

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

**Instituto de Ciências Biológicas**

**Programa Interunidades de Pós-Graduação em Bioinformática da UFMG**

Bruno de Paula Oliveira Santos

**Caracterização biofísica e estrutural de criptídeos antimicrobianos:**

Estudo da porção antigênica da proteína SmKI-1 de *Schistosoma mansoni*

Belo Horizonte  
2021

Bruno de Paula Oliveira Santos

**Caracterização biofísica e estrutural de criptídeos antimicrobianos:**

Estudo da porção antigênica da proteína SmKI-1 de *Schistosoma mansoni*

Versão Final

Tese apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais para obtenção do título de Doutor em Bioinformática.

Orientação: Profa. Dra. Mariana Torquato Quezado de Magalhães e Prof. Dr. Sérgio Costa de Oliveira.

Belo Horizonte  
2021

043

Santos, Bruno de Paula Oliveira.

Caracterização biofísica e estrutural de criptídeos antimicrobianos: estudo da porção antigênica da proteína SmKI-1 de *Schistosoma mansoni* [manuscrito] / Bruno de Paula Oliveira Santos. – 2021.

186 f. : il. ; 29,5 cm.

Orientação: Profa. Dra. Mariana Torquato Quezado de Magalhães e Prof. Dr. Sérgio Costa de Oliveira.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Peptídeos. 3. Espectroscopia de Ressonância Magnética. I. Magalhães, Mariana Torquato Quezado de. II. Oliveira, Sérgio Costa de. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
**Instituto de Ciências Biológicas**  
**Programa Interunidades de Pós-Graduação em Bioinformática da UFMG**

**PARECER Nº 28/2021**

PROCESSO Nº 23072.250218/2021-06

**FOLHA DE APROVAÇÃO**

**"Caracterização biofísica e estrutural de criptídeos antimicrobianos: Estudo da porção antigênica da proteína SmKI-1 de Schistosoma mansoni"**

**Bruno de Paula Oliveira Santos**

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Sérgio Costa Oliveira - Orientador  
Universidade Federal de Minas Gerais

Profa. Mariana Torquato Quezado de Magalhaes - Coorientadora  
Universidade Federal de Minas Gerais

Prof. Guilherme Dotto Brand  
Universidade de Brasília

Prof. Octavio Luis Franco  
Universidade Católica de Brasília

Prof. Anderson Sá Pinheiro  
Universidade Federal do Rio de Janeiro

Prof. Lucas Bleicher  
Universidade Federal de Minas Gerais  
Belo Horizonte, 01 de outubro de 2021.





Documento assinado eletronicamente por **Guilherme Dotto Brand, Usuário Externo**, em 01/10/2021, às 15:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Anderson de Sá Pinheiro, Usuário Externo**, em 01/10/2021, às 16:32, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sergio Costa Oliveira, Professor do Magistério Superior**, em 01/10/2021, às 16:41, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes, Professora do Magistério Superior**, em 01/10/2021, às 17:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Lucas Bleicher, Professor do Magistério Superior**, em 02/10/2021, às 20:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Octávio Luiz Franco, Usuário Externo**, em 03/10/2021, às 07:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0983183** e o código CRC **1119B2D3**.

Referência: Processo nº 23072.250218/2021-06 SEI nº 0983183

À Mirian, ao Alaésio e à Sandra - minha mãe, meu pai e minha tia.

## **Agradecimentos**

Em fevereiro de 2017, eu peguei um ônibus de Goiânia (GO) rumo à Belo Horizonte (MG), para participar de uma seleção de doutorado, sem ter noção do que me esperava nessa cidade. A viagem de ônibus é sempre noturna, porque demora entre 12-13 horas para atravessar os 900 km de distância entre o cerrado goianiense e os morros belo-horizontinos. Durante o final do mestrado, eu percebi que queria muito me tornar um bioinformata, então quando o professor Sérgio Oliveira respondeu meu email em que eu pedia a sua orientação, eu fiquei muito feliz com a oportunidade. Já em terras mineiras, eu fiquei sabendo que a professora Mariana Quezado participaria da minha orientação, só não sabia ainda o papel fundamental que ela teria no processo.

A seleção foi um desastre. Eu não acho que eu tenha participado antes de uma seleção em que eu tenha ficado mais branco do que aquele dia. Então já agradeço à Sheila pelo apoio moral naquele momento. Voltei para minha cidade naquele dia pensando que nunca mais eu pisaria na UFMG. Qual foi a minha surpresa ao descobrir que eu havia passado. Quatro meses depois eu estava voltando para BH com o porta mala e a parte traseira do carro do meu pai com os itens da mudança. Quatro anos depois e eu já não me reconheço mais na pele daquele garoto de 23 anos. Maturidade, conhecimento, visão de mundo, responsabilidades, novos projetos. Ganhei tudo isso no caminho e grande parte deve-se ao doutorado e eu tenho muito o que agradecer por isso.

Tive a boa sorte de ter recebido a orientação da professora Mariana, por diversos aspectos. Como professora recente na UFMG e com poucos alunos, a Mariana sempre esteve presente e disponível para ensinar. Durante esse período eu vi o laboratório crescer, nós mudamos de espaço, organizamos muito os laboratórios, estabelecemos técnicas e procedimentos padrões, mudamos equipamentos de lugar, descobrimos quem poderia fornecer reagentes, quem poderia auxiliar com técnicas (obrigado, Jamil!). Eu cheguei pelo H4 e depois fomos para o O4 125, a atual localização do laboratório, que hoje recebeu o nome de LBM. Eu tive muita sorte de ter convivido e aprendido com a Mariana. Além de sempre solícita, uma das coisas que eu mais admiro na Mariana é seu senso crítico e seu amor pela ciência. Aprendi com ela que as duas coisas não andam separadas.

Meus agradecimentos ao professor Sérgio Costa, por ter acreditado em mim e por ter me dado essa oportunidade. Espero ter atendido às expectativas e ter contribuído com o grupo

de pesquisa. O pessoal do LIDI sempre me tratou muito bem e sempre pude contar com eles quando eu tinha dúvidas ou precisava de auxílios científicos. Posso contar que a gente compartilhou muitos *happy hours*?

Eu agradeço também a todos os meus colegas de trabalho, que me ensinaram muitas coisas nesses anos. O Fábio Mambelli, pela disposição e paciência de me ensinar sobre proteínas e esquistossomose. A Camila Akemi, pela energia, alto astral e por me ensinar a ser mais organizado. O John Amaya, pela determinação e pelo apoio de sempre. O Bruno Marques, pela disposição em ajudar e por me mostrar que digitar no teclado é um processo barulhento. A Adriana Raabe, pelo apoio no laboratório multiusuário e por ter compartilhado bons momentos comigo. A Verônica Silva, por ser sempre prestativa, pela amizade e pela água deionizada nos momentos de urgência. A Sandra Fernandes, por ter me ajudado desde o início, quando eu vim prestar a seleção e ainda hoje, quando preciso de auxílio. O Jamil Oliveira, uma das pessoas mais legais, gentis e inteligentes do ICB, que eu muito admiro. A Eliane Alves, que foi a primeira pessoa que me deixou configurar experimentos de RMN sozinho. O professor Luciano Lião, por abrir as portas para mim no seu laboratório. A Carolina Ferreira, por ter me auxiliado com os experimentos de termodinâmica. O professor Rodrigo Verly, pelo bom humor e por ter aberto as portas do seu laboratório para mim. O professor Adriano Pimenta, cujo bom humor e sagacidade estão sempre presentes. A Sheila, pelo carisma e ótimo serviço junto à secretaria da bioinformática. O Tiago, pela solicitude e atendimento na secretaria. Algumas menções honrosas, que não podem deixar de ser citadas: Yala, Jaqueline, Marcela, Bárbara, Júlia, Ana, Pedro, Rodrigo e Duana. Espero não ter esquecido ninguém.

Aproveito também para agradecer aos membros da minha banca de qualificação, que me fizeram perceber as minhas dificuldades no projeto e onde eu precisava melhorar: os professores Thiago Braga, Francisco Lobo e Jarbas Magalhães. E meu agradecimento aos membros da banca de defesa, titulares e suplentes, por toparem contribuir com o meu trabalho: professores Guilherme Brand, Octávio Franco, Anderson de Sá, Lucas Bleicher, e Adolfo Moraes. Os meus agradecimentos também às agências de fomento, necessários para todo projeto de pesquisa e bolsa: CAPES, FAPEMIG e CNPQ.

Os nossos amigos são fundamentais em diversas áreas da nossa vida. Gostaria de agradecer o Danilo Resende, a Monalisa Trentini e a Fernanda Tana por terem lido versões desta tese, fazendo um ótimo trabalho editorial. Pelo apoio moral e psicológico, agradeço,

além dos nomes já citados, à Alina Roswitha, ao Carlos Silveira, à Paulianne Lima e à Ludmila Fernandes.

Finalmente, devo muito aos meus pais, à quem dedico essa tese. O meu doutoramento é um dos sonhos da minha mãe, Mirian, e que eu fico muito feliz de ter conseguido realizar. Graças ao esforço dela, e do meu pai, Alaésio, que sempre se preocupou com meu bem estar, me permitiram chegar aqui. Minha tia Sandra também foi essencial nessa etapa e eu serei eternamente grato. Fico feliz por ter pessoas na minha vida com quem eu possa contar. Tenho consciência de que investir em educação continuada é oneroso e uma realidade que não é acessível a maior parte do nosso povo. Agradeço o privilégio de ter pessoas que investiram na minha educação e que me permitiram prosseguir nesse caminho.

“Se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda...”

Paulo Freire

“A educação é o ponto em que decidimos se amamos o mundo o bastante para assumirmos a  
responsabilidade por ele.”

Hannah Arendt

## Resumo

A área de estudo de agentes antimicrobianos é extensa e diversa, uma vez que microrganismos patogênicos continuam a causar sérios problemas de saúde em todo o mundo, além de serem capazes de se tornar resistentes à maioria das drogas padronizadas em sistemas de saúde. Nesta tese, buscou-se identificar peptídeos antimicrobianos escondidos na sequência de uma proteína do parasito *Schistosoma mansoni*. Devido seu mecanismo de ação, hipotetizou-se que o inibidor do tipo kunitz SmKI-1 é clivado do tegumento e seria então capaz de liberar peptídeos membrano ativos do seu domínio C-terminal. Para este trabalho, criou-se um esquema de trabalho para 1) identificar os peptídeos escondidos, chamados criptídeos, por meio de uma programa desenvolvido *in house*; 2) avaliar a ação desses peptídeos, denominados schistocinas; 3) estudar suas características biofísicas por experimentos *in vitro* e *in silico*. Na primeira parte, o algoritmo enCrypted foi usado para realizar uma proteólise *in silico* da SmKI-1 e uma triagem para potencial atividade antimicrobiana. As schistocinas foram sintetizadas por fase sólida, testadas *in vitro* e avaliadas por estudos estruturais (CD, RMN) e termodinâmicos (ITC) para estudar sua relação estrutura-função e avaliar a interação peptídeo-membrana. O enCrypted foi capaz de prever AMPs em SmKI-1, com acurácia de 87%. As análises biofísicas descreveram uma mudança conformacional induzida por membranas zwitteriônicas e aniônicas, saindo de uma conformação de alça randômica para  $\alpha$ -hélices e um equilíbrio peptídeo-membrana para todas as schistocinas. Os dados estruturais e de dinâmica permitem sugerir um modo de ação bem conhecido de interação peptídeo-membrana em que a atração eletrostática entre os peptídeos catiônicos e as membranas aniônicas resulta no desordenamento da bicamada. Além disso, os dados de troca de H/D por RMN com a maior contribuição entrópica observada para a interação peptídeo-membrana mostraram que as schistocinas têm orientações diferentes sobre a membrana. Por fim, durante o processo de estudo, surgiu um problema cuja solução foi adicionada à tese: adicionar modificações terminais aos peptídeos durante o cálculo de estrutura com dois programas específicos: CcpNMR Analysis 2.5 e Aria 2.3. Este trabalho demonstra a relevância de combinar análises computacionais com métodos biofísicos para entender a afinidade peptídeo-membrana e melhorar outros algoritmos.

**Palavras-chaves:** preditor de peptídeos; criptídeos; interação peptídeo-membrana

## Abstract

Antimicrobial drugs design is a study field that is extensive and diverse, as pathogenic microorganisms continue to cause serious health problems around the world, in addition to being able to become resistant to most drugs standardized in healthcare systems. In this thesis, we sought to identify antimicrobial peptides hidden in the sequence of a protein from the *Schistosoma mansoni* parasite. As a mechanism of action, it was hypothesized that the kunitz-type inhibitor SmKI-1 is cleaved from the tegument and would then be able to release membrane active peptides from its C-terminal domain. For this work, a pipeline was created to 1) identify the hidden peptides, named cryptids, through an in-house software; 2) evaluate the peptides, named schistocins, activity; 3) study their biophysical characteristics by *in vitro* and *in silico* experiments. In the first part, the enCrypted algorithm was used to perform proteolysis on SmKI-1 sequence and screen for potential antimicrobial activity. Schistocins were synthesized by solid phase platform, tested *in vitro* and evaluated by structural (CD, NMR) and thermodynamic (ITC) studies to investigate their structure-function relationship and to evaluate a peptide-membrane interaction. The enCrypted was able to predict AMPs in SmKI-1, with an accuracy of 87%. Biophysical analyzes describe a conformational change induced by zwitterionic and anionic membranes, moving from a random-coil-to- $\alpha$ -helices conformation and a peptide-membrane balance for all schistocins. The structural and dynamics data allow us to suggest a well-known mode of action of peptide-membrane interaction, in which the electrostatic attraction between cationic peptides and as anionic membranes results in bilayer disorder. In addition, NMR H/D exchange data with the largest observed entropy contribution to a peptide-membrane interaction is important as schistocins have different orientations on the membrane. This multidisciplinary approach makes sense of the best characteristics of these molecules and evaluates their potential. Finally, during the study process, a problem emerged whose solution was added to the thesis: adding terminals to peptides during structure calculation with two specific programs: CcpNMR Analysis 2.5 and Aria 2.3. This work demonstrates the robustness of the use of physical-chemical features of predicted peptides in the identification of new bioactive cryptids. In addition, the work shows a combination of these analyzes with biophysical methods to better understand peptide-membrane affinity and improve other algorithms.

**Keywords:** peptide predictor; cryptides; peptide-membrane interaction



## Lista de Figuras

### Capítulo I

Figura 1 - Criptídeos de ubiquitina	22
Figura 2 - Criptídeos do rpS30 Fau.	23
Figura 3 - Criptídeos da histona H2A.	25
Figura 4 - Criptídeos da histona H4.	27
Figura 5 - Criptídeos da histona H1 e H5.	29
Figura 6 - Criptídeos da apolipoproteína B.	30
Figura 7 - Criptídeos da albumina.	32
Figura 8 - Criptídeos de hemocianina e hemoglobina.	36
Figura 9 - Criptídeos de anticorpos.	37
Figura 10 - Criptídeos de grânulos secretórios.	40
Figura 11 - Criptídeos de quimiocinas.	42
Figura 12 - Criptídeos de enzimas.	43
Figura 13 - Criptídeos de lactoferrina.	45
Figura 14 - Criptídeos de proteínas da coagulação.	48
Figura 15 - Criptídeos de proteínas da coagulação.	48
Figura 16 - Criptídeos do colágeno.	49
Figura 17 - Outros criptídeos.	53

### Capítulo II

Figura 1 - Fluxograma de desenvolvimento do modelo preditor de peptídeos antibacterianos	71
Figura 2 - Performance dos algoritmos durante o processo de validação cruzada	79
Figura 3 - Fluxograma de funcionamento do enCrypted	81
Figura 4 - Representação das schistocinas em rodas de hélices	82
Figura 5 - Alinhamento de sequências das schistocinas com sequências da base de dados APD	84
Figura 6 - Separação cromatográfica e identificação espectrométrica das schistocinas	86
Figura 7 - Análise de MS/MS dos fragmentos das schistocinas	88

Figura 8 - Espectro de dicroísmo de schistocinas de 30 $\mu$ M em vesículas POPC e POPC:POPG (3:1) (A-H)	91
Figura 9 - Espectros de dicroísmo de schistocinas de 30 $\mu$ M em TFE e SDS	93
Figura 10 - Calorimetria de titulação isotérmica das schistocinas em LUVs aniônicas	95
Figura 11 - Sinais dos experimentos de TOCSY HN-Ha e NOESY HN-HN das schistocinas	100
Figura 12 - Comparação entre hidrogênios alfa dos peptídeos schis-3 e schis-3.1	101
Figura 13 - Diagrama de conectividade sequencial das schistocinas em micelas de DPC	103
Figura 14 - Estruturas tridimensionais do criptídeo schistocina-1 obtidas por RMN de solução (400 $\mu$ M de peptídeos em 20 mM de micelas DPC- $d_{38}$ ).	104
Figura 15 - Estruturas tridimensionais do criptídeo schistocina-2 obtidas por RMN de solução (400 $\mu$ M de peptídeos em 20 mM de micelas DPC- $d_{38}$ ).	106
Figura 16 - Estruturas tridimensionais do criptídeo schistocina-3 obtidas por RMN de solução (400 $\mu$ M de peptídeos em 20 mM de micelas DPC- $d_{38}$ ).	107
Figura 17 - Estruturas tridimensionais do criptídeo schistocina-3.1 obtidas por RMN de solução (400 $\mu$ M de peptídeos em 20 mM de micelas DPC- $d_{38}$ ).	109
Figura 18 - Representação em gráfico de ajuste exponencial de troca de hidrogênio amídico	112
Figura 19 - Troca de hidrogênio amídico mostrado em proporção de volume para cada resíduo nas schistocinas.	115
Figura 20 - O desvio (RMSD) das schistocinas em interação com micelas de DPC	117
Figura 21 - O desvio (RMSD) das schistocinas em bicamadas zwitteriônicas e aniônicas	119

### Capítulo III

Figura 1 - Fluxograma dos arquivos modificados para amidação e acetilação de proteínas com CcpNmr Analysis e Aria2.3	143
Figura 2 - Esquema para uso de modificações terminais no CcpNmr Analysis	144
Figura 3 - Esquema para uso de modificações terminais no Aria2.3	146

## Lista de Tabelas

### Capítulo I

Tabela 1 - Potência de atividade antimicrobiana dos criptídeos.	55
---	----

### Capítulo II

Tabela 1 - Atributos do classificador	69
Tabela 2 - Performance de comparação dos preditores	80
Tabela 3 - Atividade antibacteriana e antifúngica de schistocinas	89
Tabela 4 - Porcentagem em hélice de estruturas de schistocinas a partir de espectro de dicroísmo circular	90
Tabela 5 - Parâmetros termodinâmicos do experimento ITC dos peptídeos em LUVs aniônicos	94
Tabela 6 - Sumário de estatísticas estrutural de experimentos NOESY	110

## Lista de abreviaturas

2D - bidimensional  
Ac - Anticorpo  
Acc - acurácia  
AMP - peptídeo antimicrobiano  
APD - base de dados de peptídeos antimicrobianos  
ATCC - do inglês, *American Type Culture Collection*  
bLF - lactoferrina bovina  
CD - Dicroísmo circular  
CNBr - brometo de cianogênio  
D<sub>2</sub>O - deutério  
DMPC - dimiristoil glicerofosfolina  
DMSO - dimetilsulfóxido  
DOPC - dioleol glicerofosfolina  
DPC-d<sub>38</sub> - dodecilsulfocolina deuterada  
DPC - dodecil sulfocolina  
DSS - 2,2-dimetil-2-silapentano-5-sulfonato  
EC - número da comissão de enzimas  
ECP - proteína catiônica eosinofílica  
FN - falso negativo  
FP - falso positivo  
H/D - Hidrogênio-Deutério  
HIV - vírus da imunodeficiência humana  
hLF - lactoferrina humana  
HSQC - espectroscopia de correlação quantum-singular heteronuclear  
ITC - Calorimetria de Titulação Isotérmica  
LPS - lipopolissacarídeo  
LUV - vesículas unilamelares grandes  
mAc - anticorpo monoclonal  
MALDI - ionização e dessorção por laser assistida por matriz  
MBC - concentração bactericida mínima  
MCC - coeficiente de correlação de *Matthews*  
MD - dinâmica molecular  
MeJA - metil jasmonato  
MIC - Concentração inibitória mínima  
MRSA - *Staphylococcus aureus* resistente à meticilina  
MS/MS - espectroscopia de massa *in tandem*  
MS - espectroscopia de massa  
Mtb - *Mycobacterium tuberculosis*  
NCCLS - Comitê Nacional de Padrões Clínicos e Laboratoriais

NOE - efeito Overhauser nuclear  
NOESY - espectroscopia de efeito Overhauser nuclear  
ORF - janela aberta de leitura  
PC - fosfatidilcolina  
PE - fosfatidiletanolamina  
PG - fosfatidilglicerol  
POPC - fosfatidilcolina  
POPG - fosfatidilglicerol  
QSAR - relação quantitativa estrutura-atividade  
RMN - Ressonância magnética nuclear  
RMSD - desvio médio quadrático  
SAXS - Espalhamentos de raios X a baixo ângulo  
SDS - dodecil sulfato de sódio  
SmKI-1 - Inibidor do tipo kunitz de *Schistosoma mansoni*  
SVM - máquina de vetor de suporte  
TFE - 2,2,2-trifluoroetanol  
TOCSY - espectroscopia de correlação total  
TOF - tempo de voo  
VN - verdadeiro negativo  
VP - verdadeiro positivo  
 $\Delta G$  - energia livre de ligação  
 $\Delta H$  - entalpia  
 $\Delta S$  - entropia

## Sumário

CAPÍTULO 1. Criptídeos antimicrobianos - Uma revisão	20
1.1 Introdução	20
1.2 Criptídeos antimicrobianos	20
1.2.1 Intracelulares	20
Ubiquitina	21
Autofagia	22
Histonas	24
1.2.2 Extracelulares	29
Apolipoproteína	29
Albumina	31
Proteínas do transporte de oxigênio	33
Anticorpos antimicrobianos	36
Grânulos secretórios	38
Enzimas	42
Lactoferrinas	44
Coagulação	45
Colágeno	49
Outros	49
1.3 Métodos de identificação	51
1.3.1 Proteômica acoplada a identificação de atividade antimicrobiana	51
1.3.2 Identificação dos criptídeos in silico	53
1.3.3 Identificação por métodos bioquímicos	54
1.4 Considerações Finais	54
HIPÓTESE	63
OBJETIVOS	65
CAPÍTULO 2. Identificação e Caracterização das schistocinas	66
2.1 Introdução	66
2.2 Material e Métodos	69
2.2.1 Preditor	69
2.2.2 Métricas de performance dos modelos	72
2.2.3 Análises de sequências	72
2.2.4 Ensaios antimicrobianos	72
2.2.5 Espectroscopia de Dicroísmo Circular	73
2.2.6 Titulação de Calorimetria Isotérmica (ITC)	74
2.2.7 Ressonância Magnética Nuclear (RMN)	75
2.2.8 Análise de dados de RMN e cálculo das estruturas	75
2.2.9 Troca de Hidrogênio-Deutério por RMN	76
2.2.10 Sistemas lipídicos para simulação	77

2.2.11	Sistemas peptídeo-lipídeo para simulação	77
2.2.12	Simulações de dinâmica molecular	78
2.3	Resultados	79
2.3.1	Encrypted, uma plataforma para obter criptídeos antimicrobianos	79
2.3.2	Análises in silico	81
2.3.3	Síntese e análise bioquímica das schistocinas	85
2.3.4	Atividade antimicrobiana das schistocinas	88
2.3.5	Análises de padrão de estrutura secundária das schistocinas em ambientes miméticos de membrana	89
2.3.6	Calorimetria de Titulação Isotérmica entre schistocinas e LUVs	93
2.3.7	Cálculo de Estrutura por RMN	96
2.3.8	Troca de Hidrogênio-Deutério entre as schistocinas e micelas de DPC	112
2.3.9	Efeito dos estruturas de RMN na interação com micelas zwitteriônicas	116
2.3.10	Comportamento das estruturas de RMN das schistocinas em bicamadas lipídicas	118
2.4	Discussão	120
2.5	Referências	125
<b>CAPÍTULO 3. Acetilação N-terminal e amidação C-terminal usando CcpNmr Analysis e Aria2.3</b>		<b>131</b>
3.1	Introdução	131
3.2	Protocolos	133
3.2.1	Protocolos modificados no CcpNmr Analysis	133
3.2.2	Protocolos modificados no Aria2.3	134
3.3	Procedimentos	142
4.4	Considerações finais	146
4.5	Referências	147
<b>CAPÍTULO 4. Conclusão Geral</b>		<b>149</b>
4.1	Considerações Finais	149
4.2	Referências	151
<b>ANEXOS</b>		<b>152</b>

## **CAPÍTULO 1. Criptídeos antimicrobianos - Uma revisão**

### **1.1 Introdução**

No campo de estratégias para produção de novas moléculas bioativas, os peptídeos se apresentam como moléculas versáteis, uma vez que o tamanho, a carga líquida e a hidrofobicidade são variáveis. A labilidade dessas moléculas é um fator vantajoso, quando em comparação com moléculas maiores, como as proteínas. Neste campo, o desenho de novos peptídeos passa por estratégias que ajustam e regulam as propriedades físico-químicas para serem específicas para diferentes tipos de microrganismos, ou melhor ainda, propriedades que criam moléculas com amplo espectro de atividade. Entre os tipos de peptídeos bioativos mais estudados estão os peptídeos antimicrobianos (1), antitumorais (2), antiinflamatórios (3), cicatrizantes (4), inibidores de canais iônicos (5), entre vários outros.

A origem dos peptídeos pode ser de dois tipos: os peptídeos podem ser sintetizados por vias ribossômicas ou podem ser oriundos de proteólises enzimáticas, sendo assim denominados criptídeos (6–8). Vários fragmentos proteicos com atividade antimicrobiana, os criptídeos antimicrobianos, já foram isolados de alimentos, após proteólise por enzimas gástricas, tais como os criptídeos da lactoferrina, da hemoglobina bovina, do músculo e de histonas de peixes, hidrolisados pela pepsina gástrica (9–12), os criptídeos de proteínas de leite bovino, tais como a  $\alpha$ -lactoalbumina e a lactoferrina bovina (13,14), e os criptídeos da ovalbumina hidrolisados pela tripsina (15). Outros são encontrados em animais e plantas que coabitam com seres humanos, ou que até os parasitam, tais como o tomate (16). Nesta revisão, os criptídeos antimicrobianos foram detalhados e pormenorizados, destacando principalmente a origem, as sequências e os alvos desses criptídeos. Além disso, eles foram classificados de acordo com a proteína de origem e pela localização da proteína no sistema (intracelular e extracelular).

### **1.2 Criptídeos antimicrobianos**

#### **1.2.1 Intracelulares**



## Ubiquitina

A ubiquitina, uma proteína ligada a processos de degradação de proteínas, não possui como atividade intrínseca uma ação antimicobacteriana. Entretanto, um criptídeo gerado pela degradação da ubiquitina durante os processos de autofagia ligados à infecção por *Mycobacterium tuberculosis* (Mtb), demonstrou possuir atividade antimicrobiana. O criptídeo foi denominado Ub2<sup>1</sup> e possui atividade contra Mtb com valor de concentração inibitória mínima (MIC) de 5 µM (17). O trabalho de Alonso e colaboradores (2007) demonstrou que esses fragmentos são gerados pela ação de catepsinas presentes no autofagossomo e que outro fragmento peptídico também é liberado, o criptídeo Ub1<sup>2</sup>, mas sem ação antimicobacteriana. Um estudo anterior já havia descrito esses criptídeos e sua atividade antimicrobiana. O criptídeo Ub2, oriundo da ubiquitina de eritrócitos bovinos e liberado pela ação da endoproteínase Glu-C (EC 3.4.21.19), possui atividade contra as bactérias gram-positivas *M. luteus* (5 µM), *B. megaterium* (4 µM), *E. coli* (20 - 50 µM). Além disso, esse criptídeo também apresenta atividade antifúngica contra fungos filamentosos, na faixa de 1-30 µM contra *A. fumigatus*, *N. haematococca*, *F. culmorum*, *F. oxysporum*, *T. mentagrophytes* e *T. viride*; e contra células fúngicas *C. albicans*, *C. tropicalis*, *C. neoformans* e *C. glabrata*, com MIC entre 15-20 µM (18) (Figura 1a).

No fluido amniótico de humanos foi encontrado um outro criptídeo de ubiquitina, derivado de Ub1, o criptídeo AFP-1<sup>3</sup> (Figura 1a). Sua sequência possui 18 resíduos de aminoácidos. O criptídeo AFP-1 foi capaz de inibir o crescimento de vários microorganismos, como as bactérias *E. coli* e *S. typhimurium*, com valor de MIC de 32 µM, *L. monocytogenes* e *S. aureus*, com valor ainda menor de MIC (8 µM), além dos fungos *Botrytis cinerea* e *Fusarium oxysporum*, com o valor de MIC de 16 µM, *Aspergillus fumigatus* (MIC de 64 µM), *Neurospora crassa* (MIC de 32 µM), *C. albicans* e *Cryptococcus neoformans*, com o menor valor de MIC (8 µM). Outra vantagem desses criptídeos é que eles não apresentam atividade hemolítica em eritrócitos. Por fim, vale ressaltar que o material amniótico foi obtido de mulheres grávidas que foram submetidas à amniocentese transabdominal (19).

Além dos criptídeos gerados pela degradação da ubiquitina, enzimas associadas a essa proteína também são alvos proteolíticos da holoenzima proteassoma, como a ligase associada à ubiquitina (E3) (Figura 1b). O estudo de Candido-Ferreira e colaboradores descreve o

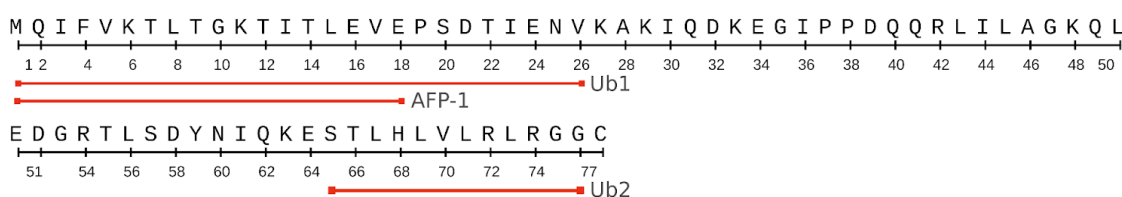
<sup>1</sup> Sequência da Ub2: STLHLVLRRLGG

<sup>2</sup> Sequência do Ub1: MQIFVKLTGKTITLEVEPSDTIENVKAKIQDK

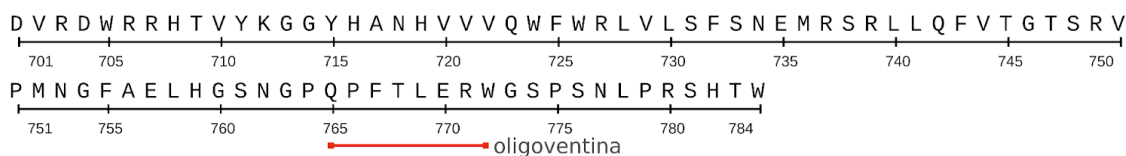
<sup>3</sup> Sequência do AFP-1: MQIFVKLTGKTITLEVE

criptídeo oligoventina<sup>4</sup>, um peptídeo antimicrobiano isolado da aranha *Phoneutria nigriventer* e que é oriundo da enzima Nedd4 E3 ligase (EC 2.3.2.26). O criptídeo foi identificado em uma fração cromatográfica de ovos da aranha *Phoneutria*. Ele demonstrou atividade contra bactérias gram-positivas *Micrococcus luteus* (47.2–94.5  $\mu$ M), *Staphylococcus aureus* (94.5–188.9  $\mu$ M) e *Staphylococcus epidermidis* (94.5–188.9  $\mu$ M), além de atividade antifúngica contra o fungo *Candida albicans* (94.5–188.9  $\mu$ M) e contra a bactéria gram-negativa *Serratia marcescens* (23.53–47.2  $\mu$ M) (20).

a) ubiquitina



b) E3 ubiquitin-protein ligase NEDD4



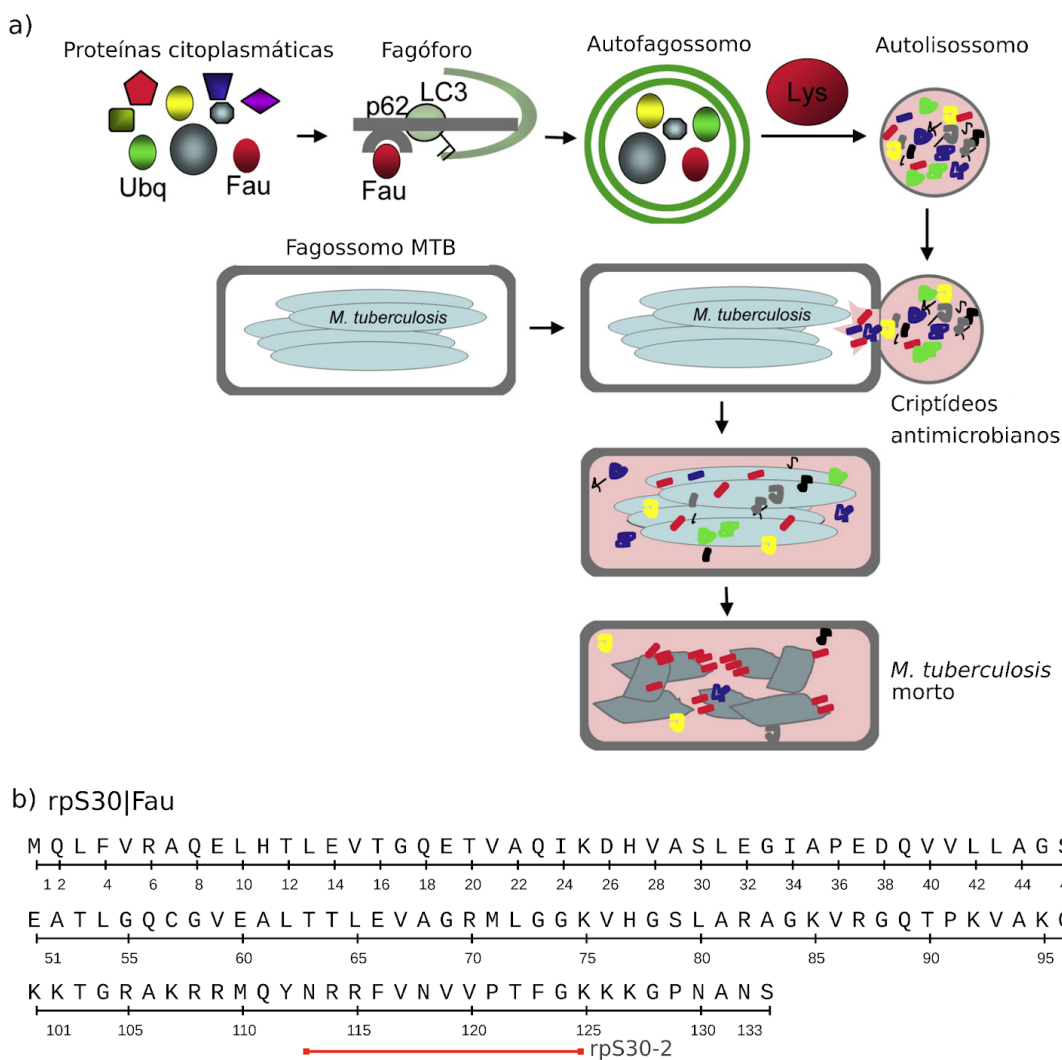
**Figura 1. Criptídeos da ubiquitina.** a) As sequências estão destacadas pelos traços: Ub1, Ub2, AFP-1. A sequência da ubiquitina está destacada em preto (XP\_005218033.1); b) A proteína E3 ubiquitin-protein ligase NEDD4 isoforma X5 (XP\_042144816.1) em preto o criptídeo oligoventina destacado pelo traço laranja.

### Autofagia

Um evento celular importante que pode ser fonte de criptídeos antimicrobianos é a autofagia. A degradação de moléculas protéicas ou organelas, num processo de reciclagem celular ou de proteção contra os organismos invasores, pode culminar na formação de autofagossomos, organelas de membranas duplas que destroem o conteúdo aprisionado pela acidificação do meio. Os processos celulares, tais como a redução dos níveis de insulina e redução de aminoácidos, se aproveitam dos produtos da lise de autofagossomos (21). Funcionalmente, os peptídeos liberados deste processo podem apresentar bioatividade e, alguns, apresentar atividade antimicrobiana. Ponpuak e colaboradores (2010) observaram que as proteínas capturadas pela proteína p62, como a proteína ribossomal de rpS30 precursora de

<sup>4</sup> Sequência da oligoventina: QPFSLERW

FAU e a ubiquitina, são degradadas em autofagossomos e originam peptídeos antimicrobianos, em processos ligados à destruição do patógeno bacteriano multidroga resistente *Mtb* (22) (Figura 2a). Dos produtos da lise da enzima catepsina L (EC 3.4.22.15), a proteína rpS30 gerou três fragmentos peptídicos, e o criptídeo rpS30-2<sup>5</sup> apresentou atividade bactericida contra o *Mtb* H37Rv (Figura 1b).



**Figura 2 - Criptídeos do rpS30|Fau.** a) Adaptado da publicação *Immunity*, volume 32, Ponpuak et al., “*Delivery of Cytosolic Components by Autophagic Adaptor Protein p62 Endows Autophagosomes with Unique Antimicrobial Properties*”, páginas 329-341, 2010. (22) com permissão da Elsevier: 5324270977176; b) Tradução da sequência gênica da proteína rpS30|Fau de *Mus musculus* (NT\_039684). Observação: A sequência da proteína possui uma valina na posição 117 que não foi descrita no peptídeo rpS30-2 como caracterizado originalmente no artigo.

<sup>5</sup> Sequência do rpS30-2: NRRFNVVPTFGK

## Histonas

As proteínas histonas também são fontes de criptídeos. A porção N-terminal da histona H2A, por exemplo, após análise de identificação de similaridade com outros peptídeos antimicrobianos previamente descritos, mostrou-se como potencial fonte de criptídeos. A molécula hiposina<sup>6</sup> foi o primeiro criptídeo de histona H2A identificado (2-52) (Figura 3a). Presente nas escamas do peixe *Hippoglossus hippoglossus L.*, o peptídeo tem o N-terminal acetilado, que auxilia na estabilidade peptídica. Sua sequência tem altíssima similaridade com outros criptídeos antimicrobianos, como a buforina e a parasina I, com identidade superior a 90% das suas sequências. A concentração inibitória mínima dos criptídeos contra bactérias foi de: 1,3-2,5 µM contra *Bacillus subtilis*, 40-80 µM contra *Enterococcus faecalis*, *Enterococcus avium* e *Staphylococcus aureus*, 0,6-1,2 µM contra *Lactobacillus coryniformis* e *Listeria ivanovii*, 10 µM contra *Staphylococcus epidermidis*, 2,5 µM contra *E. coli*, 20-40 µM contra *Pseudomonas sp.*, 10-20 µM contra *Shigella sonnei* e 2,5-20 µM contra *Vibrio sp.* (23).

O criptídeo H2A(1-39)<sup>7</sup> de 39 resíduos de aminoácidos, oriundo da histona H2A de moluscos *Chlamys farreri*, após ensaios de difusão em ágar, demonstrou atividade contra as bactérias *M. luteus*, *V. splendidus*, *V. anguillarum* e *V. vulnificus* (24) (Figura 3b). Outro animal, o sapo asiático *Bufo bufo gargarizans*, também origina criptídeos a partir da histona H2A, as buforina I<sup>8</sup> e um derivado direto, a buforina II<sup>9</sup>. A buforina I é liberada da histona H2A pela ação da enzima pepsina gástrica (EC 3.4.23.1) (12) (Figura 3c). A atividade antimicrobiana da buforina II é potencializada em comparação com a buforina I, com valores de MIC que variam de 1-4 µM e 4-8 µM, respectivamente, contra as bactérias gram-positivas *Bacillus subtilis*, *Staphylococcus aureus*, *Streptococcus mutans*, *Streptococcus pneumoniae* e *Pseudomonas putida*, as bactérias gram-negativas, *Escherichia coli*, *Salmonella typhimurium* e *Serratia sp.*, e os fungos *Candida albicans*, *Cryptococcus neoformans* e *Saccharomyces cerevisiae* (25).

O criptídeo parasina I<sup>10</sup> tem a sequência primária com 95% de identidade com o criptídeo buforina I. O criptídeo parasina I é liberado da histona H2A do peixe *Parasilurus asotus* pela ação da enzima catepsina-D (EC 3.4.23.5). Interessantemente, a forma ativa da

---

<sup>6</sup> Sequência da hiposina: Ac-SGRGKTGGKARAKAKTRSSRAGLQFPVGRVHLLLRKGN YAHRVGAG APVYL

<sup>7</sup> Sequência do H2A(1-39): MSGRGKGGKVKGKAKSRSSRAGLQFPVGRIHLLLRKGN Y

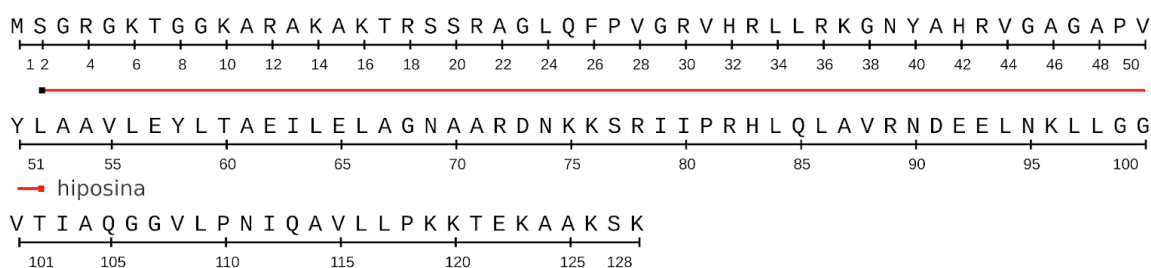
<sup>8</sup> Sequência da buforina I: AGRGKQGGKVRKAKTRSSRAGLQFPVGRVHLLLRKGN Y

<sup>9</sup> Sequência da buforina II: TRSSRAGLQFPVGRVHLLLRK

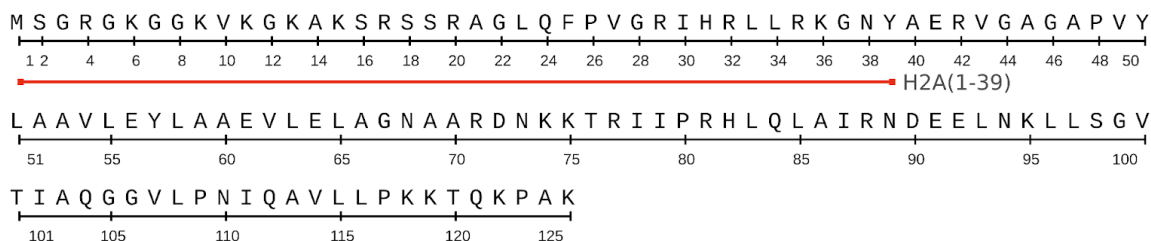
<sup>10</sup> Sequência da parasina I: KGRGKQGGKVRKAKTRSS

catepsina-D só é liberada frente à injúria na escama do peixe, o que libera o criptídeo parasina I (26). A troca da alanina-1 na sequência da buforina I pela lisina-1 na sequência da parasina I, e a remoção da porção C-terminal, que estava presente na buforina I, potencializou a atividade antimicrobiana do criptídeo contra bactérias gram-positivas, *Bacillus subtilis*, *Staphylococcus aureus*, *Streptococcus mutans* e *Pseudomonas putida* (1 à 2  $\mu\text{M}$ ), contra bactérias gram-negativas, *Escherichia coli*, *Salmonella typhimurium* e *Serratia sp.* (1 à 4  $\mu\text{M}$ ) e contra os fungos, *Cryptococcus neoformans*, *Saccharomyces cerevisiae* e *Candida albicans* (1 à 2  $\mu\text{M}$ ) (27).

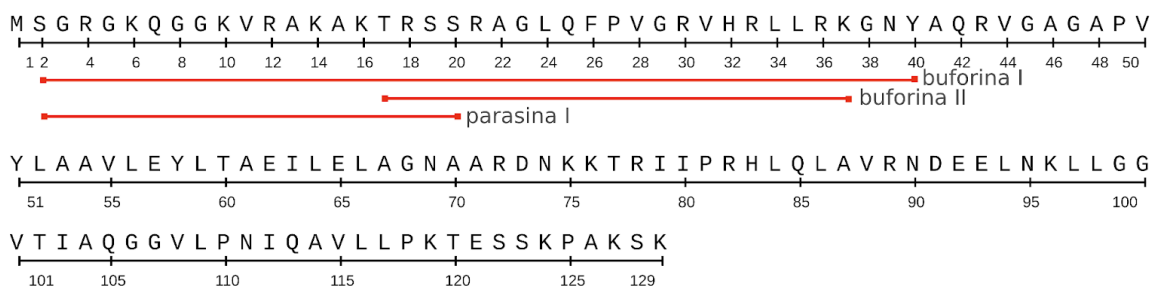
a) histona H2A [*Hippoglossus hippoglossus*]



b) histona H2A [*Azumapecten farreri*]



c) histona H2A [*Bufo gargarizans*]



**Figura 3. Criptídeos da histona H2A.** a) Histona H2A de *Hippoglossus hippoglossus* (XP\_034433694.1) e o criptídeo hiposina; b) Histona H2A de *Azumapecten farreri* (ABD76392.1) e o criptídeo H2A(1-39); c) Histona H2A de *Bufo gargarizans* (AAB57777.1) e os criptídeos buforina I, buforina II e parasina I. O criptídeo parasina I contém uma lisina na posição 1, e não uma serina. Entretanto, a sequência da histona H2A de *Parasilurus asotus* não está anotada.

Alguns criptídeos de histona desenvolveram uma forma própria de serem sintetizados. Conceitualmente pode parecer confuso, mas sua sequência peptídica ainda é parte de proteínas maiores, como neste exemplo, da histona H4. Ao invés de serem gerados pela proteólise da proteína, a sequência genômica das histonas apresentam janelas de leitura aberta (ORF, do inglês *open reading frame*) próprias, que sintetizam apenas as porções peptídicas necessárias da proteína. São exemplos os seguintes criptídeos de ratos e bovinos referentes à histona H4: HN<sub>r</sub><sup>11</sup>, Pro-HN<sub>r</sub><sup>12</sup> e HN<sub>b</sub><sup>13</sup>. Além disso, esses criptídeos podem encontrar análogos oriundos da proteólise, tais como, os criptídeos H4-(86-100)<sup>14</sup> e H4-(84-102)<sup>15</sup> (28) (Figura 4). O criptídeo HN foi primeiramente isolado da medula adrenal de bovinos (29). Já a versão do criptídeo encontrada em ratos está presente em vários tecidos, tais como o pulmão, o baço e o cérebro (30). Comparando a atividade dos criptídeos HN<sub>r</sub> e H4-(86-100), ambos apresentaram atividade similar contra *E. coli* e *Bacillus subtilis* (MIC ~ 5 µg/mL). Entretanto, o criptídeo H4-(86-100) apresentou maior atividade contra *Pseudomonas aeruginosa* (H4-(86-100): MIC de 1,75 µg/mL vs HN<sub>r</sub>: MIC de 3,5 µg/mL) e *Staphylococcus aureus* (H4-(86-100): MIC de 5,83 µg/mL vs > HN<sub>r</sub>: MIC de 30 µg/mL). A potência bactericida do análogo H4-(84-102) decaiu com a adição dos dois resíduos em cada extremidade da sequência (28).

---

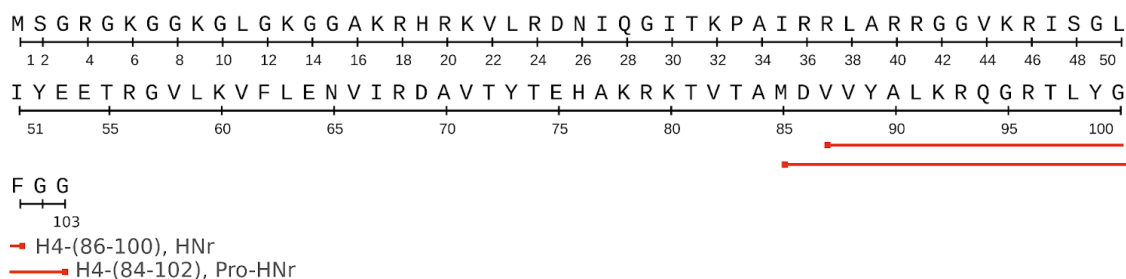
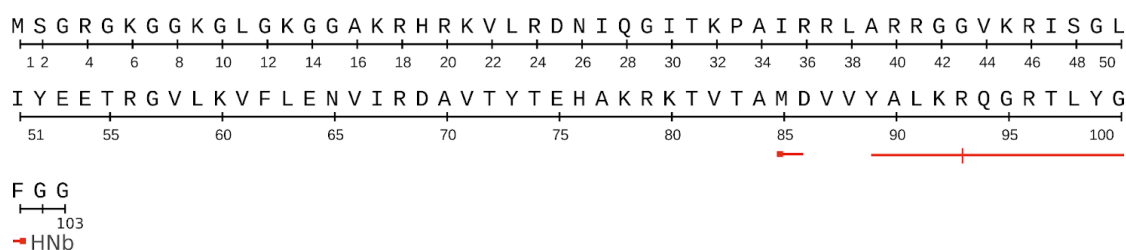
<sup>11</sup> Sequência do HN<sub>r</sub>: VVYTLKRQGRTLYGF

<sup>12</sup> Sequência do Pro-HN<sub>r</sub>: MDVVYTLKRQGRTLYGFGG

<sup>13</sup> Sequência do HN<sub>b</sub>: MNYALKGQGRTLYGF

<sup>14</sup> Sequência do H4-(86-100): VVYALKRQGRTLYGF

<sup>15</sup> Sequência do H4-(84-102): MDVVYALKRQGRTLYGFGG

a) histona H4 [*Rattus norvegicus*]b) histona H4 [*Bos taurus*]

**Figura 4. Criptídeos da histona H4.** a) Histona H4 de *Rattus norvegicus* (NP\_001116941.1) e criptídeos HNr, Pro-HNr, H4-(86-100) e H4-(84-102); b) Histona H4 de *Bos taurus* (NP\_776305.1) e o criptídeo HNb, cuja sequência não apresenta os resíduos valina 87 e 88 e uma modificação de arginina 93 para glicina.

Outro criptídeo antimicrobiano, o oncorincina II<sup>16,17</sup>, extraído do peixe *Oncorhynchus mykiss*, possui 69 resíduos de aminoácidos e atividade contra bactérias gram-positivas e gram-negativas (Figura 5a). Após uma busca global associada a um alinhamento de sequências, o peptídeo foi identificado como parte da porção C-terminal de histonas H1 (139-207), entre as posições 138-154. Além do mais, o criptídeo é fosforilado em dois resíduos. A atividade antibacteriana obteve concentrações inibitórias mínimas de 0,2-0,4 µM contra as bactérias *M. luteus* e *P. citreus*, e 0,4-0,8 µM contra *E. coli* e *L. anguillarum* (31). Outro criptídeo derivado da histona H1 foi encontrado no muco do salmão do atlântico *Salmo salar*, um peptídeo de 30 resíduos de aminoácidos, denominado SAMP H1<sup>18</sup> (Figura 5b). Interessantemente, ele é encontrado no peixe em sua forma inativa, mas sua atividade é potencializada pela exposição à enzima peptidil-prolina cis-trans-isomerase (EC 5.2.1.8). O SAMP H1 tem alta atividade antimicrobiana contra *E. coli* e *Bacillus subtilis* (MIC de 2 µM),

<sup>16</sup> Sequência da oncorincina II: KAVAANKSPKAKKPKATPKKAAKSPKVKKPAAAKKAASKPKKAT KAAKPKAAKPKAAKAKKAAPKKK

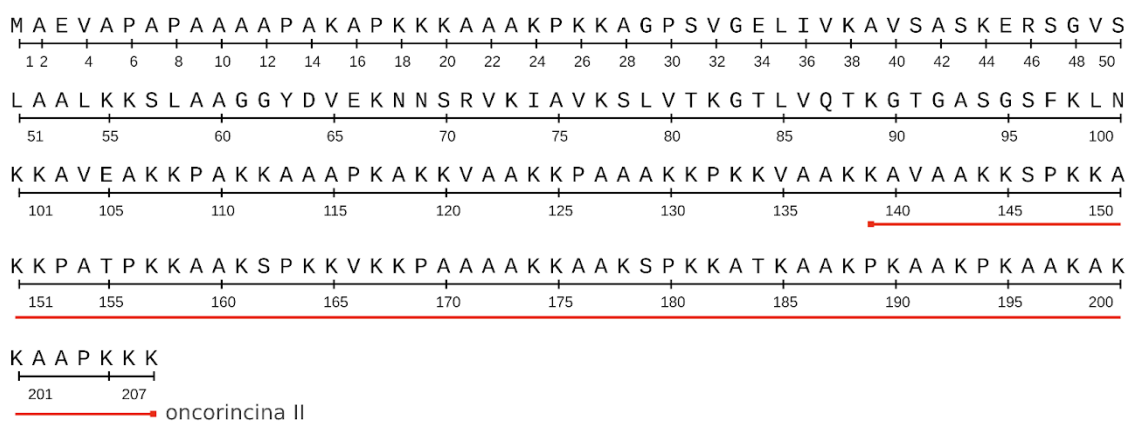
<sup>17</sup> Observação: sequência depositada na base de dados APD, referente à sequência da histona H1 depositada no UNIPROT (P06350). Entretanto, o BLASTP não consegue identificar a sequência na sua base de dados.

<sup>18</sup> Sequência do SAMP H1: AEVAPAPAAAAPAKAPKKKAAAKPKKAGPS

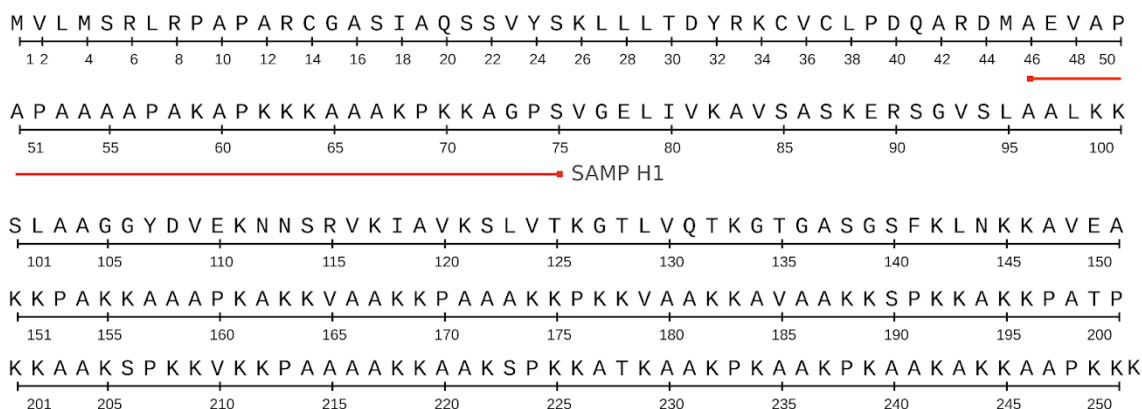
contra *Aeromonas salmonicida subsp. salmonicida*, *Salmonella enterica serovar Typhimurium* e *Listeria ivanovii* (MIC de 4  $\mu$ M), e contra *Vibrio anguillarum* (MIC de 1  $\mu$ M) (32).

O criptídeo H5(61-90)<sup>19</sup> da histona H5 de eritrócitos de galinha mostrou-se ativo contra *Listeria monocytogenes*, com valor de MIC de 23,1  $\mu$ g/mL, e contra *Pseudomonas aeruginosa*, com MIC de 16  $\mu$ g/mL (Figura 5c). A própria proteína histona apresenta atividade antimicrobiana e antibiofilme. Utilizando propriedades físico-químicas de peptídeos antimicrobianos conhecidos, seis peptídeos antimicrobianos na sequência da histona H5 foram preditos, mas apenas o criptídeo H5(61-90) apresentou atividade eficaz contra as bactérias. Em água, o peptídeo tem padrão de estrutura de transição entre randômico e alfa-hélice. Em micelas de SDS, o peptídeo apresenta-se completamente como hélice (33).

a) histona H1 [*Oncorhynchus mykiss*]

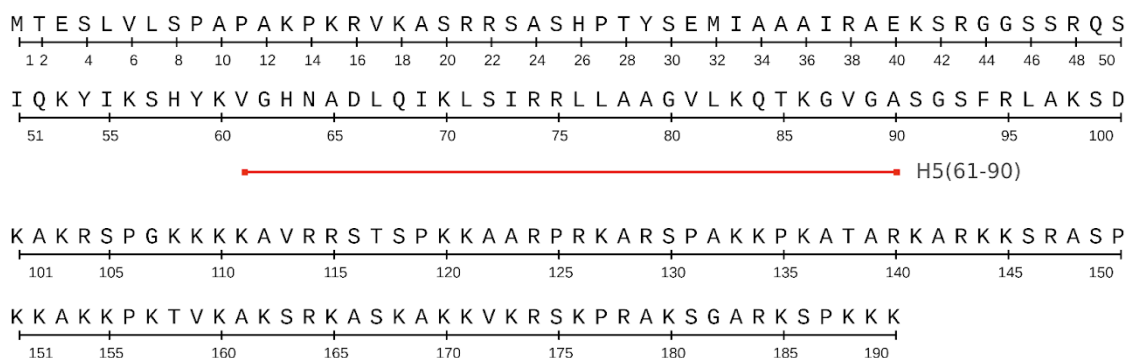


b) histona H1 [*Salmo salar*]



<sup>19</sup> Sequência do H5(61-90): VGHNADLQIKLSIRLLAAGVLKQTKGVGA



c) histona H5 [*Gallus gallus*]

**Figura 5. Criptídeos da histona H1 e H5.** a) Histona H1 de *Oncorhynchus mykiss* (CAB37646.1) e o criptídeo oncorincina II; b) Histona H1 de *Salmo salar* (XP\_045547289.1) e o criptídeo SAMP H1; c) Histona H5 de *Gallus gallus* (NP\_001038138.1) e o criptídeo H5(61-90).

## 1.2.2 Extracelulares

### Apolipoproteína

Uma vantagem na identificação de criptídeos antimicrobianos em células de mamíferos é a baixa propensão de estas moléculas serem citotóxicas, uma vez que suas sequências primárias não são estranhas ao hospedeiro. Este é o caso dos peptídeos derivados da apolipoproteína B, como o ApoB 887-923 (r(P)ApoB<sub>L</sub><sup>20</sup>) e o ApoB 887-911 (r(P)ApoB<sub>S</sub><sup>21</sup>), que possuem atividade antimicrobiana, antibiofilme e cicatrizante (Figura 6a). Estes criptídeos apresentam valores de MIC de 20 µM contra *P. aeruginosa* e 5 e 2,5 µM, respectivamente, contra *Bacillus globigii* (34). Estes criptídeos formam estruturas do tipo amilóide, cuja macroestrutura está diretamente ligada à ação antibactericida. A estrutura tridimensional desses peptídeos é do tipo folha-β, o que facilita a formação da fibra, e está em contraste ao estado da arte de peptídeos antimicrobianos com estruturação em α-hélices (35).

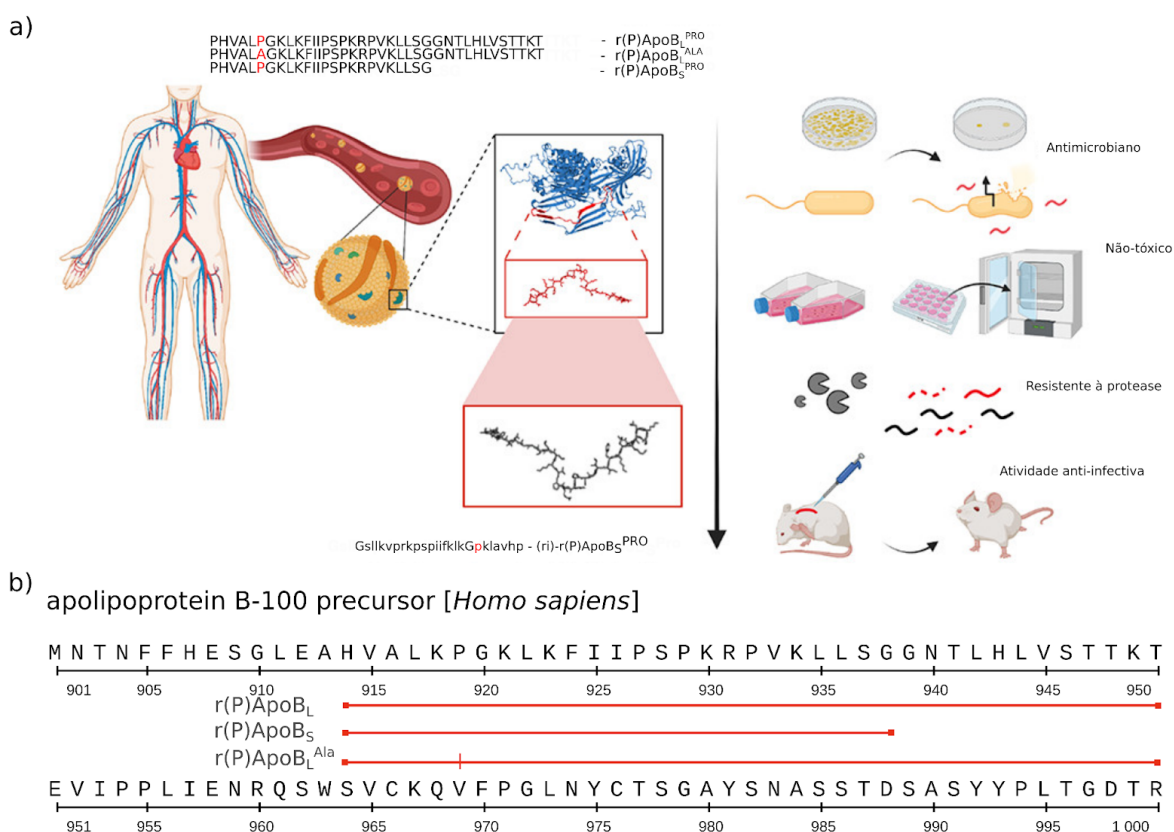
Em um estudo posterior, o grupo de Gaglione e colaboradores (2022) expressou o criptídeo com uma modificação na posição 7 (Pro7Ala), denominado r(P)ApoB<sub>L</sub><sup>Ala</sup> (Figura 6b)<sup>22</sup>. A atividade antimicrobiana desses criptídeos foi analisada contra outros patógenos: *Staphylococcus epidermidis*, *Staphylococcus aureus*, *Acinetobacter baumannii*, *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Enterococcus faecalis* e *Pseudomonas aeruginosa*. Os

<sup>20</sup> Sequência do r(P)ApoB<sub>L</sub>: PHVALKPGKLFKFIIPSPKRPVKLLSGGNTLHLVSTTKT

<sup>21</sup> Sequência do r(P)ApoB<sub>S</sub>: PHVALKPGKLFKFIIPSPKRPVKLLSG

<sup>22</sup> Sequência do r(P)ApoB<sub>L</sub><sup>Ala</sup>: PHVALKAGKLFKFIIPSPKRPVKLLSGGNTLHLVSTTKT

criptídeos apresentam maior atividade contra bactérias gram-negativas, como *A. baumannii*, *K. pneumoniae* e *P. aeruginosa*, cujos valores de MIC variam entre 2,5 a 5  $\mu\text{M}$  (36). Para as outras cepas, os valores de MIC variaram entre 5 a 20  $\mu\text{M}$ . Os ensaios para avaliar se os peptídeos encriptados são capazes de induzir resistência com tratamentos prolongados demonstraram que *A. baumannii*, *S. epidermidis* e *S. aureus* não se tornam resistentes aos criptídeos da apolipoproteína B. A síntese de um derivado retro-inverso (sequência invertida) D-enantiomérico do criptídeo r(P)ApoB<sub>S</sub>, o (ri)-r(P)ApoB<sub>S</sub><sup>23</sup>, para aumentar a resistência deste peptídeo à enzimas proteolíticas na circulação de mamíferos, não modificou sua atividade contra cepas bacterianas. Os valores de MIC encontrados contra bactérias gram-negativas e gram-positivas foram os mesmos de r(P)ApoB<sub>S</sub> (36).



**Figura 6. Criptídeos da apolipoproteína B.** a) Resumo gráfico adaptado/traduzido do trabalho de Cesaro, et al. *Synthetic Antibiotic Derived from Sequences Encrypted in a Protein from Human Plasma* (2022). Copyright 2022 American Chemical Society.; b) Apolipoproteína B-100 de *Homo sapiens* (AAA35549.1) e os criptídeos r(P)ApoB<sub>L</sub>, r(P)ApoB<sub>S</sub> e r(P)ApoB<sub>L</sub><sup>Ala</sup>. Os criptídeos foram sintetizados pelos pesquisadores com uma

<sup>23</sup> Sequência do (ri)-r(P)ApoB<sub>S</sub>: D-GSLLKVPPKPSIIFKLKGPKLAVHP

prolina adicional na posição 1 dos criptídeos. O criptídeo r(P)ApoB<sub>L</sub><sup>Ala</sup> contém uma troca da prolina 119 da apolipoproteína por uma alanina.

### Albumina

Um criptídeo endógeno, que possui atividade contra microrganismos, precisa ser gerado e liberado em grandes quantidades. A soroalbumina humana é uma boa fonte, pois libera diariamente cerca de 14 gramas de peptídeos (37). A revisão de Kragh-Hansen (2018) destaca que 28 fragmentos são liberados em hemo filtrados, 16 fragmentos na bile, 20 peptídeos na urina e 63 peptídeos no fluido cerebrospinal (38).

Apesar de a maioria dos criptídeos derivados da albumina terem sido relatados com bioatividade distinta da antimicrobiana (como anti hipertensivos (39) e inibidores de proteases (40)), estudos recentes têm demonstrados a existência de criptídeos antimicrobianos, como é o caso do criptídeo K13L<sup>24</sup>, que possui atividade antifúngica contra *C. albicans* e *C. glabrata* (41) (Figura 7a). Existem estudos que não se preocupam em caracterizar a atividade antimicrobiana por peptídeo, e sim a atividade de um conjunto (um “pool”) de peptídeos. É o caso do estudo de Al-Mohammadi e colaboradores (2020) que identificaram que a proteólise da albumina de ovos de galinha culmina com 44 peptídeos, sendo 17 dipeptídeos e 27 peptídeos entre 3 e 9 resíduos de aminoácidos. As melhores atividades desse conjunto de peptídeos foi encontrada contra as bactérias *Listeria monocytogenes*, *Bacillus cereus*, *S. aureus*, *S. typhimurium*, *S. pyogenes*, e *Klebsiella oxytoca*, com MIC de 100 µg/mL (42). Não é o escopo deste trabalho, mas vale mencionar outros criptídeos oriundos da soroalbumina humana que não possuem atividade antimicrobiana, como os criptídeos cabina-A1<sup>25</sup>, cabina-A2<sup>26</sup> e albutensina humana A<sup>27</sup>, inibidores da catepsina B (40).

