metodologias de desenho de AMP testemunharam avanços substanciais, e sua eficácia foi validada por

meio de várias abordagens. Essas incluem o desenvolvimento de AMPs usando algoritmos genéticos, redes de memória de longo prazo, redes adversárias generativas (GANs), *autoencoders* (AEs), busca em árvore de Monte Carlo, aprendizado por reforço, aprendizado de transferência e otimização bayesiana (L. Wang et al., 2024).

A utilização de modelos de *deep learning* que integram dados de sequência de peptídeos e resultados de ensaios biofísicos aprimorou o desenvolvimento de novos AMPs. Especificamente, redes adversárias generativas e *autoencoders* variacionais (VAEs) aumentaram a eficiência da geração de AMP, particularmente contra microrganismos patogênicos, com amplas aplicações na medicina e biologia (R. Wang et al., 2024). No entanto, esses modelos muitas vezes desconsideram a estabilidade da estrutura secundária dos AMPs, hemólise ou toxicidade, essenciais para a atividade antibacteriana. Nos últimos anos, vários estudos que combinaram o uso de modelos de *reinforcement learning* (RL) com outros modelos de *deep learning* mostraram resultados satisfatórios na geração de novos AMPs (Zhang et al., 2023).

Um estudo recente propõe um quadro abrangente (Diff-AMP) que integra geração, triagem, previsão e otimização de AMP, utilizando tecnologias de difusão, pré-treinamento e otimização iterativa para avançar na pesquisa de AMP. Eles usam um novo modelo de geração de AMP (RelGen) que integra o modelo de difusão termodinâmica com aprendizado por reforço. O mecanismo de atenção do modelo, informado pelo processo de difusão termodinâmica, captura efetivamente a interdependência entre aminoácidos e aprimora a caracterização da propagação da informação durante as atualizações de representação de nós. Essa abordagem permite uma simulação mais precisa da conformação de AMP e interações de aminoácidos. Isso permite a produção de candidatos a AMP que são tanto altamente diversificados quanto biologicamente ativos (R. Wang et al., 2024).

Outro modelo gerativo promissor é o modelo de partícula antimicrobiana multiobjetivo dirigido por hipervolume (HMAMP) (L. Wang et al. 2024). Este modelo inovador assume a tarefa complexa de aprender o espaço latente que abriga peptídeos com atributos conflitantes diversos. Além disso, o HMAMP filtra efetivamente os AMPs



com propriedades favoráveis de medicamento, facilitando assim sua aplicação como tratamentos eficazes. Este modelo demonstrou a capacidade de gerar AMPs com maior diversidade e potência do que outros métodos de referência, incluindo LSTM gerativo, redes neurais convolucionais profundas e algoritmos evolucionários. Esses candidatos exibiram características alinhadas aos padrões estabelecidos de AMPs eficazes, incluindo natureza catiônica, anfifilicidade e características estruturais favoráveis (L. Wang et al. 2024). É importante destacar que em ambos os estudos anteriores, os candidatos a AMP gerados com estes modelos não foram testados *in vivo* contra diferentes microrganismos.

## 2. Justificativa

O uso inadequado e indiscriminado de antibióticos, combinado com o aumento descontrolado de suas dosagens, resultou no surgimento de diversas bactérias resistentes a medicamentos. Essas bactérias tornaram-se uma grave ameaça à saúde pública no século XXI (Wang et al., 2021). Uma das estratégias promissoras para enfrentar esse problema global é o uso AMPs. Essas moléculas são geralmente catiônicas, curtas, com até 50 aminoácidos, e apresentam superfícies anfifílicas (Wang et al., 2021). Os AMPs possuem um amplo espectro de atividade contra vários microrganismos patogênicos (Boparai e Sharma, 2020), sendo mais conhecidos por suas propriedades antibacterianas de amplo espectro (Mahlapuu et al., 2016).

Os AMPs compartilham características biofísicas e um mecanismo de ação que envolve a interrupção da membrana celular, por meio de interações hidrofóbicas ou eletrostáticas, resultando na morte da célula (Witten e Witten, 2019). O avanço na descoberta de novos medicamentos tem sido significativamente impulsionado pela integração de AI com métodos tradicionais. Algoritmos ML transformaram a pesquisa ao identificar padrões complexos em grandes volumes de dados, enquanto redes neurais generativas têm sido capazes de criar novas estruturas moleculares com propriedades

desejadas, potencialmente levando a tratamentos inovadores para várias doenças. Algoritmos de RL otimizam processos como triagem virtual e ancoragem molecular, tornando a descoberta de medicamentos mais eficiente (Lavecchia, 2019).

O uso de AI pode superar muitas limitações das terapias baseadas em peptídeos, acelerando a descoberta de AMPs e aprimorando suas propriedades antimicrobianas e seletividade (Witten e Witten, 2019). No entanto, a maioria dos novos agentes antibacterianos em desenvolvimento ainda é direcionada contra bactérias Gram-positivas, enquanto bactérias Gram-negativas continuam a carecer urgentemente de novos tratamentos (Wang et al., 2021). Atualmente, quatro dos seis principais patógenos responsáveis por mortes associadas à resistência são bactérias Gram-negativas.

Diante desse cenário, este trabalho propõe o desenvolvimento de uma nova metodologia para projetar peptídeos com atividade contra bactérias Gram-negativas, utilizando *deep reinforcement learning*. Essa abordagem permitirá o desenvolvimento de estratégias terapêuticas mais personalizadas e eficazes.

### 3. Objetivos

#### 3.1. Objetivo geral

Desenvolver novos AMPs com atividade contra bactérias Gram-negativas usando um modelo de *Deep Reinforcement Learning*.

#### 3.2. Objetivos específicos

- Criar uma base de dados de peptídeos antibacterianos com atividade frente a bactérias Gram-negativas.
- Gerar novas sequências com possível atividade antibacteriana, utilizando aprendizado por reforço molecular (MRL).
- Selecionar, dentre todas as sequências geradas, aquelas com maior probabilidade de apresentar atividade antibacteriana.
- Determinar a concentração inibitória mínima das sequências selecionadas, frente às cepas bacterianas selecionadas.
- Avaliar a atividade hemolítica dos peptídeos frente a eritrócitos de camundongo.
- Simular as possíveis interações dos peptídeos com atividade comprovada, frente a modelos de membranas de bactérias Gram-negativas.

## 4. Material e métodos

### 4.1. Criação das bases de dados

Foi utilizado sequências de AMP disponíveis publicamente para treinar e testar nosso gerador e preditores de AMP. Para construir um conjunto de dados positivo não redundante, primeiro foi extraído todas as sequências com atividade frente a bactérias Gram-negativas da base de dados GRAMPA (*Giant Repository of AMP Activities*) (Witten e Witten, 2019) (Figura 1A). Uma filtragem de dados foi realizada para excluir as sequências que possuem aminoácidos não naturais. Foi utilizada o software CD-hit (Li e Godzik, 2006) para agrupamento e filtragem dos peptídeos com similaridade de 0,95 (maior a 95%), removendo sequências duplicadas, e produzindo um conjunto não redundante (positivo) de 684 sequências distintas (Figura 1B).

O treinamento e o teste de modelos de classificação binária requerem um conjunto negativo, uma coleção de peptídeos que não possuem atividade antimicrobiana. Como não há catálogos de sequências para peptídeos desprovidos de atividade antimicrobiana, estudos de campo normalmente selecionam suas sequências não-AMP do UniProt (UniProt Consortium, 2019). De UniProt foram extraídas as sequências com 50 resíduos de aminoácidos ou menos em comprimento (correspondendo ao comprimento máximo do peptídeo no conjunto AMP), excluindo aquelas com anotações contendo qualquer uma das 16 palavras-chave a seguir relacionadas a atividades antimicrobianas: antimicrobiano, antibiótico, antibacteriano, antiviral, antifúngico, antimalárico, antiparasitário, antiprotista, anticâncer, defesa, defensina, catelicidina, histatina, bacteriocina, microbicida, fungicida (Figura 1C). Depois, as sequências com resíduos diferentes dos 20 aminoácidos padrão foram removidas. Finalmente, no conjunto de dados foi aplicado, também, o software CD-hit (Li e Godzik, 2006) para removendo sequências duplicadas com similaridade de 0,95 (Figura 1D). Por fim, 684 sequências foram amostradas (aleatoriamente) do conjunto restante de 16592 não-AMPs, correspondendo ao número e à distribuição de comprimento das sequências no conjunto positivo.

Posteriormente, ambos os conjuntos de dados foram unificados para serem utilizados no treinamento dos modelos de predição (Figura 1E). Foi criada uma base de dados final de 1368 sequências.

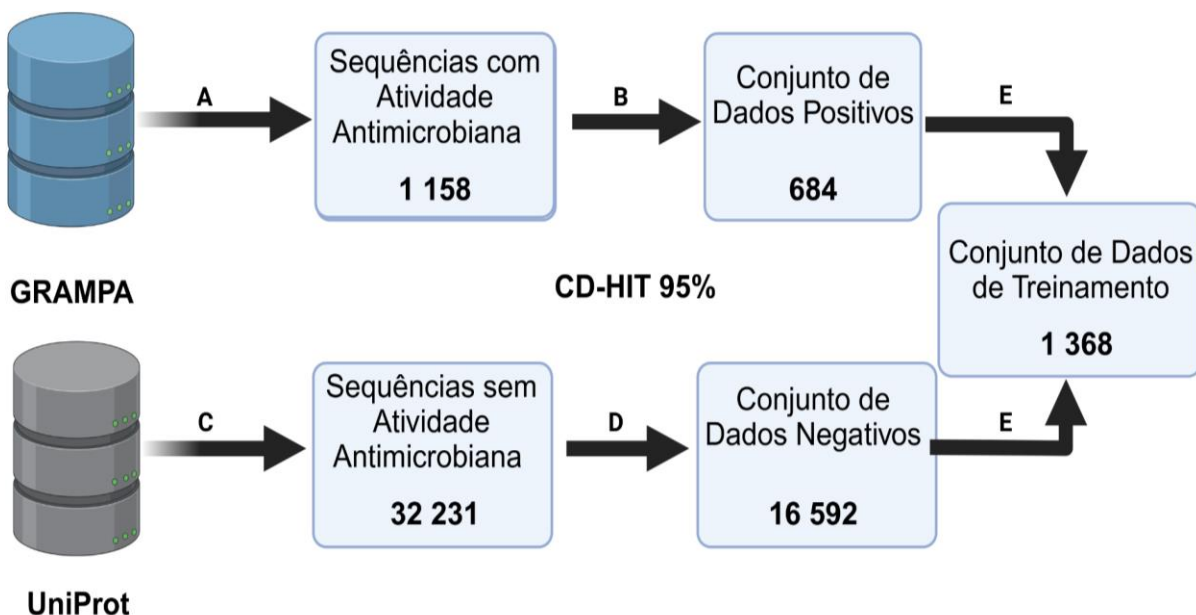


Figura 1: Criação do banco de dados para treinamento de modelos preditivos e generativos. A: seleção de sequências que possuem atividade contra bactérias Gram-negativas no banco de dados GRAMPA (foram extraídas 1.158 sequências). B: utilizou-se o software CD-hit para eliminar sequências com similaridade superior a 95%, eliminando duplicatas (o número de sequências foi reduzido para 684). C: seleção de sequências que não apresentam atividade antimicrobiana no banco de dados UniProt, conforme diversos critérios de seleção descritos no texto (foram extraídas 32.231 sequências). D: Da mesma forma, o software CD-hit foi utilizado para eliminar sequências com similaridade superior a 95%, eliminando duplicatas (o número de sequências foi reduzido para 16.592). E: união das 684 sequências do conjunto de dados positivos com 684 sequências, escolhidas aleatoriamente, do conjunto de dados negativos, para formar um conjunto final de dados de 1368 sequências.

#### 4.2. Extração de características de peptídeos

Para criar o modelo de predição, foi desenvolvida uma função de pontuação para a predição de atividade antimicrobiana. Em seguida, foi implementada uma rede neural convolucional (CNN) como codificador, com uma camada de perceptron multicamada (MLP) no topo do modelo para classificação binária de atividade antimicrobiana, utilizando as informações das sequências da base de dados criada. A base de dados criada foi dividida em conjuntos de treinamento e avaliação (80% para treinamento e

20% para avaliação) aleatoriamente. Os conjuntos de treinamento e avaliação foram utilizados para treinamento e avaliação, respectivamente, deste modelo de predição.

### **4.3. Espaço conformacional de AMP**

Em seguida, foi definido o espaço conformacional de atividade antimicrobiana, onde foi definida quais sequências são permitidas e quais são proibidas a partir da sequência de aminoácidos geradas. Foi restrito o tamanho das sequências, até um máximo de 16 aminoácidos, porque os peptídeos de maior tamanho são muito mais caros de sintetizar. A quantidade de ácido aspártico nas sequências foi restrita a zero e as de lisina a pelo menos três; isso com o propósito de obter peptídeos catiônicos, os quais possuem maior atividade frente às membranas negativas das bactérias. As restrições corretas e seus respectivos filtros são fundamentais para a otimização e geração de novas variantes do peptídeo. O processo de definição do espaço conformacional é iterativo e consiste no ajuste do mesmo visando variantes efetivas.

### **4.4. Aprendizado por reforço (MRL)**

O agente utilizado foi um modelo de linguagem com uma arquitetura de rede neural recorrente LSTM-LM (*Long Short-Term Memory-Language Models*) (Hochreiter et al., 1997) treinado em uma parte do banco de dados SwissProt (Bairoch et al., 2000). Como o modelo pré-treinado, ele pode produzir uma grande diversidade de sequências. Portanto, foi ajustado para produzir sequências que possuem apenas atividade antibacteriana, usando o banco de dados positivo criado. O método de gradiente utilizado foi o PPO (Otimização de Política Próxima) (Schulman et al., 2017). O modelo de classificação gera como saída um valor logit (regressão logística) recortado para um intervalo de  $[-10, 10]$ . Para manter uma relação entre o logaritmo da probabilidade da sequência de aminoácidos e a estabilidade dessa sequência, foi utilizado os valores de probabilidade de log do modelo pré-treinado ESM-2 (Rives et al., 2021), como um proxy para estabilidade. Este modelo de MRL utiliza esses parâmetros para atribuir um valor de recompensa às sequências geradas, sendo que quanto melhor for sua classificação

e estabilidade, maior será este valor. O valor máximo de recompensa é de 10 e o mínimo é de -10.

#### **4.5. Ajustes do modelo AMP-Bert**

Foi utilizado um novo modelo de linguagem desenvolvido por Lee et al. (2023), chamado AMP-BERT, que melhora os modelos de classificação AMP existentes, para validar a classificação das novas sequências geradas. O modelo AMP Bert é baseado na API Hugging Face Transformers (Wolf et al., 2019) e no ProtBERT-BFD, um modelo baseado em BERT que foi pré-treinado com aproximadamente 2,1 bilhões de sequências de proteínas do Big Fantastic Database (BFD) (Steinegger e Söding, 2018). Quando um processo de ajuste fino é implementado, algumas das últimas camadas do codificador do modelo pré-treinado são treinadas com a camada densa final usando um conjunto de dados específico da tarefa. As saídas do codificador de nível superior são, portanto, mais aplicáveis às tarefas fornecidas. É importante observar que o domínio de conhecimento do modelo pré-treinado deve ser relevante para o domínio da tarefa. Suponha que o conhecimento pré-treinado do modelo e a tarefa ser ajustado não sejam relacionados. Nesse caso, o processo de ajuste fino fará com que o modelo reaprenda do zero, anulando efetivamente o processo de pré-treinamento anterior.

No ajuste fino deste modelo, empregamos a base de dados, criada anteriormente. O uso desta base de dados permite o retreinamento do modelo para uma predição específica da atividade antibacteriana dos peptídeos. A base de dados foi dividida em conjuntos de treinamento e avaliação (80% para treinamento e 20% para avaliação) aleatoriamente, para treinamento e avaliação, respectivamente, deste modelo. As probabilidades previstas maiores ou iguais a 0,5 são classificadas como antibacterianas, enquanto probabilidades menores que 0,5 são classificadas como não antibacterianas.

