UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

Daniel Rocha Silva

Prospecção de Peptídeos Antimicrobianos Baseada em Aprendizado Profundo em Dados Metagenômicos do Lago Guaíba durante e após a Enchente de 2024

Porto Alegre
2026

Daniel Rocha Silva

Prospecção de Peptídeos Antimicrobianos Baseada em Aprendizado Profundo em Dados Metagenômicos do Lago Guaíba durante e após a Enchente de 2024

Trabalho de Conclusão de curso apresentado como requisito parcial para obtenção do título de Bacharel em Biotecnologia na Universidade Federal do Rio Grande do Sul.

Área de habilitação Bioinformática

Orientadora: Prof. Dra. Fabiana Quoos Mayer
Coorientador: Dr. Emerson Gonçalves Moreira

Formatação e Estruturação: Baseadas nas normas do periódico *Journal of Chemical Information and Modeling* (JCIM).

PORTO ALEGRE
2026

CIP - Catalogação na Publicação

Silva, Daniel
Prospecção de Peptídeos Antimicrobianos Baseada em Aprendizado Profundo em Dados Metagenômicos do Lago Guaíba durante e após a Enchente de 2024 / Daniel Silva. -- 2026.
64 f.
Orientadora: Fabiana Quoos Mayer.

Coorientador: Emerson Gonçalves Moreira.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Curso de Biotecnologia: Bioinformática, Porto Alegre, BR-RS, 2026.

1. Peptídeos Antimicrobianos. 2. Dados metagenômicos. 3. Mudanças climáticas. 4. Resistência a antimicrobianos. 5. Enchente de 2024. I. Quoos Mayer, Fabiana, orient. II. Gonçalves Moreira, Emerson, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Daniel Rocha Silva

Prospecção de Peptídeos Antimicrobianos Baseada em Aprendizado Profundo em Dados Metagenômicos do Lago Guaíba durante e após a Enchente de 2024

Trabalho de Conclusão de curso apresentado como requisito parcial para obtenção do título de Bacharel em Biotecnologia na Universidade Federal do Rio Grande do Sul.

Área de habilitação: Bioinformática

Orientadora: Prof. Dra. Fabiana Quoos Mayer

Coorientador: Dr. Emerson Moreira

Aprovado em: Porto Alegre, 2 de julho de 2026.

Banca Examinadora:

Profa. Dra. Fabiana Quoos Mayer
Universidade Federal do Rio Grande do Sul

Prof. Dr. Maurício Menegatti Rigo

Prof. Dr. Márcio Dorn

Dedico este trabalho à memória daqueles que partiram nas águas de maio de 2024, a todos que perderam seus lares e aos animais desamparados. Que a ciência aqui desenvolvida sirva como um tributo à resiliência do povo gaúcho e à reconstrução do estado.

AGRADECIMENTOS

Trilhar um percurso tão árduo e apaixonante em um curso superior é um sonho que se tornou realidade. Essa trajetória pareceu ser infinita e breve ao mesmo tempo, de modo que nos acostumamos a apreciá-la, cada um à sua maneira. Logo, a minha trajetória foi de altos e baixos, o que moldou minhas vivências acadêmica e pessoal. Em muitos momentos, tive que aprender a trabalhar sozinho e continuo fazendo isso para avançar nas minhas metas e dificuldades. Sou grato por essas vivências, pois foram elas que me ensinaram a não ter medo de me desafiar e errar. No mundo acadêmico competitivo e exigente, a falha é temida até entre os mais dedicados. Demora um tempo para entender que "tudo bem! Não acaba por aqui".

À minha família. Meus pais, Vanderval e Jaqueline, foram pais incríveis, atenciosos e amorosos, principalmente. Apesar da distância, estavam 24 horas presentes, eles não dormem sem dar um "Boa noite, te amo, filho!". Desde que eu dei a notícia que eu havia passado na UFRGS até este instante, eles estiveram ao meu lado. Muito obrigado, sou imensamente feliz por tê-los como pais. Além disso, cresci, desde bebê, ao lado do meu irmão gêmeo Samuel, vê-lo trilhar seu caminho, paralelamente ao meu, foi uma jornada incrível. Apoiamos um ao outro incessantemente nas nossas vidas. Minha madrinha Mag, minha amiga Day e minha prima Mariana são mais que família, são minhas amigas! Trocamos muitas risadas enquanto estamos juntos, cada segundo com vocês se transforma em um sentimento de nostalgia, amor e saudades. Amo vocês todos! Sem vocês, eu não seria o Daniel de hoje, com um coração gentil e alegre.

À minha orientadora Fabiana Quoos Mayer. Desde que eu tive a minha primeira aula com ela, tive vontade de trabalhar com ela. Por coincidência, aqui estamos e que experiência ótima! Ela me propôs um projeto de TCC diferente para o laboratório dela, e foi um projeto muito legal de se trabalhar! A Fabi foi acolhedora em todos os momentos! O pessoal do laboratório (Allice, Vini, Vitória, Luísa, Luna, Emerson e os demais) foram bem carinhosos e me acolheram muito bem nesse pequeno espaço de tempo. Emerson foi um excelente coorientador, muito obrigado por ter sido tão companheiro me ensinando! Paralelamente ao TCC, eu estava fazendo estágio no laboratório da Prof. Fê Lopes e conheci pessoas doces e realmente incríveis. Sou grato por terem me acolhido e conhecido a todos!

A graduação não vale de nada sem os meus amigos! Bruno, Carol, Arthur, Deborah, Felipe, Isabelle, Andreioz, Emerson, vocês foram os momentos felizes neste percurso. Cada rolê, cada conversa descontraída no laboratório, ir ao cinema, jogar RPG foram os melhores momentos. Sem vocês, minha graduação seria vazia e não digo isso da boca para fora! Vocês me ensinaram demais durante esses anos! Espero que eu tenha os feito rir da minha maneira. Amo vocês!

Agradeço ainda ao Centro de Biotecnologia da UFRGS, por oferecer infraestrutura e compartilhar conhecimento ao longo de minha graduação, bem como ao *National Center for Biomolecular NMR Data Processing and Analysis*, por disponibilizar recursos computacionais indispensáveis à realização deste projeto. Este trabalho foi apoiado pela Iniciativa ResGen de Vigilância Genômica da Resistência a Antimicrobianos em Saúde Única, com fomento do CNPq/MCTI/FNDCT, e pelo projeto "CLIMASANO - Resposta Integrada de Saneamento e Saúde em Emergências Climáticas", financiado pela FAPERGS. Agradecemos o suporte financeiro e institucional concedido.

RESUMO

A resistência a antimicrobianos (AMR) é reconhecida como uma pandemia silenciosa crítica, exacerbada pelo uso indevido de antibióticos e pela dispersão ambiental de genes de resistência a antimicrobianos (ARGs). Eventos ambientais extremos, como as enchentes de 2024 no Rio Grande do Sul, Brasil, podem impulsionar mudanças nas comunidades microbianas e a mobilização de determinantes de resistência em ecossistemas aquáticos. Em resposta a tais pressões ecológicas, os microrganismos podem produzir Peptídeos Antimicrobianos (AMPs) como mecanismos primários de defesa e competição. Além disso, essas pressões seletivas, frequentemente agravadas por perturbações induzidas pelas mudanças climáticas, aceleram a evolução e a seleção dessas defesas químicas microbianas. Dada a complexidade molecular decorrente desses desastres ambientais, o Aprendizado de Máquina (ML) surgiu como uma abordagem moderna para reconhecer padrões biológicos complexos em AMPs. Dentro deste campo, modelos de Aprendizado Profundo (DL) têm demonstrado desempenho superior ao identificar automaticamente motivos bioativos e padrões funcionais diretamente de dados metagenômicos brutos. Neste contexto, este trabalho partiu da hipótese de que o estresse ecológico e as perturbações físico-químicas causadas pela inundação de 2024 no Rio Grande do sul no Lago Guaíba expuseram a comunidade microbiana local com um vasto conteúdo potencial biotecnológico, sendo uma fonte promissora para a descoberta de novos candidatos a AMPs. Para testar essa hipótese, este estudo teve como objetivo prospectar e caracterizar *in silico* AMPs a partir de dados metagenômicos do Lago Guaíba coletados durante e após as enchentes de 2024, avaliando sua diversidade, origem taxonômica e estabilidade estrutural. Empregando a abordagem baseada em modelos de DL, obteve-se a identificação de 179 candidatos a AMPs (c_AMPs) não tóxicos e não hemolíticos, dos quais **7**3 sequências (40,8%) tiveram uma identidade menor ou igual a 40% a sequências do banco de dados de referência DRAMP, evidenciando um elevado índice de ineditismo molecular. Os resultados sugerem que o ambiente gravemente impactado abriga uma diversidade bioquímica singular e ainda não mapeada. Esses c_AMPs inéditos despontam como novos candidatos para substituir ou complementar antibióticos tradicionais, indicando que tais moléculas podem auxiliar no combate à AMR e validando o *pipeline* computacional como uma ferramenta eficaz para a descoberta de biomoléculas.

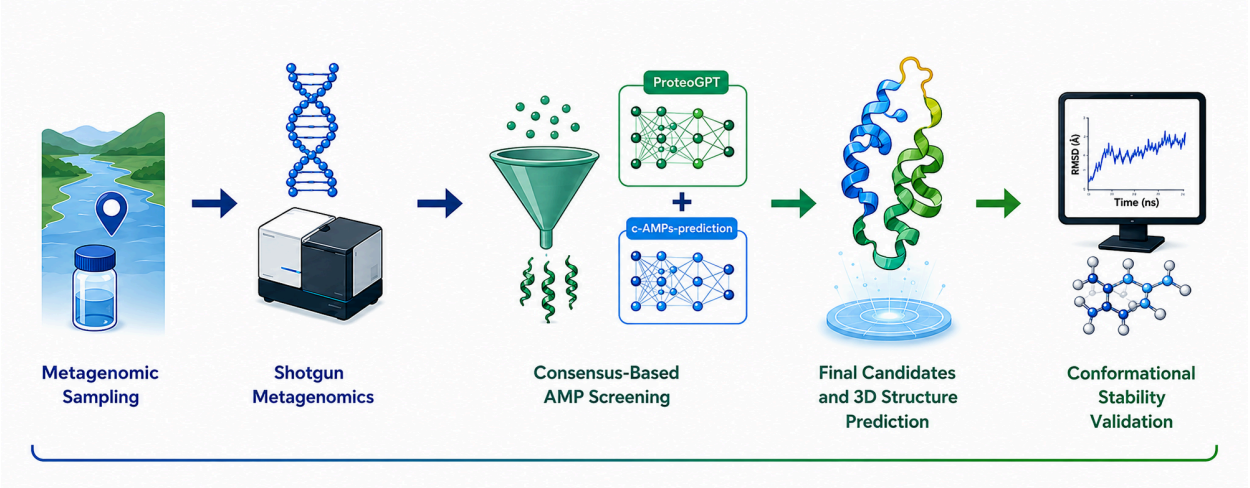
Palavras-chave: Resistência a antimicrobianos; Genes de resistência a antimicrobianos (ARGs); Rio Grande do Sul; Mudanças climáticas; Comunidades microbiana..

ABSTRACT

Antimicrobial resistance is recognized as a critical silent pandemic, exacerbated by the misuse of antibiotics and the environmental dispersion of antimicrobial resistance genes (ARGs). Extreme environmental events, such as the 2024 floods in Rio Grande do Sul, Brazil, can drive shifts in microbial communities and the mobilization of resistance determinants in aquatic ecosystems. In response to such ecological pressures, microorganisms can produce Antimicrobial Peptides (AMPs) as primary defense and competition mechanisms. Furthermore, these selective pressures, often compounded by climate change-induced disturbances, accelerate the evolution and selection of these microbial chemical defenses. Given the molecular complexity resulting from these environmental disasters, Machine Learning (ML) has emerged as a modern approach to recognize complex biological patterns in AMPs. Within this field, Deep Learning (DL) models have demonstrated superior performance by automatically identifying bioactive motifs and functional patterns directly from raw metagenomic data. In this context, this work targeted the hypothesis that the ecological stress and physicochemical disturbances caused by the 2024 flood in Rio Grande do Sul at Guaíba Lake exposed the local microbial community to a vast potential biotechnological content, serving as a promising source for discovering new antimicrobial peptide candidates. To test this hypothesis, this study aimed to prospect and characterize *in silico* antimicrobial peptides from metagenomic data collected from Guaíba Lake during and after the 2024 floods, assessing their diversity, taxonomic origin, and structural stability. Employing a DL model-based approach, 179 non-toxic and non-hemolytic AMP candidates (c_AMPs) were identified, of which 73 sequences (40.8%) shared less than or equal to 40% identity with reference database sequences (DRAMP), highlighting a high rate of molecular novelty. The findings suggest that the severely impacted environment harbors a unique and unmapped biochemical diversity. These novel c_AMPs emerge as promising candidates to replace or complement traditional antibiotics, indicating that these molecules can aid in the fight against antimicrobial resistance (AMR) while validating the computational pipeline as an effective tool for biomolecule discovery.

Keywords: Antimicrobial resistance; Antimicrobial resistance genes (ARGs); Floods; Rio Grande do Sul; Microbial community shifts.

RESUMO GRÁFICO



LISTA DE ABREVIAÇÕES E SIGLAS

ACC: Acurácia (*Accuracy*).

AMPs: Peptídeos Antimicrobianos (*Antimicrobial Peptides*).

AMR: Resistência aos Antimicrobianos (*Antimicrobial Resistance*).

ARGs: Genes de Resistência a Antimicrobianos (*Antimicrobial Resistance Genes*).

AUC: Área Sob a Curva (*Area Under the Curve*).

c_AMPs: Candidatos a Peptídeos Antimicrobianos (*Antimicrobial Peptides Candidates*).

CIM: Concentração Inibitória Mínima (*Minimum Inhibitory Concentration*)

CLI: Interface de Linha de Comando (*Command Line Interface*).

DL: Aprendizado Profundo (*Deep Learning*).

ESKAPE: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa and Enterobacter spp*.

FAIR: Encontrável (*Findable*), Acessível (*Accessible*), Interoperável (*Interoperable*) e Reutilizável (*Reusable*)

FN: Falso Negativo (*False Negative*).

FP: Falso Positivo (*False Positive*).

GDL: Aprendizado Profundo em Grafos (*Graph Deep Learning*).

JCIM: *Journal of Chemical Information and Modeling*.

MCC: Coeficiente de Correlação de Matthews (*Matthews Correlation Coefficient*).

ML: Aprendizado de Máquina (*Machine Learning*).

ORF: Matriz de Leitura Aberta (*Open Reading Frame*).

RS: Rio Grande do Sul.

SVM:Máquina de vetores de suporte (*Support Vector Machine)*

TN: Verdadeiro Negativo (*True Negative*).

TP: Verdadeiro Positivo (*True Positive*).

VPN: Valor Preditivo Negativo (*Negative Predictive Value*).

VPP: Valor Preditivo Positivo *(Positive Predictive Value)*

PCoA: Análise de Coordenadas Principais (*Principal Coordinate Analysis*)

SUMÁRIO

LISTA DE ABREVIAÇÕES E SIGLAS..... 11
1. INTRODUÇÃO..... 14
2. OBJETIVOS..... 18
2.1 Objetivo geral..... 18
2.2 Objetivos específicos..... 18
3. PROCEDIMENTOS METODOLÓGICOS.....19
3.1. Seleção dos pipelines de aprendizado profundo para predição de AMPs..... 19
3.2. Coleta, obtenção e pré-processamento de dados metagenômicos..... 19
3.3 Predição dos AMPs via consenso.....20
3.3.1 1ª Ferramenta: ProteoGPT.....20
3.3.2 2° Ferramenta: c_AMPs-prediction.....21
3.3.3 3ª Ferramenta: esm-AxP-GDL..... 22
3.4 Avaliação in silico de toxicidade e atividade hemolítica.....23
3.5 Análise Estatística da Dinâmica Espaço-Temporal..... 23
3.6 Alinhamento com banco de dados de AMPs e análise de ineditismo..... 24
3.7 Atribuição taxonômica..... 24
3.8 Análise da diversidade da comunidade microbiana..... 24
3.9 Associação dos AMPs com gêneros bacterianos.....25
3.10 Caracterização físico-química in silico.....26
3.11 Modelagem estrutural e estabilidade in silico..... 27
3.11.1 Preparação dos Sistemas.....27
3.11.2 Minimização e Equilibração.....28
3.11.3 Produção..... 28
3.11.4 Análises trajetórias e estabilidade estrutural.....29
5. RESULTADOS.....30
5.1 Pré-processamento dos metadados..... 30
5.2 Seleção dos pipelines de aprendizado profundo.....30
5.3 Predição dos AMPs via modelos de aprendizado profundo.....31
5.3.1 Predição de AMPs via consenso: 1º modelo: ProteoGPT..... 31
5.3.2 Predição de AMPs via consenso: 2º modelo: c_AMPs-prediction.....33
5.4 Filtro de Predição Estrutural Tridimensional: Ferramenta esm-AxP-GDL.....33
5.5 Testes de toxicidade e hemólise dos AMPs.....34
5.6 Caracterização físico-química dos AMPs.....34
5.7 Distribuição temporal e espacial dos AMPs.....36
5.8 Alinhamento com banco de dados de AMPs.....38
5.9 Associação dos AMPs com gêneros bacterianos.....39
5.10 Simulação Molecular.....40
5.10.1 Desvio Quadrático Médio (RMSD).....41
5.10.2 Raio de Giro (RG)..... 42
5.10.3 Flutuação Quadrática Média por Resíduo (RMSF).....43
5.10.4 Estabilidade do perfil de Estrutura Secundária (DSSP)..... 43

5.11 Atribuição taxonômica do metagenoma..... 44
5.11.1 Composição em nível de filo..... 44
5.11.2 Composição em nível de gênero..... 46
5.11.3 Avaliação da diversidade taxonômica dos metagenomas..... 46
6. DISCUSSÃO..... 48
7. CONCLUSÕES..... 54
REFERÊNCIAS..... 55

1. INTRODUÇÃO

A resistência aos antimicrobianos (AMR) é o processo biológico pelo qual microrganismos, incluindo bactérias, fungos e vírus, evoluem para resistir à ação de fármacos antimicrobianos anteriormente eficazes, tornando as infecções mais difíceis de tratar e aumentando o risco de disseminação de doenças e óbitos (GBD 2021 Antimicrobial Resistance Collaborators 2024). Esse contexto consolidou-se como uma das ameaças mais graves e alarmantes à saúde pública global no século XXI. A AMR foi diretamente responsável por 1,14 milhão de mortes em 2021 e associada a um total de 4,71 milhões de óbitos no mesmo ano (GBD 2021 Antimicrobial Resistance Collaborators 2024), estimando-se que alcance a marca de 10 milhões de óbitos até 2050 (Ho et al. 2025). Devido a esse impacto, a crise é frequentemente designada como uma "pandemia silenciosa" (Williams, 2024).

O panorama da AMR é intensificado pelas atividades dos setores veterinário e agrícola, que contribuem para a pressão seletiva por meio do uso de antimicrobianos não apenas para fins terapêuticos e metafiláticos, mas também de forma profilática e como promotores de crescimento (Pinho et al. 2025). Como consequência, observa-se a dispersão de genes de resistência a antimicrobianos (ARGs) e resíduos de antimicrobianos nos ecossistemas por meio de efluentes e excreções animais, favorecendo a aceleração da dinâmica de seleção de linhagens resistentes (Karwowska 2024). Complementarmente, na medicina humana, o uso indevido e excessivo de antimicrobianos, bem como a prescrição inadequada desses fármacos, contribuíram para a emergência e disseminação da AMR (MacNair et al. 2024).