A ovoalbumina gera diferentes criptídeos antimicrobianos a depender da enzima utilizada. A tripsina (EC 3.4.21.4) cliva a ovoalbumina em regiões com resíduos de arginina e lisina, liberando os criptídeos antimicrobianos ODT1<sup>28</sup>, ODT2<sup>29</sup>, ODT3<sup>30</sup>, ODT4<sup>31</sup> e ODT5<sup>32</sup> (Figura 7b). A quimotripsina (EC 3.4.21.1), por sua vez, cliva ligações peptídicas entre

<sup>24</sup> Sequência do peptídeo K13L: KKLVAASQAALGL

<sup>25</sup> Sequência da cabina-A1: SLHTLF

<sup>26</sup> Sequência da cabina-A2: FQNAL

<sup>27</sup> Sequência da albutensina A: AFKAWAVAR

<sup>28</sup> Sequência do ODT1: SALAM

<sup>29</sup> Sequência do ODT2: SALAMVY

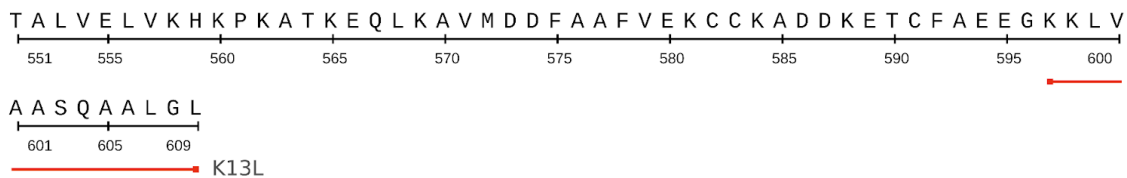
<sup>30</sup> Sequência do ODT3: YPILPEYLQ

<sup>31</sup> Sequência do ODT4: ELINSW

<sup>32</sup> Sequência do ODT5: NVLQPSS

resíduos de triptofano, leucina, tirosina e fenilalanina e libera os criptídeos antimicrobianos ODC1<sup>33</sup>, ODC2<sup>34</sup> e ODC3<sup>35</sup>. Os valores de MIC não foram calculados ou não foram disponibilizados. A atividade antimicrobiana foi demonstrada pelo valor de  $\log N_0/N_1$ , onde  $N_0$  é a quantidade de colônias no momento anterior à adição dos peptídeos, e o  $N_1$  é a quantidade de colônias duas horas depois de entrarem em contato com os peptídeos. Os criptídeos ODC1, ODC2, ODC3, ODT3 e ODT5 demonstraram atividade contra *B. subtilis*. Os outros fragmentos peptídicos apresentaram baixa atividade contra esta cepa e outras bactérias gram-positivas e gram-negativas testadas (15).

a) albumina [*Homo sapiens*]



b) ovalbumina [*Gallus gallus*]



**Figura 7. Criptídeos da albumina.** a) Albumina de *Homo sapiens* (PDB: 6ZL1\_A) e o criptídeo K13L; b) Ovalbumina de *Gallus gallus* (AAB59956.1) e os criptídeos ODT1, ODT2, ODT3, ODT4, ODT5, ODC1, ODC2 e ODC3.

<sup>33</sup> Sequência do ODC1: AEERYPILPEYL

<sup>34</sup> Sequência do ODC2: GIIRN

<sup>35</sup> Sequência do ODC3: TSSNVMEER

### Proteínas do transporte de oxigênio

Em invertebrados, o tecido responsável pelo transporte de oxigênio para as células é a hemolinfa, e a proteína hemocianina tem o papel de carregar as moléculas de oxigênio. Em aranhas do gênero *Eurypelma* sp. e *Acanthoscurria* sp. identificou-se que a proteólise da hemocianina libera o criptídeo rondonina<sup>36</sup>, cuja atividade é específica contra as espécies do fungo *Candida*: *C. albicans*, *C. krusei*, *C. glabrata*, *C. parapsilosis*, *C. tropicalis*, *C. guilliermondii*, na faixa de 8,37 à 33,5  $\mu$ M (43) (Figura 8a).

Apesar de grande em extensão, um criptídeo antimicrobiano de 61 resíduos de aminoácidos, derivado da hemoglobina, foi encontrado no endométrio de camundongos fêmeas. O criptídeo é oriundo da cadeia alfa da hemoglobina, nas posições 32-93, aqui nomeado Hb $\alpha$ (32-93)<sup>37</sup> (Figura 8b). O peptídeo apresenta atividade contra a bactéria gram-negativa *E. coli*, mas não apresenta atividade contra a bactéria *S. aureus* e o fungo *C. albicans*. Apesar do achado, os autores não encontraram como esse criptídeo poderia ser liberado da hemoglobina, mas sugerem que a hemoglobina seja degradada nos eritrócitos e as cadeias ficariam suscetíveis a outras enzimas (44,45).

Na guelra de peixes *Ictalurus punctatus* foi identificado um criptídeo com ação antibacteriana e antiparasitária. O criptídeo Hb $\beta$ P-1<sup>38</sup> é proveniente da cadeia beta de hemoglobinas (Figura 8c). É interessante notar que a sua descoberta foi estimulada pela infecção de peixes com o parasito *Ichthyophthirius multifiliis*, o que super estimulou sua liberação. Dos três fragmentos identificados, apenas a atividade do Hb $\beta$ P-1 foi avaliada. O criptídeo apresenta atividade apenas contra as bactérias gram-negativas. O valor dos MIC encontrados foram: *Vibrio alginolyticus* (13,5  $\mu$ M), *E. coli* (3,4  $\mu$ M) e *Aeromonas hydrophila* (3,4  $\mu$ M) (46).

O primeiro criptídeo antimicrobiano identificado em bovinos foi encontrado na hemolinfa do carrapato *Boophilus microplus*. Após identificarem a atividade da hemolinfa desses invertebrados contra cepas gram-negativas e contra os fungos *A. nidulans*, *N. crassa* e *C. albicans*, o tecido foi isolado e as frações protéicas foram separadas por cromatografia líquida. O criptídeo, nomeado BHP<sup>39</sup>, contém 29 resíduos de aminoácidos em sua sequência e

<sup>36</sup> Sequência da rondonina: IIIQYEGHKKH

<sup>37</sup> Sequência do Hb $\alpha$ (32-93): FLSFPTTKTYFPHFDLSHGSAQVKGQGGKVVADALNAVAHVVD

<sup>38</sup> Sequência do Hb $\beta$ P-1: AANFGPSVFTPEVHETWQKFLNVVVAALGKQYH

<sup>39</sup> Sequência do BHP: FLSFPTTKTYFPHFDLSHGSAQVKGHGAK

é fragmento da  $\alpha$ -hemoglobina bovina (regiões 33-61) (Figura 8d). Quando avaliada a atividade isolada do criptídeo BHP, os valores de MIC encontrados foram: *M. luteus* (5  $\mu$ M), *S. epidermidis* (21  $\mu$ M), *C. albicans* (5  $\mu$ M) e *S. cerevisiae* (11  $\mu$ M). Não foi encontrada atividade para bactérias gram-negativas, como *E. coli* e *P. aeruginosa* (47).

Com o intuito de fazer uma triagem em larga escala com criptídeos antimicrobianos de hemoglobina bovina, soluções de hemoglobina foram digeridas com pepsina de porcos em um grau de hidrólise de 3%. Um total de 43 peptídeos foram encontrados, em frações isoladas e mistas. A atividade antimicrobiana desses fragmentos peptídicos foi testada contra *M. luteus*, *L. innocua*, *E. coli* e *S. enteritidis*. Dos 43, 30 apresentaram alguma atividade e foram classificados em dois grupos, C1 (com valores altos de MIC; > 10  $\mu$ M) e C2 (com valores baixos de MIC;  $\leq$  10  $\mu$ M). Os 11 criptídeos que compõem o grupo C2 são:  $\alpha$  33–66<sup>40</sup>,  $\alpha$  34–66<sup>41</sup>,  $\alpha$  33–46<sup>42</sup>,  $\alpha$  33–45<sup>43</sup>,  $\alpha$  34–46<sup>44</sup>,  $\alpha$  36–45<sup>45</sup>,  $\alpha$  37–46<sup>46</sup>,  $\alpha$  133–141<sup>47</sup>,  $\alpha$  137–141<sup>48</sup>,  $\beta$  140–145<sup>49</sup> e  $\beta$  1–13<sup>50</sup> (Figura 8d e 8e). Este estudo ressalta como diferentes porções da mesma proteína podem apresentar atividade microbiana, quando separadas e produzidas em grande quantidade (48).

Indo além, um estudo de 2011 conseguiu identificar o menor criptídeo de hemoglobina alfa, um tripeptídeo Lys-Tyr-Arg (Figura 8d). Ao realizar ensaios antimicrobianos com outros criptídeos de hemoglobina, os pesquisadores observaram que o criptídeo  $\alpha$  139–141<sup>51</sup> apresentava os menores valores de MIC: 1  $\mu$ M contra *E. coli*, *S. enteritidis*, *S. aureus* e *L. innocua* e 2  $\mu$ M contra *M. luteus*. Entretanto, o criptídeo  $\alpha$  139–141 não foi possível de ser obtido via hidrólise, apenas por síntese química (49). Por outro lado, o menor criptídeo de cadeia beta da hemoglobina de bovinos é um tripeptídeo Arg-Tyr-His. O estudo de triagem criptídica em hemoglobina de bovinos havia identificado um criptídeo da cadeia beta que foi classificado no grupo C1, o  $\beta$ 114-145<sup>52</sup> (Figura 8e). Ele foi utilizado como molécula original

<sup>40</sup> Sequência do  $\alpha$  33–66: FLSFPTTKTYFPHFDLSHGSAQVKGHGAKVAAAL

<sup>41</sup> Sequência do  $\alpha$  34–66: LSFPTTKTYFPHFDLSHGSAQVKGHGAKVAAAL

<sup>42</sup> Sequência do  $\alpha$  33–46: FLSFPTTKTYFPHF

<sup>43</sup> Sequência do  $\alpha$  33–45: FLSFPTTKTYFPH

<sup>44</sup> Sequência do  $\alpha$  34–46: LSFPTTKTYFPHF

<sup>45</sup> Sequência do  $\alpha$  36–45: FPTTKTYFPH

<sup>46</sup> Sequência do  $\alpha$  37–46: PTTKTYFPHF

<sup>47</sup> Sequência do  $\alpha$  133–141: STVLTSKYR

<sup>48</sup> Sequência do  $\alpha$  137–141: TSKYR

<sup>49</sup> Sequência do  $\beta$  140–145: LAHRYH

<sup>50</sup> Sequência do  $\beta$  1–13: MLTAEKAAVTAF

<sup>51</sup> Sequência do  $\alpha$  139–141: KYR

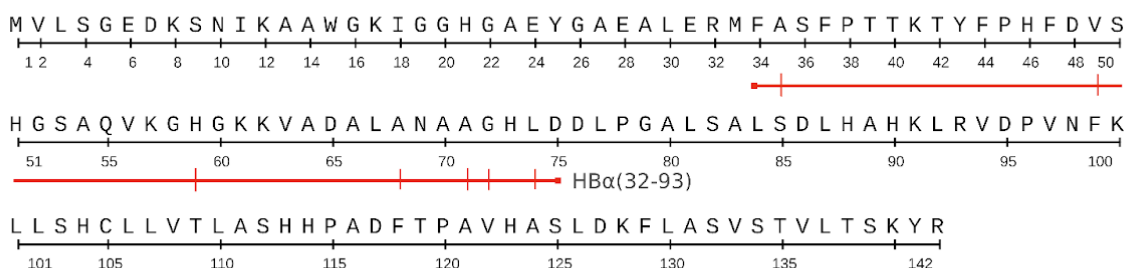
<sup>52</sup> Sequência do  $\beta$ 114-145: ARNFGKEFTPVLQADFQKVVAGVANALAHRYH

para encontrar a menor porção ativa. De acordo com os autores, as reações enzimáticas novamente envolveram a ação da pepsina de porcos (EC 3.4.23.1). Entre as quatro frações obtidas, todos os peptídeos obtiveram atividade antibacteriana com valores de MIC maiores de 18  $\mu\text{M}$ . Não fica claro o que motivou o grupo a investigar o tripeptídeo, mas como ele não foi possível de ser obtido por hidrólise, ele foi sintetizado quimicamente, o  $\beta$  143-145<sup>53</sup>. A atividade antibacteriana do “tripeptídeo”  $\beta$  143-145 foi alta: *S. enteritidis* (1  $\mu\text{M}$ ), *E. coli* (4  $\mu\text{M}$ ), *L. innocua* (1  $\mu\text{M}$ ) e *M. luteus* (4  $\mu\text{M}$ ) (50).

a) hemocianina subunidade d [*Aphonopelma californicum*]



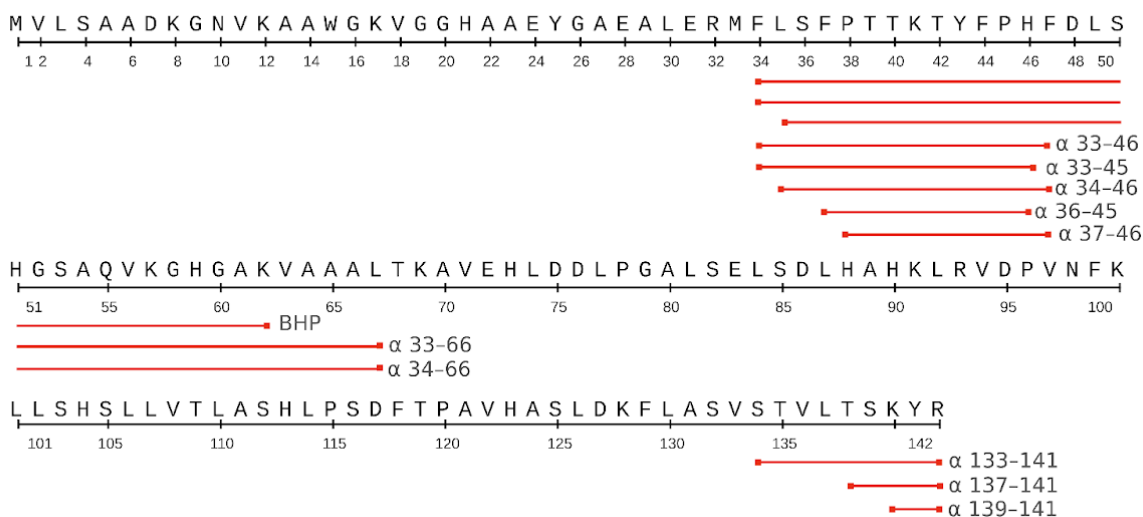
b) hemoglobina alfa [*Mus musculus*]



c) hemoglobina beta [*Ictalurus punctatus*]



d) hemoglobina alfa [*Bos taurus*]



<sup>53</sup> Sequência do  $\beta$ 143-145: RYH

e) hemoglobina beta [*Bos taurus*]



**Figura 8. Criptídeos de hemocianina e hemoglobina.** a) Porção C-terminal da proteína hemocianina subunidade D de *Aphonopelma californicum* (CAB89499.1) e o criptídeo rondonina. A hemocianina de *Eurypelma* sp. e *Acanthoscurria* não tem a sequência da hemocianina depositada e por isso foi utilizada uma sequência de outro gênero, cuja sequência encriptada é idêntica, como citado no artigo original; b) Hemoglobina alfa de *Mus musculus* (NP\_001077424.1) e o criptídeo HBα(32-93). Na sequência depositada do criptídeo, alguns resíduos estão modificados (Resíduo original/posição na proteína/resíduo no criptídeo): Ala35Leu, Val49Leu, His59Gln, Ala68Thr, Ala71Val, Gly72Ala e Leu74Asp. O motivo dessas distinções não é explicado; c) Porção C-terminal da hemoglobina beta de *Ictalurus punctatus* (NP\_001187115.1) e o criptídeo HbβP-1; d) Hemoglobina alfa de *Bos taurus* (CAB56828.1) e os criptídeos α 33–66, α 34–66, α 33–46, α 33–45, α 34–46, α 36–45, α 37–46, α 133–141, α 137–141, BHP e α 139–141; e) Hemoglobina beta de *Bos taurus* (BAJ05099.1) e os criptídeos β 140–145, β 1–13 e β114-145.

### Anticorpos antimicrobianos

Já é de conhecimento dos imunologistas, e das ciências correlatas, o papel dos anticorpos (Acs) no estímulo de respostas microbicidas por células do sistema imune inato e adaptativo (51). Entretanto, trabalhos mais recentes têm demonstrado o papel microbicida e direto de anticorpos contra microrganismos. Deste modo, os Acs microbicidas atuam sem a interferências de outras células (52).

A capacidade dos Acs monoclonais de serem identificados por outros Acs, os Acs anti idiotípicos, é uma estratégia para simular a atividade de agentes e toxinas citotóxicas e de difícil manuseio experimental (53,54). Este é o caso dos estudos com Acs anti idiotípicos da toxina assassina (do inglês, *killer toxin*) de *Pichia anomala*, que possuem atividade anti-candida. O grupo de pesquisa de Polonelli e colaboradores identificou um peptídeo nesses anticorpos que possui a mesma atividade biológica contra culturas fúngicas, o decapeptídeo P6<sup>54</sup>, cuja sequência é oriunda da região CDR do Ac (55) (Figura 9a). Como o IC<sub>50</sub> do peptídeo é da ordem de 10<sup>-4</sup> M, o grupo fez uma otimização da sequência por meio do

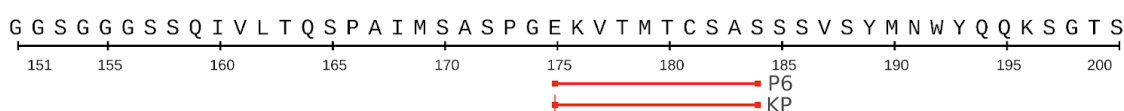
<sup>54</sup> Sequência do decapeptídeo P6: EKVTMTCSAS



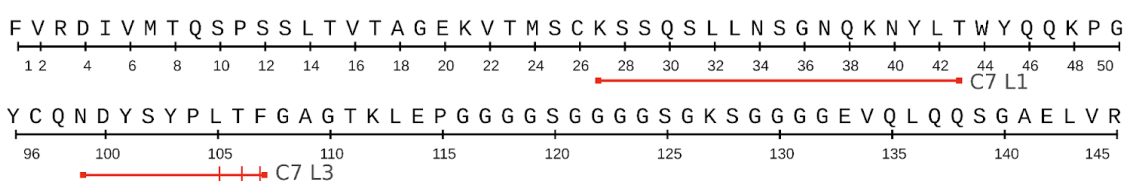
método de alanina *scan*, que substitui cada resíduo de aminoácido na sequência por alanina, e testa sua atividade. A sequência otimizada foi chamada de decapeptídeo assassino KP<sup>55</sup>, cujo IC<sub>50</sub> contra fungos *Candida* sp. é da ordem de 10<sup>-5</sup> e 10<sup>-6</sup> M (55) (Figura 9a). O peptídeo foi capaz de inibir a replicação do vírus da imunodeficiência humana (HIV-1) em células mononucleares humanas *in vitro* e *ex vivo* (56), atividade antiviral contra o vírus influenza A (57), atividade antileishmanicida (58) e antitoxoplásmica (59).

Outros criptídeos derivados de anticorpos monoclonais anti-candida foram testados a partir do mAc C7 e mAc pc42. As atividades antifúngicas demonstradas foram: C7 L1<sup>56</sup> (EC<sub>50</sub> de 1,52x10<sup>-5</sup> M), C7 L3<sup>57</sup> (EC<sub>50</sub> de 2,02x10<sup>-5</sup> M) (Figura 9b), pc42 L2<sup>58</sup> (EC<sub>50</sub> de 1,2x10<sup>-5</sup> M) (Figura 9c) e C7/pc42 H2<sup>59</sup> (EC<sub>50</sub> de 4,2x10<sup>-6</sup> M) (Figura 9d). Além da atividade antifúngica, os criptídeos demonstraram atividade antiviral contra HIV e atividade antitumoral (60).

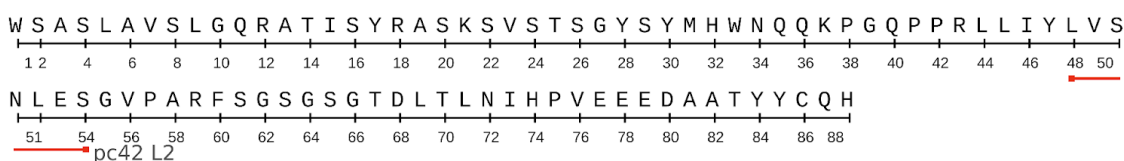
a) CD5L-OKT3scFv-CD14 [construto sintético]



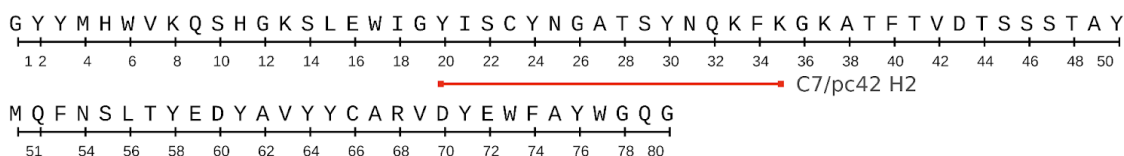
b) anticorpo cadeia FV [construto sintético]



c) anti-pneumococical polissacarídeo capsular - cadeia leve de IG [*Homo sapiens*]



d) região variável de cadeia pesada de IG [*Mus musculus*]



**Figura 9. Criptídeos de anticorpos.** a) Porção da proteína CD5L-OKT3scFv-CD14 [construto sintético] (ADN42857.1) e os criptídeos P6 e KP. A sequência de KP sofreu uma modificação *in vitro*, trocando o glutâmico 1 do peptídeo P6 por uma alanina; b) Cadeia FV de imunoglobulina (PDB: 2AP2|A) e os criptídeos C7 L1 e C7 L3. O criptídeo C7 L3

<sup>55</sup> Sequência do decapeptídeo KP: AKVTMTCSAS

<sup>56</sup> Sequência do C7 L1: KSSQSLNLSGNQKNYLT

<sup>57</sup> Sequência do C7 L3: NDYSYPRSR

<sup>58</sup> Sequência do pc42 L2: LVSNLES

<sup>59</sup> Sequência do C7/pc42 H2: YISCYNGATSYNQKFK

apresenta as seguintes alterações: Leu105Arg, Thr106Ser e Phe107Arg; c) Cadeia leve de imunoglobulina anti-pneumocócica de *Homo sapiens* (AAL04518.1) e o criptídeo pc42 L2; d) Região variável da cadeia pesada de imunoglobulina *Mus musculus* (AAQ74683.1) e o criptídeo C7/pc42 H2. As sequências dos anticorpos originais não são citadas nos trabalhos e assim, foram retiradas de buscas globais associadas a alinhamentos locais (BLAST).

### Grânulos secretórios

Os grânulos secretórios bovinos, gerados em células cromafins da medula adrenal, liberam diversas proteínas e peptídeos, tais como catecolaminas, encefalinas e cromograninas (61). Alguns deles são criptídeos antimicrobianos, como é o caso do criptídeo CGB (614-626)<sup>60</sup>, oriundo da proteína cromogranina B (62). O criptídeo CGB (614-626) apresenta atividade antibacteriana contra *M. luteus* e *B. megaterium*, com MIC de 1 µM, mas não apresenta atividade contra *E. coli*. Curiosamente, os pesquisadores relataram que a sequência peptídica apresenta similaridade com a família de peptídeos antimicrobianos cecropina, com resíduos conservados na posição 2, 5, 6, 7, 8 e 12 do CGB (614-626).

Este grupo de pesquisadores identificou outros peptídeos antimicrobianos em grânulos secretórios, tais como o criptídeo da porção C-terminal 209-237 da proteína pró encefalina-A, que recebeu o nome de encelitina<sup>61</sup> (Figura 10). A propósito, este criptídeo é bifosforilado, nas posições Ser13 e Ser15 e, assim como o criptídeo GCB (614-626), apresenta atividade contra *M. luteus* e *B. megaterium*, com MIC na faixa de 0,2-0,4 µM, mas não possui atividade contra *E. coli* (63) (Figura 10a). Um estudo posterior do grupo demonstrou a liberação de um peptídeo análogo ao encelitina, o peptídeo B<sup>62</sup>, que possui mais dois resíduos na sua porção C-terminal. Além da bifosforilação, o peptídeo B é monooxidado em um dos seus resíduos de metionina. A atividade do peptídeo B contra *M. luteus* é menor do que a atividade da encelitina: 5 µM vs 0,4 µM, respectivamente (64) (Figura 10b). Com maior espectro de ação, o criptídeo da porção 173-194 da cromogranina A de bovinos, nomeado cromacina<sup>63</sup>, apresenta atividade contra bactérias gram-positivas e gram-negativas (Figura 10c). Este criptídeo é um fragmento resultante da ação da enzima endoproteínase Lys-C (EC 3.4.21.50), que cliva proteínas ao encontrar resíduos de lisina. Além disso, o criptídeo possui modificações pós-traducionais nas posições 1 (O-glicosilação da tirosina) e 14 (fosforilação da serina) (65).

<sup>60</sup> Sequência do GCB (614-626): QKIAEKFSGTRRG

<sup>61</sup> Sequência da encelitina: FAEPLPSEEEGES(fosforilado)YS(fosforilado)KEVPEMEKRYGGFM

<sup>62</sup> Sequência do peptídeo B: FAEPLPSEEEGES(fosforilado)YS(fosforilado)KEVPEMEKRYGGFMRF

<sup>63</sup> Sequência da cromacina: Y(O-glicosilado)PGPQAKEDSEGPS(fosforilado)QGPASREK

Além das células cromafins, os eosinófilos são proteínas do sistema imune que também secretam proteínas. Uma delas é a proteína catiônica eosinofílica (ECP), que atua na parede e membranas de bactérias. Utilizando conhecimentos de propensão de atividade antimicrobiana, a porção N-terminal demonstrou ter sequências ditas antimicrobianas e três sequências foram construídas a partir deste molde. Os fragmentos ECP-(1-45)<sup>64</sup> e ECP-(24-45)<sup>65</sup> apresentaram atividade antibacteriana, quando testadas *in vitro*, contra *E. coli*, com MIC de 0,75 e 10 µM, além de atividade contra *S. aureus*, com MIC de 0,75 e 2 µM, respectivamente (66) (Figura 10d). Entretanto, o trabalho não deixa claro ou sugere como esses criptídeos poderiam ser clivados da proteína ECP. Por outro lado, o trabalho de Sánchez e colaboradores (2011) vai adiante, e testa a atividade de uma protease na proteína ECP desnaturada, a endoproteínase Lys-C (EC 3.4.21.50), e de um composto químico, o brometo de cianogênio (CNBr). Enquanto a endoproteínase Lys-C liberou um fragmento de 38 resíduos de aminoácidos (ECP-(0-38)<sup>66</sup>), a clivagem com CNBr liberou três fragmentos na porção N-terminal (ECP-(F11-M27), ECP-(R28-W35)<sup>67</sup> e ECP-(F11-W35)<sup>68</sup>) (Figura 10d). O criptídeo ECP-(0-38) apresenta atividade contra *E. coli* com um MIC de 4 µM. Com relação aos outros fragmentos, os criptídeos ECP-(R28-W35) e ECP-(F11-W35) apresentaram maior redução da viabilidade de culturas de *E. coli* (67).

---

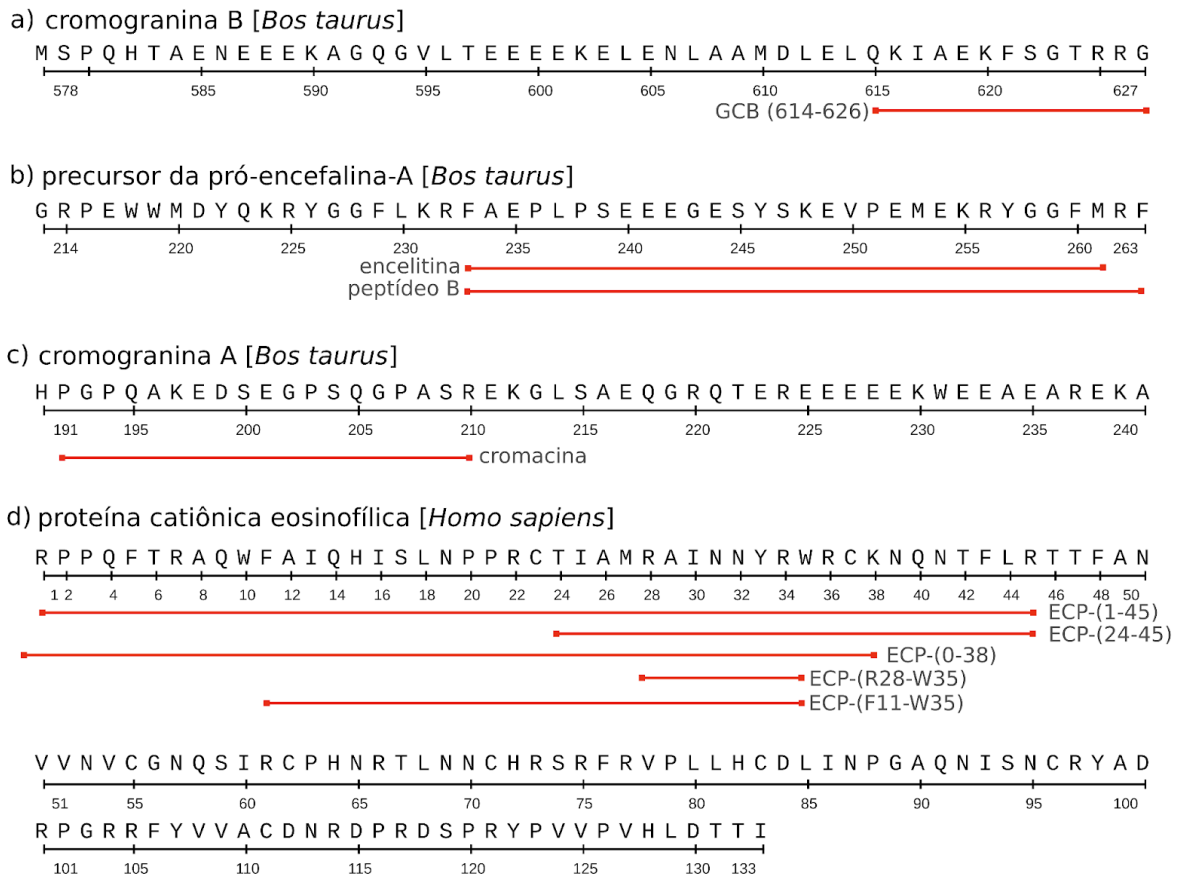
<sup>64</sup> Sequência da ECP-(1-45): RPPQFTRAQWFAIQHISLNPPRCTIAMRAINNYRWRCKNQNTFLR

<sup>65</sup> Sequência da ECP-(24-45): TIAMRAINNYRWRCKNQNTFLR

<sup>66</sup> Sequência do ECP-(0-38): MRPPQFTRAQWFAIQHISLNPPRCTIAMRAINNYRWRCK

<sup>67</sup> Sequência do ECP-(R28-W35): RAINNYRW

<sup>68</sup> Sequência do ECP-(F11-W35): FAIQHISLNPPRCTIAMRAINNYRW



**Figura 10. Criptídeos de grânulos secretórios.** a) Cromogranina B de *Bos taurus* (CAA39109.1) e o criptídeo CGB (614-626); b) Precursor da pró-encefalina-A de *Bos taurus* (NP\_776566.1) e os criptídeos encelitina e peptídeo B; c) Cromogranina A de *Bos taurus* (AAB21297.1) e o criptídeo cromacina; d) Proteína catiônica eosinofílica de *Homo sapiens* (pdb|1DYT|B).

Outras proteínas que são secretadas são as quimiocinas. A família das quimiocinas contém várias proteínas com o papel de sinalização quimiotática e um papel antimicrobiano. Analisando as sequências dessas proteínas e comparando com peptídeos antimicrobianos já descritos, os pesquisadores perceberam que a porção C-terminal dessas proteínas contém a maior parte da carga positiva. Foi com esses indícios que Su e colaboradores (2022) decidiram sintetizar criptídeos de 6 quimiocinas. O criptídeo omCK11-31<sup>69</sup>, oriundo da proteína omCK11 (quimiocina CK11 de *Oncorhynchus mykiss*) possui 31 resíduos de aminoácidos, além de quatro lisinas e cinco argininas. Os outros criptídeos são: omCCL28-like-23<sup>70</sup> (*O. mykiss* CCL28-like), drCCL27b-24<sup>71</sup> (*Danio rerio* CCL27b),

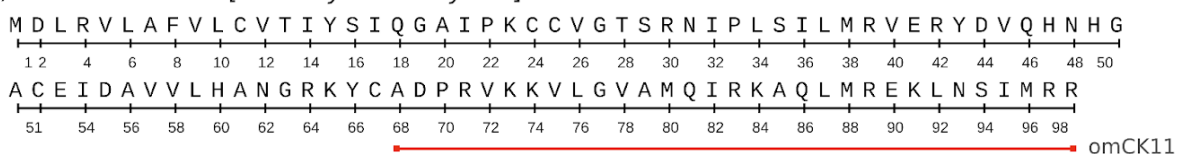
<sup>69</sup> Sequência do omCK11: ADPRVKKVLGVAMQIRKAQLMREKLNSIMRR

<sup>70</sup> Sequência do omCCL28-like23: AHPKVKRKLKRMQRKKGKQKRH

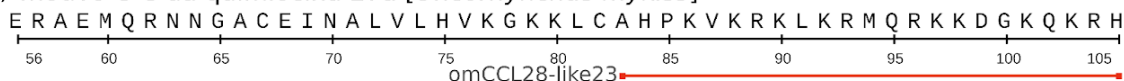
<sup>71</sup> Sequência do drCCL27b-24: ATPKMEQFLQKLMKRMARLKASAV

lcCCL28-25<sup>72</sup> (*L. crocea* CCL28), smCCL27b-25<sup>73</sup> (*Scophthalmus maximus* CCL27b), trCCL28-29<sup>74</sup> (*Takifugu rubripes* CCL28), e chCCL27a-26<sup>75</sup> (*Clupea harengus* CCL27a) (Figura 11). Os peptídeos apresentaram atividade contra bactéria gram-negativas: *E. coli* (MIC de 6,25 - 25 µg/mL), *K. pneumoniae* (25 à > 50 µg/mL), *A. hydrophila* (21 à > 50 µg/mL), *P. aeruginosa* (12,5 - 50 µg/mL); contra bactérias gram-positivas: *S. aureus*, *S. agalactiae* e *S. pneumoniae* (6,25 - 25 µg/mL), MRSA (de 6,25 - 25 µg/mL, exceto smCCL27b-25) (68).

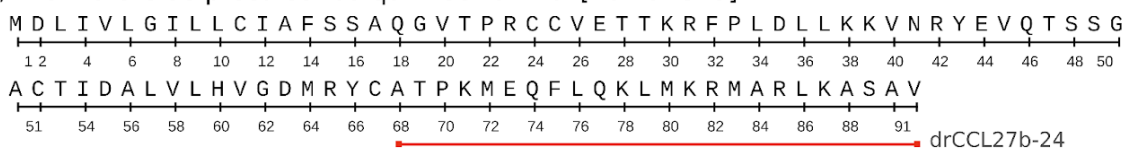
a) não identificado [*Oncorhynchus mykiss*]



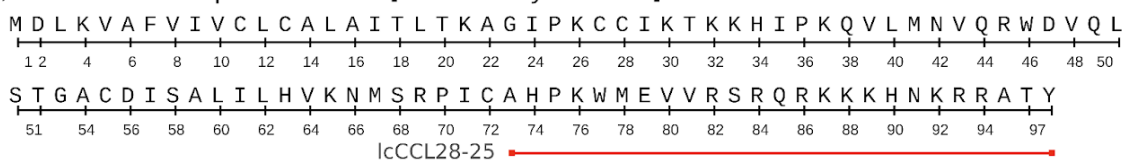
b) motivo C-C da quimiocina 27a [*Oncorhynchus mykiss*]



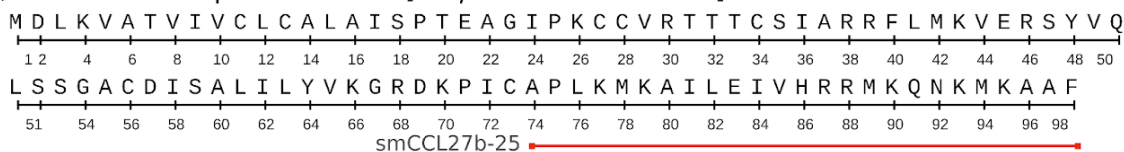
c) motivo C-C do precursor da quimiocina 27b [*Danio rerio*]



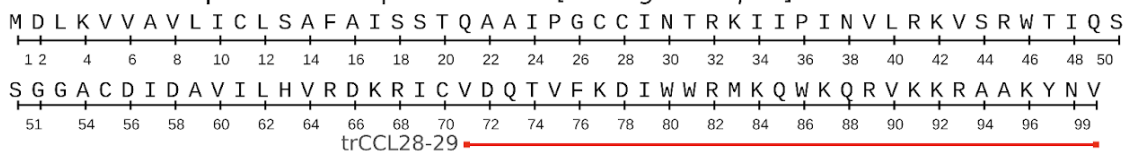
d) motivo C-C da quimiocina 28 [*Larimichthys crocea*]



e) motivo C-C da quimiocina 27b [*Scophthalmus maximus*]



f) motivo C-C do precursor da quimiocina 28 [*Takifugu rubripes*]



<sup>72</sup> Sequência do lcCCL28-25: AHPKWMEVVRSRQRKKKHNRATY

<sup>73</sup> Sequência do smCCL27b-25: APLKMKAILEIVHRRMKQNKMKAAF

<sup>74</sup> Sequência do trCCL28-29: VDQTVFKDIWWRMKQWKQRVKKRAAKYNV

<sup>75</sup> Sequência do chCCL27a-26: AHPRIKHLDLLMKKKRHLAMKIAAP

g) motivo C-C da quimiocina 27a [*Clupea harengus*]



**Figura 11. Criptídeos de quimiocinas.** a) Proteína não identificada de *Oncorhynchus mykiss* (CDQ77591.1) e o criptídeo omCK11; b) Quimiocina 27a de *Oncorhynchus mykiss* (XP\_021462169.2) e o criptídeo omCCL28-like23; c) Quimiocina 27b de *Danio rerio* (NP\_001121740.1) e o criptídeo drCCL27b-24; d) Quimiocina 28 de *Larimichthys crocea* (XP\_010731843.1) e o criptídeo lcCCL28-25; e) Quimiocina 27b de *Scophthalmus maximus* (XP\_035496366.1) e o criptídeo smCCL27b-25; f) Quimiocina 28 de *Takifugu rubripes* (NP\_001266955.1) e o criptídeo trCCL28-29; g) Quimiocina 27a de *Clupea harengus* (XP\_012669911.2) e o criptídeo chCCL27a-26.

### Enzimas

Em humanos, a enzima 11-hidroxiesteroide desidrogenase-1 do tipo beta (HSD3) (EC 1.1.1.146) pode ser clivada e liberar o criptídeo GVF27<sup>76</sup> (Figura 12a). A proteína de origem, a HSD3, tem poucos estudos que avaliam sua atividade. Entretanto, sua função é dita similar a outras desidrogenases da família, conhecidas por catalisar a conversão de glicocorticóides inativos em formas ativas (69). Este criptídeo estrutura-se em  $\alpha$ -hélice na presença de membranas e possui ampla faixa de espectro em atividade, tanto antimicrobiana quanto antibiofilme. O valor do MIC contra cepas gram-positivas de MRSA, *B. globigii*, *B. licheniformis* e *S. aureus* foi de 5  $\mu$ M, enquanto que os valores de MIC para bactérias gram-negativas, *E. coli*, *Pseudomonas aeruginosa* e *S. enteritidis*, variou entre 10-20  $\mu$ M. A inibição de biofilmes de *E. coli* foi observada em uma concentração de 32  $\mu$ M. Apesar de a sequência não ter sido identificada *in vitro*, a predição desse criptídeo indica que sua existência *in vivo* seria dependente da ação da enzima elastase neutrofílica (EC 3.4.21.37) (70).

Uma enzima bastante estudada, a lisozima (EC 3.2.1.17), também é fonte de criptídeos antimicrobianos. O hidrolisado da lisozima de ovos da galinha, obtidos por digestão péptica e e trípica, gerou 20 fragmentos peptídicos e dois se mostraram criptídeos antimicrobianos, o peptídeo 98-108<sup>77</sup> e o peptídeo 15-21<sup>78</sup> (Figura 12b). O peptídeo 98-108 (400  $\mu$ g/mL) apresentou atividade contra *E. coli*, inibindo em 70% o crescimento de colônias, e o peptídeo

<sup>76</sup> Sequência do GVF27: GVFPWRFRLCLLRRWLPRPAWFIR

<sup>77</sup> Sequência do peptídeo 98-108: IVSDGDGMNAW

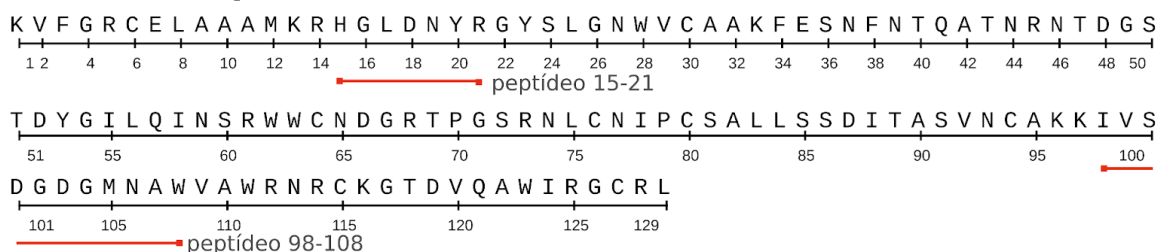
<sup>78</sup> Sequência do peptídeo 15-21: HGLDNYR

15-21 (400 µg/mL) apresentou atividade contra *S. aureus*, inibindo 35% do crescimento das colônias (71). A lisozima do leite humano também é fonte de criptídeos. Cinco criptídeos foram estudados por Ibrahim, Imazato, Ono (2011), o H1<sup>79</sup>, H2<sup>80</sup>, HLH<sup>81</sup>, H2-S12<sup>82</sup> e H2-S13<sup>83</sup> (Figura 12c). É interessante notar que a lisozima humana não apresenta atividade antimicrobiana contra bactérias gram-negativas, mas os seus criptídeos sim. Os criptídeos HLH, H1 e H2 apresentaram suas maiores atividades contra as cepas *E. coli*, *S. typhimurium*, *Bord. bronchiseptica*, *K. pneumoniae* e *P. aeruginosa*. Já em relação às bactérias gram-positivas, os criptídeos HLH e H1 apresentaram atividade contra *S. aureus*, *S. epidermidis* e *S. zooepidemicus*. Os criptídeos H2, H2-S12 e H2-S13 apresentaram atividade moderada contra *S. aureus*, *S. epidermidis*, *S. zooepidemicus*, *P. acnes* e *C. minutissimum* (72).

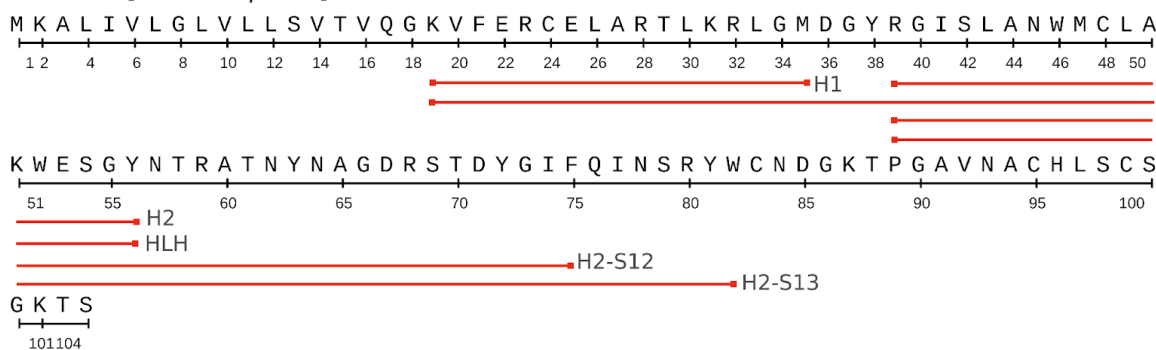
a) Proteína hidroxisteróide 11-beta-desidrogenase tipo-1 isoforma g [*Homo sapiens*]



b) lisozima [*Gallus gallus*]



c) lisozima [*Homo sapiens*]



**Figura 12. Criptídeos de enzimas.** a) Hidroxiesteróide 11-beta-desidrogenase tipo-1 isoforma g de *Homo sapiens* (NP\_001254797.1) e o criptídeo GVF27; b) Lisozima de *Gallus*

<sup>79</sup> Sequência do H1: KVFERCELARTLKRLGM

<sup>80</sup> Sequência do H2: RGISLANWMCLAKWESGY

<sup>81</sup> Sequência do HLH: KVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGY

<sup>82</sup> Sequência do H2-S12: RGISLANWMCLAKWESGYNTRATNYNAGDRSTDYGIF

<sup>83</sup> Sequência do H2-S13: RGISLANWMCLAKWESGYNTRATNYNAGDRSTDYGIFQINSRYW

*gallus* (prf|630460A) e os criptídeos peptídeo 15-21 e peptídeo 98-108; c) Lisozima de *Homo sapiens* (KAI2566955.1) e os criptídeos H1, H2, HLH, H2-S12 e H2-S13.

### Lactoferrinas

A porção N-terminal da lactoferrina humana (hLF) dá origem a alguns criptídeos. A proteína hLF está presente em mucosas e nos neutrófilos, além de ter atividade antimicrobiana já descrita (73,74). O primeiro trabalho descrevendo os criptídeos de lactoferrina é de 1992, onde os pesquisadores colocaram a hLF e a lactoferrina bovina (bLF) em contato com a pepsina gástrica (EC EC 3.4.23.1), que liberou criptídeos de 18 resíduos de aminoácido da porção N-terminal (hLF-18-40<sup>84</sup> e bLF-17-41<sup>85</sup>) (Figura 13). O criptídeo hLF-18-40 apresentou atividade apenas contra *E. coli*, com MIC de 18 µM, enquanto o criptídeo bLF-17-41 apresentou atividade contra bactérias gram-positivas e gram-negativas, como *E. coli* e *Staphylococcus aureus* (2 µM), *Klebsiella pneumoniae* e *Pseudomonas aeruginosa* (3 µM), e *Listeria monocytogenes* (0,3 µM) (14). Apesar da atividade antimicrobiana ter sido avaliada principalmente entre os resíduos 18-40 na hLF, a porção N-terminal mais extrema possui regiões características de peptídeos antimicrobianos, como a porção 2R-3R-4R-5R. Assim, foi visto que o criptídeo hLF(1-11)<sup>86</sup> possui ação contra bactérias gram-positivas, como *S. aureus*, *L. monocytogenes* e *K. pneumoniae*, mas não contra a bactéria gram-negativa *E. coli* (MIC não informado e doses testadas em 4 µM) (75) (Figura 13a). Outro estudo identificou a atividade antibacteriana do hLF(1-11) contra *Acinetobacter baumannii*, *in vitro* (36-72 µM) e *in vivo* (40 µg de peptídeo por 1 kg de animal) (76). Com relação à atividade antifúngica, o criptídeo hLF(1-11) consegue inibir o crescimento de *Candida albicans* com uma potência 10 vezes superior a um análogo derivado das porções 21-31 da hLF (77).

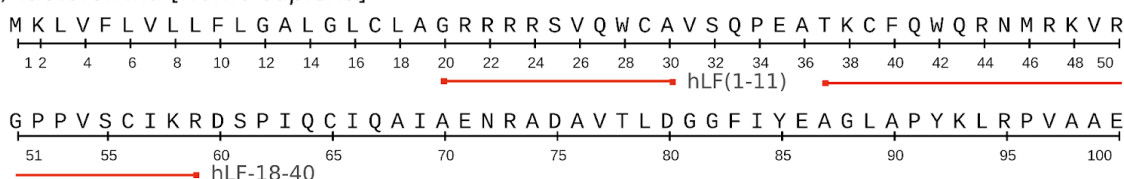
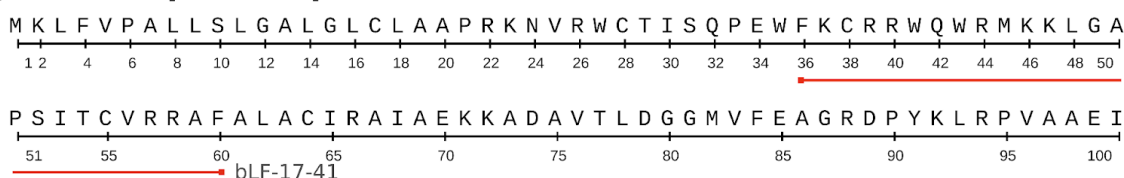
---

<sup>84</sup> Sequência do hLF-18-40: TKCFQWQRNMRKVRGPPVSCIKR

<sup>85</sup> Sequência do bLF-17-41: FKRRWQWRMKKLGAPSITCVRRAF

<sup>86</sup> Sequência do hLF(1-11): GRRRRSVQWCA



a) lactoferrina [*Homo sapiens*]b) lactoferrina [*Bos taurus*]

**Figura 13. Criptídeos de lactoferrina.** a) Porção N-terminal da lactoferrina de *Homo sapiens* (AAA59511.1) e os criptídeos hLF(1-11) e hLF-18-40; b) Porção N-terminal da lactoferrina de *Bos taurus* (AAA21722.1) e o criptídeo bLF-17-41. Numerações distintas devido a uma porção N-terminal adicional não considerada nos trabalhos originais dos criptídeos.

### Coagulação

A cascata da coagulação exerce papel na proteção do organismo contra patógenos, seja mediando a formação de coágulo ou pela ativação de células do sistema imune (78,79). A presença de um C-terminal anfipático e enovelado em  $\alpha$ -hélice foi a desconfiança inicial de que a trombina humana seria a fonte de criptídeos antimicrobianos. Como observado, a porção C-terminal da trombina humana dá origem a criptídeos, entre eles o criptídeo GKY25<sup>87</sup>, que tem ação direta e indireta na proteção do hospedeiro contra microorganismos. Além dele, o criptídeo VFR17<sup>88</sup> também apresenta atividade contra *E. coli* e *P. aeruginosa* (80) (Figura 14a). Estes criptídeos são gerados após proteólise da trombina humana pela enzima elastase neutrofílica (EC 3.4. 21.37) (81). Apesar do destaque dado para essas sequências, outros fragmentos são gerados após ação da elastase, como HVFRLKKWIKVIDQFGE e FRLKKWIKVIDQFGE. Além do mais, não apenas a elastase neutrofílica, como enzimas bacterianas podem gerar criptídeos, incluindo o próprio GKY25. A concentração inibitória mínima do criptídeo GKY25 foi testada contra vários isolados de cada cepa bacteriana e os valores encontrados foram: *E. coli* (1,2-10  $\mu$ M), *P. aeruginosa* (10-160  $\mu$ M), *S. aureus* (2,5-80  $\mu$ M), *E. faecalis* (20  $\mu$ M) e *S. pyogenes* (2,5  $\mu$ M). Ademais, o peptídeo GKY25 é conservado

<sup>87</sup> Sequência do GKY25: GKYGFYTHVFRLLKKWIKVIDQFGE

<sup>88</sup> Sequência do VFR17: VFRLKKWIKVIDQFGE

em outros fatores da coagulação humana e até em várias espécies. Os criptídeos GKY25(X)<sup>89</sup> (Figura 14b), de fator FX e GKY23<sup>90</sup> (Figura 14c), do fator FIX apresentam atividade contra as bactérias *P. aeruginosa*, *E. coli*, *S. aureus* e *B. subtilis*, além de atividade contra o fungo *C. albicans*. Os fragmentos ERP23<sup>91</sup> (Figura 14d), do fator FXI, e EQP25<sup>92</sup> (Figura 14e), da calicreína, também tiveram suas atividades microbianas testadas, mas não apresentaram atividade antibacteriana contra bactérias gram-positivas e gram-negativas, possivelmente devido à carga líquida negativa resultante desses peptídeos (80). O trabalho de Hansen e colaboradores (2015) foi adiante e avaliou como o criptídeo GKY25 atua na resposta contra patógenos. O peptídeo é capaz de aderir-se aos componentes lipopolissacarídeos de membranas de bactérias gram-negativas e atuar na ativação de células da resposta imune inata, como monócitos e macrófagos (82).

A observação de que estes peptídeos contêm uma sequência motivo, X-[PFY]-X-[AFILV]-[AFY]-[AITV]-X-[ILV]-X(5)-W-[IL]-X (sendo X um resíduo qualquer), e que ela poderia estar presente em outras moléculas, fez um grupo de pesquisadores investigar a presença de criptídeos antimicrobianos em serino proteases da família S1. Apenas a análise *in silico* buscando peptídeos usando a sequência motivo em serino proteases identificou 68 fragmentos. Da vasta triagem de peptídeos que encontraram atividade antimicrobiana, os criptídeos com melhor desempenho foram fragmentos da trombina (THRB<sup>93</sup>) (Figura 14a), fatores IX (FIX<sup>94</sup>) (Figura 14f) e X (FX<sup>95</sup>) (Figura 15a), plasminogênio (PLMN<sup>96</sup>) (Figura 15b), proteína C (PROC<sup>97</sup>) (Figura 15c), proteína ApoA (APOA<sup>98</sup>) e HABP2 (HABP2<sup>99</sup>) (Figura 15d), granzima B (GRAB<sup>100</sup>) e H (GRAH<sup>101</sup>) associados a células T citotóxicas e as calicreínas 5 (KLK5<sup>102</sup>) (Figura 15e), 8 (KLK8<sup>103</sup>)

<sup>89</sup> Sequência do GKY25(X): GKYGIYTKVTAFLKWIDRSMKTRGL

<sup>90</sup> Sequência do GKY23: GKYGIYTKVSRVNWIKKTKLT

<sup>91</sup> Sequência do ERP23: ERPGVYTNVVEYVDWILEKTQAV

<sup>92</sup> Sequência do ERP25: EQPGVYTKVAEYMDWILEKTQSSDG

<sup>93</sup> Sequência do THRB: KYGFYTHVFRLKKWIQKVID

<sup>94</sup> Sequência do FIX: KYGIYTKVSRVNWIKKTK

<sup>95</sup> Sequência do FX: KYGIYTKVTAFLKWIDRSMK

<sup>96</sup> Sequência do PLMN: KPGVYVRVSRFVTWIEGVMR

<sup>97</sup> Sequência do PROC: NYGVYTKVSRVLDWIHGHIR

<sup>98</sup> Sequência não informada

<sup>99</sup> Sequência do HABP2: RPGVYTKVTKFLNWIKATIK

<sup>100</sup> Sequência não informada

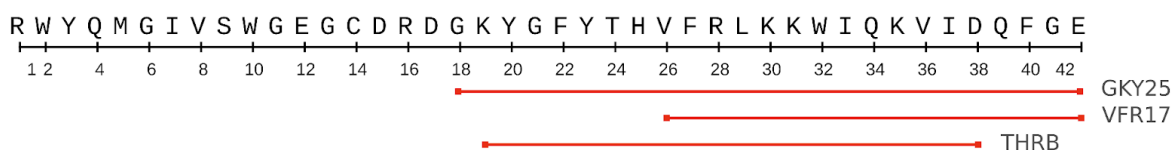
<sup>101</sup> Sequência não informada

<sup>102</sup> Sequência do KLK5: RPGVYTNLCKFTKWIQETIQ

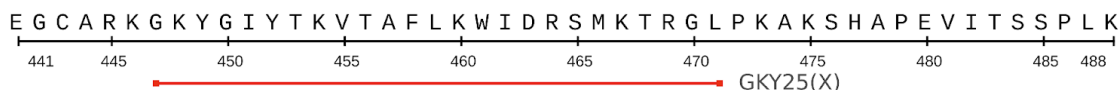
<sup>103</sup> Sequência do KLK8: KPGVYTNICRYLDWIKKIIG

(Figura 15f) e 10 (KLK10<sup>104</sup>) (Figura 15g) associadas ao epitélio. Suas atividades foram testadas em ensaios de difusão radial contra *E. coli*, *P. aeruginosa*, *S. aureus* e *C. albicans*. Os autores do trabalho sugerem que as enzimas envolvidas na liberação destes criptídeos são a elastase neutrofílica (EC 3.4. 21.37), a catepsina G (EC 3.4.21.20) e a plasmina (EC 3.4.21.7) (83). Em um estudo recente com a proteína TFPI-2, um inibidor de serino protease do tipo Kunitz proveniente do organismo *Sebastes schegellii*, o domínio C-terminal da proteína foi estudado e demonstrou atividade antimicrobiana, o criptídeo TS40<sup>105</sup>. A expressão da proteína TFPI-2 nesses organismos é aumentada frente à exposição ao patógeno *Listonella anguillarum*. Ensaio antimicrobianos com o criptídeo TS40 demonstrou que ele tem efeito bactericida em bactérias gram-positivas, como *S. aureus* (MIC de 12,5 µM) e *Streptococcus agalactiae* (MIC de 800 µM), e bactérias gram-negativas, como *L. anguillarum* (MIC de 25 µM) e *Vibrio parahaemolyticus* (MIC de 400 µM) (84).

a) pró-trombina, parcial [*Homo sapiens*]



b) pré-proteína da coagulação fator X isoforma 1 [*Homo sapiens*]



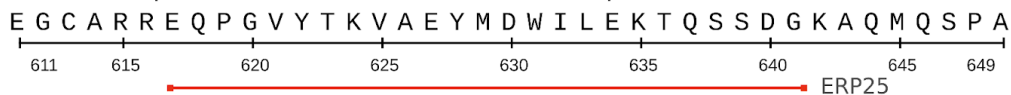
c) fator IX [*Homo sapiens*]



d) coagulação fator XI isoforma X1 [*Homo sapiens*]



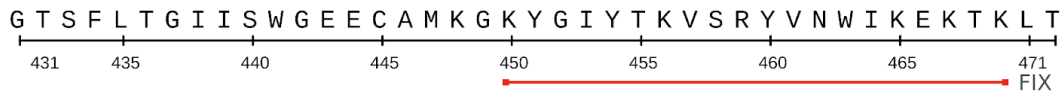
e) calicreína plasmática - isoforma X1 [*Homo sapiens*]



<sup>104</sup> Sequência do KLK10: HPAVYTQICKYMSWINKVIR

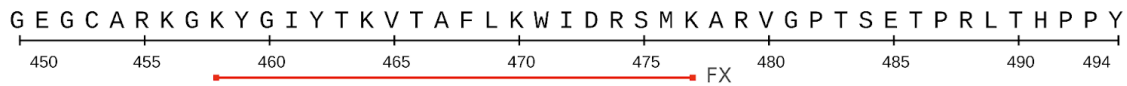
<sup>105</sup> Sequência do TS40: FVSRQSCMDVCAKGAKQHTSRGNVRRARRNRKNRITYLQA

f) pré-proteína da coagulação fator IX isoforma 1 [*Mus musculus*]

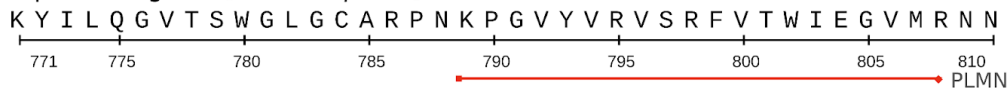


**Figura 14. Criptídeos de proteínas da coagulação.** a) Pró-trombina, parcial de *Homo sapiens* (CAD80258.1) e os criptídeos GKY25, VFR17 e THRB; b) Pré-proteína da coagulação fator X isoforma 1 *Homo sapiens* (NP\_000495.1) e o criptídeo GKY25(X); c) Fator IX de *Homo sapiens* (AAA98726.1) e o criptídeo GKY23; d) Fator de coagulação XI isoforma X1 de *Homo sapiens* (XP\_005262878.1) e o criptídeo ERP23; e) Caliceína plasmática isoforma X1 de *Homo sapiens* (XP\_011530232.1) e o criptídeo ERP25; f) Fator IX da coagulação isoforma 1 de *Mus musculus* e o criptídeo FIX.

a) fator de coagulação X [*Rattus norvegicus*]



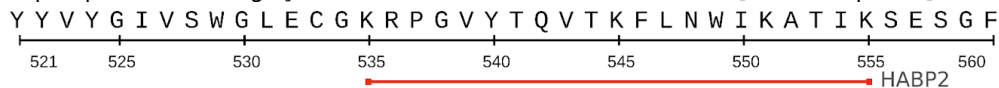
b) plasminogênio [*Homo sapiens*]



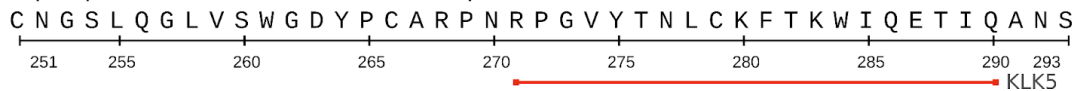
c) proteína C dependente de vitamina-K isoforma X2 [*Homo sapiens*]



d) pré-proteína de ligação com o hialuronano 2 isoforma 1 [*Homo sapiens*]



e) pré-proteína caliceína-5 [*Homo sapiens*]



f) caliceína-8 [*Homo sapiens*]



g) pré-proteína caliceína-10 [*Homo sapiens*]

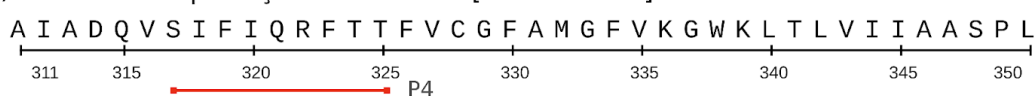


**Figura 15. Criptídeos de proteínas da coagulação.** a) Fator X da coagulação de *Rattus norvegicus* (EDM08870.1) e o criptídeo FX; b) Plasminogênio de *Homo sapiens* (AAA60113.1) e o criptídeo PLMN; c) Proteína C dependente de vitamina K isoforma X2 de *Homo sapiens* (XP\_024308771.1) e o criptídeo PROC; d) Proteína de ligação ao hialuronano 2 isoforma 1 de *Homo sapiens* (NP\_004123.1) e o criptídeo HABP2; e) Pré-proteína caliceína-5 de *Homo sapiens* (NP\_001070959.1) e o criptídeo KLK5; f) Caliceína-8 de *Homo sapiens* (pdb|5MS3|A) e o criptídeo KLK8; g) Caliceína-10 de *Homo sapiens* (NP\_001070968.1) e o criptídeo KLK10.

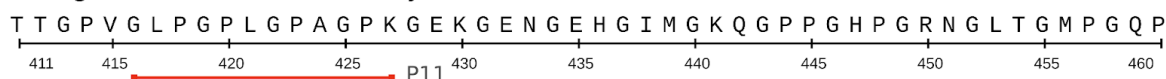
## Colágeno

As proteínas colágenas possuem em suas sequências diversos criptídeos com um amplo espectro de atividades, como antioxidante, anti hipertensivo e antimicrobiano, que são liberados por hidrolases ou pela digestão gastrointestinal (85). Um meio importante de obter essas proteínas e seus subprodutos é pela alimentação de peixes, como o peixe da espécie *Scomber scombrus*, que são ricos em colágenos. A proteólise do colágeno por uma enzima comercial (protemax) resultou na formação de treze fragmentos peptídicos. Analisando a atividade antimicrobiana desses fragmentos, quatro apresentaram atividades antimicrobianas contra *E. coli* e *L. innocua*. Em 200 µg/mL, o criptídeo P4<sup>106</sup> demonstrou-se o mais potente, inibindo completamente o crescimento das cepas bacterianas (Figura 16a). O criptídeo P11<sup>107</sup> inibiu o crescimento de *L. innocua* e 25% do crescimento de *E. coli* (Figura 16b). Os criptídeos P6<sup>108</sup>, P8.1<sup>109</sup> e P8.2<sup>110</sup> inibiram entre 19,5% e 46,8% do crescimento dessas cepas (86).

a) bomba de exportação de sal biliar [*Salmo trutta*]



b) colágeno alfa-1(XXIII) [*Petromyzon marinus*]



**Figura 16. Criptídeos do colágeno.** a) Bomba exportadora de sal biliar de *Salmo trutta* (XP\_029597354.1) e o criptídeo P4; b) Colágeno alfa-1(XXIII) de *Petromyzon marinus* e o criptídeo P11. A sequência da proteína de colágeno do peixe *Scomber scombrus* não foi encontrada na base de dados. As sequências de P8.1 e P8.2 não foram encontradas na base de dados.

## Outros

Numa abordagem peptidômica, os peptídeos de defesa da planta *Solanum lycopersicum*, o tomate, foram identificados após indução de estresse na planta com o tratamento da substância metil jasmonato (MeJA). Um deles, o criptídeo CAPE-1<sup>111</sup>, foi

<sup>106</sup> Sequência do P4: SIFIQRFTT

<sup>107</sup> Sequência do P11: GLPGPLGPAGPK

<sup>108</sup> Sequência não informada.

<sup>109</sup> Sequência do P8.1: RKSGDPLGR

<sup>110</sup> Sequência do P8.2: AKPGDGAGSGPR

<sup>111</sup> Sequência da CAPE-1: PVGNWIGQRPY

testado como *spray* para prevenção de crescimento de larvas e infecção pela bactéria *Pseudomonas syringae*, inibindo o crescimento de ambos e mostrando-se anti-larval e antimicrobiano (Figura 17a). O criptídeo CAPE-1 é proveniente da porção C-terminal da proteína PR-1b, pertencente a um grupo de proteínas ubíquas ligadas à defesa do organismo (87). Além disso, tanto a proteína quanto o criptídeo apresentam sequências conservadas entre outras plantas, como a *Arabidopsis* sp. (88).

Nós também identificamos um criptídeo antibacteriano derivado de uma proteína mitocondrial, a Romo1, proteína moduladora de espécies reativas de oxigênio. O criptídeo encontrado nesta proteína foi denominado AMPR-11<sup>112</sup>. A concentração bactericida mínima (MBC), que representa a menor concentração de um composto necessário para matar uma cepa bacteriana, variou entre 85 - 110 µg/mL contra bactérias gram-positivas *S. aureus*, *Bacillus subtilis*, *Enterococcus faecium*, *Streptomyces sindenensis*, *Enterococcus faecalis*, *Streptococcus pneumoniae*, bactérias gram-negativas, *E. coli*, *P. aeruginosa*, *K. pneumoniae*, *A. baumannii* e *Enterobacter aerogenes*, além de bactérias multidrogas resistentes, como *P. aeruginosa* resistente a carbapenêmicos, *A. baumannii* resistente a carbapenêmicos, *K. pneumoniae* resistente a carbapenêmicos, *S. aureus* resistente à vancomicina e *E. faecium* resistente à vancomicina (89). Em um outro estudo, com o intuito de aumentar a atividade antimicrobiana do AMPR-11, alguns resíduos foram substituídos por lisinas e o resíduo Met18 foi deletado, o que gerou o criptídeo AMPR-22<sup>113</sup>, cuja atividade antimicrobiana aumentou quase 100 vezes, com valores de MBC contra os mesmos patógenos variando entre 1-4 µg/mL (90).

Em peixes-zebra, a proteína ZF-RNase-3 possui funções angiogênicas e bactericidas. Interessantemente, a atividade bactericida dessa proteína não é perdida após sua desnaturação e, quando em contato com a aspartil protease OmpT (EC 3.4.23.49) de *E. coli*, um fragmento grande de 93 resíduos, denominado LF-ZF3<sup>114</sup>, é liberado e apresenta atividade antibacteriana (Figura 17b). Nas condições experimentais testadas, o criptídeo (ou “criptoproteína”) foi capaz de inibir 80% do crescimento de culturas de *E. coli* em uma concentração de 3 µM (91). É interessante notar, neste caso, como o peixe utiliza componentes na parede externa da bactéria para liberar criptídeos que serão usados na destruição deste patógeno.

---

<sup>112</sup> Sequência do AMPR-11: KTMMQSGGTFGTFMAIGMGIR

<sup>113</sup> Sequência do AMPR-22: KKMMKKGKFGTFMAIGGIR

<sup>114</sup> Sequência da LF-ZF3: RITRFPTGNDCKEVTFIQANGNHVRTVCTGGGTRQTDNRDLYMSNDQ  
FTVITCTLRSGERHPNCRYRGKESSRKIVVACEGEWPAHYERGVIV

## 1.3 Métodos de identificação

### 1.3.1 Proteômica acoplada a identificação de atividade antimicrobiana

Estratégias experimentais para a identificação de criptídeos antimicrobianos envolvem, geralmente, duas etapas. A primeira etapa é a identificação do peptidoma (conjunto de peptídeos de um organismo), por técnicas proteômicas associadas à espectrometria de massa, e uma segunda etapa que envolve a identificação da atividade antimicrobiana de cada um dos peptídeos previamente encontrados (92). Os microrganismos alvo desses peptídeos podem ser bactérias, fungos ou vírus. A escolha de qual microorganismo estudar geralmente envolve uma pergunta científica prévia do grupo de pesquisa, além da disponibilidade e de recursos financeiros. As bases de dados, tais como o APD, demonstram um maior depósito de sequências antibactericidas, seguidas por antifúngicas e depois por antivirais (93). Este dado pode estar ligado a características intrínsecas de peptídeos bioativos ou apenas uma correlação entre os microrganismos que são mais estudados e a identificação de novos peptídeos.