#### **4.6. Seleção das melhores sequências geradas**

Devido ao alto custo econômico e de tempo para avaliar a atividade antibacteriana de todas as novas sequências geradas, optou-se por selecionar apenas as sequências com maior probabilidade de ter atividade contra bactérias Gram-negativas. Entre as

sequências geradas pelo modelo de MRL, foram escolhidas todas as que obtiveram um valor de recompensa maior que 9,5. O valor de corte (9,5) foi escolhido porque sequências com o menor valor de recompensa teriam uma probabilidade muito menor de serem AMP potentes. Além disso, um valor de recompensa maior não foi escolhido, pois isso reduziria significativamente o número de sequências, perdendo várias sequências com grande potencial.

Posteriormente, foi utilizado o modelo AMP-Bert calibrado com nossa base de dados, descrito anteriormente, para validar a predição de atividade contra bactérias Gram-negativas. Descartaram-se todas as sequências que tiveram um valor de predição menor que 0,8 (sendo 1,0 o valor máximo). Com o objetivo de selecionar apenas aqueles que apresentavam maiores probabilidades de apresentar atividade antimicrobiana. As plataformas ToxinPred 3.0 (Rathore et al., 2023) e HemoPred (Win et al., 2017) foram utilizadas para prever a toxicidade e a atividade hemolítica das novas sequências, respectivamente. As sequências que podem apresentar atividade hemolítica e citotóxica, conforme os preditores, foram descartadas. Ambas plataformas foram escolhidas porque seus servidores web oferecem uma interface fácil de usar e incorporam os modelos mais eficazes em comparação com outros modelos da literatura. ToxinPred 3.0 alcança uma precisão de 0,98 e um coeficiente de correlação de Matthews de 0,81. HemoPred tem uma precisão e coeficiente de correlação de Matthews acima de 95% e 0,91, respectivamente. Também foram selecionadas as sequências com cargas líquidas positivas. Sequências com carga líquida menor que +2 foram descartadas, pois a carga positiva é uma característica essencial dos AMPs, o que lhes permite interagir com as membranas negativas das bactérias.

#### **4.7. Testes *in silico* do modelo de predição de AMP**

Antes dos testes biológicos *in vitro* das variantes dos peptídeos gerados para validação da ferramenta, foram realizados testes de avaliação *in silico*. A estrutura secundária dos peptídeos obtidos, foi gerado usando a plataforma online AlphaFold 3 (Abramson et al., 2024). O modelo AlphaFold 3 foi escolhido em detrimento do modelo anterior (AlphaFold 2), pois este novo modelo apresenta melhor precisão na previsão da estrutura de pequenas proteínas, principalmente na previsão de estruturas de alfa-hélice.



A plataforma online Heliquist (Gautier et al., 2008) foi utilizada para gerar a representação de roda helicoidal das sequências selecionadas. Esta mesma plataforma foi utilizada para calcular os valores de hidrofobicidade e momento hidrofóbico de cada peptídeo previamente selecionado. O cálculo do índice de Boman e do ponto isoelétrico, de cada peptídeo, foi realizado utilizando as plataformas APD3 (Wang et al., 2015) e Thermo Fisher (Wang et al., 2015), respectivamente.

Adicionalmente, foi realizada uma nova previsão da atividade antimicrobiana dos peptídeos selecionados utilizando 4 modelos preditivos de última geração. Os modelos utilizados foram: AMPScanner v2 (Veltri et al., 2018), AI4AMP (Lin et al., 2021), ampir (Fingerhut et al., 2020) e Macrel (Santos-Júnior et al., 2020). Essas previsões foram utilizadas para validar as previsões feitas usando o modelo AMP-Bert calibrado com nossa base de dados.

#### **4.8. Síntese dos peptídeos e determinação de massa moleculares por espectrometria de massa MALDI-ToF.**

Os 15 peptídeos selecionados foram adquiridos da empresa SYNBIOTech (New York, USA) com uma pureza de >95%. Os peptídeos purificados foram analisados qualitativamente utilizando espectrômetro de massa MALDI-ToF UltraFlex III (Bruker Daltonics). Para as análises, os peptídeos liofilizados foram dissolvidos em água ultrapura, misturados em uma solução saturada de uma matriz constituída por ácido  $\alpha$ -ciano-4-hidroxicinâmico (1:3), depositados em uma placa do tipo Anchorchip e deixados cristalizar à temperatura ambiente. A calibração foi realizada utilizando *Peptide calibration standard II* (Bruker Daltonics) como padrões de massa molecular e adquiridos em modo de operação refletido positivo (RP\_700-3000 Da).

#### **4.9. Cepas bacterianas e meios de crescimento:**

As cepas bacterianas usadas neste estudo incluíram a bactéria Gram-positiva *Staphylococcus aureus* (ATCC 25923), e as bactérias Gram-negativas *Klebsiella pneumoniae* (ATCC 13883), *Escherichia coli* (ATCC 25922) e *Acinetobacter baumannii* (ATCC 13304) e 2 cepas clinicamente isoladas apresentadas como *A. baumannii* 8892 e *A. baumannii* 5015. As bactérias foram cultivadas em caldo Mueller Hinton (MH).

#### **4.10. Determinação da concentração inibitória mínima (MIC)**

A concentração inibitória mínima (MIC) foi determinada utilizando o método de microdiluição em caldo, conforme os padrões CLSI (CLSI, 2012), em microplaca de 96 poços de fundo chato, utilizando MHC. Brevemente, os peptídeos foram preparados e estocados a uma concentração de 1 mM e, quando adicionado à placa, foi diluído em 50  $\mu\text{L}$  de MHC para a concentração de 256  $\mu\text{M}$ , seguido de diluições seriadas de 128, 64, 32, 16, 8, 4, 2 e 1  $\mu\text{M}$ . As placas foram mantidas à temperatura ambiente até à adição da suspensão bacteriana. Posteriormente, 50  $\mu\text{L}$  de suspensão bacteriana ajustada ( $1 \times 10^7 \text{ UFC.mL}^{-1}$ ) foram aplicados em cada poço da microplaca contendo os peptídeos, almejando a concentração final do ensaio de  $5 \times 10^5 \text{ UFC.mL}^{-1}$ . Culturas bacterianas em MHC foram utilizadas como controles negativo, e os antibióticos AMK e Poli b (nas mesmas concentrações dos peptídeos) foram utilizadas como controles positivo. As microplacas foram incubadas a 37 °C e a leitura da DO foi realizada a 600 nm em leitor de microplacas (Multiskan Go, Thermo Scientific), após 18 h de incubação com agitação. Os experimentos foram desenvolvidos com três réplicas biológicas.

#### **4.11. Avaliação de atividade hemolítica**

O ensaio para determinação da atividade hemolítica foi realizado conforme o método descrito nos trabalhos de Park et al. (2004), com algumas modificações. O sangue utilizado no experimento foi sangue de camundongo heterogênico Swiss, aprovado pelo Comitê de Ética no Uso de Animais (CEUA) da UCDB, com o protocolo de número 020/2018. Para a coleta do sangue utilizaram-se seringas contendo como anticoagulante ácido etilenodiamino tetra-acético (EDTA). Foram coletados 1 mL de sangue de camundongo (anestesiado com quetamina e xilazina) por punção cardíaca, e de sangue bovino foram coletados 2 mL por punção venosa por meio da veia jugular. O sangue foi centrifugado a  $400 \times g$  (centrífuga Z 326 K, HERMLE Labortechnik) por 5 min, descartando-se o sobrenadante. O precipitado (eritrócitos) foi lavado três vezes com solução salina e ressuspendido na mesma solução, com centrifugações a  $400 \times g$  por 2 min cada. Os 50  $\mu\text{L}$  da suspensão foram colocados em placa de 96 poços, adicionando 50  $\mu\text{L}$  de peptídeo na concentração de 64  $\mu\text{M}$ , com diluições seriadas de 32, 16, 8, 4 e 2  $\mu\text{M}$ . Posteriormente, as amostras foram incubadas à temperatura ambiente por 1 h.

Então, a placa foi centrifugada a  $1400 \times g$  durante 2 min e 50  $\mu\text{L}$  do sobrenadante foi transferido para uma nova placa de 96 poços. A hemoglobina liberada foi monitorada por meio do leitor de microplacas (Multiskan Go, Thermo Scientific), com medição da absorbância ao comprimento de onda de 415 nm. Para controle positivo (100% de lise) foi utilizado Triton X-100 (2% v/v), e para controle negativo, os eritrócitos foram tratados com solução salina.

#### 4.12. Simulações de dinâmica molecular

O CHARMM-GUI *membrane builder* (Jo et al., 2008) foi aplicado para montar modelos de bicamada de membrana para bactérias Gram-negativas. De acordo com estudos anteriores, as membranas internas das bactérias Gram-negativas compreendem uma proporção de 3:1 de palmitoil oleoil fosfatidiletanolamina (POPE) e palmitoil oleoil fosfatidilglicerol (POPG) (Mirnejad et al., 2023). Uma bicamada lipídica composta por 160 fosfolipídios (80 em cada monocamada) foi construída em uma caixa retangular usando o CHARMM-GUI *membrane builder*, baseada no campo de força todo-átomo CHARMM36 (Huang et al., 2016). A bicamada foi solvatada em uma caixa de água (modelo TIP3P) com tamanho de aproximadamente  $69 \times 69 \times 120 \text{ \AA}^3$ . Então, foram adicionados íons para neutralizar a caixa e também sal NaCl a 150 mM para mimetizar o ambiente fisiológico. Afim de preparar uma membrana bem equilibrada, a mesma foi simulada usando o protocolo padrão de minimização e equilíbrio em seis etapas empregado no CHARMM-GUI. Todas as simulações foram realizadas usando o pacote de simulação MD GROMACS v2023.4 (Páll et al., 2020). A pré-equilíbrio consistiu em três ocorrências de acoplamento de temperatura por 1,25 ns, duas etapas de acoplamento de temperatura e pressão (10 ns), e depois uma última fase de 50 ns, cada uma com temperatura e pressão constantes. A estrutura final foi usada como modelo inicial para duas simulações separadas de peptídeo-membrana. Na primeira simulação, o peptídeo foi colocado a 35  $\text{\AA}$  de distância da monocamada superior das bicamadas. O peptídeo, na segunda, foi introduzido no centro da membrana. Ambas as simulações foram equilibradas usando o mesmo protocolo descrito acima, com uma mudança na segunda, onde todos os tempos de simulação foram dobrados. A produção de MD de ambas as simulações foi realizada duplicando, cada uma por 200 ns. As simulações

foram analisadas pela média do desvio quadrático médio (RMSD), flutuação quadrática média (RMSF) e frequências relativas dos ângulos doador-aceitador de hidrogênio. As ocupações de ligações de hidrogênio foram calculadas usando VMD aplicando os critérios de corte: distância  $\leq 3,5$  Å; ângulo  $\leq 30$  graus.

## 5. Resultados

### 5.1. Seleção das melhores sequências

Das sequências geradas pelo modelo de MRL, foram escolhidas todas as que obtiveram um valor de recompensa maior que 9,5. Obteve-se um total de 184 novas sequências de peptídeos. No total, das 184 sequências, foram selecionados 15 sequências que atenderam as restrições anteriormente descritas. A previsão da estrutura terciária dos 15 peptídeos, prevista pelo AlphaFold 3, pode ser vista na Figura 2. Na mesma figura se observa as sequências e projeções helicoidais de cada peptídeo.

Os resultados dos cálculos das diferentes propriedades físico-químicas (carga líquida, hidrofobicidade, momento hidrofóbico, índice de Boman e ponto isoelétrico) podem ser vistos na Tabela 1. Todos os peptídeos selecionados possuem uma carga líquida entre +2 e +5. A maioria destes apresenta carga líquida de +3 e +5, apenas Pep1 e Pep4 apresentaram carga de +2 e +5, respectivamente. Quanto à hidrofobicidade, apenas Pep 4, 5, 7, 8 e 14 apresentaram valores inferiores a 0,3. Os 15 peptídeos apresentaram valores de momento hidrofóbico em torno de 0,3, Pep 4, 7 e 11 apresentaram valores ligeiramente inferiores, enquanto Pep 6 e 10 apresentaram valores superiores a 0,5. Com exceção dos Pep 1 e 2, os restantes peptídeos apresentaram valores de índice de Boman inferiores a 0. Por outro lado, os valores do ponto isoelétrico destes peptídeos oscilam entre 10,42 e 11,75.

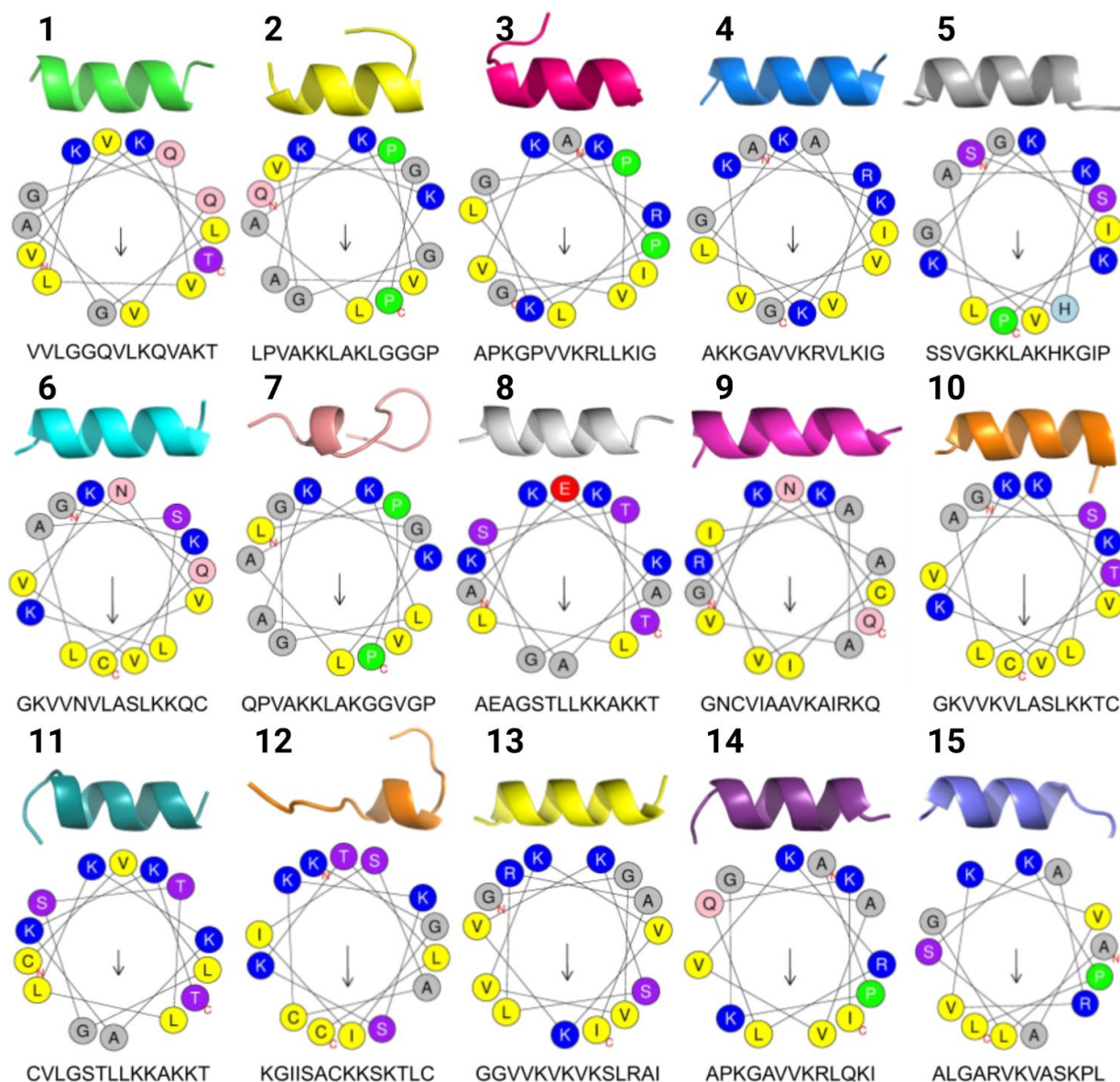


Figura 2: Estrutura tridimensional, projeções de roda helicoidal e sequências dos 15 peptídeos (Pep 1-15, respectivamente). Os círculos azuis representam os resíduos carregados positivamente arginina (R) e lisina (K); o azul-claro o resíduo de histidina (H); o vermelho o resíduo carregado negativamente ácido glutâmico (E); os roxos os resíduos polares não carregados serina (S) e treonina (T); os rosa os resíduos polares não carregados glutamina (Q) e asparagina (N); o verde o resíduo de prolina (P); os cinza os resíduos de alanina (A) e glicina (G); os amarelos representam resíduos hidrofóbicos de valina (V), isoleucina (I), leucina (L), triptofano (W) e o resíduo de cisteína (C). As setas indicam as direções e magnitudes do momento hidrofóbico: N, terminal amino; C, extremidade carboxila. As estruturas tridimensionais foram geradas usando AlphaFold 3.