Essa crise de eficácia terapêutica é evidenciada pela ascensão de patógenos de alta prioridade, cujos perfis de resistência tornam o manejo clínico extremamente complexo. Dentre os agentes etiológicos mais críticos, o grupo de bactérias *Enterobacterales, Acinetobacter baumannii e Mycobacterium tuberculosis* lideram as estatísticas do grupo crítico (World Health Organization, 2024) (Miller and Arias 2024), seguidos por outros patógenos de relevância clínica como *Klebsiella pneumoniae* e *Staphylococcus aureus* (Murray et al., 2022).

A gravidade dessas infecções é um reflexo direto da sofisticação dos mecanismos moleculares de sobrevivência bacteriana. Nesse sentido, a base genética desses fenômenos reside nos ARGs, os quais podem ser adquiridos e disseminados entre comunidades bacterianas por meio de processos de transferência horizontal de genes (HTG) (Inda-Díaz et al. 2023). Grande parte desses determinantes encontram-se

integrados a elementos genéticos móveis (MGEs), tais como plasmídeos conjugativos e transposons, que facilitam a migração de genes de bactérias ambientais para patógenos de relevância clínica (Partridge et al. 2018). Ambientes aquáticos e áreas impactadas por efluentes (como águas residuais) destacam-se, portanto, como zonas de alto risco para a mobilização e disseminação desses novos determinantes de resistência (Inda-Díaz et al. 2023).

A dinâmica de disseminação de ARGs pode ser catalisada por desastres naturais, com destaque para as enchentes, que podem atuar como promotores da dispersão em larga escala de microrganismos e ARGs do sedimento e dos sistemas de esgoto diretamente para o ambiente urbano (Wong, 2024; Owrangi *et al.*, 2014). Sob a influência das mudanças climáticas, a ocorrência e a intensidade desses fenômenos tendem a se elevar, com projeções de precipitações extremas cada vez mais frequentes (Wong, 2024). Dessa maneira, a intensificação desses eventos extremos sobrecarrega sistemas de tratamento e transforma perturbações ecológicas em crises de saúde pública sistêmicas (Barredo, 2007; Wong, 2024).

A enchente ocorrida em maio de 2024 no Estado do Rio Grande do Sul, Brasil, evidenciou a magnitude do perigo das mudanças climáticas, configurando-se como um evento histórico que demonstrou a necessidade de novas estratégias de monitoramento ambiental e vigilância genômica (Ramos et al. 2024). No contexto do Lago Guaíba, a inundação não apenas comprometeu os sistemas de tratamento de água, mas também promoveu o carreamento massivo de efluentes urbanos e resíduos de diversas origens para o corpo hídrico (Dall'Igna, 2025). Somado à carga orgânica, o aporte de resíduos de antibióticos e outros insumos farmacêuticos e químicos, provenientes da excreção humana, do descarte inadequado de medicamentos, das inundações de postos de combustíveis, indústrias, áreas agrícolas, assim como da rede de esgoto sanitário, pode acelerar a pressão seletiva sobre a microbiota local.

Esse cenário pode ter resultado em uma perturbação ecológica das comunidades microbianas, gerando um impacto ambiental de alta complexidade que pode favorecer a mobilização de ARGs nos microbiomas ambientais (Kim et al., 2022). Essa proliferação de determinantes ambientais de resistência agrava o declínio da eficácia dos tratamentos clínicos tradicionais, tornando a busca por novas entidades químicas uma urgência global.

Paralelamente ao cenário de resistência aos fármacos disponíveis no mercado, existe o histórico vazio de descoberta no qual o desenvolvimento de novas classes de antibióticos convencionais estagnou nas últimas décadas (Bourzac, 2026); emerge a

necessidade da busca por estratégias alternativas aos antibióticos convencionais, como, por exemplo, os peptídeos antimicrobianos (AMPs; entre 2 e 50 resíduos de aminoácidos) (Talapko *et al.* 2022). Ao lançar um novo olhar sobre os AMPs, visualiza-se o potencial terapêutico, uma vez que essas biomoléculas atuam de forma disruptiva contra a integridade estrutural da membrana ou camada protetora do microrganismo, estabelecendo um mecanismo de ação físico-químico contra o qual os patógenos enfrentam fortes barreiras evolutivas para desenvolver resistência (Bourzac, 2026).

Os AMPs representam um mecanismo de defesa primário e evolutivamente conservado, sendo amplamente gerados de forma natural por quase todos os organismos vivos (Haryini et al. 2026). Na microbiota ambiental, essas moléculas são sintetizadas como ferramentas de competição por nicho e resposta a estresses ecológicos, enquanto em organismos superiores atuam como pilares da imunidade inata (Haryini et al. 2026). Os AMPs apresentam baixo peso molecular (0,2 a 5,5 kDa) e surgem como uma das principais frentes biotecnológicas para o enfrentamento da crise de multirresistência — quando um microrganismo é resistente a pelo menos 3 classes de agentes antimicrobianos (MacNair et al. 2024).

Ao contrário dos antibióticos convencionais, que geralmente possuem um único alvo celular, os AMPs podem atuar em múltiplos componentes bacterianos simultaneamente (como a membrana citoplasmática, síntese de proteínas e replicação de DNA), o que pode ser uma vantagem (Bucataru and Ciobanasu 2024). Desse modo, estas moléculas são uma das soluções biotecnológicas promissoras para preencher a lacuna terapêutica deixada pelos antibióticos tradicionais.

A exploração prática desse vasto repertório químico natural, contudo, enfrenta barreiras metodológicas. As abordagens tradicionais de triagem experimental em laboratório (*wet-lab*) são altamente precisas, porém complexas, dispendiosas e ineficientes frente à escala da demanda atual, especialmente considerando que a maioria dos microrganismos ambientais não é cultivável (Santos-Júnior et al. 2024; Zhou et al. 2025). Embora o advento da metagenômica ambiental tenha superado esse obstáculo ao dar acesso direto ao material genético total de comunidades complexas, ela deslocou o gargalo para o campo computacional ao gerar um volume massivo de dados de sequenciamento (*Big Data*) (Ma et al. 2022; Santos-Júnior et al. 2024).

Rastrear sequências bioativas utilizando métodos tradicionais, como alinhamento local e análise de homologia (ex. BLAST), torna-se inviável, pois essas ferramentas consomem recursos computacionais excessivos, são propensas a omissões e falham em detectar padrões funcionais decorrentes de evolução convergente (Lin et al. 2025). Para

mitigar esse gargalo analítico, técnicas de inteligência artificial têm revolucionado a prospecção de bioprodutos, viabilizando o mapeamento do "resistoma" e a descoberta de novas moléculas no microbioma de diversos ecossistemas (Santos-Júnior et al. 2024). Enquanto o aprendizado de máquina convencional (*Machine Learning* - ML) depende fortemente de engenharia de atributos manuais (*feature engineering*) limitada pelo viés do especialista, as arquiteturas de Aprendizado Profundo (*Deep Learning* - DL) oferecem uma nova estratégia ao extraírem automaticamente representações e motivos funcionais complexos diretamente de sequências genômicas brutas (Lin et al. 2025, Zhou et al. 2025).

Neste contexto, as amostras metagenômicas coletadas no Lago Guaíba durante e após as históricas inundações de 2024 constituem um potencial repositório genético para a bioprospecção *in silico* de AMPs. Este trabalho busca, portanto, aplicar modelos de DL para minerar e predizer candidatos a peptídeos antimicrobianos codificados nesse microbioma impactado. Ao fazê-lo, visa-se converter um cenário de crise climática e sanitária regional em uma oportunidade estratégica de inovação biotecnológica, preenchendo lacunas de conhecimento científico frente ao esgotamento das terapias convencionais.

2. OBJETIVOS

2.1 Objetivo geral

Prospectar e caracterizar *in silico* peptídeos antimicrobianos (AMPs) a partir de dados metagenômicos do Lago Guaíba coletados durante e após as enchentes de 2024, avaliando sua diversidade, origem taxonômica e estabilidade estrutural.

2.2 Objetivos específicos

- Selecionar e testar *pipelines* de predição de AMPs a partir de dados metagenômicos;
- Prospectar AMPs a partir de dados metagenômicos do Lago Guaíba coletados durante e após as enchentes de 2024;
- Avaliar a abundância e diversidade de candidatos a AMPs (c_AMPs) entre as amostras coletadas durante e após o evento da enchente de 2024;
- Predizer a estrutura tridimensional (3D) das sequências peptídicas e avaliar sua estabilidade conformacional;
- Caracterizar as propriedades físico-químicas (como carga líquida, solubilidade, hidrofobicidade e ponto isoelétrico) dos peptídeos candidatos para inferir seu potencial biotecnológico.
- Realizar a análise taxonômica dos metagenomas.

3. PROCEDIMENTOS METODOLÓGICOS

3.1. Seleção dos pipelines de aprendizado profundo para predição de AMPs

Para a identificação e triagem de AMPs a partir das sequências metagenômicas, foi elaborada uma matriz de decisão focada em modelos preditivos de peptídeos antimicrobianos. A seleção metodológica baseou-se em cinco critérios de avaliação: o desempenho preditivo dos modelos, medido por métricas estatísticas como Acurácia, área sob a curva (AUC) e o Coeficiente de Correlação de Matthews (MCC), essenciais para a confiabilidade em dados desbalanceados; a adequação da arquitetura aos princípios *FAIR* (garantindo que os dados sejam encontráveis, acessíveis, interoperáveis e reutilizáveis); a maturidade da infraestrutura, evidenciada pela disponibilidade de código-fonte aberto e facilidade de integração local via interface de linha de comando (CLI); a solidez da validação *in silico* baseada em *benchmarks* independentes; e a existência de validação experimental, priorizando algoritmos cujas predições foram corroboradas por ensaios biológicos *in vitro* ou *in vivo*.

3.2. Coleta, obtenção e pré-processamento de dados metagenômicos

O conjunto de dados analisado neste estudo compreendeu 78 amostras metagenômicas coletadas em 10 regiões distintas do Lago Guaíba (Figura 1), realizadas por pesquisadores do Instituto de Pesquisas Hidráulicas da UFRGS, distribuídas ao longo de oito campanhas amostrais entre maio de 2024 e julho de 2025 (em 08/05, 21/05, 04/06, 27/06, 21/08 e 29/11 de 2024, e 29/05 e 25/07 de 2025), de modo a interpretar a diversidade espaço-temporal dos AMPs decorrente da enchentes de 2024 no Rio Grande do Sul. O sequenciamento foi executado por meio de uma abordagem metagenômica *shotgun* baseada na plataforma Illumina, gerando *reads* pareadas de 150 pb. O controle de qualidade das leituras brutas (*raw reads*) foi verificado utilizando a ferramenta FastQC (Andrews, 2010), seguido pelo processo de montagem *de novo* das sequências por meio do software SPAdes operado em modo metagenômico (*metaSPAdes*)(Nurk et al., 2017).

A predição de Janelas Abertas de Leitura (ORFs — *Open Reading Frames*) foi conduzida através do software Prodigal (Hyatt, 2010), com parâmetros configurados para a identificação direcionada de genes codificadores de peptídeos curtos, limitando o comprimento máximo das sequências a 50 aminoácidos. Para assegurar a qualidade do banco de dados e mitigar ruídos analíticos, aplicou-se uma etapa de filtragem estrita para a remoção de sequências parciais e caracteres nucleotídicos ou de aminoácidos não identificados (como "X" e asteriscos). Posteriormente, a redundância de sequência foi eliminada com o emprego da ferramenta

CD-HIT (Li e Godzik, 2006), e foi aplicado um limiar de similaridade de 90% para evitar a superestimação de moléculas idênticas.

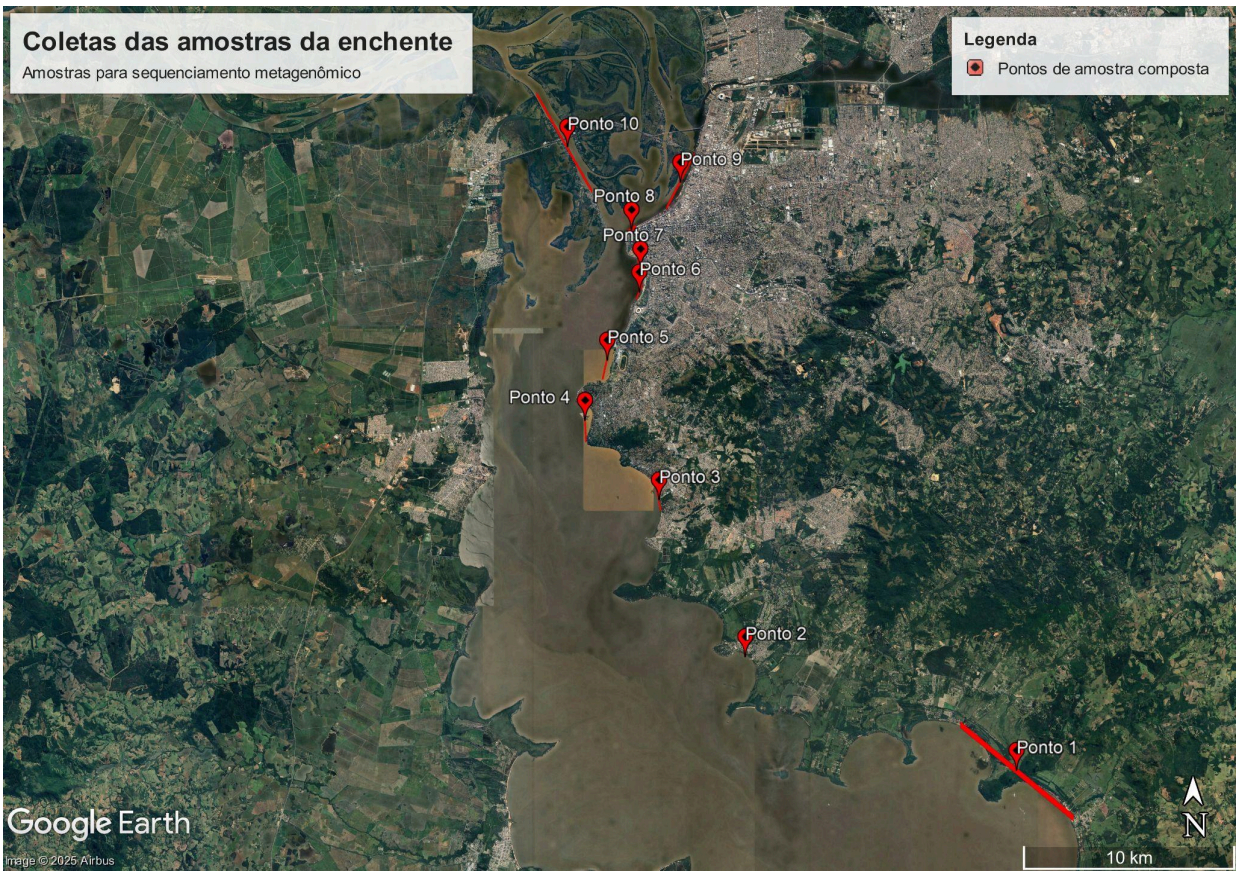


Figura 1. Mapa com as localizações amostradas durante este estudo.

3.3 Predição dos AMPs via consenso

3.3.1 1ª Ferramenta: ProteoGPT

A prospecção inicial de peptídeos antimicrobianos de alta confiança foi conduzida utilizando o ProteoGPT (Wang et al. 2025), um modelo de linguagem de proteínas de grande escala (LLM — *Large Language Model*) baseado na arquitetura autorregressiva *Transformer* (especificamente GPT-2) composto por 124 milhões de parâmetros. Desenvolvido sob o paradigma de aprendizado por transferência (*transfer learning*), o modelo base foi inicialmente pré-treinado utilizando 609.216 sequências do banco de dados UniProtKB/Swiss-Prot para a assimilação de representações contextuais profundas e padrões latentes ("gramática") das cadeias polipeptídicas. Para a aplicação neste estudo, a inferência foi estruturada em um pipeline de funil sequencial composto por submodelos especializados e refinados (*fine-tuning*), garantindo a triagem em alto rendimento com critérios estritos de bioatividade e biossegurança.

O fluxo analítico do ProteoGPT é segmentado em três etapas principais de filtragem e ranqueamento:

1. AMPSorter (Classificação Primária): Submodelo treinado de forma supervisionada com um conjunto de dados composto por 16.062 AMPs validados e 16.549 sequências não-AMPs. Esta camada atua na identificação primária das sequências brutas, discriminando fragmentos com potencial de atividade antimicrobiana.
2. BioToxiPept (Filtro de Citotoxicidade): Sub-LLM especializado na predição de toxicidade e hemólise frente a células humanas, treinado com um banco balanceado de 1.932 peptídeos tóxicos e 1.932 não tóxicos. Essa etapa funciona como um gargalo de segurança farmacológica, descartando precocemente candidatos com potencial citotóxico.
3. Módulo QSAR (Ranqueamento Físico-Químico): Os candidatos remanescentes dos filtros biológicos são submetidos a uma análise quantitativa de relação estrutura-atividade (QSAR). Esta etapa calcula propriedades físico-químicas estatísticas das sequências para mensurar índices de afinidade e estabilidade, gerando métricas de ordenação conhecidas como *Activity Score*, *MaxScore* e *Ranking Score* para priorizar as moléculas de maior interesse biotecnológico.

A predição e a confiabilidade estatística do ProteoGPT foram avaliadas por meio de testes de desempenho. Inicialmente, estruturou-se um conjunto de dados de validação balanceado composto por 1.000 sequências de AMPs verdadeiras (*true-AMPs*), obtidas a partir do *Antimicrobial Peptide Database 3* (APD3), e 1.000 sequências de controle não-AMPs (*non-AMPs*), extraídas do banco UniProt. Adicionalmente, para aferir o comportamento do modelo sob condições de desbalanceamento extremo, cenário altamente representativo do ruído analítico encontrado no processamento de dados metagenômicos ambientais, a arquitetura foi submetida a um teste de estresse contra um universo de 178.346 sequências categorizadas como não-AMP extraídas do banco UniProt.

3.3.2 2° Ferramenta: c_AMPs-prediction

Para consolidar o *pipeline* de triagem e mitigar o acúmulo de falsos positivos decorrentes do viés catiônico das classificações primárias, incorporou-se a ferramenta de consenso *c_AMPs-prediction* (Ma et al. 2022). Este modelo opera sob a estratégia de aprendizado por conjunto (*ensemble learning*), unificando três arquiteturas profundas independentes de

Processamento de Linguagem Natural (PLN): uma rede de memória de longo prazo bidirecional (*Bidirectional Long Short-Term Memory* — BiLSTM), mecanismos de Atenção (*Attention*) e modelos baseados em transformadores (*Bidirectional Encoder Representations from Transformers* — BERT) (Ma et al., 2022). Para viabilizar a entrada dos dados nas redes neurais de forma homogênea, o pré-processamento converteu as sequências lineares de aminoácidos em matrizes numéricas densas utilizando uma estratégia de preenchimento fixo (*zero-padding*) parametrizada em 300 colunas, garantindo a integridade dimensional dos vetores estruturais sem perda de contexto biológico.

A predição e a confiabilidade estatística do *c_AMPs-prediction* foram avaliadas por meio de testes de desempenho. Inicialmente, estruturou-se um conjunto de dados de validação balanceado composto por 1.000 sequências de AMPs verdadeiras (*true-AMPs*), obtidas a partir do *Antimicrobial Peptide Database 3* (APD3), e 1.000 sequências de controle não-AMPs (*non-AMPs*), extraídas do banco UniProt. Adicionalmente, para aferir o comportamento do modelo sob condições de desbalanceamento extremo, cenário altamente representativo do ruído analítico encontrado no processamento de dados metagenômicos ambientais, a arquitetura foi submetida a um teste de estresse contra um universo de 178.346 sequências categorizadas como não-AMP extraídas do banco UniProt.