A atividade pode ser predita por ferramentas que usam algoritmos de aprendizado de máquina, como os criptídeos antimicrobianos identificados no proteoma do musgo *Physcomitrella patens*, que usaram os preditores iAMPpred, CAMP e ADAM (94–96). As proteínas que, neste caso, deram origem a estes criptídeos são a gliceraldeído-3-fosfato desidrogenase, a cucumicina e a  $\alpha$ -expansina. Dois peptídeos foram selecionados para teste de atividade antibacteriana em *E. coli* e *B. subtilis*, e o peptídeo 8 de Pp3c14\_22870V3<sup>115</sup> apresentou o menor valor de MIC: 16 e 32  $\mu\text{g/mL}$ , respectivamente (16) (Figura 17c).

A atividade também pode ser avaliada *in vitro*, como no estudo que avaliou o criptoma (“ômica” dos criptídeos) do soro de humanos. Seis criptídeos foram avaliados: G17K<sup>116</sup>, oriundo da proteína do complemento C4-A (Figura 17d), S17K<sup>117</sup>, oriundo da proteína do complemento C3 (Figura 17e), D15R<sup>118</sup> e D15T<sup>119</sup>, oriundos do fibrinogênio da cadeia alfa (Figura 17f), e K13L<sup>120</sup>, oriundo da soroalbumina (Figura 7a). Os peptídeos S17K, D15T e D15R não demonstraram atividade antifúngica em *Candida albicans*. Os peptídeos K13L e

<sup>115</sup> Sequência do Pp3c14\_22870V3: INIINAPLQGFKIA

<sup>116</sup> Sequência do G17K: GLEEELQFSLGSKINVK

<sup>117</sup> Sequência do S17K: SEETKENEFTVTAEGK

<sup>118</sup> Sequência do D15R: DSGEGFLAEGGGVR

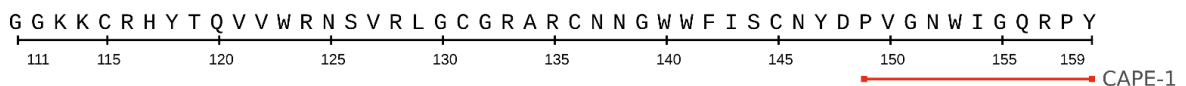
<sup>119</sup> Sequência do D15T: DEAGSEADHEGTHST

<sup>120</sup> Sequência do K13L: KKLVAASQAALGL

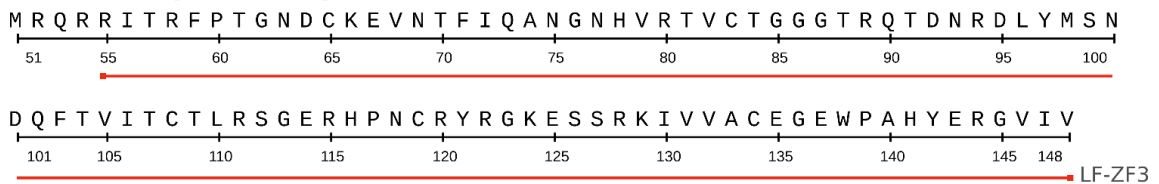
G17K apresentaram atividade e os valores de EC50 do peptídeo K13L foram menor do que os valores de EC50 do peptídeo G17K nas cepas testadas (*C. albicans* e *C. glabrata*). Nenhum dos peptídeos apresentou atividade hemolítica em eritrócitos (41).

Um método distinto foi realizado por Cerrato e colaboradores (2020). O grupo utilizou como modelo o músculo de peixes da espécie *Thunnus albacares* para passar por um processo de digestão similar ao gastrointestinal. Assim, os pesquisadores avaliaram a atividade em conjunto de peptídeos liberados por ação das enzimas tripsina, pepsina e pancreatina (EC 232-468-9) (11), e o conjunto de peptídeos foi separado por colunas de cromatografia C18. O conjunto de criptídeos, um total de 403 peptídeos, teve atividade contra *S. aureus*. Os peptídeos não tiveram suas sequências exibidas, exceto por dois que foram destacados após pontuação da atividade avaliada pelo programa CPPpred, que são as sequências KKLGELLK e KLGELLK. Entretanto, nenhum peptídeo foi testado separadamente (97).

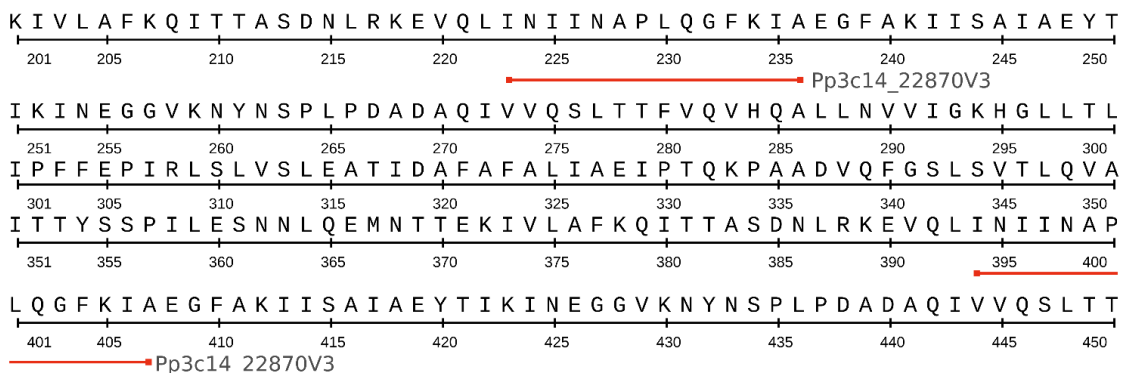
a) Proteína da folha relacionada à patogênese [*Solanum lycopersicum*]



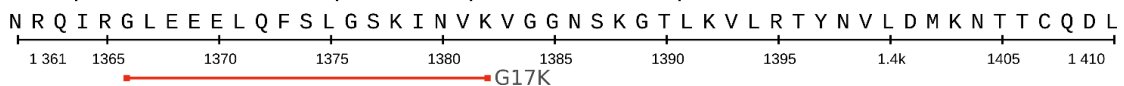
b) ZF-RNase-3 [*Danio rerio*]



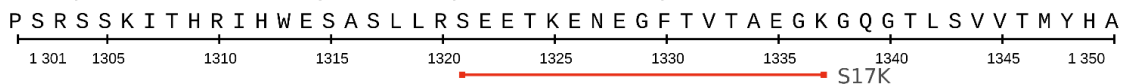
c) LOC112291799 [*Physcomitrium patens*]



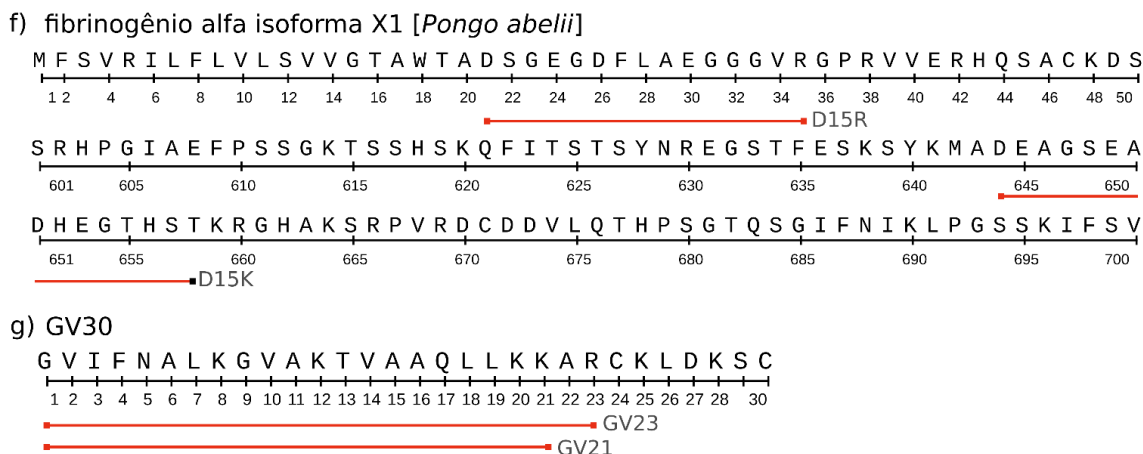
d) componente C4A do complemento, parcial [*Homo sapiens*]



e) componente C3 do complemento, parcial [*Homo sapiens*]







**Figura 17. Outros criptídeos.** a) Proteína da folha relacionada à patogênese 6 de *Solanum lycopersicum* (NP\_001234314.1) e o criptídeo CAPE-1; b) ZF-Rnase 3 de *Danio rerio* (pdb|2VQ9|A) e a criptoproteína LF-ZF3; c) LOC112291799 de *Physcomitrium patens* (XP\_024395469.1) e o criptídeo Pp3c14\_22870V3; d) Componente C4a do complemento de *Homo sapiens* (AAB59537.1) e o criptídeo G17K; e) Componente C3 do complemento de *Homo sapiens* (AAA85332.1) e o criptídeo S17K; f) Fibrinogênio alfa isoforma X1 de *Pongo abelii* (XP\_002815279.3) e os criptídeos D15R e D15K; g) GV30 de *Hylarana guentheri* e os criptídeos GV23 e GV21.

### 1.3.2 Identificação dos criptídeos *in silico*

A identificação de criptídeos não precisa ser experimental. O trabalho de Bosso e colaboradores (2017) identificou um criptídeo na enzima HSD3 após fazer uma varredura em sequências enzimas com um algoritmo de pontuação para atividade antimicrobiana: potência antimicrobiana (AP) =  $\text{Log}(1000/\text{MIC})$ , que está linearmente relacionado a uma atividade absoluta, dependente de propriedades físico-químicas (98). Além disso, a liberação desses criptídeos pode ser estipulada pela ação de enzimas proteolíticas, identificando os possíveis sítios de clivagem das proteínas que carregam os potenciais criptídeos (69).

Além da identificação de criptídeos em proteínas, a identificação de novos criptídeos pode ser feita em peptídeos. Foi o trabalho feito por Ma e colaboradores (2021). Após uma tripsinolise *in silico* do peptídeo GV30, derivado de secreções do sapo *Hylarana guentheri*, culminou com dois criptídeos com atividade aumentada em relação ao peptídeo precursor, os criptídeos GV23<sup>121</sup> e GV21<sup>122</sup>. O criptídeo GV23 apresentou MIC de 2  $\mu\text{M}$  contra *S. aureus*, 4  $\mu\text{M}$  contra *E. coli*, 2  $\mu\text{M}$  contra *S. aureus* resistente à metilicina, 32  $\mu\text{M}$  contra *K. pneumoniae*, 16  $\mu\text{M}$  contra *E. faecium* e 16  $\mu\text{M}$  contra *P. aeruginosa*. Já o criptídeo GV21

<sup>121</sup> Sequência do GV23: GVIFNALKGVAKTVAAQLLKKAR –NH<sub>2</sub>

<sup>122</sup> Sequência do GV21: GVIFNALKGVAKTVAAQLLKK –NH<sub>2</sub>

apresentou MIC de 2  $\mu\text{M}$  contra *S. aureus*, 4  $\mu\text{M}$  contra *E. coli*, 4  $\mu\text{M}$  contra *S. aureus* resistente à metilicina, 16  $\mu\text{M}$  contra *K. pneumoniae*, 4  $\mu\text{M}$  contra *E. faecium* e 16  $\mu\text{M}$  contra *P. aeruginosa*) (99).

### 1.3.3 Identificação por métodos bioquímicos

Em alguns casos, a identificação de criptídeos pode ser um tiro no escuro. Um exemplo, é a identificação de criptídeos em animais marinhos, sem conhecimento prévio de dados genômicos e proteômicos. Apesar desse desconhecimento, estudos mostram como esses organismos podem ser fontes de substâncias antimicrobianas depois de ingeridos. Nesse caso, métodos bioquímicos de geração de lisados enzimáticos e separação dos componentes hidrolisados são essenciais. Um exemplo é a ostra do pacífico, *Crassostrea gigas*, que demonstrou ser uma fonte de criptídeos antimicrobianos, mesmo sem indícios diretos para essa hipótese. Após a separação do músculo da ostra e a proteólise com alcalase e bromelina, um grupo de pesquisadores partiu para a separação destes componentes hidrolisados. Três frações foram identificadas na cromatografia de troca iônica, sendo a fração D3 aquela com atividade antimicrobiana, que depois foi separada por gel filtração e a fração S3 demonstrou ser a fração ativa. Por cromatografia líquida de fase reversa, o criptídeo *CgPep33*<sup>123</sup> foi identificado. Esse criptídeo demonstrou atividade contra *E. coli* (MIC de 80  $\mu\text{g/mL}$ ), *P. aeruginosa* (MIC de 80  $\mu\text{g/mL}$ ), *B. subtilis* (MIC de 60  $\mu\text{g/mL}$ ), *S. aureus* (MIC de 40  $\mu\text{g/mL}$ ), *B. cinerea* (MIC de 120  $\mu\text{g/mL}$ ) e *Penicillium expansum* (MIC de 140  $\mu\text{g/mL}$ ) (100).

## 1.4 Considerações Finais

Em uma era com grandes avanços na área da biotecnologia, usar as moléculas e os dados da natureza é uma forma de reciclar os recursos naturais, ganhando novas perspectivas sobre as macromoléculas com funções previamente conhecidas e resolvendo problemas que compostos amplamente usados, hoje em dia, não foram capazes de superar. É nesse contexto que o conceito de criptídeos ganha força e o interesse de alguns grupos de pesquisa em todo o mundo. Os dados da atividade de alguns criptídeos citados estão compilados abaixo. A potência antimicrobiana dos criptídeos está resumida na Tabela 1.

---

<sup>123</sup> Sequência do criptídeo *CgPep33* não determinada.

Alguns supostos criptídeos não foram considerados nesta revisão por conflitos de definição. Um deles, por exemplo, é o peptídeo LL-37, que apesar de ser liberado após clivagem de outra proteína, a hCAP-18, é o objetivo único e final da proteína precursora, a liberação do peptídeo antimicrobiano LL-37. Nessa área, além do estabelecimento do conceito de criptídeos e a identificação de novas moléculas, precisa ficar claro como elas são liberadas na natureza, qual é o impacto que essas moléculas possuem na defesa dos organismos que os geram e que, mesmo que alguns criptídeos só sejam capazes de ser gerados *in vitro*, os mesmos representam sim uma classe de criptídeos. Isto porque este autor defende que criptídeos são quaisquer moléculas peptídicas (até 50 resíduos) com bioatividade oriundas de moléculas proteicas, cuja atividade pode ou não estar relacionada à atividade do fragmento liberado, mas cujo intuito da proteína precursora não seja, unicamente, a liberação do fragmento peptídico (92).

Tabela 1. Potência de atividade antimicrobiana dos criptídeos.\*

Bactérias/MIC	≤ 1 μM	> 1 e ≤ 5 μM	> 5 e ≤ 10 μM	> 10 μM
<i>E. coli</i>	oncorincina II, α 33–66, α 139–141, ECP-(1-45)	SAMP H1, HbβP-1, β 140–145, β 1–13, β 143-145, ECP-(0-38), bLF-17-41, GKY25, LF-ZF3, parasina I, buforina II	α 34–66, α 33–46, α 33–45, α34–46, α 36–45, α37–46, α 133–141, α 137–141, ECP-(24-45), GVF27, buforina I	Ub2, AFP-1, hiposina, hLF-18-40
<i>P. aeruginosa</i>		r(P)ApoB <sub>L</sub> <sup>Ala</sup> , bLF-17-41		GKY25, r(P)ApoB <sub>L</sub> , r(P)ApoB <sub>S</sub> , omCK11-31, omCCL28-like- 23, drCCL27b-24, lcCCL28-25, smCCL27b-25,

---

				trCCL28-29, chCCL27a-26, H5(61-90), GVF27
<i>S. aureus</i>	$\alpha$ 139–141, ECP-(1-45)	ECP-(24-45), GVF27	AFP-1,	GKY25

---

\* Os dados na tabela foram retirados de artigos científicos que disponibilizam os valores de MIC em micromolar.

## Referências

1. Yount NY, Bayer AS, Xiong YQ, Yeaman MR. Advances in antimicrobial peptide immunobiology. *Peptide Science*. 2006;84(5):435–58.
2. Chernysh S, Kim SI, Bekker G, Pleskach VA, Filatova NA, Anikin VB, et al. Antiviral and antitumor peptides from insects. *Proceedings of the National Academy of Sciences*. 2002 Oct;99(20):12628–32.
3. Dadar M, Shahali Y, Chakraborty S, Prasad M, Tahoori F, Tiwari R, et al. Antiinflammatory peptides: current knowledge and promising prospects. *Inflamm Res*. 2019 Feb 1;68(2):125–45.
4. Mangoni ML, McDermott AM, Zasloff M. Antimicrobial peptides and wound healing: biological and therapeutic considerations. *Experimental Dermatology*. 2016;25(3):167–73.
5. Diochot S, Salinas M, Baron A, Escoubas P, Lazdunski M. Peptides inhibitors of acid-sensing ion channels. *Toxicon*. 2007 Feb 1;49(2):271–84.
6. Mukai H, Hokari Y, Seki T, Takao T, Kubota M, Matsuo Y, et al. Discovery of mitocryptide-1, a neutrophil-activating cryptide from healthy porcine heart. *J Biol Chem*. 2008 Nov 7;283(45):30596–605.
7. Mukai H, Seki T, Nakano H, Hokari Y, Takao T, Kawanami M, et al. Mitocryptide-2: purification, identification, and characterization of a novel cryptide that activates neutrophils. *J Immunol*. 2009 Apr 15;182(8):5072–80.
8. Ueki N, Someya K, Matsuo Y, Wakamatsu K, Mukai H. Cryptides: functional cryptic peptides hidden in protein structures. *Biopolymers*. 2007;88(2):190–8.
9. Dionysius DA, Milne JM. Antibacterial Peptides of Bovine Lactoferrin: Purification and Characterization. *Journal of Dairy Science*. 1997 Apr 1;80(4):667–74.
10. Daoud R, Dubois V, Bors-Dodita L, Nedjar-Arroume N, Krier F, Chihib NE, et al. New antibacterial peptide derived from bovine hemoglobin. *Peptides*. 2005 May 1;26(5):713–9.
11. Zenezini Chiozzi R, Capriotti AL, Cavaliere C, La Barbera G, Piovesana S, Laganà A. Identification of three novel angiotensin-converting enzyme inhibitory peptides derived from cauliflower by-products by multidimensional liquid chromatography and bioinformatics. *Journal of Functional Foods*. 2016 Dec 1;27:262–73.
12. Kim HS, Yoon H, Minn I, Park CB, Lee WT, Zasloff M, et al. Pepsin-Mediated Processing of the Cytoplasmic Histone H2A to Strong Antimicrobial Peptide Buforin I. *The Journal of Immunology*. 2000 Sep 15;165(6):3268–74.
13. Pihlanto-Leppälä A, Marnila P, Hubert L, Rokka T, Korhonen HJT, Karp M. The effect of  $\alpha$ -lactalbumin and  $\beta$ -lactoglobulin hydrolysates on the metabolic activity of *Escherichia coli* JM103. *Journal of Applied Microbiology*. 1999;87(4):540–5.
14. Bellamy W, Takase M, Yamauchi K, Wakabayashi H, Kawase K, Tomita M. Identification of the bactericidal domain of lactoferrin. *Biochim Biophys Acta*. 1992 May 22;1121(1–2):130–6.
15. Pellegrini A, Hülsmeier AJ, Hunziker P, Thomas U. Proteolytic fragments of ovalbumin display antimicrobial activity. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2004 May 3;1672(2):76–85.
16. Fesenko I, Azarkina R, Kirov I, Kniazev A, Filippova A, Grafaskaia E, et al. Phytohormone treatment induces generation of cryptic peptides with antimicrobial activity in the Moss *Physcomitrella patens*. *BMC Plant Biology*. 2019 Jan 7;19(1):9.
17. Alonso S, Pethe K, Russell DG, Purdy GE. Lysosomal killing of *Mycobacterium* mediated by ubiquitin-derived peptides is enhanced by autophagy. *Proceedings of the National Academy of Sciences*. 2007 Apr 3;104(14):6031–6.
18. Kieffer AE, Goumon Y, Ruh O, Chasserot-Golaz S, Nullans G, Gasnier C, et al. The N- and C-terminal fragments of ubiquitin are important for the antimicrobial activities. *The FASEB Journal*. 2003;17(6):776–8.
19. Kim JY, Lee SY, Park SC, Shin SY, Choi SJ, Park Y, et al. Purification and antimicrobial activity studies of the N-terminal fragment of ubiquitin from human amniotic fluid. *Biochimica et*

- Biophysica Acta (BBA) - Proteins and Proteomics. 2007 Sep 1;1774(9):1221–6.
20. Candido-Ferreira IL, Kronenberger T, Sayegh RSR, Batista I de FC, da Silva Junior PI. Evidence of an Antimicrobial Peptide Signature Encrypted in HECT E3 Ubiquitin Ligases. *Frontiers in Immunology* [Internet]. 2017 [cited 2022 Jun 7];7. Available from: <https://www.frontiersin.org/article/10.3389/fimmu.2016.00664>
  21. Ichimiya T, Yamakawa T, Hirano T, Yokoyama Y, Hayashi Y, Hirayama D, et al. Autophagy and Autophagy-Related Diseases: A Review. *International Journal of Molecular Sciences*. 2020 Jan;21(23):8974.
  22. Ponpuak M, Davis AS, Roberts EA, Delgado MA, Dinkins C, Zhao Z, et al. Delivery of cytosolic components by autophagic adaptor protein p62 endows autophagosomes with unique antimicrobial properties. *Immunity*. 2010 Mar 26;32(3):329–41.
  23. Birkemo GA, Lüders T, Andersen Ø, Nes IF, Nissen-Meyer J. Hippusin, a histone-derived antimicrobial peptide in Atlantic halibut (*Hippoglossus hippoglossus* L.). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2003 Mar 21;1646(1):207–15.
  24. Li C, Song L, Zhao J, Zhu L, Zou H, Zhang H, et al. Preliminary study on a potential antibacterial peptide derived from histone H2A in hemocytes of scallop *Chlamys farreri*. *Fish & Shellfish Immunology*. 2007 Jun 1;22(6):663–72.
  25. Park CB, Kim MS, Kim SC. A Novel Antimicrobial Peptide from *Bufo bufo gargarizans*. *Biochemical and Biophysical Research Communications*. 1996 Jan 5;218(1):408–13.
  26. Cho JH, Park IY, Kim HS, Lee WT, Kim MS, Kim SC. Cathepsin D produces antimicrobial peptide parasin I from histone H2A in the skin mucosa of fish. *The FASEB Journal*. 2002;16(3):429–31.
  27. Park IY, Park CB, Kim MS, Kim SC. Parasin I, an antimicrobial peptide derived from histone H2A in the catfish, *Parasilurus asotus*. *FEBS Letters*. 1998;437(3):258–62.
  28. Lemaire S, Trinh TT, Le HT, Tang SC, Hincke M, Wellman-Labadie O, et al. Antimicrobial effects of H4-(86–100), histogranin and related compounds – possible involvement of DNA gyrase. *The FEBS Journal*. 2008;275(21):5286–97.
  29. Lemaire S, Shukla VK, Rogers C, Ibrahim IH, Lapierre C, Parent P, et al. Isolation and characterization of histogranin, a natural peptide with NMDA receptor antagonist activity. *European Journal of Pharmacology: Molecular Pharmacology*. 1993 May 15;245(3):247–56.
  30. Lemaire S, Rogers C, Dumont M, Shukla VK, Lapierre C, Prasad J, et al. Histogranin, a modified histone h4 fragment endowed with N-methyl-D-aspartate antagonist and immunostimulatory activities. *Life Sciences*. 1995 Mar 3;56(15):1233–41.
  31. Fernandes JMO, Molle G, Kemp GD, Smith VJ. Isolation and characterisation of oncorhyncin II, a histone H1-derived antimicrobial peptide from skin secretions of rainbow trout, *Oncorhynchus mykiss*. *Developmental & Comparative Immunology*. 2004 Feb 1;28(2):127–38.
  32. Lüders T, Birkemo GA, Nissen-Meyer J, Andersen Ø, Nes IF. Proline Conformation-Dependent Antimicrobial Activity of a Proline-Rich Histone H1 N-Terminal Peptide Fragment Isolated from the Skin Mucus of Atlantic Salmon. *Antimicrobial Agents and Chemotherapy*. 2005 Jun;49(6):2399–406.
  33. Jodoin J, Hincke MT. Histone H5 is a potent Antimicrobial Agent and a template for novel Antimicrobial Peptides. *Sci Rep*. 2018 Feb 5;8(1):2411.
  34. Gaglione R, Cesaro A, Dell’Olmo E, Della Ventura B, Casillo A, Di Girolamo R, et al. Effects of human antimicrobial cryptides identified in apolipoprotein B depend on specific features of bacterial strains. *Sci Rep*. 2019 Apr 30;9(1):6728.
  35. Gaglione R, Smaldone G, Cesaro A, Rumolo M, De Luca M, Di Girolamo R, et al. Impact of a Single Point Mutation on the Antimicrobial and Fibrillogenic Properties of Cryptides from Human Apolipoprotein B. *Pharmaceuticals*. 2021 Jul;14(7):631.
  36. Cesaro A, Torres MDT, Gaglione R, Dell’Olmo E, Di Girolamo R, Bosso A, et al. Synthetic Antibiotic Derived from Sequences Encrypted in a Protein from Human Plasma. *ACS Nano*. 2022 Feb 22;16(2):1880–95.

37. Kragh-Hansen U. Human Serum Albumin: A Multifunctional Protein. In: Otagiri M, Chuang VTG, editors. *Albumin in Medicine: Pathological and Clinical Applications* [Internet]. Singapore: Springer; 2016 [cited 2022 Jun 7]. p. 1–24. Available from: [https://doi.org/10.1007/978-981-10-2116-9\\_1](https://doi.org/10.1007/978-981-10-2116-9_1)
38. Kragh-Hansen U. Possible Mechanisms by Which Enzymatic Degradation of Human Serum Albumin Can Lead to Bioactive Peptides and Biomarkers. *Frontiers in Molecular Biosciences* [Internet]. 2018 [cited 2022 Jun 7];5. Available from: <https://www.frontiersin.org/article/10.3389/fmolb.2018.00063>
39. Nakagomi K, Fujimura A, Ebisu H, Sakai T, Sadakane Y, Fujii N, et al. Acein-1, a novel angiotensin-I-converting enzyme inhibitory peptide isolated from tryptic hydrolysate of human plasma. *FEBS Lett.* 1998 Nov 6;438(3):255–7.
40. Nakagomi K, Takatsu K, Takagi S, Ebisu H, Sadakane Y, Fujii N, et al. Isolation of cathepsin B inhibitory peptides, Cabin-A1 and -A2, from a tryptic and chymotryptic hydrolysate of human serum albumin. *Peptides.* 2002 Sep 1;23(9):1567–71.
41. Ciociola T, Zanello PP, D’Adda T, Galati S, Conti S, Magliani W, et al. A Peptide Found in Human Serum, Derived from the C-Terminus of Albumin, Shows Antifungal Activity In Vitro and In Vivo. *Microorganisms.* 2020 Oct 21;8(10).
42. Al-Mohammadi AR, Osman A, Enan G, Abdel-Shafi S, El-Nemer M, Sitohy M, et al. Powerful Antibacterial Peptides from Egg Albumin Hydrolysates. *Antibiotics.* 2020 Dec;9(12):901.
43. Riciluca KCT, Sayegh RSR, Melo RL, Silva PI. Rondonin an antifungal peptide from spider (*Acanthoscurria rondoniae*) haemolymph. *Results in Immunology.* 2012 Jan 1;2:66–71.
44. Deng L, Pan X, Wang Y, Wang L, Zhou XE, Li M, et al. Hemoglobin and its derived peptides may play a role in the antibacterial mechanism of the vagina. *Human Reproduction.* 2009 Jan 1;24(1):211–8.
45. Ivanov VT, Karelin AA, Philippova MM, Nazimov IV, Pletnev VZ. Hemoglobin as a source of endogenous bioactive peptides: the concept of tissue-specific peptide pool. *Biopolymers.* 1997;43(2):171–88.
46. Ullal AJ, Wayne Litaker R, Noga EJ. Antimicrobial peptides derived from hemoglobin are expressed in epithelium of channel catfish (*Ictalurus punctatus*, Rafinesque). *Developmental & Comparative Immunology.* 2008 Jan 1;32(11):1301–12.
47. Fogaça AC, da Silva PI, Miranda MTM, Bianchi AG, Miranda A, Ribolla PEM, et al. Antimicrobial Activity of a Bovine Hemoglobin Fragment in the Tick *Boophilus microplus* \*. *Journal of Biological Chemistry.* 1999 Sep 3;274(36):25330–4.
48. Nedjar-Arroume N, Dubois-Delval V, Adje EY, Traisnel J, Krier F, Mary P, et al. Bovine hemoglobin: An attractive source of antibacterial peptides. *Peptides.* 2008 Jun 1;29(6):969–77.
49. Catiau L, Traisnel J, Delval-Dubois V, Chihib NE, Guillochon D, Nedjar-Arroume N. Minimal antimicrobial peptidic sequence from hemoglobin alpha-chain: KYR. *Peptides.* 2011 Apr 1;32(4):633–8.
50. Catiau L, Traisnel J, Chihib NE, Le Flem G, Blanpain A, Melnyk O, et al. RYH: A minimal peptidic sequence obtained from beta-chain hemoglobin exhibiting an antimicrobial activity. *Peptides.* 2011 Jul 1;32(7):1463–8.
51. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies.* 2019 Dec;8(4):55.
52. Magliani W, Giovati L, Ciociola T, Sperindè M, Santinoli C, Conti G, et al. Antibodies as a source of anti-infective peptides: an update. *Future Microbiol.* 2015;10(7):1163–75.
53. Pan Y, Yuhasz SC, Amzel LM. Anti-idiotypic antibodies: biological function and structural studies. *The FASEB Journal.* 1995;9(1):43–9.
54. Pettoello-Mantovani M, Nocerino A, Polonelli L, Morace G, Conti S, di Martino L, et al. *Hansenula anomala* killer toxin induces secretion and severe acute injury in the rat intestine. *Gastroenterology.* 1995 Dec 1;109(6):1900–6.
55. Polonelli L, Magliani W, Conti S, Bracci L, Lozzi L, Neri P, et al. Therapeutic Activity of an

- Engineered Synthetic Killer Antiidiotypic Antibody Fragment against Experimental Mucosal and Systemic Candidiasis. *Infection and Immunity*. 2003 Nov;71(11):6205–12.
56. Casoli C, Pilotti E, Perno CF, Balestra E, Polverini E, Cassone A, et al. A killer mimotope with therapeutic activity against AIDS-related opportunistic micro-organisms inhibits ex-vivo HIV-1 replication. *AIDS*. 2006 Apr 24;20(7):975–80.
  57. Conti G, Magliani W, Conti S, Nencioni L, Sgarbanti R, Palamara AT, et al. Therapeutic Activity of an Anti-Idiotypic Antibody-Derived Killer Peptide against Influenza A Virus Experimental Infection. *Antimicrobial Agents and Chemotherapy*. 2008 Dec;52(12):4331–7.
  58. Savoia D, Scutera S, Raimondo S, Conti S, Magliani W, Polonelli L. Activity of an engineered synthetic killer peptide on *Leishmania major* and *Leishmania infantum* promastigotes. *Experimental Parasitology*. 2006 Jul 1;113(3):186–92.
  59. Giovati L, Santinoli C, Mangia C, Vismarra A, Belletti S, D’Adda T, et al. Novel Activity of a Synthetic Decapeptide Against *Toxoplasma gondii* Tachyzoites. *Frontiers in Microbiology* [Internet]. 2018 [cited 2022 Jun 7];9. Available from: <https://www.frontiersin.org/article/10.3389/fmicb.2018.00753>
  60. Polonelli L, Pontón J, Elguezabal N, Moragues MD, Casoli C, Pilotti E, et al. Antibody Complementarity-Determining Regions (CDRs) Can Display Differential Antimicrobial, Antiviral and Antitumor Activities. *PLOS ONE*. 2008 Jun 11;3(6):e2371.
  61. Simon JP, Aunis D. Biochemistry of the chromogranin A protein family. *Biochemical Journal*. 1989 Aug 15;262(1):1–13.
  62. Strub JM, Garcia-Sablone P, Lonning K, Taupenot L, Hubert P, Van Dorsselaer A, et al. Processing of chromogranin B in bovine adrenal medulla. Identification of secretolytin, the endogenous C-terminal fragment of residues 614-626 with antibacterial activity. *Eur J Biochem*. 1995 Apr 15;229(2):356–68.
  63. Goumon Y, Strub JM, Moniatte M, Nullans G, Poteur L, Hubert P, et al. The C-terminal Bisphosphorylated proenkephalin-A-(209–237)-peptide from Adrenal Medullary Chromaffin Granules Possesses Antibacterial Activity. *European Journal of Biochemistry*. 1996;235(3):516–25.
  64. Goumon Y, Lugardon K, Kieffer B, Lefèvre JF, Van Dorsselaer A, Aunis D, et al. Characterization of Antibacterial COOH-terminal Proenkephalin-A-derived Peptides (PEAP) in Infectious Fluids: IMPORTANCE OF ENKELYTIN, THE ANTIBACTERIAL PEAP209–237 SECRETED BY STIMULATED CHROMAFFIN CELLS\*. *Journal of Biological Chemistry*. 1998 Nov 6;273(45):29847–56.
  65. Strub JM, Goumon Y, Lugardon K, Capon C, Lopez M, Moniatte M, et al. Antibacterial Activity of Glycosylated and Phosphorylated Chromogranin A-derived Peptide 173-194 from Bovine Adrenal Medullary Chromaffin Granules\*. *Journal of Biological Chemistry*. 1996 Nov 8;271(45):28533–40.
  66. Torrent M, de la Torre BG, Nogués VM, Andreu D, Boix E. Bactericidal and membrane disruption activities of the eosinophil cationic protein are largely retained in an N-terminal fragment. *Biochemical Journal*. 2009 Jul 15;421(3):425–34.
  67. Sánchez D, Moussaoui M, Carreras E, Torrent M, Nogués V, Boix E. Mapping the eosinophil cationic protein antimicrobial activity by chemical and enzymatic cleavage. *Biochimie*. 2011 Feb 1;93(2):331–8.
  68. Su J, Li H, Hu J, Wang D, Zhang F, Fu Z, et al. LcCCL28-25, Derived from Piscine Chemokine, Exhibits Antimicrobial Activity against Gram-Negative and Gram-Positive Bacteria In Vitro and In Vivo. *Microbiology Spectrum*. 2022;0(0):e02515-21.
  69. Sukhija R, Kakar P, Mehta V, Mehta JL. Enhanced 11 $\beta$ -Hydroxysteroid Dehydrogenase Activity, the Metabolic Syndrome, and Systemic Hypertension. *The American Journal of Cardiology*. 2006 Aug 15;98(4):544–8.
  70. Bosso A, Pirone L, Gaglione R, Pane K, Del Gatto A, Zaccaro L, et al. A new cryptic host defense peptide identified in human 11-hydroxysteroid dehydrogenase-1  $\beta$ -like: from in silico



- identification to experimental evidence. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2017 Sep 1;1861(9):2342–53.
71. Mine Y, Ma F, Lauriau S. Antimicrobial Peptides Released by Enzymatic Hydrolysis of Hen Egg White Lysozyme. *J Agric Food Chem*. 2004 Mar 1;52(5):1088–94.
  72. Ibrahim HR, Imazato K, Ono H. Human Lysozyme Possesses Novel Antimicrobial Peptides within Its N-terminal Domain that Target Bacterial Respiration. *J Agric Food Chem*. 2011 Sep 28;59(18):10336–45.
  73. Nuijens JH, van Berkel PHC, Schanbacher FL. Structure and biological actions of lactoferrin. *J Mammary Gland Biol Neoplasia*. 1996 Jul 1;1(3):285–95.
  74. Bullen JJ. The Significance of Iron in Infection. *Reviews of Infectious Diseases*. 1981 Nov 1;3(6):1127–38.
  75. Nibbering PH, Ravensbergen E, Welling MM, van Berkel LA, van Berkel PHC, Pauwels EKJ, et al. Human Lactoferrin and Peptides Derived from Its N Terminus Are Highly Effective against Infections with Antibiotic-Resistant Bacteria. *Infection and Immunity*. 2001 Mar;69(3):1469–76.
  76. Dijkshoorn L, Brouwer CPJM, Bogaards SJP, Nemec A, van den Broek PJ, Nibbering PH. The Synthetic N-Terminal Peptide of Human Lactoferrin, hLF(1-11), Is Highly Effective against Experimental Infection Caused by Multidrug-Resistant *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*. 2004 Dec;48(12):4919–21.
  77. Lupetti A, Paulusma-Annema A, Welling MM, Senesi S, van Dissel JT, Nibbering PH. Candidacidal Activities of Human Lactoferrin Peptides Derived from the N Terminus. *Antimicrobial Agents and Chemotherapy*. 2000 Dec;44(12):3257–63.
  78. Bode W. The Structure of Thrombin: A Janus-Headed Proteinase. *Semin Thromb Hemost*. 2006 Feb;32(S 1):016–31.
  79. Davie EW, Kulman JD. An Overview of the Structure and Function of Thrombin. *Semin Thromb Hemost*. 2006 Feb;32(S 1):003–15.
  80. Papareddy P, Rydengård V, Pasupuleti M, Walse B, Mörgelin M, Chalupka A, et al. Proteolysis of human thrombin generates novel host defense peptides. *PLoS Pathog*. 2010 Apr 22;6(4):e1000857.
  81. Brower M, Walz D, Garry K, Fenton J 2d. Human neutrophil elastase alters human alpha-thrombin function: limited proteolysis near the gamma-cleavage site results in decreased fibrinogen clotting and platelet-stimulatory activity. *Blood*. 1987 Mar 1;69(3):813–9.
  82. Hansen FC, Kalle-Brune M, Plas MJA van der, Strömdahl AC, Malmsten M, Mörgelin M, et al. The Thrombin-Derived Host Defense Peptide GKY25 Inhibits Endotoxin-Induced Responses through Interactions with Lipopolysaccharide and Macrophages/Monocytes. *The Journal of Immunology*. 2015 Jun 1;194(11):5397–406.
  83. Kasetty G, Papareddy P, Kalle M, Rydengård V, Walse B, Svensson B, et al. The C-Terminal Sequence of Several Human Serine Proteases Encodes Host Defense Functions. *JIN*. 2011;3(5):471–82.
  84. Liu H, Wang G, Hao D, Wang C, Zhang M. Antimicrobial and Immunoregulatory Activities of TS40, a Derived Peptide of a TFPI-2 Homologue from Black Rockfish (*Sebastes schlegelii*). *Mar Drugs*. 2022 May 26;20(6):353.
  85. Gómez-Guillén MC, Giménez B, López-Caballero ME, Montero MP. Functional and bioactive properties of collagen and gelatin from alternative sources: A review. *Food Hydrocolloids*. 2011 Dec 1;25(8):1813–27.
  86. Ennaas N, Hammami R, Beaulieu L, Fliss I. Purification and characterization of four antibacterial peptides from protamex hydrolysate of Atlantic mackerel (*Scomber scombrus*) by-products. *Biochemical and Biophysical Research Communications*. 2015 Jul 3;462(3):195–200.
  87. Lawrence CB, Joosten MH AJ, Tuzun S. Differential induction of pathogenesis-related proteins in tomato by *Alternaria solani* and the association of a basic chitinase isozyme with resistance. *Physiological and Molecular Plant Pathology*. 1996 Jun 1;48(6):361–77.

88. Chen YL, Lee CY, Cheng KT, Chang WH, Huang RN, Nam HG, et al. Quantitative Peptidomics Study Reveals That a Wound-Induced Peptide from PR-1 Regulates Immune Signaling in Tomato. *The Plant Cell*. 2014 Oct 1;26(10):4135–48.
89. Lee HR, You DG, Kim HK, Sohn JW, Kim MJ, Park JK, et al. Romo1-Derived Antimicrobial Peptide Is a New Antimicrobial Agent against Multidrug-Resistant Bacteria in a Murine Model of Sepsis. *mBio*. 2020 Apr 14;11(2):e03258-19.
90. You DG, Lee HR, Kim HK, Lee GY, Yoo YD. A Novel Peptide Derived from the Transmembrane Domain of Romo1 Is a Promising Candidate for Sepsis Treatment and Multidrug-Resistant Bacteria. *International Journal of Molecular Sciences*. 2021 Jan;22(15):8243.
91. Zanfardino A, Pizzo E, Di Maro A, Varcamonti M, D'Alessio G. The bactericidal action on *Escherichia coli* of ZF-RNase-3 is triggered by the suicidal action of the bacterium OmpT protease. *The FEBS Journal*. 2010;277(8):1921–8.
92. Samir P, Link AJ. Analyzing the cryptome: uncovering secret sequences. *AAPS J*. 2011 Jun;13(2):152–8.
93. Wang Z, Wang G. APD: the Antimicrobial Peptide Database. *Nucleic Acids Res*. 2004 Jan 1;32(Database issue):D590–2.
94. Meher PK, Sahu TK, Saini V, Rao AR. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Sci Rep*. 2017 Feb 13;7:42362.
95. Waghu FH, Gopi L, Barai RS, Ramteke P, Nizami B, Idicula-Thomas S. CAMP: Collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res*. 2014 Jan;42(Database issue):D1154-1158.
96. Lee HT, Lee CC, Yang JR, Lai JZC, Chang KY. A large-scale structural classification of antimicrobial peptides. *Biomed Res Int*. 2015;2015:475062.
97. Cerrato A, Capriotti AL, Capuano F, Cavaliere C, Montone AMI, Montone CM, et al. Identification and Antimicrobial Activity of Medium-Sized and Short Peptides from Yellowfin Tuna (*Thunnus albacares*) Simulated Gastrointestinal Digestion. *Foods*. 2020 Sep;9(9):1185.
98. Pane K, Durante L, Crescenzi O, Cafaro V, Pizzo E, Varcamonti M, et al. Antimicrobial potency of cationic antimicrobial peptides can be predicted from their amino acid composition: Application to the detection of “cryptic” antimicrobial peptides. *J Theor Biol*. 2017 Apr 21;419:254–65.
99. Ma Y, Yao A, Chen X, Wang L, Ma C, Xi X, et al. Generation of truncated derivatives through in silico enzymatic digest of peptide GV30 target MRSA both in vitro and in vivo. *Computational and Structural Biotechnology Journal*. 2021 Jan 1;19:4984–96.
100. Liu Z, Dong S, Xu J, Zeng M, Song H, Zhao Y. Production of cysteine-rich antimicrobial peptide by digestion of oyster (*Crassostrea gigas*) with alcalase and bromelin. *Food Control*. 2008 Mar 1;19(3):231–5.

## HIPÓTESE

Os criptídeos, peptídeos bioativos previamente presentes em sequências proteicas maiores, podem ser encontrados em uma vasta gama de proteínas suscetíveis à ação proteolítica de enzimas ou do meio (ex: pH, temperatura). Um meio propício para a liberação de criptídeos é o trato gastrointestinal do parasito *Schistosoma mansoni*, local conhecido pela presença de enzimas. Além do mais, sabe-se da presença de diversas proteínas no produto excretório do parasito, que são responsáveis pela modulação da resposta imune do hospedeiro (1). Entre essas proteínas, de ação modulatória da resposta imune e que pode estar presente nesse produto excretório, acredita-se ser a proteína SmKI-1, que inibe elastase neutrofilica e cuja presença no tegumento de *S. mansoni* já foi previamente comprovada na fase adulta (2) e na fase de esquistossômula (3).

A proteína SmKI-1 apresenta dois domínios, o domínio kunitz, que atua como inibidor de serino proteases, e um domínio C-terminal, cuja atividade é desconhecida e não tem similaridade com outras proteínas descritas nos bancos de dados (4). Sendo assim, subentende-se que o domínio C-terminal seja o domínio responsável pela adesão da SmKI-1 no tegumento de *S. mansoni* e hipotetiza-se que esse domínio seja clivado, em diferentes regiões, para a liberação do domínio Kunitz, gerando diversos peptídeos membrano-ativos. Como propriedade geral de peptídeos membrano ativos, como descrito na introdução, esses peptídeos podem ter ação antimicrobiana. A principal hipótese desta tese recai na possibilidade de que a clivagem do domínio C-terminal da proteína SmKI-1 para liberação do domínio kunitz seja responsável pela liberação de criptídeos antimicrobianos e que a identificação desses peptídeos pode ocorrer de modo integrado com métodos computacionais e biofísico, por meio de um fluxograma de trabalho que leve em conta a identificação, a verificação da ação antimicrobiana e pela caracterização estrutural dessas sequências.

No meio do estudo, nos deparamos com um problema que não estava no planejamento inicial: a plataforma de assinalamento de sinais de RMN e cálculo de estruturas que estava sendo utilizada não contemplava as modificações terminais, amidação do C-terminal e a acetilação do N-terminal. Devido à diversas outras vantagens que essa plataforma integrada apresenta, ao invés de trocar para outros programas para o cálculo de estrutura, acreditou-se que essa seria uma oportunidade para criar melhorias aos programas e adicionar, aos seus

respectivos códigos, modificações, além de criar um *patch* para adição das modificações protéicas, nos programas CcpnNMR Analysis v2.5 e Aria2.3.

### Referências

1. Lightowers MW, Rickard MD. Excretory-secretory products of helminth parasites: effects on host immune responses. *Parasitology*. 1988;96 Suppl:S123-166.
2. Ranasinghe SL, Fischer K, Gobert GN, McManus DP. Functional expression of a novel Kunitz type protease inhibitor from the human blood fluke *Schistosoma mansoni*. *Parasites & Vectors*. 2015 Aug 4;8(1):408.
3. Morais SB, Figueiredo BC, Assis NRG, Homan J, Mambelli FS, Bicalho RM, et al. *Schistosoma mansoni* SmKI-1 or Its C-Terminal Fragment Induces Partial Protection Against *S. mansoni* Infection in Mice. *Front Immunol*. 2018;9:1762.
4. Morais SB, Figueiredo BC, Assis NRG, Alvarenga DM, de Magalhães MTQ, Ferreira RS, et al. *Schistosoma mansoni* SmKI-1 serine protease inhibitor binds to elastase and impairs neutrophil function and inflammation. *PLoS Pathog*. 2018 Feb;14(2):e1006870.

## OBJETIVOS

### Objetivo geral

O objetivo geral deste estudo foi identificar modelos de criptídeos antimicrobianos no domínio C-terminal da proteína SmKI-1 de *S. mansoni* e caracterizar os componentes destes peptídeos responsáveis pela atividade antimicrobiana, além dos componentes que prejudicam essa ação. Além disso, objetivou-se aprimorar as plataformas de cálculo de estrutura, CcpNMR Analysis e Aria2.3, para levar em consideração, durante os cálculos, as modificações terminais amidação C-terminal e acetilação N-terminal.

### Objetivos específicos

- Criar um método de identificação de criptídeos em sequências proteicas
- Selecionar sequências para análise de atividade antimicrobiana
- Identificar a afinidade dos peptídeos por membranas aniônicas e zwitteriônicas
- Propor um modo de ação para os criptídeos
- Criar um *patch* para adição das modificações terminais no cálculo de estrutura

## CAPÍTULO 2. Identificação e Caracterização das schistocinas

### 2.1 Introdução

Os criptídeos são uma nova pletera de peptídeos com funções biológicas que estão escondidas nas estruturas primárias de várias proteínas (1). De acordo com essa ideia, as proteínas podem ter múltiplas funções e atividades, dependendo do processamento enzimático, alvo metabólico ou da sua concentração relativa. Alguns exemplos de criptídeos são os peptídeos derivados da hemoglobina, que têm várias atividades biológicas distintas daquelas descritas a proteína precursora (2–5), além dos peptídeos antimicrobianos (AMPs) da apolipoproteína-B, r(P)ApoBL<sup>124</sup>/r(P)ApoBs<sup>125</sup> (6), os derivados da lactoferrina (7) e os da trombina humana, como GKY25<sup>126</sup> (8). Esses peptídeos são gerados após a maturação de proteínas, processos metabólicos (9,10) ou durante os ensaios *in vitro* (11,12). A maioria desses peptídeos foi previamente negligenciada e talvez isso ocorra porque eles são ativos apenas após a ação proteolítica de enzimas, mudança de pH do ambiente ou podem estar ocultos dentro de conjuntos de dados proteômicos (13,14).

A diversidade de peptídeos biologicamente ativos encriptados em proteínas pode ser explorada por vários modelos teóricos e computacionais (15–17). Isso leva os pesquisadores a obter o máximo de informações das sequências de proteínas já conhecidas, examinando sua multifuncionalidade e remodelando seu papel na natureza, especialmente aquelas com domínios desconhecidos. Além disso, os peptídeos bioativos com potenciais aplicações terapêuticas (por exemplo, inflamação, câncer) podem ser descobertos (18).

Diversas abordagens aplicadas ao campo do *design* de peptídeos já foram propostas para recuperar criptídeos de moléculas naturais e, geralmente, consideram as propriedades da estrutura primária e as características tridimensionais (14,19–21). As propriedades físico-químicas dos peptídeos influenciam seu comportamento e a afinidade com seus alvos (22). O fator estrutural inclui não apenas o enovelamento peptídico (estrutura secundária), mas também a disposição dos resíduos na estrutura (anfipaticidade), e ambos são influenciados pelas interações iônicas e hidrofóbicas com o alvo (ex .: membrana biológica). Além disso, a citotoxicidade é refletida diretamente pela capacidade dos peptídeos de se ligar

---

<sup>124</sup> Sequência da r(P)ApoBL: PHVALKPGKCLKFIIPSPKRVPKLLSGGNTLHLVSTTKT

<sup>125</sup> Sequência da r(P)ApoBs: PHVALKPGKCLKFIIPSPKRVPKLLSG

<sup>126</sup> Sequência do GKY25: GKYGFYTHVFRLLKWKIQQVIDQFGE

e destruir células eucarióticas, que deve ser cuidadosamente regulada no momento do *design* (23,24). Durante a racionalização de peptídeos, é possível inserir tantos recursos quanto desejado em relação ao que pode ser crucial para uma atividade biológica específica em um peptídeo (25). No entanto, consolidar e analisar todas as informações não é simples, pois selecionar as propriedades é mais relevante do que inserir muitos recursos na etapa de racionalizar os peptídeos.

Alguns preditores para peptídeos antimicrobianos, incluindo iAMPpred (26), AmPEP (27), ClassAMP (28), amPEPpy 1.0 (29), AntiMPmod (30), MAMPs-Pred (31), AntiBP (32), são relatados com frequência. No entanto, essas ferramentas não levam em consideração o conceito de criptídeo em seus algoritmos, o que dificulta a descoberta de peptídeos ocultos. Como o conceito de criptídeo é inovador na exploração de proteínas e seus domínios correspondentes sem função conhecida, nos concentramos em proteínas expressas pelo parasita humano *Schistosoma mansoni*. Nos últimos anos, nosso grupo de pesquisa tem estudado as propriedades de uma dessas proteínas, o inibidor de tipo Kunitz de *S. mansoni* SmKI-1 (uniprot CCD77156) (33,34), uma proteína com um único domínio do tipo Kunitz, que também possui um domínio C-terminal, sem similaridade fora do gênero, com aplicabilidade vacinal contra infecção por cercárias de *S. mansoni* em modelo murino (35). A presença do SmKI-1 no tegumento do esquistossômulo (35,36) gerou a hipótese de que seu domínio C-terminal possui afinidade por membranas.

Para abordar essa possibilidade, recuperou-se do SmKI-1 um tipo muito específico de peptídeos membrano ativos: os AMPs. Essas classes de peptídeos são amplamente distribuídas na natureza e têm amplo espectro de atividade (antibiótico, antiinflamatório, imunomodulador) (18). A maioria dos AMPs são moléculas catiônicas, uma propriedade que culmina em uma grande afinidade com membranas carregadas negativamente. Sua interação com as membranas pode ser por meio de: barril, onde a interação de peptídeos cria poros na membrana; toroidal, com formação de poros sem interação de peptídeos; modelo de tapete, onde os peptídeos não entram na célula, mas desestabilizam a membrana; ou semelhantes a detergentes, formando micelas com os lipídios da membrana (37–40). O conhecimento desses processos culminou no estudo de milhares de peptídeos antimicrobianos nos últimos anos e permitiu classificá-los de acordo com suas propriedades físico-químicas, como o conteúdo estrutural, a hidrofobicidade, a carga líquida e a estrutura em membrana (41,42).

A maioria dos peptídeos antimicrobianos sintéticos são projetados como moléculas catiônicas com interfaces hidrofílicas e hidrofóbicas para permitir a interação com as cabeças e caudas fosfolipídicas, garantindo uma boa interação peptídeo-membrana. No entanto, isso não é o estado da arte, e o equilíbrio entre as diferentes propriedades físico-químicas precisa ser avaliado: os peptídeos extremamente carregados podem ser tóxicos para as células eucarióticas; os peptídeos hidrofóbicos podem ser difíceis de administrar e a mistura desses recursos também pode culminar em AMPs ineficientes. Isso reforça a necessidade de estudar as características dos AMPs e a relevância do *design* de peptídeos. Outro fator a ser levado em consideração é a estratégia de encurtamento de peptídeos, que visa melhorar a atividade do peptídeo apenas mantendo na sequência o núcleo do peptídeo e removendo resíduos que poderiam prejudicar o melhor desempenho do peptídeo (43,44). A redução pode ser realizada removendo resíduos N-terminal ou C-terminal e as abordagens levam em consideração a anfipaticidade do peptídeo, a carga e outras características físico-químicas. Por último, mas não menos importante, a substituição de resíduos também pode ser a chave na racionalização de peptídeos, uma vez que a identificação de regiões estratégicas para modificação é fundamental na melhoria do peptídeo.

Neste trabalho, o principal objetivo foi identificar criptídeos escondidos no domínio C-terminal SmKI-1 e avaliar suas potenciais aplicações biotecnológicas como peptídeos antimicrobianos. O algoritmo *encrypted* possui 18 atributos de propriedades físico-químicas e o uso de um algoritmo de aprendizado de máquina permitiu descrever uma nova classe de criptídeos antimicrobianos: as schistocinas. Para apoiar essa hipótese, testou-se sua atividade antimicrobiana e realizou-se a caracterização biofísica usando dicroísmo circular, ressonância magnética nuclear, calorimetria de titulação isotérmica e dinâmica molecular para acessar sua relação estrutura-função. Os dados sugerem que a abordagem computacional e experimental utilizada foi capaz de descobrir novos criptídeos de um domínio desconhecido de SmKI-1 e é eficaz para ser aplicado a muitas outras proteínas ou domínios com atividade desconhecida.



## 2.2 Material e Métodos

### 2.2.1 Preditor

A figura 1 ilustra o fluxograma de construção do preditor de peptídeos antibacterianos. Com o intuito de gerar um preditor que fosse capaz de discriminar sequências proteicas em antibacterianas e não antibacterianas, foram utilizados os dados da base APD, como controle positivo, e da base PepBank, como controle negativo. Para a classificação binária em peptídeos antibacterianos e não antibacterianos, foram calculados atributos a partir das sequências proteicas que são as propriedades físico-químicas presentes na Tabela 1.

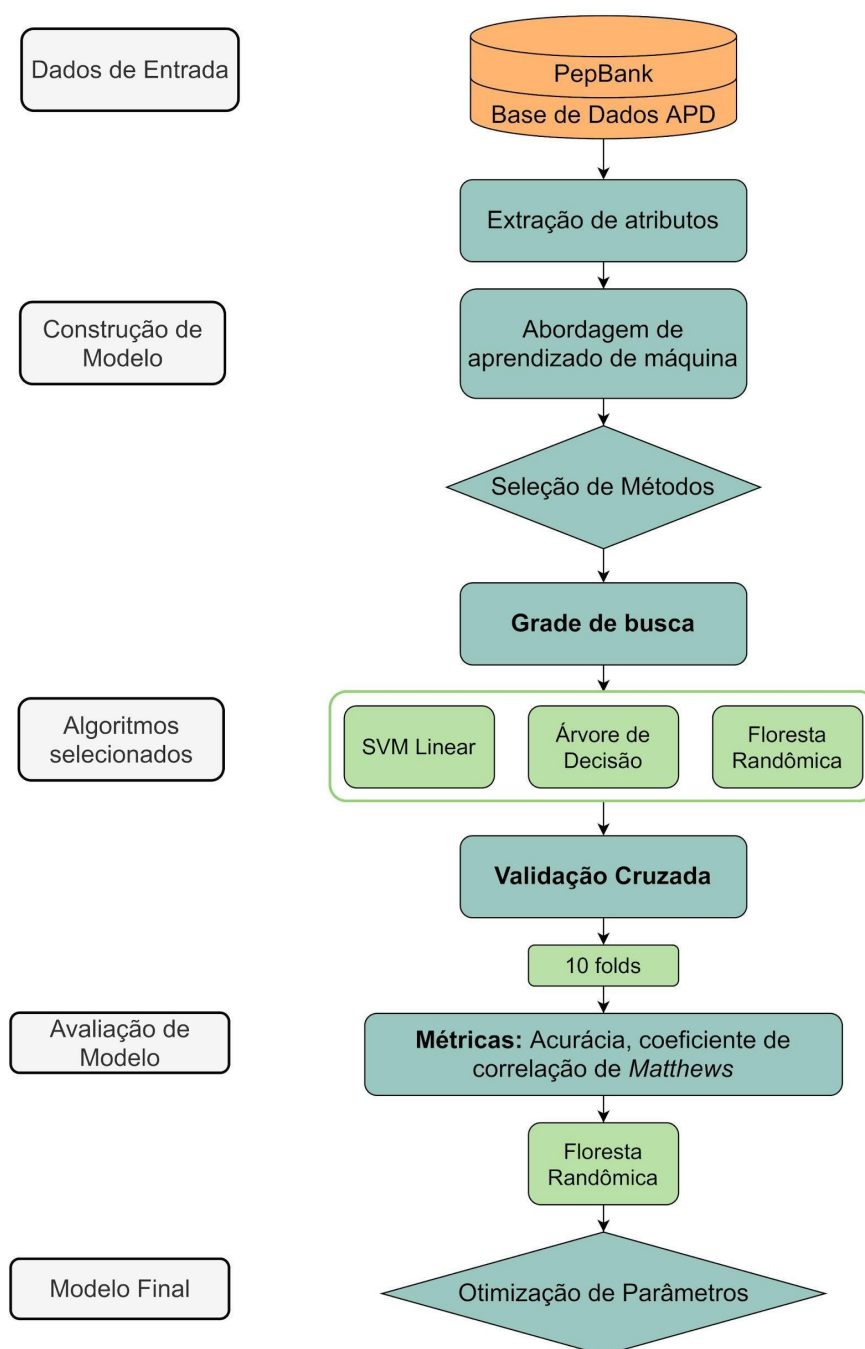
**Tabela 1 - Atributos do classificador.**

	Atributos	Referências
1	Massa molecular	-
2	Ponto isoelétrico	(45)
3	Hidrofobicidade	(46)
4	Momento hidrofóbico	(47)
5	Resíduos ácidos (%)	-
6	Resíduos básicos (%)	-
7	Resíduos carregados (%)	-
8	Resíduos polares (%)	-
9	Resíduos apolares (%)	-
10	Resíduos aromáticos	-
11	Coefficiente de alfa-hélice	(48)
12	Coefficiente de folha beta	(48)
13	Coefficiente de randômicos	(48)
14	Volume	(49)
15	Porcentagem enterrada	(50)
16	Acessibilidade da cadeia principal ao solvente	(51)
17	Acessibilidade de resíduos apolares ao solvente	(51)
18	Acessibilidade de resíduos polares ao solvente	(51)

Para a construção de um modelo preditor, três algoritmos de aprendizado de máquina foram previamente selecionados: SVM linear, árvores de decisão e floresta randômica. Uma grade de busca foi executada junto à validação cruzada, para buscar os melhores parâmetros para cada algoritmo.

O objetivo do SVM linear é criar um hiperplano que separa os dados em classes. Os parâmetros testados para o SVM foram: parâmetro de regularização (do inglês “regularization parameter” (‘C’) [0.001, 0.01, 0.1, 1, 10] e a perda de função (‘loss’) [“squared hinge”, “hinge”]. A árvore de decisão consiste de um método de classificação que particiona binariamente e recursivamente os dados e é composto de nós, arestas e folhas. Os parâmetros testados para este algoritmo foram: a função de critério (‘criterion’) [“gini”, “entropy”], a estratégia de divisão (‘splitter’) [“best”, “random”], a máxima profundidade da árvore (‘max\_depth’) [3, 4, 5, 7, 8, 9], o número de características por divisão (‘max\_features’) [3, 5, 8, 9, 10]. A floresta randômica tem os mesmos princípios que a árvore de decisão, mas em larga escala, onde cada árvore testa uma classe de predição. Os parâmetros testados foram o número de árvores (‘n\_estimators’) [50, 100, 200], a função de critério (‘criterion’) [‘gini’, ‘entropy’] e a máxima profundidade de uma árvore (‘max\_depth’) [4, 5, 6, 7, 8].

A validação cruzada utiliza diferentes subconjuntos dos dados para o treinamento e para o teste, o que permite reduzir vieses com porções dos dados. Foram realizadas dez iterações com os hiperparâmetros e os três algoritmos selecionados (Figura 1). Os dados foram divididos em 10 partes e em cada iteração, foram utilizadas nove partes para o treinamento e uma parte para o teste. Os resultados da última décima iteração foram utilizados.



**Figura 1 - Fluxograma de desenvolvimento do modelo preditor de peptídeos antibacterianos.** Na primeira etapa, um banco de dados de peptídeos antimicrobianos (APD) e um banco com peptídeos com funções distintas, como apoptose e câncer (PepBank) foram selecionados para treinamento do classificador supervisionado. Na segunda etapa, foram calculadas 18 características físico-químicas dessas sequências que serviram como atributos para treinamento/teste de classificadores binários: SVM linear, árvore de decisão e floresta randômica. Na terceira etapa, uma validação cruzada realizou 10 iterações com os três classificadores usando nossa base de dados e a última iteração foi selecionada para escolha do algoritmo de melhor performance, cujas métricas avaliadas foram acurácia e coeficiente de correlação de Matthews. O modelo escolhido, RF, foi otimizado e extraído para ser utilizado como preditor.

### 2.2.2 Métricas de performance dos modelos

Para avaliar as performances de todos os classificadores, foram utilizadas as seguintes métricas: acurácia (Acc) e coeficiente de correlação de Matthews (MCC).

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP) \times (VP + FN) \times (VN + FP) \times (VN + FN)}} \quad (2)$$

Onde VP é o número de verdadeiros positivos, VN o número de verdadeiros negativos, FP é o número de falso positivos e FN é o número de falso negativos.

### 2.2.3 Análises de sequências

A disposição dos resíduos polares e apolares em estrutura secundária foi verificada por meio do arranjo em roda de hélice (52). O alinhamento de sequências foi performed utilizando o alinhador MUSCLE (53) e o visualizador de alinhamentos uGENE (54).

### 2.2.4 Ensaios antimicrobianos

Os ensaios antibacterianos - *Escherichia coli* (ATCC 25922), *Pseudomonas aeruginosa* (ATCC 27853) e *Staphylococcus aureus* (ATCC 29213) da American Type Culture Collection (Manassas, VA, EUA), fornecida pelo Instituto Nacional de Controle de Qualidade em Saúde (INCQS, FIOCRUZ, Rio de Janeiro, Brasil), foram cultivadas a 37 °C em placa de ágar contendo meio BHI (Brain Heart Infusion) da HiMedia Laboratories (Mumbai, Índia) por 12 horas. Posteriormente, as colônias isoladas foram selecionadas e transferidas para um meio de cultura Mueller-Hinton de Kasvi (São José do Pinhais, PR Brasil) de acordo com o Comitê Nacional de Padrões Clínicos e Laboratoriais (NCCLS) para a realização dos bioensaios (protocolos do Instituto NCCLS) [57]. O péptido foi diluído 10 vezes em caldo líquido Mueller-Hinton. A maior concentração de péptido utilizada foi de 264 µM em uma suspensão composta por aproximadamente 5x10<sup>5</sup> UFC.mL<sup>-1</sup> (unidades formadoras de colônia.mL<sup>-1</sup>), e o volume final foi de 100 µL por poço 1:1 (péptido:inóculo). Os experimentos foram realizados por 20 horas em cultura estacionária a 37 °C, e a densidade

óptica foi medida (A 595 nm). A ampicilina e o cloranfenicol foram utilizados como antibióticos padrão para o controle da morte celular, e o controle positivo foi obtido pela adição de 50 µL do inóculo e 50 µL do meio de cultura estéril. Além disso, usou-se um controle interno do peptídeo antimicrobiano DS-01 peptídeo antimicrobiano [58].

Os ensaios antifúngicos - as leveduras *Candida albicans* ATCC 10231 (resistentes a Anidulafungina, Voriconazol, Itraconazol e Fluconazol), *Candida krusei* ATCC 6258, *Candida tropicalis* ATCC 750, *Candida parapsilosis* ATCC 90018 e *Candida glabrata* ATCC 2001, fornecidos pelo Instituto Nacional de Controle de Qualidade na Saúde (INCQS, FIOCRUZ, Rio de Janeiro, Brasil), para atividade antifúngica. A concentração inibitória mínima (MIC) foi determinada por um teste de sensibilidade para microdiluição em caldo para a determinação de MIC para leveduras (CSLI 2012) [59], com modificações. Os testes também foram realizados em caldo Sabouraud dextrose (Kasvi, Itália). As diluições em série de 0,5 a 0,9765 mg·mL<sup>-1</sup> foram preparadas a partir de compostos em placas de microtitulação de 96 poços estéreis. A densidade de leveduras por poço correspondeu a 1x10<sup>3</sup> a 5x10<sup>3</sup> CFU·mL<sup>-1</sup>. As placas foram incubadas a 35 ± 2 °C por 44 h, e posteriormente avaliadas por leitura visual. O controle positivo foi a nistatina (50 a 0,0225 µg·mL<sup>-1</sup>). Os controles também foram preparados para a avaliação da viabilidade dos microrganismos de teste e a avaliação da esterilidade do meio de cultura. Todos os testes foram realizados em triplicata.

## 2.2.5 Espectroscopia de Dicroísmo Circular

As preferências conformacionais das schistocinas (schis-1, schis-2, schis-3 e schis-3.1) em soluções tampão (50 mM Tris-HCl, pH 7,0 e 100 mM NaCl), em soluções de mistura TFE:água, na presença de micelas de dodecilsulfato de sódio (SDS) e LUVs foram avaliados por espectroscopia de dicroísmo circular no espectropolarímetro JASCO® J-810 (Tóquio, Japão), equipado com o sistema de controle de temperatura Peltier Jasco® - PFD-425S (Tóquio, Japão). Os espectros de CD foram registrados a 20 °C usando uma cubeta de quartzo retangular com caminho de 0,1 mm (Uvonic Instruments, Plainview, NY). Todos os espectros foram registrados com três varreduras de 260 a 190 nm usando uma largura de banda espectral de 1,0 nm, velocidade de varredura de 200 nm·min<sup>-1</sup> e tempo de resposta de 1s. Experimentos semelhantes com as respectivas soluções em branco também foram realizados a fim de permitir a subtração do fundo. Todas as amostras foram preparadas com concentração

de peptídeo de 30  $\mu\text{M}$ . Os espectros foram analisados usando o *software Spectra Analysis*, para subtração e suavização do branco, e algoritmo próprio para deconvolução usando a seguinte fórmula:

$$\theta, \text{ em graus} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}, = \frac{\text{miligraus} \times \left( \frac{\text{peso molecular}}{\text{número de amidas da cadeia principal}} \right)}{\text{caminho óptico [mm]} \times \text{concentração [mg/mL]}} \quad (3)$$

O cálculo da elipticidade foi feito como previamente descrito (55). A deconvolução dos espectros e a obtenção dos valores médios de helicidade foi feita pelo programa K2D2 (56).