Tabela 1. Propriedades físico-químicas dos peptídeos utilizados neste estudo.

<b>Peptídeo</b>	<b>Carga líquida</b>	<b>Hidrofobicidade</b>	<b>Momento hidrofóbico</b>	<b>Índice de Boman</b>	<b>Ponto isoeletrico</b>
Pep 1	+2	0,46	0,32	-0,35	10,70
Pep 2	+3	0,37	0,35	-0,61	10,99
Pep 3	+4	0,37	0,30	0,36	11,68
Pep 4	+5	0,20	0,30	0,69	11,75
Pep 5	+4	0,13	0,31	1,15	11,16
Pep 6	+3	0,36	0,54	0,45	10,42
Pep 7	+3	0,22	0,29	0,20	10,99
Pep 8	+3	0,02	0,38	1,52	10,69
Pep 9	+3	0,34	0,39	0,90	10,68
Pep 10	+4	0,37	0,56	0,16	10,69
Pep 11	+4	0,34	0,28	0,57	10,69
Pep 12	+4	0,35	0,45	0,82	10,27
Pep 13	+4	0,33	0,40	0,38	11,68
Pep 14	+4	0,24	0,40	1,13	11,68
Pep 15	+3	0,34	0,38	0,40	11,57

Além disso, a probabilidade de cada peptídeo apresentar atividade antimicrobiana foi calculada usando 4 modelos de predição de última geração. Os resultados dessas previsões estão representados na Figura 3. Os valores de predição (representado na Figura 3) são indicadores numéricos gerados por modelos de aprendizado de máquina que representam a probabilidade de que o peptídeo, possua atividade antimicrobiana. Esses valores geralmente variam entre 0 e 1, sendo que valores próximos a 1 sugerem uma alta probabilidade de que o peptídeo apresente a atividade antimicrobiana, enquanto valores próximos a 0 indicam o contrário. A Figura 3 inclui também a previsão obtida através do modelo AMP\_Bert, ajustado através do conjunto de dados criado neste estudo (AMP\_Bert (GRAMPA)). Ao analisar essas previsões, podemos ver como os modelos AMP\_Bert, AI4AMP e ampir previram todas as sequências como prováveis peptídeos antimicrobianos (valores de previsão > 0,5). Por outro lado, embora os modelos

de predição Macrel e AMPScanner v2 previram a maioria das sequências como antimicrobianas, diversas sequências foram previstas como não antimicrobianas (valores de predição  $<0,5$ ). Eles foram previstos como não antimicrobianos, Pep 1, 5 e 13 de acordo com AMPScanner v2 e Pep 5, 7, 8, 12, 14 e 15, de acordo com Macrel. Curiosamente, apenas o Pep 5 foi previsto por ambos os modelos como não antimicrobiano.

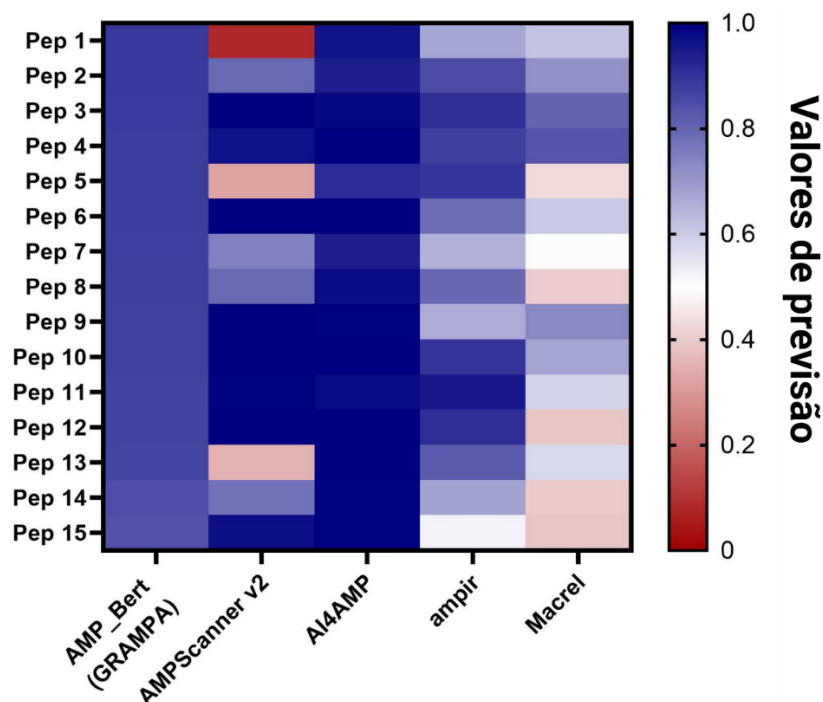


Figura 3: *Heatmap* de previsão da atividade antimicrobiana dos quinze peptídeos (Pep 1-15) usando 5 modelos preditivos de última geração. O azul representa valores de predição superiores a 0,6. O vermelho representa valores de predição inferiores a 0,4. O branco representa valores de previsão entre 0,4 e 0,6. Os valores de predição representam a probabilidade de que uma sequência possua atividade antimicrobiana. Os valores de predição  $> 0,5$  significa que o modelo prevê com mais de 50% de confiança que o peptídeo tem atividade antimicrobiana. Em contrapartida, valores de predição  $< 0,5$ , indicam que é mais provável que o peptídeo não tenha essa atividade.

## 5.2. Caracterização por MALDI-ToF dos peptídeos antimicrobianos obtidos comercialmente.

Os peptídeos antimicrobianos foram analisados por espectrometria de massa em modo MALDI-ToF. Os espectros de massa, após adquiridos, mostraram o íon molecular  $[M+H]^+$  que correspondem à massa molecular mais um átomo de hidrogênio que

correspondente a um próton. A Tabela 2, mostra a sequência, a fórmula molecular, o íon  $[M+H]^+$  encontrado e calculado para cada peptídeo antimicrobiano.

Tabela 2. Dados correspondente a sequência, a fórmula molecular, e os íons  $[M+H]^+$  encontrado e calculado para os AMPs. Os valores de  $[M+H]^+$  calculados e os obtidos experimentalmente praticamente os mesmos.

Peptídeo	Sequência (N-terminal amidado)	$[M+H]^+$ Encontrada (Da)	$[M+H]^+$ Calculada (Da)	Formula Molecular
Pep 1	VVLGGQVLKQVAKT-NH2	1438,71	1439,88	C <sub>65</sub> H <sub>119</sub> N <sub>19</sub> O <sub>17</sub>
Pep 2	LPVAKKLAKLGGGP-NH2	1347,61	1348,80	C <sub>63</sub> H <sub>114</sub> N <sub>18</sub> O <sub>14</sub>
Pep 3	APKGPVVKRLLKIG-NH2	1474,82	1475,99	C <sub>69</sub> H <sub>127</sub> N <sub>21</sub> O <sub>14</sub>
Pep 4	AKKGAVVKRVLKIG-NH2	1466,04	1466,98	C <sub>67</sub> H <sub>128</sub> N <sub>22</sub> O <sub>14</sub>
Pep 5	SSVGKKLAKHKGIP-NH2	1449,17	1449,86	C <sub>65</sub> H <sub>117</sub> N <sub>21</sub> O <sub>16</sub>
Pep 6	GKVVNVLASLKKQC-NH2	1486,41	1486,94	C <sub>65</sub> H <sub>120</sub> N <sub>20</sub> O <sub>17</sub> S
Pep 7	QPVAKKLAKGGVGP-NH2	1349,80	1349,74	C <sub>61</sub> H <sub>109</sub> N <sub>19</sub> O <sub>15</sub>
Pep 8	AEAGSTLLKKAKKT-NH2	1445,83	1445,81	C <sub>63</sub> H <sub>117</sub> N <sub>19</sub> O <sub>19</sub>
Pep 9	GNCVIAAVKAIRKQ-NH2	1469,49	1470,89	C <sub>63</sub> H <sub>116</sub> N <sub>22</sub> O <sub>16</sub> S
Pep 10	GKVVKVLASLKKTC-NH2	1472,71	1473,98	C <sub>66</sub> H <sub>125</sub> N <sub>19</sub> O <sub>16</sub> S
Pep 11	CVLGSTLLKKAKKT-NH2	1488,74	1489,98	C <sub>66</sub> H <sub>125</sub> N <sub>19</sub> O <sub>17</sub> S
Pep 12	KGIISACKKSKTLC-NH2	1476,84	1479,95	C <sub>63</sub> H <sub>119</sub> N <sub>19</sub> O <sub>17</sub> S <sub>2</sub>
Pep 13	GGVVKVKVKSRLAI-NH2	1453,22	1453,94	C <sub>66</sub> H <sub>125</sub> N <sub>21</sub> O <sub>15</sub>
Pep 14	APKGAVVKRLQKI-NH2	1407,37	1407,86	C <sub>64</sub> H <sub>119</sub> N <sub>21</sub> O <sub>14</sub>
Pep 15	ALGARVKVASKPL-NH2	1309,47	1309,70	C <sub>59</sub> H <sub>109</sub> N <sub>19</sub> O <sub>14</sub>

No Apêndice B, se encontram os espectros de massa onde se observa o respectivo íon  $[M+H]^+$  encontrado para cada composto. Ao modo de exemplo, a Figura S3 I, mostra o espectro de massa do composto Pep 10, com um íon molecular  $[M+H]^+$  a  $m/z$  1472,71 Da, que sugere a fórmula molecular C<sub>66</sub>H<sub>125</sub>N<sub>19</sub>O<sub>16</sub>S ( $[M+H]^+$  calculado igual 1473,98).



### 5.3. Concentração Inibitória Mínima (MIC)

Com o intuito de avaliar a eficácia dos peptídeos frente a diferentes cepas bacterianas, realizou-se um primeiro ensaio de MIC com os 15 peptídeos frente às cepas suscetíveis de *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Escherichia coli* e *Acinetobacter baumannii*. Observou-se que, para todos os peptídeos frente às cepas suscetíveis de *S. aureus* e *K. pneumoniae* não houve atividade antimicrobiana, até a concentração máxima testada (256  $\mu$ M) (Figura 4). Para a cepa suscetível de *E. coli* apenas o Pep 10 apresentou resultado na concentração de 256  $\mu$ M. É interessante notar que o peptídeo Pep 10 apresentou atividade contra cepas suscetíveis de *Acinetobacter baumannii*, demonstrando eficácia a uma concentração de 64  $\mu$ M.

Também foi realizado o MIC com o peptídeo Pep 10 frente às cepas resistentes de *Acinetobacter baumannii* (5015 e 8892). O peptídeo mostrou atividade contra essas cepas resistentes, com eficácia observada na concentração de 32  $\mu$ M. (Figura 5).

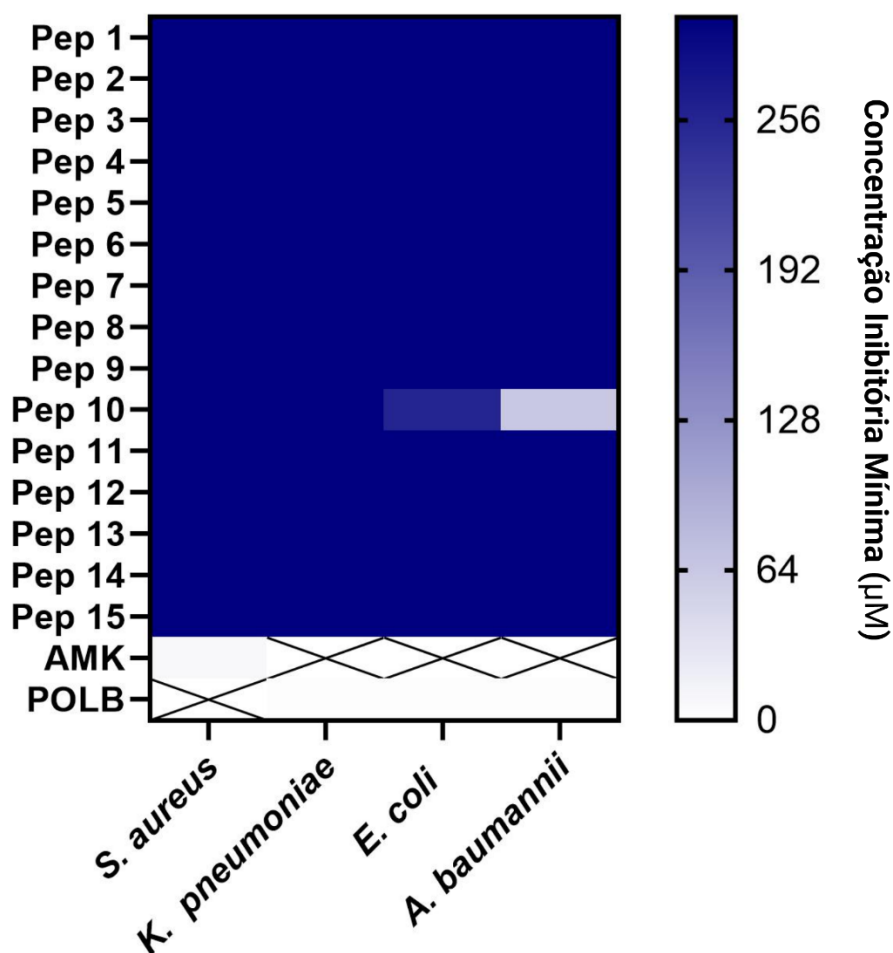


Figura 4: *Heatmap* de avaliação da atividade antimicrobiana *in vitro* dos quinze peptídeos (Pep 1-15) frente as cepas bacterianas suscetíveis (ATCC). Foi usado AMK (amicacina) como controle positivo para as bactérias *S. aureus* (Gram-positivo) e POLB (polimixina B) foi o controle positivo para as bactérias *K. pneumoniae*, *E. coli*, *A. baumannii* (Gram-negativo). Pep 10 apresentou atividade contra *E. coli* em 256  $\mu\text{M}$  e contra *A. baumannii* em 64  $\mu\text{M}$ . A cruz representa que nenhum teste foi feito.