3.3.3 3ª Ferramenta: esm-AxP-GDL

Como última camada de refinamento e validação bioinformática do *pipeline* de consenso, os candidatos a peptídeos remanescentes foram submetidos à ferramenta avançada esm-AxP-GDLv (Cordoves-Delgado e García-Jacas, 2024). Diferente das abordagens estritamente baseadas em sequências lineares (1D) discutidas anteriormente, este modelo incorpora a informação da conformação espacial tridimensional (3D) e propriedades evolutivas por meio de arquiteturas de Aprendizado Profundo em Grafos (*Graph Deep Learning*, GDL). O algoritmo opera convertendo as coordenadas atômicas tridimensionais dos peptídeos em grafos moleculares, onde os resíduos de aminoácidos são mapeados como nós, enriquecidos com embutimentos estruturais do modelo de linguagem ESM-2, e as distâncias espaciais euclidianas entre os carbonos alfa (Cα) definem as arestas conectivas. Esse mapeamento geométrico permite capturar motivos estruturais tridimensionais discretos que determinam a atividade biológica de anfipaticidade e interação com membranas-alvo.

Devido à alta demanda computacional intrínseca aos algoritmos de dobramento atômico, a execução direta da modelagem 3D estrutural em larga escala para o universo total de centenas de milhares de sequências metagenômicas iniciais mostrou-se inviável. Para contornar esse gargalo e viabilizar a execução do *pipeline* sob restrições severas de *hardware*, o fluxo de trabalho foi adaptado. A rotina de predição espacial tridimensional utilizando a rede ESMFold foi configurada para operar em sessões baseadas estritamente em processamento em CPU, funcionando como uma camada subsequente de filtragem aplicada apenas ao grupo selecionado de prováveis candidatos bioativos.

Para atestar a confiabilidade preditiva do modelo geométrico sob essa adaptação computacional, implementou-se uma validação baseada em uma amostragem piloto balanceada contendo inicialmente 2.000 sequências polipeptídicas. Contudo, devido às restrições arquiteturais intrínsecas dos servidores e modelos de linguagem da Meta (ESM2/ESMFold) em relação aos limites de tamanho e delimitação física das janelas de leitura, sequências consideradas excessivamente longas ou fora das fronteiras de estabilidade geométrica do algoritmo foram automaticamente desconsideradas pelo sistema.

3.4 Avaliação in silico de toxicidade e atividade hemolítica

Para assegurar a viabilidade farmacológica e a segurança biológica dos c_AMPs selecionados, as sequências foram submetidas a análises de toxicidade e potencial hemolítico. A toxicidade foi avaliada pelo ToxinPred3 (Rathore et al. 2024), adotando-se o modelo preditivo baseado em *Random Forest* (RF) com um limiar de corte de escore $\geq 0{,}50$ para a designação de moléculas tóxicas. De forma complementar, o potencial hemolítico foi analisado pela plataforma HemoPi 2.0 (Chaudhary et al. 2016), empregando o modelo preditivo baseado em *Support Vector Machine* (SVM) com valor de corte de escore SVM $\geq 0{,}0$ para a determinação de atividade hemolítica. Por fim, as sequências que excederam os limiares de corte pré-estabelecidos por ambos os algoritmos foram excluídas, refinando a biblioteca de candidatos para aqueles com maior potencial terapêutico e segurança biológica.

3.5 Análise Estatística da Dinâmica Espaço-Temporal

A variação na frequência de detecção dos cAMPs ao longo das campanhas amostrais foi avaliada mediante estatística não paramétrica, justificando-se pela natureza de contagem discreta e distribuição não normal dos dados metagenômicos. O teste de Kruskal-Wallis (Kruskal; Wallis,

1952) foi aplicado para determinar a existência de variância global significativa na abundância de sequências entre as diferentes datas de coleta. Subsequentemente, as diferenças específicas entre os períodos amostrais foram identificadas por meio do teste *post-hoc* de Dunn para comparações par a par. Para evitar o viés de falsos positivos decorrente de múltiplas comparações, os valores de probabilidade foram ajustados utilizando a correção de Benjamini-Hochberg (FDR - *False Discovery Rate*) (Benjamini; Hochberg, 1995). O nível de significância adotado para todas as análises foi de 5% ($p < 0,05$).

3.6 Alinhamento com banco de dados de AMPs e análise de ineditismo

As sequências candidatas a AMPs foram submetidas a alinhamento local utilizando o algoritmo BLASTp contra os bancos de dados de referência DRAMP (FAN et al., 2016) e AMPSphere (Santos-Júnior et al., 2024). A partir dos alinhamentos, extraiu-se o valor de identidade máxima de sequência (*pident*) de cada candidato. Estabeleceu-se um limiar de corte de 40% para a classificação estrutural das moléculas: peptídeos com identidade máxima inferior a 40% em ambos os bancos de dados foram classificados como sequências inéditas. Em contrapartida, sequências com identidade igual ou superior a este limiar foram categorizadas como variantes ou homólogos de peptídeos antimicrobianos previamente descritos.

3.7 Atribuição taxonômica

A classificação taxonômica baseada em *k-mers* foi realizada nos *contigs* montados e filtrados utilizando o software Kraken2 (Wood et al., 2019). O processamento computacional foi executado no servidor de alto desempenho Dayhoff (Centro de Biotecnologia - UFRGS). Utilizou-se o banco de dados *Standard* do Kraken2 versão de 15/10/2025, composto por genomas de referência do NCBI referentes aos domínios Archaea e Bacteria, além de vírus, plasmídeos e o genoma humano. O algoritmo operou com 8 *threads*, formatação de saída nominal *(--use-names)* e limiar de confiança padrão (*confidence score* = 0). Os relatórios individuais (*reports*) foram processados via scripts customizados em Python para a extração e criação das matrizes de abundância nos níveis taxonômicos de gênero e filo.

3.8 Análise da diversidade da comunidade microbiana

As análises de diversidade ecológica e os testes estatísticos foram conduzidos em ambiente R, empregando os pacotes *phyloseq* (Mcmurdie; Holmes, 2013), *vegan* (Oksanen et al., 2022) e *FSA* (Ogle et al., 2024). Visando uma avaliação temporal de resolução da sucessão microbiana, as amostras metagenômicas foram analisadas considerando as oito campanhas de

coleta distintas, que englobam desde a fase aguda da inundação (maio de 2024) até o período de restabelecimento do corpo hídrico (julho de 2025).

A avaliação da diversidade estrutural baseou-se na matriz de abundância taxonômica do nível de gênero gerada pelo Kraken2 (Wood et al., 2019) . A diversidade alfa (riqueza e equitabilidade) da comunidade microbiana foi mensurada pelo Índice de Shannon (H′). A normalidade da distribuição dos dados foi inspecionada por meio do teste de Shapiro-Wilk (Shapiro; Wilk, 1965). Em função da ausência de normalidade paramétrica, as variações temporais da diversidade alfa foram avaliadas pelo teste global de Kruskal-Wallis (Kruskal; Wallis, 1952), seguido pelo teste *post-hoc* de comparações múltiplas de Dunn (Dunn, 1964) com correção da taxa de falsa descoberta de Benjamini-Hochberg (FDR) (Benjamini; Hochberg, 1995).

A diversidade beta, indicadora das discrepâncias na composição da comunidade entre os períodos, foi estimada a partir da matriz de distâncias de Bray-Curtis (Bray; Curtis, 1957). A significância estatística das alterações estruturais ao longo do tempo foi testada por análise de variância permutacional (PERMANOVA) (Anderson, 2001) com 999 permutações. Para assegurar as premissas do modelo permutacional e descartar eventuais vieses decorrentes de variâncias intragrupais desiguais, conduziu-se o teste de homogeneidade de dispersão multivariada (PERMDISP) (Anderson, 2006). A visualização espacial do padrão de sucessão ecológica foi construída através da ordenação por análise de coordenadas principais (PCoA).

3.9 Associação dos AMPs com gêneros bacterianos

A investigação da influência dos candidatos a peptídeos antimicrobianos (c_AMPs) sobre gêneros bacterianos da microbiota do Lago Guaíba foi conduzida por meio de uma análise estatística de abundância diferencial. O método buscou identificar associações entre a detecção de peptídeos específicos e a variação na abundância de gêneros bacterianos.

Para cada peptídeo, as amostras foram segregadas em um modelo binário: grupo de "presença" (valor 1.0) e grupo de "ausência" (valor 0.0). A comparação das medianas da abundância relativa dos gêneros bacterianos entre esses dois grupos foi realizada mediante o teste não paramétrico U de Mann-Whitney (Conover, 1999), justificado pela distribuição não normal dos dados metagenômicos. A análise foi configurada como um teste unilateral (*alternative='less'*), visando identificar gêneros cuja abundância fosse significativamente menor nas amostras contendo o peptídeo. Foram consideradas apenas as associações que atenderam aos critérios de valor de $p < 0,05$ e *delta de mediana* negativo. Posteriormente, para a construção da

rede de interações no software Cytoscape (Shannon et al. 2003), aplicou-se um filtro de significância ($p < 0{,}001$) para minimizar a ocorrência de falsos positivos.

3.10 Caracterização físico-química in silico

Após a etapa de filtragem e seleção dos candidatos a AMPs, foi realizada a caracterização das propriedades físico-químicas de cada sequência peptídica. A análise compreendeu o cálculo de quatro parâmetros:

- Carga Líquida: Determinada pelo balanço entre os resíduos carregados positivamente (Lisina e Arginina) e negativamente (Ácido Aspártico e Ácido Glutâmico) em pH fisiológico (Lehninger *et al.*, 2019). A carga determina a interação eletrostática entre o peptídeo e as membranas bacterianas negativas.

  Calculada pela soma das cargas formais dos resíduos de aminoácidos em pH 7,0. Considerou-se a carga +1 para Lisina (K) e Arginina (R) e a carga -1 para Ácido Aspártico (D) e Ácido Glutâmico (E), conforme as propriedades de ionização das cadeias laterais.

- Momento Hidrofóbico (μH): Calculado com base na escala de hidrofobicidade de (Dayhoff *et al.*, 1978), este parâmetro avalia a anfifilicidade da estrutura primária. O μH é obtido projetando-se os valores de hidrofobicidade (hi) dos resíduos em uma escala angular de 100º (período da hélice-α), conforme a equação:

$$\mu H = \sqrt{(\sum hi \cdot sin(i \cdot 100°))^2 + (\sum hi \cdot cos(i \cdot 100°))^2},$$

  onde hi é o valor de hidrofobicidade do aminoácido na posição i, 100° é o ângulo de rotação da hélice-alfa e n é o número total de resíduos.

- Índice de Boman: Este índice reflete a propensão de uma proteína em interagir com outras moléculas, sendo calculado pela soma das energias de solubilidade de cada resíduo de aminoácido (Boman, 2003):

$$Boman = \frac{\sum Ei}{n}$$

Valores elevados deste índice indicam potencial de interação proteína-proteína, enquanto valores baixos sugerem maior propensão à interação com membranas lipídicas.

- Índice de Hidropaticidade (GRAVY): Calculado pela média aritmética dos valores de hidropaticidade de cada resíduo (Gi) segundo a escala de Kyte-Doolittle (Kyte; Doolittle, 1982):

$$GRAVY = \frac{\sum Gi}{n},$$

em que Gi é o valor de hidropaticidade do aminoácido na posição i.

Este índice fornece uma estimativa da hidrofobicidade global da sequência, sendo um indicador relevante para a compreensão da solubilidade e estabilidade do peptídeo em meio aquoso.

- O ponto isoelétrico (pI) das sequências de peptídeos foi calculado utilizando o módulo ProtParam da biblioteca Biopython (Cock et al., 2009). Com base nos valores de pI, os peptídeos foram classificados em três categorias estruturais: catiônicos/básicos (pI > 8,0), neutros (6,5 ≤ pI ≤ 8,0) e aniônicos/ácidos (pI < 6,5). A associação linear entre os valores de carga líquida e os valores de pI calculados foi avaliada por meio do coeficiente de correlação de Pearson (r). Esta análise foi aplicada de forma idêntica à biblioteca global de ineditismo (n = 179).

3.11 Modelagem estrutural e estabilidade in silico

3.11.1 Preparação dos Sistemas

As estruturas tridimensionais dos peptídeos foram inicialmente preditas utilizando o AlphaFold3 (Jumper et al., 2021; Abramson et al., 2024). A preparação dos sistemas para as simulações de Dinâmica Molecular (DM) foi realizada por meio do servidor web CHARMM-GUI (Jo et al., 2008), empregando o módulo *Solution Builder*. Os peptídeos foram solvatados em uma caixa contendo o modelo de água explícita TIP3P e contra-íons (Na^+ e Cl^-) foram adicionados a uma concentração de 0,15 M para neutralização do sistema. A topologia do sistema foi descrita utilizando o campo de força CHARMM36m (Huang et al., 2017). Todas as etapas de simulação foram conduzidas utilizando o software GROMACS, versão 2025.2.

3.11.2 Minimização e Equilibração

A minimização da energia inicial do sistema foi realizada através do algoritmo *Steepest Descent*. Em seguida, a termalização e o equilíbrio do solvente foram realizados em etapas sucessivas de restrição posicional gradativa sobre os átomos pesados do soluto. Inicialmente, o sistema foi submetido ao *ensemble* canônico (NVT) a 298,15 K por 1 ns, acoplado ao termostato *V-rescale* com constante de acoplamento térmico (τt) de 0,1 ps. Posteriormente, a equilibração foi realizada no *ensemble* isotérmico-isobárico (NPT) a 1 bar por 1 ns em seis etapas. Nas etapas iniciais de equilibração (passos 1 a 5), aplicou-se o barostato de *Berendsen* para a estabilização do volume da caixa. As constantes de força de restrição posicional foram reduzidas progressivamente de 5000 kJ/mol/nm² para 1000 kJ/mol/nm² sobre os átomos pesados do esqueleto (*backbone*) e das cadeias laterais. Na etapa final de equilibração (NPT, passo 6), o sistema foi transicionado para o barostato de *Parrinello-Rahman* com acoplamento isotrópico (pressão de referência de 1,0 bar e constante de acoplamento de pressão τp = 2,0 ps). Nessa fase, a restrição harmônica de 1000 kJ/mol/nm² foi mantida sobre o esqueleto peptídico, permitindo a liberação e acomodação das cadeias laterais antes do início da etapa de produção.

3.11.3 Produção

A simulação de produção foi conduzida no ensemble NPT, mantendo-se a temperatura de referência estável a 298 K e a pressão a 1,0 bar, sob o controle dos algoritmos de acoplamento *V-rescale* (τt=0,1 ps) e *Parrinello-Rahman* (τp=2,0 ps). Para cada peptídeo, coletou-se uma trajetória contínua de 500 ns com um passo de integração (*time step*) de 2 fs. De forma consistente com as etapas prévias de minimização e equilibração, as interações eletrostáticas de longo alcance foram tratadas via método *Particle Mesh Ewald* (PME), enquanto, para as forças de van der Waals e as interações eletrostáticas de curto alcance, foi estabelecido um raio de corte de 1,2 nm. Todas as ligações covalentes envolvendo átomos de hidrogênio foram restringidas por meio do algoritmo *Linear Constraint Solver* (LINCS) (HESS *et al.*, 1997).

3.11.4 Análises trajetórias e estabilidade estrutural

Para a avaliação da estabilidade conformacional, flexibilidade local e manutenção das estruturas secundárias dos peptídeos ao longo dos 500 ns de simulação de produção, as trajetórias foram inicialmente processadas para a eliminação de artefatos de condições periódicas de contorno (PBC). Utilizou-se o módulo *gmx trjconv*, aplicando-se a centralização molecular do soluto (-pbc mol -center) para garantir a integridade geométrica dos peptídeos ao longo de toda a coordenada temporal.

O desvio posicional global das estruturas em relação ao modelo predito inicial foi quantificado por meio do cálculo do desvio quadrático médio (*Root-Mean-Square Deviation* - RMSD), executado pelo módulo *gmx rms* com unidades expressas em Ångströms (Å). Concomitantemente, a compactação molecular e o relaxamento estrutural das cadeias foram monitorados pelo cálculo do raio de giro (Rg) através da ferramenta *gmx gyrate*, permitindo correlacionar a estabilidade macroscópica com o volume ocupado pelo peptídeo ao longo do tempo.

A flexibilidade local e as flutuações dinâmicas individuais de cada aminoácido foram determinadas pelo cálculo da Flutuação Quadrática Média (*Root-Mean-Square Fluctuation* - RMSF) por resíduo, empregando-se o módulo *gmx rmsf -res*. Por fim, a evolução temporal da estrutura secundária (como α-hélices e folhas β) foi mapeada numericamente pelo algoritmo *Define Secondary Structure of Proteins* (DSSP) por meio da interface *gmx do_dssp*, acoplada ao binário externo do programa DSSP. Os dados brutos gerados foram integrados e plotados utilizando scripts customizados em linguagem Python 3 (bibliotecas *Matplotlib* e *Seaborn*).

5. RESULTADOS

5.1 Pré-processamento dos metadados

A etapa inicial de predição de ORFs através do software Prodigal (HYATT, 2010), direcionada a peptídeos curtos (≤ 50 aminoácidos) e submetida à filtragem estrita de qualidade resultou em um montante inicial de 876.601 sequências proteicas filtradas. A subsequente aplicação da ferramenta CD-HIT (Li e Godzik, 2006) para eliminar a redundância de moléculas idênticas reduziu com precisão o volume analítico para 665.630 sequências de proteínas únicas e não redundantes. Um refinamento de 24,07% do banco de dados inicial.

5.2 Seleção dos pipelines de aprendizado profundo

Após a aplicação e avaliação da matriz de decisão, conforme descrito na seção 3.1, três ferramentas destacaram-se e foram selecionadas para compor o *pipeline* analítico (Tabela 1). É importante destacar que outras ferramentas como *Macrel* (2020) (Santos-Júnior et al. 2020), *deep-AMPpred* (2025) (Zhao et al. 2025) e *UniAMP* (2025) (Chen et al. 2025) foram avaliadas, porém não se mostraram adequadas aos critérios estabelecidos. Os critérios para essa exclusão fundamentaram-se em entraves técnicos como o risco de viés em dados de treinamento, a falta de conformidade com os princípios *FAIR* (o que inviabilizaria a reprodutibilidade dos dados), a limitação com o *hardware* disponível e, em última análise, a obtenção de métricas de desempenho inferiores em *benchmarks* independentes quando comparadas aos modelos selecionados.

Neste contexto, o *ProteoGPT* foi escolhido devido à sua arquitetura de Modelos de Linguagem Generativa (*Transformers*), que oferece alta sensibilidade para escanear grandes volumes de dados metagenômicos e extrair padrões biológicos ocultos. De forma complementar, o *c_AMPs-prediction* foi integrado ao fluxo por sua abordagem conservadora de aprendizado profundo, utilizando um funil de predição baseado no consenso simultâneo de três redes neurais distintas (BERT, LSTM e mecanismo de *Attention*), o que garante especificidade aos resultados. Por fim, a terceira ferramenta selecionada foi o modelo de Aprendizado Profundo em Grafos (GDL / *Graph* DAMP), chamado *esm-AxP-GDL*, escolhido especificamente por sua capacidade de incorporar informações estruturais tridimensionais (preditas pelo ESMFold) em representações de grafos, permitindo a análise das interações espaciais e topológicas entre os resíduos de aminoácidos, compensando e mitigando as limitações dos modelos baseados puramente em sequências lineares.

Tabela 1. Modelos selecionados para prospecção de cAMPs nos dados metagenômicos do Lago Guaíba.

Nome da ferramenta	DOI / Referência	Arquitetura	Dados de treinamento	Fonte
c_AMPs-prediction (2022)	10.1038/s41587-022-01226-0	*Ensemble (BERT, Attention, LSTM)*	8.290 AMPs; 74.838 não-AMPs	https://github.com/mayuefine/c_AMPs-prediction
ProteoGPT (2025)	10.1038/s41564-025-02114-4	*Transformer (GPT-2)*	609k seq; 16k AMPs; 1.9k tóxicos	https://github.com/W1V1995/AMP_Project
esm-AxP-GDL (2024)	10.1021/acs.jcim.3c02061	*Graph DL (GAT)*	67.058 peptídeos	https://github.com/cicese-biocom/esm-AxP-GDL

Fonte: Elaborado pelo autor (2026).