### 2.2.6 Titulação de Calorimetria Isotérmica (ITC)

Os experimentos de ITC foram realizados à 35 °C com um VP-ITC Microcalorimeter® de alta precisão (Malvern Instrument Ltd, Worcestershire, UK) com soluções de peptídeo à 50 mM em Tris-HCl 50 mM, pH 8,5, contendo NaCl a 100 mM. O microcalorímetro foi eletricamente e quimicamente calibrado antes de realizar os experimentos (57). Todas as soluções e tampões empregados nos experimentos foram filtrados e desgaseificados sob pressão reduzida (140 mbar) durante cinco minutos. As soluções de peptídeo na célula do calorímetro foram tituladas com 51 injeções sucessivas de LUVs de POPC 20 mM ou POPC:POPG (3:1, mol:mol). A primeira injeção de 1  $\mu\text{L}$  foi descartada para eliminar os efeitos de difusão do material da seringa para a célula do calorímetro, e essa injeção foi seguida por cinquenta injeções de 5  $\mu\text{L}$ . Os tempos de injeção de 2 s com intervalos de 240 s foram usados no experimento. Os experimentos de base foram realizados titulando a solução tampão com as respectivas suspensões de LUVs para determinar os calores de diluição correspondentes. Para medir o calor produzido pelas interações peptídeo-lipídeo, cada pico da curva calorimétrica foi integrado seguindo procedimentos padrão descritos por Wieprecht e colaboradores (58). A concentração total de lipídios foi usada para estimar o grau de associação da membrana, bem como na determinação dos parâmetros de ligação termodinâmica. Os dados brutos foram analisados com o *software Microcal Origin 5.0* (OriginLab Corporation, Northampton, MA) para ITC, fornecido juntamente com o equipamento.

### 2.2.7 Ressonância Magnética Nuclear (RMN)

As amostras para os experimentos de RMN foram preparadas dissolvendo 1 mM de cada peptídeo em uma solução micelar contendo 100 mM de dodecilsulfocolina deuterada (DPC-d<sub>38</sub>), 10% de D<sub>2</sub>O e 2,2-dimetil-2-silapentano-5-sulfonato (DSS) a 1,0 mM foi usado como referência interna. O pH foi ajustado, usando tampão fosfato, para 3,5 usando tampão fosfato. Os espectros de RMN bidimensionais (2D) de schis-1, schis-2, schis-3 e schis-3.1 foram registrados em um espectrômetro de RMN Bruker AVANCE III 500 operando a 500,013 MHz e equipado com uma sonda de 5 mm de banda larga inversa (BBI) (Instituto de Química da Universidade Federal de Goiás, Goiânia, Brasil) em 300K. A supressão de água foi alcançada usando técnicas de pré-saturação. Os espectros de espectroscopia de correlação total (TOCSY) foram adquiridos usando a sequência de pulso DIPSI-2 (59) com supressão de água (dipsi2esgpph) com um tempo de spin-lock de 80 ms. A largura espectral foi de 9615 Hz, incrementos de 512 t1 foram coletados com 16 transientes de 4096 pontos. Os espectros de espectroscopia Nuclear Overhauser (NOESY) foram adquiridos usando uma sequência de pulso (60) com supressão de watergate (noesygpwh5) e tempos de mistura de 150 e 200 ms. A largura espectral foi 9615 Hz, 512 incrementos t1 foram coletados com 16 transientes de 4096 pontos para cada FID. Os espectros de coerência quântica única heteronuclear (HSQC) de <sup>1</sup>H-<sup>13</sup>C foram adquiridos no modo editado com HSQC editado por multiplicidade de fase sensível (61) usando sequência de eco-antieco (hsqcedetgp) com larguras espectrais F1 e F2 de 72010 Hz e 9615 Hz, respectivamente. Ao todo, 128 incrementos de t1 foram coletados com 128 transientes de 1024 pontos. Os experimentos de HSQC permitiram ao usuário observar grupos metil de diferentes maneiras: correlações CH e CH<sub>3</sub> mostram fase positiva e correlações CH<sub>2</sub> mostram fase negativa.

### 2.2.8 Análise de dados de RMN e cálculo das estruturas

Todos os dados de RMN foram processados utilizando o *software* TopSpin 4.1 (licença acadêmica, Bruker, Alemanha). As atribuições de ressonância de <sup>1</sup>H bidimensional foram realizadas analisando-se, simultaneamente, os espectros 2D <sup>1</sup>H-<sup>1</sup>H TOCSY e NOESY, conforme descrito por Wüthrich (62), usando o software de análise CcpNmr Analysis (63). As intensidades de NOEs obtidas em tempos de mistura de 200 ms foram convertidas em

restrições de distância semiquantitativas usando calibrações por Hyberts e colaboradores (64). Os limites superiores das distâncias assim obtidas foram 2,8, 3,4 e 8,0 Å (para NOE forte, médio e fraco, respectivamente). As restrições de ângulos diedro ( $\varphi$  e  $\psi$ ) foram obtidas a partir da análise dos deslocamentos químicos  $C\alpha$ ,  $H\alpha$ ,  $C\beta$  e  $H\beta$  com o *software* DANGLE, parte da plataforma CcpNmr (65). Os cálculos de estruturas foram realizados usando o *software* Aria2.3, juntamente com o CNS (66,67). Começando com uma conformação estendida, 200 estruturas foram geradas usando um protocolo de anelamento simulado. Isto foi seguido por 20000 etapas de anelamento simulado a 1000 K e uma diminuição subsequente na temperatura em 15000 etapas no primeiro estágio de anelamento de resfriamento lento. As vinte estruturas de menor energia foram refinadas usando um protocolo sa\_new.inp e a qualidade estereoquímica foi analisada por WhatCheck (68), Procheck (69). A exibição, análise, contagem de ligações de hidrogênio potenciais e manipulação das estruturas tridimensionais foram realizadas com o *software* Chimera (UCSF) (70). O cálculo da energia potencial de solvatação foi medido convertendo os arquivos PDB em PQR, que serviram de entrada para o APBD, uma calculadora Solver Poisson-Boltzmann (71,72).

### 2.2.9 Troca de Hidrogênio-Deutério por RMN

Os dados de troca de hidrogênio foram registrados no espectrômetro de RMN Bruker AVANCE III 500 operando a 500,013 MHz (Instituto de Química da Universidade Federal de Goiás, Goiânia, Brasil). Os experimentos de troca de hidrogênio-deutério foram iniciados usando peptídeo (1 mM) e lípidos DPC- $d_{38}$  liofilizados (100 mM) em 95% de  $D_2O$ . Em seguida, uma série de espectros de TOCSY foram adquiridos para monitorar o decaimento dos sinais HN- $H\alpha$  conforme os sinais de hidrogênio foram substituídos pelos sinais de deutério em intervalos de 1,33 horas. Os espectros foram adquiridos com supressão de água usando sequência de pulso dipsi2esgpph com tamanho de dados no domínio do tempo de 4.096, 512 varreduras, 16 varreduras simuladas e 32 número de varreduras, com um tempo de bloqueio de rotação de 80 ms. Os dados foram ajustados para decaimento exponencial usando a função de ajuste  $A.exp(-Bx)$ . Os dados foram processados usando TopSpin 4.1 (Bruker) e analisados usando análise CcpNMR para atribuir e acompanhar as mudanças de intensidade.



### 2.2.10 Sistemas lipídicos para simulação

Para estudos de peptídeos em membrana zwitteriônica por simulação de dinâmica molecular (MD), uma bicamada de POPC com 40 moléculas de lipídios em cada monocamada foi montada a partir do sistema CHARMM-GUI *input generator* (73,74), na função de construção de membranas (75–77). Foram utilizadas 80 moléculas lipídicas e 3751 moléculas de água. Para os estudos em membrana aniônica, uma bicamada POPC:POPG (3:1) com 30 moléculas de POPC e 10 moléculas de POPG, aleatoriamente distribuídas, em cada monocamada foi formada também a partir do CHARMM-GUI. Para esse sistema, foram utilizadas 60 moléculas de POPC, 20 moléculas de POPG e 3613 moléculas de água. A área de POPC é de 68.3 Å e a área de POPG é de 62 Å. A área total média é de 2662 Å. O método de troca por íons do CHARMM foi usado para neutralizar a carga da bicamada aniônica com 28 moléculas de potássio e 8 moléculas de cloro. Para estudos com micelas de dodecilmfosfolina (DPC), o sistema de construção de micelas do CHARMM-GUI *input generator* também foi utilizado. Os sistemas foram minimizados em 1000 passos e equilibrados em condições normais de temperatura e pressão em 600 ps.

### 2.2.11 Sistemas peptídeo-lipídeo para simulação

Os peptídeos foram estudados de modo estruturado, a partir dos dados de cálculo de estrutura por RMN, e desenovelados, a partir de estruturas geradas no UCSF Chimera com ângulos planares (70). Essas estruturas foram usadas para construir os sistemas contendo os peptídeos schistocinas (schis-1, schis-2, schis-3 e schis-3.1) e as bicamadas lipídicas, como descrito no tópico 3.2.1. Os sistemas foram neutralizados com contra íons considerando as diferentes cargas de cada sistema. Os peptídeos foram colocados em aproximadamente 30 Å de distância dos sistemas peptídeo-vesículas, e em contato próximo com as cabeças polares da camada superior, nos sistemas peptídeo-bicamada, como sugerido no trabalho de Wang e colaboradores (2012) (78).

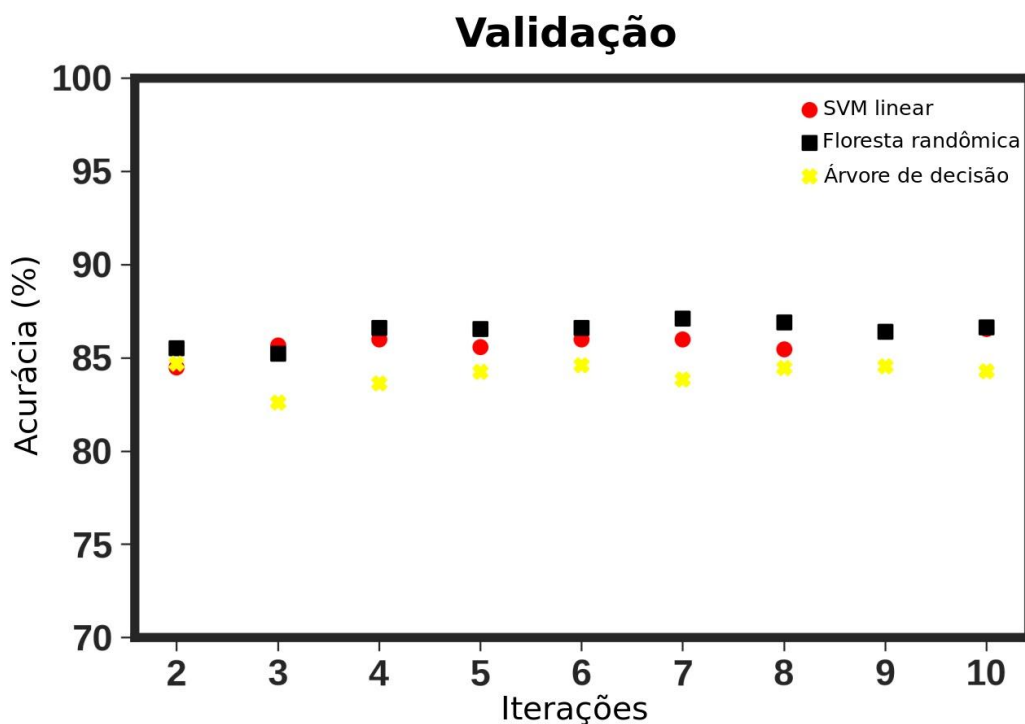
### 2.2.12 Simulações de dinâmica molecular

Um total de 16 sistemas peptídeos-lipídeos foram criados e, para cada um, foram realizadas simulações em triplicatas de 500 ns, culminando com um total de 48 simulações. Todas as simulações foram realizadas em condições de pressão e temperaturas constantes, em temperatura de 300K. O campo de força CHARMM 36 (79) foi utilizado dentro do programa Gromacs (80) para realizar as simulações de MD. O algoritmo *steepest-descent* foi utilizado para minimização de energia e a temperatura mantida a 300K durante a corrida, usando dinâmica de Langevin, enquanto a pressão foi mantida a 1 bar pelo algoritmo de Berendsen (81).

## 2.3 Resultados

### 2.3.1 Encrypted, uma plataforma para obter criptídeos antimicrobianos

Com o intuito de criar um preditor de peptídeos antibacterianos, testamos três algoritmos com dados da base APD e Pepbank. A figura 2 mostra os testes de performance dos três algoritmos. A acurácia dos algoritmos manteve-se praticamente linear durante as iterações e terminou com valores de 87,01% para o SVM linear e floresta randômica (Figura 2). O mesmo comportamento se observa com os dados do coeficiente de correlação de Matthews, que se mantiveram no entorno de 0,73 (Tabela 2). Como os resultados foram muito próximos, escolheu-se continuar com a construção do modelo utilizando-se o algoritmo de floresta randômica. Uma nova grade de busca foi realizada no processo de hiperparametrização e o preditor foi finalizado.



**Figura 2 - Performance dos algoritmos durante o processo de validação cruzada.** Três algoritmos foram testados: SVM Linear (bola vermelha), floresta randômica (quadrado preto) e árvore de decisão (cruzado amarelo). Duas métricas foram utilizadas: Acc (A) e MCC (B).

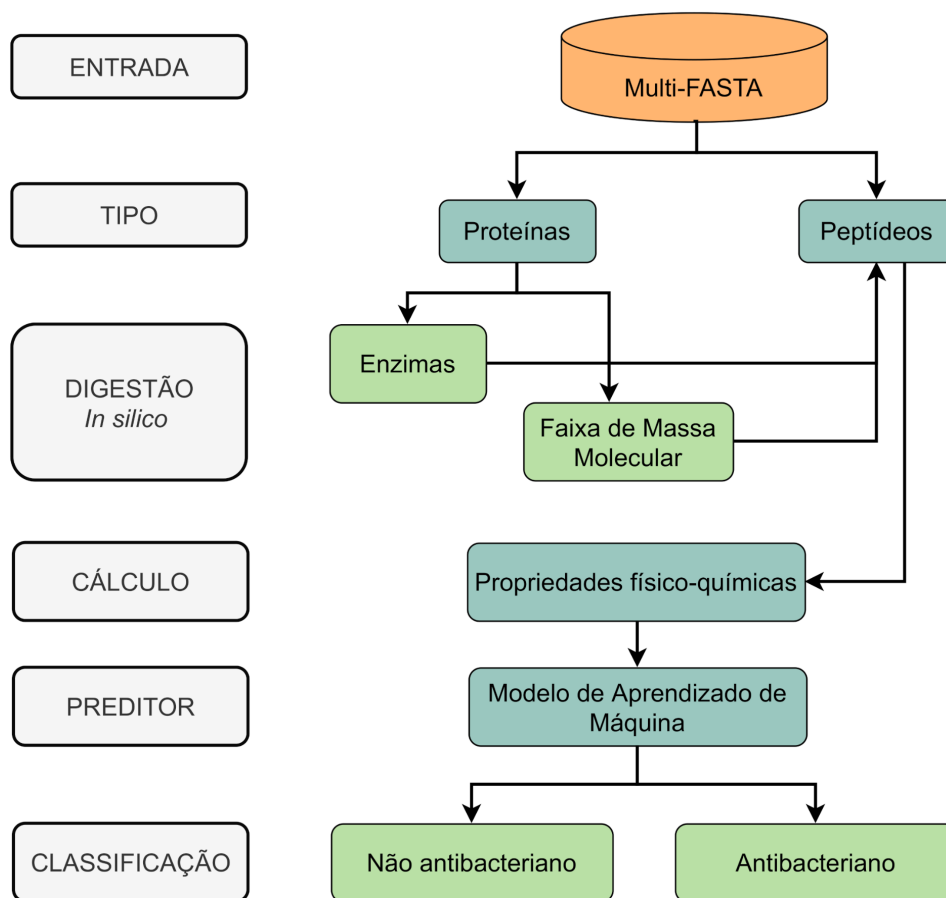
Com o intuito de comparar a performance do nosso classificador com outros preditores de peptídeos antibacterianos, selecionou-se os dados de Acc e MCC de modelos cujos dados estão disponíveis na literatura. A tabela 2 resume os dados dos preditores AMPEP, amPEPpy, antiMPmod, MAMP-Pred e Anti-BP. Os dados variam de 89.89% (antiMPmod) a 96% (AMPEP e amPEPpy) de acurácia. O *encrypted* apresenta os menores dados de Acc (87,01%) e MCC (0,73). Entretanto, apesar dos menores índices, os valores são assertivos e favoráveis, o que permite realizar a tarefa de classificação com segurança.

**Tabela 2 - Performance de comparação dos preditores.** Informações de performance de outros preditores foram obtidas de dados publicados. Métrica avaliadas: acurácia (em porcentagem) e coeficiente de correlação de Matthew. \*Anti-BP não publicaram dados de coeficiente de correlação de Matthews.

Preditores	Acurácia	MCC
enCrypted	87,01%	0,73
AMPEP	96%	0,9
amPEPpy	96%	0,93
antiMPmod	89,89%	0,8
MAMP-Pred	93,91%	0,88
Anti-BP	92,11%	-

Para identificar criptídeos antimicrobianos escondidos em sequências proteicas, usando um programa simples e direto, nosso grupo de pesquisa decidiu escrever, em linguagem python, um programa próprio. Considerando satisfatória as métricas do nosso modelo de classificador, o mesmo foi incorporado na construção do programa *encrypted*. A figura 3 resume as funcionalidades do programa. Primeiramente, o programa recebe como entrada arquivos fasta que podem conter uma ou mais sequências proteicas. O usuário indica se as sequências são proteicas ou peptídicas, pois isso determina as próximas etapas. Caso a opção escolhida seja “proteína”, o usuário indica o método de proteólise de preferência: usando janelas deslizantes que quebram as proteínas em fragmentos proteicos com tamanhos específicos ou se escolhe alguma enzima para simular um proteólise enzimática. As próximas etapas são semelhantes para proteínas e peptídeos: o cálculo das propriedades físico-químicas e a predição da atividade peptídica. O resultado final é uma tabela no formato “csv” (do

inglês, *comma separated values*) com os dados calculados e preditos. O código do programa foi depositado na plataforma github com o endereço: <https://github.com/bposantos/encrypted>.

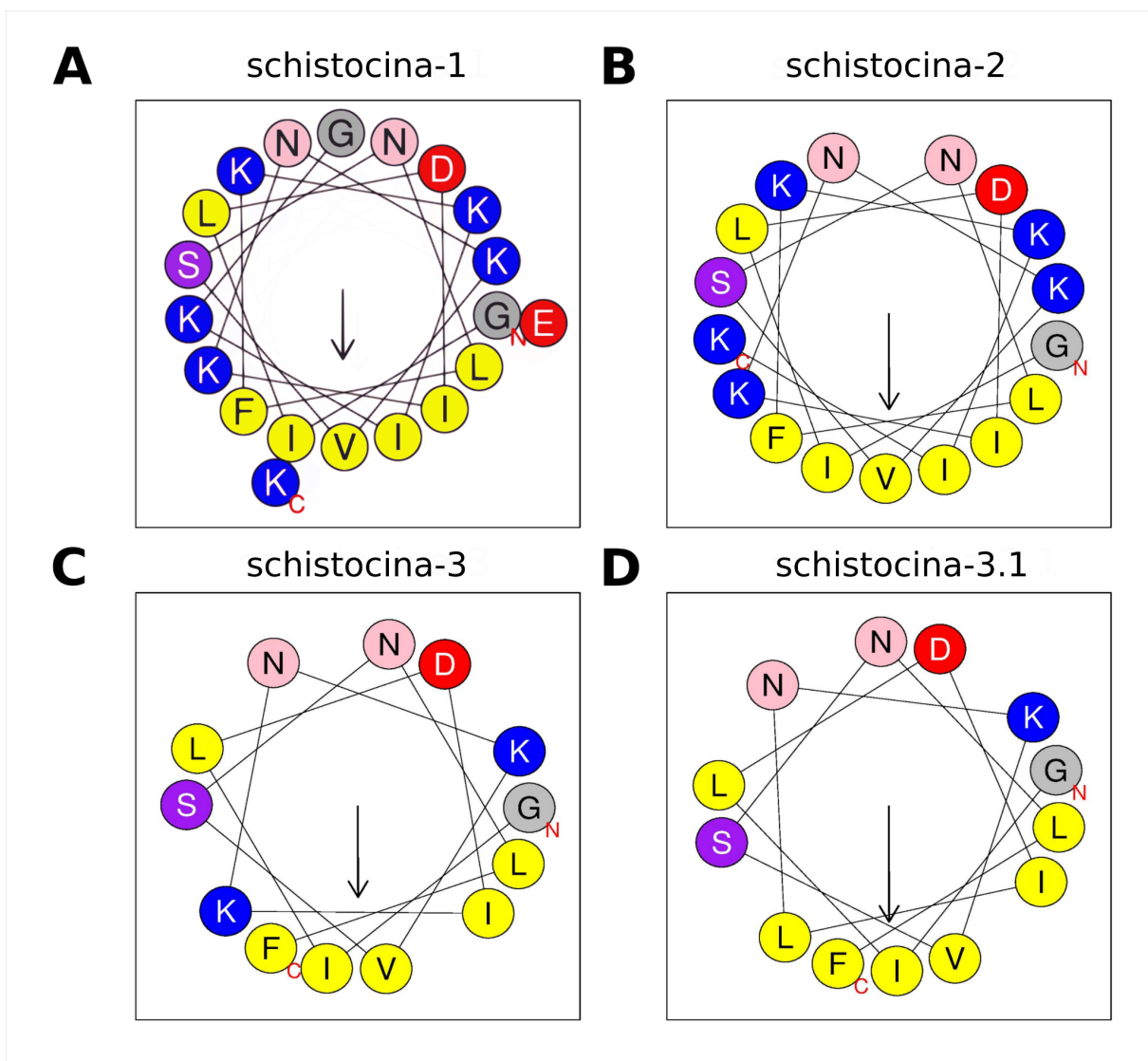


**Figura 3 - Fluxograma de funcionamento do enCrypted.** Na primeira parte, as sequências proteicas são inseridas pelo usuário num arquivo FASTA. O conteúdo do arquivo é indicado como PROTEÍNA ou PEPTÍDEO. Em caso de proteínas, o passo seguido pelo programa é a proteólise *in silico*, que precisa ser selecionada pelo usuário: fragmentos peptídeos de tamanho pré-definido, ação enzimática ou ambos. Na etapa seguinte, os procedimentos são iguais para arquivos FASTA do tipo proteína ou peptídeos, uma vez que agora ambos contém peptídeos. O modelo preditor de peptídeos antibacterianos realiza a predição e gera uma tabela CSV contendo os peptídeos antibacterianos e não-antibacterianos.

### 2.3.2 Análises *in silico*

Para a análise do programa, utilizou-se uma proteína com um domínio que tem função desconhecida, mas que parece ser responsável por aderir a proteína em membranas. Como peptídeos antimicrobianos são, geralmente, membrano-ativos, hipotetizou-se que o domínio C-terminal da proteína SmKI-1 seja fonte potencial de peptídeos antimicrobianos. O domínio

C-terminal da SmKI-1 foi digerido em tamanhos entre 1400-2300 Da, o que resultou em 183 peptídeos antibacterianos e 265 peptídeos não antibacterianos. Primeiro, o peptídeo com maior similaridade com os peptídeos da base APD foi selecionado, após ser realizado um P-BLAST. Posteriormente, utilizando uma estratégia de minimização, foram selecionados outros dois peptídeos com atividade antibacteriana predita, denominados schistocinas (Figura 4).



**Figura 4 - Representação das schistocinas em rodas de hélices.** Cores: amarelo para hidrofóbicos, violeta para Ser, roxo para Lys, vermelho para acídicos, rosa claro para Asn e cinza para resíduos pequenos (Ala e Gly).

Ao observar as cargas dos resíduos em roda de hélice, observou-se que a schis-1<sup>127</sup> apresenta anfipaticidade parcial, com uma face hidrofóbica composta pelos resíduos 2Ile, 5Ile, 9Val, 12Leu, 13Phe e 16Ile. Um resíduo hidrofóbico adicional está na face hidrofílica, o 3Leu, separado da face hidrofóbica por duas lisinas, 6Lys e 17Lys. A face hidrofílica possui quatro resíduos carregados positivamente, as lisinas 6Lys, 8Lys, 14Lys e 17Lys. Além disso, nessa porção estão os resíduos 1Gly, 3Leu, 4Asp, 7Asn, 10Ser, 11Asn, 18Gly e 19Glu. A quinta lisina, a 20Lys, está na face hidrofóbica. Com isso, destaca-se a presença de uma face hidrofílica maior que a face hidrofóbica e três lisinas que estão prejudicando uma face completamente hidrofóbica e reduzindo a anfipaticidade (Figura 4a).

Com relação à schis-2<sup>128</sup>, a predição da estrutura em hélice mostra uma face hidrofílica com cinco lisinas e um resíduo hidrofóbico, a 3Leu, interposto. A face hidrofóbica é a mesma da schis-1: 2Ile, 5Ile, 9Val, 12Leu, 13Phe e 16Ile. Os resíduos perdidos da schis-1 para schis-2 foram: 20Lys, o que aumentou a hidrofobicidade da face hidrofóbica, 19Glu, que estava na face hidrofílica e a 18Gly, também presente na face hidrofílica, aproximando os resíduos carregados (Figura 4b).

Já a schis-3<sup>129</sup> perdeu três resíduos de lisina: 14Lys, 15Lys e 17Lys. Além disso, o peptídeo perde o resíduo 16Ile. Assim, os criptídeos schis-1 e schis-2 que tinham uma face predominantemente hidrofílica, passam a distinguir predominantemente da schis-3, que possui uma maior quantidade de resíduos hidrofóbicos. Agora, o resíduo 3Leu, que estava prejudicando a face hidrofílica dos peptídeos schis-1 e schis-2 pode ser visto de outra maneira: como parte da face hidrofóbica, que estaria sendo prejudicada pelo resíduo 6Lys. Assim, soluções para a anfipaticidade desse peptídeo passam ou pela substituição do resíduo 3Leu por um resíduo hidrofílico, ou pela substituição da 6Lys por um resíduo hidrofóbico (Figura 4c).

Ao realizar uma análise comparativa das sequências das schistocinas com os peptídeos antibacterianos mais similares, obtidos na base de dados APD, observou-se uma similaridade da região N-terminal dos criptídeos com os peptídeos antimicrobianos conhecidos: ranateurinas, brevininas, uperinas e esculetinas (Figura 5). Os resíduos conservados no alinhamento estão nas posições: 1, 2, 3, 4 e 15. Assim, fica claro que a posição 3 desses peptídeos é composta por resíduos hidrofóbicos e não hidrofílicos. Interessante também notar

<sup>127</sup> Sequência da schis-1: GILDIKKNKVS NFLFKKIKGEK-NH<sub>2</sub>

<sup>128</sup> Sequência da schis-2: GILDIKKNKVS NFLFKKIK-NH<sub>2</sub>

<sup>129</sup> Sequência da schis-3: GILDIKKNKVS NFLF-NH<sub>2</sub>

que a maioria desses peptídeos na posição 6 possuem resíduos hidrofóbicos e a lisina nos peptídeos schis-1, schis-2 e schis-3 destoa. Assim, com o intuito de aumentar a anfipaticidade do peptídeo schistocina-3, criou-se um análogo com a substituição Lys6Leu, denominado schis-3.1<sup>130</sup> (Figura 4d).

	cov	pid	
1 Schistocin-1	100.0%	100.0%	G I L D I K N K V S N L F K K I K G E K - - - - -
2 Schistocin-2	85.0%	100.0%	G I L D I K N K V S N L F K K I K - - - - -
3 Schistocin-3	65.0%	100.0%	G I L D I K N K V S N L F - - - - -
4 Schistocin-4	65.0%	92.3%	G I L D I L N K V S N L F - - - - -
5 Aurein_2.1	80.0%	43.8%	G L L D I V K K V V G A F G S L - - - - -
6 Uperin-2.4	95.0%	31.6%	G I L D F A K T V V G G I R N A L G I - - - - -
7 Uperin-2.7	95.0%	31.6%	G I L D I A K K L V G G I R N V L G I - - - - -
8 Uperin-2.8	95.0%	26.3%	G I L D V A K T L V G K L R N V L G I - - - - -
9 Brevinin-2CG1	75.0%	10.3%	G I L D K L K E F G I S A A R - - - - G V A Q S L L N T T A S C K L A - - K T C -
10 Brevinin-20A5	75.0%	10.5%	G L L D G I L N A N F N A A K - - - - S A G T S V L N A - L S C K L S - - K T C -
11 Brevinin-2HSa	75.0%	10.5%	G L L D S L K N L A I N A A K - - - - G A G Q S V L N T - L S C K L S - - K T C -
12 Brevinin-2HSb	75.0%	10.5%	G L L D T L K N M A I N A A K - - - - G A G Q S V L N T - L S C K L S - - K T C -
13 Brevinin-2ISA	75.0%	7.9%	S L L D T F K N L A V N A A K - - - - S A G V S V L N A - L S C K I S - - R T C -
14 Brevinin-20A3	75.0%	7.9%	G L L D T F K N M A I N A A H - - - - G A G V S V L N A - L S C K L K - - K T C -
15 Brevinin-20R2	75.0%	10.5%	G L L D T I K N M A L N A A K - - - - S A G V S V L N S - L S C K L S - - K T C -
16 Brevinin-20R6	75.0%	10.5%	G L L D T I K N M A L N A A K - - - - S A G V S V L N S - L S C K D S - - K T C -
17 Brevinin-20R3	75.0%	7.9%	G L I D T I K N M A L N A A K - - - - S A G V S V L N T - L S C K L S - - K T C -
18 Brevinin-20R7	75.0%	10.5%	G L L D T I K N M A L N A A K - - - - S A G V S V L N T - L S C K L S - - K T C -
19 Brevinin-20A1	75.0%	13.2%	G I L D T F K N M A L N A A K - - - - S A G V S V L N A - L S C K L S - - K T C -
20 Brevinin-20A2	75.0%	10.5%	G L L D T F K N M A L N A A K - - - - S A G V S V L N A - L S C K L S - - K T C -
21 Brevinin-20A4	75.0%	7.9%	G L L D T F K N L A I N A A E - - - - S A G V S V L N S - L S C K L S - - K T C -
22 Odoranain-D1	75.0%	10.5%	G F L D T F K N L A L N A A K - - - - S A G V S V L N S - L S C K L F - - K T C -
23 Brevinin-2GRa	75.0%	10.5%	G L L D T F K N L A L N A A K - - - - S A G V S V L N S - L S C K L S - - K T C -
24 Ranateurin-2La	75.0%	16.2%	G I L D S F K G V A K G V A K - - - - D L A G K L L D K - L K C K I T G - - C -
25 Ranateurin-2ARb	55.0%	13.5%	G I L D T I K N A A - - - - K - - - - T V A V G L L E K - I K C K M T G - - C -
26 Ranateurin-2SEc	55.0%	10.8%	G I M D T I K D T A - - - - K - - - - T V A V G L L N K - L K C K I T G - - C -
27 Ranateurin-2CPa	55.0%	10.8%	G I M D T I K D T A - - - - K - - - - T V A V G L L D K - I K C K I T G - - C -
28 Ranateurin-2CPb	55.0%	10.8%	G I M D T I K N T A - - - - K - - - - T V A V G L L D K - I K C K I T G - - C -
29 Brevinin-20A7	75.0%	7.9%	G V L G T V K D L L I G A G K - - - - S A A Q S T L K T - L S C K I S - - N D C -
30 Brevinin-20A6	75.0%	5.3%	G V L A T V K N L L I G T G D - - - - G A A Q S V L K T - L S C K L S - - N D C -
31 Brevinin-20A8	75.0%	7.9%	G V L A T V K N L I N G T G D - - - - G A A Q S V L K T - L S C K L S - - N D C -
32 Hadrurin	100.0%	12.2%	G I L D T I K S I A S K V W N S K T V Q D L K R K G I N W - V A N K L G V S P Q A A
33 Esculentin-2EM	100.0%	13.5%	G I L D T L K Q F A K G V G K D L V K G - A A Q G V L S T - V S C K L A - - K T C -
34 Rugosin-C	100.0%	13.5%	G I L D S F K Q F A K G V G K D L I K G - A A Q G V L S T - M S C K L A - - K T C -
35 Hylaseptin-P1	70.0%	28.6%	G I L D A I K A I A K A A G -
36 Brevinin-2ED	70.0%	11.4%	G I L D S L K N L A K N A G -
37 Brevinin-2EC	80.0%	8.1%	I L L D K L K N F A K T A G K - - - - G - V L Q S L L N T - A S C K L S - - G Q C -
38 Brevinin-2EA	80.0%	13.5%	G I L D T L K N L A I S A A K - - - - G - A A Q G L V N K - A S C K L S - - G Q C -
39 Brevinin-2EB	80.0%	13.5%	G I L D T L K N L A K T A G K - - - - G - A L Q G L V K M - A S C K L S - - G Q C -
40 Andersonin-D1	75.0%	33.3%	F I F P K K N I I N S L F G R -
41 Esculentin-1c	75.0%	0.0%	L L I S G L K N I G K E V G M -
42 Esculentin-1B	75.0%	2.6%	L L I S G L K N V G K E V G M -
43 Esculentin-1A	75.0%	2.6%	L L I S G L K N V G K E V G M -
44 CAP18(106-142)	100.0%	13.6%	K L K K I G Q K I Q G F V P K L A P R T D Y -
45 RL-37	100.0%	9.1%	G L K K V G Q K I K D F L G N L V P R T A S -

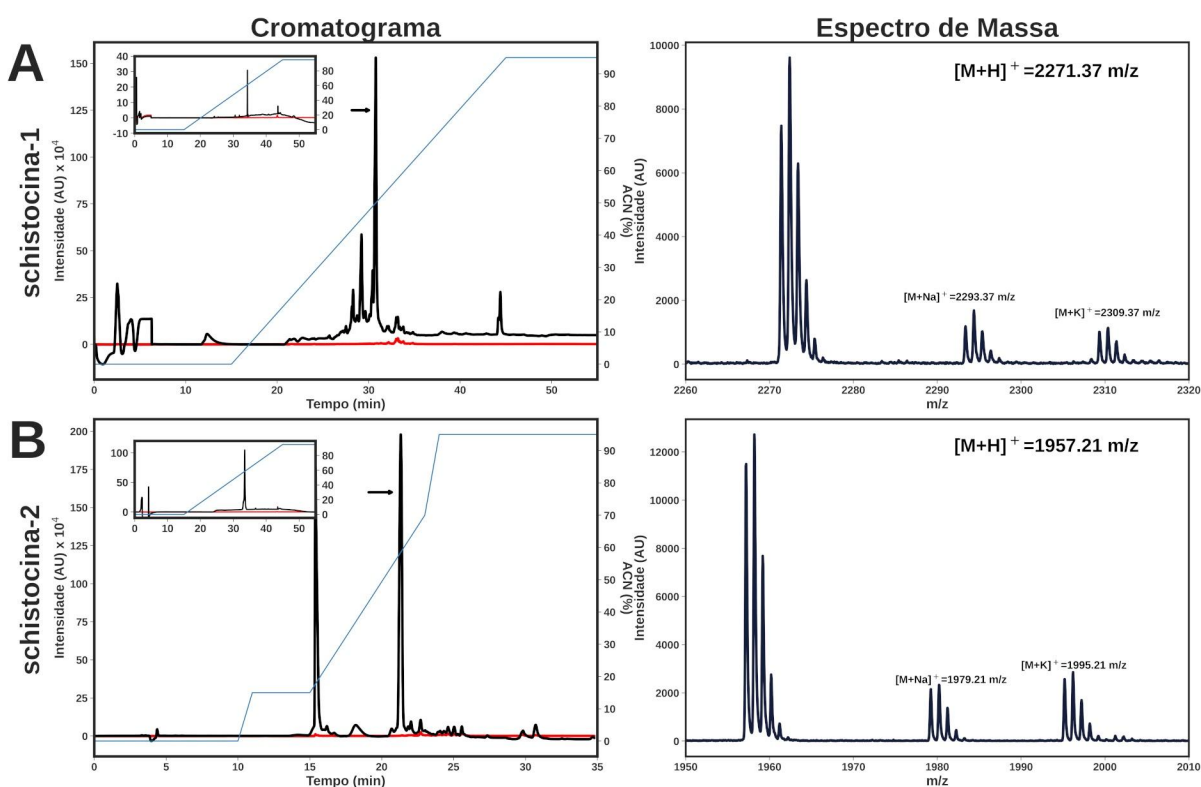
**Figura 5 - Alinhamento de seqüências das schistocinas com seqüências da base de dados APD.** Siglas - cov: cobertura; pid: percentual de identidade. Seqüências com fundo vermelho: 90% de identidade entre as seqüências analisadas. Seqüências com as letras vermelhas: 70% de similaridade entre as seqüências analisadas. Quadro azul: região de resíduos similares consecutivos.

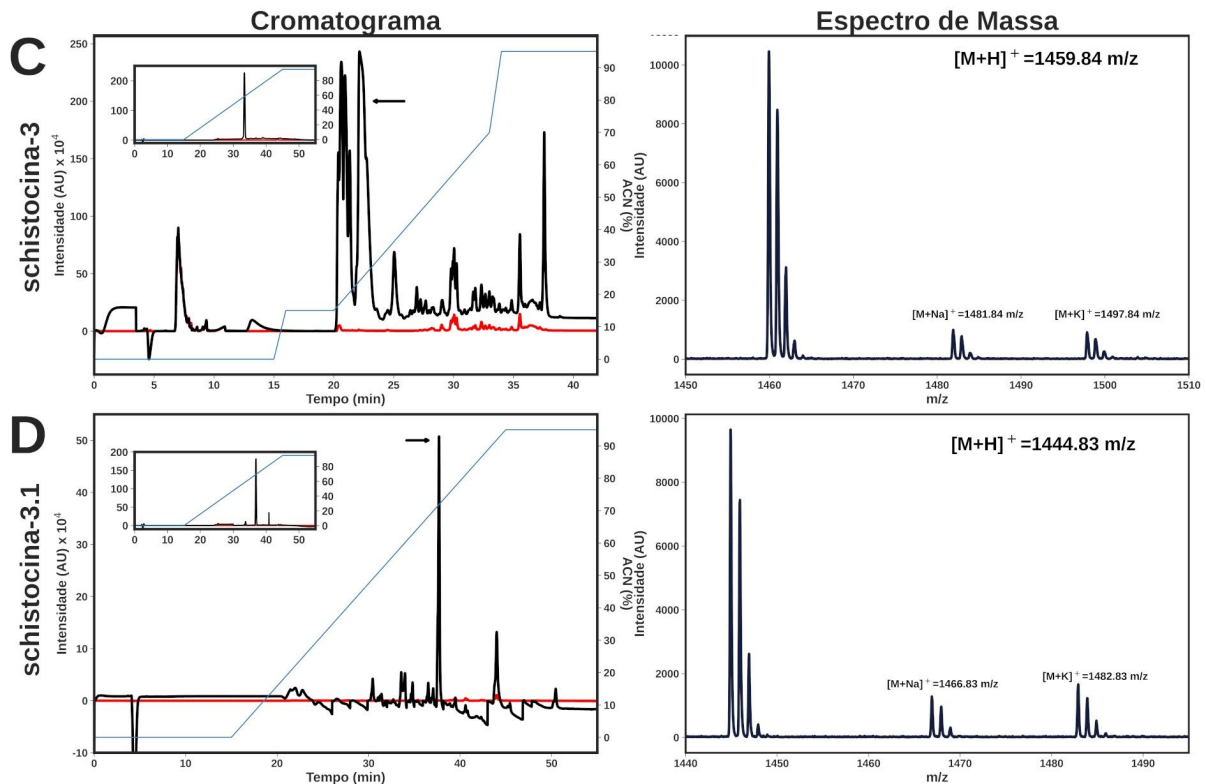
<sup>130</sup> Seqüência da schis-3.1: GILDILNKVSNLF-NH<sub>2</sub>



### 2.3.3 Síntese e análise bioquímica das schistocinas

Com a finalidade de estudar a atividade antimicrobiana e as propriedades biofísicas desses peptídeos, as schistocinas foram quimicamente sintetizadas em fase sólida, purificadas por cromatografia líquida de fase reversa e as respectivas massas aferidas por espectrometria de massa. A figura 6 constitui um painel com as cromatografias e os espectros de massa dos quatro criptídeos. Analisando os cromatogramas, é possível observar que a schis-1 e schis-2 tem hidrofobicidade média (eluição com ~ 50% de acetonitrila) (Figura 6a e 6b), a schistocina-3 o mais hidrofílico (eluição com ~ 30% de acetonitrila) (Figura 6d) e a schistocina-3.1 é o peptídeo mais hidrofóbico (eluição com ~ 70% de acetonitrila) (Figura 6d). A massa aferida dos peptídeos é: a) schis-1  $[M+H]^+ = 2271,37$  m/z; b) schis-2  $[M+H]^+ = 1957,21$  m/z; c) schis-3  $[M+H]^+ = 1459,84$  m/z e d) schis-3.1  $[M+H]^+ = 1444,83$  m/z.





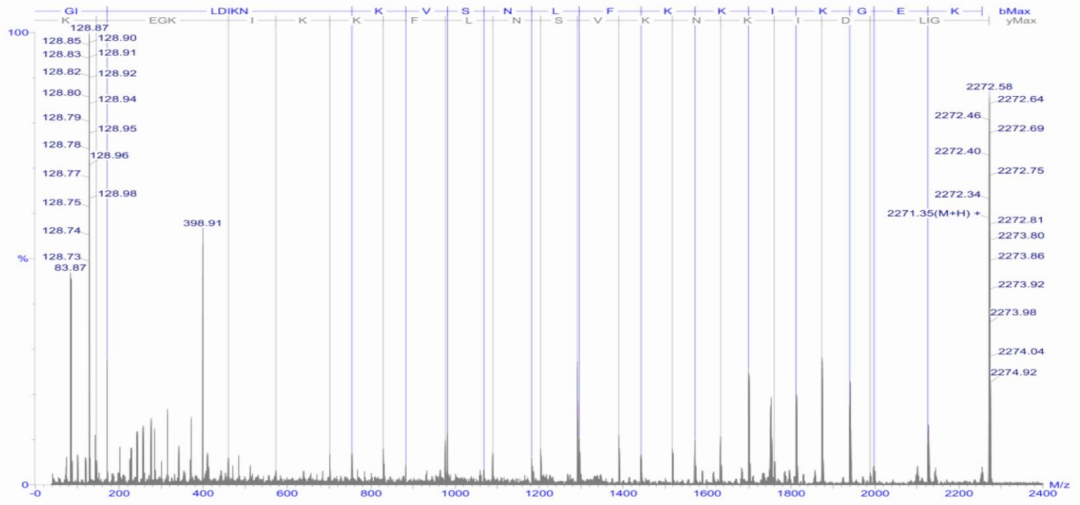
**Figura 6 - Separação cromatográfica e identificação espectrométrica das schistocinas.** As schistocinas foram sintetizadas em fase sólida, purificadas por cromatografia líquida (HPLC), como mostrado no painel à esquerda, e as massas foram aferidas por espectrometria de massa (MALDI-TOF), como mostrado no painel à direita. A linha de acetonitrila está representada em azul e os cromatogramas internos foram realizados após purificação.

Para verificar a constituição dos resíduos de aminoácidos dos peptídeos, uma proteólise via incidência de laser no espectrômetro de massa, técnica denominada MALDI-TOF MS/MS, foi realizada e os espectros resultantes estão resumidos na Figura 7. As fragmentações encontradas:

<p>schis-1 GI/L I/LDI/LKN K V S N I/L F K K I/L K G E K K EGK I/L K K F I/L K K F I/L N S V K N K I/L D I/LI/LG), GI/L I/L D I/L KN K V S N I/L F K K I/L K K I/L K K F I/L N S V K N K I/L D I/LI/LG), GI/L I/L D I/L KN K V S N I/L F e série y = F I/L N S V K N K I/L D I/LI/LG), schis-3.1 (série b = GI/L I/L D I/L I/L N K V S N I/L F e série y = FI/LNSVK N I/L I/L D I/LI/LG).</p>	<p>schis-2 (série b = e série y = e série y = e série b =</p>
--	---

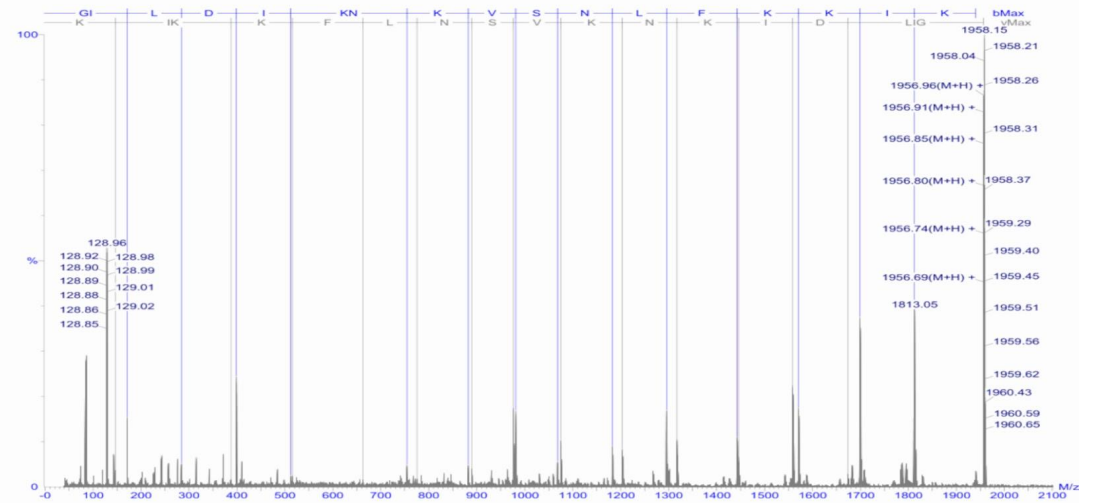
**A**

schistocina-1



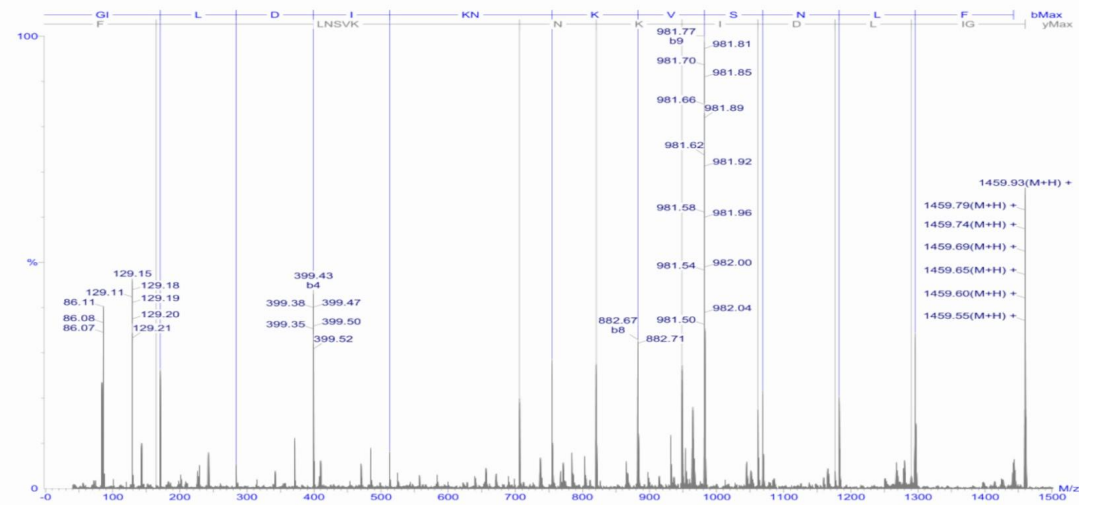
**B**

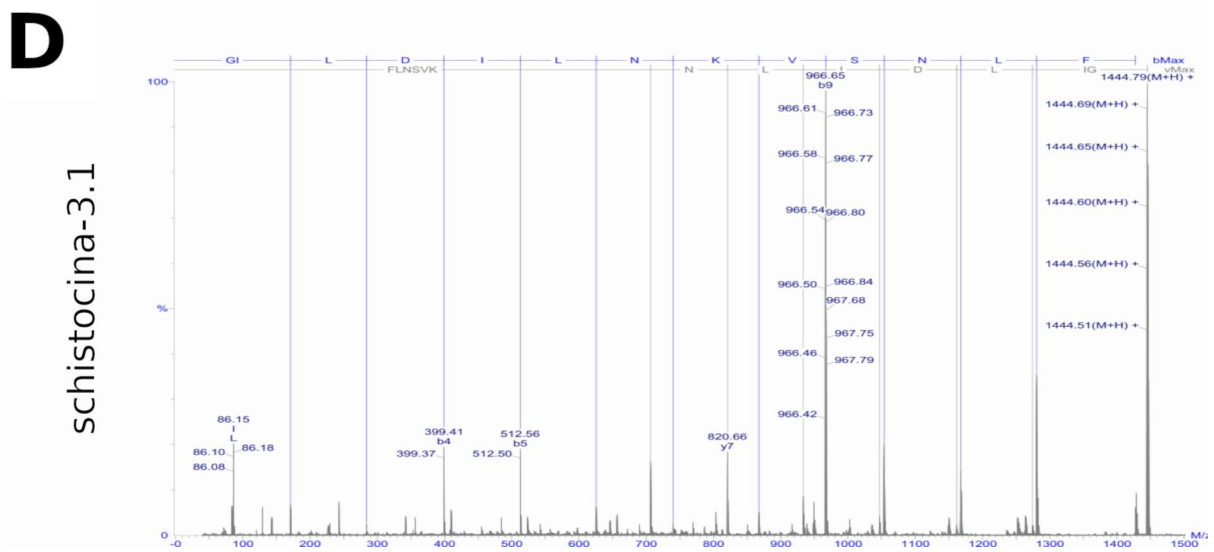
schistocina-2



**C**

schistocina-3





**Figura 7 - Análise de MS/MS dos fragmentos das schistocinas.** As massas moleculares observadas dos peptídeos foram: schistocina-1 = 2271,35 Da; schistocina-2 = 1956,96 Da; schistocina-3 = 1459,93 Da; schistocina-3.1 = 1444,79 Da. Os peptídeos foram fragmentados pelo MALDI-TOF MS/MS depois de análise prévia das massas peptídicas por meio de MALDI-TOF MS. Os dados resultantes foram analisados por meio de Pepseq (Walters Co.) e Bruker Daltonics flexAnalysis 3.3.

### 2.3.4 Atividade antimicrobiana das schistocinas

Com o intuito de avaliar a capacidade antimicrobiana das schistocinas, realizou-se ensaios antimicrobianos contra cepas bacterianas (*S. aureus* ATCC 29313, *E. coli* ATCC 25922 e *P. aeruginosa* ATCC 27853) e fúngicas (*C. albicans* ATCC 10231, *C. krusei* ATCC 6258, *C. tropicalis* ATCC 22019, *C. parapsilosis* ATCC 90018 e *C. glabrata* ATCC 2001). Os resultados das atividades antimicrobianas das schistocinas são apresentados na Tabela 3. Os peptídeos schis-3 e schis-3.1 exibiram atividades semelhantes quando comparados ao grupo controle (DS01, ampicilina e cloranfenicol), enquanto valores mais elevados de MIC foram detectados nos criptídeos schis-1 (>130) e schis-2 (>130) contra as cepas bacterianas testadas. Um comportamento semelhante foi observado nos ensaios antifúngicos, exceto para *C. krusei*, onde todos os peptídeos foram ineficazes como antifúngicos. No geral, schis-3 e schis-3.1 foram mais efetivos contra bactérias e fungos do que schis-1 e schis-2. Mesmo assim, todos foram menos ativos do que DS01 (26,5  $\mu$ M contra *S. aureus*, 6,6  $\mu$ M contra *E. coli*, <1,42  $\mu$ M contra *C. albicans*, 32,3  $\mu$ M contra *C. krusei*).

**Tabela 3 - Atividade antibacteriana e antifúngica de schistocinas.** A atividade de peptídeos foi testada contra três cepas bacterianas e 5 cepas fúngicas. Nistatina, DS01, Ampicilina e Cloranfenicol foram usados como drogas de controle.

Microrganismo	MIC( $\mu\text{M}$ ) <sup>a</sup>				Nistatina	°DS01	Amp.	Cloran.
	schistocina-1	schistocina-2	schistocina-3	schistocina-3.1			<sup>d</sup>	<sup>d</sup>
<i>S. aureus</i> ATCC 29313	>130	>130	31.2	8	ND <sup>b</sup>	26.5	<11	ND
<i>E. coli</i> ATCC 25922	>130	>130	62.5	62.5	ND	6.6	46	25
<i>P. aeruginosa</i> ATCC 27853	>130	>130	62.5	31.2	ND	NT <sup>c</sup>	20	30
<i>C. albicans</i> ATCC 10231	250	250	62.5	31.2	125	<1.42	NT	NT
<i>C. krusei</i> ATCC 6258	125	62.5	62.5	62.5	125	32.3	NT	NT
<i>C. tropicalis</i> ATCC 22019	500	500	31.2	31.2	125	42	NT	NT
<i>C. parapsilosis</i> ATCC 90018	>500	>500	125	125	62.5	32.3	NT	NT
<i>C. glabrata</i> ATCC 2001	>500	>500	250	125	7.8	26.5	NT	NT

<sup>a</sup>MIC - Concentração mínima de peptídeos requerida para inibir o crescimento de microrganismos

<sup>d</sup>Amp and Cloran - ampicilina e cloranfenicol

<sup>b</sup>ND - Atividade antimicrobiana não detectada

<sup>c</sup>DS01 - Dermaseptina-1

<sup>e</sup>NT - Peptídeo não testado

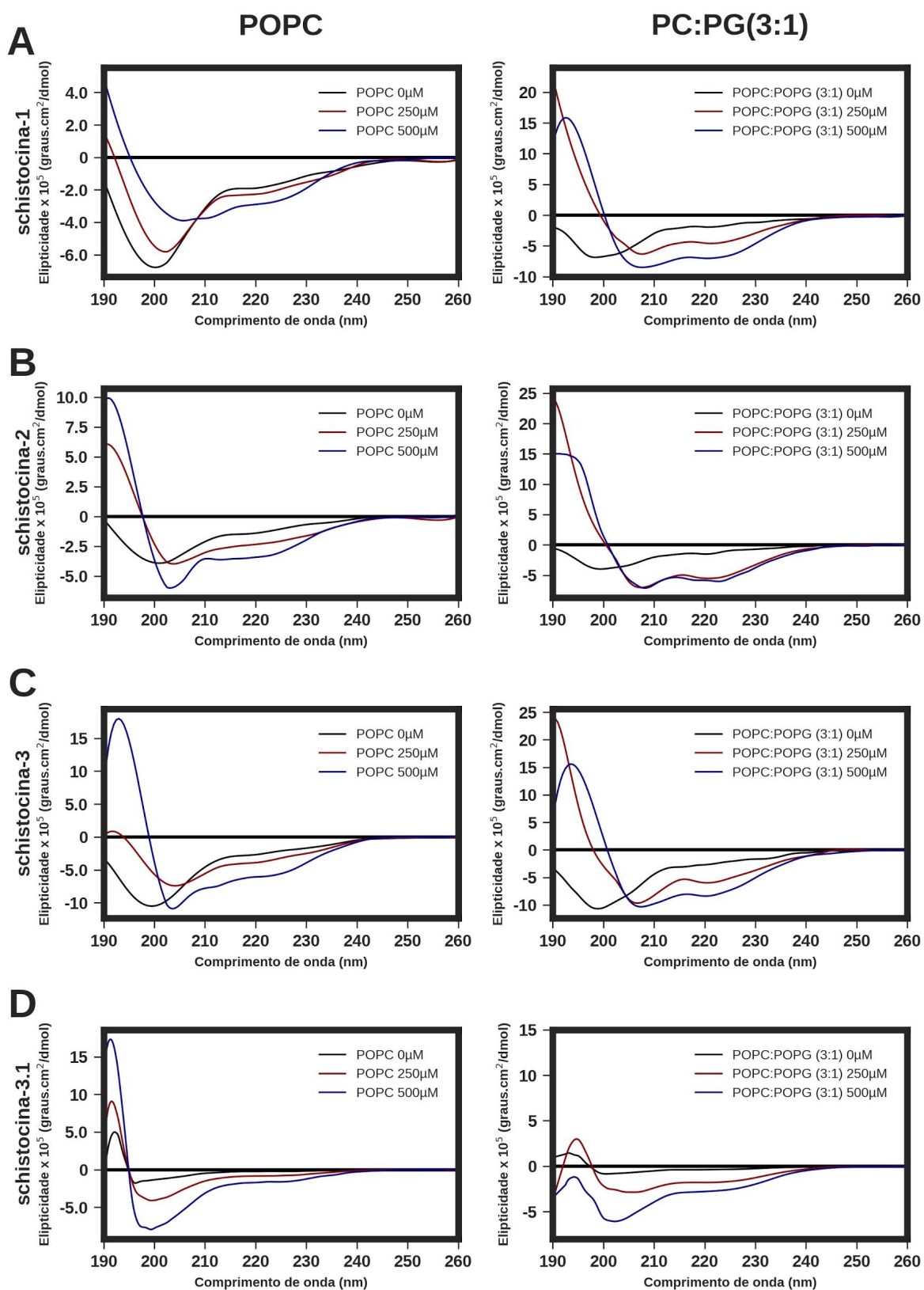
### 2.3.5 Análises de padrão de estrutura secundária das schistocinas em ambientes miméticos de membrana

A fim de avaliar se o algoritmo *encrypted* foi capaz de prever peptídeos membrano ativos capazes de enovelar-se em  $\alpha$ -hélices, foram realizados experimentos de dicroísmo circular em solventes miméticos de membrana e em vesículas de fosfolípidos (Figura 8 e Figura 9). É possível observar que, na presença do tampão, as schistocinas adotaram uma conformação randômica; entretanto, quando na presença de vesículas zwitteriônicas e

aniônicas, elas adquiriram conformação em  $\alpha$ -hélice (Figura 8 e Tabela 4). Além disso, na presença de vesículas aniônicas POPC:POPG (3:1), o conteúdo da hélice de schis-1, schis-2 e schis-3 é alto (84-27%), mesmo em concentração média, enquanto schis-3.1 exibiu níveis mais baixos (250 mM: 50,16%, 500 mM: 75,57%) (Figura 8).

**Tabela 4 - Porcentagem em hélice de estruturas de schistocinas a partir de espectro de dicroísmo circular.** Dados foram ajustados utilizando o programa online K2D2.

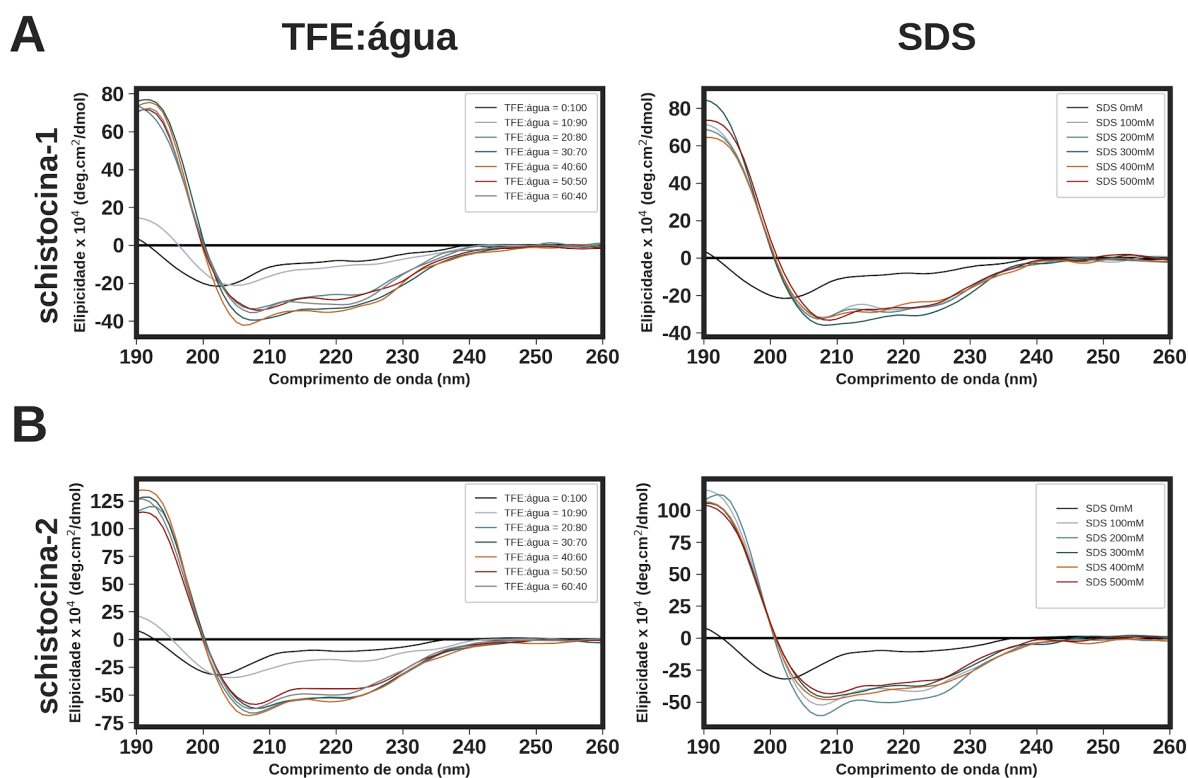
Conteúdo em Hélice (%)							
SDS (mM)							
	0	100	200	300	400	500	
schistocina-1	11.33	69.46	69.46	75.57	69.46	69.46	
schistocina-2	8.6	84.27	84.27	84.27	84.27	84.27	
schistocina-3	6.27	47.81	47.93	47.81	47.93	36.77	
schistocina-3.1	56.82	75.57	80.68	69.46	80.68	69.46	
TFE (%)							
	0	10	20	30	40	50	60
schistocina-1	11.33	20.49	69.46	80.68	80.68	69.46	75.57
schistocina-2	8.6	50.16	84.27	84.27	84.27	84.27	84.27
schistocina-3	6.27	10.33	12.98	47.81	47.93	47.81	47.81
schistocina-3.1	56.82	27.25	84.27	84.27	84.27	84.27	84.27
POPC (mM)							
	0		250		500		
schistocina-1	7.96		8.6		75.57		
schistocina-2	8.6		56.82		75.57		
schistocina-3	6.27		7.84		11.27		
schistocina-3.1	6.71		7.96		7.96		
PCPG(3:1) (mM)							
	0		250		500		
schistocina-1	6.27		84.27		84.27		
schistocina-2	8.6		84.27		84.27		
schistocina-3	7.96		84.27		84.27		
schistocina-3.1	6.71		50.16		75.57		



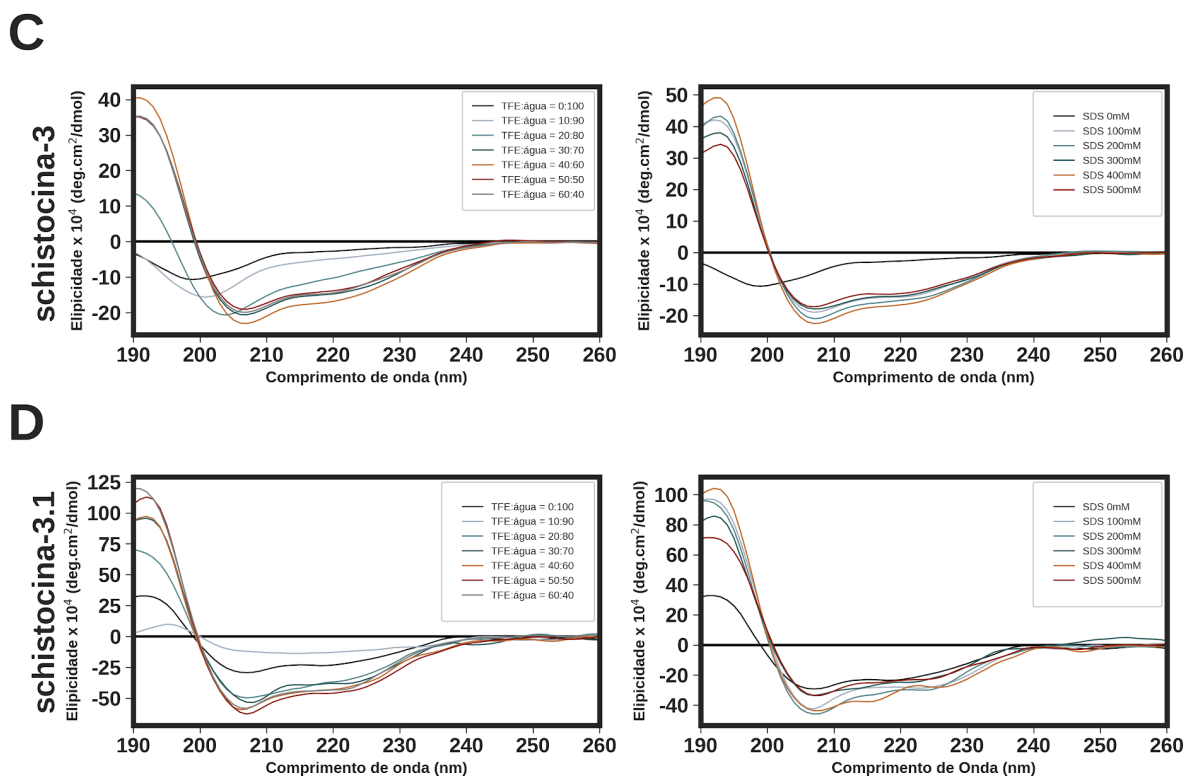
**Figura 8 - Espectro de dicroísmo de schistocinas de 30  $\mu$ M em vesículas POPC e POPC:POPG (3:1) (A-H). Coluna da esquerda: vesículas POPC (250 $\mu$ M, 500 $\mu$ M); coluna da direita: vesículas POPC:POPG(3:1) (250 $\mu$ M, 500 $\mu$ M). A) schistocina-1. B) schistocina-2.**

C) schistocina-3. D) schistocina-3.1. Todos os espectros foram gravados em 50 mM Tris-HCl pH 7.0 e 100 mM NaCl. Espectros foram gravados de 260 nm até 190 nm. As linhas pretas representam peptídeos sem vesículas, a linha cinza representa peptídeos em 250  $\mu$ M de vesículas, e as linhas azuis representam peptídeos em 500  $\mu$ M de vesículas. Bandas negativas no 222 nm e 208 nm são características de peptídeos alfa-hélices, como visto em peptídeos-vesículas POPC:POPG; bandas negativas abaixo de 200 nm e menores são características dos peptídeos randômicos, como visto em peptídeos-vesículas POPC.

Além disso, em misturas de TFE:água e na presença de micelas de SDS, todas as schistocinas têm um padrão de  $\alpha$ -hélice característico (conteúdo entre 47-84%), com dois mínimos, em 208 nm e 222 nm (Figura 8). Os peptídeos schis-3 e schis-3.1 também mostraram um ponto isodicroico bem evidenciado, próximo a 200 nm, indicando uma transição de alça randômica para hélice na interação com ambientes hidrofóbicos. Em conjunto, esses resultados nos permitiram avaliar o estado geral de conformação  $\alpha$ -helicoidal das schistocinas após a adsorção por membrana.