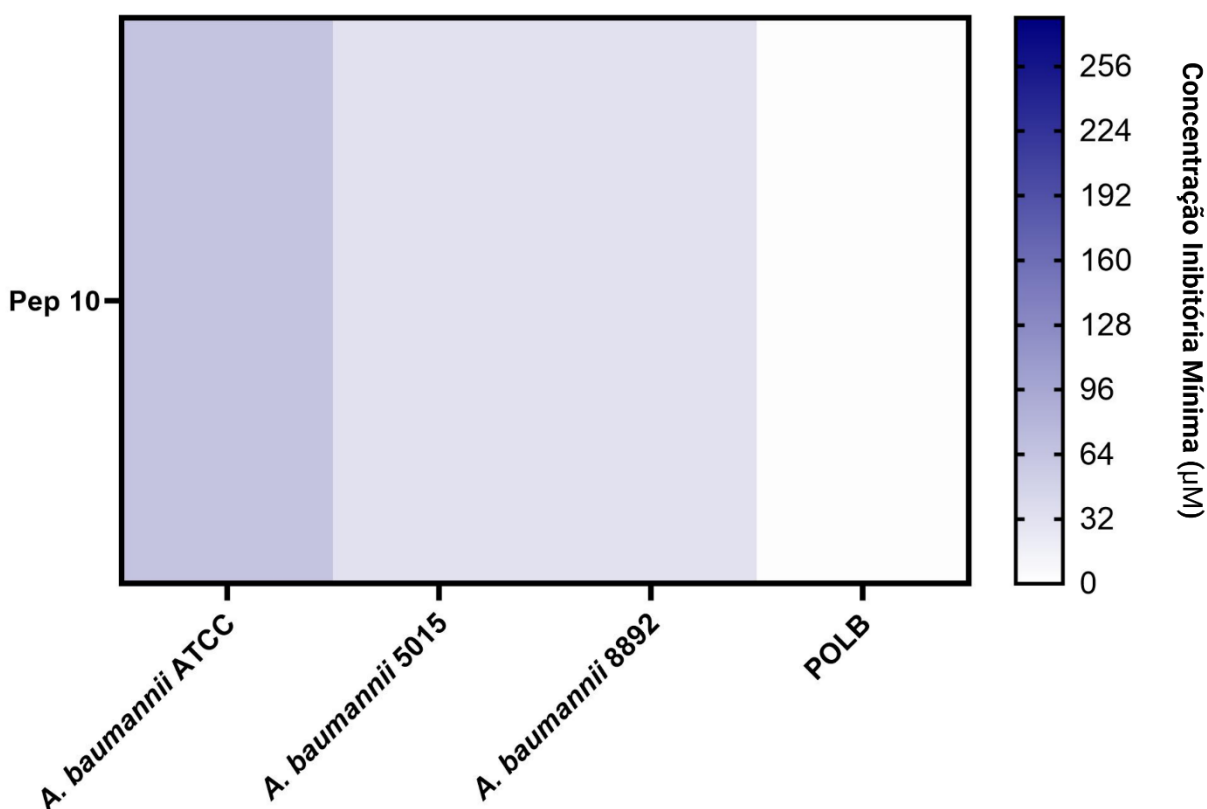


Figura 5: *Heatmap* de avaliação da atividade antimicrobiana *in vitro* do peptídeo 10 frente a cepa suscetível de *A. baumannii* (ATCC) e as cepas resistentes de *A. baumannii* (5015 e 8892, isolados clínicos de hospitais de Brasília). Foi usado POLB (polimixina B) como controle positivo.

#### 5.4. Atividade Hemolítica

Os peptídeos em estudo não apresentaram atividade hemolítica frente aos eritrócitos de camundongos até à concentração máxima testada de 64 µM, com exceção de Pep 3, que apresentou efeito hemolítico de 100% em concentrações começando em 32 µM (Figura 6).

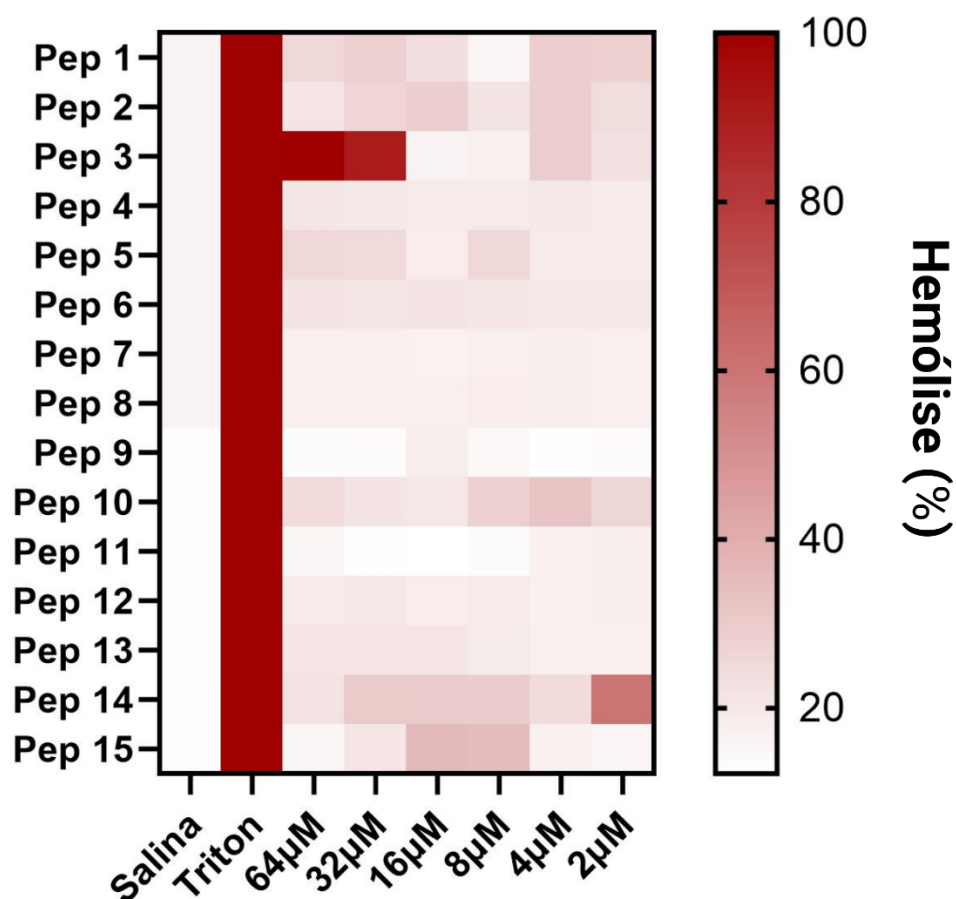


Figura 6: *Heatmap* de avaliação da atividade hemolítica dos quinze peptídeos (Pep 1-15) frente a eritrócitos de camundongos. Foi usado para controle positivo Triton X-100 (2% v/v), e para controle negativo, os eritrócitos foram tratados com solução salina.

### 5.5. Dinâmica molecular do Pep 10 na membrana lipídica

As simulações do Pep 10 na presença de membrana lipídica bacteriana (3 POPE: 1 POPG) foram realizadas em dois momentos distintos. Na primeira simulação, o Pep 10 estava localizado a uma distância de 35 Å, na folha externa da membrana. O Pep 10 foi inserido dentro da bicamada, perpendicularmente a ela, na segunda simulação. Ambas as simulações foram realizadas a 200 ns. A seguir, as trajetórias das simulações são analisadas quanto ao uso da raiz do desvio quadrático médio (RMSD) do *backbone* das proteínas e da flutuação média da raiz quadrada (RMSF) de cada resíduo; para validação da movimentação da cadeia principal da estrutura e de cada resíduo, respectivamente.

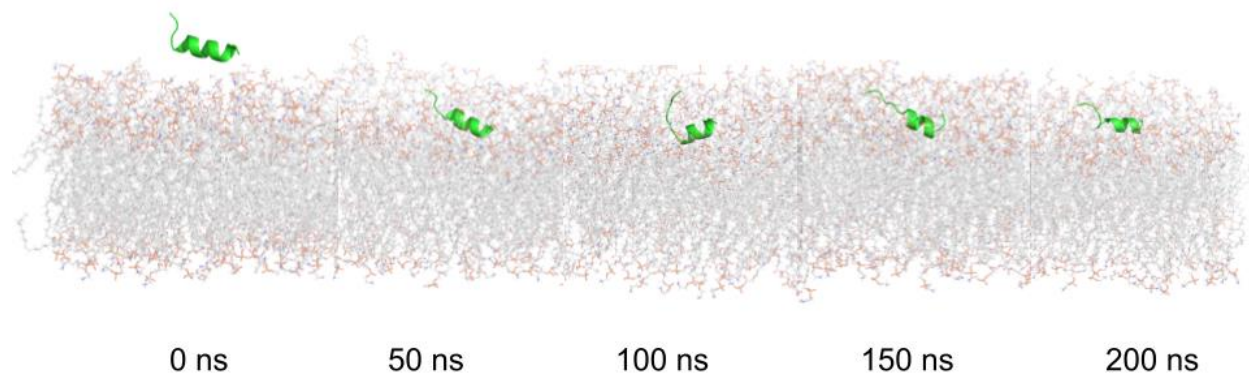


Figura 7: Simulações de MD de Pep 10 em um sistema de bicamadas lipídicas, POPE/POPG (3:1). As simulações foram realizadas durante 200 ns e aqui representadas em intervalos de 50 ns. A estrutura peptídica inicial foi colocada a ~35 Å das bicamadas de membrana.

Observando a trajetória da simulação do Pep 10 (Figura 7), localizado na parte externa da membrana, podemos perceber como logo após o início esse peptídeo começa a interagir com a membrana. O Pep 10 permanece interagindo com ele durante todo o tempo de simulação, o que coincide com a análise do número de interações do hidrogênio (Figura 10A).

Pep 10 exibiu um RMSD da amplitude entre 0,05 e 0,5 nm (Figura 8A). Nos primeiros 80 ns de simulação o RMSD aumentou gradativamente até atingir valores próximos a 0,5 nm. Posteriormente, a variação do RMSD se estabilizou, até que após 150 ns a amplitude permanecesse entre 0,2 e 0,35 nm. Essas variações do RMSD correspondem a uma perda da estrutura alfa-hélice do Pep 10 no C- terminal, ao entrar em contato com a membrana (Figura 7). Por outro lado, o Pep 10 inserido na membrana, apresentou uma variação RMSD abaixo de 0,2 nm ao longo da simulação, indicando uma preservação da estrutura  $\alpha$ -helicoidal inicial (Figura 8B).

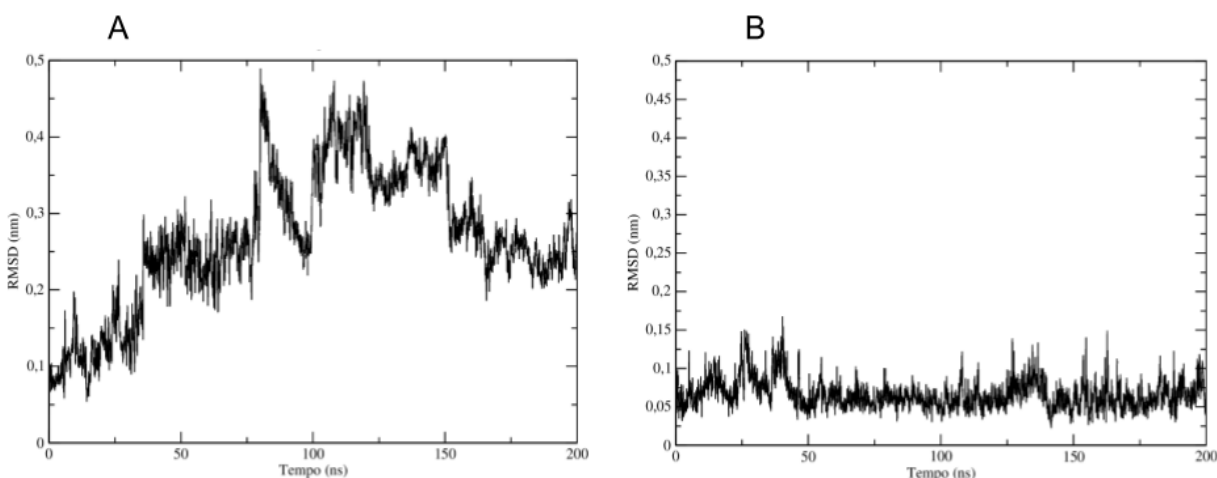


Figura 8: RMSD do *backbone* do Pep 10 durante o tempo de simulação de 200ns. A variação foi plotada tendo como ponto inicial a estrutura em 0ns. (A) O peptídeo a 35 Å da bicamada lipídica. (B) O peptídeo inserido na bicamada lipídica.

A análise RMSF de ambas as simulações (Figura 9) mostrou que os resíduos terminais geraram a maioria da variação. Uma diferença entre o RMSF de ambas as simulações é que o Pep 10 a 35 Å da membrana apresentou variação nos resíduos 8 e 9 (Figura 9A), não presente na outra simulação (Figura 9B). Além disso, o RMSF da primeira simulação apresentou maiores valores de variação (até 0,4 nm) em relação à segunda (até 0,1).

Ao analisar o número de ligações de hidrogênio entre o Pep 10 (fora da membrana) e os lipídios (POPE e POPG) da membrana (Figura 10A), observa-se que desde o início da simulação o Pep 10 começa a interagir com a membrana, mantendo um intervalo entre 4 e 10 interações. Esses valores permanecem constantes durante toda a simulação, exceto por um pequeno intervalo entre os tempos de simulação 180 e 190 ns, onde o Pep 10 atinge mais de 15 ligações de hidrogênio entre peptídeos e lipídios. No caso do Pep 10, no interior da membrana (Figura 10B), no início da simulação, há um aumento no número de ligações de hidrogênio, até atingir no máximo 20 interações. Posteriormente, esse número diminui, até que um intervalo entre 6 e 15 seja mantido durante a maioria da simulação.

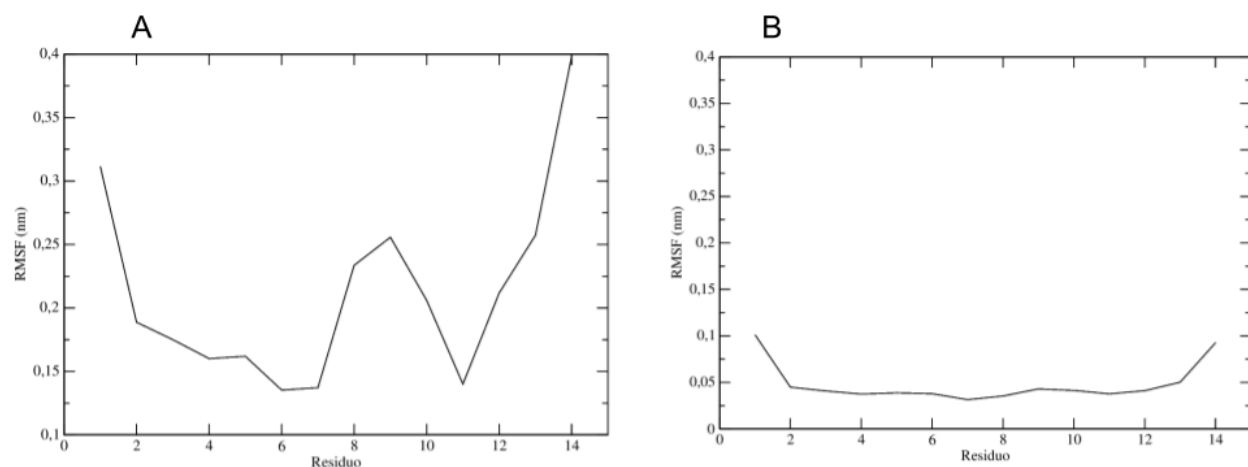


Figura 9: RMSF de todos os resíduos de Pep 10 durante o tempo de simulação de 200ns. A variação foi plotada tendo como ponto inicial a estrutura em 0ns. (A) O peptídeo a 35 Å da bicamada lipídica. (B) O peptídeo inserido na bicamada lipídica.

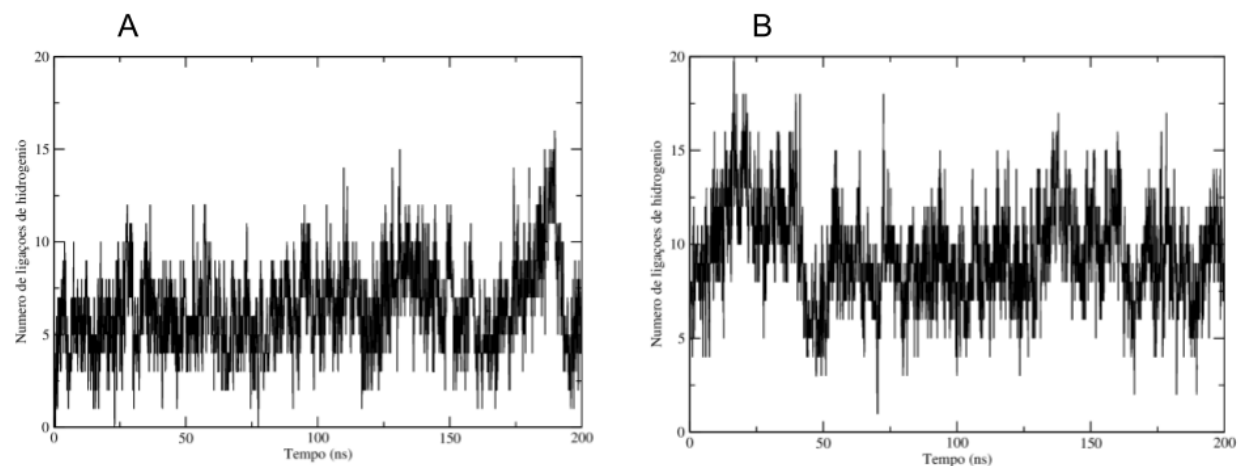


Figura 10: Número de interações de hidrogênio entre Pep 10 e a bicamada lipídica durante o tempo de simulação de 200ns. A variação foi plotada tendo como ponto inicial a estrutura em 0ns. (A) O peptídeo a 35 Å da bicamada lipídica. (B) O peptídeo inserido na bicamada lipídica.