5.3 Predição dos AMPs via modelos de aprendizado profundo

5.3.1 Predição de AMPs via consenso: 1º modelo: ProteoGPT

No ensaio de desempenho, o modelo *ProteoGPT* demonstrou uma capacidade discriminatória alta, atingindo 96,86% de acurácia, 95% de sensibilidade e um coeficiente de correlação de *Matthews* (MCC) de 0,938. O espectro completo de métricas avaliadas, englobando acurácia, sensibilidade, especificidade, precisão, valor preditivo negativo (VPN) e MCC (Figura 2).

Para mapear a distribuição das predições e avaliar a eficiência do classificador na segregação das classes sem vieses, gerou-se a matriz de confusão do ensaio (Figura 3). A matriz evidencia o equilíbrio alcançado pelo funil analítico entre verdadeiros positivos e verdadeiros negativos, permitindo a identificação exata de como os erros residuais foram distribuídos pelo modelo. Mesmo diante desse cenário, o ProteoGPT demonstrou um desempenho robusto, exibindo uma especificidade de 99,57% e limitando a ocorrência de falsos positivos a apenas 751 sequências.

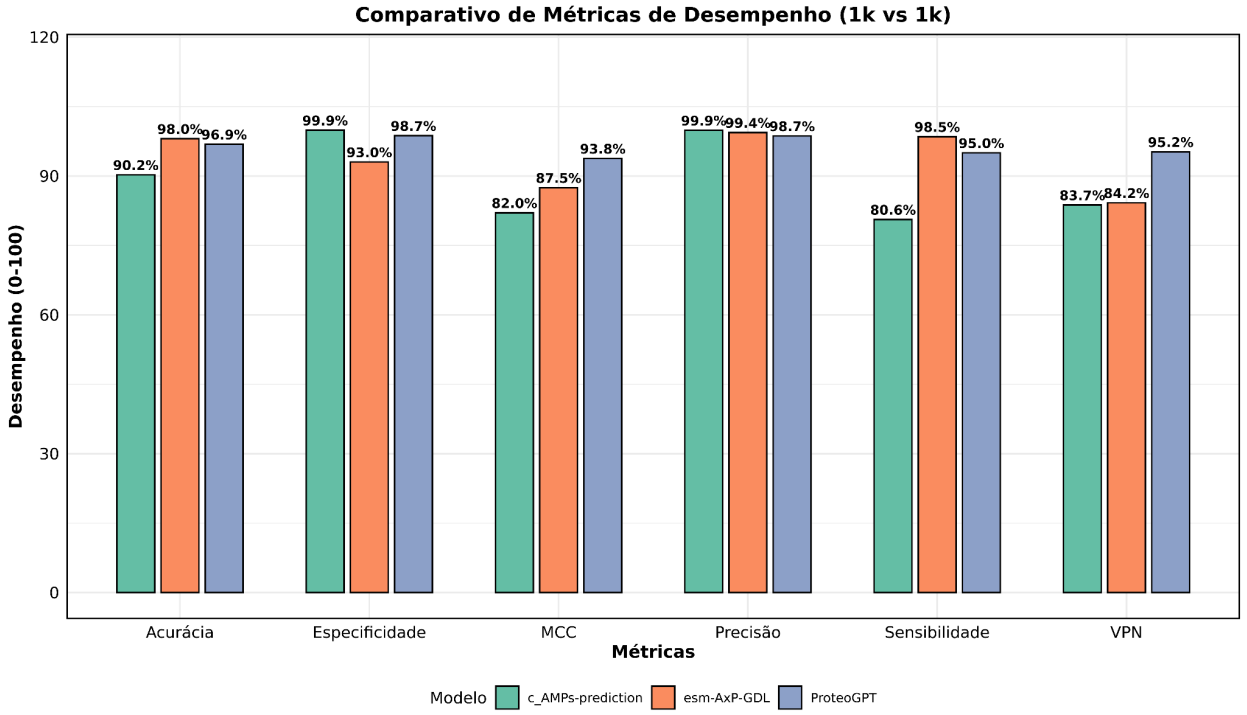


Figura 2. Comparativo das métricas de desempenho dos modelos de *deep learning* avaliados (ProteoGPT, c_AMPs-prediction e esm-AxP-GDL). Os resultados refletem os testes de validação balanceados (1.000 *true-AMPs* da APD3 versus 1.000 *non-AMPs* da UniProt) para os modelos ProteoGPT e c_AMPs-prediction, e a validação no lote piloto estrutural (n=1.076 sequências cruzadas) para o modelo geométrico esm-AxP-GDL.

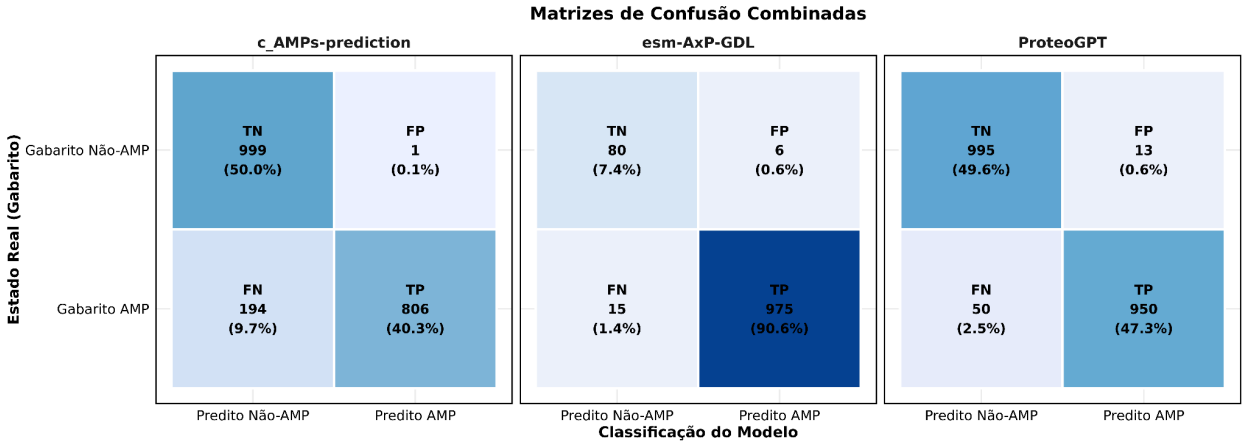


Figura 3. Matrizes de confusão referentes aos testes de validação preditiva dos modelos ProteoGPT, c_AMPs-prediction e esm-AxP-GDL. Os quadrantes em cada painel indicam a distribuição absoluta e percentual dos acertos e erros de classificação do respectivo algoritmo: Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), Falsos Positivos (FP) e Falsos Negativos (FN).

Iniciando o fluxo analítico do trabalho atual, um total de 665.630 sequências de proteínas foi submetido à avaliação do modelo *ProteoGPT*. A partir desta prospecção em larga escala, a ferramenta extraiu 1.936 peptídeos com características compatíveis com a atividade

antimicrobiana. Esse rendimento representa o isolamento de aproximadamente 0,29% do montante original.

5.3.2 Predição de AMPs via consenso: 2º modelo: c_AMPs-prediction

A arquitetura do *c_AMPs-prediction* é um filtro conservador, conferindo elevada segurança analítica à prospecção biotecnológica. Em um ensaio de validação balanceado composto por 1.000 sequências de controle, o modelo exibiu uma capacidade de discriminação elevada, identificando apenas 1 falso positivo em todo o universo analisado. Esse comportamento refletiu-se em índices altos de especificidade e valor preditivo positivo (VPP) de 99,9%, além de um coeficiente de correlação de Matthews (MCC) de 0,820 (Figura 2). Para detalhar o mapeamento das classificações e atestar a ausência de desvios sistemáticos nas predições do consenso neural, a matriz de confusão correspondente a este ensaio foi gerada e está apresentada na Figura 3. A concordância quase absoluta entre as classes previstas e os alvos reais (*ground truth*) validou o uso da ferramenta para a prospecção.

A validação em condições de grande volume de sequências confirmou a robustez do modelo frente a bancos de dados massivos e ruidosos. Ao ser submetido a um teste de grande volume de dados contra um universo desbalanceado de 178.346 peptídeos, o *c_AMPs-prediction* sustentou sua estabilidade estatística, limitando a taxa de descoberta de falsos positivos a 0,11% (com apenas 202 sequências erroneamente sinalizadas).

Na etapa paralela de triagem, o *pipeline c_AMPs-prediction*, compondo a segunda vertente da prospecção, processou a totalidade das 665.630 sequências proteicas e resultou na identificação de 7 peptídeos candidatos. O modelo reteve 0,001% do total do montante original. Conforme o consenso simultâneo de três redes neurais independentes (BERT, LSTM e Attention) da sua estrutura, obteve-se uma redução da ocorrência de falsos positivos.

5.4 Filtro de Predição Estrutural Tridimensional: Ferramenta esm-AxP-GDL

O processo piloto de validação do modelo estrutural resultou em um conjunto final de 1.076 sequências peptídicas cruzadas, cujas estruturas 3D foram completamente resolvidas e submetidas à inferência de validação. O modelo alcançou um índice de Verdadeiros Positivos (TP) de 975 sequências corretamente identificadas como AMPs e Verdadeiros Negativos (TN)

de 80 sequências não-AMP validadas. A taxa de erro residual foi minimizada a apenas 6 falsos positivos (FP; alarmes falsos) e 15 falsos negativos (FN; sequências omitidas). A distribuição matemática completa dessas métricas de desempenho, incluindo a Acurácia Geral de 98,05%, Sensibilidade de 98,48%, Especificidade de 93,02%, Precisão (VPP) de 99,39% e o Valor Preditivo Negativo (VPN) de 84,21% (Figura 2).

A sua matriz de confusão do lote piloto de validação do modelo geométrico *esm-AxP-GDL* (n=1.076 sequências), evidenciou uma acurácia geral de 98,05% apoiada por uma taxa de erro global restrita a apenas 2,0%. O classificador demonstrou capacidade discriminativa ao confirmar 975 verdadeiros positivos (90,6%) e 80 verdadeiros negativos (7,4%). O principal destaque da *pipeline* reside na minimização de falsos positivos, registrando apenas 6 falsos positivos (0,6%), o que eleva a precisão (valor preditivo positivo - VPP) para 99,39%, enquanto os falsos negativos ficaram limitados a 15 sequências (1,4%).

Para a etapa final de triagem dos dados metagenômicos do Lago do Guaíba, o conjunto combinado de 1.943 peptídeos, resultante das ferramentas anteriores baseadas puramente em sequências lineares, foi submetido à avaliação do framework esm-AxP-GDL. Atuando como um filtro de validação tridimensional e evolutiva baseado em grafos, o modelo confirmou 711 peptídeos como candidatos viáveis. Nesta etapa, houve a retenção de aproximadamente 36,6% das moléculas de entrada. A exclusão da maior parte das entradas evidencia a limitação das análises estritamente unidimensionais, demonstrando que a incorporação de coordenadas espaciais é importante para mitigar falsos positivos ao descartar sequências que não adotam o dobramento tridimensional necessário para a interação com membranas bacterianas.

5.5 Testes de toxicidade e hemólise dos AMPs

Após a triagem estrutural realizada pelo modelo *esm-AxP-GDL*, que identificou 711 sequências com potencial antimicrobiano, os candidatos foram submetidos a uma avaliação de toxicidade e potencial hemolítico. Esse processo de filtragem resultou na seleção final de 179 peptídeos classificados como não hemolíticos e não tóxicos.

5.6 Caracterização físico-química dos AMPs

A análise físico-química dos 179 candidatos a AMPs revelou um perfil de ampla variabilidade estrutural. A carga líquida média observada foi de 6,53 (± 5,66), com valores situados entre -7,00 e +25,00, evidenciando uma tendência à natureza catiônica, característica

favorável à interação com membranas bacterianas. O momento hidrofóbico (μH) apresentou média de 0,17 (± 0,09), indicando a presença de anfifilicidade estrutural, atributo necessário para a inserção em bicamadas lipídicas (Figura 4A) .

Em relação aos índices de estabilidade e solubilidade, o índice de Boman apresentou média de 3,38 (± 0,57) (Figura 4B), enquanto o índice de hidropaticidade (GRAVY) obteve média de -0,57 (± 0,76) (Figura 4A). Estes resultados demonstram um balanço entre a solubilidade em meio aquoso e o potencial de interação molecular dos peptídeos.

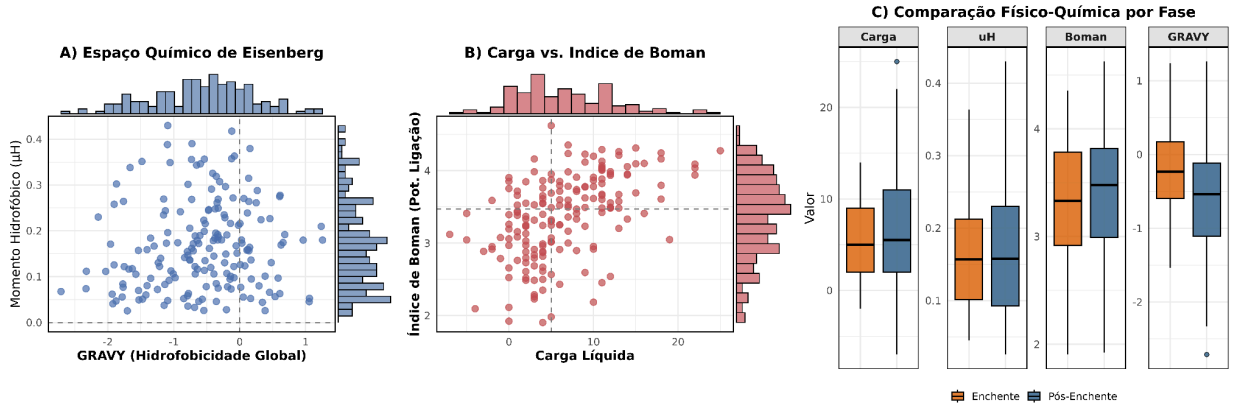


Figura 4. Perfil físico-químico dos peptídeos validados (n=179). **(A)** Espaço químico de Eisenberg relacionando o momento hidrofóbico (μH) e a hidrofobicidade global (GRAVY), com densidades marginais. **(B)** Relação entre a carga líquida e o índice de Boman (potencial de ligação a proteínas). **(C)** Distribuição comparativa das propriedades físico-químicas (carga, momento hidrofóbico, índice de Boman e GRAVY) entre as amostras obtidas no pico da enchente e no período de monitoramento pós-enchente.

O comportamento físico-químico global dos peptídeos foi avaliado por meio da relação entre o momento hidrofóbico (μH) e a hidrofobicidade global (GRAVY) (Figura 4A) . A análise indica diversidade estrutural, com GRAVY médio de -0,57 (± 0,76), variando entre -2,71 e 1,26, e μH médio de 0,17 (± 0,10), com amplitude de 0,03 a 0,43. A análise comparativa entre as fases de coleta (Figura 4C) aponta variações nos perfis médios dos candidatos. Os peptídeos identificados apresentam carga líquida média de 6,53 (± 5,66), com valores situados entre -7,00 e 25,00.

A análise físico-química das 179 sequências registrou média aritmética de *pI* de 9,64, mediana de 10,11 e amplitude de distribuição entre 4,05 e 12,00. A classificação quanto ao perfil de carga indicou 153 peptídeos (85,5%) como catiônicos, 9 (5,0%) como neutros e 17 (9,5%) como aniônicos. O coeficiente de correlação de Pearson entre a carga líquida e o *pI* foi de r = 0,7665.

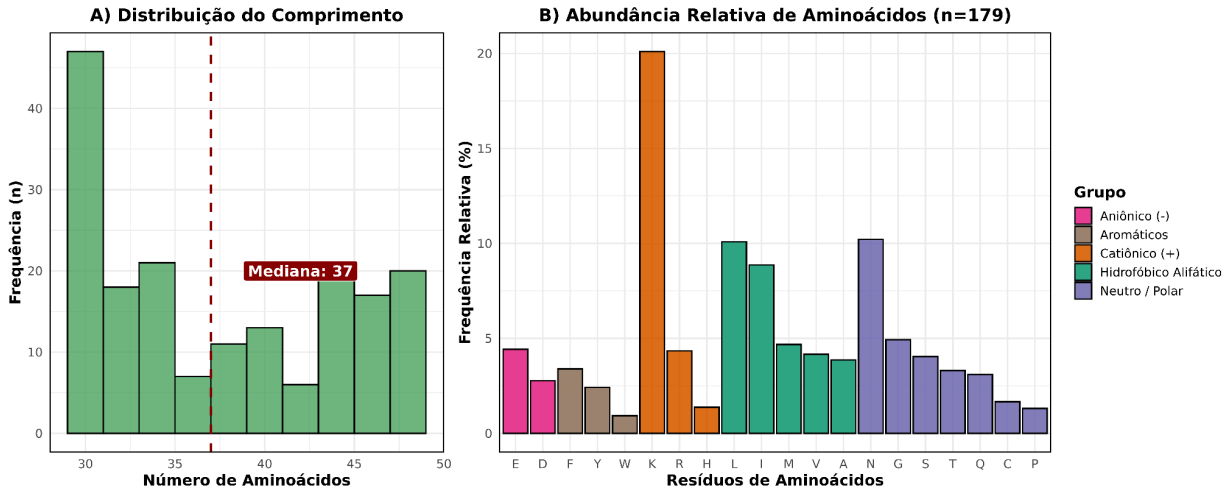


Figura 5. Características estruturais e de composição dos peptídeos antimicrobianos prospectados (n=179). **(A)** Distribuição de frequência do comprimento das sequências peptídicas, com indicação da mediana (linha tracejada). **(B)** Abundância relativa percentual dos resíduos de aminoácidos nas sequências analisadas, classificados segundo suas propriedades químicas: catiônicos, aniônicos, hidrofóbicos, aromáticos e neutros/polares.

As 179 sequências peptídicas selecionadas apresentaram um comprimento médio de 37,85 (± 7,02) aminoácidos, com uma amplitude que variou de 29 a 49 aminoácidos. A análise da distribuição (Figura 5A) revela um perfil bimodal, com concentrações de sequências na faixa de 30 a 48 aminoácidos. Conforme a análise da composição dos 179 AMPs, a frequência relativa de aminoácidos destaca o enriquecimento de resíduos catiônicos, notadamente lisina (K), que representa 20% da composição total, seguida por resíduos polares e hidrofóbicos, como asparagina (N), leucina (L) e isoleucina (I) (Figura 5B).

5.7 Distribuição temporal e espacial dos AMPs

A análise da dinâmica espaço-temporal e da evolução das ocorrências (Figura 6) baseou-se no conjunto de 1943 peptídeos identificados pelas ferramentas *ProteoGPT* e *c_AMPs-prediction*. Esta etapa exploratória teve como objetivo mapear a diversidade e a abundância de sequências com potencial antimicrobiano antes da aplicação dos filtros de segurança, toxicidade e ineditismo. A média de identificações por ponto de coleta foi de 17,90 (± 17,58), com produtividade máxima registrada no Ponto 5 (67 ocorrências) e mínima no Ponto 9 (5 ocorrências). No âmbito temporal, a média observada foi de 22,38 (± 19,87) ocorrências por data, com coeficiente de variação de 0,89, evidenciando a heterogeneidade das detecções ao longo do monitoramento.