**Figura 9 - Espectros de dicroísmo de schistocinas de 30  $\mu$ M em TFE e SDS.** Coluna da esquerda: espectros foram recordados para os peptídeos em soluções tampão aquosa: 50 mM Tris-HCl, pH 7.0, 100 mM NaCl, TFE:água (0:100) (linha preta), TFE:água (10:90) (linha cinza clara), TFE:água (20:80) (linha cinza média), TFE:água (30:70) (linha cinza escura), TFE:água (40:60) (linha amarela), TFE:água (50:50) (linha vermelha), TFE:água (60:40) (linha cinza); coluna da direita: espectros foram gravados para os peptídeos em soluções tamponadas aquosas: 50mM Tris-HCl, pH 7.0, 100mM NaCl (linha preta) e na presença de micelas de SDS: 100mM (linha cinza clara), 200mM (linha cinza média), 300mM (linha cinza escura), 400mM (linha amarela) e 500mM (linha vermelha).

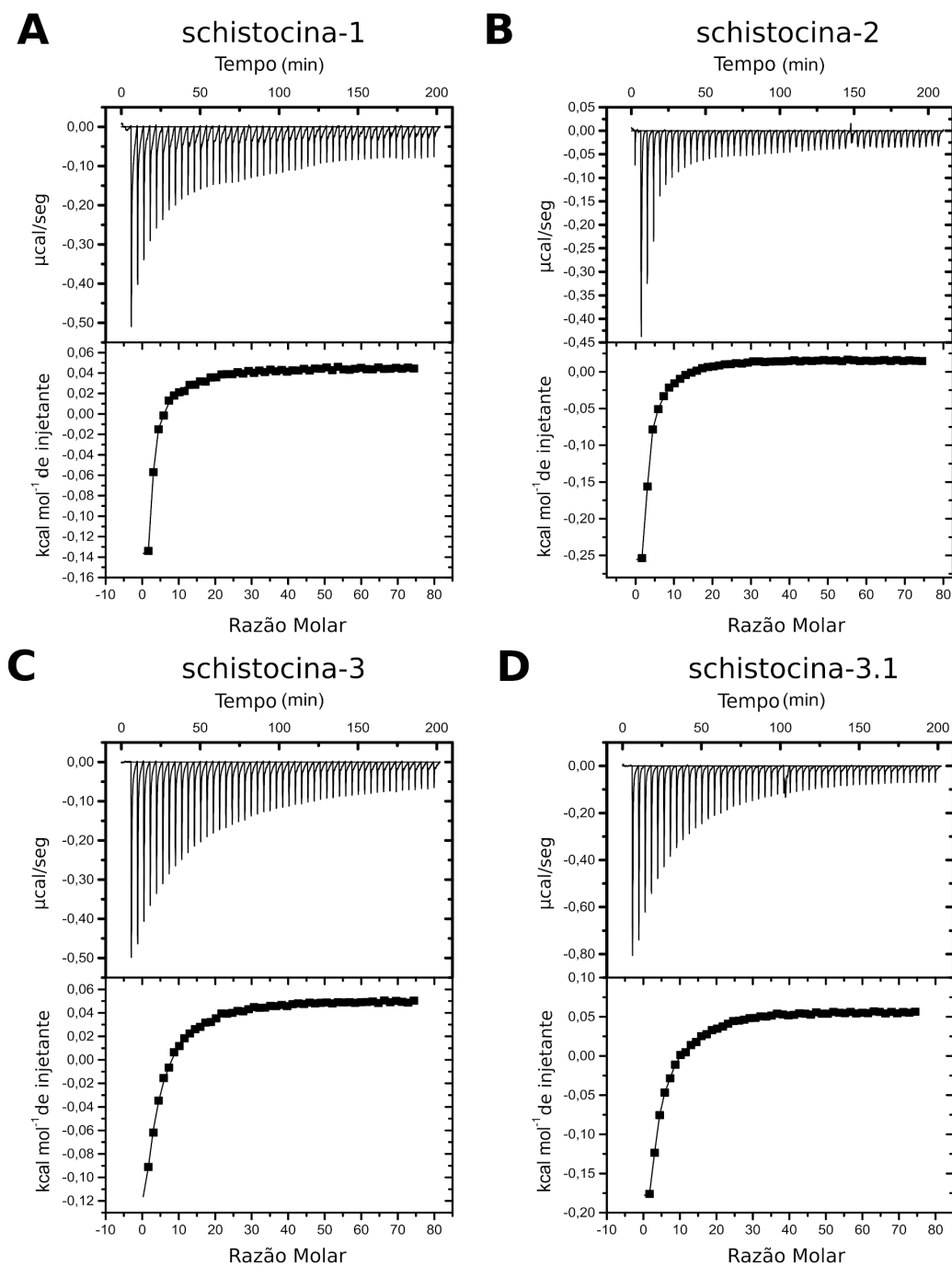
### 2.3.6 Calorimetria de Titulação Isotérmica entre schistocinas e LUVs

Como as análises de dicroísmo circular revelaram maior conteúdo de hélice das schistocinas na presença de bicamadas de membrana aniônica, realizou-se uma medição termodinâmica quantitativa da interação peptídeo-membrana por calorimetria de titulação isotérmica (ITC) e estimou-se a afinidade de ligação. As isotermas de ligação revelaram que as interações gerais de schistocinas com vesículas aniônicas são de natureza exotérmica, conforme sugerido pelos valores de calor integrados (Figura 10). As diminuições no fluxo de calor até valores próximos à diluição são devidos à menor quantidade de peptídeo livre em solução ao final do experimento, o que evidencia a afinidade das schistocinas pelos LUVs de

POPC:POPG. Os parâmetros de ligação de ITC para as schistocinas são fornecidos na Tabela 5. Os processos exergônicos ocorrem para a interação do peptídeo com as vesículas e as maiores alterações de energia livre de Gibbs foram observadas para schis-3.1 (-6,23 kcal.mol<sup>-1</sup>). Consequentemente, uma maior constante de ligação ( $k_{app}$ ) é observada para schis-3-1 se comparada com outras schistocinas. Embora os valores de entalpia negativa indiquem ligação exotérmica dos peptídeos às vesículas aniônicas e a predominância de forças coulombianas atrativas, o processo de ligação para todos os peptídeos é principalmente impulsionado por fatores entrópicos, uma vez que  $|\Delta H| \ll |T\Delta S|$ . Portanto, esses dados sugerem um modo bem conhecido de interação peptídeo-membrana em que a atração eletrostática entre os peptídeos catiônicos e as membranas aniônicas resulta no desordenamento da bicamada.

**Tabela 5 - Parâmetros termodinâmicos do experimento ITC dos peptídeos em LUVs aniônicos.**

<b>Peptídeos</b>	<b>K</b>	<b>DH Kcal/mol</b>	<b>TDS Kcal.mol<sup>-1</sup></b>	<b>DG</b>
schistocina-1	4.06E+03	-1.03	4.07	-5.09
schistocina-2	9.34E+03	-0.40	5.17	-5.57
schistocina-3	2.38E+03	-1.14	3.60	-4.74
schistocina-3.1	2.55E+04	-0.11	6.13	-6.23



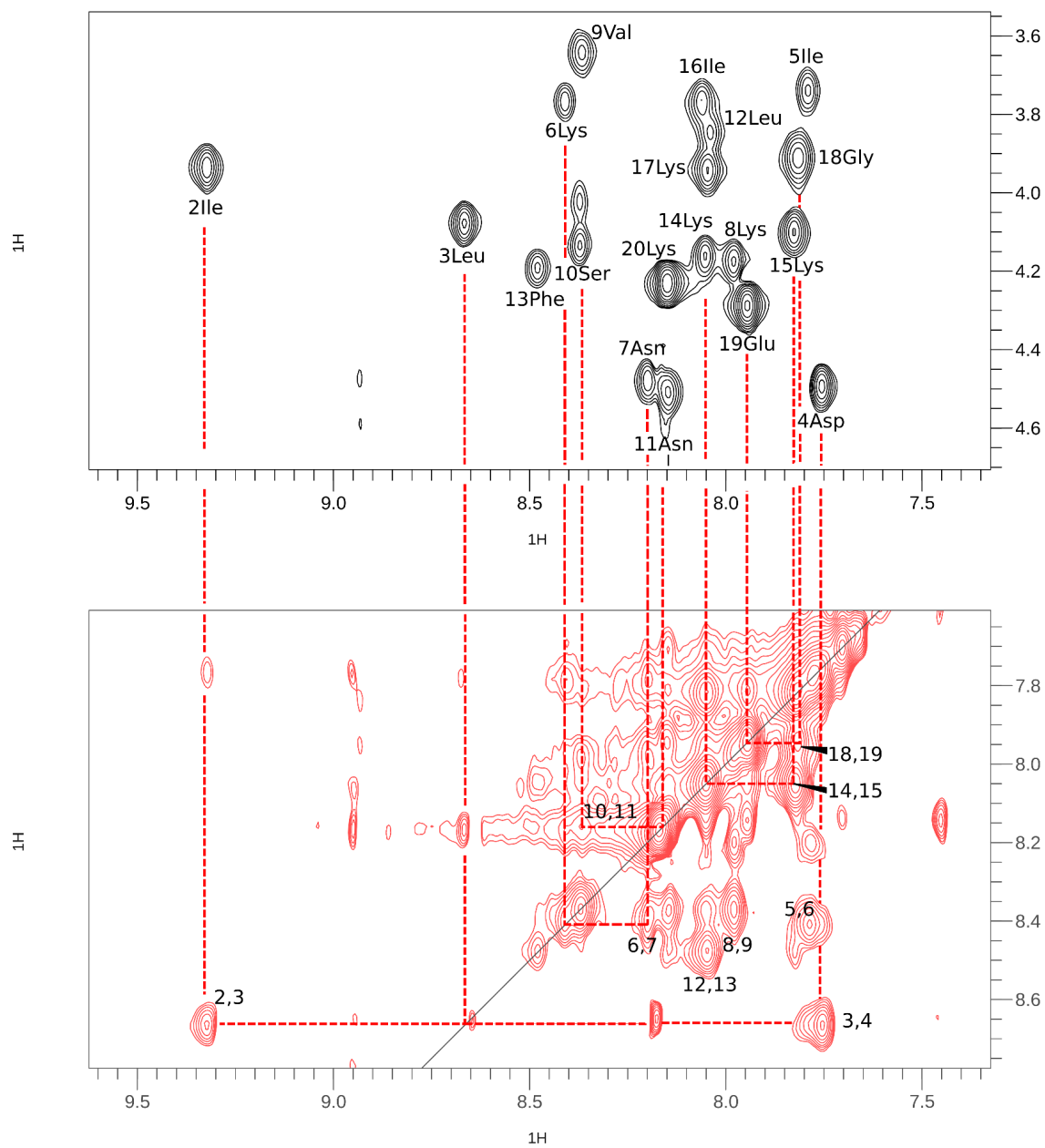
**Figura 10 - Calorimetria de titulação isotérmica das schistocinas em LUVs aniônicas.** Calorimetria de titulação isotérmica das schistocinas em vesículas de POPC:POPG (3:1) vesículas em tampão (35°C; 50 mM Tris-HCl, 100 mM NaCl, pH 8.5). Um total de 5  $\mu\text{L}$  de dispersão de lipídios (concentração = 20 mM) foi injetada em célula de calorimetria contendo a schistocina-1 (A) ou schistocina-2 (B) ou schistocina-3 (C) ou schistocina-3.1 (D) (concentração = 50  $\mu\text{M}$ ) no tampão mesmo. A imagem de cima de cada letra corresponde aos dados brutos. A imagem abaixo corresponde a mudança de calor integrada por injeção com a melhor curva de ajuste. Os sítios de ligação, a constante de dissociação, a entalpia, a energia livre de Gibbs e os resultados de temperatura vezes a entropia em cada letra.

### 2.3.7 Cálculo de Estrutura por RMN

Presumindo que, em meio tamponado, as schistocinas se comportam como estruturas randômicas, mas na presença de meios miméticos de membranas, suas estruturas mudam de conformação estrutural e se enovelam, para uma melhor compreensão estrutural dos peptídeos e uma correlação estrutura-função, partiu-se para o cálculo de estruturas das schistocinas na presença de micelas de dodecilsfosfolina ( $\text{DPC-d}_{38}$ ). Experimentos bidimensionais de RMN foram realizados. A figura 11 constitui um painel com a região de correlação hidrogênio amídico e hidrogênio alfa dos experimentos de TOCSY e a região de correlação hidrogênio amídico e hidrogênio amídico dos experimentos de NOESY para as quatro schistocinas. Com exceção do resíduo Gly-1, que foi identificado apenas nos experimentos com a schis-3, os demais sinais de resíduos na região de hidrogênio alfa foram todos identificados. Devido a mobilidade desses peptídeos na temperatura analisada, 25 °C, alguns sinais extras foram identificados em algumas regiões, mas que não permitiram fazer correlações com resíduos vizinhos. As correlações entre resíduos amídicos (i,i+1) foram fundamentais para fazer o assinalamento sequencial dos peptídeos.

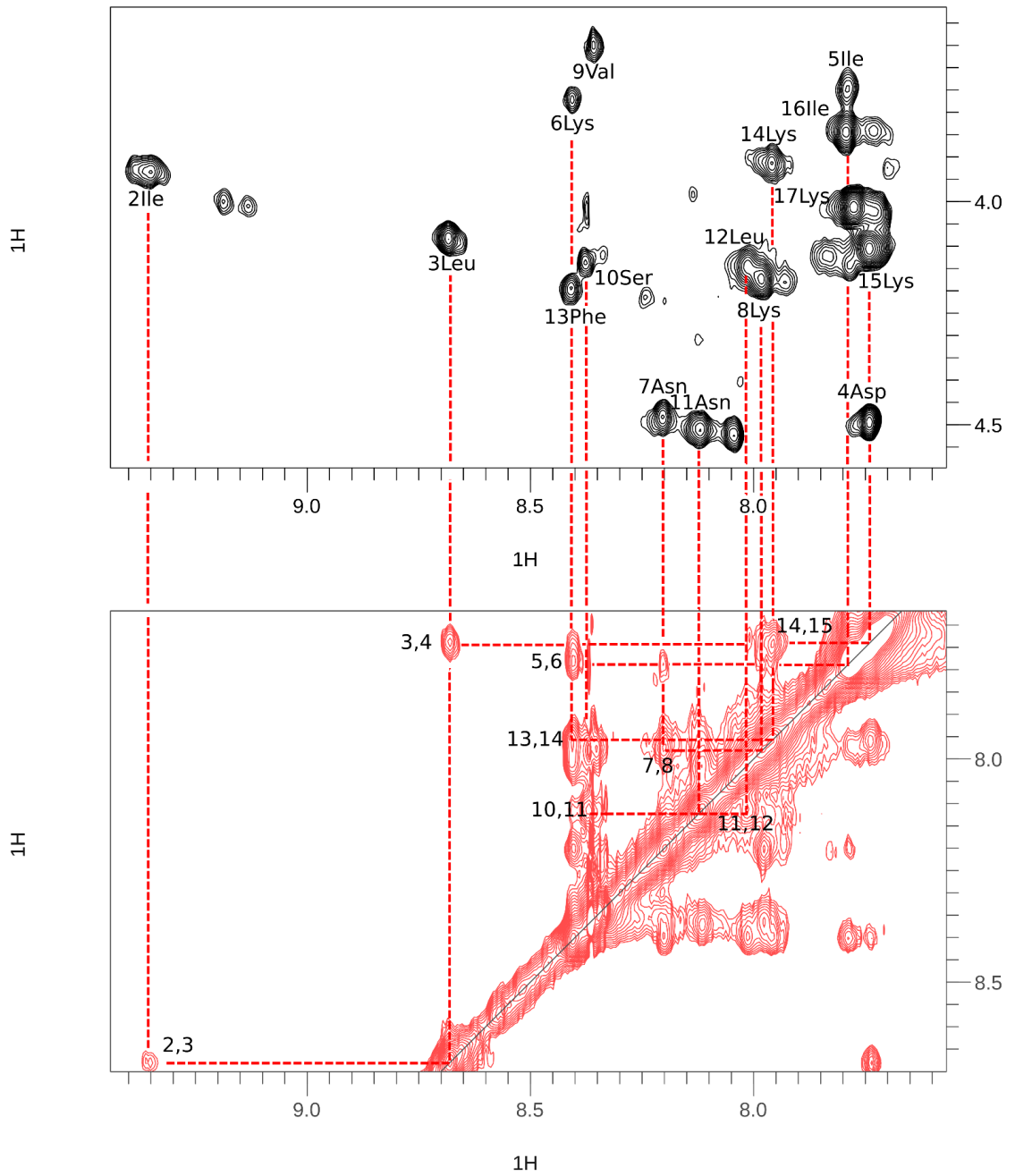
**A**

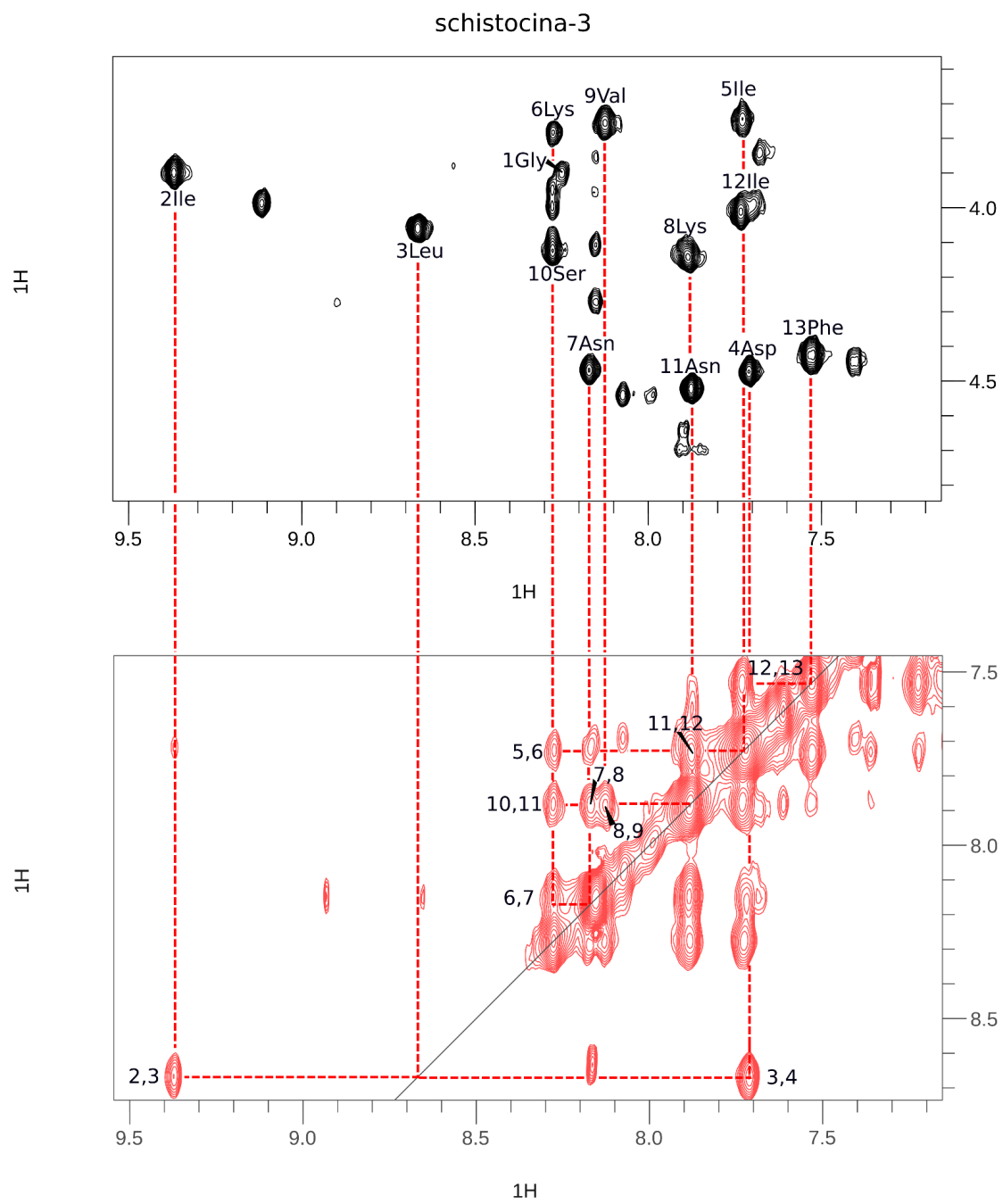
schistocina-1

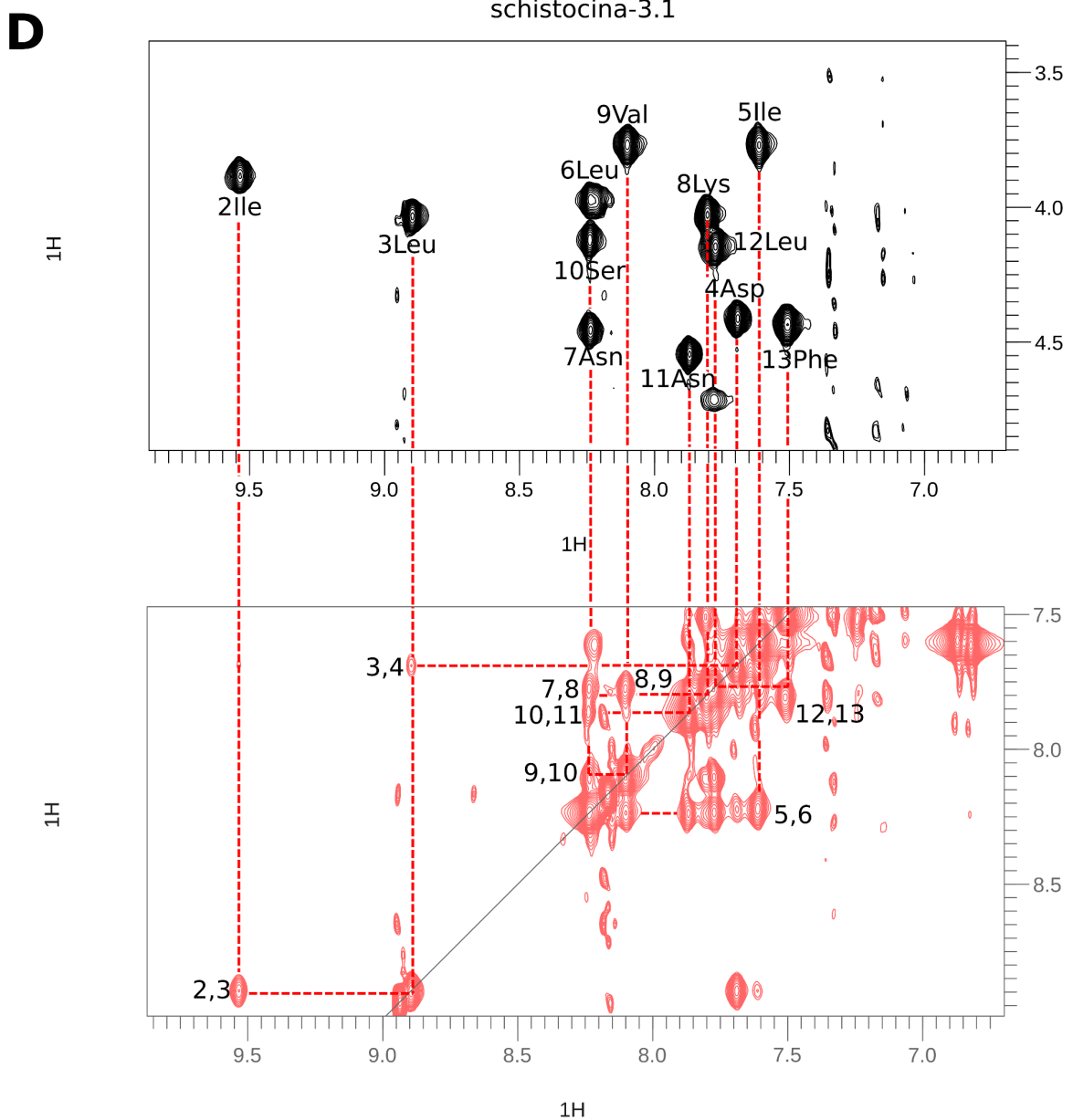


**B**

schistocina-2



**C**

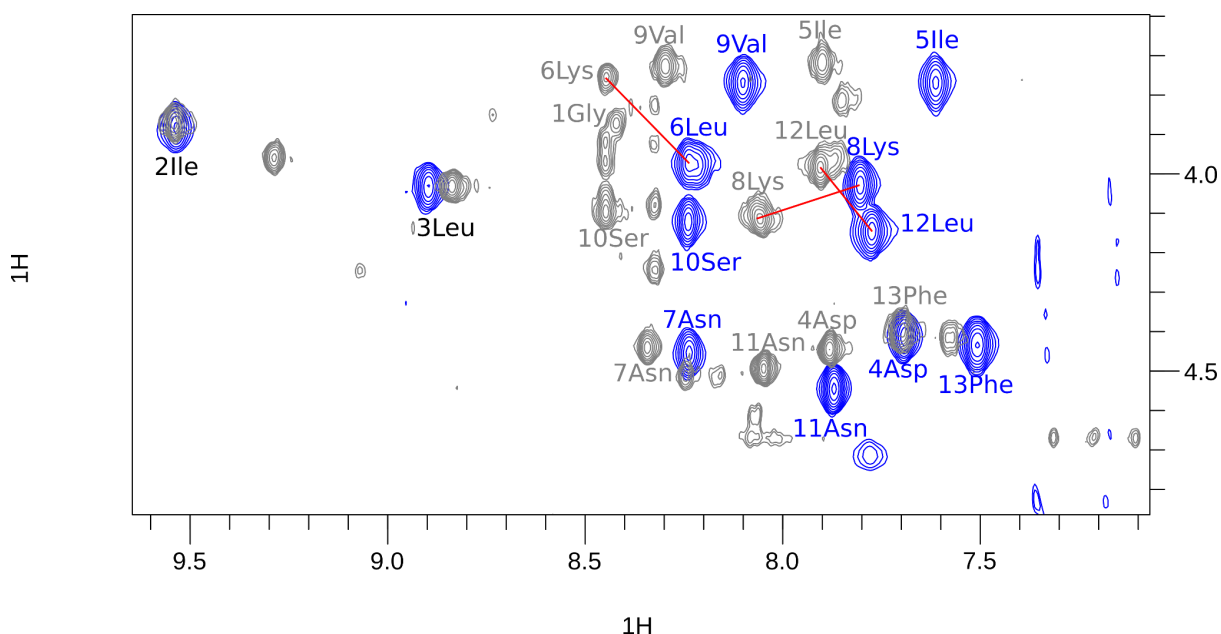


**Figura 11 - Sinais dos experimentos de TOCSY HN-Ha e NOESY HN-HN das schistocinas.** Experimentos bidimensionais de RMN 1H-1H. A região de TOCSY está evidenciando a interação entre átomos HN-Ha dos resíduos dos peptídeos schistocina-1 (A), schistocina-2 (B), schistocina-3 (C) e schistocina-3.1 (D). A região de NOESY é a interconexão  $i,i+1$  dos átomos HN-HN, responsável pelo assinalamento sequencial.

Com exceção dos resíduos a partir da posição 12 até o C-terminal, o deslocamento químico dos sinais de HN-H $\alpha$  foi bem similar, o que, além de facilitar parte do processo de assinalamento, indica que não houve muitas mudanças estruturais na região N-terminal de um peptídeos para o outro (Figura 11). Mesmo as correlações de sinais do resíduo que foi trocado entre os peptídeos schis-3 e schis-3.1, Lys-6-Leu, não teve mudança significativa de deslocamento químico. Como observado na figura 12, os espectros da schis-3 e schis-3.1



tiveram como principais diferenças os deslocamentos químicos dos resíduos 6Lys-6Leu, 8Lys e 12Leu, além da 1Gly, que está presente no espectro da schis-3, mas não no espectro da schis-3.1.



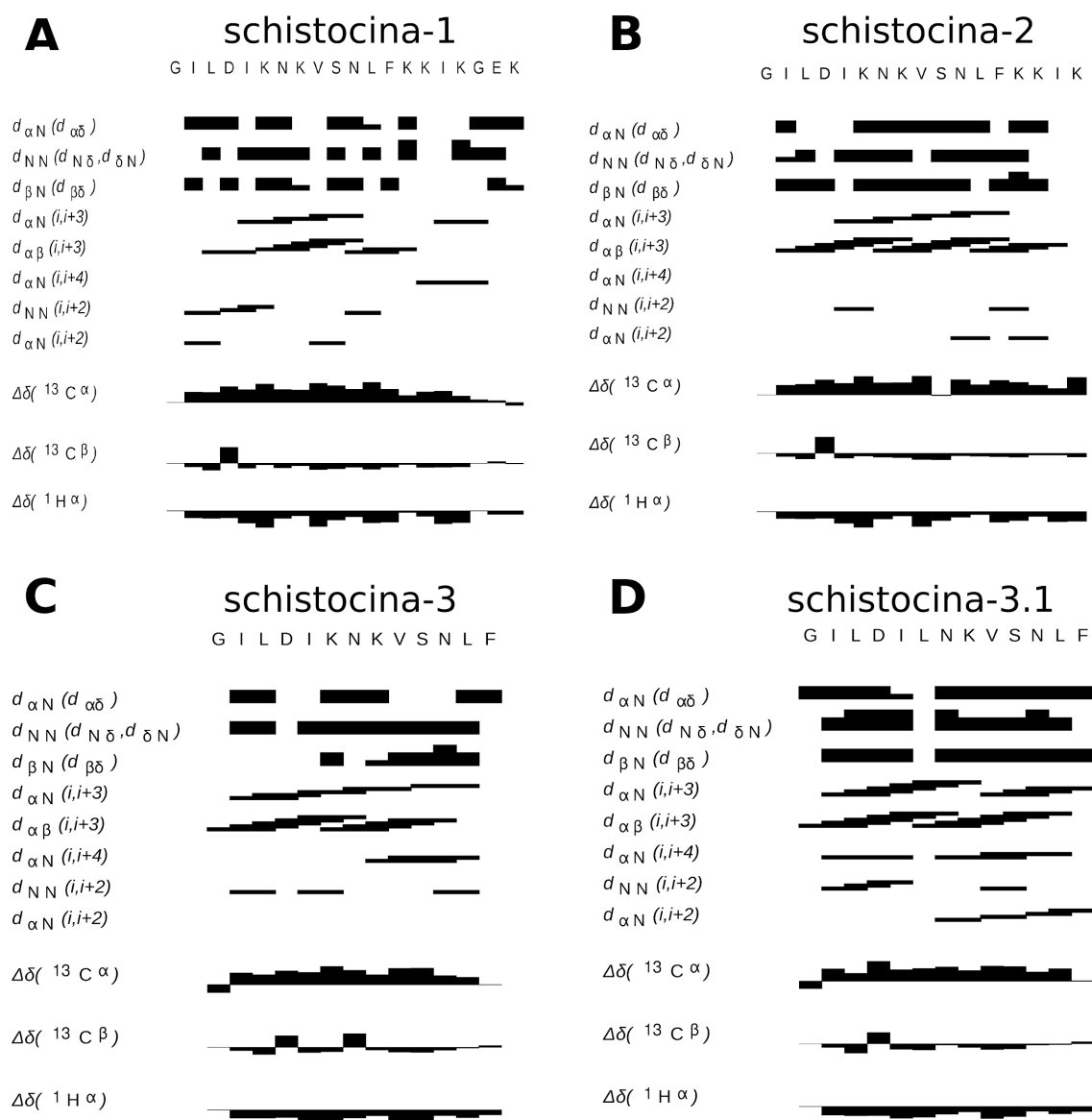
**Figura 12 - Comparação entre hidrogênios alfa dos peptídeos schis-3 e schis-3.1.** Região de correlação hidrogênios amídicos e hidrogênios alfa do peptídeo schis-3 (cinza) e schis-3.1 (azul). Nomenclatura em preto para os sinais com deslocamento químico compartilhado. Os resíduos com deslocamento distintos: Lys6Leu, 8Lys e 12Leu, destacados por traços em vermelho.

A figura 13 resume os mapas de restrições dos NOEs provenientes dos experimentos de NOESY. O peptídeo schis-1, de 21 resíduos, têm sinais intensos  $i, i+1$  ( $\alpha N$ , NN,  $\beta N$ ) sequenciais até o resíduo 14 e depois prossegue a partir do resíduo 16 até o resíduo 20. A ausência de sinais consecutivos nessa região deve-se a sobreposição de sinais no espectro, mas que é compensado pela identificação de sinais  $i, i+3$  e  $i, i+4$  ( $\alpha N$ ). As correlações do tipo  $\alpha N(i, i+3)$  estão presentes entre os resíduos 5Ile-8Lys, 7Asn-10Ser, 9Val-12Leu e 16Ile-19Glu. As correlações do tipo  $\beta N(i, i+3)$  estão presentes entre os resíduos 3Leu-6Lys, 6Lys-9Val, 7Asn-10Ser, 8Lys-11Asn, 9Val-12Leu, 11Asn-14Lys e 12Leu-15Lys. Um único sinal do tipo  $\alpha N(i, i+4)$  foi identificado: 15Lys-19Glu. Na região das correlações entre hidrogênios amídicos, quatro restrições do tipo NN( $i, i+2$ ) foram identificadas: 2Ile-4Asp, 4Asp-6Lys, 5Ile-7Asn e 11Asn-13Phe. Duas correlações do tipo  $\alpha N(i, i+2)$  foram identificadas: 2Ile-4Asp e 9Val-11Asn (Figura 13a).

O peptídeo schis-2, de 17 resíduos, apresenta mais sinais intensos  $i, i+1$  do que a schis-1, entretanto, é possível observar uma ausência de sinais  $\alpha N(i, i+1)$  nas posições 3, 4, 5, 16 e 17. O resíduo Lys17 não tem correlações de sinais assinaladas. As restrições que indicam o enovelamento  $\alpha\beta(i, i+3)$  estão presente na maior parte da sequência peptídica (de 2Ile à 16Ile). As correlações do tipo  $\alpha N(i, i+3)$  estão presentes entre os resíduos: 5Ile-8Lys, 7Asn-10Ser, 9Val-12Leu e 11Asn-14Lys. Além disso, duas correlações  $NN(i, i+2)$  foram identificadas (5Ile-7Asn e 13Phe-15Lys) e duas correlações do tipo  $\alpha N(i, i+2)$  (11Asn-13Phe e 14Lys-16Ile) (Figura 13b).

Com relação à schis-3, de 13 resíduos, os sinais  $NN(i, i+1)$  cobrem a maior parte da sequência peptídica, diferentemente dos sinais  $\alpha N(i, i+1)$ , ausentes nas posições 4, 5, 9, 10 e 11. Os sinais do tipo  $i, i+3$  e  $i, i+4$  estão presentes da posição Gly1 até Leu12. São observados sinais  $\alpha N(i, i+3)$  entre 2Ile-5Ile, 3Leu-6Lys, 5Ile-8Lys, 7Asn-10Ser e 10Ser-13Phe. Sinais do tipo  $\alpha\beta(i, i+3)$  estão presentes em maior quantidade. São eles: 1Gly-4Asp, 2Ile-5Ile, 3Leu-6Lys, 4Asp-7Asn, 5Ile-8Lys, 6Lys-9Val, 7Asn-10Ser, 8Lys-11Asn e 9Val-12Leu. Duas restrições do tipo  $\alpha N(i, i+4)$  foram identificadas na porção C-terminal do criptídeo: 8Lys-12Leu e 9Val-13Phe. Além disso, três correlações  $NN(i, i+2)$  estão presentes, sendo elas 2Ile-4Asp, 5Ile-7Asn e 11Asn-13Phe (Figura 13c).

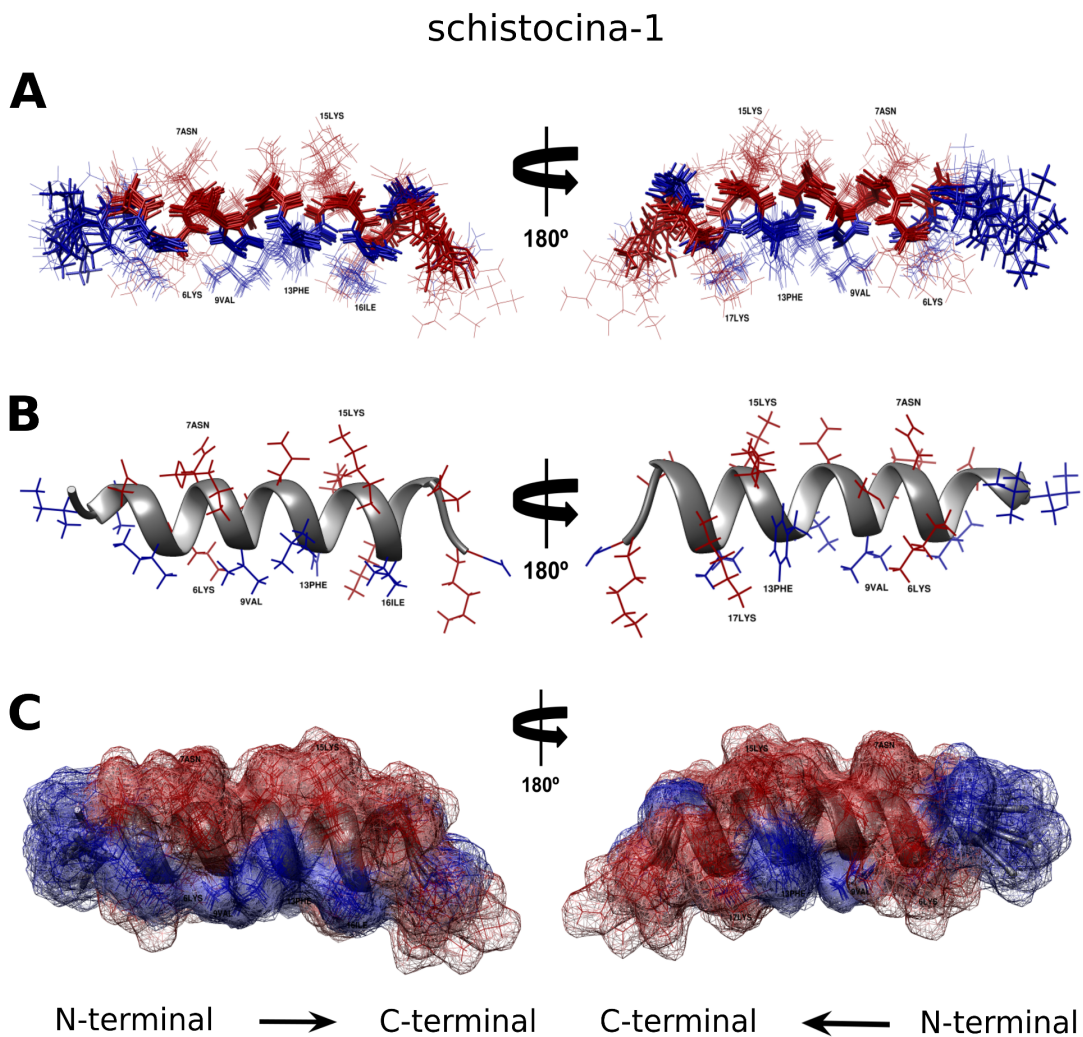
O criptídeo schis-3.1 apresenta restrições fortes em toda a sequência, exceto no resíduo 6Leu, que é justamente o resíduo que difere a schis-3 (6Lys) e a schis-3.1. As correlações do tipo  $\alpha N(i, i+3)$  vão desde o resíduo 2Ile até o resíduo 13Phe. As restrições do tipo  $\alpha\beta(i, i+3)$  vão desde o resíduo 1Gly até o resíduo 12Leu. Três correlações do tipo  $\alpha N(i, i+4)$  estão presentes: 2Ile-6Leu, 7Asn-11Asn e 9Val-13Phe. Na porção N-terminal estão a maior parte das correlações  $NN(i, i+2)$ : 2Ile-4Asp, 3Leu-5Ile, 4Asp-6Leu, além de 9Val-11Leu. Na porção C-terminal existem restrições do tipo  $\alpha N(i, i+2)$ : 7Asn-9Val, 9Val-11Asn, 11Ser-13Phe e 12Asn com a amidação do resíduo 13Phe (peptídeo amidado na porção C-terminal) (Figura 13d).



**Figura 13 - Diagrama de conectividade sequencial das schistocinas em micelas de DPC.** Os peptídeos: schistocina-1 (A); schistocina-2 (B); schistocina-3 (C); schistocina-3.1 (D). O diagrama mostra os contatos (i,i+1), (i,i+2), (i,i+3) e (i,i+4) observados no experimento de NOESY. A espessura das barras indicam a intensidade aproximada dos NOE.

Em ambiente hidrofóbico, como visto no experimento de difração circular, os peptídeos adotam um enovelamento em alfa-hélice. Este padrão de estruturação foi confirmado pelos experimentos de RMN. A figura 14 resume as representações gráficas do peptídeo schistocina-1 a partir dos cálculos de estrutura dos dados experimentais de RMN. A representação em varetas (do inglês, *sticks*) das dez estruturas de menor energia demonstra que, de modo geral, os dois primeiros resíduos e os dois últimos resíduos apresentam baixa

superposição estrutural, característico de porções randômicas (Figura 14a). Os resíduos hidrofóbicos 9Val e 13Phe são aqueles cuja sobreposição das dez estruturas encontram melhor ajuste. Os resíduos hidrofílicos, entretanto, apresentam-se mais dispersos. Apesar disso, o conjunto de estruturas de menor energia mostram valores de RMSD de todos os resíduos relativamente pequenos para ambos os peptídeos: schis-1 ( $0,32\text{\AA} \pm 0,11\text{\AA}$ ). Os resíduos 15Lys/19Glu e 19Glu/20Lys tiveram interações de hidrogênio responsáveis por estabilizar o C-terminal da schis-1.

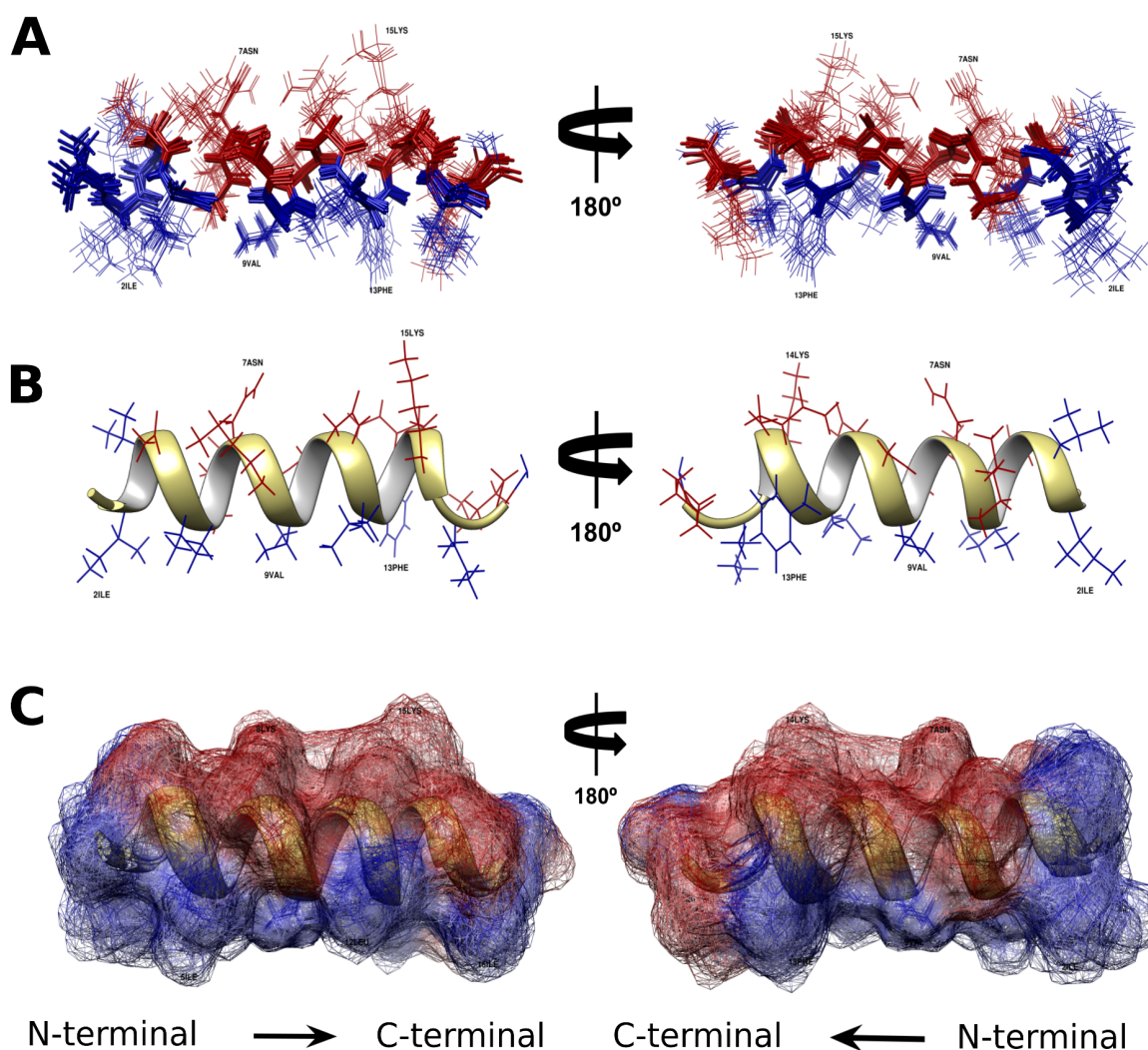


**Figura 14 - Estruturas tridimensionais do criptídeo schistocina-1 obtidas por RMN de solução (400  $\mu\text{M}$  de peptídeos em 20 mM de micelas DPC- $d_{38}$ ). As estruturas tridimensionais (400  $\mu\text{M}$ ) foram obtidas em micelas de DPC- $d_{38}$  (20 mM) no Bruker Avance III 500 (11,75T). Em a) as dez estruturas de menor energia (representação em varetas), evidenciando os resíduos hidrofóbicos, em azul, e os resíduos hidrofílicos, em vermelho; b) as estruturas de menor energia (representação em desenho); c) por último, a estrutura de menor energia (representação de superfície), com as regiões hidrofóbicas, em azul, e em regiões hidrofílicas, em vermelho.**

As dez estruturas de menor energia da schis-2 apresentam uma ótima superposição da cadeia principal (RMSD da schis-2 de  $0,10 \text{ \AA} \pm 0,02 \text{ \AA}$ ) e superposição média das cadeias laterais, sendo que o resíduo 9Val é aquele que melhor se sobrepõe entre as dez estruturas (Figura 15a). A representação em desenho da estrutura de menor energia demonstra que os resíduos 16Ile e 17Lys estão desenovelados, enquanto o restante da cadeia está em conformação de alfa-hélice (Figura 15b). As interações de hidrogênio entre os resíduos 13Phe/16Ile, 13Phe/17Lys e 14Lys/17Lys estabilizaram o C-terminal da schis-2.

Analisando os peptídeos schis-1 e schis-2, ambos têm uma superfície apolar com resíduos hidrofóbicos e alifáticos: Gly1, Ile2, Leu3, Ile5, Val9, Leu12, Phe-13, Ile-16 (Figura 14c e Figura 15c). Além disso, o potencial eletrostático de superfície dos peptídeos gerou uma energia potencial de solvatação de  $1,16 \cdot 10^4 \text{ kJ} \cdot \text{mol}^{-1}$  para schis-1 e  $1,33 \cdot 10^4 \text{ kJ} \cdot \text{mL}^{-1}$  para schis-2, demonstrando que são altamente catiônicos. Schis-1 tem uma conformação parcialmente anfipática, prejudicada por Lys-6 e Gly-18. Além disso, os resíduos hidrofílicos das cadeias laterais apresentam-se menos rígidos do que os resíduos hidrofóbicos.

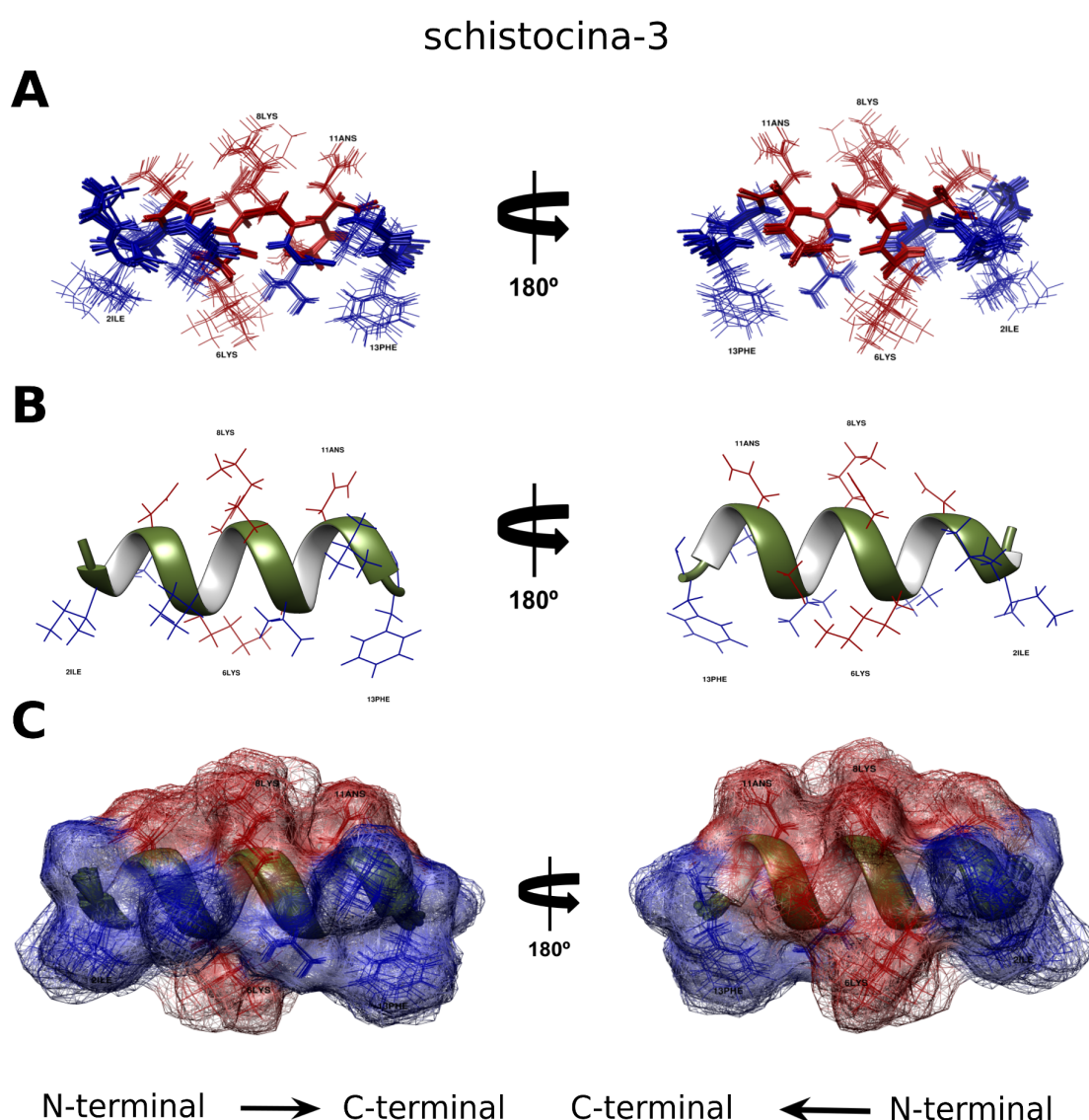
## schistocina-2



**Figura 15 - Estruturas tridimensionais do criptídeo schistocina-2 obtidas por RMN de solução (400  $\mu\text{M}$  de peptídeos em 20 mM de micelas DPC- $d_{38}$ ). As estruturas tridimensionais (400  $\mu\text{M}$ ) foram obtidas em micelas de DPC- $d_{38}$  (20 mM) no Bruker Avance III 500 (11,75T). Em a) as dez estruturas de menor energia (representação em varetas), evidenciando os resíduos hidrofóbicos, em azul, e os resíduos hidrofílicos, em vermelho; b) as estruturas de menor energia (representação em desenho); c) por último, a estrutura de menor energia (representação de superfície), com as regiões hidrofóbicas, em azul, e em regiões hidrofílicas, em vermelho.**



Como esperado, a schis-3 apresenta uma estrutura completamente enovelada em  $\alpha$ -hélice (Figura 16). A representação em varetas das dez estruturas de menor energia mostra uma maior dispersão dos resíduos 2Ile, 6Lys, 8Lys e 13Phe. A cadeia lateral dos outros resíduos apresentam melhor superposição, com RMSD de  $0,16 \text{ \AA} \pm 0,05 \text{ \AA}$  (Figura 16a). A representação em desenho evidencia que a estrutura de menor energia apresenta uma estrutura helicoidal com porção N-terminal mais afunilada (Figura 16b). A representação de superfície evidencia a quebra da anfipaticidade pelo resíduo 6Lys, que está presente na porção hidrofóbica do peptídeo (Figura 16c).



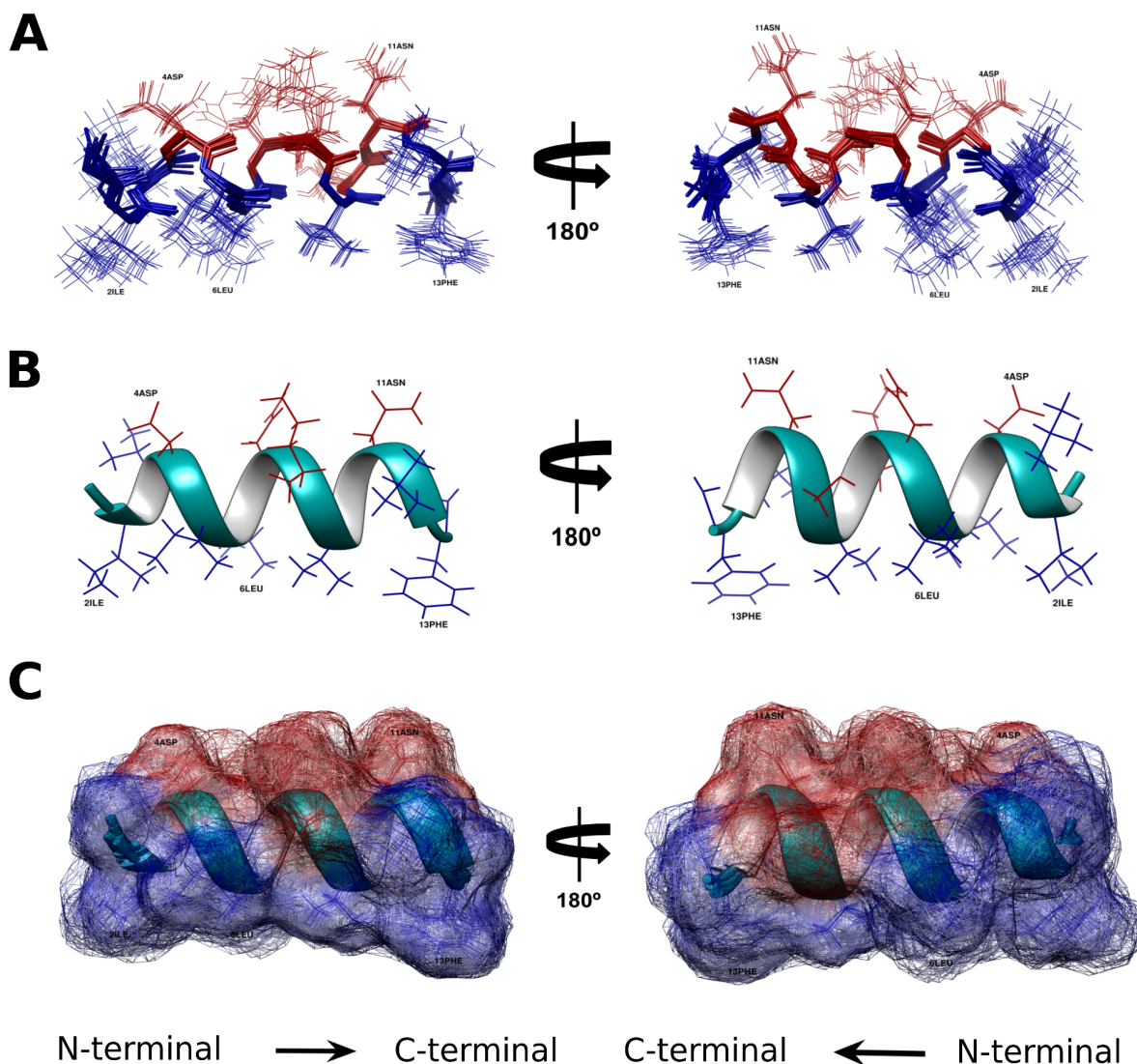
**Figura 16 - Estruturas tridimensionais do criptídeo schistocina-3 obtidas por RMN de solução (400  $\mu\text{M}$  de peptídeos em 20 mM de micelas DPC- $d_{38}$ ). As estruturas tridimensionais (400  $\mu\text{M}$ ) foram obtidas em micelas de DPC- $d_{38}$  (20 mM) no Bruker Avance**

III 500 (11,75T). Em a) as dez estruturas de menor energia (representação em varetas), evidenciando os resíduos hidrofóbicos, em azul, e os resíduos hidrofílicos, em vermelho; b) as estruturas de menor energia (representação em desenho); c) por último, a estrutura de menor energia (representação de superfície), com as regiões hidrofóbicas, em azul, e em regiões hidrofílicas, em vermelho.

O peptídeo derivado da schis-3, o schis-3.1, também apresenta uma estrutura enovelada em  $\alpha$ -hélice na presença de vesículas de DPC (Figura 17). A representação em varetas das dez estruturas de menor energia mostra uma maior dispersão dos resíduos 6Leu e 8Lys. A cadeia lateral dos outros resíduos apresentam melhor superposição, com RMSD de  $0,11 \text{ \AA} \pm 0,05 \text{ \AA}$  (Figura 17a). Além disso, há uma clara justaposição dos anéis da fenilalanina, que sugerem um papel deste resíduo na interação com as micelas. A representação em desenho evidencia que a estrutura de menor energia apresenta uma estrutura helicoidal com porção N-terminal mais afinada, porém menos do que o peptídeo schis-3 (Figura 17b). A representação de superfície evidencia que a substituição da 6Lys pela 6Leu aumentou a anfipaticidade do peptídeo, com as faces hidrofóbicas e hidrofílicas claramente separadas (Figura 17c).



## schistocina-3.1



**Figura 17 - Estruturas tridimensionais do criptídeo schistocina-3.1 obtidas por RMN de solução (400  $\mu\text{M}$  de peptídeos em 20 mM de micelas DPC- $d_{38}$ ). As estruturas tridimensionais (400  $\mu\text{M}$ ) foram obtidas em micelas de DPC- $d_{38}$  (20 mM) no Bruker Avance III 500 (11,75T). Em a) as dez estruturas de menor energia (representação em varetas), evidenciando os resíduos hidrofóbicos, em azul, e os resíduos hidrofílicos, em vermelho; b) as estruturas de menor energia (representação em desenho); c) por último, a estrutura de menor energia (representação de superfície), com as regiões hidrofóbicas, em azul, e em regiões hidrofílicas, em vermelho.**

A Tabela 6 resume a análise dos diagramas de Ramachandran (estatísticas) com os resíduos nas regiões mais favorecidas e os resíduos nas regiões permitidas conforme determinado por PROCHECK, além dos Z scores, mensurados pelo WHATCHECK, indicando uma alta qualidade estereoquímica das estruturas mais estáveis obtidas para todos os peptídeos.

**Tabela 6 - Sumário de estatísticas das estruturas da schistocina.**

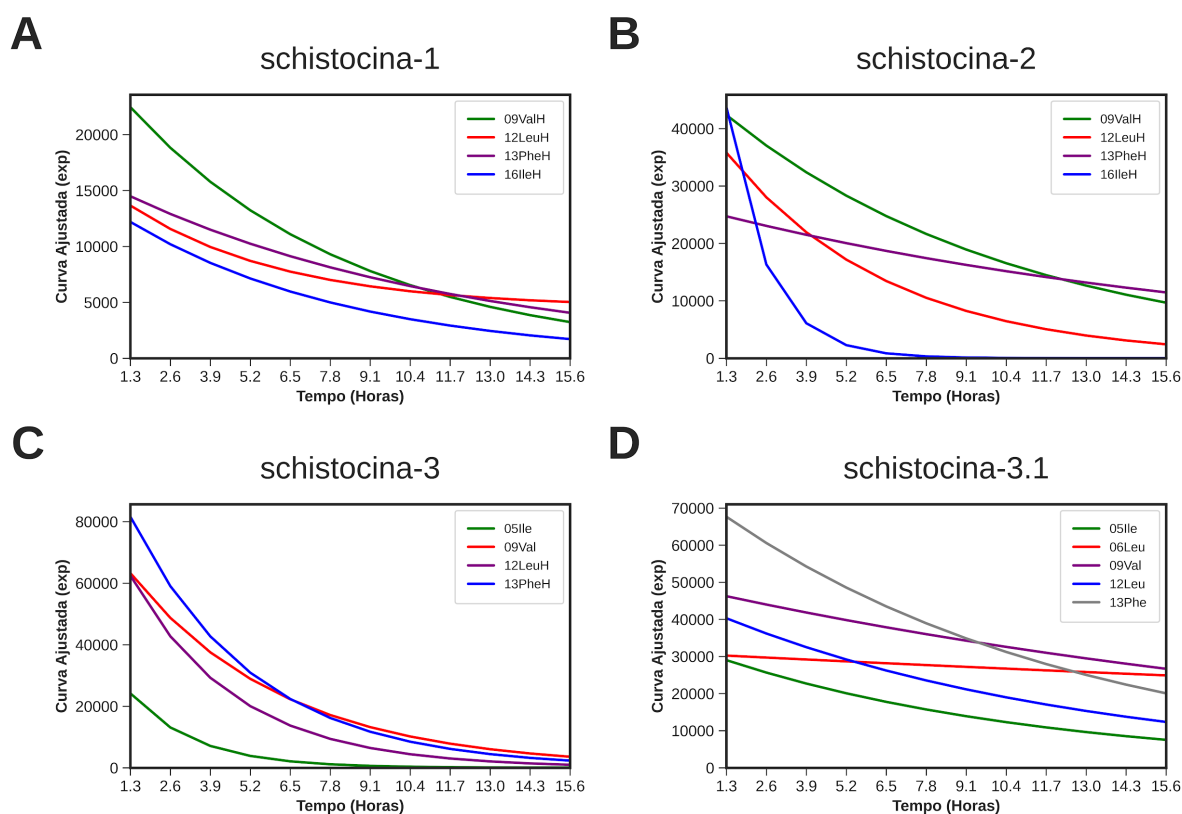
Sumário estatístico das estruturas																	
		schistocina-1				schistocina-2				schistocina-3				schistocina-3.1			
Restrições de distância e Diedros																	
NOEs																	
Intra Residuais		101				101				97				100			
Inter Residuais		71				70				48				76			
Sequencial		39				44				25				44			
Alcance pequeno		28				25				21				28			
Alcance médio		4				1				2				4			
Alcance longo		0				0				0				0			
Total de																	
não-ambíguos		172				171				145				176			
Ambíguos		45				65				14				29			
Ângulos (psi + phi)		34				28				20				20			
Estatística estrutural																	
Precisão estrutural																	
RMSD Cadeia principal (tudo/ estrutura secundária)		0.32 ± 0.12 / 0.75 ± 0.30				0.10 ± 0.02 / 0.24 ± 0.05				0.16 ± 0.05 / 0.37 ± 0.05				0.11 ± 0.05 / 0.45 ± 0.19			
RMSD Átomo pesado (tudo/ estrutura secundária)		0.90 ± 0.11 / 1.31 ± 0.38				0.40 ± 0.03 / 0.47 ± 0.03				0.60 ± 0.09 / 0.63 ± 0.11				0.68 ± 0.09 / 0.93 ± 0.28			
		valor	erro	min	max	valor	erro	min	max	valor	erro	min	max	valor	erro	min	max
PROCHECK																	
regiões mais favorecidas		99.4	1.82	94.1	100	100	0	100	100	98.3	3.41	91.7	100	95	5.67	83.3	100
regiões permitidas		0.59	1.82	0	5.9	0	0	0	0	1.66	3.41	0	8.3	4.99	5.67	0	16.7
regiões generosamente permitidas		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
regiões não permitidas		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WHATCHECK																	
Ramachandran plot Z score (RAMCHK)		-6.63	1.35	-9.73	-4.69	-4.27	1.28	-6.32	-2.39	-1.84	1.18	-3.66	0.751	-1.12	2.23	-4.62	1.76
Normalidade de rotâmetro x-1 x-2																	
Z-score (C12CHK)		-7.28	1.21	-9.88	-4.92	-9.7	1.09	-11.4	-7.65	-8.08	0.978	-9.54	-5.84	-7.19	1.7	-10.1	-4.79

Conformação de cadeia principal																	
Z-score (BBCCHK)	-1.84	0.597	-3.62	-1.38	0.38	0.079	0.288	0.445	0.305	0	0.305	0.305	0.303	0	0.303	0.303	
Tamanho das ligações																	
RMS Z-score (BNDCHK)	0.3	0.128	0.278	0.325	0.38	0.027	0.323	0.419	0.312	0.014	0.284	0.335	0.282	0.011	0.261	0.307	
Ângulos de ligação																	
RMS Z-score (ANGCHK)	0.395	0.168	0.367	0.43	0.49	0.076	0.377	0.578	0.325	0.02	0.302	0.361	0.271	0.034	0.261	0.383	
Restrições de ângulos																	
omega RMS Z-score (OMECHK)	0.734	0.707	0.606	0.877	0.64	0.060	0.563	0.733	0.455	0.055	0.329	0.549	0.447	0.057	0.323	0.549	
Planaridade das cadeias laterais RMS																	
Z-score (PLNCHK)	0.301	0.82	0.18	0.536	0.31	0.12	0.103	0.566	0.384	0.125	0.117	0.56	0.248	0.111	0.075	0.52	
Distribuição de diedros impróprios																	
RMS Z-score (HNDCHK)	0.51	0.194	0.471	0.549	0.50	0.049	0.418	0.579	0.468	0.025	0.423	0.505	0.458	0.034	0.41	0.562	

---

### 2.3.8 Troca de Hidrogênio-Deutério entre as schistocinas e micelas de DPC

Os experimentos de troca H/D de RMN permitiram avaliar as intensidades da correlação HN-H $\alpha$  e ajustar os decaimentos exponenciais para determinar as taxas de troca de hidrogênio amídico dos resíduos com grande decaimento (Figura 18). A maioria das trocas H/D dos resíduos para as schistocinas ocorreu em uma taxa muito rápida (1,33 horas). No entanto, para schis-1 e schis-2, as correlações HN-H $\alpha$  envolvendo 9Val, 12Ile, 13Phe e 16Ile estavam em troca lenta, sugerindo que eles estão localizados na interface da micela com baixa acessibilidade às moléculas de água. O 6Lys apresentado em schis-1, schis-2 e schis-3 trocou rapidamente com deutério, enquanto em schis-3.1 o 6Leu estava em troca lenta, indicando uma maior inserção do N-terminal deste peptídeo em micelas.



**Figura 18 - Representação em gráfico de ajuste exponencial de troca de hidrogênio amídico.** Curvas de ajuste para decaimento exponencial de resíduos lentos para a) schistocina-1; b) schistocina-2; c) schistocina-3 e d) schistocina-3.1.