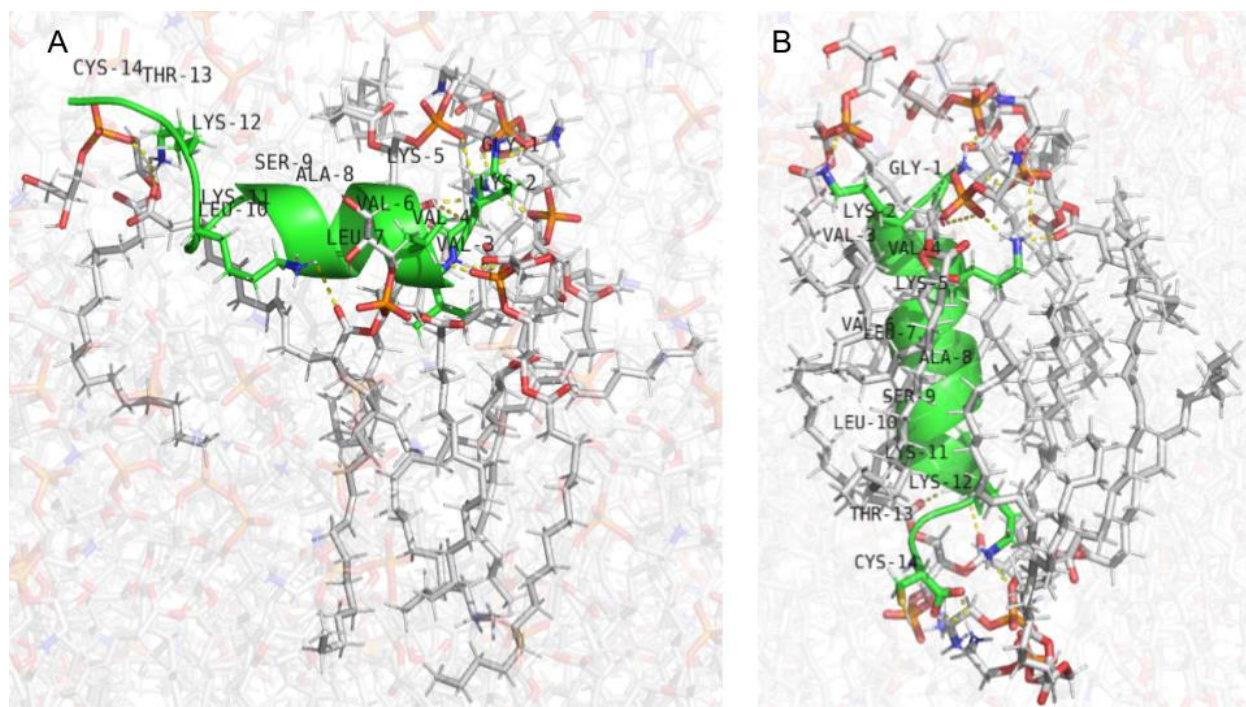


Figura 11: Interações entre Pep 10 e a membrana lipídica em 200 ns de simulação. (A) O peptídeo a 35 Å da bicamada lipídica, (B) o peptídeo inserido na bicamada lipídica, no início da simulação. As interações entre eles são representadas como linhas tracejadas amarelas. Pep 10 é de cor verde e os lipídios que interagem com ele são de cor cinza. Os átomos de oxigênio são coloridos em vermelho, os átomos de nitrogênio em azul e os átomos de hidrogênio em branco.

Na Figura 11A pode-se observar como o Pep 10 interage principalmente com o POPG (fosfatidilglicerol) (Tabela 3) localizado na camada externa da bicamada lipídica. A maioria dessas interações são ligações de hidrogênio entre o peptídeo e o lipídio. Da mesma forma, as interações do Pep 10 inseridas na membrana a 200 ns de simulação (Figura 10), também são formadas principalmente por ligações de hidrogênio e pontes salinas (Tabela 4). Os átomos NZ das cadeias laterais das lisinas (Lys) estão formando pontes salinas com os grupos fosfato do POPG. As demais interações observadas são pontes de hidrogênio. As Tabelas 3 e 4 representam os resíduos que interagem com os lipídios da membrana, em cada caso especificamente. Para cada resíduo está representada sua posição no peptídeo, o átomo que está interagindo e a distância de interação, bem como o respectivo lipídeo e átomo específico com o qual interagem.



Tabela 3. Interações entre Pep 10 e a membrana lipídica, em 200ns de simulação. O peptídeo estava localizado, no início da simulação, a 35 Å de distância da membrana. Os átomos de HT1, HT2 e HT3 referem-se aos hidrogênios das extremidades da cadeia lateral do resíduo de glicina (Gly). O HN é o átomo de hidrogênio da amina nos resíduos de lisina (Lys) e valina (Val). O HZ1 é um hidrogênio específico da lisina (Lys). Os termos POPE e POPG representam tipos de lipídios na membrana: Fosfatidilcolina de etanolamina e Fosfatidilglicerol, respectivamente. Os códigos O12, O14, O11, O32, O13, O22 e O21 correspondem aos átomos de oxigênio nos lipídios da membrana. Os números representam sua posição específica dentro da estrutura do composto.

Resíduo	Posição	Átomo	Distância (Å)	Resíduo	Posição	Átomo
Pep 10				Membrana		
Gly	1	HT1	2,4	POPE	67	O12
Gly	1	HT1	1,6	POPE	67	O14
Gly	1	HT2	1,7	POPG	147	O11
Gly	1	HT3	2,0	POPE	41	O32
Lys	2	HN	2,1	POPE	41	O14
Val	3	HN	2,1	POPG	142	O13
Val	4	HN	1,7	POPG	142	O14
Lys	11	NZH1	2,6	POPG	153	O22
Lys	12	NZH1	1,6	POPG	152	O11
Lys	12	NZH1	2,3	POPG	152	O21

Tabela 4. Interações entre Pep 10 e a membrana lipídica, em 200ns de simulação. O peptídeo estava localizado, no início da simulação, inserido na membrana. Os átomos de HT1, HT2 e HT3 referem-se aos hidrogênios das extremidades da cadeia lateral do resíduo de glicina (Gly). O HZ1, HZ2 e HZ3 são hidrogênios da lisina (Lys). O OT1 é um oxigênio da cisteína (Cys). Os termos POPE e POPG representam tipos de lipídios na membrana: Fosfatidilcolina de etanolamina e Fosfatidilglicerol, respectivamente. Os códigos O14, O13, O22, O32, O11 e HN1 correspondem aos átomos de oxigênio e hidrogênio nos lipídios da membrana. Os números representam sua posição específica dentro da estrutura do composto.

Resíduo	Posição	Átomo	Distância (Å)	Resíduo	Posição	Átomo
Pep 10				Membrana		
Gly	1	HT1	1,8	POPE	32	O14
Gly	1	HT2	1,8	POPG	150	O14
Gly	1	HT3	1,6	POPE	18	O13
Lys	2	NZH2	1,8	POPG	140	O13
Lys	5	NZH1	1,6	POPE	32	O13
Lys	5	NZH2	2,8	POPE	18	O14
Lys	5	NZH3	2,3	POPE	18	O22
Lys	11	NZH2	2,3	POPE	74	O22
Lys	11	NZH3	2,0	POPE	74	O32
Lys	12	NZH1	2,0	POPE	133	O11
Cys	14	OT1	2,0	POPE	103	HN1
Cys	14	OT1	1,7	POPE	103	HN1

## 6. Discussão

Os métodos de aprendizado de máquina geralmente têm bom desempenho na previsão de AMPs, mas a falta de dados detalhados de sequência de AMP disponíveis para treinamento pode limitar seu desempenho (Li et al., 2022). O tamanho dos dados de treinamento ainda é pequeno em relação aos dados normalmente usados na maioria das aplicações de aprendizagem profunda. Este poderá ter sido um fator decisivo nos resultados obtidos. O tamanho dos dados utilizados e o tamanho limite escolhido (16 aminoácidos) impediram a implementação de mais restrições durante a geração de novas sequências utilizando MRL. As restrições corretas e seus respectivos filtros são essenciais para a otimização e geração de novas variantes peptídicas. Portanto, para solucionar isso, um banco de dados maior permitiria restrições mais extensas, como: aumentar o número de aminoácidos que favorecem a atividade antimicrobiana (arginina e lisina), diminuir a presença de aminoácidos que não favorecem a atividade antimicrobiana (ácidos glutâmico e aspártico). Aumentar o tamanho máximo para 30 aminoácidos também poderia ter um impacto positivo na geração de novas sequências, apesar do aumento no custo de produção desses peptídeos.

Vários estudos demonstraram que a função dos AMPs está altamente correlacionada com a estabilidade das estruturas secundárias, como componentes da alfa-hélice ou da folha beta (Cao et al., 2023); portanto, usamos a disponibilidade da estrutura secundária como uma base importante para avaliar os peptídeos gerados. Na Figura 2 podemos ver como os peptídeos gerados apresentam uma conformação de alfa-hélice na maior parte de sua estrutura secundária. Os Pep 7 e 12 são os únicos que não atendem a esta afirmação, ambos os peptídeos apresentam principalmente uma conformação estendida e uma pequena região alfa helicoidal.

A presença de cargas positivas é fundamental para a função dos AMPs, porque as membranas celulares bacterianas exibem uma carga aniônica líquida, de modo que os peptídeos podem atingir as membranas celulares bacterianas mediante interações eletrostáticas (Yan et al., 2021). Embora todos os peptídeos gerados apresentassem cargas positivas, sua baixa atividade, o que pode sugerir que é necessário aumentar sua carga para melhorar sua atividade. Além das suas cargas positivas, os AMPs dominados

por grupos hidrofóbicos também podem exercer melhores efeitos antibacterianos. O triptofano é um aminoácido altamente hidrofóbico. Os resíduos de triptofano (W) têm a capacidade única de interagir com a região de interface de uma membrana, permitindo que o peptídeo se ancore na superfície da bicamada (Bi et al., 2014). Portanto, os peptídeos ricos em triptofano são candidatos atraentes para o desenvolvimento de novas gerações de agentes antibacterianos. Vários estudos demonstraram que o motivo WWW é essencial para matar *S. aureus* resistente à meticilina e destruir biofilmes bacterianos pré-formados (Yan et al., 2021). Levando em consideração o dito acima, ao analisar os peptídeos gerados (Figura 2) podemos observar falta de triptofano nas sequências. A ausência de triptofano pode ser um fator importante no fato destes peptídeos apresentarem em sua maioria baixos valores de hidrofobicidade e momento hidrofóbico, contribuindo para sua baixa atividade.

Ao comparar os valores de predição de diferentes modelos de predição de atividade antimicrobiana para os 15 peptídeos (Figura 3), podemos observar todas as probabilidades apresentadas de atividade antimicrobiana maiores que 0,5 em pelo menos 3 modelos de predição diferentes. Apenas Pep 5 apresentou valores de predição inferiores a 0,5 para dois modelos, AMPScanner v2 (Veltri et al., 2018) e Macrel (Santos-Júnior et al., 2020). Dos 15 peptídeos, 8 peptídeos apresentaram probabilidades de atividade antimicrobiana superiores a 0,5 nos 5 modelos avaliados. Tendo em conta estes resultados, a utilização de modelos de predição de atividade antimicrobiana, embora úteis, não são suficientes para selecionar possíveis AMPs. Os modelos de predição devem ser combinados com outros critérios de seleção (parâmetros físico-químicos) para escolher possíveis AMPs. Embora vários candidatos identificados usando este pipeline não tenham mostrado nenhuma atividade *in vitro* contra as bactérias testadas, especulamos, com base nesses resultados, que eles ainda podem possuir atividade contra outras espécies bacterianas ou outros microrganismos (por exemplo, fungos, vírus).

AMPsphere é um catálogo de AMPs composto por 863.498 peptídeos não redundantes. Para sua criação, o autor propõe quatro importantes características físico-químicas que os AMPs devem apresentar. Essas características são: ter carga positiva

(+2 - +7), alto ponto isoelétrico (9,7 - 11,1), momento hidrofóbico (0,5 - 0,7) e potencial para se ligar a membranas ou outras proteínas (índice de Boman, 0,04 - 2,24). Para verificar se os quinze peptídeos projetados atendem a esses parâmetros, comparamos esses valores com os calculados na Tabela 1. Os valores de carga e ponto isoelétrico dos 15 peptídeos estão dentro das faixas destas características físico-químicas propostas por Santos-Júnior et al. (2024), enquanto para o índice de Boman, 13 dos 15 peptídeos apresentam valores nesta faixa, apenas Pep1 e Pep 2 apresentam valores inferiores (-0,35 e -0,61, respectivamente) ao mínimo do AMPSphere para esta característica. Por outro lado, a maior diferença entre os peptídeos desenhados e o AMPSphere encontra-se no momento hidrofóbico. Apenas dois dos peptídeos projetados (Pep 6 e 10) excede o valor mínimo do momento hidrofóbico (0,5) do AMPSphere, este pode ser o ponto crítico para explicar a baixa atividade antimicrobiana dos peptídeos projetados (Santos-Júnior et al., 2024).

Para validar o AMPSphere, foi sintetizado e testado 100 AMPs contra patógenos resistentes a medicamentos clinicamente relevantes. Um total de 79 peptídeos demonstraram atividade. As bactérias Gram-negativas *A. baumannii* e *E. coli* apresentaram maior suscetibilidade aos AMPs, com 39 e 24 acertos, respectivamente. Pelo contrário, as bactérias Gram-negativas *K. pneumoniae* e as Gram-positivas *S. aureus* apresentaram menor suscetibilidade aos AMPs, com 3 e 4 acertos, respectivamente (Santos-Júnior et al., 2024).

O pequeno AMP sintético com estrutura alfa-helicoidal (RR) (Figura 12) demonstrou a capacidade de inibir o crescimento, em concentrações de 16-64  $\mu\text{M}$ , de cepas sensíveis e resistentes de *A. baumannii*. Durante a criação de análogos de RR, a substituição dos resíduos hidrofóbicos (leucina) em RR por resíduos catiônicos (lisina) (para gerar RR1) (Figura 12) aumentou a carga catiônica de +5 (RR) para +7 (RR1) e diminuiu a hidrofobicidade de 0,45 para 0,24. Como resultado, descobriu-se que RR1 era inativo contra todas as cepas bacterianas testadas ( $\text{MIC} > 256 \mu\text{M}$ ). Pelo contrário, o análogo com maior conteúdo helicoidal que manteve a anfipaticidade (RR4) (Figura 12) teve a atividade antibacteriana mais potente (4  $\mu\text{M}$ ) contra cepas sensíveis e resistentes de *A. baumannii*. Levando em consideração esses resultados, o autor propõe que a

hidrofobicidade desempenhe um papel mais crítico na atividade antibacteriana dos AMPs helicoidais do que o aumento no número de cargas catiônicas (Mohamed et al., 2017).