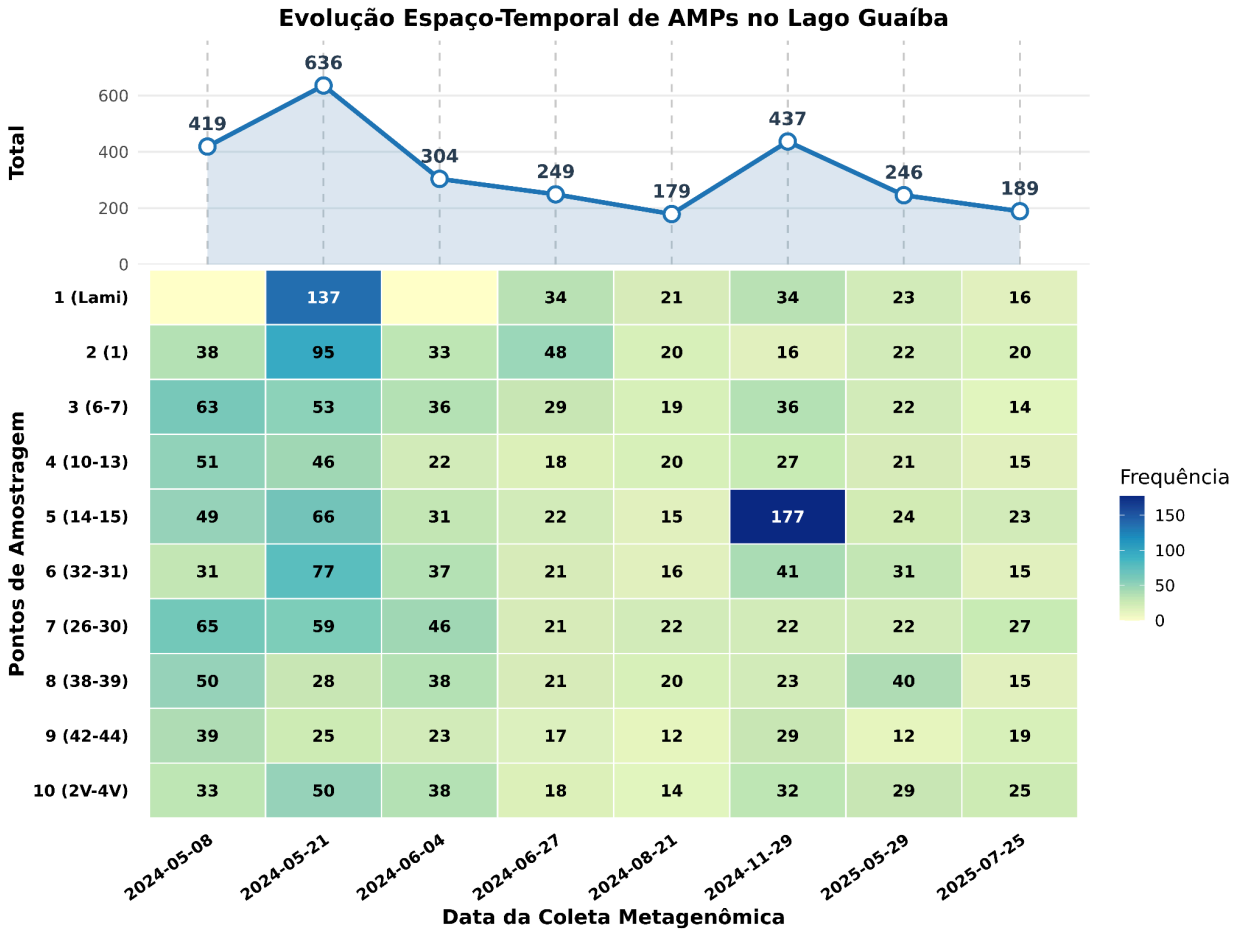


Figura 6. Dinâmica espaço-temporal e evolução das ocorrências de candidatos a peptídeos antimicrobianos (cAMPs) identificados no Lago Guaíba. O painel superior apresenta a evolução temporal da frequência absoluta acumulada de predições. O painel inferior detalha a distribuição dessa frequência nos dez pontos de amostragem georreferenciados ao longo das campanhas de coleta.

A amostra referente ao Ponto 5 (data 29/11/2024) apresentou um volume absoluto de candidatos superior aos demais pontos. Para assegurar a consistência técnica deste resultado, realizou-se a normalização da abundância de peptídeos pelo número total de *contigs* montados em cada amostra. A comparação entre a amostra F5 e a amostra de referência A5 demonstrou uma proporção de predição idêntica de 0,178 para ambos os casos. Este dado confirma que a maior contagem observada na amostra F5 é proporcional à densidade de sequências na montagem metagenômica, validando a representatividade estatística das amostras incluídas nesta análise exploratória.

A análise temporal revelou oscilações quantitativas expressivas na identificação de cAMPs no Lago Guaíba, sugerindo uma forte influência de fatores ambientais na frequência de detecção das sequências. O teste de Kruskal-Wallis confirmou uma diferença global altamente significativa na frequência de peptídeos entre as datas de coleta ($p < 0,001$). A avaliação

post-hoc de Dunn (p ajustado < 0,05) validou que as duas primeiras datas de amostragem, 08/05/2024 (n = 419 sequências) e 21/05/2024 (n = 636 sequências), correspondentes ao período de maior impacto da enchente, juntamente com a coleta de 29/11/2024 (n = 437 sequências totais), apresentaram abundâncias estatisticamente superiores às dos demais períodos de monitoramento. Em contraste, coletas realizadas em períodos de estabilidade hidrológica, como agosto de 2024 (n = 179) e julho de 2025 (n = 189), registraram as menores frequências de detecção do estudo.

5.8 Alinhamento com banco de dados de AMPs

O banco de dados DRAMP é composto por sequências com atividade antimicrobiana comprovada *in vitro* ou *in vivo*, além de peptídeos patenteados ou avaliados clinicamente. A análise das 179 sequências candidatas permitiu a classificação quanto à homologia estrutural, utilizando o limiar de 40% de identidade como critério para definição de sequências inéditas, conforme ilustrado na Figura 7. Esta análise mostrou que das 179 sequências, 106 (59,2%) apresentaram alinhamento superior ou igual a 40%, sendo classificadas como homólogas a peptídeos validados experimentalmente. As 73 sequências restantes (40,8%) não atingiram o limiar, sendo categorizadas como inéditas nesta base de dados. Este resultado indica que uma parcela dos candidatos identificados no ecossistema amostrado não possui contraparte direta em registros experimentais consolidados, sugerindo a diversidade de peptídeos em ambientes impactados por eventos climáticos extremos.

O confronto com o banco AMPSphere, que reúne um catálogo global de AMPs oriundos de estudos metagenômicos, revelou que 119 peptídeos (66,5%) apresentam identidade superior ou igual a 40%. Em contrapartida, 60 sequências (33,5%) não demonstraram homologia significativa com este catálogo global. A maior proporção de homologia observada no AMPSphere, comparada ao DRAMP, reflete a natureza metagenômica do conjunto de dados, uma vez que a maioria dos candidatos já foi identificada em outros levantamentos baseados em sequenciamento ambiental.

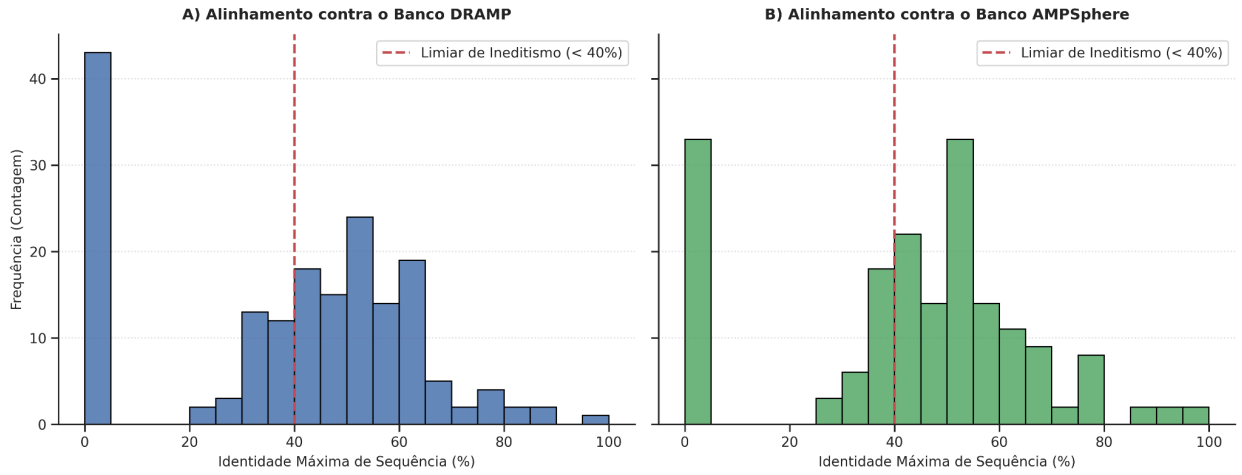


Figura 7. O gráfico apresenta o percentual de sequências homólogas (identidade ≥ 40%) e inéditas (identidade < 40%) identificadas a partir do confronto com os bancos de dados DRAMP e AMPSphere.

A análise de interseção de sequências inéditas demonstrou que 29 peptídeos (16,2%) não apresentaram homologia significativa com nenhum dos dois bancos utilizados (DRAMP e AMPSphere simultaneamente).

5.9 Associação dos AMPs com gêneros bacterianos

A aplicação do teste de Mann-Whitney U permitiu identificar um conjunto de associações entre peptídeos antimicrobianos e a supressão de gêneros bacterianos. Após a análise estatística inicial e a aplicação do filtro de elite no software Cytoscape, foram validadas 287 associações entre peptídeos e gêneros microbianos, como demonstrado na rede da Figura 8. Estes resultados demonstram que, em pares onde a detecção do peptídeo foi confirmada, a abundância dos gêneros bacterianos correspondentes apresentou uma redução estatisticamente significativa em comparação às amostras onde tais peptídeos não foram identificados. A rede de interações gerada a partir do filtro de significância ($p < 0,001$) evidencia uma estrutura complexa de associações, na qual a presença de determinados peptídeos coincide com a menor representatividade de táxons específicos no ecossistema do Lago Guaíba.

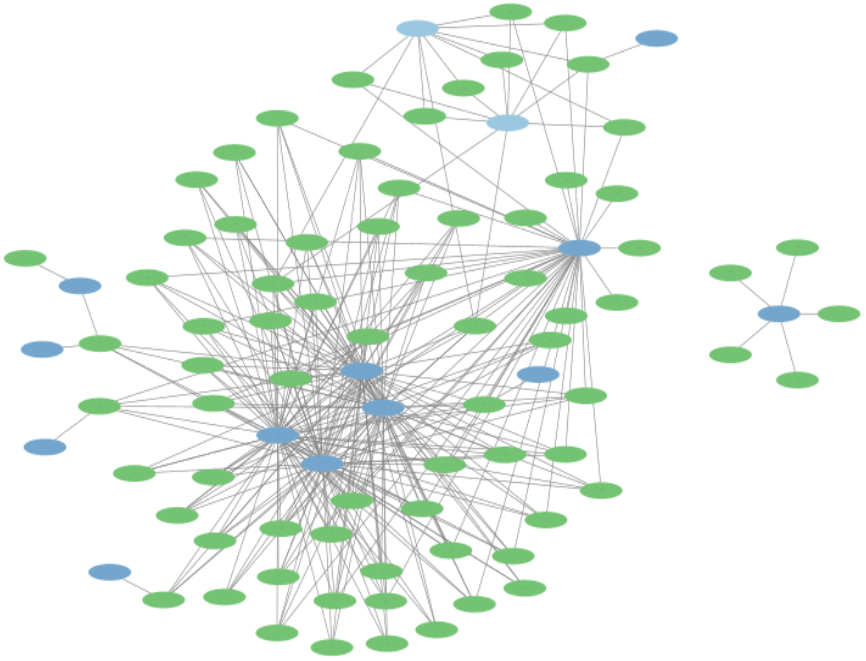


Figura 8. Rede de supressão de AMPs entre gêneros bacterianos. Os nós azuis representam peptídeos antimicrobianos candidatos (c_AMPs) e os nós verdes representam gêneros bacterianos suprimidos. As conexões (arestas) ilustram padrões de exclusão mútua estatisticamente significativos (P≤0,001) inferidos a partir de perfis de abundância metagenômica em nível de sistema.

5.10 Simulação Molecular

A seleção das 15 variantes peptídicas submetidas às simulações de dinâmica molecular foi fundamentada nos perfis de potencial biológico estimados pelas ferramentas de aprendizado profundo (Figura 9). Esses algoritmos preditivos avaliaram as sequências primárias e determinaram a probabilidade de cada estrutura atuar como um AMP, permitindo a triagem sistemática e a escolha dos sistemas que apresentaram os maiores índices de viabilidade teórica para a posterior caracterização de sua estabilidade conformacional.

;

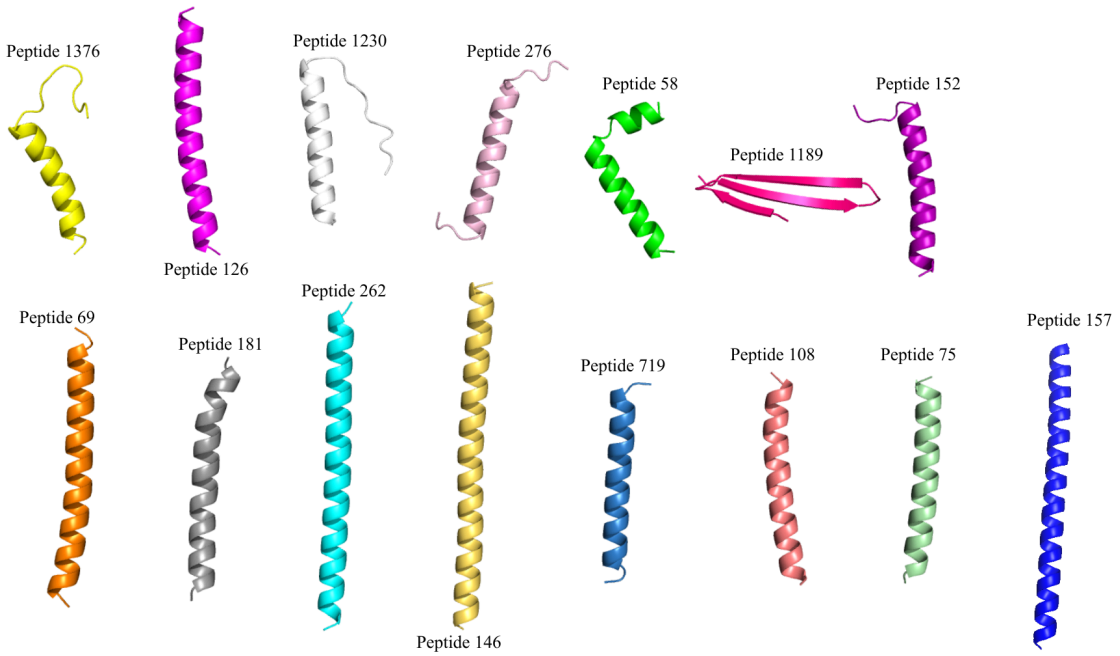


Figura 9. Representação tridimensional das 15 variantes dos c_AMPs selecionadas para a etapa de simulação por dinâmica molecular.

5.10.1 Desvio Quadrático Médio (RMSD)

O monitoramento temporal do RMSD dos átomos pesados do esqueleto peptídico é apresentado de forma comparativa na Figura 9. Na Figura 10A(1), correspondente às variantes Pep_108, Pep_1189, Pep_1230, Pep_126 e Pep_1376, registraram-se médias globais de trajetória situadas entre 3,88 ± 0,76 Å (Pep_1376) e 8,15 ± 2,73 Å (Pep_1230); nesse mesmo conjunto, durante os últimos 100 ns de simulação (fase de platô), as variantes Pep_108 e Pep_1230 atingiram os maiores desvios médios estáveis, com valores de 9,79 ± 1,40 Å e 10,05 ± 0,98 Å, respectivamente, exibindo desvios máximos pontuais de 14,20 Å e 12,02 Å, enquanto as variantes Pep_1189 (5,52 ± 0,29 Å), Pep_126 (4,70 ± 0,33 Å) e Pep_1376 (4,04 ± 0,78 Å) mantiveram desvios de platô contidos e perfis lineares com baixas amplitudes de flutuação.

Em paralelo, os dados dispostos na Figura 10A(2), constituídos pelas variantes Pep_146, Pep_152, Pep_157, Pep_181 e Pep_262, apresentaram médias gerais ao longo de toda a trajetória flutuando na faixa de 4,92 ± 0,83 Å (Pep_152) a 6,79 ± 1,89 Å (Pep_146), sendo que, na fase terminal de equilíbrio, a variante Pep_152 manteve o menor desvio estrutural (5,58 ± 0,80 Å), ao passo que as variantes Pep_181, Pep_262, Pep_146 e Pep_157 apresentaram valores de platô progressivamente maiores, estabelecidos em 7,19 ± 2,62 Å, 7,65 ± 0,75 Å, 7,74 ± 1,51

Å e 7,87 ± 1,38 Å, respectivamente, com o desvio máximo absoluto desse perfil registrado para a Pep_146, alcançando 13,29 Å.

Por fim, as trajetórias avaliadas na Figura 10A(3), formadas pelas variantes Pep_276, Pep_58, Pep_69, Pep_719 e Pep_75, evidenciaram o maior desvio estrutural absoluto de toda a amostragem na variante Pep_276, que registrou média de trajetória de 7,63 ± 3,11 Å, média de platô final de 12,96 ± 1,20 Å e um pico máximo de afastamento geométrico de 15,54 Å, em contraposição às quatro variantes restantes do gráfico, que se mantiveram em patamares estáveis e inferiores de RMSD na janela final de 100 ns, apresentando médias de platô de 4,57 ± 0,87 Å (Pep_75), 4,66 ± 0,39 Å (Pep_719), 5,54 ± 1,55 Å (Pep_69) e 5,57 ± 0,75 Å (Pep_58).

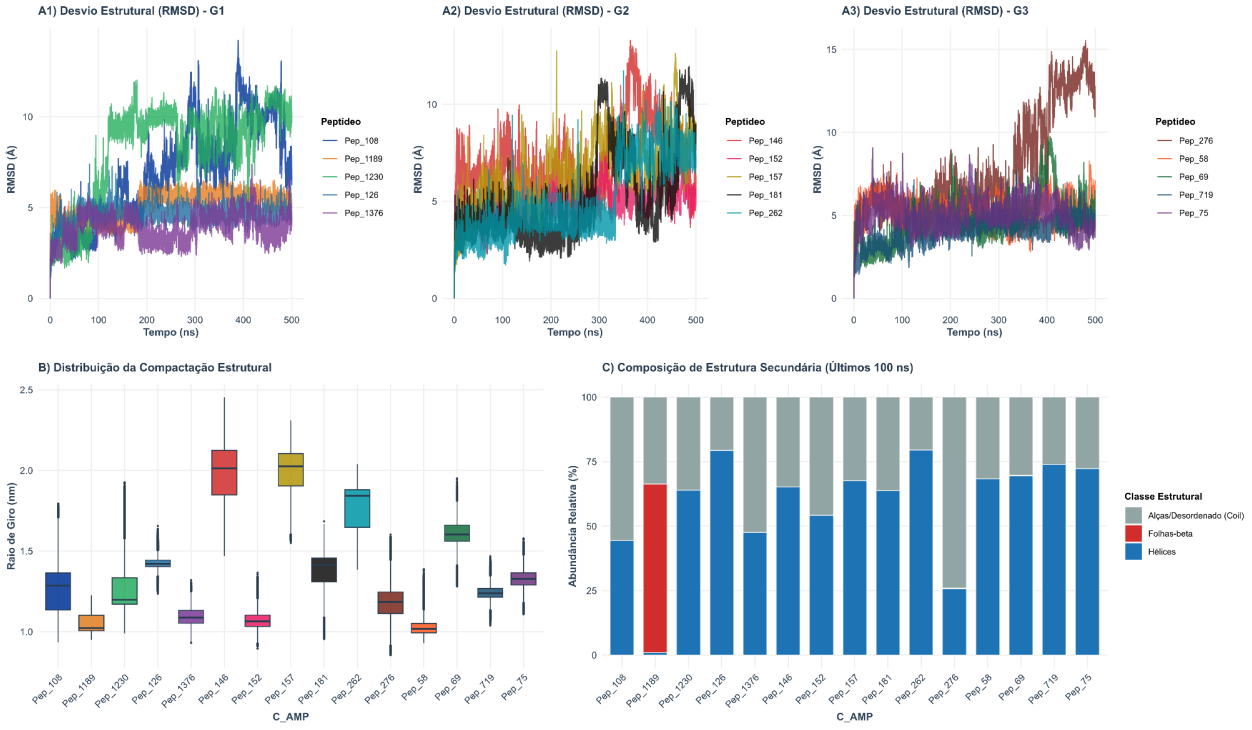


Figura 10. Métricas comparativas de estabilidade estrutural e conformacional das 15 variantes peptídicas obtidas por dinâmica molecular (500 ns). *(A1, A2, A3)* Trajetórias temporais do RMSD (Å) do esqueleto peptídico, divididas em três subgrupos de cinco variantes para otimização da resolução visual. *(B)* Distribuição do Raio de Giro (nm) de cada variante ao longo de toda a simulação, representada em diagramas de caixas (*boxplots*). *(C)* Abundância relativa percentual (%) das classes de estrutura secundária (Hélices em azul, Folhas-beta em vermelho e Alças/Desordenado em cinza) obtidas via DSSP, calculada sobre os últimos 100 ns de simulação.