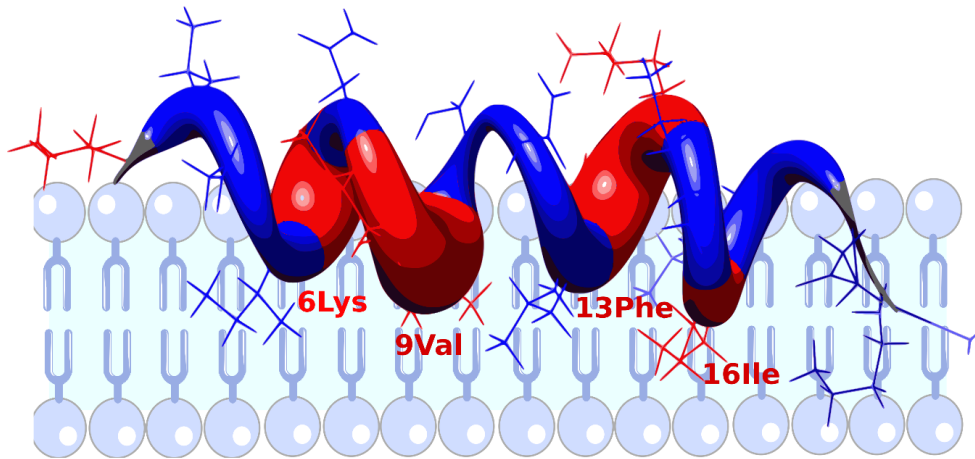
Esses dados, juntamente com a maior contribuição entrópica observada para a interação peptídeo-membrana, nos permitiram propor um modelo de inserção de peptídeo na

interface da membrana (Figura 19), sugerindo que as schis-1, schis-2 e schis-3 parecem estar fixados de forma quase paralela com a superfície da membrana, mas com a porção N-terminal (resíduos 1-6) do lado de fora da membrana, enquanto o schis-3.1 apresenta uma maior fixação na membrana, com auxílio do resíduo Leu-6, evidenciando a importância dessa substituição de resíduos para a atividade antimicrobiana da schis-3.1.

Na figura 19a, foi possível observar na estrutura da schis-1 que os resíduos 6Lys, 9Val, 13Phe e 16Ile são os resíduos que menos trocaram no tempo experimental, o que também coincide com a face mais hidrofóbica do peptídeo. As extremidades N-terminal e C-terminal tiveram maior taxa de troca, o que pode estar relacionado com o fato das extremidades serem menos enveladas. Na figura 19b, o peptídeo schis-2 mostra-se mais enterrado nas micelas pela porção C-terminal, com os resíduos 8Lys, 9Val, 12Leu, 13Phe e 16Ile como aqueles com menor taxa de troca. Novamente, a maior parte dos resíduos hidrofóbicos parecem ser responsáveis pelo ancoramento do peptídeo na membrana. O criptídeo schis-3 evidencia uma partição entre a porção C-terminal inserida na membrana, com os resíduos 9Val, 12Ile e 13Phe, e a porção N-terminal com altas taxas de troca, podendo estar trocando entre posições intra e extra-membranares (Figura 19c). Por fim, o peptídeo schis-3 apresenta as maiores taxas de resíduos que não trocaram: 2Ile, 5Ile, 6Leu, 9Val, 12Leu e 13Phe, demonstrando que os resíduos hidrofóbicos do peptídeos são responsáveis pela inserção do peptídeo de modo paralelo ao eixo da membrana (Figura 19d).

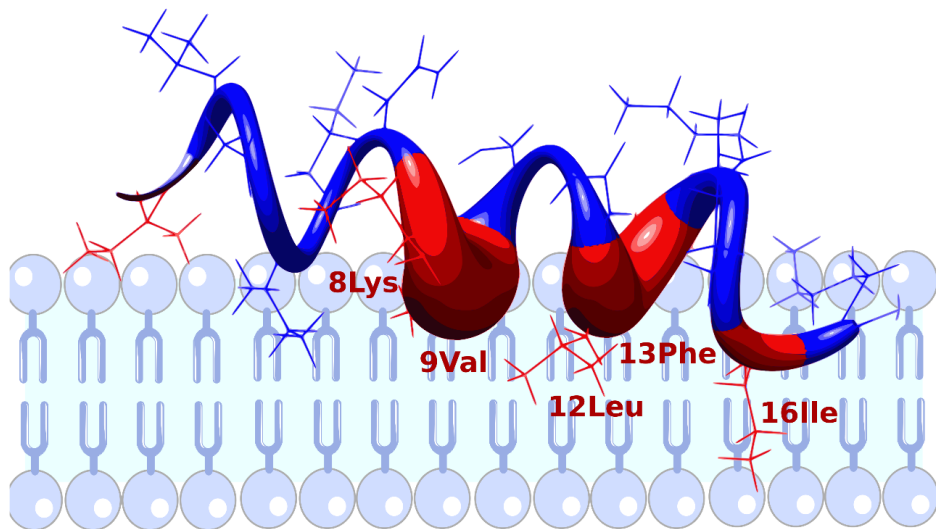
**A**

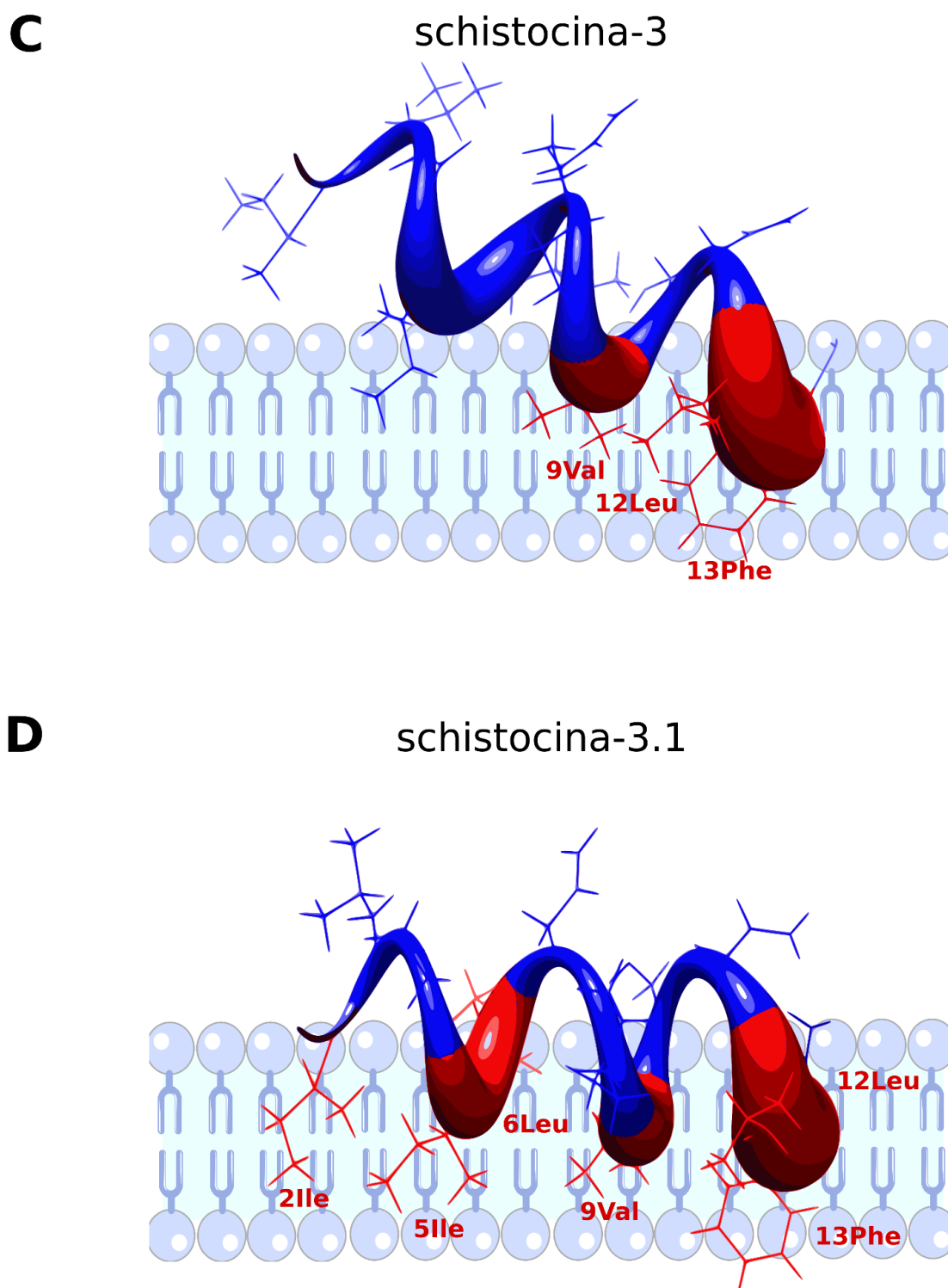
schistocina-1



**B**

schistocina-2



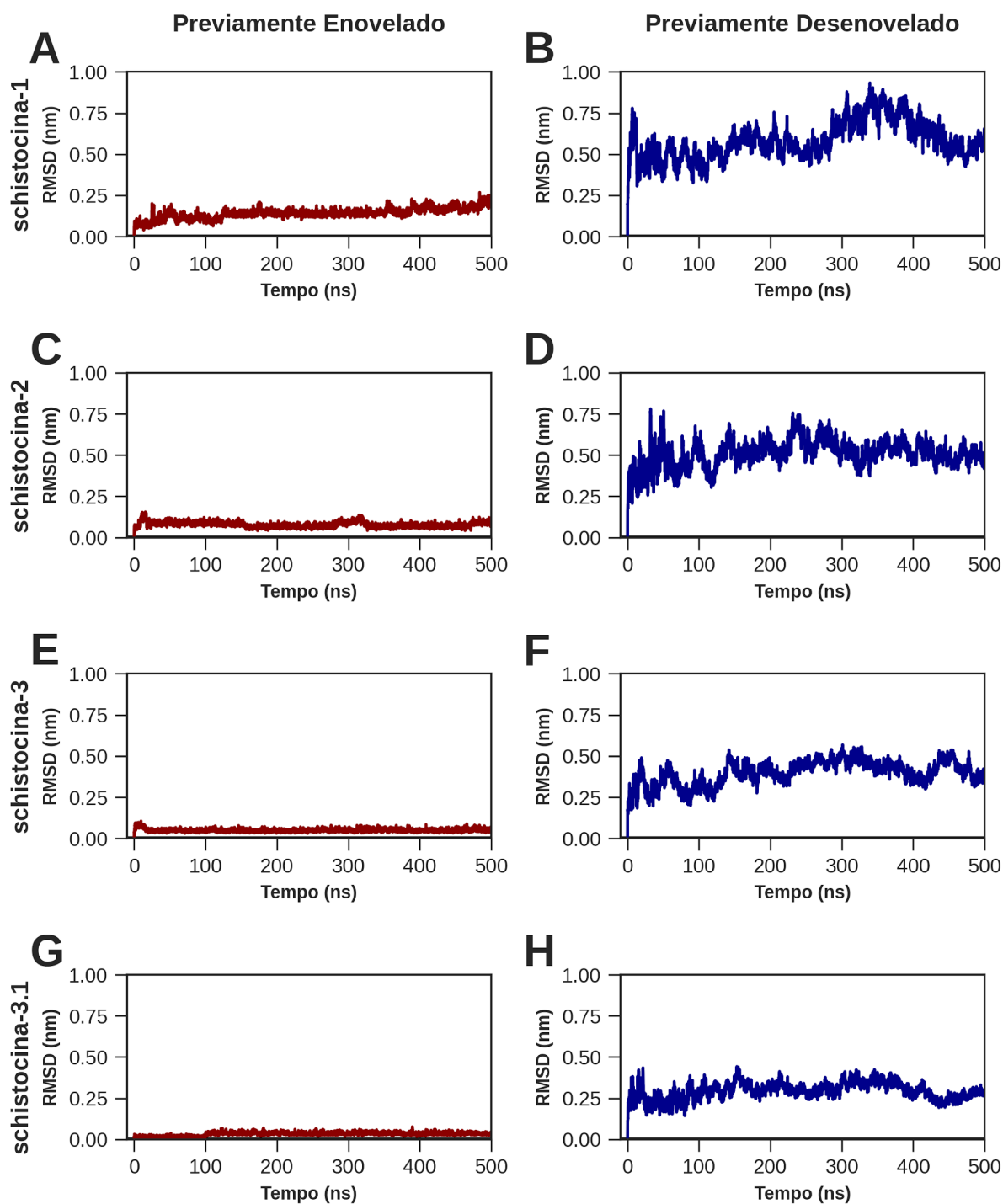


**Figura 19 - Troca de hidrogênio amídico mostrado em proporção de volume para cada resíduo nas schistocinas.** Resíduos com baixa taxa de troca são mostrados com volume maior, enquanto resíduos com menor volume são estão em azul e tem troca rápida durante experimento de TOCSY hidrogênio-deutério.

### 2.3.9 Efeito dos estruturas de RMN na interação com micelas zwitteriônicas

Uma outra etapa no estudo de interação peptídeo-membrana pode ser feita por meio de simulações de dinâmica. É comum usar as estruturas de RMN para estes estudos. Entretanto, as schistocinas atuam como a maioria dos peptídeos membrano-ativos: são desenoveladas em solução e formam  $\alpha$ -hélices em membranas lipídicas. Como descrito previamente, as estruturas das schistocinas foram resolvidas em micelas de DPC. Para verificar se as estruturas obtidas nos experimentos de RMN de solução poderiam interferir em ensaios de interação peptídeo-micelas, foram feitas simulações de dinâmica para entender como ocorre a interação entre as schistocinas desenoveladas e enoveladas e as micelas. A figura 20 resume os dados de RMSD total das quatro schistocinas (schis-1, schis-2, schis-3 e schis-3.1) em micelas de DPC. Os peptídeos foram colocados em caixas de água à distância de 30 Å da micelas, mas rapidamente entraram em contato com as micelas. Na parte esquerda do painel da figura 20 é possível observar que, se tratando dos peptídeos previamente estruturados, os quatro peptídeos obtiveram desvios de RMS menor do 0.25 nm (Figura 20a, 20c, 20e, 20g). Dos quatro peptídeos, a schis-1, de 20 resíduos e o maior dos quatros, é o que tem o maior valor de RMSD, sendo que os outros apresentaram valores muito próximos a zero. O mesmo não se conclui a partir dos resultados obtidos por peptídeos desenovelados, como observado na coluna direita da figura 20. Como é possível observar, tanto a schis-1 e a schis-2 (20-mer e 17-mer) possuem RMSD médio de 0,5 Å nos 500 ns observados (Figura 20b e 20d), enquanto a schis-3 e schis-3.1 apresentam RMSD médio de 0,3 Å e 0,25 Å (Figura 20f e 20h). Esses dados indicam a importância do tamanho do peptídeo no tempo de enovelamento e sugere-se que o estado de enovelamento interfere sim no estudo de interação peptídeo-micela, considerando-se um tempo de 500 ns.

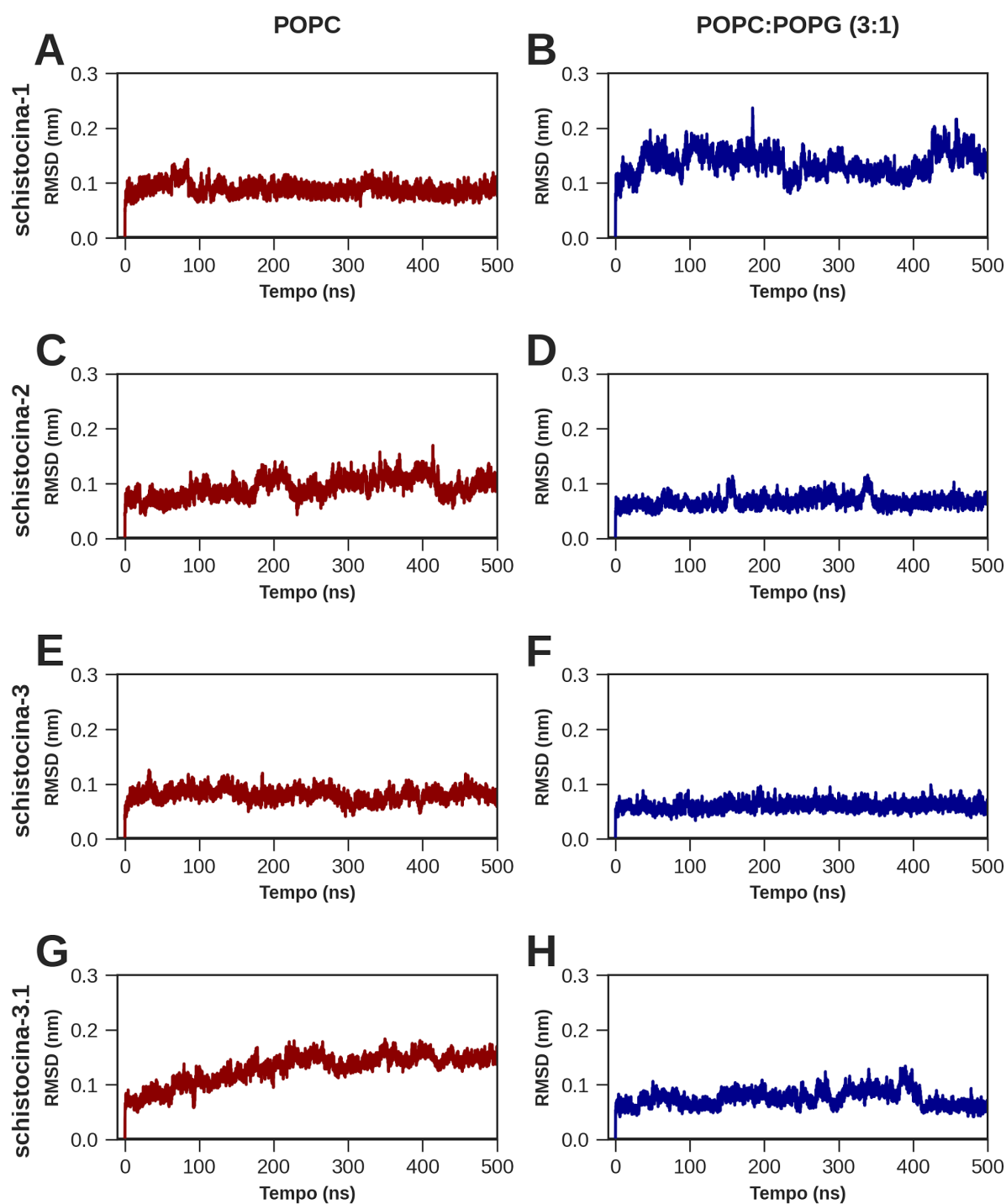




**Figura 20 - O desvio (RMSD) das schistocinas em interação com micelas de DPC.** Os criptídeos previamente enovelados (coluna da esquerda, em vermelho) e previamente desenovelados (coluna da direita, em azul) foram posicionados a 30 Å de uma micela de DPC e avaliados em estabilidade por 500 ns. Os resultados são os valores médios de simulações realizadas em triplicatas.

### 2.3.10 Comportamento das estruturas de RMN das schistocinas em bicamadas lipídicas

Um dos empecilhos do estudo de peptídeo por RMN de solução é observar a estrutura dos peptídeos em vesículas, formadas por bicamadas lipídicas, devido a limitações da técnica. Por isso, as estruturas das schistocinas obtidas por RMN de solução foram colocadas em contato com bicamadas zwitteriônicas (POPC) e aniônicas (POPC:POPG 3:1) para estudar essa interação. Diferentemente do estudo com micelas, em que os peptídeos foram colocados distantes das micelas, aqui os peptídeos foram colocados na caixa de água próximos às cabeças polares dos fosfolipídios, na porção da camada superior. A figura 21 resume os resultados de RMSD total das schistocinas em contato com os dois tipos de bicamadas. Com relação a interação com a bicamada zwitteriônica (coluna à esquerda), é possível observar um RMSD médio próximo a 0,1 nm, exceto no caso da schis-3.1 (Figura 21g), cujo RMSD se estabiliza a partir de 250 ns, em 0,15 nm. Por outro lado, na interação das schistocinas com a bicamada aniônica (coluna à direita), é possível observar um RMSD médio menor do que 0,1 nm para schis-2, schis-3 e schis-3.1 (Figura 21c, 21e e 21g). O RMSD médio das triplicatas entre schis-1 e POPC:POPG (3:1) ficou entre 0,15 e 0,1 nm (Figura 21b). Estes resultados sugerem que a interação das schistocinas com bicamadas lipídicas é estável, com valores de RMSD baixos (próximos a 0,1 nm), com uma tendência de maior afinidade por bicamadas aniônicas, corroborados por estudos experimentais.



**Figura 21 - O desvio (RMSD) das schistocinas em bicamadas zwitteriônicas e aniônicas.** Os criptídeos enovelados foram posicionados próximos a bicamadas zwitteriônicas (coluna da esquerda, em vermelho) e aniônicas (coluna da direita, em azul), e foram avaliados em estabilidade por 500 ns. Os resultados são os valores médios de simulações realizadas em triplicatas.

## 2.4 Discussão

Neste estudo, foi realizado o *design*, a síntese, a caracterização antimicrobiana e biofísica de uma nova família de criptídeos denominados schistocinas. Esses novos criptídeos foram selecionados entre diversos peptídeos preditos pelo nosso algoritmo após proteólise da proteína SmKI-1 *in silico*. Usando a abordagem de aprendizado de máquina (floresta randômica), atingiu-se uma predição com 87% de acurácia e descobriu-se com sucesso novos peptídeos antimicrobianos encriptados na proteína inibidor do tipo Kunitz de *Schistosoma mansoni*, SmKI-1. Também demonstrou-se como as propriedades físico-químicas podem ser usadas para encontrar novos peptídeos bioativos encriptados dentro de proteínas previamente conhecidas.

As abordagens experimentais e computacionais na descoberta de criptídeos foram descritas anteriormente, e alguns estudos mostraram que os criptídeos podem ser encontrados por proteômica *up-down* de tecidos, seguido por ensaios de bioatividade (14), como a descoberta de peptídeos semelhantes à hemorfina (82), criptídeos de *Physcomitrella patens* (83) e os peptídeos salivares humanos (84). Além disso, diversos grupos de pesquisa vêm desenvolvendo metodologias computacionais para a identificação de criptídeos. Por exemplo, os mitocriptídeos foram identificados em proteínas mitocondriais obtidas da base de dados *Swiss-Prot*, que foi seguido por uma clivagem proteolítica *in silico*, e pela seleção de peptídeos com propriedades físico-químicas semelhantes aos peptídeos mastoparanos (85), como a disposição de cargas positivas e a presença de resíduos específicos (1,9,10). Outro exemplo é o *software* Kamal, proposto por Brand et al. (2012), que faz a análise de componentes principais (PCA) de propriedades físico-químicas, utilizadas para agrupamento de peptídeos bioativos e que permite a identificação de novos peptídeos antimicrobianos localizados em regiões semelhantes aos agrupamentos dos PCAs. O Kamal incluiu a seleção de informações de todo o genoma de uma grande variedade de organismos, junto com AMPs selecionados como um controle positivo (19–21). O *software* *PeptideLocator* tem como objetivo geral a identificação de peptídeos bioativos em sequências proteicas, sendo que a predição do peptídeo bioativo é realizada por algoritmos que identificam as posições dos resíduos e as probabilidades estruturais (hélice, beta ou curva) (86–88). O *PeptideDB* possui um fluxograma semelhante, mas com finalidade distinta: identificar peptídeos bioativos de

proteínas de alimentos em sequências de entrada, porém, limitando para peptídeos já presentes em sua base de dados, sem uma ferramenta de predição (89,90).

O *software encrypted* se encaixa neste campo de pesquisa pois segue o procedimento de análise de sequências proteicas, seguido por proteólise das sequências proteicas e do cálculo de propriedades físico-químicas, culminando em uma lista de peptídeos, que utiliza um modelo de aprendizado de máquina de classificação supervisionado, floresta randômica, para previsão antibacteriana. O conceito do *software* se propõe a unir tarefas que geralmente são realizadas separadamente: (i) a digestão *in silico* e (ii) a predição físico-química e de bioatividade; e abre um campo para expandir a previsão de bioatividade: antifúngica, antiparasitária, imunostimuladora e outras, já que as bases de dados atuais são curadas e disponibilizadas. Além disso, este estudo propôs uma nova abordagem para explorar proteínas desconhecidas, dando uma olhada mais de perto nas informações encriptadas.

Como prova de conceito, iniciamos um estudo usando o conhecimento de que a proteína SmKI-1 deve ser clivada do tegumento para apresentar atividade (35,36), e seu outro domínio poderia ser digerido por degradação ou proteólise das proteínas e liberar fragmentos bioativos na circulação do hospedeiro. Selecionou-se três peptídeos de um conjunto de 351 e, após uma análise de padrão secundário *in silico*, projetou-se o quarto peptídeo (schis-3.1) com uma única substituição para aumentar a sua anfipaticidade (Figura 4). Esses peptídeos foram sintetizados em fase sólida e testados *in vitro* contra cepas de bactérias e fungos que são patogênicas para humanos, e os MICs foram comparáveis a outros peptídeos antimicrobianos (91,92). Os peptídeos schis-3 e schis-3.1 exibiram um amplo espectro antimicrobiano, enquanto schis-1 e schis-2 apresentaram atividade leve contra cepas de *Candida krusei*. Comparando nossos achados com outros criptídeos antimicrobianos com atividade, pôde-se observar que o schis-3 e o schis-3.1 são semelhantes em eficiência aos criptídeos ApoB<sup>131,132</sup> (de apolipoproteína), que têm atividade antimicrobiana/antibiofilme com valores de MIC de 20  $\mu$ M (6). Eles também são semelhantes a rpS30-2<sup>133</sup>, da proteína intracelular eucariótica p62, com valor MIC de 100  $\mu$ M contra atividade antimicobacteriana (93). A fim de obter informações que pudessem explicar a divergência de atividade biológica entre as schistocinas, a interação de peptídeos com membrana foi investigada por um conjunto de técnicas biofísicas (94–96). Membranas e micelas biomiméticas são úteis para entender a estruturação

<sup>131</sup> Sequência da r(P)ApoB<sub>L</sub>: PHVALKPGKCLKFIIPSPKRPVKLLSGGNTLHLVSTTKT

<sup>132</sup> Sequência da r(P)ApoB<sub>S</sub>: PHVALKPGKCLKFIIPSPKRPVKLLSG

<sup>133</sup> Sequência do rpS30-2: NRRFNVPVPTFGK

de peptídeos em ambientes semelhantes às bicamadas encontradas em procariotos e eucariotos, além de simular a eletrostática interfacial responsável pela afinidade dos peptídeos (97).

Como diversos outros peptídeos catiônicos com atividade antimicrobiana (98–100) quando na presença de modelos miméticos de membrana, as schistocinas exibem estrutura de  $\alpha$ -hélice (Figura 8 e Figura 9). Esses peptídeos anfipáticos com carga líquida positiva adotam uma estrutura  $\alpha$ -helicoidal em contato com a superfície hidrofílica/hidrofóbica, como a membrana de microorganismos (101). Um exemplo é o peptídeo GK Y25<sup>134</sup>, um criptídeo antimicrobiano derivado da proteína trombina humana (102). Os resultados de CD descrevem a transição de alça randômica para  $\alpha$ -hélice das schistocinas em tampão aquoso para LUVs (3:1) POPC:POPG (Figura 8). Os peptídeos schis-1 e schis-2 têm fraca intensidade de formação de hélice com POPC e helicidade moderada com POPC:POPG. Por outro lado, schis-3 e schis-3.1 após titulação da membrana mostraram helicidade peptídeo-fosfolípido significativa, destacada em vesículas aniônicas (Figura 8). Esses resultados são corroborados pelas simulações de MD, que analisou a interação dos peptídeos com bicamadas aniônicas e zwitteriônicas (Figura 21). O peptídeo schis-1 é o que mais demora para atingir níveis similares de RMSD igual ao dos outros peptídeos e a interação das schistocinas é mais forte com membranas aniônicas. Outras análises com essas corridas de MD ainda precisam ser feitas para obter informações eletrostáticas e sobre quais resíduos são responsáveis pela interação com a bicamada.

Em geral, os estudos termodinâmicos sobre as interações membranares das schistocinas indicaram que esses peptídeos apresentam afinidade de moderada a alta em vesículas carregadas negativamente apenas em 35 °C e que resultam em processo exotérmico. A interação das schistocinas à 25 °C não mostrou qualquer ligação significativa e nenhum parâmetro termodinâmico pôde ser obtido (dados não mostrados). A dependência da temperatura para a interação peptídeo-membrana sugere que o componente entrópico é predominante, como consequência do efeito hidrofóbico causado pela dessolvatação da água que ocorre durante a interação peptídeo-membrana. De fato, as principais forças motrizes da interação entre as schistocinas e a membrana fosfolípida são as contribuições entrópicas (103). Curiosamente, a interação schis3.1 revela maior fator entrópico e maior constante de ligação, que está de acordo com os dados de RMN, que mostram melhor particionamento dos

---

<sup>134</sup> Sequência do GK Y25: GK YGFYTHVFRLLKKWIQKVIDQFGE

resíduos hidrofílicos e hidrofóbicos dentro da interface de bicamada. Por outro lado, uma maior contribuição entálpica observada na interação schis-1, schis-2 e schis-3 sugere que a atração eletrostática pode ainda ser um fenômeno importante para o processo de ligação à membrana desses peptídeos (104). Mesmo que o schis-3.1 tenha carga líquida +1, o schis-1, schis-2 e schis-3 apresentam superfícies hidrofílicas catiônicas mais altas para estabelecer forças coulombianas atrativas com a superfície carregada negativamente das membranas POPC:POPG [102,103]. Assim, a maior atividade antimicrobiana de schis-3.1 poderia estar relacionada à maior afinidade de membrana devido a uma inserção ótima dos resíduos hidrofóbicos, apontando para inserções na membrana por N-terminais ancorados pelo resíduo 6Leu. Os dados de RMN confirmaram que os peptídeos se enovelam em estruturas tridimensionais de  $\alpha$ -hélice bem definidas, na presença de DPC- $d_{38}$ , e corroboraram a hipótese de que a 6Lys estava prejudicando a anfipaticidade do peptídeo, além de fornecer informações sobre a flexibilidade do peptídeo em micelas (Figura 17). Schis-1 e schis-2 têm um C-terminal mais flexível, enquanto schis-3 e schis-3.1 têm uma hélice rígida, com a cadeia lateral da 13Phe, na extremidade do C-terminal, em troca lenta, mostrando sua afinidade com a membrana e aumentando a especificidade destes peptídeos. Além disso, resíduos hidrofóbicos são conhecidos por serem importantes na estabilização de peptídeos na membrana (105,106).

Conforme descrito anteriormente, o conjunto de dados do modelo foi construído com peptídeos de até 20 resíduos de aminoácidos, porque os peptídeos usados como controles negativos foram obtidos de um banco de dados que limitava o tamanho de seus peptídeos. Restringir os tamanhos de AMPs foi essencial para não construir um conjunto de dados desbalanceado, levando a erros de predição (107). Uma abordagem alternativa poderia ser a geração de peptídeos a partir de bancos de dados Uniprot com uma grande faixa de massa molecular e usá-los como controle negativo. Ainda assim, uma extensa curadoria seria necessária para distinguir as atividades e não prejudicar a previsão da bioatividade. Além disso, os dados obtidos sugerem, como mostrado na tabela 3, que dois dos peptídeos selecionados não exibiram forte atividade antimicrobiana contra as cepas bacterianas e fúngicas estudadas. No entanto, isso não é uma desvantagem, pois as sequências de schis-1 e schis-2 são precursoras de schis-3, que apresentaram atividade antimicrobiana e assim, essenciais para entender quais diferenças estruturais foram responsáveis pela divergência das atividades. Considerando como um impedimento do algoritmo, isso poderia ser resolvido

adicionando novas camadas à previsão do peptídeo bioativo, conforme proposto por Lin et al. (2019), bem como outros aspectos em previsões de peptídeos, como a preferência por posição resíduo específico (27,28). A robustez deste estudo é a abordagem de combinar o esforço computacional e a verificação experimental, o que resultou em um melhor entendimento sobre identificação de criptídeos.

Até onde sabe-se, este é o primeiro relato de criptídeos da SmKI-1 com atividade antimicrobiana. Este trabalho fortalece vários outros que encontraram criptídeos derivados de vias responsáveis pela degradação de proteínas ou outras com muitas funções, muitas vezes completamente diferentes daquelas da proteína original (6,10,108). Além disso, ele traz uma nova funcionalidade para o domínio C-terminal SmKI-1 usando o conceito de criptídeos. Demonstrou-se que é possível olhar para as sequências de proteínas de forma diferente e obter mais informações sobre as funções das proteínas. Sobre esta proteína multifacetada, tem mais por vir e a ser aprendido, e muito provavelmente como muitos outros peptídeos antimicrobianos descritos na literatura, ainda existem vários experimentos a serem feitos para melhorar seu uso potencial como biofarmacêutico. Portanto, descrevemos uma nova ferramenta para procurar novos peptídeos e descobrir funções ocultas de proteínas. Além disso, observou-se que a afinidade dos peptídeos originados do domínio C-terminal só interagem com as vesículas à temperatura semelhante à do corpo humano. Este efeito pode estar relacionado com: a) a energia necessária para as interações peptídeo-peptídeo em uma forma agregada para dar origem às interações peptídeo-fosfolípido durante o processo de ligação; b) especificidade do parasita na adaptação ao hospedeiro humano, fazendo com que os esquistossômulos retenham o SmKI-1 no tegumento e liberem apenas em oscilações específicas de temperatura, variação de pH ou ação enzimática externa.

Assim, estudos futuros devem considerar um olhar mais profundo sobre a orientação das schistocinas em micelas DPC- $d_{38}$  usando sondas paramagnéticas em solução NMR. Um estudo anterior mostrou que o gadolínio é inerte com proteínas e micelas DPC- $d_{38}$ . Sua adição ao tubo de NMR resulta em aprimoramentos de relaxamento paramagnético que tornam possível informar a orientação de estruturas  $\alpha$ -helicoidais (109–111). Quanto ao *encrypted*, o algoritmo está sendo aprimorado para prever outros peptídeos bioativos e oferecer modificações de resíduos que podem levar a um melhor desempenho. A união da análise de criptídeos e do *design* racional de peptídeos permitirá concentrar nossos esforços na



caracterização biofísica e no uso de dados experimentais para criar, em um futuro próximo, um preditor de criptídeos mais eficiente.

## 2.5 Referências

1. Ueki N, Someya K, Matsuo Y, Wakamatsu K, Mukai H. Cryptides: functional cryptic peptides hidden in protein structures. *Biopolymers*. 2007;88(2):190–8.
2. Ivanov VT, Karelina AA, Philippova MM, Nazimov IV, Pletnev VZ. Hemoglobin as a source of endogenous bioactive peptides: the concept of tissue-specific peptide pool. *Biopolymers*. 1997;43(2):171–88.
3. Abou-Diab M, Thibodeau J, Deracinois B, Flahaut C, Fliss I, Dhulster P, et al. Bovine Hemoglobin Enzymatic Hydrolysis by a New Eco-Efficient Process-Part II: Production of Bioactive Peptides. *Membranes (Basel)*. 2020 Sep 29;10(10).
4. Groß R, Bauer R, Krüger F, Rücker-Braun E, Olari L-R, Ständker L, et al. A Placenta Derived C-Terminal Fragment of  $\beta$ -Hemoglobin With Combined Antibacterial and Antiviral Activity. *Front Microbiol*. 2020;11:508.
5. Wei F, Zhao L, Jing Y. Hemoglobin-derived peptides and mood regulation. *Peptides*. 2020 Feb 15;127:170268.
6. Gaglione R, Cesaro A, Dell’Olmo E, Della Ventura B, Casillo A, Di Girolamo R, et al. Effects of human antimicrobial cryptides identified in apolipoprotein B depend on specific features of bacterial strains. *Sci Rep*. 2019 Apr 30;9(1):6728.
7. van der Kraan MIA, Groenink J, Nazmi K, Veerman ECI, Bolscher JGM, Nieuw Amerongen AV. Lactoferrampin: a novel antimicrobial peptide in the N1-domain of bovine lactoferrin. *Peptides*. 2004 Feb;25(2):177–83.
8. Papareddy P, Rydengård V, Pasupuleti M, Walse B, Mörgelin M, Chalupka A, et al. Proteolysis of human thrombin generates novel host defense peptides. *PLoS Pathog*. 2010 Apr 22;6(4):e1000857.
9. Mukai H, Hokari Y, Seki T, Takao T, Kubota M, Matsuo Y, et al. Discovery of mitocryptide-1, a neutrophil-activating cryptide from healthy porcine heart. *J Biol Chem*. 2008 Nov 7;283(45):30596–605.
10. Mukai H, Seki T, Nakano H, Hokari Y, Takao T, Kawanami M, et al. Mitocryptide-2: purification, identification, and characterization of a novel cryptide that activates neutrophils. *J Immunol*. 2009 Apr 15;182(8):5072–80.
11. Iavarone F, Desiderio C, Vitali A, Messana I, Martelli C, Castagnola M, et al. Cryptides: latent peptides everywhere. *Crit Rev Biochem Mol Biol*. 2018 Jun;53(3):246–63.
12. Autelitano DJ, Rajic A, Smith AI, Berndt MC, Ilag LL, Vadas M. The cryptome: a subset of the proteome, comprising cryptic peptides with distinct bioactivities. *Drug Discov Today*. 2006 Apr;11(7–8):306–14.
13. Pimenta DC, Lebrun I. Cryptides: Buried secrets in proteins. *Peptides*. 2007 Dec;28(12):2403–10.
14. Samir P, Link AJ. Analyzing the cryptome: uncovering secret sequences. *AAPS J*. 2011 Jun;13(2):152–8.
15. Nielsen H, Brunak S, von Heijne G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Engineering, Design and Selection*. 1999 Jan 1;12(1):3–9.
16. Cunningham JM, Koytiger G, Sorger PK, AlQuraishi M. Biophysical prediction of protein–peptide interactions and signaling networks using machine learning. *Nature Methods*. 2020 Feb;17(2):175–83.
17. Plisson F, Ramírez-Sánchez O, Martínez-Hernández C. Machine learning-guided discovery and

- design of non-hemolytic peptides. *Scientific Reports*. 2020 Oct 6;10(1):16581.
18. Seo M-D, Won H-S, Kim J-H, Mishig-Ochir T, Lee B-J. Antimicrobial peptides for therapeutic applications: a review. *Molecules*. 2012 Oct 18;17(10):12276–86.
  19. Brand GD, Magalhães MTQ, Tinoco MLP, Aragão FJL, Nicoli J, Kelly SM, et al. Probing protein sequences as sources for encrypted antimicrobial peptides. *PLoS One*. 2012;7(9):e45848.
  20. Brand GD, Ramada MHS, Genaro-Mattos TC, Bloch C. Towards an experimental classification system for membrane active peptides. *Sci Rep*. 2018 Jan 19;8(1):1194.
  21. Brand GD, Ramada MHS, Manickchand JR, Correa R, Ribeiro DJS, Santos MA, et al. Intragenic antimicrobial peptides (IAPs) from human proteins with potent antimicrobial and anti-inflammatory activity. *PLoS One*. 2019;14(8):e0220656.
  22. Li S, Wang Y, Xue Z, Jia Y, Li R, He C, et al. The structure-mechanism relationship and mode of actions of antimicrobial peptides: A review. *Trends in Food Science & Technology*. 2021 Mar 1;109:103–15.
  23. Greco I, Molchanova N, Holmedal E, Jenssen H, Hummel BD, Watts JL, et al. Correlation between hemolytic activity, cytotoxicity and systemic in vivo toxicity of synthetic antimicrobial peptides. *Sci Rep*. 2020 Aug 6;10(1):13206.
  24. Oddo A, Hansen PR. Hemolytic Activity of Antimicrobial Peptides. *Methods Mol Biol*. 2017;1548:427–35.
  25. Wheeler LC, Perkins A, Wong CE, Harms MJ. Learning peptide recognition rules for a low-specificity protein. *Protein Sci*. 2020 Nov;29(11):2259–73.
  26. Meher PK, Sahu TK, Saini V, Rao AR. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Sci Rep*. 2017 Feb 13;7:42362.
  27. Bhadra P, Yan J, Li J, Fong S, Siu SWI. AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Sci Rep*. 2018 Jan 26;8(1):1697.
  28. Joseph S, Karnik S, Nilawe P, Jayaraman VK, Idicula-Thomas S. ClassAMP: a prediction tool for classification of antimicrobial peptides. *IEEE/ACM Trans Comput Biol Bioinform*. 2012 Oct;9(5):1535–8.
  29. Lawrence TJ, Carper DL, Spangler MK, Carrell AA, Rush TA, Minter SJ, et al. amPEPpy 1.0: A portable and accurate antimicrobial peptide prediction tool. *Bioinformatics*. 2020 Nov 2;
  30. Agrawal P, Raghava GPS. Prediction of Antimicrobial Potential of a Chemically Modified Peptide From Its Tertiary Structure. *Front Microbiol*. 2018;9:2551.
  31. Lin Y, Cai Y, Liu J, Lin C, Liu X. An advanced approach to identify antimicrobial peptides and their function types for penaeus through machine learning strategies. *BMC Bioinformatics*. 2019 Jun 10;20(Suppl 8):291.
  32. Lata S, Sharma BK, Raghava GPS. Analysis and prediction of antibacterial peptides. *BMC Bioinformatics*. 2007 Jul 23;8:263.
  33. Mambelli F, Santos BPO, Morais SB, Gimenez EGT, Astoni DCDS, Braga AD, et al. *S. mansoni* SmKI-1 Kunitz-domain: Leucine point mutation at P1 site generates enhanced neutrophil elastase inhibitory activity. *PLoS Negl Trop Dis*. 2021 Jan;15(1):e0009007.
  34. Morais SB, Figueiredo BC, Assis NRG, Alvarenga DM, de Magalhães MTQ, Ferreira RS, et al. *Schistosoma mansoni* SmKI-1 serine protease inhibitor binds to elastase and impairs neutrophil function and inflammation. *PLoS Pathog*. 2018 Feb;14(2):e1006870.
  35. Morais SB, Figueiredo BC, Assis NRG, Homan J, Mambelli FS, Bicalho RM, et al. *Schistosoma mansoni* SmKI-1 or Its C-Terminal Fragment Induces Partial Protection Against *S. mansoni* Infection in Mice. *Front Immunol*. 2018;9:1762.
  36. Ranasinghe SL, Duke M, Harvie M, McManus DP. Kunitz-type protease inhibitor as a vaccine candidate against schistosomiasis mansoni. *Int J Infect Dis*. 2018 Jan;66:26–32.
  37. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev*

- Microbiol. 2005 Mar;3(3):238–50.
38. Matsuzaki K. Membrane Permeabilization Mechanisms. *Adv Exp Med Biol.* 2019;1117:9–16.
  39. Park Y, Hahm K-S. Antimicrobial peptides (AMPs): peptide structure and mode of action. *J Biochem Mol Biol.* 2005 Sep 30;38(5):507–16.
  40. Sepehri A, PeBenito L, Pino-Angeles A, Lazaridis T. What Makes a Good Pore Former: A Study of Synthetic Melittin Derivatives. *Biophys J.* 2020 Apr 21;118(8):1901–13.
  41. Hancock REW, Rozek A. Role of membranes in the activities of antimicrobial cationic peptides. *FEMS Microbiol Lett.* 2002 Jan 10;206(2):143–9.
  42. Bechinger B, Zasloff M, Opella SJ. Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. *Protein Sci.* 1993 Dec;2(12):2077–84.
  43. Andraö J, Monreal D, de Tejada GM, Olak C, Brezesinski G, Gomez SS, et al. Rationale for the Design of Shortened Derivatives of the NK-lysine-derived Antimicrobial Peptide NK-2 with Improved Activity against Gram-negative Pathogens\*. *Journal of Biological Chemistry.* 2007 May 18;282(20):14719–28.
  44. Hoang HN, Song K, Hill TA, Derksen DR, Edmonds DJ, Kok WM, et al. Short Hydrophobic Peptides with Cyclic Constraints Are Potent Glucagon-like Peptide-1 Receptor (GLP-1R) Agonists. *J Med Chem.* 2015 May 14;58(9):4080–5.
  45. Bjellqvist B, Basse B, Olsen E, Celis JE. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *ELECTROPHORESIS.* 1994;15(1):529–39.
  46. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982 May 5;157(1):105–32.
  47. Fauchere J-L, Pliska V. Hydrophobic parameters  $\pi$  of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur J Med Chem.* 1983;18:369–75.
  48. Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem.* 1978;47:251–76.
  49. Richards FM. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng.* 1977;6:151–76.
  50. Schein CH. Solubility as a Function of Protein Structure and Solvent Components. *Bio/Technology.* 1990 Apr;8(4):308–17.
  51. Hubbard SJ, Thornton JM. Naccess: Department of biochemistry and molecular biology, university college london. Software available at <http://www.bioinf.manchester.ac.uk/naccess/nacdownload.html>. 1993;
  52. Gautier R, Douguet D, Antony B, Drin G. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics.* 2008 Sep 15;24(18):2101–2.
  53. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
  54. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012 Apr 15;28(8):1166–7.
  55. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc.* 2006;1(6):2876–90.
  56. Perez-Iratxeta C, Andrade-Navarro MA. K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Structural Biology.* 2008 May 13;8(1):25.
  57. Abraham T, Lewis RNAH, Hodges RS, McElhaney RN. Isothermal titration calorimetry studies of the binding of the antimicrobial peptide gramicidin S to phospholipid bilayer membranes. *BIOCHEMJ.* 2005 Aug 23;44(33):11279–85.
  58. Wieprecht T, Apostolov O, Beyermann M, Seelig J. Interaction of a mitochondrial presequence with lipid membranes: role of helix formation for membrane binding and perturbation. *Biochemistry.* 2000 Dec 19;39(50):15297–305.
  59. Shaka AJ, Lee CJ, Pines A. Iterative schemes for bilinear operators; application to spin

- decoupling. *Journal of Magnetic Resonance* (1969). 1988 Apr 1;77(2):274–93.
60. Kumar A, Ernst RR, Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochemical and Biophysical Research Communications*. 1980 Jul 16;95(1):1–6.
  61. Willker W, Leibfritz D, Kerssebaum R, Bermel W. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magnetic Resonance in Chemistry*. 1993;31(3):287–92.
  62. Wüthrich K. *NMR with Proteins and Nucleic Acids*. *Europhys News*. 1986;17(1):11–3.
  63. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*. 2005;59(4):687–96.
  64. Hyberts SG, Goldberg MS, Havel TF, Wagner G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci*. 1992 Jun;1(6):736–51.
  65. Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson*. 2010 Feb;202(2):223–33.
  66. Bardiaux B, Malliavin T, Nilges M. ARIA for solution and solid-state NMR. *Methods Mol Biol*. 2012;831:453–83.
  67. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*. 1998 Sep 1;54(Pt 5):905–21.
  68. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature*. 1996 May 23;381(6580):272.
  69. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 1993;26(2):283–91.
  70. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004 Oct;25(13):1605–12.
  71. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004 Jul 1;32(Web Server issue):W665-667.
  72. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*. 2001 Aug 28;98(18):10037–41.
  73. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*. 2008;29(11):1859–65.
  74. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput*. 2016 Jan 12;12(1):405–13.
  75. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*. 2014;35(27):1997–2004.
  76. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal*. 2009 Jul 8;97(1):50–8.
  77. Jo S, Kim T, Im W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE*. 2007 Sep 12;2(9):e880.
  78. Wang Y, Schlamadinger DE, Kim JE, McCammon JA. Comparative molecular dynamics simulations of the antimicrobial peptide CM15 in model lipid bilayers. *Biochimica et*

- Biophysica Acta (BBA) - Biomembranes. 2012 May 1;1818(5):1402–9.
79. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B*. 2010 Jun 17;114(23):7830–43.
  80. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*. 1995 Sep 2;91(1):43–56.
  81. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984 Oct 15;81(8):3684–90.
  82. Ianzer D, Konno K, Xavier CH, Stöcklin R, Santos RAS, de Camargo ACM, et al. Hemorphin and hemorphin-like peptides isolated from dog pancreas and sheep brain are able to potentiate bradykinin activity in vivo. *Peptides*. 2006 Nov;27(11):2957–66.
  83. Fesenko I, Azarkina R, Kirov I, Kniazev A, Filippova A, Grafskaja E, et al. Phytohormone treatment induces generation of cryptic peptides with antimicrobial activity in the Moss *Physcomitrella patens*. *BMC Plant Biology*. 2019 Jan 7;19(1):9.
  84. Messana I, Cabras T, Pisano E, Sanna MT, Olianias A, Manconi B, et al. Trafficking and postsecretory events responsible for the formation of secreted human salivary peptides: a proteomics approach. *Mol Cell Proteomics*. 2008 May;7(5):911–26.
  85. Ho CL, Hwang LL. Structure and biological activities of a new mastoparan isolated from the venom of the hornet *Vespa basalis*. *Biochem J*. 1991 Mar 1;274 ( Pt 2):453–6.
  86. Mooney C, Haslam NJ, Holton TA, Pollastri G, Shields DC. PeptideLocator: prediction of bioactive peptides in protein sequences. *Bioinformatics*. 2013 May 1;29(9):1120–6.
  87. O'Brien KT, Mooney C, Lopez C, Pollastri G, Shields DC. Prediction of polyproline II secondary structure propensity in proteins. *R Soc Open Sci*. 2020 Jan;7(1):191239.
  88. Ruhanen H, Hurley D, Ghosh A, O'Brien KT, Johnston CR, Shields DC. Potential of known and short prokaryotic protein motifs as a basis for novel peptide-based antibacterial therapeutics: a computational survey. *Front Microbiol*. 2014;5:4.
  89. Panyayai T, Ngamphiw C, Tongsimma S, Mhuantong W, Limsripraphan W, Choowongkamon K, et al. PeptideDB: A web application for new bioactive peptides from food protein. *Heliyon*. 2019 Jul 1;5(7):e02076.
  90. Wang L, Niu D, Wang X, Khan J, Shen Q, Xue Y. A Novel Machine Learning Strategy for the Prediction of Antihypertensive Peptides Derived from Food with High Efficiency. *Foods*. 2021 Mar 6;10(3).
  91. Morikawa N, Hagiwara K, Nakajima T. Brevinin-1 and -2, unique antimicrobial peptides from the skin of the frog, *Rana brevipoda porsa*. *Biochem Biophys Res Commun*. 1992 Nov 30;189(1):184–90.
  92. Nunes LO, Munhoz VHO, Sousa AA, de Souza KR, Santos TL, Bemquerer MP, et al. High-resolution structural profile of hylaseptin-4: Aggregation, membrane topology and pH dependence of overall membrane binding process. *Biochim Biophys Acta Biomembr*. 2021 May 1;1863(5):183581.
  93. Ponpuak M, Davis AS, Roberts EA, Delgado MA, Dinkins C, Zhao Z, et al. Delivery of cytosolic components by autophagic adaptor protein p62 endows autophagosomes with unique antimicrobial properties. *Immunity*. 2010 Mar 26;32(3):329–41.
  94. Sonnichsen FD, Van Eyk JE, Hodges RS, Sykes BD. Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry*. 1992 Sep 22;31(37):8790–8.
  95. Gierasch LM, Lacy JE, Thompson KF, Rockwell AL, Watnick PI. Conformations of model peptides in membrane-mimetic environments. *Biophys J*. 1982 Jan;37(1):275–84.
  96. Damberg P, Jarvet J, Gräslund A. Micellar systems as solvents in peptide and protein structure determination. *Methods Enzymol*. 2001;339:271–85.
  97. Khandelia H, Kaznessis YN. Molecular dynamics investigation of the influence of anionic and

- zwitterionic interfaces on antimicrobial peptides' structure: implications for peptide toxicity and activity. *Peptides*. 2006 Jun;27(6):1192–200.
98. Halverson T, Basir YJ, Knoop FC, Conlon JM. Purification and characterization of antimicrobial peptides from the skin of the North American green frog *Rana clamitans*. *Peptides*. 2000 Apr;21(4):469–76.
  99. Porto WF, Irazazabal L, Alves ESF, Ribeiro SM, Matos CO, Pires AS, et al. In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. *Nature Communications*. 2018 Apr 16;9(1):1490.
  100. Hori Y, Demura M, Iwadate M, Ulrich AS, Niidome T, Aoyagi H, et al. Interaction of mastoparan with membranes studied by <sup>1</sup>H-NMR spectroscopy in detergent micelles and by solid-state <sup>2</sup>H-NMR and <sup>15</sup>N-NMR spectroscopy in oriented lipid bilayers. *Eur J Biochem*. 2001 Jan;268(2):302–9.
  101. Huang Y, Huang J, Chen Y. Alpha-helical cationic antimicrobial peptides: relationships of structure and function. *Protein Cell*. 2010 Feb;1(2):143–52.
  102. Kasetty G, Papareddy P, Kalle M, Rydengård V, Mörgelin M, Albiger B, et al. Structure-Activity Studies and Therapeutic Potential of Host Defense Peptides of Human Thrombin<sup>∇</sup>. *Antimicrob Agents Chemother*. 2011 Jun;55(6):2880–90.
  103. Verly RM, Rodrigues MA, Daghananli KRP, Denadai AML, Cuccovia IM, Bloch C, et al. Effect of cholesterol on the interaction of the amphibian antimicrobial peptide DD K with liposomes. *Peptides*. 2008 Jan;29(1):15–24.
  104. Seelig J. Thermodynamics of lipid-peptide interactions. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2004 Nov 3;1666(1):40–50.
  105. Chang T-W, Wei S-Y, Wang S-H, Wei H-M, Wang Y-J, Wang C-F, et al. Hydrophobic residues are critical for the helix-forming, hemolytic and bactericidal activities of amphipathic antimicrobial peptide TP4. *PLoS One*. 2017;12(10):e0186442.
  106. Chen Y, Guarnieri MT, Vasil AI, Vasil ML, Mant CT, Hodges RS. Role of Peptide Hydrophobicity in the Mechanism of Action of  $\alpha$ -Helical Antimicrobial Peptides. *Antimicrobial Agents and Chemotherapy*. 2007 Apr 1;51(4):1398–406.
  107. Sales AP, Tomaras GD, Kepler TB. Improving peptide-MHC class I binding prediction for unbalanced datasets. *BMC Bioinformatics*. 2008 Sep 19;9(1):385.
  108. Ciociola T, Zanello PP, D'Adda T, Galati S, Conti S, Magliani W, et al. A Peptide Found in Human Serum, Derived from the C-Terminus of Albumin, Shows Antifungal Activity In Vitro and In Vivo. *Microorganisms*. 2020 Oct 21;8(10).
  109. Zangger K, Respondek M, Göbl C, Hohlweg W, Rasmussen K, Grampp G, et al. Positioning of micelle-bound peptides by paramagnetic relaxation enhancements. *J Phys Chem B*. 2009 Apr 2;113(13):4400–6.
  110. Zangger K, Gössler R, Khatai L, Lohner K, Jilek A. Structures of the glycine-rich diastereomeric peptides bombinin H2 and H4. *Toxicon*. 2008 Aug 1;52(2):246–54.
  111. Pintacuda G, Otting G. Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *J Am Chem Soc*. 2002 Jan 23;124(3):372–3.

## CAPÍTULO 3. Acetilação N-terminal e amidação C-terminal usando *CcpNmr Analysis* e *Aria2.3*

### 3.1 Introdução

Durante e após a etapa final de tradução proteica, os polipeptídeos podem ser alvo de diversas enzimas, como quinases e fosfatases, e conseqüentemente, alvos de clivagens e modificações em suas cadeias laterais, um processo conhecido como modificação pós-traducional (1). Os exemplos mais conhecidos são fosforilações (2), metilações (3), glicosilações (4), ubiquitações (5), acetilações (6), amidações (7) e outros. Essas modificações são responsáveis por encaminhar proteínas para localizações intracelulares (8), fixar proteínas em membranas específicas (9) ou estabilizar a conformação tridimensional ou quaternária das proteínas (10). Assim, torna-se indispensável o estudo dessas modificações em biologia estrutural.

A acetilação N-terminal, por exemplo, modifica a carga líquida e a hidrofobicidade proteica, culminando na modificação da vida útil da molécula (11). O processo de acetilação N-terminal é irreversível (12), uma vez que enzimas desacetilases N-terminais não foram descritas, e consiste de adicionar um grupamento acetil ( $\text{CH}_3\text{CO}$ ) ao aminoácido N-terminal livre que contém um grupamento positivo  $\alpha$ -amino ( $\text{NH}_3^+$ ) (13). Esta modificação é extremamente comum em seres humanos, uma vez que aproximadamente 80% das proteínas apresentam o N-terminal acetilado (14). Funcionalmente, o grupamento acetil no N-terminal pode aumentar ou diminuir a vida útil proteica. Neste último caso, por exemplo, o grupamento acetil funcionaria como alvo para poliubiquitinação e conseqüente degradação proteossômica (15). Mas isso é dependente do meio e dos resíduos presentes na porção N-terminal. O grupamento também auxilia no enovelamento de algumas proteínas, como a  $\alpha$ -sinucleína, cuja estruturação em  $\alpha$ -hélice depende da acetil (16). Além disso, a acetilação N-terminal está envolvida com formação de complexos proteicos (17), ligação com membrana (18) e modificação de histonas (19).

Outro exemplo, é a amidação do C-terminal de proteínas. A amidação peptídica é responsável pelo aumento de atividade biológica (20) e está presente em endorfinas (21), neurotensinas (22), endotelina-1 (23) e diversos peptídeos antimicrobianos (24). Assim como no processo de acetilação do N-terminal, a amidação proteica ocorre com enzimas resíduos

específica e em mais de uma etapa, onde a primeira é a clivagem da proteína na região próxima a uma glicina e dois resíduos polares, e depois a clivagem da glicina por outra enzima, a peptidilglicina alfa-amidante monooxigenase (número E.C.: 1.14.17.3), que culmina com o resíduo predecessor sendo o resíduo terminal e amidado (20). Funcionalmente, a amidação auxilia na estabilização estrutural, aumento das pontes de hidrogênio e aumento da carga positiva total.

Como descrito, tanto a acetilação N-terminal quanto a amidação C-terminal tem papel na estabilização proteica. A presença dessas modificações em proteínas podem ser sutis, e assim indetectáveis pelos métodos de validação, ou causar alterações consideráveis nas estruturas proteicas, sendo o ponto chave no enovelamento ou na ausência de enovelamento de porções da molécula. Levando em consideração que mudanças estruturais podem ocorrer sob modificações pós-traducionais, é relevante considerar essas modificações durante o assinalamento e cálculo de estrutura. Entretanto nem todas as plataformas estão adequadas para adição destas modificações, como é o caso do *CcpNmr Analysis* (assinalamento e derivação de restrições) (25) e *Aria2.3* (26) acoplado ao *CNS* (27) (cálculo de estruturas), quando usadas em conjunto. *CcpNmr Analysis* é uma plataforma de assinalamento de experimentos de RMN, derivação de restrições e validação de estruturas calculadas em outros programas. Oferece portabilidade com o *Cyana* e o *Aria*, além de ferramentas de conversão de dados para outros programas. O programa *Aria* teve sua primeira ideia em um trabalho de Nilges e colaboradores em 1997, que implementaram um fluxo de trabalho para interpretar sinais de RMN ambíguos e usou o *X-PLOR* para o cálculo da estrutura (28). Em 2001, o grupo lançou como um programa para NOEs ambíguos e cálculo estrutural, integrado ao *CNS*, denominado *Aria v1.0*. *Aria* recebeu várias atualizações ao longo dos anos e é uma das principais opções acadêmicas gratuitas para cálculo estrutural atualmente (26,29).

Neste trabalho foi desenvolvido um protocolo para adição de modificações acetil N-terminal e amida C-terminal em moléculas protéicas. Assim, o protocolo permite assinalar essas modificações no espectro, gerar suas restrições para serem consideradas durante o cálculo e durante o processo de refinamento em água ou dimetilsulfóxido (DMSO). O protocolo foi desenvolvido em plataforma linux e é uma ferramenta simples, que possibilita que essas modificações não sejam ignoradas no cálculo, o que culminaria em alterações nas qualidades estruturais.



## 3.2 Protocolos

### 3.2.1 Protocolos modificados no *CcpNmr Analysis*

A principal estratégia para incluir as modificações (grupos acetil e amida) foi considerá-las como resíduos. O Aria possui algumas configurações para acetilação e amidação como modificação dos resíduos, mas os mesmos são incompatíveis com as configurações do *CcpNmr Analysis*.

No *CcpNmr Analysis*, os novos resíduos foram modificados na pasta *molecule/ChemComp*. Os arquivos *others\_acy* e *others\_nh3* foram alterados para incluir um código *ccp\_3\_letter*, igual aos arquivos Aria *atomnames* e *iupac*. As quatro primeiras linhas do arquivo *other+Acy+ChemBuild\_tjs23\_2011-03-14-23-20-23-515\_00001.xml*:

```
<?xml version="1.0" encoding="UTF-8"?>
<_StorageUnit release="2.0.b3" originator="CCPN Python XmlIO"
packageGuid="www.ccpn.ac.uk_Fogh_2006-08-16-14:22:51_00046" time="Mon Mar
14 23:20:26 2011">

<CHEM.NonStdChemComp _ID="_1" ccpCode="Acy" code3Letter="ACY"
createdBy="tjs23" guid="ChemBuild_tjs23_2011-03-14-23-20-23-515_00001"
molType="other">
```

As quatro primeiras linhas do arquivo com as informações do *patch* Nh3, *other+Nh3+ChemBuild\_tjs23\_2011-03-14-23-20-04-832\_00001.xml*:

```
<?xml version="1.0" encoding="UTF-8"?>
<_StorageUnit release="2.0.b3" originator="CCPN Python XmlIO"
packageGuid="www.ccpn.ac.uk_Fogh_2006-08-16-14:22:51_00046"
time="Mon Mar 14 23:20:08 2011">

<CHEM.NonStdChemComp _ID="_1" ccpCode="Nh3" code3Letter="NH2"
createdBy="tjs23"
guid="ChemBuild_tjs23_2011-03-14-23-20-04-832_00001"
molType="other">
```

### 3.2.2 Protocolos modificados no Aria2.3

Nesse sentido, os “novos resíduos” foram incluídos nos arquivos *atomnames.xml* e *iupac.xml*, das pastas *Aria2.3*, com a mesma nomenclatura de átomos encontrada no arquivo dessas moléculas da CCPN. No arquivo *atomnames.xml*, os átomos de cada resíduo foram adicionados no modo *iupac*, *cns* e *dyana*, e em diferentes possibilidades atômicas, como visto abaixo:

```
<residue residue_type="AMINO_ACID" iupac_name="NH2"
cns_name="NH2" dyana_name="Nh3">
  <atom iupac_name="H" cns_name="H" dyana_name=""/>
  <atom iupac_name="H'" cns_name="H" dyana_name=""/>
  <atom iupac_name="H1" cns_name="H" dyana_name=""/>
  <atom iupac_name="H2" cns_name="H" dyana_name=""/>
  <atom iupac_name="NT" cns_name="N" dyana_name=""/>
  <atom iupac_name="CA" cns_name="CA" dyana_name="CA"/>
  <atom iupac_name="O" cns_name="O" dyana_name="O"/>
  <atom iupac_name="C" cns_name="C" dyana_name="C"/>
  <atom iupac_name="N" cns_name="N" dyana_name=""/>
</residue>
<residue residue_type="AMINO_ACID" iupac_name="ACY"
cns_name="ACY" dyana_name="Acy">
  <atom iupac_name="H" cns_name="H" dyana_name=""/>
  <atom iupac_name="H21" cns_name="H" dyana_name=""/>
  <atom iupac_name="H22" cns_name="H" dyana_name=""/>
  <atom iupac_name="H23" cns_name="H'" dyana_name=""/>
  <atom iupac_name="N" cns_name="N" dyana_name=""/>
  <atom iupac_name="CA" cns_name="CA" dyana_name="CA"/>
  <atom iupac_name="O" cns_name="O" dyana_name="O"/>
  <atom iupac_name="C1" cns_name="C" dyana_name="C"/>
  <atom iupac_name="C2" cns_name="C" dyana_name="C"/>
  <atom iupac_name="N" cns_name="N" dyana_name=""/>
</residue>
```

O arquivo *iupac.xml* contém o nome atômico e o tipo atômico. Além do mais, fornece informações sobre cadeia principal e cadeia lateral. Como são adaptações de grupos químicos para resíduos, algumas informações foram colocadas em posições estratégicas. Elas não foram removidas para não gerar incompatibilidade nos arquivos. Assim, com o grupo amídico, o átomo de nitrogênio ficou como cadeia principal e os átomos de hidrogênio como cadeias

laterais. No caso do grupo acetil, os átomos de carbono e oxigênio como cadeia principal e os três hidrogênio como cadeia lateral. Nada disso interfere em posições e coordenadas: apenas auxilia na troca de informações entre *Aria2.3* e *CcpNmr Analysis*.

```
<topology description="Amino acid definitions following the IUPAC
convention" name="IUPAC">
  <residue residue_type="AMINO_ACID" name="NH2">
    <atom name="C" atom_type="C" hetero="None"/>
    <atom name="CA" atom_type="C" hetero="None"/>
    <atom name="O" atom_type="O" hetero="None"/>
    <atom name="NT" atom_type="N" hetero="None"/>
    <atom name="N" atom_type="N" hetero="None"/>
    <atom name="H" atom_type="H" hetero="None"/>
    <atom name="H'" atom_type="H" hetero="None"/>
    <atom name="H1" atom_type="H" hetero="None"/>
    <atom name="H2" atom_type="H" hetero="None"/>
    <backbone>
      <atom_name name="N"/>
    </backbone>
    <sidechain>
      <atom_name name="H"/>
      <atom_name name="H'"/>
    </sidechain>
    <terminus terminus_type="N_STANDARD">
      <atom_name name="H1"/>
    </terminus>
    <terminus terminus_type="N_AMINYL">
      <atom_name name="H1"/>
    </terminus>
    <terminus terminus_type="C_STANDARD">
      <atom_name name="H1"/>
    </terminus>
    <terminus terminus_type="C_AMIDO">
      <atom_name name="H1"/>
    </terminus>
    <terminus terminus_type="C_CARBOXYL">
      <atom_name name="H1"/>
    </terminus>
    <equivalent_group group_type="METHYLENE">
      <atom_name name="H1"/>
    </equivalent_group>
  </residue>
</topology>
```

```
</equivalent_group>
<equivalent_group group_type="NTERMINUS">
  <atom_name name="H1"/>
</equivalent_group>
</residue>
<residue residue_type="AMINO_ACID" name="ACY">
  <atom name="C1" atom_type="C" hetero="None"/>
  <atom name="C2" atom_type="C" hetero="None"/>
  <atom name="O" atom_type="O" hetero="None"/>
  <atom name="N" atom_type="N" hetero="None"/>
  <atom name="H" atom_type="H" hetero="None"/>
  <atom name="H21" atom_type="H" hetero="C2"/>
  <atom name="H22" atom_type="H" hetero="C2"/>
  <atom name="H23" atom_type="H" hetero="C2"/>
  <backbone>
    <atom_name name="C1"/>
    <atom_name name="O"/>
  </backbone>
  <sidechain>
    <atom_name name="H21"/>
    <atom_name name="H22"/>
    <atom_name name="H23"/>
  </sidechain>
  <terminus terminus_type="N_STANDARD">
    <atom_name name="H21"/>
  </terminus>
  <terminus terminus_type="N_AMINYL">
    <atom_name name="H21"/>
  </terminus>
  <terminus terminus_type="C_STANDARD">
    <atom_name name="H21"/>
  </terminus>
  <terminus terminus_type="C_AMIDO">
    <atom_name name="H21"/>
  </terminus>
  <terminus terminus_type="C_CARBOXYL">
    <atom_name name="H21"/>
  </terminus>
  <equivalent_group group_type="METHYLENE">
    <atom_name name="H21"/>
  </equivalent_group>
```

```

<equivalent_group group_type="NTERMINUS">
  <atom_name name="H21"/>
</equivalent_group>
</residue>

```

O arquivo *AminoAcid.py* foi modificado para incluir a acetilação e a amidação. O código de uma letra para ambos foi definido como “X”, como já visto nas estruturas depositadas no PDB. O código de três letras usado foi “NH2” e “ACY” e o nome do “resíduo” como “amidation” e “acetylation”, respectivamente. Como o nome no arquivo *atomnames.xml* contém as formas “NH3” e “NH2” e o *ChemComp* do grupo amídico tem o código de três letras “NH2”, não há problemas na trocas de nomes de “NH3” para “NH2”.

```

one2all = {'A': ('A', 'ALA', 'alanine'),
           'R': ('R', 'ARG', 'arginine'),
           'N': ('N', 'ASN', 'asparagine'),
           'D': ('D', 'ASP', 'aspartic acid'),
           'C': ('C', 'CYS', 'cysteine'),
           'Q': ('Q', 'GLN', 'glutamine'),
           'E': ('E', 'GLU', 'glutamic acid'),
           'G': ('G', 'GLY', 'glycine'),
           'H': ('H', 'HIS', 'histidine'),
           'I': ('I', 'ILE', 'isoleucine'),
           'L': ('L', 'LEU', 'leucine'),
           'K': ('K', 'LYS', 'lysine'),
           'M': ('M', 'MET', 'methionine'),
           'F': ('F', 'PHE', 'phenylalanine'),
           'P': ('P', 'PRO', 'proline'),
           'S': ('S', 'SER', 'serine'),
           'T': ('T', 'THR', 'threonine'),
           'W': ('W', 'TRP', 'tryptophan'),
           'Y': ('Y', 'TYR', 'tyrosine'),
           'V': ('V', 'VAL', 'valine'),
           'X': ('X', 'NH2', 'amidation')}
           'X': ('X', 'ACY', 'acetylation')}
three2all = {'ALA': ('A', 'ALA', 'alanine'),
             'ARG': ('R', 'ARG', 'arginine'),
             'ASN': ('N', 'ASN', 'asparagine'),

```

```
'ASP': ('D', 'ASP', 'aspartic acid'),
'CYS': ('C', 'CYS', 'cysteine'),
'GLN': ('Q', 'GLN', 'glutamine'),
'GLU': ('E', 'GLU', 'glutamic acid'),
'GLY': ('G', 'GLY', 'glycine'),
'HIS': ('H', 'HIS', 'histidine'),
'ILE': ('I', 'ILE', 'isoleucine'),
'LEU': ('L', 'LEU', 'leucine'),
'LYS': ('K', 'LYS', 'lysine'),
'MET': ('M', 'MET', 'methionine'),
'PHE': ('F', 'PHE', 'phenylalanine'),
'PRO': ('P', 'PRO', 'proline'),
'SER': ('S', 'SER', 'serine'),
'THR': ('T', 'THR', 'threonine'),
'TRP': ('W', 'TRP', 'tryptophan'),
'TYR': ('Y', 'TYR', 'tyrosine'),
'VAL': ('V', 'VAL', 'Valine'),
'NH2': ('X', 'NH2', 'Amidation'),
'ACY': ('X', 'ACY', 'Acetylation')}]
```

As principais alterações estão nos arquivos do programa CNS, que trabalha junto com o *Aria2.3* no cálculo de estruturas, para inclusão de arquivos de topologias, ligações e parâmetros. Os arquivos de topologia contêm os “novos resíduos” e as ligações das moléculas.