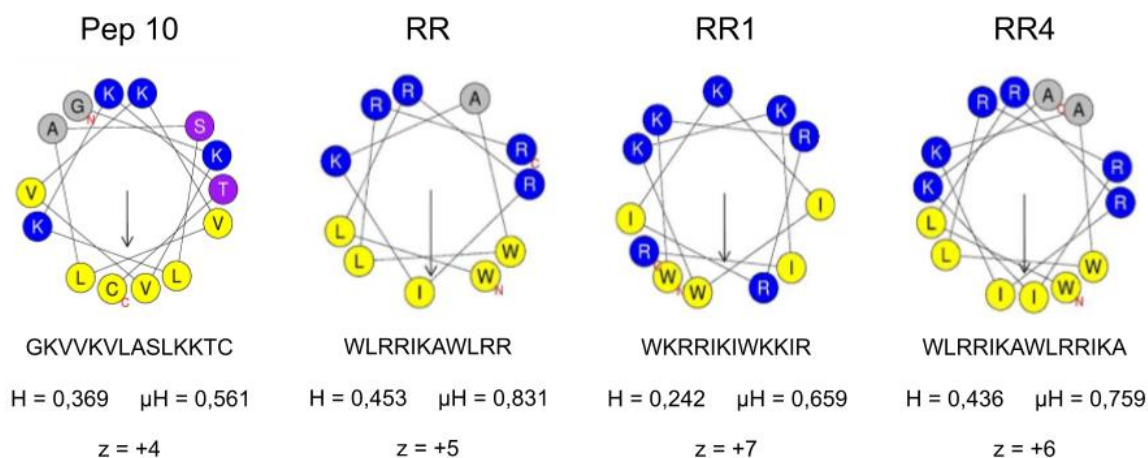


Figura 12: Projeções de roda helicoidal do peptídeo projetado Pep 10 e os peptídeos RR e RR4 (Mohamed et al., 2017). Os círculos azuis representam os resíduos carregados positivamente arginina (R) e lisina (K) e os roxos representam resíduos polares não carregados serina (S) e treonina (T). Os círculos cinza representam resíduos de alanina (A) e glicina (G) e os amarelos representam resíduos hidrofóbicos de valina (V), isoleucina (I), leucina (L), triptofano (W) e o resíduo de cisteína (C). As setas indicam as direções e magnitudes do momento hidrofóbico; N, terminal amino; C, extremidade carboxila; H, hidrofobicidade;  $\mu H$ , momento hidrofóbico.

Ao comparar o Pep10 com o peptídeo RR e seus análogos (Figura 12), podemos ver como os valores de hidrofobicidade são ligeiramente inferiores aos do RR e do seu análogo RR4, mas o seu momento hidrofóbico é muito inferior ao de ambos. Além disso, na representação helicoidal podemos ver como o número de aminoácidos positivos (círculos azuis) e, portanto, a carga líquida é maior no RR e seus análogos. Estas diferenças entre Pep 10 e RR podem ser uma possível causa entre os valores mínimos de MIC contra cepas resistentes de *A. baumannii*. O RR apresenta atividade em concentrações de até 16  $\mu M$ , enquanto Pep 10 a 32  $\mu M$ . Embora também deva ser levado em consideração que RR e seus análogos foram testados contra três cepas suscetíveis e 13 resistentes de *A. baumannii*; enquanto o Pep 10 foi testado contra uma cepa suscetível e duas cepas resistentes.

Por outro lado, RR1, o análogo de RR, também possui momento hidrofóbico e carga líquida maiores que Pep 10, mas sua hidrofobicidade é muito menor. Mohamed et

al. (2017) propuseram que sua baixa hidrofobicidade era a principal causa, portanto, esse peptídeo era inativo contra as cepas bacterianas testadas. Ao comparar o valor de hidrofobicidade do RR1 com os valores dos peptídeos desenhados que não apresentaram atividade antibacteriana (Tabela 1), podemos observar que a maioria apresenta valores semelhantes. Levando isso em consideração e o fato de apenas Pep 6 e 10 apresentarem valores de momento hidrofóbico maiores que 0,5, podemos inferir que essas são as razões prováveis pelas quais apenas Pep 10 apresentou atividade contra as cepas *A. baumannii*.

Em um estudo recente redes adversárias geradoras de sequência (SeqGAN) são utilizadas para gerar peptídeos e, em seguida, usam um preditor AMP baseado em representações de codificadores bidirecionais de transformadores (BERT) e perceptrons multicamadas (MLP) para identificar peptídeos gerados por SeqGAN. Dos peptídeos gerados, o A-222 foi o que apresentou melhores resultados, apresentando MIC de 32  $\mu$ M contra a bactéria Gram-negativa *P. aeruginosa*. Posteriormente, eles projetaram análogos do A-222 usando metodologia de varredura de alanina e aumentando o número de cargas positivas. Os análogos gerados aumentaram sua atividade antimicrobiana, atingindo CIMs de 3  $\mu$ M contra as mesmas cepas de *P. aeruginosa*. Uma metodologia semelhante pode ser utilizada para melhorar a atividade do Pep10 e obter análogos com melhor atividade antimicrobiana (Cao et al., 2023).

Levando em consideração o que foi afirmado pelos estudos mencionados acima, é necessário criar um filtro de hidrofobicidade no modelo MRL. Para isso, o banco de dados utilizado, GRAMPA, deverá ser curado e conferido. Além disso, as novas sequências incluídas nas diferentes bases de dados devem ser adicionadas desde a criação do GRAMPA. As características físico-químicas de cada sequência do banco de dados devem ser calculadas para fornecer dados mais precisos ao modelo.

Ao analisar o número de interações de hidrogênio entre o Pep 10 e a bicamada lipídica (Figura 10), podemos ver como o Pep 10 mantém um elevado número de interações ao longo do tempo de simulação de 200 ns. Além disso, ao analisar mais de perto essas interações (Figura 11), pode-se observar que quase todas são fortes ligações de hidrogênio. As conexões de hidrogênio são geralmente mais fortes quando

a distância entre o dispensador de hidrogênio e o lubrificador de hidrogênio está na faixa de 1,5 a 2,5 Å, correspondendo ao mostrado na Tabela 3. As interações entre os oxigênios da glicina (Gly 1) e POPE (41) (Tabela 3), poderiam ser uma interação de Van der Waals muito fraca, mas a esta distância (3,3 Å), uma interação significativa não seria esperada. Da mesma forma, as interações entre os oxigênios da glicina (Gly 1) e o lipídeo, neste caso, POPE (Tabela 4); com uma distância de interação de 3,3 Å também não apresentaria interação significativa.

Ambas as análises (RMSD e RMSF) da simulação Pep 10, na parte externa da membrana, amostra ou peptídeo apresentam alguma instabilidade em sua estrutura, principalmente nas extremidades. Pep 10 perde a parte do sua conformação  $\alpha$ -helicoidal inicial, prevista pelo AlphaFold 3, adquirindo uma conformação estendida no final de sua estrutura (Figura 7). Por outro lado, quando o Pep 10 é inserido na membrana, a estrutura inicial permanece estável durante toda a simulação. Embora sejam necessárias escalas de tempo mais longas para a reorientação de peptídeos em sistemas de bicamada usando simulações MD, nossos dados computacionais apoiam nossas descobertas experimentais e fornecem uma visão inicial sobre como o Pep 10 exibe preferência de ligação para ambientes semelhantes a bactérias.



## 7. Considerações Finais

Este estudo relata a avaliação de um novo modelo de desenho de AMPs utilizando aprendizagem por reforço profundo, visando a possível concepção de novos medicamentos com potencial terapêutico, para o tratamento de doenças causadas por bactérias Gram-negativas de importância clínica. O estudo revelou resultados promissores e desafiadores.

Das 184 novas sequências selecionadas, 15 atenderam a todos os critérios de seleção propostos neste estudo. Destes novos peptídeos, Pep10 apresentou atividade antibacteriana em concentrações de 64  $\mu$ M e 32  $\mu$ M contra cepas de *A. baumannii* suscetíveis e resistentes, respectivamente. Além disso, Pep 10 não apresentou efeito hemolítico em eritrócitos de camundongos. Isto nos leva a destacar o potencial deste peptídeo para o desenvolvimento de novos fármacos para o controle e/ou tratamento de doenças causadas por *A. baumannii*. Isto pode ser conseguido através da criação de análogos ao Pep 10 com diferentes características físico-químicas que melhoram a sua atividade.

A dinâmica molecular do Pep 10 dentro e fora da membrana interna das bactérias Gram-negativas mostra que ele mantém diversas interações fortes com os lipídios da membrana. Portanto, a baixa atividade contra as outras cepas de bactérias Gram-negativas testadas pode ser devida a dificuldades de interação com a membrana externa dessas bactérias.

Foi realizada uma análise mais aprofundada em dois peptídeos que não apresentaram atividade antimicrobiana contra as cepas bacterianas testadas, para descobrir as razões desse resultado. A atividade antimicrobiana foi prevista utilizando os mais avançados modelos de predição disponíveis. Além disso, foram calculados diferentes parâmetros físicos e químicos. Estas análises mostraram que, a maioria destes peptídeos apresentassem bons valores de predição para os diferentes modelos testados, mais seus valores de hidrofobicidade e momento hidrofóbico estão abaixo de outros peptídeos com atividade comprovada para bactérias Gram-negativas. Portanto, o uso dessas propriedades é imperativo para projetos futuros de novos possíveis AMPs.

## 8. Perspectivas

O novo peptídeo Pep 10 demonstrou atividade antibacteriana contra isolados de *A. baumannii*, e não apresentou atividade hemolítica em eritrócitos de camundongos, demonstrando potencial biotecnológico. Desta forma, é importante verificar o efeito bactericida frente aos isolados de *A. baumannii*, e testar se este peptídeo não apresenta atividade citotóxica. Além disso, o Pep 10 pode ser testado em experimentos de mecanismo de ação, como a microscopia de força atômica. Novas dinâmicas do Pep 10 também poderiam ser executadas em ambientes com membranas externas de diferentes bactérias Gram-negativas para melhor compreender seu mecanismo de ação. Isto é para confirmar que o Pep 10 é um possível candidato para o desenvolvimento de novos medicamentos.

O Pep 10 demonstre boa atividade antibacteriana contra isolados de *A. baumannii*, mais os valores apresentados podem ser melhorados. Utilizando um design racional, com o auxílio de ferramentas de AI, é possível criar novos AMPs análogos ao Pep 10, que apresentam atividade antibacteriana em concentrações mais baixas. Para aumentar a carga do peptídeo, podemos substituir os resíduos de serina e/ou treonina por resíduos de lisina ou arginina, mantendo o tamanho do peptídeo. Adicionalmente, para aumentar a sua hidrofobicidade, podem ser adicionados resíduos de triptofano à sua estrutura. É importante que ao adicionar e/ou substituir alguns resíduos em sua estrutura, numa representação helicoidal, os resíduos positivos fiquem em cima e os resíduos hidrofóbicos em baixo, mantendo certa simetria na estrutura. Se isto não for cumprido, pode afetar o momento hidrofóbico do peptídeo e, portanto, reduzir a sua atividade antibacteriana.

Também é possível melhorar o Pep 10 utilizando XAI (*Explainable Artificial Intelligence*). Técnicas de XAI, como SHAP e LIME, são aplicadas para interpretar os os modelos de previsão utilizados, revelando quais características dos peptídeos são mais importantes para a eficácia. Com esses dados, ajustes podem ser feitos no Pep 10, como substituir resíduos que influenciam negativamente sua atividade antimicrobiana por lisina ou arginina para aumentar a carga positiva e adicionar resíduos hidrofóbicos como triptofano para melhorar a hidrofobicidade. É essencial manter a estrutura helicoidal do

peptídeo, com resíduos positivos externamente e hidrofóbicos internamente. Após essas modificações, os novos peptídeos devem ser sintetizados e testados experimentalmente para validar as melhorias. O ciclo de refinamento é contínuo, utilizando XAI para interpretar novos dados e otimizar o peptídeo continuamente.

Adicionalmente, um novo desenho poderia ser feito utilizando a metodologia proposta neste estudo. Este novo desenho deverá incorporar as propriedades físicas-químicas (hidrofobicidade e momento hidrofóbico), nas etapas de espaço conformacional e seleção das sequências geradas. Aumentar a quantidade do aminoácido triptofano pode ser uma boa maneira de aumentar a hidrofobicidade de novos peptídeos. Também seria prudente modificar alguns dos parâmetros utilizados neste estudo, para o desenho das novas sequências, tais como: uma carga líquida (limitar o valor mínimo a +4) e aumentar a sequência máxima de até 30 aa. Para implementar com sucesso estas mudanças no modelo, é imperativa a criação de uma base de dados nova e melhor. Um novo banco de dados, com maior número de sequências com atividade comprovada contra bactérias Gram-negativas, é necessário para poder modificar o espaço conformacional no modelo MRL. O novo espaço conformacional, com as restrições levantadas acima, permitiria a geração de novos peptídeos com melhor atividade antibacteriana.

## 9. Referências

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024 630:8016, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>

Alba, A., López-Abarrategui, C., & Otero-González, A. J. (2012). Host defense peptides: an alternative as antiinfective and immunomodulatory therapeutics. *Biopolymers*, 98(4), 251–267. <https://doi.org/10.1002/BIP.22076>

Algammal, A., Hetta, H. F., Mabrok, M., & Behzadi, P. (2023). Editorial: Emerging multidrug-resistant bacterial pathogens “superbugs”: A rising public health threat. *Frontiers in Microbiology*, 14, 1135614. <https://doi.org/10.3389/FMICB.2023.1135614/BIBTEX>

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, 28(1), 45–48. <https://doi.org/10.1093/NAR/28.1.45>

Bi, X., Wang, C., Dong, W., Zhu, W., & Shang, D. (2014). Antimicrobial properties and interaction of two Trp-substituted cationic antimicrobial peptides with a lipid bilayer. *The Journal of Antibiotics* 2014 67:5, 67(5), 361–368. <https://doi.org/10.1038/ja.2014.4>

Boman, H. G. (2003). Antibacterial peptides: basic facts and emerging concepts. *Journal of Internal Medicine*, 254(3), 197–215. <https://doi.org/10.1046/J.1365-2796.2003.01228.X>

Brandenburg, L. O., Merres, J., Albrecht, L. J., Varoga, D., & Pufe, T. (2012). Antimicrobial Peptides: Multifunctional Drugs for Different Applications. *Polymers*, 4(1), 539–560. <https://doi.org/10.3390/POLYM4010539>

Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews. Microbiology*, 3(3), 238–250. <https://doi.org/10.1038/NRMICRO1098>

Caneiras, C., Lito, L., Melo-Cristino, J., & Duarte, A. (2019). Community- and Hospital-Acquired *Klebsiella pneumoniae* Urinary Tract Infections in Portugal: Virulence and Antibiotic Resistance. *Microorganisms*, 7(5). <https://doi.org/10.3390/MICROORGANISMS7050138>

Cao, L., Coventry, B., Goreshnik, I., Huang, B., Sheffler, W., Park, J. S., Jude, K. M., Marković, I., Kadam, R. U., Verschueren, K. H. G., Verstraete, K., Walsh, S. T. R., Bennett, N., Phal, A., Yang, A., Kozodoy, L., DeWitt, M., Picton, L., Miller, L., ... Baker, D. (2022). Design of protein-binding proteins from the target structure alone. *Nature* 2022 605:7910, 605(7910), 551–560. <https://doi.org/10.1038/s41586-022-04654-9>

Cao, Q., Ge, C., Wang, X., Harvey, P. J., Zhang, Z., Ma, Y., Wang, X., Jia, X., Mobli, M., Craik, D. J., Jiang, T., Yang, J., Wei, Z., Wang, Y., Chang, S., & Yu, R. (2023). Designing antimicrobial peptides using deep learning and molecular dynamic simulations. *Briefings in Bioinformatics*, 24(2), 1–13. <https://doi.org/10.1093/BIB/BBAD058>

Cardoso, M. H., de la Fuente-Nunez, C., Santos, N. C., Zasloff, M. A., & Franco, O. L. (2024). Influence of antimicrobial peptides on the bacterial membrane curvature and vice versa. *Trends in Microbiology*, 32(7), 624–627. <https://doi.org/10.1016/J.TIM.2024.04.012>