5.10.2 Raio de Giro (RG)

A distribuição dos valores do raio de giro ao longo da coordenada temporal foi mapeada comparativamente em formato de diagramas de caixas para todos os sistemas (Figura 10B). Os

menores valores médios de raio de giro, característicos de estruturas com alta compactação molecular, foram identificados nas variantes Pep_58 (1,025 ± 0,042 nm), Pep_1189 (1,049 ± 0,056 nm), Pep_152 (1,071 ± 0,063 nm) e Pep_1376 (1,092 ± 0,050 nm). Nessas variantes, os limites mínimos absolutos detectados durante a dinâmica atingiram marcas inferiores a 1,0 nm, com destaque para a variante Pep_152, que registrou o mínimo absoluto de 0,896 nm.

Perfis lineares de menor compactação e com maiores valores médios de raio de giro foram observados nas variantes Pep_157 (2,000 ± 0,130 nm), Pep_146 (1,988 ± 0,188 nm) e Pep_262 (1,784 ± 0,128 nm). As amplitudes máximas registradas alcançaram 2,454 nm na variante Pep_146 e 2,310 nm na variante Pep_157. O restante das variantes peptídicas demonstrou distribuições intermediárias, situadas na faixa de 1,170 nm (Pep_276) a 1,602 nm (Pep_69).

5.10.3 Flutuação Quadrática Média por Resíduo (RMSF)

Os perfis de flexibilidade local revelaram flutuações médias globais e picos locais distintos entre as estruturas primárias. Os menores índices de flutuação média global foram calculados para as variantes Pep_126 (2,22 Å), Pep_1376 (2,37 Å), Pep_1189 (2,47 Å), Pep_58 (2,49 Å) e Pep_719 (2,53 Å). Em contrapartida, os maiores valores de flutuação média ao longo da cadeia ocorreram nas variantes Pep_276 (5,98 Å) e Pep_108 (5,62 Å).

No que tange aos picos discretos de máxima flexibilidade molecular, os maiores valores absolutos foram localizados na região N-terminal da variante Pep_146 (14,64 Å no resíduo 1), seguidos por regiões de alças e terminais nas variantes Pep_1230 (13,86 Å no resíduo 33) e Pep_108 (12,50 Å no resíduo 29). Os picos de flutuação máxima mais contidos foram identificados nas variantes Pep_58 (5,50 Å no resíduo 1) e Pep_126 (5,86 Å no resíduo 33).

5.10.4 Estabilidade do perfil de Estrutura Secundária (DSSP)

A análise quantitativa da composição estrutural acumulada na janela terminal de equilíbrio (últimos 100 ns) revelou distribuições heterogêneas entre as variantes (Figura 10C). A classe conformacional correspondente a hélices (α-hélices, hélices 310 e hélices π) apresentou dominância na maioria dos sistemas estudados. Os maiores percentuais de helicidade estável foram identificados nas variantes Pep_262 (79,6%), Pep_126 (79,4%), Pep_719 (73,9%),

Pep_75 (72,3%), Pep_69 (69,5%) e Pep_58 (68,3%). A formação e manutenção de estruturas organizadas em folhas-β e pontes-β ocorreu de forma quase exclusiva na variante Pep_1189, registrando 65,3% de abundância relativa. Sinais residuais dessa classe estrutural foram detectados apenas nas variantes Pep_276 (0,3%) e Pep_157 (0,0% nominal, com flutuações nulas), enquanto as 13 variantes restantes apresentaram 0,0% de conteúdo de folhas-β no platô final. A fração correspondente a alças, voltas e novelos aleatórios (*coils, turns* e *bends*) representou a classe majoritária nas variantes Pep_276 (74,0%), Pep_108 (55,6%) e Pep_1376 (52,4%). Nas variantes caracterizadas por alto conteúdo helicoidal, como Pep_262 e Pep_126, a presença de alças reduziu-se aos menores patamares registrados na amostragem, estabelecendo-se em 20,4% e 20,6%, respectivamente.

5.11 Atribuição taxonômica do metagenoma

Por fim, foi realizada a caracterização taxonômica dos metagenomas totais dos quais os c_AMPs foram projetados, conforme resultados abaixo.

5.11.1 Composição em nível de filo

Realizaram-se a identificação e a quantificação taxonômica de 78 amostras metagenômicas do Lago Guaíba nos perfis biológicos em níveis de filo e gênero. Os dados foram estruturados de forma contínua e cronológica para cada uma das dez estações fixas de monitoramento (Figura 11). No escopo do metagenoma global, a estrutura de filos foi amplamente dominada por dois grupos principais, que juntos acumularam mais de 88% da abundância relativa total do ambiente: *Pseudomonadota* (50,51%) e *Actinomycetota* (38,42%). Os demais componentes principais registraram ocorrências individuais inferiores a 3%, distribuídas de forma decrescente entre *Bacillota* (2,63%), *Bacteroidota* (2,47%), *Planctomycetota* (1,15%), *Myxococcota* (*0,94%*), *Chordata* (0,92%), *Cyanobacteriota* (0,56%) e *Methanobacteriota* (0,48%). A fração agregada de táxons minoritários ("Outros") correspondeu a apenas 0,03% do perfil acumulado.

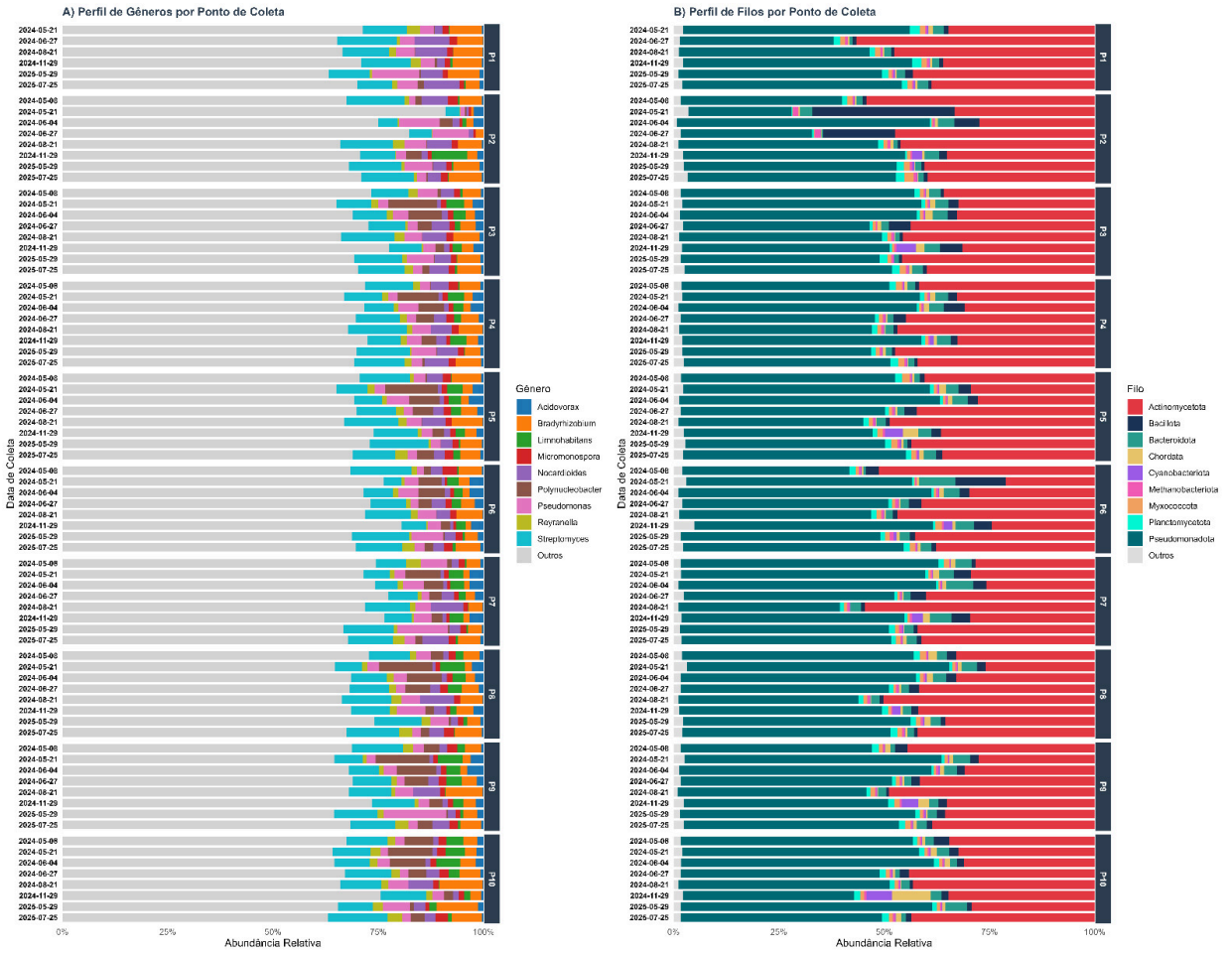


Figura 11. Perfil taxonômico horizontal do microbioma do Lago Guaíba estruturado por estação de monitoramento (maio/2024 a julho/2025). (A) Dinâmica temporal da abundância relativa emno nível de Gênero. (B) Dinâmica temporal da abundância relativa emno nível de Filo. O eixo vertical (Y) indica a evolução cronológica das campanhas amostrais dentro de cada ponto de monitoramento (P1 a P10), compreendendo as 78 amostras individuais analisadas. As colorações distinguem os nove táxons mais abundantes em cada nível e a categoria "Outros".

Ao estratificar a ocorrência dos filos entre os dez pontos amostrais (Figura 11B), observa-se que *Pseudomonadota* manteve-se estável acima de 44% em todos os pontos, exibindo sua taxa máxima de representatividade no ponto P7 (52,62%) e a menor proporção no ponto P2 (44,27%). De maneira inversa, o filo *Actinomycetota* apresentou sua maior abundância relativa no ponto P1 (42,81%), decrescendo gradualmente até atingir sua menor quota territorial no ponto P7 (36,38%).

Variações espaciais foram documentadas para os táxons de abundância intermediária. O filo *Bacillota* registrou um incremento acentuado no ponto P2, alcançando 7,73% de abundância relativa, valor substancialmente superior à média observada nos demais pontos, onde oscilou entre 1,05% (P1) e 3,50% (P6). Adicionalmente, verificou-se que o filo *Bacteroidota* obteve seu

maior percentual no ponto P7 (3,46%), enquanto *Chordata* atingiu seu pico de detecção nas amostras coletadas no ponto P10, contabilizando 1,94% de abundância relativa.

5.11.2 Composição em nível de gênero

A análise detalhada em nível de gênero revelou que os nove táxons mais abundantes responderam pela fração majoritária da microbiota identificada (Figura 11A). O gênero *Streptomyces* despontou como o grupo mais representativo do metagenoma total, apresentando uma abundância média global de 9,73%. Em patamares intermediários (entre 2% e 5%), situaram-se os gêneros *Bradyrhizobium* (4,20%), *Pseudomonas* (3,89%), *Polynucleobacter* (3,17%) e *Nocardioides* (2,97%). Os gêneros *Limnohabitans* (1,66%), *Reyranella* (1,49%), *Micromonospora* (1,43%) e *Acidovorax* (1,36%) completaram o ranking dos principais componentes biológicos, deixando 0,02% para a fração residual ("Outros").

A distribuição individual desses gêneros revelou assimetrias entre os as dez pontos estações de monitoramento (Figura 11A). O gênero dominante *Streptomyces* variou de 8,17% no ponto P7 até um pico de 10,92% no ponto P1. O gênero *Bradyrhizobium* registrou sua maior representatividade no ponto P1 (6,43%) e no ponto P10 (5,60%), apresentando sua menor densidade no ponto P7 (2,76%). Padrões divergentes de ocorrência foram identificados para os gêneros *Pseudomonas* e *Nocardioides*, que exibiram simultaneamente suas maiores abundâncias relativas nas amostras pertencentes ao ponto P1, com 5,12% e 5,52%, respectivamente. Em contrapartida, os menores valores quantificados para *Pseudomonas* ocorreram no ponto P5 (2,97%) e para *Nocardioides* no ponto P6 (2,14%). Por fim, variações lineares foram observadas para os gêneros *Polynucleobacter* e *Limnohabitans*. O gênero *Polynucleobacter* apresentou escassa detecção no ponto P1 (0,46%), contudo, mostrou um aumento em direção aos pontos P9 (4,81%) e P10 (4,59%). Comportamento similar foi verificado para o gênero *Limnohabitans*, cuja abundância relativa inicial de 0,21% no ponto P1 expandiu-se continuamente ao longo das estações até atingir o ápice de ocorrência no ponto P10, com 2,61% da composição local.

5.11.3 Avaliação da diversidade taxonômica dos metagenomas

A análise da diversidade alfa (Índice de Shannon) demonstrou flutuações temporais significativas na riqueza e equitabilidade global ao longo das campanhas de monitoramento (Kruskal-Wallis) (Figura 12A). Contrariando a premissa de estabilidade, observou-se que durante a fase mais aguda das inundações (maio e junho de 2024), a diversidade manteve-se em

platôs elevados (médias entre 4,99 e 5,08). No entanto, ocorreu uma queda abrupta da diversidade na amostragem de 21 de agosto de 2024, que registrou a menor média temporal (4,83). O teste de comparações par a par de Dunn (correção FDR) confirmou que a estrutura comunitária em agosto de 2024 foi estatisticamente inferior àquelas observadas em 08 de maio (p = 0,003), 27 de junho (p = 0,003) e 29 de novembro de 2024 (p = 0,006). Posteriormente, a microbiota iniciou um processo de restabelecimento gradual, atingindo seu ápice de diversidade em julho de 2025 (média de 5,14).

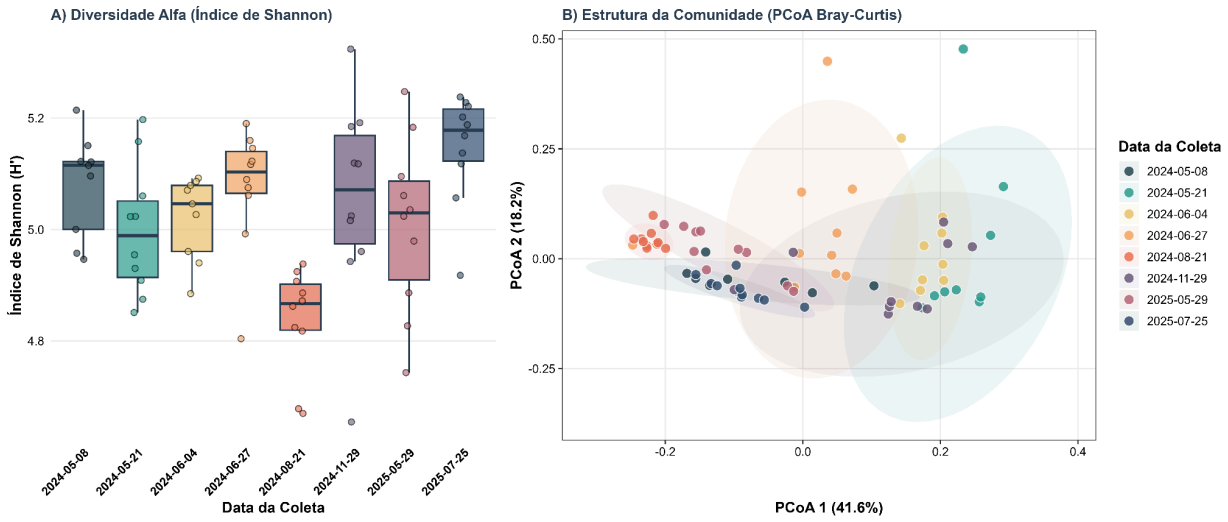


Figura 12. Dinâmica temporal da diversidade taxonômica do microbioma global do Lago Guaíba (maio/2024 a julho/2025). **(A)** Diversidade alfa mensurada pelo Índice de Shannon (H′). A distribuição dos dados é apresentada em *boxplots* interpostos aos pontos amostrais individuais (*jitter*). Observaram-se flutuações temporais significativas (Teste de Kruskal-Wallis, p = 0,00015), com destaque para a queda em agosto de 2024. **(B)** Ordenação da diversidade beta via Análise de Coordenadas Principais (PCoA) baseada na matriz de distâncias de Bray-Curtis. As elipses sombreadas delimitam o agrupamento das amostras por campanha. A cronologia das coletas explicou quase 50% da variação composicional da comunidade (PERMANOVA, p = 0,001; R^2 = 0,498).

Com relação à diversidade beta (matriz de dissimilaridade de Bray-Curtis), os resultados revelaram alterações na estrutura global do microbioma ao longo do tempo (PERMANOVA, p = 0,001) (Figura 12B). O fator temporal cronológico despontou como o principal modulador da composição microbiana, sendo capaz de explicar aproximadamente 49,8% de toda a variação biológica do ecossistema do Lago Guaíba (R^2 = 0,4979). A robustez deste achado foi ratificada pelo teste PERMDISP (p = 0,183), que demonstrou a homogeneidade da dispersão multivariada intragrupo, confirmando que as variações apontadas representam autênticos deslocamentos ecológicos estruturais (migração de centróides) em detrimento de ruídos amostrais ou artefatos de dispersão.

6. DISCUSSÃO

A resistência aos antimicrobianos tornou-se uma das ameaças à saúde pública contemporânea, com projeções que apontam os patógenos multirresistentes, principalmente os do grupo ESKAPE (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* e *Enterobacter* spp.) (Miller e Arias 2024), como potenciais líderes em estatísticas de mortalidade global até o ano de 2050 (Cordoves-Delgado e García-Jacas, 2024; Ho et al., 2025). Diante do declínio na descoberta de novos fármacos, os AMPs surgem como alternativas terapêuticas promissoras devido a sua ação de amplo espectro e à menor probabilidade de induzir resistência (Wang et al. 2025; Ma et al. 2022).

Historicamente, a identificação dessas moléculas esbarrava na dificuldade de cultivar a maioria dos microrganismos em laboratório (Du et al., 2025; Xu et al., 2024). Contudo, a prospecção de novas substâncias tem sido inspirada pela mineração de dados metagenômicos em larga escala associada ao aprendizado de máquina (Ma et al., 2022; Wei et al., 2025). Estudos recentes demonstraram a eficácia dessa abordagem computacional ao validar *in vitro* centenas de peptídeos ativos derivados de microbiomas diversos, que vão do intestino humano até ambientes desafiadores como lodos de esgoto, e extraindo milhões de sequências candidatas a partir do material genético de microrganismos não cultiváveis (Xu et al., 2024; Li et al., 2025). Inspirado por esses avanços, o presente estudo buscou preencher a lacuna ao investigar o potencial biotecnológico de um ecossistema aquático dinâmico sob intenso estresse físico-químico.