No arquivo *topallhdg5.3.pro*, o resíduo ACE foi alterado para ACY e junto as nomenclaturas dos átomos e além disso, o resíduo NH2 foi incluído. As cargas atômicas do acetil não foram alteradas. As cargas do grupo amídico são as mesmas do desse grupo presente nos resíduos GLN, ASN e ARG. As informações presentes nesta etapa são: átomos, tipo, carga, ligações e diedros impróprios e doadores de ligação. Os nomes atômicos correspondem àqueles presentes no nome dos grupamentos do *CcpNmr Analysis*.

```
residue ACY
  group
    atom C2  type=CH3E  charge= 0.000 end
    atom H21 type=HA    charge= 0.000 end
    atom H22 type=HA    charge= 0.000 end
```

```

atom H23 type=HA charge= 0.000 end
atom C1 type=C charge= 0.500 end
atom O type=O charge=-0.500 end

bond C1 C2 bond C2 H21 bond C2 H22 bond C2 H23
bond C1 O

improper H21 H22 C1 H23 !methyl C2

ACCE O C1
end

residue NH2
group
atom N type=NH2 charge=-0.850 end
atom H type=H charge= 0.425 end
atom "H'" type=H charge= 0.425 end

bond N H bond N "H'"

DONO H N
end

```

Neste mesmo arquivo, as informações sobre o “presidue PEPT”, um *patch* estrutural, foi incorporado com as novas ligações e ângulos a serem incluídos na cadeia principal. Entre elas estão as informações destacadas abaixo, que incluem uma ligação entre carbono e nitrogênio e informações angulares envolvendo um carbono C-terminal do último resíduo peptídico e os átomos do grupo amídico. Informações do grupo acetil não precisaram ser adicionadas.

```

presidue PEPT ! PEPTide bond link, for all except the *(-) -
(+)PRO link
! "*(-) - PEPT - (+)*:

add bond -C +N
add bond -C1 +N

add angle -CA -C +N
add angle -O -C +N
add angle -C +N +H

```

```

add angle -C +N +"H'"
add angle -C +N +HN
add angle -C +N +CA

add angle -C1 +N +CA
add angle -C1 +N +HN
add angle -O -C1 +N
add angle -C2 -C1 +N

add improper -C -CA +N -O ! planar -C
add improper +N -C +CA +HN ! planar +N
add improper -CA -C +N +CA ! angle across peptide plane

! phi/psi related topology UCL 05-JUL-01

add dihedral -C +N +CA +C mult 6
add dihedral -C +N +CA +CB mult 6
add dihedral -CB -CA -C +N mult 6
add dihedral -CB -CA -C +O mult 6
end

```

O arquivo de parâmetro, *parallhdg5.3.pro*, inclui os novos átomos, ligações, ângulos, diedros e diedros impróprios. Esses parâmetros foram reciclados de ligações e átomos com as mesmas características dos grupos amidas e acetila. O mesmo foi considerado na avaliação dos diedros e dos diedros impróprios. As informações são tamanhos de ligações, em angstrom, valores angulares, valores de ângulos diedros e impróprios. Esses valores foram acrescentados, pois não estavam presentes com os nomes atômicos especificados abaixo:

```

BOND C CH3E 1000.000 {sd= 0.001} 1.521

ANGLe C CH3E HA 500.00 {sd= 0.031} 108.9914
ANGLe CH3E C O 500.00 {sd= 0.031} 120.9106
ANGLe CH1E C NH2 500.00 {sd= 0.031} 120.0000
ANGLe CH3E C NH1 500.00 {sd= 0.031} 110.3844

IMPRoper HA C HA HA 500.00 {sd= 0.031} 0
-66.5934
IMPRoper C CH1E NH2 O 500.00 0 0.0000 !

DIHEdral CH1E CH1E C NH2 MULT 6 0.5 1 0.0000 0.4 2
0.0000 0.3 3 0.0000 0.2 4 0.0000 0.1 5 0.0000 0.05 6

```



```
0.0000 !(ile,thr,val)-aaa
DIHEdral CH2E   CH1E   C       NH2   MULT 6   0.5   1 0.0000 0.4   2
0.0000 0.3   3 0.0000 0.2   4 0.0000 0.1   5 0.0000 0.05  6
0.0000 !(ile,thr,val)-aaa
```

O arquivo *topallhdg5.3.pep* contém informações sobre como serão dispostas o N-terminal e o C-terminal nas estruturas. Para isso, uma informação para adição do grupo acetil e uma informação para adição do grupo amídico foram adicionados no modo comentário. Deste modo, caso não haja alterações, o peptídeo tem um N-terminal amídico e um C-terminal carboxilado. Caso o *patch* de modificação seja executado, essas informações são seletivamente inseridas (saem do *status* de comentário).

```
!first NTER           tail + *   end
!last CTER head - *   end
```

A figura 3 resume os arquivos modificados, as suas extensões e onde estão localizados. Os caminhos presentes nos instaladores podem ser modificados de acordo com os locais que os programas foram instalados. O arquivo *./run\_patches*:

```
#!/bin/bash
#Parameters that will be modficated in any conditions:
topdir=${1:-.}

modx=$2

cp peptide_modification/topallhdg5.3.pro "$topdir"/run1/cns/toppar/
cp peptide_modification/parallhdg5.3.pro "$topdir"/run1/cns/toppar/

#=====

#N-terminal Acetylation
if [[ $modx == "N" ]]; then
    cp peptide_modification/n_term_acet/topallhdg5.3.pep
"$topdir"/run1/cns/toppar/
```

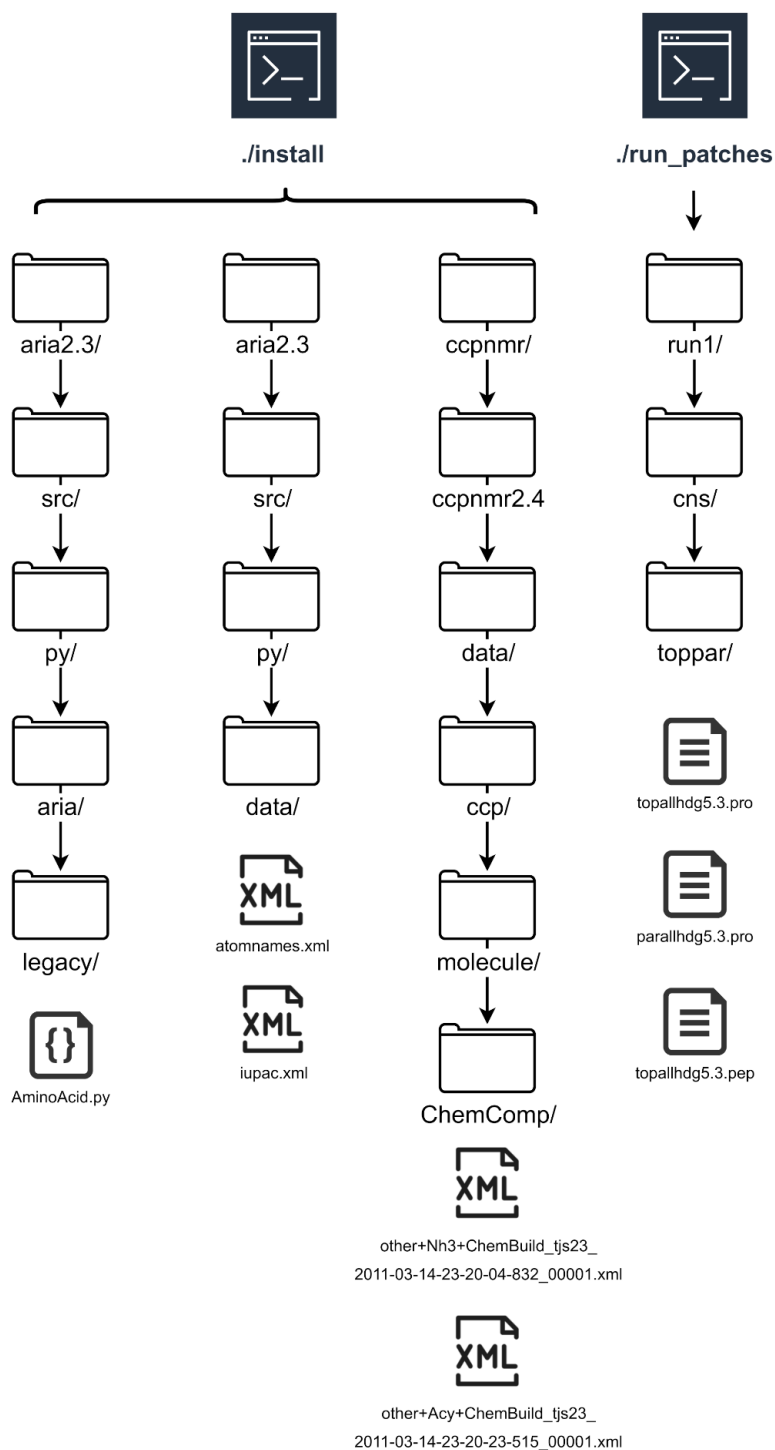
```
fi

#C-terminal Amidation
if [[ $modx == "C" ]]; then
    cp peptide_modification/c_term_amid/topallhdg5.3.pep
"$topdir"/run1/cns/toppar/
fi

#N-terminal and C-terminal modifications
if [[ $modx == "NC" ]]; then
    cp peptide_modification/n_term_c_term/topallhdg5.3.pep
"$topdir"/run1/cns/toppar/
```

### 3.3 Procedimentos

O primeiro passo para adicionar as modificações N-terminal e C-terminal é inseri-las na molécula dentro de um projeto CcpNmr (Figura 1). Em "adicionar sequência", o usuário deve colocar sua sequência de resíduos de aminoácidos e na porção N-terminal o termo "Acy" e/ou na porção C-terminal, o termo "Nh3" (Figura 2a). Essas extensões vêm diretamente no CcpNmr e nada precisa ser instalado ou modificado nesta etapa. No navegador de átomos, os átomos de hidrogênio estão disponíveis para o grupo acetil como "H2 \*" e para o grupo amídico como, "H" e "H'" (Figura 2b). O processo de assinalamento de sinais consiste em avaliar os sistemas de spin de cada resíduo em experimentos de correlações intra residuais (por exemplo, COSY, TOCSY) (30), as correlações entre resíduos (por exemplo, NOESY) (31) e correlações heteronucleares ( $^{13}\text{C}/^{15}\text{N}$  HSQC) (32). A plataforma CcpNmr possui o pacote DANGLE que permite o cálculo dos ângulos  $\varphi$  e  $\psi$  (33). Na aba "Estrutura", o usuário tem a opção "Fazer restrições de distância", que converte a intensidade/volume dos NOEs em restrições de distância de átomos.



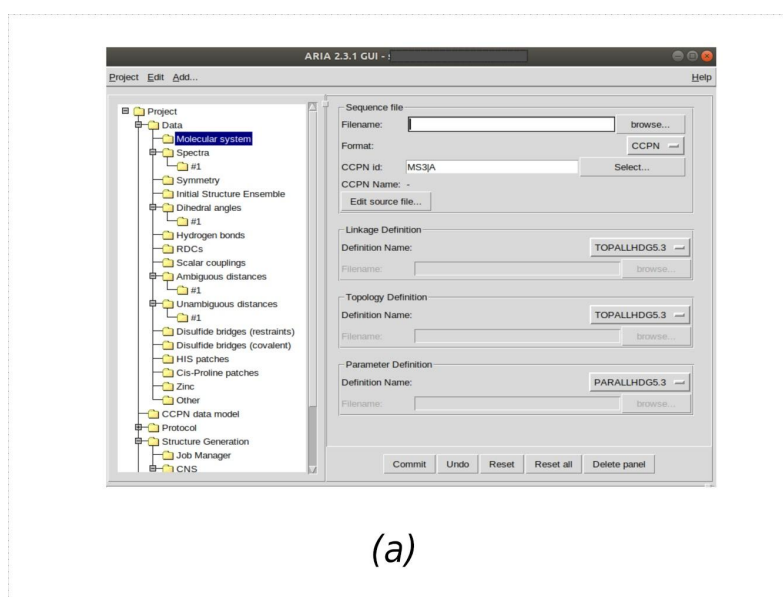
**Figura 1 - Fluxograma dos arquivos modificados para amidação e acetilação de proteínas com *CcpNmr Analysis* e *Aria2.3*.** Os comandos utilizados estão expostos no topo, os caminhos representados pelas pastas e os arquivos modificados representados na porção inferior, em diferentes extensões (.py, .xml, .pro e .pep).



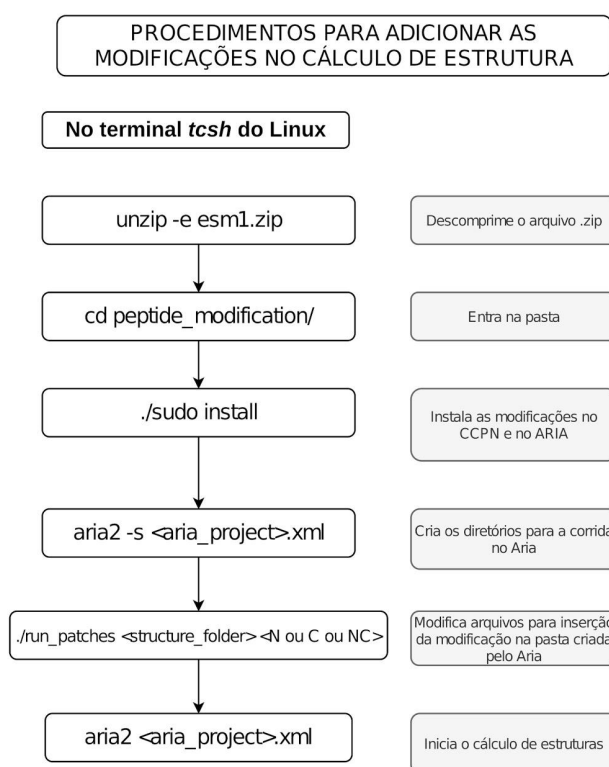
**Figura 2 - Esquema para uso de modificações terminais no CcpNmr Analysis.** (a) Adição dos grupos acetil e amida na sequência proteica: *molecule* > *add sequence* > [Acy] ou [Nh3]. (b) Representação atômica para assinalamento dos hidrogênios dos grupos acetil (H2\*) e amídico (H $\alpha$  e H $\beta$ ).

Um novo projeto Aria é criado sob o comando “aria2 -g” no terminal da linha de comando. Na aba *CCPN data model* deve ser inserido o caminho do projeto CcpNmr. Na guia *CNS*, o caminho do arquivo executável CNS. Com o projeto CcpNmr escolhido, a lista é reconhecida pelo Aria. O usuário é então capaz de inserir a lista de deslocamento químico, sistema molecular, espectro, ângulos diedros e distâncias ambíguas e não ambíguas diretamente do projeto CcpNmr. Em relação às modificações de proteínas, o usuário deve inserir na aba *Molecular system*: a) Definição de ligação: TOPPALLHDG5.3, b) Definição de topologia: TOPPALLHDG5.3, c) Definição dos parâmetros: PARALLHDG5.3, como ilustrado (Figura 3a). Esses são os arquivos modificados para adicionar a acetilação N-terminal e a amidação C-terminal. Também é importante criar uma nova pasta onde ocorrerão os cálculos: na aba “Projeto”, insira o novo caminho da pasta em “diretório de trabalho” e “caminho temporário”. Dessa forma, o patch de modificação funcionará sem alterações adicionais.

Para inserção das modificações, um arquivo zip com os arquivos comprimidos foi criado e é preciso descompactá-lo na mesma pasta onde se encontra o arquivo Aria (arquivo .xml). O usuário deve executar o arquivo de instalação, criar as pastas onde ocorrerá os cálculos, executar o *patch* de modificação e iniciar o cálculo, como demonstrado no fluxograma (Figura 3b).



(a)



**Figura 3 - Esquema para uso de modificações terminais no Aria2.3.** a) As versões 5.3 dos arquivos de ligação, topologia e parâmetros selecionados na aba *Molecular system* da interface gráfica do ARIA. b) fluxograma dos comandos para instalação e modificação dos arquivos a serem usados no cálculo de estrutura.

#### 4.4 Considerações finais

Após modificação de alguns protocolos presentes nos programas *CcpNmr Analysis*, *Aria2.3* e *CNS*, tornou-se viável a adição dos grupamentos acetil e amida nas terminações N-terminal e C-terminal, respectivamente. Devido a importância dessas modificações nas estruturas, que impactam a qualidade das estruturas resolvidas finais, e sua importância biológica, envolvida diretamente com o mecanismo de ação dessas proteínas, a opção de adicionar essas nas porções terminais em programas de assinalamento e cálculo de estrutura são fundamentais. O resultado final é um produto executável que modifica arquivos e pastas, que é também portátil e computacionalmente leve.

## 4.5 Referências

1. Walsh C. Posttranslational Modification of Proteins: Expanding Nature's Inventory. Roberts and Company Publishers; 2006. 524 p.
2. Ubersax JA, Ferrell Jr JE. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol.* 2007 Jul;8(7):530–41.
3. Biggar KK, Li SS-C. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat Rev Mol Cell Biol.* 2015 Jan;16(1):5–17.
4. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol.* 2020 Dec;21(12):729–49.
5. Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol.* 2005 Aug;6(8):599–609.
6. Narita T, Weinert BT, Choudhary C. Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol.* 2019 Mar;20(3):156–74.
7. Eipper BA, Mains RE. Peptide  $\alpha$ -Amidation. *Annual Review of Physiology.* 1988;50(1):333–44.
8. Eisenhaber B, Eisenhaber F. Posttranslational modifications and subcellular localization signals: indicators of sequence regions without inherent 3D structure? *Curr Protein Pept Sci.* 2007 Apr;8(2):197–203.
9. Ahearn I, Zhou M, Philips MR. Posttranslational Modifications of RAS Proteins. *Cold Spring Harb Perspect Med.* 2018 Nov;8(11):a031484.
10. Müller MM. Post-Translational Modifications of Protein Backbones: Unique Functions, Mechanisms, and Challenges. *Biochemistry.* 2018 Jan 16;57(2):177–85.
11. Ree R, Varland S, Arnesen T. Spotlight on protein N-terminal acetylation. *Exp Mol Med.* 2018 Jul;50(7):1–13.
12. Varland S, Osberg C, Arnesen T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics.* 2015 Jul;15(14):2385–401.
13. Pestana A, Pitot HC. N-terminal Acetylation of Histone-like Nascent Peptides on Rat Liver Polyribosomes in vitro. *Nature.* 1974 Jan;247(5438):200–2.
14. Arnesen T, Van Damme P, Plevoda B, Helsens K, Evjenth R, Colaert N, et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci U S A.* 2009 May 19;106(20):8157–62.
15. Lee K-E, Heo J-E, Kim J-M, Hwang C-S. N-Terminal Acetylation-Targeted N-End Rule Proteolytic System: The Ac/N-End Rule Pathway. *Mol Cells.* 2016 Mar;39(3):169–78.
16. Iyer A, Roeters SJ, Schilderink N, Hommersom B, Heeren RMA, Woutersen S, et al. The Impact of N-terminal Acetylation of  $\alpha$ -Synuclein on Phospholipid Membrane Binding and Fibril Structure. *J Biol Chem.* 2016 Sep 30;291(40):21110–22.
17. Shemorry A, Hwang C-S, Varshavsky A. Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Molecular Cell.* 2013 May 23;50(4):540–51.
18. Runfola M, De Simone A, Vendruscolo M, Dobson CM, Fusco G. The N-terminal Acetylation of  $\alpha$ -Synuclein Changes the Affinity for Lipid Membranes but not the Structural Properties of the Bound State. *Sci Rep.* 2020 Jan 14;10(1):204.
19. Demetriadou C, Koufaris C, Kirmizis A. Histone N-alpha terminal modifications: genome regulation at the tip of the tail. *Epigenetics & Chromatin.* 2020 Jul 17;13(1):29.
20. Merkler DJ. C-terminal amidated peptides: production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzyme Microb Technol.* 1994 Jun;16(6):450–6.
21. Eipper BA, Park L, Keutmann HT, Mains RE. Amidation of joining peptide, a major pro-ACTH/endorphin-derived product peptide. *Journal of Biological Chemistry.* 1986 Jul

- 5;261(19):8686–94.
22. Folkers K, Chang D, Humphries J, Carraway R, Leeman SE, Bowers CY. Synthesis and activities of neurotensin, and its acid and amide analogs: possible natural occurrence of [Gln4]-neurotensin. *Proc Natl Acad Sci U S A*. 1976 Nov;73(11):3833–7.
  23. Nakajima K, Kumagaye S, Nishio H, Kuroda H, Watanabe TX, Kobayashi Y, et al. Synthesis of endothelin-1 analogues, endothelin-3, and sarafotoxin S6b: structure-activity relationships. *J Cardiovasc Pharmacol*. 1989;13 Suppl 5:S8-12; discussion S18.
  24. Mura M, Wang J, Zhou Y, Pinna M, Zvelindovsky AV, Dennison SR, et al. The effect of amidation on the behaviour of antimicrobial peptides. *Eur Biophys J*. 2016 Apr;45(3):195–207.
  25. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*. 2005;59(4):687–96.
  26. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*. 2007 Feb 1;23(3):381–2.
  27. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*. 1998 Sep 1;54(Pt 5):905–21.
  28. Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol*. 1997 Jun 13;269(3):408–22.
  29. Linge JP, O'Donoghue SI, Nilges M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol*. 2001;339:71–90.
  30. Shaka AJ, Lee CJ, Pines A. Iterative schemes for bilinear operators; application to spin decoupling. *Journal of Magnetic Resonance (1969)*. 1988 Apr 1;77(2):274–93.
  31. Kumar A, Ernst RR, Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochemical and Biophysical Research Communications*. 1980 Jul 16;95(1):1–6.
  32. Willker W, Leibfritz D, Kerssebaum R, Bermel W. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magnetic Resonance in Chemistry*. 1993;31(3):287–92.
  33. Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson*. 2010 Feb;202(2):223–33.



## CAPÍTULO 4. Conclusão Geral

### 4.1 Considerações Finais

A geração de novas biotecnologias é uma etapa essencial no desenvolvimento de moléculas antimicrobianas. Os antibióticos clássicos, desenvolvidos no último século, são ineficazes contra diversos patógenos multidroga resistentes e extensivamente resistentes à drogas, como no caso de algumas cepas de *Mycobacterium tuberculosis*, em que o patógeno é resistente a drogas convencionais e drogas adicionais do tratamento (1). Atualmente, os antibióticos e as vacinas são as principais estratégias no combate a microrganismos bacterianos. Entretanto, apesar do investimento de grandes companhias, o desenvolvimento de novos antibióticos com novos mecanismos contra bactérias gram-negativas, por exemplo, não teve sucesso nos últimos 40 anos (2). Assim, os principais problemas das áreas de antimicrobianos são a resistência dos patógenos às drogas estabelecidas, o tratamento empírico de pacientes e profissionais da saúde, sem a devida identificação do patógeno, além da ausência de novas classes de antibióticos.

Uma das dificuldades durante o desenvolvimento de novas drogas está na composição variada das membranas de vírus, bactérias, fungos e parasitos, que faz necessário o uso de moléculas com propriedades físico-químicas específicas para esses patógenos e suas cepas distintas (3). Estudos de bioinformática e quimioinformática são parte de estratégias para a racionalização dessas novas drogas, que combina modelos teóricos, preditores e ensaios antibacterianos para testar e analisar a atividade das moléculas, como o método de análise de acúmulo de antibióticos no envelope bacteriano (3), construção de modelos que correlacionam a estrutura de peptídeos e atividade (QSAR) (4), e preditores que usam algoritmos de aprendizado de máquina e redes neurais para identificar sequências a partir de atributos, como propriedades físico químicas e estruturas (5,6). Ainda assim, melhorias nessas ferramentas precisam considerar, além da atividade antimicrobiana, o metabolismo da droga, a farmacocinética e a citotoxicidade.

Neste estudo, foi criado um fluxograma de análise que inclui a identificação de criptídeos, com o teste de atividade antimicrobiana e caracterização de estrutura e afinidade peptídeo-membrana. O algoritmo enCrypted usa códigos para ler sequências proteicas e quebrar em peptídeos, por faixas de massa molecular ou por simulação de atividade enzimática, gerando quebras nas sequências em pontos específicos. O resultado de saída

dessas proteólises *in silico* são usados para o cálculo de propriedades físico-químicas, que são usados como atributos para predição de atividade antimicrobiana. Esse modelo de identificação de criptídeos e teste de atividade antimicrobiana e ensaios de afinidade por membranas permitiu responder nossa hipótese principal, de que a sequência do C-terminal de SmKI-1 possui sequências membrano-ativas e é fonte de criptídeos antimicrobianos. Esses resultados nos permitem conjecturar que talvez o domínio C-terminal seja de fato o domínio responsável pela adesão da proteína SmKI-1 no tegumento de *S. mansoni* e um evento de degradação dessa sequência liberaria esses criptídeos, com ação antimicrobiana escondida na sequência de uma proteína com outra função.

A estratégia de minimização das schistocinas foi um método eficaz para pensar em modos de reduzir custo de produção, avaliar se a atividade se manteria ao se manter um centro ativo dos peptídeos e entender melhor sobre a composição resídica dos criptídeos. Com os experimentos de troca H/D de RMN foi possível observar que a face hidrofílica dos peptídeos reduz a velocidade de inserção dos peptídeos schis-1, schis-2 e schis-3 em micelas, e que o principal fator que prejudica essa inserção desses peptídeos é o resíduo de lisina na posição 6, cuja ausência na schis-3.1 permitiu uma maior inserção do peptídeo na membrana. Ainda assim, schis-3 e schis-3.1 possuem atividade antimicrobiana em níveis similares e os ensaios de termodinâmica comprovaram que em condições favoráveis de temperatura, todas as schistocinas interagem com membranas zwitteriônicas e aniônicas. Essas diferenças propõem que as schistocinas possuem graus variados de afinidade por membranas e suas propriedades físico-químicas distintas podem ser aproveitadas para avaliar a atividade em outros microrganismos, não avaliados neste estudo.

Por fim, a relevância de modificações pós-traducionais, como a amidação C-terminal, para a estabilidade dos peptídeos e de suas estruturas, foi um problema durante o cálculo de estrutura, devido a ausência dos patches para os programas de CcpNMR Analysis v2.5 e Aria 2.3. Desenvolvemos um patch que contorna esse problema, ao considerar as modificações durante o assinalamento da proteína, durante o cálculo e refinamento no ARIA e CNS, e durante a avaliação de estruturas e violações no CCPNMR Analysis. Com isso, espera-se contribuir com o desenvolvimento dessas plataformas que são distribuídas livremente para a academia e auxiliar outros espectroscopistas de RMN de peptídeos e proteínas.

Novas etapas podem ser incluídas nesse fluxograma de identificação de criptídeos para torná-lo mais eficaz, como a predição de citotoxicidade e sua avaliação *in vitro*, predição de

outras atividades biológicas e adição de novas camadas de algoritmos, que permitam gerar predições mais acuradas e, talvez, direcionadas a diferentes cepas de microrganismos. Como relatado em estudos de biofarmacêutica, ferramentas que incluam o estudo de farmacocinética e biodisponibilidade desses peptídeos, podem ser os elementos essenciais para ter uma plataforma robusta de identificação de criptídeos.

## 4.2 Referências

1. Seung KJ, Keshavjee S, Rich ML. Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. *Cold Spring Harb Perspect Med*. 2015 Sep;5(9):a017863.
2. Baker SJ, Payne DJ, Rappuoli R, Gregorio ED. Technologies to address antimicrobial resistance. *PNAS*. 2018 Dec 18;115(51):12887–95.
3. Richter MF, Drown BS, Riley AP, Garcia A, Shirai T, Svec RL, et al. Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature*. 2017 May;545(7654):299–304.
4. Haney EF, Brito-Sánchez Y, Trimble MJ, Mansour SC, Cherkasov A, Hancock REW. Computer-aided Discovery of Peptides that Specifically Attack Bacterial Biofilms. *Sci Rep*. 2018 Jan 30;8(1):1871.
5. Bhadra P, Yan J, Li J, Fong S, Siu SWI. AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Sci Rep*. 2018 Jan 26;8(1):1697.
6. Lin Y, Cai Y, Liu J, Lin C, Liu X. An advanced approach to identify antimicrobial peptides and their function types for penaeus through machine learning strategies. *BMC Bioinformatics*. 2019 Jun 10;20(Suppl 8):291.







## Schistocins: Novel antimicrobial peptides encrypted in the *Schistosoma mansoni* Kunitz Inhibitor SmKI-1

B.P.O. Santos<sup>a,b</sup>, E.S.F. Alves<sup>c</sup>, C.S. Ferreira<sup>d</sup>, A. Ferreira-Silva<sup>e</sup>, A. Góes-Neto<sup>b,e</sup>, R.M. Verly<sup>d</sup>, L.M. Lião<sup>c</sup>, S.C. Oliveira<sup>b,f,g,1</sup>, M.T.Q. de Magalhães<sup>a,b,\*,1</sup>

<sup>a</sup> Laboratório de Biofísica de Macromoléculas, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil

<sup>b</sup> Programa Interunidades de Pós-graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil

<sup>c</sup> Laboratório de Ressonância Magnética Nuclear, Instituto de Química, Universidade Federal de Goiás, Goiânia, Goiás 74690-900, Brazil

<sup>d</sup> Laboratório de Síntese e Estrutura de Biomoléculas, Departamento de Química, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, Minas Gerais 39100-000, Brazil

<sup>e</sup> Laboratório de Biologia Molecular e Computacional de Fungos, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil

<sup>f</sup> Laboratório de Imunologia de Doenças Infecciosas, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil

<sup>g</sup> Instituto Nacional de Ciência e Tecnologia em Doenças Tropicais (INCT-DT), Conselho Nacional de Desenvolvimento Científico e Tecnológico, Ministério da Ciência Tecnologia e Inovação Salvador, Bahia, Brazil

### ARTICLE INFO

#### Keywords:

Cryptides  
Peptide conformation  
Peptide design  
Nuclear magnetic resonance (NMR)

### ABSTRACT

**Background:** Here we describe a new class of cryptides (peptides encrypted within a larger protein) with antimicrobial properties, named schistocins, derived from SmKI-1, a key protein in *Schistosoma mansoni* survival. This is a multi-functional protein with biotechnological potential usage as a therapeutic molecule in inflammatory diseases and to control schistosomiasis.

**Methods:** We used our algorithm enCrypted, to perform an *in silico* proteolysis of SmKI-1 and a screening for potential antimicrobial activity. The selected peptides were chemically synthesized, tested *in vitro* and evaluated by both structural (CD, NMR) and biophysical (ITC) studies to access their structure-function relationship.

**Results:** EnCrypted was capable of predicting AMPs in SmKI-1. Our biophysical analyses described a membrane-induced conformational change from random coil-to- $\alpha$ -helix and a peptide-membrane equilibrium for all schistocins. Our structural data allowed us to suggest a well-known mode of peptide-membrane interaction in which electrostatic attraction between the cationic peptides and anionic membranes results in the bilayer disordering. Moreover, the NMR H/D exchange data with the higher entropic contribution observed for the peptide-membrane interaction showed that schistocins have different orientations upon the membrane.

**Conclusions:** This work demonstrate the robustness for using the physicochemical features of predicted peptides in the identification of new bioactive cryptides. Besides, it demonstrates the relevance of combining these analyses with biophysical methods to understand the peptide-membrane affinity and improve further algorithms.

**Abbreviations:** 2D, two-dimensional; AMPs, antimicrobial peptides; APD, antimicrobial peptide database; ATCC, American type culture collection; CD, circular dichroism; CFU, colony forming units; D<sub>2</sub>O, deuterium; DPC, dodecylphosphocholine; FID, free induction decay; FN, False Negative; FP, False Positive; HSQC, heteronuclear single quantum coherence; ITC, isothermal titration calorimetry; KD, equilibrium dissociation constant; LUV, large unilamellar vesicles; MALDI-TOF/MS, matrix-assisted laser desorption/ionization-time of flight/mass spectrometry; MCC, Matthews correlation coefficient; MIC, minimum inhibitory concentration; NCCLS, national committee for clinical and laboratory standards; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPG, 2-oleoyl-1-palmitoyl-sn-glycero-3-phospho-rac-(1-glycerol); RF, random forest; RP-HPLC, reverse phase-high performance liquid chromatography; SDS, sodium dodecyl sulphate; SVM, support vector machine; TFE, 2,2,2-trifluoroethanol; TN, True Negative; TOCSY, total correlation spectroscopy; TP, True Positive;  $\Delta G$ , Change in Gibbs free energy;  $\Delta H$ , enthalpy change;  $\Delta S$ , entropic change.

\* Corresponding author at: Avenida Antônio Carlos, 6627, Instituto de Ciências Biológicas, Dep. Bioquímica e Imunologia, Laboratório de Biofísica de Macromoléculas, Bloco O4, Sala 125, Belo Horizonte, Minas Gerais, Brazil.

E-mail address: [mquezado@icb.ufmg.br](mailto:mquezado@icb.ufmg.br) (M.T.Q. de Magalhães).

<sup>1</sup> Both authors share senior authorship.

<https://doi.org/10.1016/j.bbagen.2021.129989>

Received 31 May 2021; Received in revised form 30 July 2021; Accepted 6 August 2021

Available online 10 August 2021

0304-4165/© 2021 Elsevier B.V. All rights reserved.

*General significance:* Bioprospecting cryptides can be conducted through data mining of protein databases demonstrating the success of our strategy. The peptides-based agents derived from SmKI-1 might have high impact for system-biology and biotechnology.

## 1. Introduction

Antimicrobial peptides (AMPs) are widely distributed in nature and have broad spectrum of activity (antibiotic, anti-inflammatory, immunomodulatory) [1]. Most known AMPs are cationic molecules, a property that culminates in great affinity with negatively charged membranes. Their interaction with membranes can be through: barrel-stave, in which peptides interaction creates pores in the membrane (eg. pardaxins) [2,3]; detergent-like, forming micelles with the membrane lipids (eg. magainins and MSI-78) [4,5]; carpet model, in which the membrane is disrupted but peptides do not enter the cell (eg. LL-37) [6]; or toroidal, with pores formation without peptides interaction [7–10]. The knowledge of those processes culminated in the study of thousands of antimicrobial peptides in the past few years and allowed researchers to classify them according to their physicochemical properties, such as secondary structural content, hydrophobicity, net charge, and structure upon interaction with a membrane [11,12].

Most of synthetic antimicrobial peptides are designed as cationic molecules with hydrophilic and hydrophobic interfaces to allow the interaction with the phospholipids heads and tails, assuring a good peptide-membrane interaction. However, this is not state of art, and the balance between different physicochemical properties need to be evaluated such as: toxicity to eukaryotic cells of extremely charged peptides [13,14]; difficulties in administration of hydrophobic peptides and mixing these features can also culminate in inefficient AMPs [15]. This reinforces the necessity of studying AMPs features and the relevance of peptide design. Another factor to take into account is the peptide shortening strategy, which aims to improve the peptide activity by only keeping in the sequence and removing residues that could hinder the peptide's best performance. Shortening can be accomplished by removing N-terminal or C-terminal residues and the approaches take in consideration the peptide amphipathicity, charge and other physicochemical features [16–18]. Residue substitution can also be the key in peptide design since identifying strategic regions for modification are fundamental in peptide improvement.

Cryptides describe peptide fragments encrypted within a larger protein, displaying dissimilar biological functions from the parental, originary primary structures [20]. According to this idea, proteins may have multiple functions and activities, depending on enzymatic processing, metabolic target or relative concentration. A few relevant examples of cryptides are the hemoglobin-derived peptides that display several biological activities distinct from those well-described to its precursor [20–23], the antimicrobial peptides (AMPs) from the apolipoprotein-B, r(P)ApoBL/r(P)ApoBs [24], the lactoferrin-derived ones [25], the cryptides found in the C-terminal domain hemocyanins, a multi-subunit oxygen carrier proteins found in the hemolymph of arthropods and molluscs [26–28], and those from the human thrombin, such as GKY25 [29]. These peptides are generated either after protein maturation and metabolic processes [30,31] or during *in vitro* assays [32,33]. Most of these peptides have been overlooked and maybe this occurs because they are only active after proteolytic action of enzymes, environmental pH change or can be converted inside large sets of proteomic data [34,35]. The diversity of encrypted biologically active peptides in known proteins can be exploited by several theoretical and computational models [36–38]. This might lead researchers to get the most out of the already known protein sequences, looking into their multifunctionality, and reshaping their role in nature, especially those with unknown domains. In addition, bioactive peptides with potential therapeutic applications (e.g inflammation, cancer) can also be discovered [39].

Several approaches applied to the peptide design field have already been proposed to retrieve cryptides from natural molecules, and, usually, they consider the primary structure properties and the three-dimensional features [35,40–42]. The physicochemical properties of peptides influence their behavior and affinity to their targets [43]. The structure factor includes not only the peptide folding (secondary structure) but also the residues disposition in the structure (amphipathicity), and both will be influenced by the ionic and hydrophobic interactions with the target (e.g.: biological membrane). Atomically, membrane disruption may occur through the activity of AMPs that reside in the polar-apolar interface, forcing the lipid groups to distance from each other [44]; through peptide-peptide oligomerization and changes in the electrostatic surface potential in the lipid bilayer [45]. Also, peptide orientation is a crucial factor in its ability to cause membrane disruption, as seen for the amyloidogenic peptides IAPP, where the rat analogue was more capable of burying deeper in the membrane, while the human analogue was oriented in the surface of the micelle [46]. In the peptide rational design, it is possible to input as many features as desired regarding what may be crucial for a specific biological activity in a peptide [47]. Nonetheless, consolidating and analyzing all the information is not straightforward since it is more than inputting too many features in peptide rational design, as selecting the properties is more relevant.

Some predictors for antimicrobial peptides including iAMPpred [48], AmPEP [49], ClassAMP [50], amPEPpy 1.0 [51], AntiMPmod [52], MAMPs-Pred [53], AntiBP [54] have been frequently reported. Nevertheless, these tools do not take into account the cryptide concept in their algorithms, which impairs the discovery of peptides hidden within a protein. Since the cryptide concept is innovative in exploring proteins and their corresponding domains without known function, we focused on the proteins expressed by the water-borne human parasite, *Schistosoma mansoni*. For the past few years, our research group has been studying the properties of the *S. mansoni* Kunitz Inhibitor SmKI-1 (uni-prot CCD77156) [55,56], a protein with a single Kunitz-type domain, which is also a C-terminus domain with no similarity outside the genera, with a vaccine applicability against *S. cercariae* infection in murine model [57]. The presence of the SmKI-1 in the schistosomula tegument [57,58] lead us to hypothesize that its C-terminal domain has membrane affinity features.

In this work, we developed and validated an algorithm to predict cryptides in protein sequences – enCrypted. Our main goal was to use our tool to uncover the cryptides hidden in the SmKI-1 C-terminal domain and evaluate their potential biotechnological applications as antimicrobial peptides. Our approach retrieved 18 attributes and the use of machine-learning algorithms allowed us to describe a new class of antimicrobial cryptides: the schistocins. In order to validate our hypothesis, we tested their antimicrobial activity and performed biochemical characterization by using circular dichroism, calorimetry and nuclear magnetic resonance to access their structure-function relationship. Our data suggests that our computational and experimental framework was able to uncover new cryptides from an uncharacterized domain from SmKI-1 and it is effective to be applied to many other proteins domains of unknown function.

## 2. Material and methods

### 2.1. enCrypted: *in silico* prediction, selection, and downstream analyses of bioactive cryptides

enCrypted is a software that reads and identifies peptides from

FASTA protein files. A predictor model was built using curated sequences (3245 FASTA sequences) extracted from the APD database [59] and PepBank [60] (Date of Access: 20/06/2020). For the algorithm selection, we performed a grid-search to obtain a hyperparameter space for the best cross validation score. The machine learning classification was performed using scikit-learn with python codes [61]. The selected algorithms were linear Support Vector Machines (SVM), decision tree, and random forest (RF). The linear SVM creates an optimal hyperplane to separate data into classes [62]. The parameters tested were: regularization parameter ('C') [0.001, 0.01, 0.1, 1, 10] and loss function ('loss') [squared hinge, hinge]. The decision tree consists of a binary recursive partitioning classification method composed of nodes, edges and leaf nodes that split data into partitions and possibilities to test the different values [63]. The parameters tested were the criterion function ('criterion') [gini, entropy], the splitter strategy (splitter) [best, random], the maximum depth of the tree ('max\_depth') [3,4,5,7,8,9], the number of features per split ('max\_features') [3,5,8,9,10]. The RF has the same principle of decision trees, but with a large number, which one testing a class prediction [64]. The parameters tested were the number of trees ('n\_estimators') [50, 100, 200], the criterion function ('criterion') ['gini', 'entropy'] and the maximum depth of the tree ('max\_depth') [4,5,6,7,8]. The grid-search was performed 10 fold (Fig. S1). The metrics were evaluated using the Eqs. (1) and (2):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Matthews correlation coefficient} = \frac{TP \times TN - FP \times FN}{\sqrt{((TP+FP) \times (TN+FN)) \times ((TP+FN) \times (TN+FP))}} \quad (2)$$

The TP as true positive, the TN as true negative, the FP as false positive and the FN as the false negative. After selection, the method had its parameters optimized and exported to be used as the predictor model. The C-terminal SmKI-1 sequences were filtered by molecular range between 1400 and 2300 Da and a potential biological activity was predicted for each one of them. At the end, the algorithm generates a peptide list with two distinct groups: (a) antimicrobial activity and (b) non-antimicrobial activity.

*In silico* analyses for the selected sequences were also performed. The positive results obtained were compared with the original C-terminal domain of SmKI-1 sequence, and the visualization was provided by Jalview [65]. A multiple sequence alignment was performed with MUSCLE [66], using sequences from the APD database (similarity), previously selected using NCBI-Blast+ [67,68], and the visualization and pid (percent sequence identity) were provided by MView [69]. The amino acid residues disposition in potential  $\alpha$ -helices were studied through helical wheels generated in the webserver Heliquest [70].

## 2.2. Peptide synthesis, purification and characterization

Four *S. mansoni* – derived antimicrobial peptides (cryptides) were designed and synthesized by solid-phase synthesis, purified by RP-HPLC, and the monoisotopic molecular mass and primary sequences were verified by mass spectrometry (Fig. S3 and Fig. S4).

The peptides schistocin-1 (GILDIKKNVSNLFFKKIKGEK-NH<sub>2</sub>), schistocin-2 (GILDIKKNVSNLFFKKIK-NH<sub>2</sub>), schistocin-3 (GILDIKKNVSNLFF-NH<sub>2</sub>), and schistocin-3.1 (GILDILNKVSNLFF-NH<sub>2</sub>) were obtained by solid-phase peptide synthesis *via* Fmoc (9-fluorenylmethoxycarbonyl) strategy [71]. Fmoc Rink-Amide® resin (0.63 mmol·g<sup>-1</sup>) was used for the synthesis of all the peptides resulting in C-terminal amidated peptides. The amino acid derivatives were provided by Iris Biotech GmbH. Cleavage and final deprotection were conducted with TFA:Triisopropylsilane:Water (95.0:2.5:2.5, v:v:v) solution for 1.5 h at room temperature. Peptide purification was performed through RP-HPLC with a C18 column (Discovery® C18 HPLC Column 5  $\mu$ m particle size, L  $\times$  I.D. 25 cm  $\times$  10 mm) with a linear gradient of water containing 0.1% (v:v) TFA and acetonitrile containing 0.1% TFA (v:v). The peptide

concentrations were determined according to the Murphy & Kies method, by measuring the absorbance of stock solutions at 205, 215, and 225 nm [56]. The identity of the peptides (0.001 mg·mL<sup>-1</sup> in 0.1% TFA) were assessed by MALDI-TOF/MS (Bruker Daltonics Autoflex III) in the positive reflected mode, using  $\alpha$ -cyano-4-hydroxycinnamic acid (5 mg·mL<sup>-1</sup>, in 50% AcN, 0.1% TFA) as an ionization matrix in a proportion of 1:1 (v/v). Ions of interest were fragmented in the LIFT mode (MS/MS). The purity of the synthetic peptides was accessed by RP-C<sub>18</sub>-HPLC (not shown).

## 2.3. Antibacterial assays

Certified American Type Culture Collection (ATCC; USA) strains of *E. coli* (ATCC 25922), *P. aeruginosa* (ATCC 27853) and *Staphylococcus aureus* (ATCC 29213) were provided by the National Institute of Quality Control in Health (INCQS-FIOCRUZ, Brazil), and were cultured at 37 °C on agar plate containing BHI medium (Brain Heart Infusion) from HiMedia (Mumbai, India) for 12 h. Subsequently, the isolated colonies were selected and transferred to a Mueller-Hinton culture medium (Kasvi; Brazil) according to the National Committee for Clinical and Laboratory Standards (NCCLS; USA) in order to perform the bioassays (NCCLS Institute protocols) [72]. The peptide was diluted 10-fold in Mueller-Hinton liquid broth. The highest peptide concentration used was 264  $\mu$ M in a suspension composed of approximately  $5 \times 10^5$  CFU·mL<sup>-1</sup> (colony forming units·mL<sup>-1</sup>), and the final volume was 100  $\mu$ L per well 1:1 (peptide:inoculum). The experiments were performed for 20 h in stationary culture at 37 °C, and the optical density was measured at the beginning and at the end of the assay (A<sub>595</sub> nm). Ampicillin and chloramphenicol (50  $\mu$ M) were used as the standard antibiotics for the control of cell death, and the positive control was achieved by the addition of 50  $\mu$ L of the inoculum and 50  $\mu$ L of the sterile culture medium. All assays were performed in triplicate.

## 2.4. Antifungal assays

Antifungal activity were screened using *C. albicans* (ATCC 10231; resistant to anidulafungin, voriconazole, itraconazole, and fluconazole), *C. krusei* (ATCC 6258), *C. tropicalis* (ATCC 750), *C. parapsilosis* (ATCC 90018) and *Candida glabrata* (ATCC 2001) as provided by the National Institute of Quality Control in Health (INCQS-FIOCRUZ, Brazil). The minimum inhibitory concentration (MIC) was determined by a susceptibility test for microdilution in broth for the determination of MIC to yeasts (CSLI 2012) [73], with modifications. Tests were also performed in Sabouraud dextrose broth (Kasvi, Brazil). Serial dilutions of 20–500  $\mu$ M were prepared from compounds on sterile 96-well microtitration plates. The density for yeasts per well corresponded to  $1 \times 10^3$  to  $5 \times 10^3$  CFU·mL<sup>-1</sup>. The plates were incubated at  $35 \pm 2$  °C for 44 h, and subsequently evaluated by visual reading. The positive control was nystatin (50 to 0.0225  $\mu$ g·mL<sup>-1</sup>). Controls were also prepared for viability assessment of the test microorganisms and sterility assessment of the culture medium. All assays were performed in triplicate.

## 2.5. Vesicle preparation

Large unilamellar vesicles (LUVs) were prepared using 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 2-oleoyl-1-palmitoyl-sn-glycero-3-phospho-rac-(1-glycerol) (POPG) lipids. The appropriate amounts of phospholipids required for the preparation of POPC or POPC:POPG (3:1, mol:mol) LUVs were solubilized in 1.0 mL of chloroform, which was evaporated using nitrogen gas inside the fume hood and the resulting lipid film was resuspended in an aqueous 50 mM Tris-HCl buffer (pH 7.5), containing 100 mM NaCl buffer. In order to form unilamellar vesicles, 8 cycles of freezing and warming the solution was performed: the tube was heated in a water bath (40 °C) and flash-frozen using liquid nitrogen. The vesicles size was regulated by extrusion procedure, where the lipid solution was pushed 11 times between a



mini-extruder and two syringes. The mini-extruder filter contained filters with 0.1  $\mu\text{m}$ . The LUVs diameter was confirmed using the Dynamic Light Scattering analyzer/Malvern® Zetasizer nano ZS particles model BI-900 (Worcestershire, United Kingdom) from the Multi-user Laboratory of Research in Pharmacy (Department of Pharmacy – UFVJM). The hydrodynamic diameter analysis was accessed through the Malvern® model DTS1060 cuvette, using monochromatic light (Ne 4 mW laser,  $\lambda$  633 nm) and was carried out in triplicates. LUVs were maintained at 4 °C and used within 24 h from preparation [74].

## 2.6. Circular dichroism (CD) spectroscopy

The schistocins 1, 2, 3 and 3.1 were evaluated for secondary structure contents by circular dichroism (CD) measurements using a JASCO® J-810 spectropolarimeter (Jasco Inc., Japan), equipped with Peltier temperature control system - PFD-425S (Jasco Inc., Japan). CD spectra were recorded at 20 °C using a 100  $\mu\text{m}$  path length quartz cuvette (Uvonic Instruments, NY), with peptides (30  $\mu\text{M}$ ) in i) 50 mM Tris-HCl pH 7.0 and 100 mM NaCl, ii) TFE:water solutions, iii) presence of 100-500  $\mu\text{M}$  sodium dodecyl sulphate (SDS) micelles, and iv) LUVs. All the spectra were recorded with three scans from 260 to 190 nm using a 1.0 nm spectral bandwidth, 200 nm.min<sup>-1</sup> scan speed, and 1 s response time, and properly corrected for their respective blank measurements performed without peptides. Blank subtraction was performed with Spectra Analysis software (Jasco Inc., Japan). The resulting curves were smoothed without data distortion by using the Savitzky-Golay filter [74] with a polynomial order of 3 and a smoothing window of 51 points. A customized algorithm for deconvolution was used using the following Eq. (3):

$$\text{Ellipticity, in degrees}\cdot\text{cm}^2\cdot\text{dmol}^{-1} = 100 \times \theta\lambda/m \times d \quad (3)$$

The ellipticity calculation was proceeded as previously informed [76]. The K2D2 method was used for spectra deconvolution and average helicity value was extracted [77].

## 2.7. Isothermal titration calorimetry (ITC)

ITC experiments were performed at 35 °C with a high-precision VP®-ITC microcalorimeter (Malvern Instrument Ltd., UK) for 50  $\mu\text{M}$  peptide solutions in 50 mM Tris-HCl, pH 8.5, containing 100 mM NaCl. The ITC equipment was electrically and chemically calibrated before performing the experiments [62]. All the solutions and buffers employed in the experiments were filtered and degassed under reduced pressure (140 mbar) for 5 min. The peptide solutions in the calorimeter cell were titrated with fifty-one successive injections of 20 mM POPC or POPC:POPG (3:1, mol:mol) LUVs (the first 1  $\mu\text{L}$  injection was discarded in order to eliminate diffusion effects of the material from syringe to calorimeter cell, and this injection was followed by fifty injections of 5  $\mu\text{L}$ ). Injection times of 2 s with intervals of 240 s have been used in the experiment. The background experiments were performed titrating the buffer solution with the respective LUVs suspensions to determine the corresponding heats of dilution. In order to measure the heat produced by the peptide-to-lipid interactions, each peak of the calorimetric curve was integrated following standard procedures as described elsewhere [78]. The total lipid concentration was used to estimate the degree of membrane association as well as in the determination of the thermodynamic binding parameters. The raw data were analyzed with Microcal Origin 5.0 (OriginLab Corporation, MA).

## 2.8. Nuclear magnetic resonance spectroscopy (NMR)

The NMR measurements were performed at 1 mM peptide in a micellar solution containing 100 mM of deuterated dodecylphosphocholine (DPC-*d*<sub>38</sub>), 10% D<sub>2</sub>O, 1.0 mM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS; internal reference) and 50 mM sodium phosphate pH

3.5. Two-dimensional (2D) NMR spectra of schistocin-1, 2, 3 and 3.1 were recorded on a Bruker AVANCE III 500 NMR spectrometer operating at 500.013 MHz equipped with a 5 mm broadband inverse (BBI) probe head (Chemistry Institute, Federal University of Goiás - UFG, Brazil) at 300 K, and pre-saturation for proper water suppression.

Total correlation spectroscopy (TOCSY) spectra were acquired using the DIPSI-2 pulse sequence [75,79]. The water suppression was performed using the pulse sequence dipsi2esgpph (BRUKER nomenclature), using a spin-lock of 80 ms. The spectral width was 9615 Hz, and 512 t<sub>1</sub> increments were collected with 16 transients of 4096 points. Nuclear Overhauser Spectroscopy (NOESY) spectra were acquired using a pulse sequence with watergate suppression (noesygpphw5) and mixing times of 150 and 200 ms [80]. The spectral width was 9615 Hz, 512 t<sub>1</sub> increments were collected with 16 transients of 4096 points for each FID. <sup>1</sup>H-<sup>13</sup>C Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired in edited mode with phase sensitive multiplicity-edited HSQC using echo-antiecho sequence (hsqcedetgp) with F1 and F2 spectral widths of 72,010 Hz and 9615 Hz, respectively [81]. A total of 128 t<sub>1</sub> increments were collected with 128 transients of 1024 points. The HSQC experiments allowed the user to observe methyl groups in different ways: CH and CH<sub>3</sub> correlations show positive phase and CH<sub>2</sub> correlations show negative phase.

## 2.9. NMR data analysis and structure calculations

All the NMR data were processed using TopSpin® software 4.1 (academic license, Bruker, Germany). Two-dimensional <sup>1</sup>H resonance assignments were performed by simultaneously analyzing 2D <sup>1</sup>H-<sup>1</sup>H TOCSY and NOESY spectra [82], using the CCPNMR analysis software [83]. NOE intensities obtained at mixing times of 200 ms were converted into semi-quantitative distance restraints [84], resulting in upper limits of distances of 2.8, 3.4 and 8.0 Å (for strong, medium, and weak NOEs, respectively). The dihedral angle ( $\varphi$  (phi) and  $\psi$  (psi)) restraints have been obtained from analysis of C<sub>α</sub>, H<sub>α</sub>, C<sub>β</sub> and H<sub>β</sub> chemical shifts with the DANGLE software, part of CCPNMR platform [85]. Structure calculations were performed using the Aria2.3 software, coupled with CNS [86,87]. Starting with an extended conformation, 200 structures were generated using a simulated annealing protocol, followed by 20,000 steps of simulated annealing at 1000 K and a subsequent decrease in temperature in 15,000 steps in the first slow-cool annealing stage. The twenty lowest energy structures were refined by using a sa\_new.inp protocol and the stereochemical quality was analyzed by MolProbity [88], Procheck [89] and Prosa [90]. The display, analysis, potential hydrogen bonds counting and manipulation of the three-dimensional structures were performed with Chimera (UCSF, USA) [91]. The solvation potential energy calculation was measured by converting the PDB files into PQR, that served as input for APBD, a Poisson-Boltzmann Solver calculator [92,93].

## 2.10. Hydrogen-deuterium exchange

Hydrogen exchange data were recorded on a Bruker AVANCE III 500 NMR spectrometer operating at 500.013 MHz (Chemistry Institute, Federal University of Goiás – IQ/UFG, GO, Brazil). Hydrogen-deuterium exchange experiments were initiated by gently reconstituting dry lyophilized peptide (originally 1 mM) in dry DPC-*d*<sub>38</sub> lipids (originally 100 mM) with 95% of D<sub>2</sub>O sufficient to achieve their original concentration. A series of TOCSY spectra were acquired in intervals of 1.33 h to monitor the decay of the HN-H<sub>α</sub> due to exchange for deuterium. All spectra were acquired with water suppression by using dipsi2esgpph pulse sequence with 4096 time domain data size, 512 scans, 16 dummy scans and 32 number of scans, with a spin-lock time of 80 ms. All data were processed using TopSpin 4.1 (Bruker) and analyzed using CCPNMR to assign and to follow the intensity changes. The data exchange kinetic was adjusted for a monoexponential decay using  $I = A \cdot \exp(-B \cdot t)$ , where A and B are the fit parameters and t is time after complete reconstitution

with D<sub>2</sub>O.

### 3. Results

#### 3.1. Identification of new cryptides from SmKI-1 C-terminal domain, their solid-phase synthesis, and characterization

SmKI-1 is located in the schistosomula tegument bound to a plasma membrane and overlain by a secreted membranocalyx (a second membrane). Considering its affinity to membranes as desired feature for novel AMP, we attempted to identify potential bioactive cryptides from the SmKI-1 protein by using enCrypted (details in the *Material and Methods* section). A set of theoretical, potential AMP (183 peptides) and non-AMP (265 peptides) were obtained (Fig. 1A), with the AMP typically starting at the position Ser-21 and ending at Leu-69 (numbers corresponds to full SmKI-1 C-terminal domain amino acid sequence).

In order to select potential new antimicrobial cryptides from SmKI-1, and study their biophysical properties, we selected one out of the three peptides presenting similarity with the greatest number of sequences of another known antimicrobial peptides (Supporting CSV File). Firstly, we chose the theoretical AMP 20-mer peptide found by the C-terminal of SmKI-1, hereby named schistocin-1. Secondly, we used the minimization strategy to shorten the peptide and evaluate its effects caused by changing important features such as charge, molecular weight and hydrophobicity, resulting in the theoretical peptides schistocin-2 (17 mer), and schistocin-3 (13 mer). The net charge in the peptides schistocin-1 and schistocin-2 is the same (+5), but schistocin-3 presents a lower value (+2) (Fig. 1B). Despite the lowest charge, schistocin-3 is the most hydrophobic of the three peptides (0.463) and has similar hydrophobic moment with schistocin-2 (0.551 and 0.579, respectively).

The schistocins exhibited primary structure similarity to previously deposited antimicrobial peptides in the APD database, which were originally identified in amphibians (percentage of identity: 26.3–31.6% uperin, 7.9–10.5% brevinin, 10.8–16.7% ranateurin), and in scorpions (9.1% hadrurin) (Fig. 1C). A helical wheel plot of schistocin-1, schistocin-2, schistocin-3 were obtained to confirm and compare the cryptides amphiphilic nature when in an  $\alpha$ -helical configuration (Fig. S2). After this analysis, we increased the  $\alpha$ -helix amphipathicity (0.701) by replacing Lys6 > Leu in schistocin-3, leading to schistocin-3.1. The quality of the predictor for this analysis can be evaluated through its accuracy value of 87% and its efficacy compared with other predictors reported elsewhere (Table S1). The physical-chemical features used in the model construction and full peptides selection are listed in Table S2.

#### 3.2. Schistocins exhibit different antimicrobial activities against human pathogenic bacteria and yeasts

The schistocins peptides were evaluated for antimicrobial activity against both gram-positive and gram-negative bacterial (*S. aureus* ATCC 29213, *E. coli* ATCC 25922, *P. aeruginosa* ATCC 7853) and yeast strains (*C. albicans* ATCC 10231, *C. krusei* ATCC 6258, *C. tropicalis* ATCC 750, *C. parapsilosis* ATCC 90018, *C. glabrata* ATCC 2001). Schistocin-3 and schistocin-3.1 displayed a similar antibacterial activity when compared to the controls (ampicillin and chloramphenicol) whereas higher MIC values were detected in cryptides schistocin-1 (>130  $\mu$ M) and schistocin-2 (>130  $\mu$ M) against the tested bacterial strains (Table 1). A similar behavior was observed in antifungal assays, except for *C. krusei*, to which all the peptides were ineffective. Overall, schistocin-3 and schistocin-3.1 were more effective either against bacteria or fungi than schistocin-1 and schistocin-2.

#### 3.3. Schistocins conformation depends on environmental hydrophobicity

In order to evaluate whether the resulting enCrypted output provided membrane-active peptides capable of folding into an ordered  $\alpha$ -helix, we performed circular dichroism measurements in the presence

of membrane-mimetic solvents and phospholipid vesicles. The four schistocins adopted a random coil conformation when free in solution, and transitioned into  $\alpha$ -helix-rich structures when in the presence of zwitterionic (a.k.a. amphoteric) or anionic POPC:POPG (3:1) vesicles (Table 2 and Fig. 2). In POPC:POPG vesicles the helix content of schistocin-1, schistocin-2 and schistocin-3 are high (84%), even at mid concentration, while schistocin-3.1 exhibited lower helix content (50% in P/L = 1/8, 76% in P/L = 1/16) (Fig. 2). In addition, in TFE:water mixtures and in the presence of SDS micelles, all the schistocins have a characteristic  $\alpha$ -helix pattern, with two minimums, in 208 nm and 222 nm (Fig. S5). Altogether, these findings demonstrate the self-assembly of schistocins into helix-rich structure upon recognition and binding into hydrophobic environment.

#### 3.4. Interaction and binding affinity of schistocins are favored by negative enthalpy changes

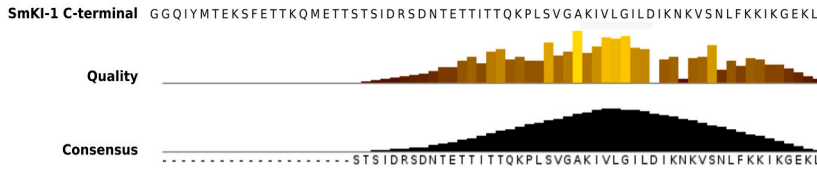
In order to access the coupled energetics of schistocins binding to membrane and peptide folding into helix we conducted isothermal titration calorimetry (ITC). The overall interactions of schistocins with anionic vesicles were observed to be exothermic in nature as suggested by the integrated heat values (Fig. 3). Decreases in the heat flow until values near to the dilution are due to the smaller amount of free peptide in solution at the end of the experiment, which indicates the affinity of schistocins to the POPC:POPG LUVs (Table 3). Exergonic processes take place for the interaction of the peptide with the vesicles, and the highest change in Gibbs free energy was observed for schistocin-3.1 (-6.2 kcal. mol<sup>-1</sup>). Consequently, greater binding constant ( $k_{app}$ ) is observed for schistocin-3.1 when compared with others schistocins. Although, the negative enthalpy values indicate exothermic binding of the peptides to the anionic vesicles and the predominance of attractive coulomb forces, the binding process for all peptides are mainly driven by entropic factors, since  $|\Delta H| < < |\Delta S|$ . Therefore, these data suggest a well-known mode of peptide-membrane interaction in which electrostatic attraction between the cationic peptides and anionic membranes results in the bilayer disordering.

#### 3.5. NMR solution structures of schistocins revealed peptide interaction and orientation upon the mimetic-membrane model

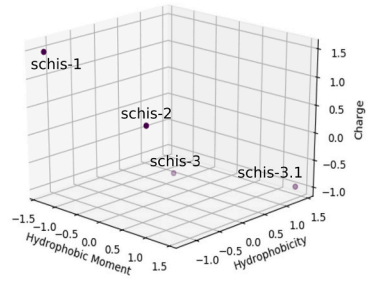
CD spectroscopy and ITC analyses jointly described a membrane-induced conformational change from random coil-to- $\alpha$ -helix and a peptide-membrane equilibrium for schistocins. In order to gain more detailed insights on the structural basis of schistocin interaction with membranes, solution NMR spectra were recorded in the presence of dodecylphosphocholine (DPC-*d*<sub>38</sub>) micelles. As expected, several strong sequential connectivity  $d_{NN}(i,i + 1)$  and medium range  $d_{\alpha N}(i,i + 3)$  correlations characteristic of  $\alpha$ -helix were observed in all the schistocins NOESY spectra. All inter-residual NOE correlations are graphically summarized in Fig. S6.