Castañeda-Casimiro, J., Ortega-Roque, J. A., Venegas-Medina, A. M., Aquino-Andrade, A., Serafín-López, J., Estrada-Parra, S., & Estrada, I. (2009). péptidos antimicrobianos: péptidos con múltiples funciones . *Alergias, Asmas e Inmunologías Pediátricas*, 18(1), 16–29. [www.medigraphic.com](http://www.medigraphic.com)

Cesaro, A., Bagheri, M., Torres, M., Wan, F., & de la Fuente-Nunez, C. (2023). Deep learning tools to accelerate antibiotic discovery. *Expert Opinion on Drug Discovery*, 18(11), 1245–1257. <https://doi.org/10.1080/17460441.2023.2250721>

Chen, N., & Jiang, C. (2023). Antimicrobial peptides: Structure, mechanism, and modification. *European Journal of Medicinal Chemistry*, 255, 115377. <https://doi.org/10.1016/J.EJMECH.2023.115377>

Ciulla, M. G., & Gelain, F. (2023). Structure–activity relationships of antibacterial peptides. *Microbial Biotechnology*, 16(4), 757–777. <https://doi.org/10.1111/1751-7915.14213>

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1, 4171–4186. <https://arxiv.org/abs/1810.04805v2>

Du, Z., Ding, X., Hsu, W., Munir, A., Xu, Y., & Li, Y. (2024). pLM4ACE: A protein language model based predictor for antihypertensive peptide screening. *Food Chemistry*, 431, 137162. <https://doi.org/10.1016/J.FOODCHEM.2023.137162>

Fingerhut, L. C. H. W., Miller, D. J., Strugnell, J. M., Daly, N. L., & Cooke, I. R. (2021). ampir: an R package for fast genome-wide prediction of antimicrobial peptides. *Bioinformatics* (Oxford, England), 36(21), 5262–5263. <https://doi.org/10.1093/BIOINFORMATICS/BTAA653>

Fjell, C. D., Hiss, J. A., Hancock, R. E. W., & Schneider, G. (2011). Designing antimicrobial peptides: form follows function. *Nature Reviews Drug Discovery* 2012 11:1, 11(1), 37–51. <https://doi.org/10.1038/nrd3591>

Fleitas, O., Fontes, W., de Souza, C. M., da Costa, M. C., Cardoso, M. H., Castro, M. S., Sousa, M. v., Ricart, C. A. O., Ramada, M. H. S., Duque, H. M., Porto, W. F., Silva, O. N., & Franco, O. L. (2024). A proteomic perspective on the resistance response of *Klebsiella pneumoniae* to antimicrobial peptide PaDBS1R1. *The Journal of Antimicrobial Chemotherapy*, 79(1), 112–122. <https://doi.org/10.1093/JAC/DKAD354>

Galdiero, S., Falanga, A., Berisio, R., Grieco, P., Morelli, G., & Galdiero, M. (2015). Antimicrobial peptides as an opportunity against bacterial diseases. *Current Medicinal Chemistry*, 22(14), 1665–1677. <https://doi.org/10.2174/0929867322666150311145632>

Gautier, R., Douguet, D., Antonny, B., & Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* (Oxford, England), 24(18), 2101–2102. <https://doi.org/10.1093/BIOINFORMATICS/BTN392>

Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2021). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology* 2021 23:1, 23(1), 40–55. <https://doi.org/10.1038/s41580-021-00407-0>

Guntuboina, C., Das, A., Mollaei, P., Kim, S., & Barati Farimani, A. (2023). PeptideBERT: A Language Model Based on Transformers for Peptide Property Prediction. *Journal of Physical Chemistry Letters*, 14(46), 10427–10434. [https://doi.org/10.1021/ACS.JPCLETT.3C02398/ASSET/IMAGES/LARGE/JZ3C02398\\_0003.JPEG](https://doi.org/10.1021/ACS.JPCLETT.3C02398/ASSET/IMAGES/LARGE/JZ3C02398_0003.JPEG)

Gutiérrez, P., & Orduz, S. (2017). Péptidos antimicrobianos: Estructura, función y aplicaciones. *Actualidades Biológicas*, 25(78), 1–12. <https://doi.org/10.17533/UDEA.ACBI.329497>

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/NECO.1997.9.8.1735>

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2016). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 2016 14:1, 14(1), 71–73. <https://doi.org/10.1038/nmeth.4067>

Huang, J., Xu, Y., Xue, Y., Huang, Y., Li, X., Chen, X., Xu, Y., Zhang, D., Zhang, P., Zhao, J., & Ji, J. (2023). Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nature Biomedical Engineering* 2023 7:6, 7(6), 797–810. <https://doi.org/10.1038/s41551-022-00991-2>

Huerta-Cantillo, J., & Navarro-García, F. (2016). Properties and design of antimicrobial peptides as potential tools against pathogens and malignant cells. *Investigación En Discapacidad*, 5(2), 96–115. [www.medigraphic.org.mx](http://www.medigraphic.org.mx)

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/S12525-021-00475-2/TABLES/2>

Jenssen, H., Hamill, P., & Hancock, R. E. W. (2006). Peptide antimicrobial agents. *Clinical Microbiology Reviews*, 19(3), 491–511. <https://doi.org/10.1128/CMR.00056-05>

Jesus Otero-González, A., Simas Magalhães, B., Garcia-Villarino, M., López-Abarategui, C., Amaro Sousa, D., Campos Dias, S., & Luiz Franco, O. (2010). Antimicrobial peptides from marine invertebrates as a new frontier for microbial infection control. *The FASEB Journal*, 24(5), 1320–1334. <https://doi.org/10.1096/FJ.09-143388>

Ji, S., An, F., Zhang, T., Lou, M., Guo, J., Liu, K., Zhu, Y., Wu, J., & Wu, R. (2024). Antimicrobial peptides: An alternative to traditional antibiotics. *European Journal of Medicinal Chemistry*, 265, 116072. <https://doi.org/10.1016/J.EJMECH.2023.116072>

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence? *Discover Artificial Intelligence* 2022 2:1, 2(1), 1–19. <https://doi.org/10.1007/S44163-022-00022-8>

Jin, S., Zeng, X., Xia, F., Huang, W., & Liu, X. (2021). Application of deep learning methods in biological networks. *Briefings in Bioinformatics*, 22(2), 1902–1917. <https://doi.org/10.1093/BIB/BBAA043>

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11), 1859–1865. <https://doi.org/10.1002/JCC.20945>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ...



Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/S41586-021-03819-2>

Kaynak, O. (2021). The golden age of Artificial Intelligence: Inaugural Editorial. *Discover Artificial Intelligence*, 1(1), 1–7. <https://doi.org/10.1007/S44163-021-00009-X/FIGURES/3>

Kumar, N., Bhagwat, P., Singh, S., & Pillai, S. (2024). A review on the diversity of antimicrobial peptides and genome mining strategies for their prediction. *Biochimie*. <https://doi.org/10.1016/J.BIOCHI.2024.06.013>

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). <https://doi.org/10.3390/BIOM8010004>

Ladosz, P., Weng, L., Kim, M., & Oh, H. (2022). Exploration in deep reinforcement learning: A survey. *Information Fusion*, 85, 1–22. <https://doi.org/10.1016/J.INFFUS.2022.03.003>

Lavecchia, A. (2019). Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discovery Today*, 24(10), 2017–2032. <https://doi.org/10.1016/J.DRUDIS.2019.07.006>

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, C., Sutherland, D., Hammond, S. A., Yang, C., Taho, F., Bergman, L., Houston, S., Warren, R. L., Wong, T., Hoang, L. M. N., Cameron, C. E., Helbing, C. C., & Birol, I. (2022). AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens. *BMC Genomics*, 23(1), 1–15. <https://doi.org/10.1186/S12864-022-08310-4/TABLES/3>

Li, S. E. (2023). Reinforcement Learning for Sequential Decision and Optimal Control. *Reinforcement Learning for Sequential Decision and Optimal Control*, 1–462. <https://doi.org/10.1007/978-981-19-7784-8/COVER>

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, 22(13), 1658–1659. <https://doi.org/10.1093/BIOINFORMATICS/BTL158>

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). <https://doi.org/10.1128/MSYSTEMS.00299-21/ASSET/C5D82C8C-CFDD-4C3D-A925-18334516FDEA/ASSETS/IMAGES/MEDIUM/MSYSTEMS.00299-21-F006.GIF>

Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3), 773–780. <https://doi.org/10.1016/J.DRUDIS.2018.11.014>

Marr, A. K., Gooderham, W. J., & Hancock, R. E. (2006). Antibacterial peptides for therapeutic use: obstacles and realistic outlook. *Current Opinion in Pharmacology*, 6(5), 468–472. <https://doi.org/10.1016/J.COPH.2006.04.006>

Min, B., Ross, H., Sulem, E., Veyseh, A. P. ben, Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., & Roth, D. (2023). Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey. *ACM Computing Surveys*, 56(2). <https://doi.org/10.1145/3605943>

Mirnejad, R., Fasihi-Ramandi, M., Behmard, E., Najafi, A., & Moosazadeh Moghaddam, M. (2023). Interaction of antibacterial CM11 peptide with the gram-positive and gram-negative bacterial membrane models: a molecular dynamics simulations study. *Chemical Papers*, 77(7), 3727–3735. <https://doi.org/10.1007/S11696-023-02735-1/FIGURES/4>

Mohamed, M. F., Brezden, A., Mohammad, H., Chmielewski, J., & Seleem, M. N. (2017). A short D-enantiomeric antimicrobial peptide with potent immunomodulatory and antibiofilm activity against multidrug-resistant *Pseudomonas aeruginosa* and

*Acinetobacter baumannii*. Scientific Reports, 7(1). <https://doi.org/10.1038/S41598-017-07440-0>

Molecular Reinforcement Learning | mrl. (n.d.). Retrieved August 26, 2024, from <https://darkmatterai.github.io/mrl/>

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. <https://doi.org/10.1038/S41573-019-0058-8>

Morris, S., & Cerceo, E. (2020). Trends, Epidemiology, and Management of Multi-Drug Resistant Gram-Negative Bacterial Infections in the Hospitalized Setting. *Antibiotics* 2020, Vol. 9, Page 196, 9(4), 196. <https://doi.org/10.3390/ANTIBIOTICS9040196>

Moya, C., & Maicas, S. (2020). Antimicrobial Resistance in *Klebsiella pneumoniae* Strains: Mechanisms and Outbreaks. *Proceedings* 2020, Vol. 66, Page 11, 66(1), 11. <https://doi.org/10.3390/PROCEEDINGS2020066011>

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)

Nedyalkova, M., Paluch, A. S., Vecini, D. P., & Lattuada, M. (2024). Progress and future of the computational design of antimicrobial peptides (AMPs): bio-inspired functional molecules. *Digital Discovery*, 3(1), 9–22. <https://doi.org/10.1039/D3DD00186E>

Nguyen, L. T., Haney, E. F., & Vogel, H. J. (2011). The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in Biotechnology*, 29(9), 464–472. <https://doi.org/10.1016/J.TIBTECH.2011.05.001>

Ofer, D., Brandes, N., & Linial, M. (2021). The language of proteins: NLP, machine learning & protein sequences. *Computational and Structural Biotechnology Journal*, 19, 1750–1758. <https://doi.org/10.1016/J.CSBJ.2021.03.022>

Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1). <https://doi.org/10.1186/S13321-017-0235-X>

Páll, S., Zhmurov, A., Bauer, P., Abraham, M., Lundborg, M., Gray, A., Hess, B., & Lindahl, E. (2020). Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *Journal of Chemical Physics*, 153(13), 134110. <https://doi.org/10.1063/5.0018516/199476>

Peptide Analyzing Tool | Thermo Fisher Scientific - BR. (n.d.). Retrieved August 26, 2024, from <https://www.thermofisher.com/br/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html>

Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science Advances*, 4(7). [https://doi.org/10.1126/SCIADV.AAP7885/SUPPL\\_FILE/AAP7885\\_SM.PDF](https://doi.org/10.1126/SCIADV.AAP7885/SUPPL_FILE/AAP7885_SM.PDF)

Prestinaci, F., Pezzotti, P., & Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and Global Health*, 109(7), 309. <https://doi.org/10.1179/2047773215Y.0000000030>

Rathore, A. S., Choudhury, S., Arora, A., Tijare, P., & Raghava, G. P. S. (2024). ToxinPred 3.0: An improved method for predicting the toxicity of peptides. *Computers in Biology and Medicine*, 179, 108926. <https://doi.org/10.1016/J.COMPBIOMED.2024.108926>

Reddy, K. V. R., Yedery, R. D., & Aranha, C. (2004). Antimicrobial peptides: premises and promises. *International Journal of Antimicrobial Agents*, 24(6), 536–547. <https://doi.org/10.1016/J.IJANTIMICAG.2004.09.005>

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 118(15). <https://doi.org/10.1073/PNAS.2016239118/-/DCSUPPLEMENTAL>

Santos-Júnior, C. D., Pan, S., Zhao, X. M., & Coelho, L. P. (2020). Macrel: Antimicrobial peptide screening in genomes and metagenomes. *PeerJ*, 8, e10555. <https://doi.org/10.7717/PEERJ.10555/SUPP-7>

Santos-Júnior, C. D., Torres, M. D. T., Duan, Y., Rodríguez del Río, Á., Schmidt, T. S. B., Chong, H., Fullam, A., Kuhn, M., Zhu, C., Houseman, A., Somborski, J., Vines, A., Zhao, X. M., Bork, P., Huerta-Cepas, J., de la Fuente-Nunez, C., & Coelho, L. P. (2024). Discovery of antimicrobial peptides in the global microbiome with machine learning. *Cell*, 187(14), 3761-3778.e16. <https://doi.org/10.1016/J.CELL.2024.05.013>

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *ArXiv*, arXiv:1707.06347. <https://doi.org/10.48550/ARXIV.1707.06347>

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Openai, O. K. (2017). Proximal Policy Optimization Algorithms. <https://arxiv.org/abs/1707.06347v2>

Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(07), 3897–3904. <https://ssrn.com/abstract=4458723>

Steinegger, M., & Söding, J. (2018). Clustering huge protein sequence sets in linear time. *Nature Communications* 2018 9:1, 9(1), 1–8. <https://doi.org/10.1038/s41467-018-04964-5>

Tietz, J. I., Schwalen, C. J., Patel, P. S., Maxson, T., Blair, P. M., Tai, H. C., Zakai, U. I., & Mitchell, D. A. (2017). A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature Chemical Biology* 2017 13:5, 13(5), 470–478. <https://doi.org/10.1038/nchembio.2319>

van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8), 5929–5955. <https://doi.org/10.1007/S10462-020-09838-1/TABLES/1>

Veltri, D., Kamath, U., & Shehu, A. (2018). Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, 34(16), 2740–2747. <https://doi.org/10.1093/BIOINFORMATICS/BTY179>

Wang, Y. bin, You, Z. H., Yang, S., Yi, H. C., Chen, Z. H., & Zheng, K. (2020). A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. *BMC Medical Informatics and Decision Making*, 20(Suppl 2). <https://doi.org/10.1186/S12911-020-1052-0>

Wang, C., Garlick, S., & Zloh, M. (2021). Deep Learning for Novel Antimicrobial Peptide Design. *Biomolecules* 2021, Vol. 11, Page 471, 11(3), 471. <https://doi.org/10.3390/BIOM11030471>

Wang, G., Li, X., & Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research*, 44(D1), D1087–D1093. <https://doi.org/10.1093/NAR/GKV1278>

Wang, L., Li, Y., Fu, X., Ye, X., Shi, J., Yen, G. G., & Zeng, X. (2024). HMAMP: Hypervolume-Driven Multi-Objective Antimicrobial Peptides Design. *CoRR*, abs/2405.00753. <https://doi.org/10.48550/ARXIV.2405.00753>

Wang, R., Wang, T., Zhuo, L., Wei, J., Fu, X., Zou, Q., & Yao, X. (2024). Diff-AMP: tailored designed antimicrobial peptide framework with all-in-one generation, identification, prediction and optimization. *Briefings in Bioinformatics*, 25(2). <https://doi.org/10.1093/BIB/BBAE078>

Wang, X., Wang, S., Liang, X., Zhao, D., Huang, J., Xu, X., Dai, B., & Miao, Q. (2024). Deep Reinforcement Learning: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 35(4), 5064–5078. <https://doi.org/10.1109/TNNLS.2022.3207346>

Wang, Y., He, H., Wen, C., & Tan, X. (2019). Truly Proximal Policy Optimization. ArXiv E-Prints, arXiv:1903.07940. <https://doi.org/10.48550/arXiv.1903.07940>

Win, T. S., Malik, A. A., Prachayasittikul, V., S Wikberg, J. E., Nantasenamat, C., & Shoombuatong, W. (2017). HemoPred: a web server for predicting the hemolytic activity of peptides. *Future Medicinal Chemistry*, 9(3), 275–291. <https://doi.org/10.4155/FMC-2016-0188>

Witten, J., & Witten, Z. (2019). Deep learning regression model for antimicrobial peptide design. *BioRxiv*, 692681. <https://doi.org/10.1101/692681>

World Health Organization. (2022). Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2022. <https://www.who.int/publications/book-orders>.