Feita a observação da dinâmica da comunidade microbiana e da frequência de detecção de AMPs no Lago do Guaíba, percebe-se que o impacto de eventos climáticos extremos sobre a microbiota pode não ter ocorrido de forma linear ou imediata. Os dados estatísticos observados tanto na estruturação da comunidade (PERMANOVA, $p = 0{,}001$) quanto nas flutuações da diversidade alfa (Kruskal-Wallis, $p = 0{,}00015$) descrevem a cronologia das coletas como o principal fator direcionador do ecossistema, sendo capaz de explicar 49,8% ($R^2 = 0{,}4979$) da variância biológica total. Esse comportamento temporal sugere a possibilidade de que a resposta biológica da comunidade não tenha acompanhado imediatamente o pico da cheia. É possível que o ecossistema tenha mantido uma diversidade aparente no início para, semanas após a baixa das águas, sofrer uma redução súbita de espécies.

Durante a fase aguda das inundações (maio e início de junho de 2024), a manutenção de índices elevados de diversidade alfa (médias de Shannon entre 4,99 e 5,08) correlaciona-se com

os maiores volumes absolutos de c_AMPs minerados (419 sequências em 08/05 e 636 em 21/05). Do ponto de vista ecológico, hipotetiza-se que esse cenário não reflita uma estabilidade homeostática do ecossistema, mas sim um reflexo do efeito de lavagem do solo e lixiviação superficial causados pelo escoamento pluviométrico extremo (Perryman et al. 2011). Sob essa perspectiva, a introdução de matéria orgânica, sedimentos suspensos e microrganismos derivados do solo e dos efluentes urbanos podem ter gerado uma sobreposição temporária de nichos ecológicos (Rillig and Mansour 2017). Essa mistura física entre comunidades terrestres e aquáticas poderia explicar a sobreposição temporária de nichos ecológicos, resultando na riqueza taxonômica elevada (Macias et al. 2026), o que justificaria a ampla amostragem funcional de c_AMPs detectada neste período.

Em contrapartida, em agosto de 2024, o Índice de Shannon atingiu o valor mínimo de 4,83, validado pelo teste de Dunn como estatisticamente inferior ao início das cheias, coincidindo com a redução para 179 sequências de c_AMPs. Esse declínio sugere uma fase de transição ecológica, potencialmente induzida pela combinação de fatores ambientais, como a possibilidade de exaustão de nutrientes, sedimentação da matéria que estava suspensa, estresse térmico sazonal e redução da radiação solar, típico do inverno em agosto, que podem ter atuado de forma conjunta para influenciar na dinâmica da estrutura da comunidade microbiana (Liu et al. 2020; Shade et al. 2012). Essa combinação de fatores poderia explicar a redução observada tanto na diversidade alfa global quanto na detecção dos c_AMPs.

A reversão desse cenário iniciou-se em novembro de 2024 com um aumento do índice de Shannon para 5,05 e a elevação para 437 c_AMPs. Esse período coincidiu com a transição térmica para a primavera e o verão. Essa mudança sazonal, associada à estabilização do regime hidrológico, pode ter permitido a reorganização das interações comunitárias e o restabelecimento da complexidade metabólica do lago (Zhao et al. 2020; Xu et al. 2024). Nesse sentido, esse processo pode ter influenciado o pico absoluto de diversidade em julho de 2025 (Shannon de 5,14), indicando restabelecimento da comunidade após o evento climático extremo (Shabarova et al. 2021; Duan et al. 2025).

A dominância global observada dos filos *Pseudomonadota* (50,51%) e *Actinomycetota* (38,42%), associada à proeminência do gênero *Streptomyces* (9,73%), sugere a existência de um núcleo microbiano central amplamente adaptado às características hidrológicas do Lago Guaíba. É provável que, frente às variações climáticas, a comunidade tenha tido uma versatilidade metabólica para suportar as perturbações no ecossistema (Comte and del Giorgio 2009; Comte et al. 2013). Neste contexto, o gênero *Streptomyces*, amplamente reconhecido por sua capacidade de formação de esporos estáveis e pela degradação de compostos orgânicos complexos (McCormick e Flärdh 2012; Chater et al. 2016), destaca-se por permanecer abundante mesmo

sob condições de estresse ambiental. Apesar da distribuição contínua dessa estrutura taxonômica principal, assimetrias pontuais nos dados sugerem a existência de microambientes específicos entre os pontos de amostragem. Nesse cenário, o incremento acentuado do filo *Bacillota* no ponto P2, onde atingiu 7,73% de abundância relativa em contraste com as baixas médias dos demais pontos, levanta a hipótese de uma influência ecológica de caráter estritamente local. Sugere-se que o ponto P2 atue como uma zona de impacto antrópico imediato ou de deposição sazonal, uma vez que o filo *Bacillota* é um marcador clássico de aportes de efluentes urbanos (Cheung et al. 2025; Ibekwe et al. 2016). É possível que esse comportamento possa estar associado a aportes de drenagem particulares daquela região ou a condições hidrodinâmicas sazonais que teriam favorecido a deposição de sedimentos ricos em matéria orgânica (Wang et al. 2022; Fermani et al. 2022; Smits et al. 2023).

A heterogeneidade espacial se acentua ao analisar o gradiente de transição taxonômica entre o ponto P1 e os pontos mais distantes P9 e P10. A concentração de gêneros como *Pseudomonas* (5,12%), *Nocardioides* (5,52%) e *Bradyrhizobium* (6,43%) no ponto P1 pode indicar uma área sob maior influência de aportes fluviais e carreamento de solos, fatores frequentemente associados à introdução de táxons não nativos (Ibekwe et al. 2016; Shabarova et al. 2021; Newton et al. 2011). Por outro lado, o aumento progressivo de *Polynucleobacter* e *Limnohabitans* em direção aos pontos P9 (4,81% e 2,31%) e P10 (4,59% e 2,61%) sugere uma transição para dinâmicas tipicamente lacustres (Hahn et al. 2010; Jezbera et al. 2011). Esse cenário sugere que o distanciamento das fontes de perturbação e o tempo de residência hídrica reduzem o sinal de microrganismos terrestres, viabilizando o desenvolvimento desses táxons especializados de água doce.

Essa organização espacial e temporal da microbiota pode se conectar de maneira estreita com a distribuição dos peptídeos antimicrobianos inferidos no sistema. A identificação de 287 associações estatisticamente significativas pelo teste de Mann-Whitney U ($p<0,001$) sugere que a presença de determinados c_AMPs correlaciona-se inversamente com a abundância de gêneros bacterianos específicos na rede de interações. A detecção simultânea de um peptídeo e a redução concomitante de um táxon-alvo abrem espaço para a hipótese de que essas moléculas atuem como moduladores da estrutura comunitária local, exercendo pressões seletivas discretas sobre a sobrevivência de grupos específicos durante as oscilações ambientais do lago (Riley and Wertz 2002) (Granato et al. 2019; Wang et al. 2025). Embora a validação de uma causalidade direta dependa de ensaios experimentais complementares, os padrões de conectividade observados na rede revelam que a ação inibitória teórica não é isolada, mas compartilhada entre diferentes candidatos e táxons bacterianos.

Para identificar esse cenário biológico a partir de dados metagenômicos brutos, foi indispensável adotar uma estratégia de triagem baseada em diferentes modelos computacionais. A escolha por evitar *softwares* tradicionais serviu para garantir a reprodutibilidade dos dados (seguindo os princípios FAIR) e reduzir erros que pudessem comprometer o ineditismo das moléculas. Assim, unir a alta capacidade de detecção do ProteoGPT com o filtro de consenso do c_AMPs-prediction permitiu uma varredura ampla das sequências lineares, equilibrando a descoberta de novos candidatos com o descarte precoce de falsos positivos.

Apesar da eficiência dessa etapa, a triagem baseada apenas em sequências unidimensionais ainda apresentava limitações. Diante desse cenário, tornou-se necessário o refinamento por meio do modelo em grafos esm-AxP-GDL. Esse filtro tridimensional representou um afunilamento, retendo apenas 36,6% dos 1.943 peptídeos iniciais. Essa redução expressiva indica que a conformação espacial pode atuar como um gargalo seletivo para a viabilidade de um c_AMP. Hipotetiza-se que a exclusão da maior parte das sequências decorra da incapacidade teórica dessas moléculas de adotar o dobramento anfipático necessário para interagir e romper membranas bacterianas (Cordoves-Delgado e García-Jacas, 2024). Dessa forma, a incorporação de coordenadas estruturais preditas mostrou-se importante para aumentar a precisão da triagem, garantindo que o conjunto final de 711 peptídeos concentre apenas os candidatos com maior probabilidade de estabilidade biofísica. Posteriormente, esse conjunto de 711 resultou na seleção final de 179 peptídeos classificados como não hemolíticos e não tóxicos, através das ferramentas HEMOPI 2.0 e ToxinPred3.

A triagem final que resultou no conjunto de 179 candidatos classificados como não hemolíticos e não tóxicos revelou um perfil molecular com forte predisposição ao caráter catiônico. O elevado ponto isoelétrico médio (pI = 9,64), somado ao fato de mais de 85% das sequências serem básicas, alinha-se às propriedades clássicas de peptídeos antimicrobianos funcionais (Zasloff 2002). Essa característica eletrostática deve-se ao enriquecimento de lisina (K), responsável por 20% da composição da biblioteca. Sob uma perspectiva biofísica, é plausível que essa forte carga positiva funcione como o principal motor para a aproximação e a atração eletrostática entre o peptídeo e a membrana bacteriana negativa (Brogden 2005; Cresca et al. 2026).

A forte relação entre a carga líquida e o pI é confirmada estatisticamente pelo coeficiente de Pearson, que atingiu $r = 0,7665$ na biblioteca global. Além da carga, o comportamento dessas moléculas na água do lago é explicado pelo equilíbrio entre o índice GRAVY médio (–0,57) e o alto índice de Boman (3,38). Enquanto o valor negativo de GRAVY indica que os compostos são hidrofílicos e solúveis, o escore elevado de Boman aponta uma alta capacidade de interação molecular (Boman, 2003). Sugere-se que esse balanço impeça os peptídeos de sofrerem uma

agregação precoce no ambiente hídrico, mantendo-os livres e ativos para interagir com as membranas-alvo (Dathe and Wieprecht 1999).

Essas características físico-químicas ganham relevância prática quando comparadas aos bancos de dados de referência. O alinhamento contra o banco de dados DRAMP revelou que 40,8% das sequências obtidas não atingiram o limite de 40% de identidade com AMPs já validados experimentalmente ou patenteados. Essa ausência de homologia direta sugere que uma parte expressiva da biblioteca minerada é composta por sequências novas. Diante disso, a identificação dessas sequências inéditas demonstra que o monitoramento de um ambiente gravemente impactado permitiu mapear uma diversidade bioquímica singular, revelando compostos que diferem daqueles catalogados nos bancos de dados tradicionais. Essa percepção ganha ainda mais força quando os dados são comparados ao catálogo metagenômico AMPSphere, no qual 33,5% dos peptídeos avaliados ficaram abaixo do limiar de similaridade. A ausência de correspondentes para essas moléculas, mesmo em inventários globais de sequenciamento ambiental, indica que o ecossistema do Lago Guaíba abriga peptídeos funcionais ainda não mapeados. Essa baixa homologia reforça o valor inovador do trabalho, demonstrando que o *pipeline* aplicado foi capaz de superar a redundância dos bancos de dados públicos e capturar candidatos originais com real potencial biotecnológico.

Para avaliar o potencial biotecnológico dos peptídeos preditos, avaliou-se como os candidatos se comportavam estruturalmente por meio de simulações de dinâmica molecular. O monitoramento do Desvio Quadrático Médio (RMSD) revelou perfis de estabilização bastante distintos entre as quinze variantes estudadas. Supõe-se que as flutuações e os altos valores médios de RMSD observados na fase de equilíbrio em sistemas como a variante Pep_276 (12,96 ± 1,20 Å) e Pep_108 (9,79 ± 1,40 Å) reflitam a alta flexibilidade intrínseca dessas cadeias lineares quando expostas ao solvente (Ulmschneider and Ulmschneider 2018). Sem uma interface de ancoragem, essas moléculas tendem a explorar uma ampla amostragem de conformações, o que justifica os altos valores de raio de giro e RMSF (Ulmschneider and Ulmschneider 2018;Cresca et al. 2026) .

Por outro lado, as variantes Pep_126, Pep_262 e Pep_719 mantiveram seus valores de RMSD baixos e estáveis, sugerindo que alcançaram uma estabilização. Esse comportamento está diretamente ligado aos dados do DSSP, que mostraram o predomínio de estruturas em alfa-hélice nesses sistemas, especialmente na Pep_262 (79,6%) e Pep_126 (79,4%). Hipotetiza-se que, ao interagirem com a membrana lipídica, essas variantes sofram a influência de interações hidrofóbicas na interface que podem também induzir o dobramento do peptídeo, promovendo a organização dessas estruturas e reduzindo a presença de regiões desorganizadas (Melo et al. 2009; Cresca et al. 2026).

Essa mudança na estrutura também se reflete na compactação das moléculas, medida pelo raio de giro (RG). Enquanto os candidatos em alfa-hélice mostraram um empacotamento estável, a estratégia computacional também foi capaz de identificar outras formas de estabilização. É o caso da variante Pep_1189, que adotou quase exclusivamente a conformação em folhas beta (65,3%), resultando em um dos menores raios de giro do grupo (mínimo de 0,896 nm). No geral, esses resultados sugerem que as sequências selecionadas possuem potencial para adotar conformações estáveis e organizadas, o que funcionaria como um requisito essencial para que possam interagir com as membranas lipídicas e, potencialmente, rompê-las (Melo et al. 2009; Cresca et al. 2026).

Apesar dos indícios estruturais favoráveis e do rigor computacional aplicado, a interpretação desses resultados deve considerar as restrições de uma abordagem estritamente preditiva. Primeiramente, os algoritmos de aprendizado de máquina dependem de bancos de dados pré-existentes, o que pode gerar um viés de treinamento e limitar a identificação de mecanismos de ação totalmente novos que fujam dos padrões já descritos (Lin et al. 2025). Além disso, as simulações de dinâmica molecular foram realizadas estritamente em meio aquoso, sem a presença de uma membrana lipídica. Diante da complexidade do Lago Guaíba ou de um alvo biológico real, a ausência dessa interface celular e de fatores como a força iônica, oscilações térmicas e barreiras adicionais, a exemplo dos lipopolissacarídeos em bactérias Gram-negativas, restringe a avaliação de como esses peptídeos de fato interagiriam com uma célula (Cresca et al. 2026). Por fim, as janelas de tempo na escala de nanossegundos capturam bem o dobramento inicial e a estabilização térmica no solvente, mas não conseguem refletir processos macroscópicos de longo prazo, como a aproximação, a ancoragem e a lise celular completa.

Para superar essas restrições e validar o real potencial biotecnológico mapeado, os desdobramentos deste trabalho devem avançar de forma gradual para a bancada experimental. O primeiro passo prático consiste em selecionar um subconjunto reduzido de candidatos de destaque. Esses peptídeos selecionados serão submetidos a testes *in silico* e a ensaios *in vitro* de concentração inibitória mínima para verificar a atividade antimicrobiana contra bactérias-alvo do lago e de interesse clínico. Por fim, ensaios iniciais de hemólise e citotoxicidade serão conduzidos para avaliar se o perfil de segurança previsto pelos modelos computacionais se confirma em testes celulares básicos, servindo como um primeiro filtro para futuros estudos farmacêuticos.

7. CONCLUSÕES

O presente estudo demonstrou que houve flutuações de diversidade microbiana e assimetrias espaciais ao longo dos pontos de coleta. Esse cenário de forte estresse ecológico e perturbação ambiental revelou-se um ambiente promissor para a prospecção de novas biomoléculas. Através de um *pipeline* computacional multimodal e rigoroso, foi possível superar a redundância das bases de dados públicas e identificar 179 c_AMPs inéditos, não tóxicos e não hemolíticos, demonstrando o alto valor de inovação biotecnológica do projeto.

Do ponto de vista molecular, os candidatos prospectados exibiram características físico-químicas altamente favoráveis para aplicação biotecnológica, caracterizados por um perfil predominantemente catiônico, impulsionado pelo enriquecimento de resíduos de lisina, e um equilíbrio hidrofílico ideal para estabilidade em meio aquoso. Além disso, as simulações de dinâmica molecular em solvente indicaram o potencial dessas sequências para adotar dobramentos estruturais organizados, como alfa-hélices e folhas-beta, cumprindo requisitos biofísicos importantes para a futura interação com membranas e proteínas-alvo.

Por fim, embora a abordagem seja estritamente *in silico* e possua limitações preditivas naturais, este trabalho cumpre o papel de mapear o potencial biotecnológico oculto do ecossistema em um momento de perturbação hídrica. As sequências encontradas despontam como novos candidatos para substituir ou complementar os antibióticos tradicionais, que enfrentam uma perda global de eficácia. Como desdobramento prático, aponta-se para a seleção de um grupo restrito dessas moléculas para síntese química e ensaios *in vitro* de concentração inibitória mínima. Esse passo é fundamental para validar o fluxo computacional utilizado e confirmar se esses compostos podem auxiliar, no futuro, no combate à resistência aos antimicrobianos.

REFERÊNCIAS

GBD 2021 ANTIMICROBIAL RESISTANCE COLLABORATORS. Global burden of bacterial antimicrobial resistance 1990-2021: a systematic analysis with forecasts to 2050. The Lancet, [*s. l.*], vol. 404, no. 10459, p. 1199–1226, 2024. Disponível em: https://doi.org/10.1016/S0140-6736(24)01867-1

HO, Charlotte S *et al.* Antimicrobial resistance: a concise update. **The Lancet. Microbe**, [*s. l.*], vol. 6, no. 1, p. 100947, 2025. Available at: https://doi.org/10.1016/j.lanmic.2024.07.010

WILLIAMS, Sarah C. P. **The Antimicrobial Resistance Pandemic: Breaking the Silence**. American Society for Microbiology, 2024. Disponível em: https://asm.org/articles/2024/october/antimicrobial-resistance-pandemic-breaking-silence. Acesso em: 8 fev. 2026.

PINHO, Jacinta Oliveira *et al.* Evaluation of guidelines on antimicrobials use in food-producing animals: a systematic review. **One Health Outlook**, [*s. l.*], vol. 7, no. 1, p. 39, 2025. Disponível em: https://doi.org/10.1186/s42522-025-00160-w

KARWOWSKA, Ewa. Antibiotic resistance in the farming environment. **Applied Sciences**, [*s. l.*], vol. 14, no. 13, p. 5776, 2024. Available at: https://doi.org/10.3390/app14135776

MACNAIR, Craig R; RUTHERFORD, Steven T; TAN, Man-Wah. Alternative therapeutic strategies to treat antibiotic-resistant pathogens. **Nature Reviews Microbiology**, [*s. l.*], vol. 22, no. 5, p. 262–275, 2024. Available at: https://doi.org/10.1038/s41579-023-00993-0

WORLD HEALTH ORGANIZATION. **WHO Bacterial Priority Pathogens List, 2024**: bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Geneva: World Health Organization, 2024.

MURRAY, Christopher J. L. *et al*. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. **The Lancet**, [s. l.], v. 399, n. 10325, p. 629-655, 12 fev. 2022. Disponível em: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02724-0/fulltext. Acesso em: 8 fev. 2026.