Yan, J., Cai, J., Zhang, B., Wang, Y., Wong, D. F., & Siu, S. W. I. (2022). Recent Progress in the Discovery and Design of Antimicrobial Peptides Using Traditional Machine Learning and Deep Learning. *Antibiotics* 2022, Vol. 11, Page 1451, 11(10), 1451. <https://doi.org/10.3390/ANTIBIOTICS11101451>

Yan, Y., Li, Y., Zhang, Z., Wang, X., Niu, Y., Zhang, S., Xu, W., & Ren, C. (2021). Advances of peptides for antibacterial applications. *Colloids and Surfaces B: Biointerfaces*, 202, 111682. <https://doi.org/10.1016/J.COLSURFB.2021.111682>

Zhang, Q. Y., Yan, Z. bin, Meng, Y. M., Hong, X. Y., Shao, G., Ma, J. J., Cheng, X. R., Liu, J., Kang, J., & Fu, C. Y. (2021). Antimicrobial peptides: mechanism of action, activity and clinical potential. *Military Medical Research*, 8(1). <https://doi.org/10.1186/S40779-021-00343-2>

Zhang, Y., Li, S., Xing, M., Yuan, Q., He, H., & Sun, S. (2023). Universal Approach to De Novo Drug Design for Target Proteins Using Deep Reinforcement Learning. *ACS Omega*, 8(6), 5464–5474. [https://doi.org/10.1021/ACSOMEGA.2C06653/SUPPL\\_FILE/AO2C06653\\_SI\\_001.PDF](https://doi.org/10.1021/ACSOMEGA.2C06653/SUPPL_FILE/AO2C06653_SI_001.PDF)

Zhou, Z., Kearnes, S., Li, L., Zare, R. N., & Riley, P. (2019). Optimization of Molecules via Deep Reinforcement Learning. *Scientific Reports*, 9(1). <https://doi.org/10.1038/S41598-019-47148-X>



## 10. Apêndice A - Banco de dados e modelos utilizados neste estudo.

O conjunto de dados usado para treinar os modelos preditivos e generativos pode ser encontrado aqui ([https://github.com/Ragn95/MRL\\_Gram\\_negativa/blob/main/data\\_train.xlsx](https://github.com/Ragn95/MRL_Gram_negativa/blob/main/data_train.xlsx)). O modelo MRL utilizado na geração dos peptídeos ([https://github.com/Ragn95/MRL\\_Gram\\_negativa/blob/main/MRL\\_tutorials.proteins.amp\\_Abel.ipynb](https://github.com/Ragn95/MRL_Gram_negativa/blob/main/MRL_tutorials.proteins.amp_Abel.ipynb)). Todos os peptídeos gerados com MRL podem ser encontrados em ([https://github.com/Ragn95/MRL\\_Gram\\_negativa/blob/main/new\\_desing\\_peptídeos.xlsx](https://github.com/Ragn95/MRL_Gram_negativa/blob/main/new_desing_peptídeos.xlsx)). Além disso, o modelo utilizado para ajustar o AMP-BERT está disponível em ([https://github.com/Ragn95/MRL\\_Gram\\_negativa/blob/main/MRL\\_tutorials.proteins.amp\\_Abel.ipynb](https://github.com/Ragn95/MRL_Gram_negativa/blob/main/MRL_tutorials.proteins.amp_Abel.ipynb)).

## 11. Apêndice B - Espectros de massa dos peptídeos antimicrobianos obtidos comercialmente

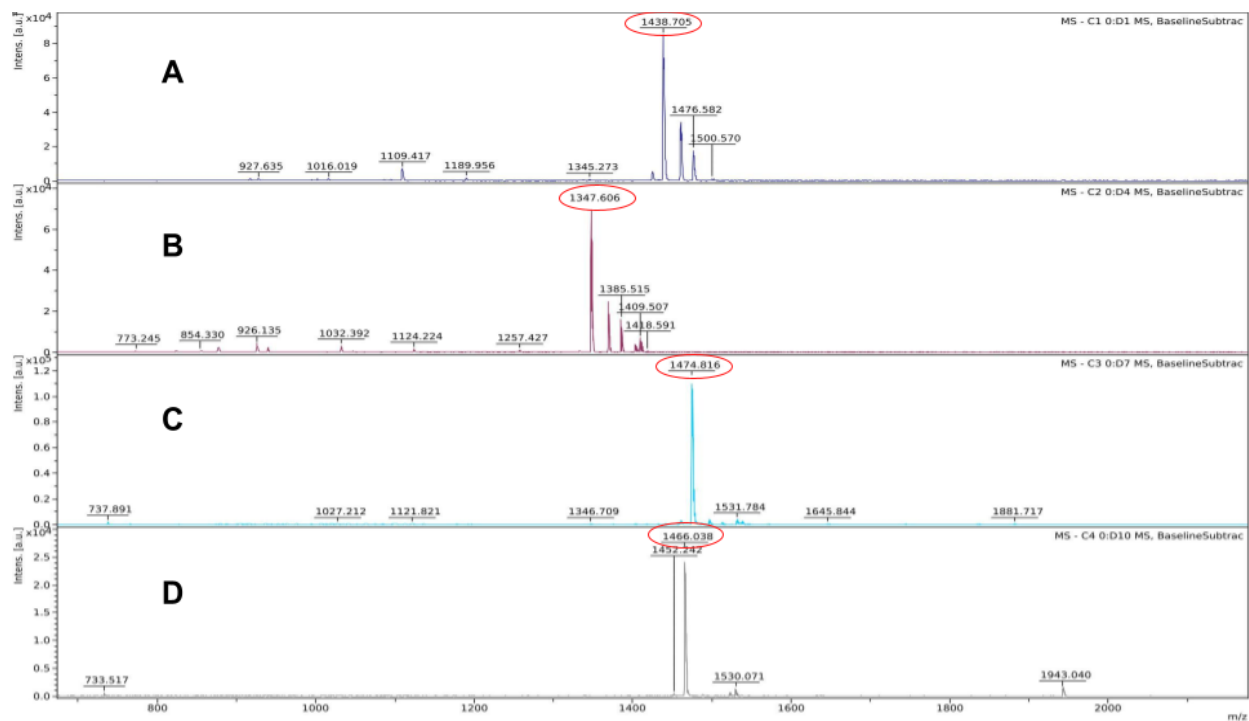


Figura S1: Espectros MALDI-TOF dos compostos A (Pep 1), B (Pep 2), C (Pep 3), e D (Pep 4).

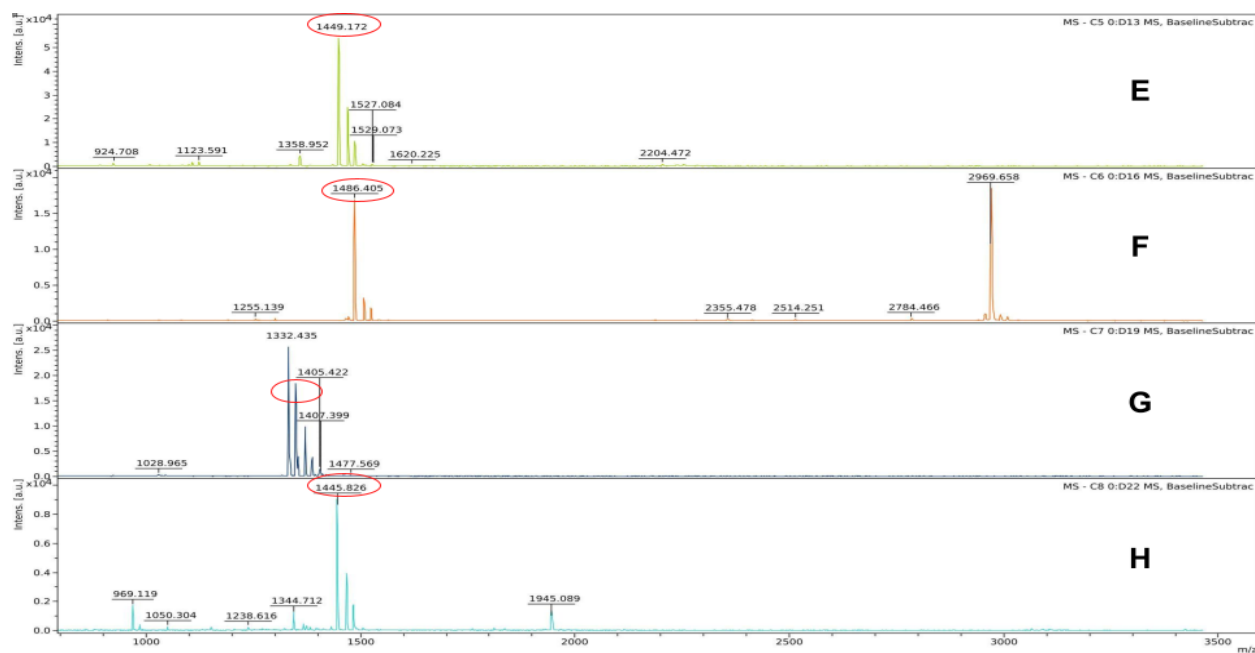


Figura S2: Espectros MALDI-TOF dos compostos E (Pep 5), F (Pep 6), G (Pep 7), e H (Pep 8).

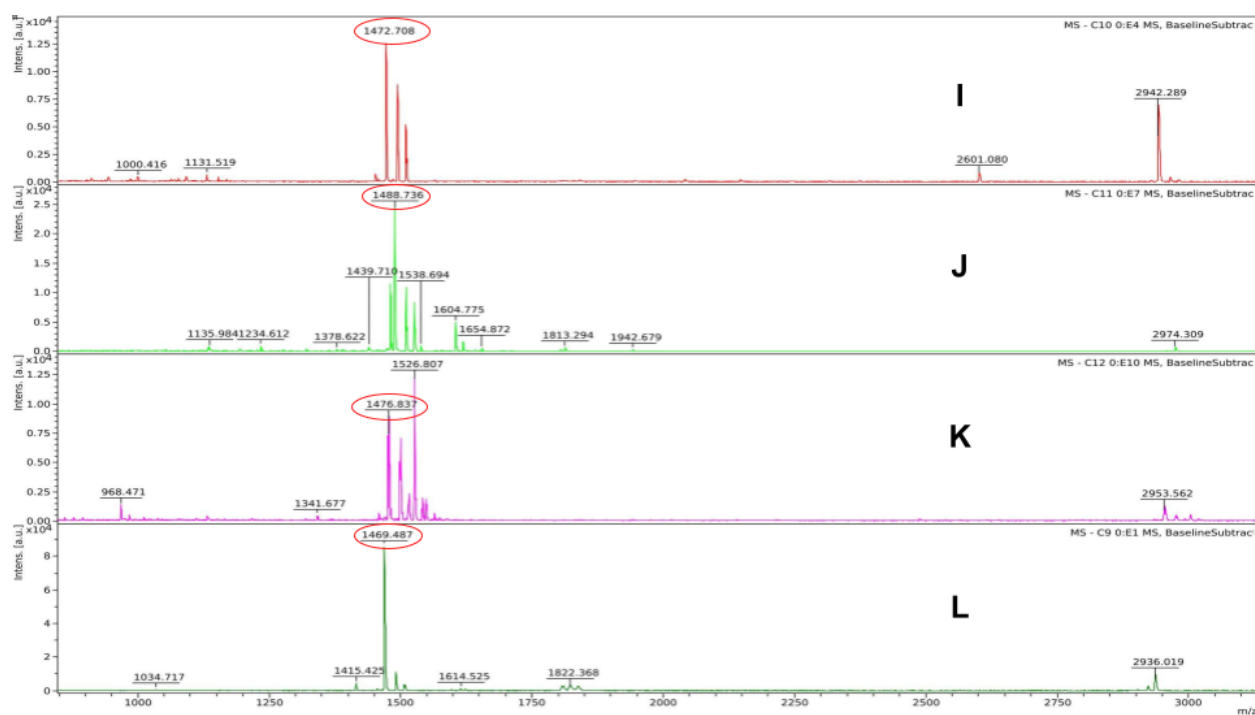


Figura S3: Espectros MALDI-TOF dos compostos I (Pep 10), J (Pep 11), K (Pep 12), e L (Pep 9).

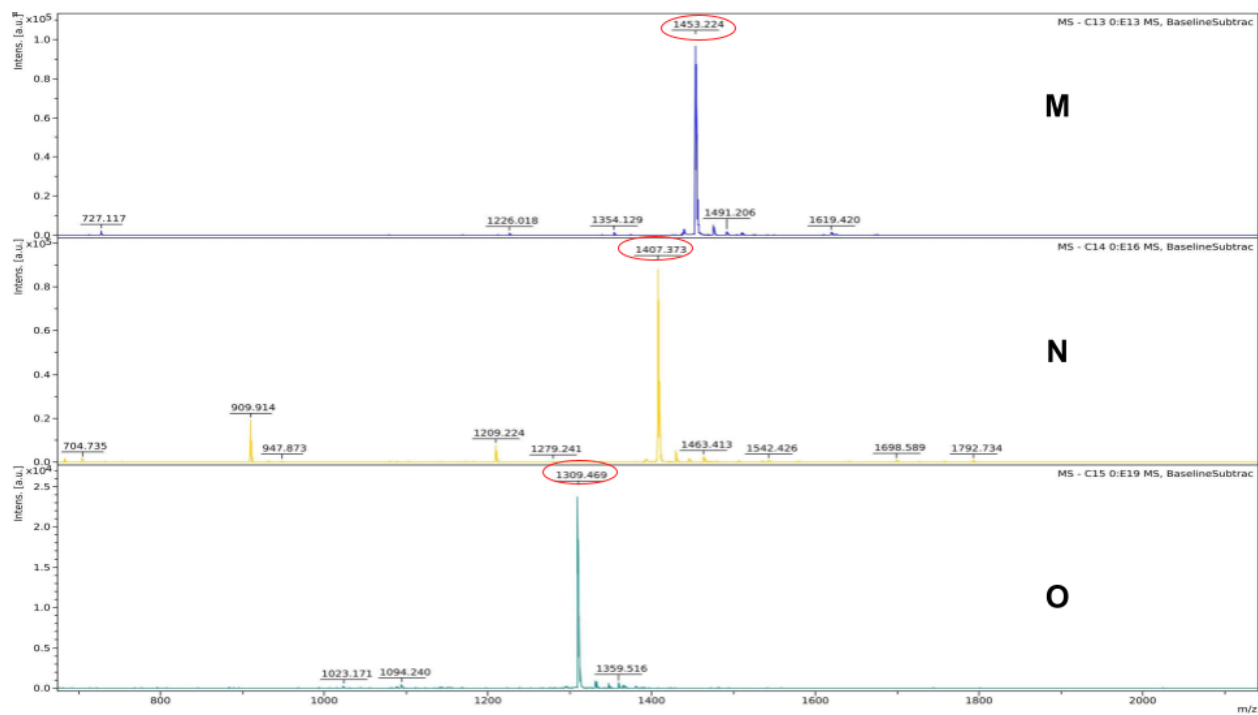


Figura S4: Espectros MALDI-TOF dos compostos M (Pep 13), N (Pep 14) e O (Pep 15).