ANTIMICROBIAL RESISTANCE COLLABORATORS. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. **The Lancet**, [*s. l.*], vol. 399, no. 10325, p. 629–655, 2022. Available at: https://doi.org/10.1016/S0140-6736(21)02724-0

INDA-DÍAZ, Juan Salvador *et al.* Latent antibiotic resistance genes are abundant, diverse, and mobile in human, animal, and environmental microbiomes. **Microbiome**, [*s. l.*], vol. 11, no. 1, p. 44, 2023. Available at: https://doi.org/10.1186/s40168-023-01479-0

PARTRIDGE, S R *et al.* Mobile genetic elements associated with antimicrobial resistance. **Clinical Microbiology Reviews**, [*s. l.*], vol. 31, no. 4, 2018. Available at: https://doi.org/10.1128/CMR.00088-17

WONG, Carissa. **Antibiotic resistance is a growing threat — is climate change making it worse?** Nature, [s. l.], v. 625, p. 224-225, 12 jan. 2024. Disponível em: https://doi.org/10.1038/d41586-023-04077-0. Acesso em: 8 fev. 2026.

BARREDO, José I. Major flood disasters in Europe: 1950–2005. **Natural Hazards**, [*s. l.*], vol. 42, no. 1, p. 125–148, 2007. Available at: https://doi.org/10.1007/s11069-006-9065-2

RAMOS, Marília Patta *et al.* As enchentes de 2024 no Rio Grande do Sul e a capacidade de resposta dos municípios às inundações. **Revista Eletrônica Direito e Sociedade - REDES**, [*s. l.*], vol. 29, 2024. Available at: https://doi.org/10.17058/redes.v29i1.19566

DALL' IGNA, Eduardo. **Análise dos impactos das inundações em Porto Alegre sobre o sistema de abastecimento de água**. Orientador: Salatiel Wohlmuth da Silva. Coorientador: Louidi Lauer Albornoz. 2025. Trabalho de Conclusão de Curso (Graduação em Engenharia Civil) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2025. Available at: https://lume.ufrgs.br/handle/10183/289905

KIM, Minjae *et al.* Metagenomics indicate that public health risk may be higher from flooding following dry versus rainy periods. **Environmental microbiology reports**, [*s. l.*], vol. 14, no. 2, p. 265–273, 2022. Available at: https://doi.org/10.1111/1758-2229.13047

TALAPKO, Jasminka *et al.* Antimicrobial Peptides-Mechanisms of Action, Antimicrobial Effects and Clinical Applications. **Antibiotics (Basel, Switzerland)**, [*s. l.*], vol. 11, no. 10, 2022. Available at: https://doi.org/10.3390/antibiotics11101417

BOURZAC, Katherine. Old antibiotics are being revived to fight new threats. **Nature**, v. 653, p. S18-S19, 2026. Disponível em: https://www.nature.com/articles/d41586-026-01379-x. Acesso em: 16 jun. 2026.

HARYINI, Sree *et al.* Antimicrobial peptides: Bioinformatic advances and translational therapeutics to combat antibiotic resistance. **Advances in protein chemistry and structural biology**, [*s. l.*], vol. 149, p. 1–59, 2026. Available at: https://doi.org/10.1016/bs.apcsb.2025.10.017

BUCATARU, Cezara; CIOBANASU, Corina. Antimicrobial peptides: Opportunities and challenges in overcoming resistance. **Microbiological Research**, [*s. l.*], vol. 286, p. 127822, 2024. Available at: https://doi.org/10.1016/j.micres.2024.127822

SANTOS-JÚNIOR, Célio Dias *et al.* Discovery of antimicrobial peptides in the global microbiome with machine learning. **Cell**, [*s. l.*], vol. 187, no. 14, p. 3761-3778.e16, 2024. Available at: https://doi.org/10.1016/j.cell.2024.05.013

ZHOU, Xiangrun *et al.* Deep learning for antimicrobial peptides: computational models and databases. **Journal of Chemical Information and Modeling**, [*s. l.*], vol. 65, no. 4, p. 1708–1717, 2025. Available at: https://doi.org/10.1021/acs.jcim.5c00006

MA, Yue *et al.* Identification of antimicrobial peptides from the human gut microbiome using deep learning. **Nature Biotechnology**, [*s. l.*], vol. 40, no. 6, p. 921–931, 2022. Available at: https://doi.org/10.1038/s41587-022-01226-0

XU, Jiaqi *et al.* Waste to resource: Mining antimicrobial peptides in sludge from metagenomes using machine learning. **Environment International**, [*s. l.*], vol. 186, p. 108574, 2024. Available at: https://doi.org/10.1016/j.envint.2024.108574

ANDREWS, S. **FastQC**: a quality control tool for high throughput sequence data. [s. l.], 2010. Disponível em: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Acesso em: 8 fev. 2026.

FU, Limin *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, [*s. l.*], vol. 28, no. 23, p. 3150–3152, 2012. Available at: https://doi.org/10.1093/bioinformatics/bts565

NURK, S. *et al*. metaSPAdes: a new versatile metagenomic assembler. **Genome Research**, [s. l.], v. 27, n. 7, p. 824-834, 2017. Disponível em: https://doi.org/10.1101/gr.213959.116. Acesso em: 8 fev. 2026.

NURK, Sergey; MELESHKO, Dmitry; KOROBEYNIKOV, Anton; *et al.* metaSPAdes: a new versatile metagenomic assembler. **Genome Research**, vol. 27, no. 5, p. 824–834, 2017.

CORDOVES-DELGADO, Greneter; GARCÍA-JACAS, César R. Predicting Antimicrobial Peptides Using ESMFold-Predicted Structures and ESM-2-Based Amino Acid Features with Graph Deep Learning. **Journal of Chemical Information and Modeling**, [*s. l.*], vol. 64, no. 10, p. 4310–4321, 2024. Available at: https://doi.org/10.1021/acs.jcim.3c02061

RATHORE, Anand Singh *et al.* ToxinPred 3.0: An improved method for predicting the toxicity of peptides. **Computers in biology and medicine**, [*s. l.*], vol. 179, p. 108926, 2024. Available at: https://doi.org/10.1016/j.compbiomed.2024.108926

CHAUDHARY, Kumardeep *et al.* A web server and mobile app for computing hemolytic potency of peptides. **Scientific Reports**, [*s. l.*], vol. 6, p. 22843, 2016. Available at:

https://doi.org/10.1038/srep22843

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 57, n. 1, p. 289-300, 1995.

FAN, Linlin *et al.* DRAMP: a comprehensive data repository of antimicrobial peptides. **Scientific Reports**, [*s. l.*], vol. 6, p. 24482, 2016. Available at: https://doi.org/10.1038/srep24482

WOOD, Derrick E; LU, Jennifer; LANGMEAD, Ben. Improved metagenomic analysis with Kraken 2. **Genome Biology**, [*s. l.*], vol. 20, no. 1, p. 257, 2019. Available at: https://doi.org/10.1186/s13059-019-1891-0

CONOVER, W. J. **Practical nonparametric statistics**. 3. ed. New York: Wiley, 1999.

SHANNON, Paul *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome Research**, [*s. l.*], vol. 13, no. 11, p. 2498–2504, 2003. Available at: https://doi.org/10.1101/gr.1239303

BOMAN, H. G. Antibacterial peptides: basic facts and emerging concepts. **Journal of Internal Medicine**, [s. l.], v. 254, n. 3, p. 197-215, 2003.

DAYHOFF, M. O.; SCHWARTZ, R. M.; ORCUTT, B. C. A model of evolutionary change in proteins. **Atlas of Protein Sequence and Structure**, [s. l.], v. 5, n. 3, p. 345-352, 1978.

KYTE, J.; DOOLITTLE, R. F. A simple method for displaying the hydropathic character of a protein. **Journal of Molecular Biology**, [s. l.], v. 157, n. 1, p. 105-132, 1982.

LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Lehninger**. 7. ed. Porto Alegre: Artmed, 2019.

JUMPER, J. *et al*. Highly accurate protein structure prediction with AlphaFold. **Nature**, [s. l.], v. 596, n. 7873, p. 583-589, 2021. Disponível em: https://doi.org/10.1038/s41586-021-03819-2. Acesso em: 8 fev. 2026.

LIN, Z. *et al*. Evolutionary-scale prediction of atomic-level protein structure with a language model. **Science**, [s. l.], v. 379, n. 6637, p. 1123-1130, 2023. Disponível em: https://doi.org/10.1126/science.ade2574. Acesso em: 8 fev. 2026.

HYATT, Doug; CHEN, Gwo-Liang; LOCASCIO, Philip F; *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, vol. 11, no. 1, 2010.

ANDERSON, M. J. A new method for non-parametric multivariate analysis of variance. **Austral Ecology**, [s. l.], v. 26, n. 1, p. 32-46, 2001. Disponível em: https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x. Acesso em: 8 fev. 2026.

ANDERSON, Marti J. Distance-based tests for homogeneity of multivariate dispersions. **Biometrics**, v. 62, n. 1, p. 245-253, 2006.

CHEN, Zixin *et al.* UniAMP: enhancing AMP prediction using deep neural networks with inferred information of peptides. **BMC Bioinformatics**, [*s. l.*], vol. 26, no. 1, p. 10, 2025. Available at: https://doi.org/10.1186/s12859-025-06033-3

SANTOS-JÚNIOR, Célio Dias *et al.* Macrel: antimicrobial peptide screening in genomes and metagenomes. **PeerJ**, [*s. l.*], vol. 8, p. e10555, 2020. Available at: https://doi.org/10.7717/peerj.10555

ZHAO, Jun *et al.* deep-AMPpred: A Deep Learning Method for Identifying Antimicrobial Peptides and Their Functional Activities. **Journal of Chemical Information and Modeling**, [*s. l.*], vol. 65, no. 2, p. 997–1008, 2025. Available at: https://doi.org/10.1021/acs.jcim.4c01913

BRAY, J. R.; CURTIS, J. T. An ordination of the upland forest communities of southern Wisconsin. **Ecological Monographs**, [s. l.], v. 27, n. 4, p. 325-349, 1957. Disponível em: https://doi.org/10.2307/1942268. Acesso em: 8 fev. 2026

HESS, B.; BEKKER, H.; BERENDSEN, H. J. C.; FRAAIJE, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, 1997. Disponível em: https://www.researchgate.net/publication/2790343_LINCS_A_Linear_Constraint_Solver_for_molecular_simulations. Acesso em: 15 junho. 2026

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, [s. l.], v. 52, n. 3/4, p. 591-611, 1965. Disponível em: https://doi.org/10.2307/2333709. Acesso em: 8 fev. 2026.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, [s. l.], v. 47, n. 260, p. 583-621, 1952. Disponível em: https://doi.org/10.2307/2280779. Acesso em: 8 fev. 2026.

DUNN, Olive J. Multiple comparisons using rank sums. **Technometrics**, v. 6, n. 3, p. 241-252, 1964.

MCMURDIE, Paul J.; HOLMES, Susan. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. **PLoS ONE**, v. 8, n. 4, p. e61217, 2013.

OGLE, Derek H. et al. **FSA**: Fisheries Stock Analysis. R package version 0.10.1. 2024. Disponível em: https://CRAN.R-project.org/package=FSA. Acesso em: 19 jun. 2026.

OKSANEN, Jari et al. **vegan**: Community Ecology Package. R package version 2.6-4. 2022. Disponível em: https://CRAN.R-project.org/package=vegan. Acesso em: 19 jun. 2026.

DU, Rubing *et al.* Metagenomics reveals potential antimicrobial peptides in Chinese baijiu fermentation. **Food microbiology**, [*s. l.*], vol. 134, p. 104918, 2026. Available at: https://doi.org/10.1016/j.fm.2025.104918

PERRYMAN, Shane E *et al.* Urban stormwater runoff drives denitrifying community composition through changes in sediment texture and carbon content. **Microbial Ecology**, [*s. l.*], vol. 61, no. 4, p. 932–940, 2011. Available at: https://doi.org/10.1007/s00248-011-9833-8

RILLIG, Matthias C; MANSOUR, India. Microbial ecology: community coalescence stirs things up. **Current Biology**, [*s. l.*], vol. 27, no. 23, p. R1280–R1282, 2017. Available at: https://doi.org/10.1016/j.cub.2017.10.027

MACIAS, Patricia Yong *et al.* Flooding events reshape bacterioplankton assemblages and functional properties in a hypertrophic Shallow Lake (Laguna Chascomús, Argentina). **Hydrobiologia**, [*s. l.*], vol. 853, no. 6, p. 1807–1827, 2026. Available at: https://doi.org/10.1007/s10750-025-06031-9

LIU, Yang *et al.* Microbial community structure and functional properties in permanently and seasonally flooded areas in Poyang Lake. **Scientific Reports**, [*s. l.*], vol. 10, no. 1, p. 4819, 2020. Available at: https://doi.org/10.1038/s41598-020-61569-z

SHADE, Ashley *et al.* Fundamentals of microbial community resistance and resilience. **Frontiers in Microbiology**, [*s. l.*], vol. 3, p. 417, 2012. Available at: https://doi.org/10.3389/fmicb.2012.00417

ZHAO, Man *et al.* Dynamics of bacterioplankton community structure in response to seasonal hydrological disturbances in Poyang Lake, the largest wetland in China. **FEMS Microbiology Ecology**, [*s. l.*], vol. 96, no. 8, 2020. Available at: https://doi.org/10.1093/femsec/fiaa064

XU, Biying *et al.* Spatial and Seasonal Dynamics of Plankton Community and Its Relationship with Environmental Factors in an Urban River: A Case Study of Wuxi City, China. **Water**, [*s. l.*], vol. 17, no. 1, p. 51, 2024. Available at: https://doi.org/10.3390/w17010051

DUAN, Xinyi *et al.* Effects of extreme flooding on soil characteristics, soil enzyme activity, and microbial structure in shengjin lake. **Water**, [*s. l.*], vol. 17, no. 12, p. 1789, 2025. Available at: https://doi.org/10.3390/w17121789

SHABAROVA, Tanja *et al.* Recovery of freshwater microbial communities after extreme rain events is mediated by cyclic succession. **Nature Microbiology**, [*s. l.*], vol. 6, no. 4, p. 479–488, 2021. Available at: https://doi.org/10.1038/s41564-020-00852-1

COMTE, Jérôme; FAUTEUX, Lisa; DEL GIORGIO, Paul A. Links between metabolic plasticity and functional redundancy in freshwater bacterioplankton communities. **Frontiers in Microbiology**, [*s. l.*], vol. 4, p. 112, 2013. Available at: https://doi.org/10.3389/fmicb.2013.00112

BARKA, Essaid Ait *et al.* Taxonomy, physiology, and natural products of actinobacteria. **Microbiology and Molecular Biology Reviews**, [*s. l.*], vol. 80, no. 1, p. 1–43, 2016. Available at: https://doi.org/10.1128/MMBR.00019-15

CHEUNG, Calum *et al.* The spread of antimicrobial resistance in the aquatic environment from faecal pollution: a scoping review of a multifaceted issue. **Environmental Monitoring and Assessment**, [*s. l.*], vol. 197, no. 4, p. 467, 2025. Available at: https://doi.org/10.1007/s10661-025-13860-7

IBEKWE, A Mark; MA, Jincai; MURINDA, Shelton E. Bacterial community composition and structure in an Urban River impacted by different pollutant sources. **The Science of the Total Environment**, [*s. l.*], vol. 566–567, p. 1176–1185, 2016. Available at: https://doi.org/10.1016/j.scitotenv.2016.05.168

MCCORMICK, Joseph R; FLÄRDH, Klas. Signals and regulators that govern Streptomyces development. **FEMS Microbiology Reviews**, [*s. l.*], vol. 36, no. 1, p. 206–231, 2012. Available at: https://doi.org/10.1111/j.1574-6976.2011.00317.x

CHATER, Keith F. Recent advances in understanding Streptomyces. [version 1; peer review: 4 approved]. **F1000Research**, [*s. l.*], vol. 5, p. 2795, 2016. Available at: https://doi.org/10.12688/f1000research.9534.1

WANG, Yi *et al.* Shifts in microbial communities in shallow lakes depending on trophic states: Feasibility as an evaluation index for eutrophication. **Ecological Indicators**, [*s. l.*], vol. 136, p. 108691, 2022. Available at: https://doi.org/10.1016/j.ecolind.2022.108691

FERMANI, Paulina *et al.* Divergent dynamics of microbial components in two temperate shallow lakes with contrasting steady states in the Southern Hemisphere. **Anais da Academia Brasileira de Ciências**, [*s. l.*], vol. 94, no. 1, p. e20191545, 2022. Available at: https://doi.org/10.1590/0001-3765202220191545

SMITS, Adrianne P. *et al.* Hydrodynamics structure plankton communities and interactions in a

freshwater tidal estuary. **Ecological Monographs**, [*s. l.*], vol. 93, no. 2, 2023. Available at: https://doi.org/10.1002/ecm.1567

NEWTON, Ryan J *et al.* A guide to the natural history of freshwater lake bacteria. **Microbiology and Molecular Biology Reviews**, [*s. l.*], vol. 75, no. 1, p. 14–49, 2011. Available at: https://doi.org/10.1128/MMBR.00028-10

HAHN, Martin W *et al.* Polynucleobacter cosmopolitanus sp. nov., free-living planktonic bacteria inhabiting freshwater lakes and rivers. **International Journal of Systematic and Evolutionary Microbiology**, [*s. l.*], vol. 60, no. Pt 1, p. 166–173, 2010. Available at: https://doi.org/10.1099/ijs.0.010595-0

JEZBERA, Jan *et al.* Ubiquity of Polynucleobacter necessarius subspecies asymbioticus results from ecological diversification. **Environmental Microbiology**, [*s. l.*], vol. 13, no. 4, p. 922–931, 2011. Available at: https://doi.org/10.1111/j.1462-2920.2010.02396.x

RILEY, Margaret A; WERTZ, John E. Bacteriocins: evolution, ecology, and application. **Annual Review of Microbiology**, [*s. l.*], vol. 56, p. 117–137, 2002. Available at: https://doi.org/10.1146/annurev.micro.56.012302.161024

GRANATO, Elisa T; MEILLER-LEGRAND, Thomas A; FOSTER, Kevin R. The evolution and ecology of bacterial warfare. **Current Biology**, [*s. l.*], vol. 29, no. 11, p. R521–R537, 2019. Available at: https://doi.org/10.1016/j.cub.2019.04.024

ZASLOFF, Michael. Antimicrobial peptides of multicellular organisms. **Nature**, [*s. l.*], vol. 415, no. 6870, p. 389–395, 2002. Available at: https://doi.org/10.1038/415389a

BROGDEN, Kim A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? **Nature Reviews Microbiology**, [*s. l.*], vol. 3, no. 3, p. 238–250, 2005. Available at: https://doi.org/10.1038/nrmicro1098

CRESCA, Sofia *et al.* Current Status of Molecular Dynamics Simulations of Membrane Permeabilization by Antimicrobial Peptides and Pore-Forming Proteins: A Review. **Journal of Chemical Information and Modeling**, [*s. l.*], vol. 66, no. 4, p. 1982–2005, 2026. Available at: https://doi.org/10.1021/acs.jcim.5c02731

DATHE, Margitta; WIEPRECHT, Torsten. Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. **Biochimica et Biophysica Acta (BBA) - Biomembranes**, [*s. l.*], vol. 1462, no. 1–2, p. 71–87, 1999. Available at: https://doi.org/10.1016/S0005-2736(99)00201-1

ULMSCHNEIDER, Jakob P; ULMSCHNEIDER, Martin B. Molecular Dynamics Simulations

Are Redefining Our View of Peptides Interacting with Biological Membranes. **Accounts of Chemical Research**, [*s. l.*], vol. 51, no. 5, p. 1106–1116, 2018. Available at: https://doi.org/10.1021/acs.accounts.7b00613

MILLER, William R; ARIAS, Cesar A. ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. **Nature Reviews Microbiology**, [*s. l.*], vol. 22, no. 10, p. 598–616, 2024. Available at: https://doi.org/10.1038/s41579-024-01054-w