The NOEs distance constraints in both schistocin-1 and schistocin-2 showed short-range constraints and medium-range constraints comprising residues Ile-2 to Lys-14 for schistocin-1 and Ile-2 to Ile-16 for schistocin-2 (Fig. S6). The calculated structures show an  $\alpha$ -helix in these segments with a random-coil tail in the C-terminal. The ensemble of the lowest-energy structures shows relatively small all residues RMSD value for both schistocin-1 (0.32  $\pm$  0.11) and schistocin-2 (0.10  $\pm$  0.02) (Table 4). The residues Lys-15/Glu-19 and Glu-19/Lys-20 had hydrogen bond interactions responsible to stabilize the C-terminus of schistocin-1, while the Phe-13/Ile-16, Phe-13/Lys-17, Lys-14/Lys-17 hydrogen bonds stabilized the C-terminus of schistocin-2. Schistocin-1 and schistocin-2 have an apolar cluster with aliphatic residues Gly-1, Ile-2, Leu-3, Ile-5, Val-9, Leu-12, Phe-13, Ile-16 (Fig. 4A and Fig. 4B). Furthermore, the electrostatic potential on the surface of the peptides generated a solvation potential energy of 1.16  $\times$  10<sup>4</sup> kJ·mL<sup>-1</sup> for schistocin-1 and 1.33·10<sup>4</sup> kJ·mL<sup>-1</sup> for schistocin-2, demonstrating that they are highly

**A**



**B**



**C**

	cov	pid	1 [ . . . . ] 42
1 Schistocin-1	100.0%	100.0%	G I D I K N K V S N L F K K I K G E K
2 Schistocin-2	85.0%	100.0%	G I D I K N K V S N L F K K I K
3 Schistocin-3	65.0%	100.0%	G I D I K N K V S N L F
4 Schistocin-3.1	65.0%	92.3%	G I D I L N K V S N L F
5 Aurein_2.1	80.0%	43.8%	G L D I V K K W G A F G S L
6 Uperin-2.4	95.0%	31.6%	G I D F A K T V V G G R N A L G I
7 Uperin-2.7	95.0%	31.6%	G I D I A K K I V G G R M L G I
8 Uperin-2.8	95.0%	26.3%	G I D V A K T I V G K I R M L G I
9 Brevinin-2CG1	75.0%	10.3%	G I D K L K E F G S A A R
10 Brevinin-2OA5	75.0%	10.5%	G L D G I I N A F N A A K
11 Brevinin-2HSa	75.0%	10.5%	G L D S L K N L A I N A A K
12 Brevinin-2HSb	75.0%	10.5%	G L D T L K N M A I N A A K
13 Brevinin-2ISa	75.0%	7.9%	S L I D T F K N L A V N A A K
14 Brevinin-2OA3	75.0%	7.9%	G L I D T F K N M A I N A A H
15 Brevinin-2OR2	75.0%	10.5%	G L I D T I K N M A I N A A K
16 Brevinin-2OR6	75.0%	10.5%	G L I D T I K N M A I N A A K
17 Brevinin-2OR3	75.0%	7.9%	G L I C T I K N M A I N A A K
18 Brevinin-2OR7	75.0%	10.5%	G L I D T I K N M A I N A A K
19 Brevinin-2OA1	75.0%	13.2%	G I D T F K N M A I N A A K
20 Brevinin-2OA2	75.0%	10.5%	G L I D T F K N M A I N A A K
21 Brevinin-2OA4	75.0%	7.9%	G L I D T F K N L A I N A A E
22 Odoranain-D1	75.0%	10.5%	G I D T F K N L A I N A A K
23 Brevinin-2GRa	75.0%	10.5%	G L I D T F K N L A I N A A K
24 Ranateurin-2La	75.0%	16.2%	G I D S F K G V A K G V A K
25 Ranateurin-2ARb	55.0%	13.5%	G I D T I K N A A
26 Ranateurin-2SEc	55.0%	10.8%	G I N C T I K C T A
27 Ranateurin-2CPa	55.0%	10.8%	G I N C T I K C T A
28 Ranateurin-2CPb	55.0%	10.8%	G I N C T I K A T A
29 Brevinin-2OA7	75.0%	7.9%	G M I C T V K D L I G A C K
30 Brevinin-2OA6	75.0%	5.3%	G M A T V K N L L I C T C D
31 Brevinin-2OA8	75.0%	7.9%	G M A T V K N L L N E T C D
32 Hadrurin	100.0%	12.2%	G I D C T I K S A S K V W S K T V Q D L K R K G N W V A N K L G V S P Q A A
33 Esculentin-2EM	100.0%	13.5%	G I D T L K G F A K G V K D L V K G A A C G M L S T V S C K L A
34 Rugosin-C	100.0%	13.5%	G I D S F K G F A K G V K D L I K G A A C G M L S T V S C K L A
35 Hylaseptin-P1	70.0%	28.6%	G I D A I K A I A K A A G
36 Brevinin-2ED	70.0%	11.4%	G I D S L K N L A K N A G
37 Brevinin-2EC	80.0%	8.1%	I L I D K L K N F A K T A G K
38 Brevinin-2EA	80.0%	13.5%	G I D T L K N L A I S A A K
39 Brevinin-2EB	80.0%	13.5%	G I D T L K N L A K T A G K
40 Andersonin-D1	75.0%	33.3%	F I F P K K N I N S L I F G R
41 Esculentin-1c	75.0%	0.0%	L L I S G L K N V C K E V G M
42 Esculentin-1B	75.0%	2.6%	L L I S G L K N V C K E V G M
43 Esculentin-1A	75.0%	2.6%	L L I S G L K N V C K E V G M
44 CAP18(106-142)	100.0%	13.6%	K L K K I G Q K I Q F V F K L A P R T D Y
45 RL-37	100.0%	9.1%	G I K K V G Q K I K D F L G N L V P R T A S
consensus/100%			. h h . . h . . h . . . . .
consensus/90%			h l h s h p p h . . . h . . . . .
consensus/80%			G I D h h k p h s . . h t p . . . . .
consensus/70%			G I D s h k s h u h t h u c . . . . . h h . t h l p . . h t C K h t . . . . . C

(caption on next page)



**Fig. 1.** Peptides Identification. Antibacterial prediction was performed using *in house* enCrypted software. Cryptides were generated through simulation of PAM enzyme activity and limiting them between 1400 Da and 2300 Da. A) Alignment of predicted peptides with the C-terminal sequence of SmKI-1. The analysis shows a preference for a fixed region of C-terminal SmKI-1, between residues 22 and 60. B) Physicochemical distribution of schistocins, taking into account hydrophobic moment, hydrophobicity and charge. C) Sequence alignment with AMPs sequences. A total of forty one sequences, in a rank of similarity, from the APD database were compared with schistocins peptides. The MUSCLE tool was used for multiple sequence alignment. The coloured residues are those with a consensus threshold of 70% and identical to the reference sequence, the schistocin-1. Colors: A - bright-green, C - yellow, D - bright-blue, E - bright-blue, F - dark-green, G - bright-green, H - dark-green, I - bright-green, K - bright-red, L - bright-green, M - bright-green, N - purple, P - bright-green, Q - Purple, R - bright-red, S - dull-blue, T - dull-blue, V - bright-green, W - dark-green, Y - dark-green. Color schemes from MView.

**Table 1**

Antibacterial and antifungal activity of schistocins using determination of the Minimum Inhibitory Concentration (MIC). Peptides activity was tested against three bacterial strains and 5 fungi strains. Nystatin, DS01, Ampicillin and Chloramphenicol were used as control drugs.

Microorganisms	MIC( $\mu$ M) <sup>a</sup>				Nystatin	Amp. <sup>c</sup>	Chloram. <sup>c</sup>
	schistocin-1	schistocin-2	schistocin-3	schistocin-3.1			
<i>S. aureus</i> ATCC 29213	>130	>130	31.2	8	NT <sup>b</sup>	<11	>30
<i>E. coli</i> ATCC 25922	>130	>130	62.5	62.5	NT <sup>b</sup>	46	25
<i>P. aeruginosa</i> ATCC 27853	>130	>130	62.5	31.2	NT <sup>b</sup>	20	30
<i>C. albicans</i> ATCC 10231	250	250	62.5	31.2	0.39	NT <sup>b</sup>	NT <sup>b</sup>
<i>C. krusei</i> ATCC 6258	125	62.5	62.5	62.5	1.56	NT <sup>b</sup>	NT <sup>b</sup>
<i>C. tropicalis</i> ATCC 750	500	500	31.2	31.2	0.78	NT <sup>b</sup>	NT <sup>b</sup>
<i>C. parapsilosis</i> ATCC 90018	>500	>500	125	125	0.78	NT <sup>b</sup>	NT <sup>b</sup>
<i>C. glabrata</i> ATCC 2001	>500	>500	250	125	0.09	NT <sup>b</sup>	NT <sup>b</sup>

<sup>a</sup> MIC - Minimal agent concentrations required for total inhibition of cell growth.

<sup>b</sup> NT- Antimicrobial activity not tested.

<sup>c</sup> Amp and Chloram - ampicillin and chloramphenicol.

**Table 2**

Helix percentage of schistocins structures from circular dichroism spectra. Data were fitted using K2D2 online software.

Helix content (%)							
SDS (mM)							
	0	100	200	300	400	500	
schistocin-1	11.33	69.46	69.46	75.57	69.46	69.46	
schistocin-2	8.6	84.27	84.27	84.27	84.27	84.27	
schistocin-3	6.27	47.81	47.93	47.81	47.93	36.77	
schistocin-3.1	56.82	75.57	80.68	69.46	80.68	69.46	
TFE (%)							
	0	10	20	30	40	50	60
schistocin-1	11.33	20.49	69.46	80.68	80.68	69.46	75.57
schistocin-2	8.6	50.16	84.27	84.27	84.27	84.27	84.27
schistocin-3	6.27	10.33	12.98	47.81	47.93	47.81	47.81
schistocin-3.1	56.82	27.25	84.27	84.27	84.27	84.27	84.27
POPC (mM)							
	0			250		500	
schistocin-1	7.96			8.6		75.57	
schistocin-2	8.6			56.82		75.57	
schistocin-3	6.27			7.84		11.27	
schistocin-3.1	6.71			7.96		7.96	
POPC:POPG(3:1) (mM)							
	0			250		500	
schistocin-1	6.27			84.27		84.27	
schistocin-2	8.6			84.27		84.27	
schistocin-3	7.96			84.27		84.27	
schistocin-3.1	6.71			50.16		75.57	

cationic.

Schistocin-1 has a partially amphipathic conformation, impaired by Lys-6 and Gly-18. In addition, the hydrophilic side chains residues were less rigid than the hydrophobic residues (Fig. 4A). On the other hand, for schistocin-3 and schistocin-3.1 (13-mer peptides), the NOEs distance restraints evidenced sequential and medium-range constraints in virtually all backbone from Gly-1 up to Phe-13 (Fig. S6). A high stereochemical quality of the most stable structures obtained for all the peptides was obtained as depicted by the validation analysis (Table 4). The calculated structures show a full  $\alpha$ -helix structure, compatible with the previous solution results by CD. The ensemble of the 10 lowest energy structure show that the N-terminal portion (2-12 amino acids

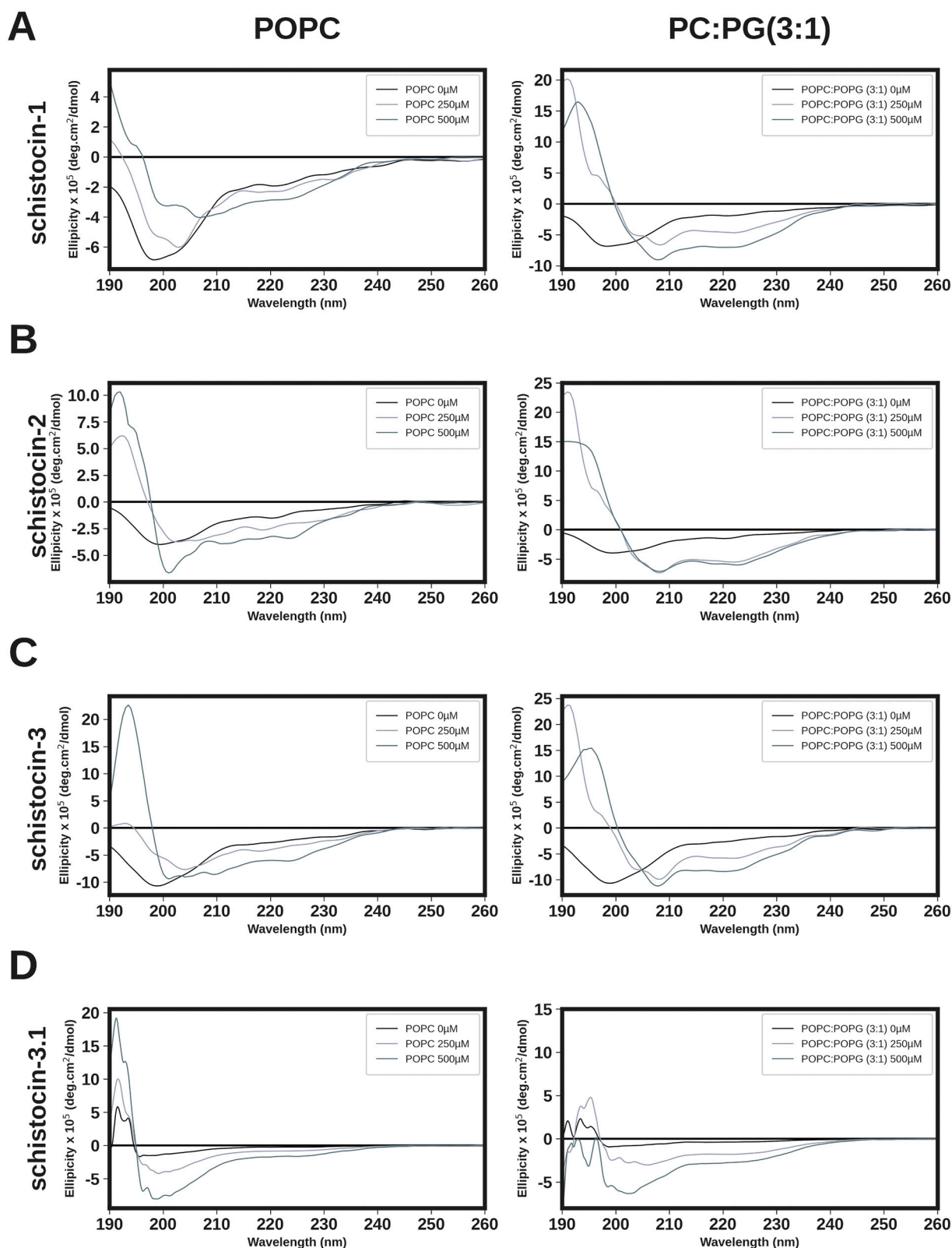
residues) has relatively low conformational flexibility, including the side chains for both peptides (Fig. 4C and Fig. 4D) with RMSD values of  $0.16 \pm 0.05$  schistocin-3 and  $0.11 \pm 0.05$  for schistocin-3.1. As for schistocin-3.1, the substitution of Lys6 > Leu improved the helix amphipathicity. Taken together, these results led us to suggest that the schistocins display a well-defined, stable  $\alpha$ -helix conformation.

Accordingly, NMR H/D exchange experiments allowed us to evaluate the intensities of HN-H $\alpha$  correlation and fit exponential decays to determine the exchange rates of amidic hydrogen from the residues with large decay (Fig. S7). In all schistocins investigated here, most residues exchange occurred at a very fast rate (1.33 h). Nonetheless, in schistocin-1 and schistocin-2, the HN-H $\alpha$  correlations involving Val-9, Ile-12, Phe-13, and Ile-16 were in slow exchange, suggesting that they are located in the micelle interface with low water accessibility. The Lys-6 presented in schistocin-1, schistocin-2, and schistocin-3 rapidly exchanged with deuterium, while in schistocin-3.1 the Leu-6 was in slow exchange indicating a deeper insertion of the N-terminus of this peptide into the micelles. This data together with the higher entropic contribution observed for the peptide-membrane interaction led us to propose a model of peptide insertion into the membrane interface (Fig. 5), suggesting that schistocin-1 and schistocin-3.1 appeared to attach almost in parallel to the membrane surface, while schistocin-2 and schistocin-3 present a slight tilted angle in relation to the membrane surface.

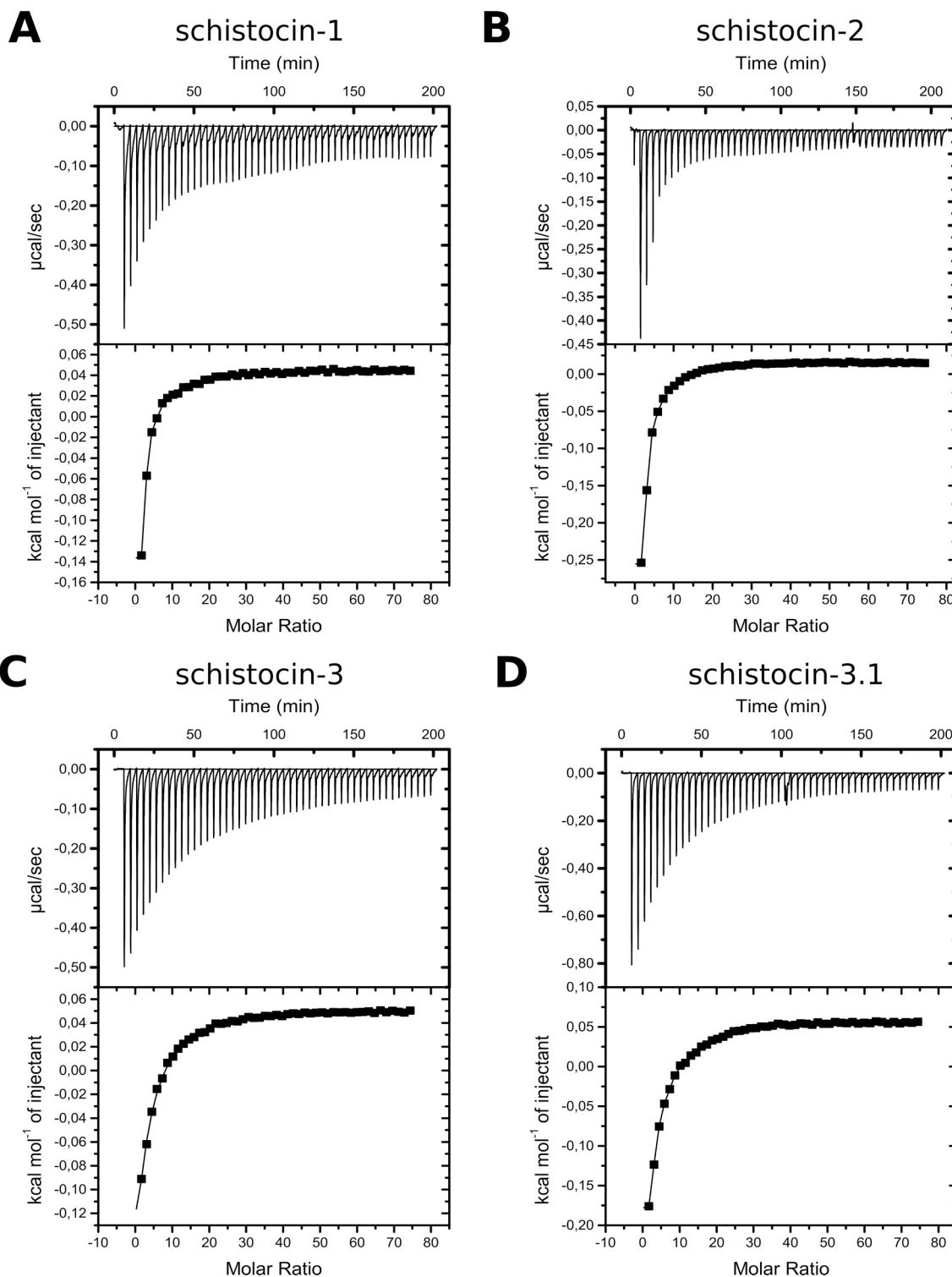
#### 4. Discussion

In this paper, we developed and validated a machine learning algorithm (a random forest model) for the prospection of novel peptides encrypted within protein sequences. From the results obtained with enCrypted, we further designed, produced and evaluated the antimicrobial and biophysical properties of this new family of cryptides in the *Schistosoma mansoni* Kunitz Inhibitor protein, SmKI-1, hereby named schistocins. Our present data demonstrate the robustness for using the physical-chemical features of predicted peptides in the identification of new bioactive peptides encrypted within larger proteins.

Experimental and computational approaches in discovering cryptides have been previously described, and some studies have shown that cryptides can be found by top-down proteomics in tissues followed by bioactivity assays [35], such as the discovery of hemorphin-like peptides



**Fig. 2.** CD spectra of 30  $\mu\text{M}$  of schistocins in POPC and POPC:POPG (3:1) vesicles (A-H). Left column: POPC vesicles ( $P/L = 1/8$ ,  $P/L = 1/16$ ); right column: POPC:POPG (3:1) vesicles ( $P/L = 1/8$ ,  $P/L = 1/16$ ). A) schistocin-1. B) schistocin-2. C) schistocin-3. D) schistocin-3.1. All spectra were recorded in 50 mM Tris-HCl pH 7.0 and 100 mM NaCl. Spectra recorded from 260 nm until 190 nm. The black line represents peptides without vesicles, the gray line represents peptides in  $P/L = 1/8$  peptide, and the blue line represents peptides in  $P/L = 1/16$ . Negative bands at 222 nm and 208 nm are characteristic of alpha-helix peptides, as seen in peptide-POPC:POPG vesicles; negative bands lower at 200 nm and lower are characteristic of random coil peptides, as seen in peptides-POPC vesicles.



**Fig. 3.** POPC:POPG (3:1) vesicles - schistocins isothermal titration calorimetry. Titration calorimetry of schistocins solutions with unilamellar POPC:POPG (3:1) vesicles in buffer (35 °C; 50 mM Tris-HCl, 100 mM NaCl, pH 8.5). A total of 5  $\mu$ l of the lipid dispersion (concentration = 20 mM) was injected into the calorimeter cell containing the schistocin-1 (A) or schistocin-2 (B) or schistocin-3 (C) or schistocin-3.1 (D) (concentration = 50  $\mu$ M) in the same buffer. The upper image of each letter corresponds to the raw data. The lower image corresponds to the integrated heat change per injection with the best fitting curve.

[94], from *P. patens* cryptides [95], and human salivary peptides [96]. Furthermore, several research groups have been developing computational methodologies for the identification of cryptides. For instance, mitocryptides were identified in mitochondrial proteins obtained from the *Swiss-Prot* database, which followed by proteolytic cleavage *in silico*, and the selection of peptides with physical-chemical properties similar

to *mastoparans* [97], such as identification of positive charges and the presence of specific residues [19,30,31]. Another example is the *Kamal* software proposed by Brand et al. (2012), which identifies physico-chemical properties and resumes them as principal component analysis, which are used for bioactive peptides clusterization and allows identifying new antimicrobial peptides located in similar regions. The

**Table 3**

Thermodynamics parameters of ITC experiment of schistocin peptides in anionic LUVs.

Peptides	K	DH Kcal/mol	TDS Kcal.mol <sup>-1</sup>	DG
schistocin-1	4.06E+03	-1.03	4.07	-5.09
schistocin-2	9.34E+03	-0.40	5.17	-5.57
schistocin-3	2.38E+03	-1.14	3.60	-4.74
schistocin-3.1	2.55E+04	-0.11	6.13	-6.23

forementioned software included selection of whole-genome information of a large range of organisms, along with selected AMPs as a positive control [36–38]. The software *PeptideLocator* has the general goal of identifying bioactive peptides in protein sequences, with the bioactive peptide prediction being conducted by algorithms that identify residue positions and structural probability (helix, beta or turn) [98–100]. The *PeptideDB* has a similar framework but with different purpose: it identifies bioactive peptides of food proteins in input sequences, but limits for peptides already present in their database, having no prediction tool [101,102]. The present enCrypted algorithm has a pipeline for protein sequence screening, followed by sequence proteolysis and physical-chemical properties extraction, culminating in a list of peptides, and uses a supervised classification machine-learning algorithm (RF) for antibacterial prediction. The software concept proposes to join tasks that generally are separately performed: (i) *in silico* digestion and (ii) physical-chemical and bioactivity prediction; and opens up a field to expand the bioactivity prediction: antifungal, antiparasitic, immunostimulatory and others, provide that the current databases are curated and available. In addition, we propose a new approach to exploring unknown proteins by taking a closer look at encrypted information.

We hypothesised that SmKI-1 has to be cleaved from the tegument to exhibit an anti-inflammatory activity [57,58], and its other domain could be digested by degradation or proteolysis of the proteins resulting in the release of bioactive fragments in the host circulation. We selected three peptides from a set of 351 and, after a secondary pattern *in silico* analysis, we designed the fourth (schistocin-3.1) with a single substitution to improve amphipathicity (Fig. S2). These peptides were chemically synthesized and *in vitro* tested against bacterial and fungal strains that are pathogenic to humans, and the MICs were comparable to other antimicrobial peptides [103–105]. The peptides schistocin-3 and schistocin-3.1 displayed a wide antimicrobial spectrum, while schistocin-1 and schistocin-2 showed mild activity against *C. krusei* strains. Comparing our findings to other antimicrobial cryptides with known high activity, we can observe that schistocin-3 and schistocin-3.1 are nearly as efficient as apoB cryptides (from apolipoprotein), which have antimicrobial/antibiofilm activity with MIC values of 20  $\mu$ M [24]. They are also similar to rpS30-2, from p62 eukaryotic intracellular protein, with MIC value of 100  $\mu$ M against antimycobacterial activity [106]. In order to gain insights to explain the biological activity divergence among schistocins, the peptides-membrane interactions were investigated by a set of biophysical techniques [107–109]. Biomimetic membranes and micelles are useful to understand the peptides structuring in environments similar to bilayers found in both prokaryotes and eukaryotes, besides simulating the interfacial electrostatics responsible for the peptides' affinity and the analyses of different peptide-membrane stages [110,111].

As many other cationic peptides with antimicrobial activity [105,112–114] when in the presence of membrane mimetic models, schistocins display an  $\alpha$ -helical fold (Fig. 2). These amphipathic peptides with net positive charge adopt  $\alpha$ -helical structure upon contact with hydrophilic/hydrophobic surface, such as the membrane of microorganisms [115]. Examples are the GKY25, an AMP derived from the human thrombin protein [115], and the 23-mer PvhCt peptide from the shrimp *L. vannamei* hemocyanin [25]. The CD results describe schistocins random-to-helix transition from aqueous buffer to (3:1) POPC:POPG LUVs (Fig. 2). The peptides schistocin-1 and schistocin-2

presented weak helix-forming intensity with POPC and moderate helicity with POPC:POPG. Conversely, schistocin-3 and schistocin-3.1 upon membrane titration showed significant peptide-phospholipid helicity, highlighted in anionic vesicles (Fig. 2). In general, thermodynamic studies regarding the membrane interactions of schistocins indicated that these peptide presents from moderated to higher affinity to negatively charged vesicles only ranging from 35 °C and that result in exothermic process. The schistocins interaction at 25 °C did not show any significant binding and no thermodynamic parameters could be obtained (data not shown). The temperature-dependence of peptide-membrane interaction suggest the entropic component is predominant, as a consequence of the hydrophobic effect caused by water desolvation, which occurs during the peptide-membrane interaction. In fact, the major driving forces of interaction between schistocins and phospholipid membranes are the entropic contributions [117]. Interestingly, schistocin-3.1 interaction reveals the highest entropic factor and binding constant, which is in accordance with NMR data that shows better partitioning of the hydrophilic and hydrophobic residues within the bilayer interface. On the other hand, a higher enthalpic contribution observed in the schistocin-1, schistocin-2 and schistocin-3 interaction suggests that the electrostatic attraction may still be an important phenomenon to the membrane binding process of these peptide [118]. Even though the schistocin-3.1 has net charge +1, the schistocin-1, 2 and 3 presents higher cationic hydrophilic surfaces to establish attractive coulomb forces with the negatively charged surface of the POPC:POPG membranes [119,120]. Thus, the higher antimicrobial activity of schistocin-3.1 could be related to the higher membrane affinity due an optimal insertion of the hydrophobic residues, pointing to membrane insertions of N-terminus anchored by Leu-6 residue. Recent studies showed that peptides derived from the C-terminal of SARS CoV E protein fold into an  $\alpha$ -helix, which is correlated to the precursor protein host specificity and its inherent function to bind to the membrane [121]. In a similar way, the  $\alpha$ -helical structure observed for schistocins when upon the membrane may also be related to the potential C-terminal domain adhering function of the kunitz domain in the *S. mansoni* tegument.

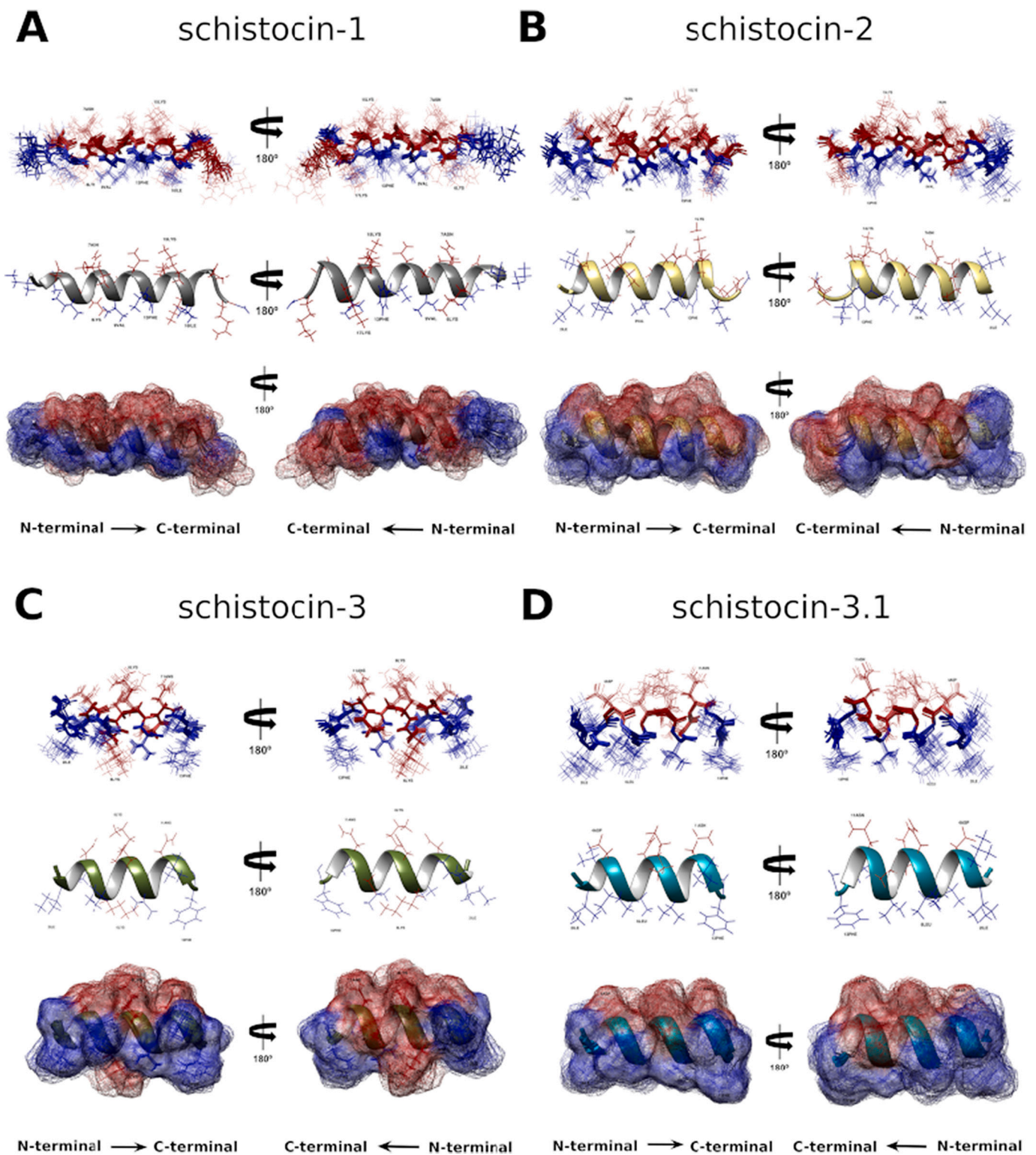
Our NMR data confirmed that the peptides fold into well-defined  $\alpha$ -helix three-dimensional structures in DPC-*d*<sub>38</sub>, and corroborated the hypothesis that the Lys-6 was impairing the peptide amphipathicity, besides giving information on peptide flexibility in micelles (Fig. S5). The peptides schistocin-1 and schistocin-2 have a more flexible C-terminal, while schistocin-3 and schistocin-3.1 have a rigid helix, with a Phe-13 C-terminal side chain in slow exchange, showing its affinity to the membrane and increasing the specificity of these peptides. Moreover, hydrophobic residues are known to be important in peptide stabilization upon the membrane [122,123]. Future studies should consider taking a deeper look on the schistocins orientation in DPC-*d*<sub>38</sub> micelles using paramagnetic probes in solution NMR. A previous study showed that gadolinium is inert with proteins and DPC-*d*<sub>38</sub> micelles. Its addition to the NMR tube results in paramagnetic relaxation enhancements that make it possible to inform the orientation of  $\alpha$ -helical structures [124,125].

Our dataset was constructed with peptides of up to 20 amino acid residues, provided that the peptides used as negative controls were obtained from a database that limited their peptides size. Restricting the AMPs sizes was essential to not construct an imbalanced dataset, leading to prediction errors [126]. An alternative approach could be the generation of peptides from Uniprot databases with a large molecular weight range and use them as the negative control. Still, an extensive curation would be necessary to distinguish the activities and to not impair the bioactivity prediction. Furthermore, our data suggests, as shown in Table 1, that two of the selected peptides did not exhibit strong antimicrobial activity against the studied bacterial and fungal strains. Nonetheless, this is not a drawback because the sequences of schistocin-1 and schistocin-2 are precursors of schistocin-3, which presented antimicrobial activity and helped us to understand which structural differences were responsible for the activity divergence. If taken as an

**Table 4**  
Summary of structural statistics from NOE experiments.

Summary of Structural Statistics																	
	schistocin-1				schistocin-2				schistocin-3				schistocin-3.1				
Distance and dihedral restraints																	
NOEs																	
Intraresidue	101				101				97				100				
Interresidue	71				70				48				76				
Sequential	39				44				25				44				
Short range	28				25				21				28				
Medium range	4				1				2				4				
Long range	0				0				0				0				
Total unambiguous	172				171				145				176				
Ambiguous	45				65				14				29				
Dihedral angles (psi + phi)	34				28				20				20				
Structural statistics																	
Structural precision																	
RMSD Backbone (all/2ndary structure)	0.32 (+/- 0.12) / 0.75 (+/- 0.30)				0.10 (+/- 0.02) / 0.24 (+/- 0.05)				0.16 (+/- 0.05) / 0.37 (+/- 0.05)				0.11 (+/- 0.05) / 0.45 (+/- 0.19)				
RMSD Heavyatom (all/2ndary structure)	0.90 (+/- 0.11) / 1.31 (+/- 0.38)				0.40 (+/- 0.03) / 0.47 (+/- 0.03)				0.60 (+/- 0.09) / 0.63 (+/- 0.11)				0.68 (+/- 0.09) / 0.93 (+/- 0.28)				
	value	error	min	max	value	error	min	max	value	error	min	max	value	error	min	max	
PROCHECK:																	
most favored regions	99.4	1.82	94.1	100	100	0	100	100	98.3	3.41	91.7	100	95	5.67	83.3	100	
allowed regions	0.59	1.82	0	5.9	0	0	0	0	1.66	3.41	0	8.3	4.99	5.67	0	16.7	
generously allowed regions	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
disallowed regions	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
labelled residues (all Ramachandrans)	0	0	0	0	0	0	0	0	0.1	0.308	0	1	0.25	0.55	0	2	
labelled residues (Chi1-chi2 plots)	1.2	0.523	1	3	0.65	0.489	0	1	0.15	0.366	0	1	0	0	0	0	
bad contacts	0.05	0.224	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
G-factor dihedrals	-0.145	0.578	-0.24	0	-0.124	0.055	-0.2	-0.01	0.131	0.0759	-0.01	0.27	0.154	0.123	-0.12	0.31	
G-factor covalent	0.609	0.671	0.59	0.62	0.528	0.048	0.47	0.6	0.604	0.0131	0.58	0.63	0.641	0.008	0.62	0.65	
G-factor overall	0.135	0.352	0.08	0.22	0.118	0.044	0.06	0.2	0.307	0.0456	0.22	0.39	0.339	0.073	0.17	0.43	
WHATCHECK:																	
Ramachandran plot appearance Z-score (RAMCHK)	-6.63	1.35	-9.73	-4.69	-4.27	1.28	-6.32	-2.39	-1.84	1.18	-3.66	0.751	-1.12	2.23	-4.62	1.76	
Chi-1 chi-2 rotamer normality Z-score (C12CHK)	-7.28	1.21	-9.88	-4.92	-9.7	1.09	-11.4	-7.65	-8.08	0.978	-9.54	-5.84	-7.19	1.7	-10.1	-4.79	
Backbone conformation Z-score (BBCCHK)	-1.84	0.597	-3.62	-1.38	0.382	0.0789	0.288	0.445	0.305	0	0.305	0.305	0.303	0	0.303	0.303	
Bond lengths RMS Z-score (BNDCHK)	0.3	0.128	0.278	0.325	0.382	0.0272	0.323	0.419	0.312	0.0137	0.284	0.335	0.282	0.011	0.261	0.307	
Bond angles RMS Z-score (ANGCHK)	0.395	0.168	0.367	0.43	0.492	0.0757	0.377	0.578	0.325	0.02	0.302	0.361	0.271	0.034	0.261	0.383	
Omega angle restraints RMS Z-score (OMECHK)	0.734	0.707	0.606	0.877	0.648	0.0601	0.563	0.733	0.455	0.0548	0.329	0.549	0.447	0.057	0.323	0.549	
Side chain planarity RMS Z-score (PLNCHK)	0.301	0.82	0.18	0.536	0.306	0.12	0.103	0.566	0.384	0.125	0.117	0.56	0.248	0.111	0.075	0.52	
Improper dihedral distribution RMS Z-score (HNDCHK)	0.51	0.194	0.471	0.549	0.499	0.0492	0.418	0.579	0.468	0.0245	0.423	0.505	0.458	0.034	0.41	0.562	
MOLPROBITY:																	
Clashscore	17.7	4.6	11.6	29	43.1	9.23	26.4	59.2	18	3.47	13.7	22.8	8.8	5.6	0	18.5	
Clashscore percentile	41.4	14.3	16	64	7.5	5.38	2	19	39.4	11.6	25	55	75.3	22.1	36	100	
Clashscore Z-score	-0.195	0.365	-0.947	0.365	-1.48	0.336	-1.97	-0.809	-0.242	0.28	-0.605	0.125	nan	nan	-0.306	inf	



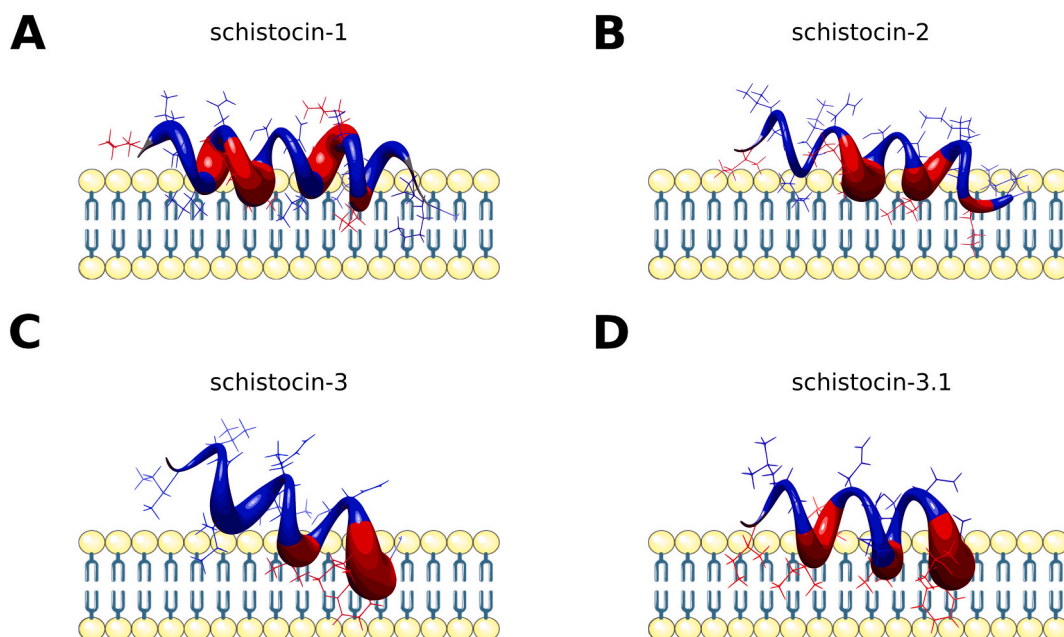


**Fig. 4.** Solution NMR structure of 400  $\mu\text{M}$  in 20 mM DPC- $d_{38}$  micelles. (A) schistocin-1, (B) schistocin-2, (C) schistocin-3, (D) schistocin-3.1. All peptide (400  $\mu\text{M}$ ) structures were obtained in Bruker Avance III 500 (11.75 T) in DPC- $d_{38}$  micelles (20 mM). First, the ten lowest energy structures (atoms superposition representation), evidencing the hydrophobic residues, in blue, and the hydrophilic residues, in red. Second, the lowest energy structure (ribbon representation). Third, the lowest energy structure (surface representation), with the hydrophobic regions, in blue, and the hydrophilic regions, in red.

algorithm impediment, this could be solved by adding new layers to the bioactive peptide prediction, as proposed by Lin et al. (2019), as well as other aspects in peptide predictions, such as position-specific residue preference [49,53]. This study robustness is the approach of combining the computational effort and the experimental verification, which led us

to a great understanding of cryptides identification.

To our knowledge, this is the first report on SmKI-1 cryptides with antimicrobial activity. This work strengthens several others that came across cryptides derived from pathways responsible for protein degradation or other ones with many functions, often completely different



**Fig. 5.** Amide hydrogen exchange shown in volume rate for each residue of the schistocins. Residues with low exchange rate are shown in higher volume, while the residues with low volume are those with fast exchange during the hydrogen-deuterium TOCSY experiments. Cartoon representation in blue: residues with fast exchange. Cartoon representation in red: residues with slow exchange.

from those of the parent protein [24,31,127]. Additionally, it brings a new functionality for SmKI-1 C-terminal domain by using the concept of cryptides. We demonstrated that one could look into protein sequences differently and get more information about proteins' function. This multifaceted protein has more to come and to be learned, and most likely many other antimicrobial peptides described in the literature, further design is still necessary until converging into potential drug candidates. Therefore, we described a new tool to hunt for new peptides and uncover proteins' hidden functions. In addition, we interestingly observed the affinity of the peptides originated from the membrane domain to interact with vesicles only at a temperature similar to that in the human body. This effect could be related to: a) the energy required for peptide-peptide interactions in an aggregated form to give rise to peptide-phospholipid interactions during the binding process; b) a parasite specificity in the adaptation to the human host, making the schistosomula retain the SmKI-1 in the tegument and liberating only at specific temperature oscillation, pH variation or external enzymatic action.

The presented enCrypted algorithm is being improved to extend its application to other classes of bioactive peptides and the design of point-mutation and chemical modifications. Our perspective is that cryptides screening and peptide rational design will enable us to focus our efforts on biophysical characterization and using experimental data to create, in a near future, a more efficient encrypted bioactive peptide predictor.

## 5. Conclusions

While the identification of novel bioactive peptides can be performed by a top-down, untargeted approach by high throughput solution proteomic, we showed here that a bioprospection of encrypted, "hidden" peptides can satisfactorily be conducted by data mining using predictive analyses of protein database based on physicochemical properties of the sequence-only approach. The enCrypted algorithm presented here was capable of identifying hundreds of predicted AMPs, successfully electing three representatives for biological validation by antimicrobial assay and structural and thermodynamic analysis, and further optimization by point mutation. These observations demonstrated that SmKI-1 protein is a protein with multiple biotechnological applications and the bioactive

peptide-based agents derived from it might have high impact for system biology and biotechnology can be a successful strategy. The extension of this study for other biological systems and further optimization of enCrypted will assist in the rapid identification and design of novel bioactive peptide-based agents. Future investigations will focus in learning the mechanism of action of release of schistocins and how the protein gets accessible to the main proteases able to cut them off.

## Author contributions

Conceptualization: BS, SO, MM, Methodology: BS, EA, CF, RV, LL, SO, MM, Formal analyses: BS, EA, RV, MM. Investigation: BS, EA, CF, RV, LL, SO, MM. Writing: BS, MM, Writing-revision and edition: BS, EA, AG-N, RV, LL, SO, MM. All the authors contributed to the article and approved the manuscript.

## Funding

This research was supported by grants from Universidade Federal de Minas Gerais (UFMG/PRPQ/PQ; grant #23853-405 01/2017 to MM), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grants #465229/2014-0 to SO, and #303044/2020-9 to SO, by Fundação de Amparo a Pesquisa de São Paulo (FAPESP; grant #2017/24832-6 to SO), Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG; grant# REDE-00140-16 to SO and MM). BS was supported by a DSc fellowship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; grant #88882.349264/2019-01; Financial code #001). The funding agencies had no role in the study design, data collection and analysis, or decision to publish or prepare the manuscript.

## Data availability

All coordinates presented in this paper have been deposited in the Protein Data Bank (PDB) with the following codes: schistocin-1: 7M67; schistocin-2: 7M73; schistocin-3: 7M77; schistocin-3.1: 7M79. All NMR data have been deposited in the Biological Magnetic Resonance Bank (BMRB) with the following accession numbers: schistocin-1: 30891,

schistocin-2: 30892, schistocin-3: 30893 and schistocin-3.1: 30894. All other datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

#### **Declaration of Competing Interest**

None.

#### **Acknowledgements**

We thank Dr. Talita Lopes Santos for her technical support on peptide synthesis at Universidade Federal dos Vales do Jequitinhonha e Mucuri

(UFVJM). We thank Jamil de Oliveira for its technical support in some experiments. We thank Dr. Yulia Pustovalova and Dr. L. Maurício T. R. Lima for critical reading of the manuscript. We would like to thank Laboratório Multiusuário de Proteômica (LMProt/UFMG), Laboratório de Ressonância Magnética Nuclear (RMN/UFVJM), LASEB (UFVJM) and their staff for the access to their analytical facilities. This publication is part of BS DSc thesis.

#### **Appendix A. Supplementary data**

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbagen.2021.129989>.



## **Workflow for Protein N-terminal Acetylation and C-terminal Amidation Application using CcpNmr Analysis v2.4 Platform and Aria2.3 Structure Calculation Software**

**Bruno de Paula Oliveira Santos<sup>ab</sup>, Bruno Marques Silva<sup>ab</sup> and Mariana Torquato Quezado de Magalhães<sup>ab\*</sup>**

<sup>a</sup>Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, Av. Antonio Carlos, 6627, Belo Horizonte, Minas Gerais, 31270901, Brazil

<sup>b</sup>Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Av. Antonio Carlos, 6627, Belo Horizonte, Minas Gerais, 31270901, Brazil

\*Correspondence email: [mquezado@icb.ufmg.br](mailto:mquezado@icb.ufmg.br)

**Synopsis** During the past decades, different tools were developed for NMR data processing, peaks assignment, structure elucidation and data submission. Here we highlight CCPNMR2.4 analysis and ARIA2.3, responsible for peak assignment and structure calculation, respectively, and can work associated.

### **Abstract**

Structural biology is a field that enables a better understanding of proteins from scratch. From the available techniques, solution NMR is one well established that provides structure, dynamics and protein-molecules interaction. In a NMR lab routine, from data acquisition until protein/mechanisms elucidation comes a process that can undergo months. During the past decades, different tools were developed for NMR data processing, peaks assignment, structure elucidation and data submission. Since many of these programs demand great computational skills, a few groups have tried to combine those programs and make them more friendly and useful, what can possibilite a faster process. Here we highlight CCPNMR2.4 analysis and ARIA2.3, responsible for peak assignment and structure calculation, respectively, and can work associated. Although being academic free and the possibility of working with a GUI interface, the common N-terminal acetylation and C-terminal amidation modifications are not implemented in a way that possibilities to work with them in combination, what results in a dilemma. This work brings visual data that evidences the low usability of CCPN and ARIA with N-terminal acetylated and C-terminal amidated proteins and propose a workflow to overcome this problem, which may improve the usage of both software in the mentioned versions and facilitate the lab users already used to these programs. As a proof of concept, we have chosen a N-terminal amidated peptide, L-Phenylseptin, whose structure has already been solved with other programs. Statistical data

shows that no significant difference was found with the structure obtained with the new protocol. In conclusion, we exhibit a new protocol that can be used in combination with CCPNMR2.4 and ARIA2.3 for protein with the mentioned modifications and it successfully works and manipulates these molecules.

Keywords: workflows, NMR spectroscopy, experimental protocol, characterization, L-phenylseptin

## 1. Introduction

Applications for NMR structural biologists have been in development inside big NMR groups according to their necessities. Starting from assignments in papers and migrating to computer interface spectra interactions, NMR software have grown in usage and tools for researchers since the end of the 80's. If the established software do not reach your needs, why not develop your own system? Quality is synonym of improved producing, results and also, better work. When your group is able to develop your own software, features as usability can be implemented considering their own tastes and new implementations can go according to predefined models. Sometimes, the available platforms already accomplish the groups needs and just need some adaptations, what, in time, results in new versions of the same software. However, scientific computation is not an easy task, and multi-collaborative groups are set up to integrate NMR data and software development.

Historically, NMR software for protein assignment was first developed in 1989, when Ansig was published and made available ([Kraulis 1989](#)). Kraulis describes the laborious job of peak assignment in papers and how an interface guided program written in Fortran 77 could make the process faster and more practical, allowing the user to check the lists before exporting them for the structure software (in the work, X-PLOR was cited). In 1991, EASY was published as a new software for 2D spectra interactive assignment ([Eccles et al. 1991](#)) that was quickly substituted in 1995 by XEASY, an improved version that contented the assignment of 3D and 4D spectra and with the concept of 'strips', to visualize 3D spectra in a 2D manner ([Bartels et al. 1995](#)). Johnson and Blevins, in 1994, worked in NMR analysis integration with the structural data and also, integrate multiple spectra, instead of working with them "one at a time", originating the NMR View ([Johnson and Blevins 1994](#)). Sparky, from UCSF, had its first published version in 1999, first released for Linux systems (Goddard and Kneller 1999), and the content was similar what the other ones already provided: 2D and 3D NMR analysis and assignment. In 2007, Burrow-owl software was created and its procedure relays on transposing the NMR spectra, projecting a slice, combination of two spectra, diagonal projection and



integration of the spectra in an one fewer order dimension ([Benison et al. 2007](#)). Nowadays, Ansig was substituted by the collaborative platform CcpNmr Analysis, NMR View was upgraded for NMRFX and SPARKY maintenance was ceased in 2001, being substituted in 2015 by NMRFAM Sparky ([Vranken et al. 2005](#); [Norris et al. 2016](#); [Lee et al. 2015](#)). They work in most known and used operating systems (Linux, Windows and MacOs) and are supported in more recent script languages, such as Python and Java.

Even before the Ansig development, a software for protein structure calculation was already available in 1987, the first release of X-PLOR (Brünger 1993). X-PLOR (1.0) had evolved from the CHARMM force field program, and was capable of using X-ray crystallographic and NMR data for determine structural models, using molecular dynamics for structure refinement. In 1998, the same author of X-PLOR and new collaborators created a new definition of structure calculation software, the CNS. Using an own low-level script language, CNS is versatile in the source experiments, being used solution NMR, X-ray crystallography and also for solid state NMR and electron microscopy. CNS, over the X-PLOR 1.0, improved with the possibility of using more computational power and speeding the calculation process ([Brünger et al. 1998](#)). Both software are no longer maintained and were incorporated by NIH group, creating a new tool, the Xplor-NIH, which uses a variety of minimizations process to improve structure calculation ([Schwieters et al. 2003](#)). In other vertent, Nilges and collaborators in 1997, implemented a workflow for working with ambiguous NMR signs and it used X-PLOR for structure calculation ([Nilges et al. 1997](#)). In 2001 the situation changed, when the workflow was released as a software for NOE ambiguities and structural calculation, integrated with CNS, named Aria v1.0. Aria received several updates through the years and is one of the main free academic options for structural calculation nowadays ([Linge et al. 2001](#); [Rieping et al. 2007](#)). Cyana is also a main established software and its precursor, CANDID, was developed in 2002, that had as a main new component at the time, the automated NOE assignment ([Herrmann et al. 2002](#)). CANDID was reformulated as CYANA in 2004 to improve the handling of NMR chemical shift ambiguities ([Güntert 2004](#)).

New advances of 2010 until nowadays includes the outcoming of virtual machines, such as NMRbox, that allows researchers around the world to use a supermachine located in an US cluster. The platform provides pre-installed NMR software that allows the user to skip a dispendious times trying to install each software ([Maciejewski et al. 2017](#)). Another interesting outcome is the possibility of making your tasks in an online web server, such as the new version of Aria that, besides the downloadable version, allows the user to use their super machines for structure calculation and provides structure visualization for data evaluation ([Allain et al. 2020](#)). It is important to notice, however, that some software lose

maintenance since they are created by students during graduate processes, and the research group may lose capacity to keep the development and update of their tools. That is why collaborative projects have become of particular interest in NMR groups. Free academic software are available for each group to use for their own purposes but also allows more people to test and contribute in possible flaws or missing tools, that can be of interest of many other researchers. Here we highlight the importance of two of the quoted software in this process: CcpNmr Analysis v2.4 and Aria 2.3. Like a great union, CcpNMR analysis allows NMR assignment, Aria directly communicates to import the assigned data and make the structure calculation and validation and the results are exported back to Aria for structure analysis ([Vranken et al. 2005](#)). For NMR spectroscopists, this combination is very positive and provides a possibility to standardize a method to solve the structure of different proteins.

Here, we report on a workflow that allows to consider the N-terminal acetylation and C-terminal amidation of proteins during structure calculation, using CcpNmr Analysis v2.4 as the assignment platform and Aria2.3 as the structure calculation software. The implementation of these extremities molecules are missing in an integrated manner, what makes CcpNmr does not recognize this information on Aria and vice versa. The N-terminal acetylation and C-terminal amidation interferes in the final structures models and can not lack during the steps of NMR analysis. They guarantee to proteins features as: durability, endurance, binding and recognition. The C-terminal amidation is produced by the action of the peptidylglycine  $\alpha$ -amidating monooxygenase (PAM) enzyme when it recognizes dipeptides or tripeptides in proteins sequences, such as GK, GR, GKK, GKR ([Prigge et al. 2000](#)). The C-terminal amidation have been found in many bioactive peptides, leading the peptides to have a less ionizable C-terminal, what gives the molecule more chances for executing their biological actions ([Dennison and Phoenix 2011](#); [Dennison et al. 2012](#)). Examples of biological activity includes antimicrobial activity, where the C-terminal amidation impairs mechanisms of protein degradation ([Mura et al. 2016](#)); neuroactivity, since neuropeptides not amidated are not recognized by nervous systems cells ([Eipper et al. 1992](#)); hGLP-1 peptides that are responsible for insulin release and post-prandial plasma glucose lowering ([Zhang et al. 2004](#)) and others. Regarding protein acetylation, this modification can occur in the N-terminal portion or at the H $\epsilon$  of lysine residues. The focus of this study remains on the first option. N-terminal acetylation occurs by a series of acetyltransferases, on translation process of the protein, post-translationally on prokaryotes and on processed regulatory peptides of eukaryotes ([Driessen et al. 1985](#); [Dockray 1987](#)). This modification occurs in 85 percent of eukaryotic proteins and in bioactive peptides, such as human defensin-3 ([Papanastasiou et al. 2009](#)).

Here we demonstrate a direct and easy-to-implement workflow, that was tested to be used intercalating the already known Aria2 steps for structure calculation. Our observations provides a basis to suggest



that this workflow will improve the use of the combination CcpNmr-Aria by researchers that work with post-translational modified proteins and also reduce the use of different software when working with different types of molecules. Aiming to validate this project, we solved the structure of an antimicrobial peptide, the L-phenylseptin. This peptide, from *Hypsiboas punctatus* amphibian, have 14 amino acid residues and is known for its predator warning function, which structure have been previously solved using NmrView and Xplor-NIH and serve as comparison model to the new structures ([de Magalhães et al. 2013](#)). Our conclusions suggest that L-phenylseptin final structures are slightly changed from those that are not C-terminal amidated, and highlight the importance of considering these modifications during protein analysis.

## 2. Modifications Implementation

### 2.1. Data Survey

With the purpose of analyse and compare the usage of assignment and structure calculation software, pdb files were collected from Protein Data Bank webserver. The terms used in the search were: a) QUERY: Full Text = "antimicrobial peptide" AND ( Experimental Method = "SOLUTION NMR" AND Number of Polymer Residues per Assembly <= 40 AND Polymer Entity Type = "Protein" ); b) QUERY: Full Text = "ccpnmr" AND ( Experimental Method = "SOLUTION NMR" AND Polymer Entity Type = "Protein" ); c) QUERY: Full Text = "ARIA" AND ( Experimental Method = "SOLUTION NMR" AND Polymer Entity Type = "Protein" ); d) QUERY: Full Text = "xeasy" AND ( Experimental Method = "SOLUTION NMR" AND Polymer Entity Type = "Protein" ); e) QUERY: Full Text = "sparky" AND ( Experimental Method = "SOLUTION NMR" AND Polymer Entity Type = "Protein" ); f) QUERY: Full Text = "nmrview" AND ( Experimental Method = "SOLUTION NMR" AND Polymer Entity Type = "Protein" ). A threshold of 40 amino acids residues was arbitrarily established for the antimicrobial peptides class, to limit the size of these proteins and the quantity of files analyzed. The informations of the purchased files were extracted: a) using command-line Linux tools; b) manually assessed at the PDB webserver, when the “software used” data was not available in the files; c) in the published article, when the option “a” and “b” were not available. When none of the three options were indisposable, the target information was considered unavailable. Graphical visualization was provided by matplotlib library in python 3.6.

### 2.2. NMR Acquisition

This is a subsection (style: IUCr body text)

### 2.3. NMR assignment in CcpNmr Analysis

The first step to add the N-terminal and C-terminal modifications is to insert them in the molecule inside a CcpNmr project. In “add sequence”, the user must put their amino acid residues sequence and at the N-terminal portion the term “Acy” and/or at the C-terminal portion, the term “Nh3”. This extensions come direct in CcpNmr and nothing needs to be installed in this step. In atom browser, the hydrogens atoms are available for the acetyl as “H2\*”, and for the amidation, “H” and “H’”. The peak picking process consists of evaluating the spin systems of each residue in intra-residue correlations experiments (e.g. COSY, TOCSY), the inter-residue correlations (e.g. NOESY) and heteronuclear correlations ( $^{13}\text{C}/^{15}\text{N}$  HSQC) ([Sugiki et al. 2017](#)). CcpNmr have combined the DANGLE package that allows the prediction of psi and phi angles. In the “Structure” tab, the user have the “Make distance restraints” option, that converts volume/height NOESY information in atoms distance restraints.

#### 2.4. Aria2.3 Project

A new ARIA project is created under the command “aria2 -g” on command line shell. In the tab “CCPN data model”, the path of the CcpNmr project must be inserted. In the CNS tab, the CNS executable path. With the CcpNmr project chosen, the list are recognizable by Aria. The user is then capable of inserting the chemical shift list, molecular system, spectrum, dihedral angles and ambiguous and unambiguous distances directly from the CcpNmr project. Regarding the protein modifications, the user must insert in the tab “Molecular System”: a) Linkage Definition: TOPPALLHDG5.3, b) Topology definition: TOPPALLHDG5.3, c) Parameters Definition: PARALLHDG5.3. Those are the files modified to add the N-terminal acetylation and C-terminal amidation. Is also important to create a new folder to where the calculations will occur: in the tab “Project”, insert the new folder path in “working directory” and “temporary path”. This way, the modification patch will work without further alteration.

To insert the modification, the user must download the zip file (S1) and uncompress it in the same folder where the aria file is located (.xml file). The user must execute the install file, using the following command:

```
sudo ./install
```

It is default, after creating the aria project, in the xml extension, to execute the commands:

```
aria2 -s ariafile.xml
```

```
aria2 ariafile.xml
```

The first command allows ARIA to create the folder where the iterations will occur. The second one is where the ARIA starts the structure calculation. Here, we implement a new step:

```
aria2 -s ariafile.xml
```

```
./run_patches <structure_folder> <N or C or NC>
```

```
aria2 ariafile.xml
```

The second command implements the changes in topology and parameters to the new run. The second argument must be the name or path of the folder where the structure runs will occur. The third argument must be N or C or NC, if you want to implement the N-terminal modification or the C-terminal modification or both, respectively. Example:

```
./run_patches folder01 NC
```

## 2.5. Files Changes

### 2.5.1. Aria/CcpNmr

The main strategy to include these modifications was to consider them as residues. That in mind, the “new residues” were included in the files “atomnames.xml” and “iupac.xml”, from aria2.3 folders, with the same atoms nomenclature found in these molecules file of CCPN. They were named “ACY” and “NH2”. The file “AminoAcid.py” was modified to include the acetylation and amidation: The one letter code for both were set as “X”, as already seen in PDB deposited structures. In CcpNmr, the modification was in the folder “molecules”, where the files “others\_acy” and “others\_nh3” were changed to include a ccp\_3\_letter code, same as aria2.3 “atomnames” and “iupac” files.

### 2.5.2. CNS

The main changes are in the CNS files, to include the topologies and parameters. The topology files contain the “new residues” and the molecule bonds. The “topallhdg5.3.pro” has changed the residue ACE to ACY and the atoms nomenclatures. The residue NH2 was included and the “presidue PEPT” was increased with the new bonds and angles to be included in the backbone chain. The parameter file, parallhdg5.3.pro, include the new atoms, bonds, angles, dihedrals and impropers. These parameters were recycled from bonds and atoms with the same features of the amides and acetyl groups. The same was considered when evaluating the dihedral and impropers. (Fig S1).

## 3. Results and Discussion

### 3.1. CcpNmr users tend to choose ARIA during protein solving and vice-versa, but not for N-terminal acetylated and C-terminal amidated proteins

Since the frequency of C-terminal amidation and N-terminal acetylation in antimicrobial peptide is high, because of the importance of these modifications on the functionality of these type of molecules, the antimicrobial peptides N-terminal acetylated and/or C-terminal amidated with  $\leq 40$  residues were collected from PDB database to evaluate the most frequent software used for assignment and structure calculation (Table 2). Regarding the assignment software, SPARKY (19.7%), XEASY (16.1%), NMRVIEW (13.7%) and CCPNMR (12.0%), are the most used ones (Figure 1a). Also a great slice of

the sector graph belongs to uninformed software (17.1%), which information misses on PDB files and in the correspondent publication, when available. Regarding structure calculation software, CYANA (28.1%), XPLOR-NIH (26.1%), DYANA (a previous version of CYANA, 10.1%) and X-PLOR (a previous version of XPLOR-NIH, 8.5%), are the most used structure calculation software when considering the antimicrobial peptide class of peptides. ARIA represents only 3.3% of this proportion (Figure 1b).

To characterize the CcpNmr profile and its impact on PDB deposited structures, the protein structures assigned with CcpNmr were collected from PDB webserver (Table S2). The PDB data allow us to observe that the highest frequency of structure calculation software used with CcpNmr belongs to CYANA (46.7%), with ARIA in the second place (31.2%) (Figure 2a). This result is probably due to the pre-compiled intercommunication between CcpNmr and CYANA/ARIA software. Also, when evaluating the rate of N-terminal acetylated and C-terminal amidated structures in comparison with the other proteins assigned with CcpNmr, we observe an index of 5.1% of 430 proteins, in contrast with XEASY (6.5% of 976 proteins), SPARKY (3.7% of 2100 proteins) and NMRVIEW (2.5% of 2428 proteins) (Figure 2b, Table S3). Even though it is not a low rate, when observing the other tools, the absolute number of proteins of CcpNmr is the lowest. Changing the observation point to ARIA, CcpNmr represents the second software of choice in ARIA absolute numbers registered in PDB database (25%), only behind of NMRVIEW, with 29% (Figure 2c). The scenario changed since CcpNMR came out, and a great burden of the combination CcpNmr/ARIA is observed when CcpNmr Analysis version 2 was released, what made one of its competitor, SPARKY, be relegated from the second to the third place, with 20.5% (Figure 2c, Table S4). In regard to the N-terminal acetylated and C-terminal amidated proteins, the ratio of proteins solved by ARIA is 3.4% (Figure 2d). These results lead us to suggest that, in both ways, CcpNmr and ARIA are one of the main choices for the users of these tools. However, comparing the ratio of N-terminal acetylated and C-terminal amidated proteins to the total proteins assigned in CcpNmr and/or solved by ARIA, along with the relative low ratio of these tools in modified AMPs, we have informations to support the hypothesis that these tools are underused for N-terminal acetylated and C-terminal amidated proteins. A well established protocol for these kind of molecules could improve the usage of CcpNmr and ARIA for N-terminal acetylated and C-terminal amidated proteins.

### **3.2. L-Phenylseptin structures solved with ARIA have similar models with the structures solved with XPLOR-NIH**

To verify the implementation of the workflow in Aria2.3, we solved the structure of L-Phenylseptin peptide in DPC micelles. This peptide is found in amphibians with its C-terminal amidated. The NMR spectra assignment was performed by simultaneous analysis of  $H^1$ - $H^1$  TOCSY and NOESY, as proposed by Wüthrich. The  $H^\alpha$  and  $H^\beta$  were used by Dangle tool, present in CcpNmr Analysis, to predict dihedral angles. Except the first phenylalanine, all residue were identified in the experiments and were correlated with at least one more residue (Figure 3). The amide hydrogens allowed us to perform a sequential assignment, which confirms the peptide sequence and provides characteristic correlations on peptide two-dimensional structure. In the NOESY spectra, the most present correlations were NN(i,i+1) and NN(i,i+2). The NN(i,i+1) correlations are present between 3Phe and 4Asp and, after that, 8Asn beyond. The missing residues are complemented with NN(i,i+2) correlations, present in 2Phe with 4Asp and 5Thr with 7Lys. A new NN(i,i+2) correlation is only seen again along 11Lys and 13Ile. A suspect of alpha-helix structure is supported when  $\alpha$ N(i,i+3) is observed from 2Phe until 16Leu. Also, from 6Leu until 15Ala, there are  $\alpha$  $\beta$ (i, i+3) correlations, supported by two  $\alpha$ N(i, i+4) correlations in the same region, 9Leu with 13Val and 13Val with 17Leu. This results suggest a more strong folded C-terminal portion in comparison with the N-terminal. The alpha region showed 27 short-range NOEs (HN,  $H^\alpha$  - i, i + 1, i+2) and 8 medium range NOEs (HN,  $H^\alpha$  - i, i+3 and i,+4), that, along with short range amide correlations, are characteristic of alpha-helix structures (Fig. 3, table 1).

Once the NOEs were converted in distance restrictions in CcpNmr, the data was filled in ARIA and the peptide structure solved (Figure 4a). As expected, the molecule is an alpha-helix, with randomized portion among the first two residues and the last residue, due to their lack of correlation, as seen previously (Figure 3). The molecule presents a partial amphipacity, as the hydrophilic residues, 36% of the peptide composition, are in the opposite face to the hydrophobic residue (Figure 4a). The lowest energy structures evidence side-chain rigidity of hydrophobic residues: 3F, 10A, 13V, 14I, 16A, 17L. However, the hydrophilic residues have more free disposition and flexibility (Figure 4c). Since the structure of L-Phenylseptin was already solved in XPLOR-NIH ([de Magalhães et al. 2013](#)), we can observe that in macro visualization, no differences were observed between ARIA and X-PLOR-NIH structures (Figure 4b and 4d). The same residues have side-chain rigidity or flexibility.

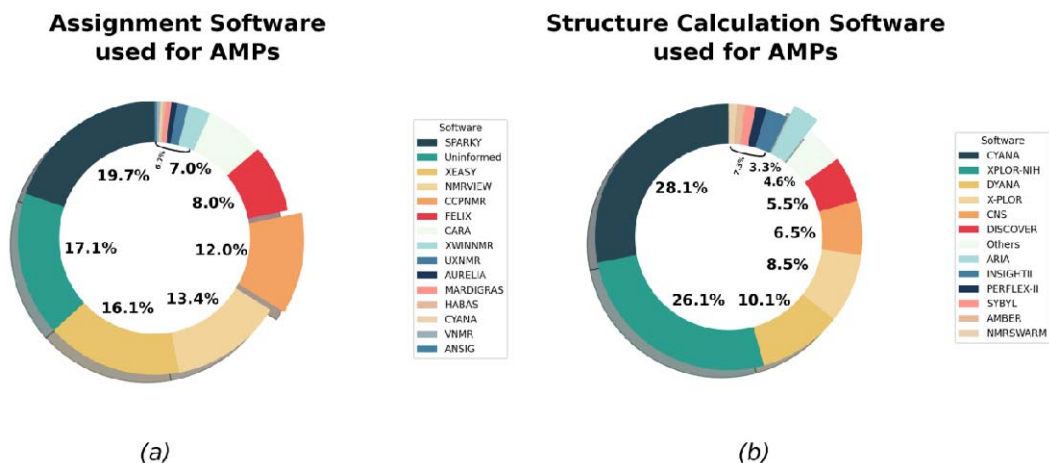
The statistical data are present in table 1 and the chemical shift table is present in table S5. The ensemble of the 20 lowest-energy XPLOR-NIH structures showed: a) all residues backbone RMSD of  $0.51 \pm 0.17$ ; b) all residues heavy atoms RMSD of  $1.55 \pm 0.4$  ([de Magalhães et al. 2013](#)). Aria2.3 ensemble of 20 lowest-energy C-terminal amidated structures showed: a) all residues backbone RMSD of  $0.81 \pm 0.22$ ; b) all residues heavy atoms RMSD of  $1.47 \pm 0.28$ . Aria2.3 structures appears to have backbone atoms more disperse and heavy atoms more rigid. However, considering the standard

deviations, our analysis suggests that there is no statistic differences between the Xplor-NIH and Aria2.3 structures.

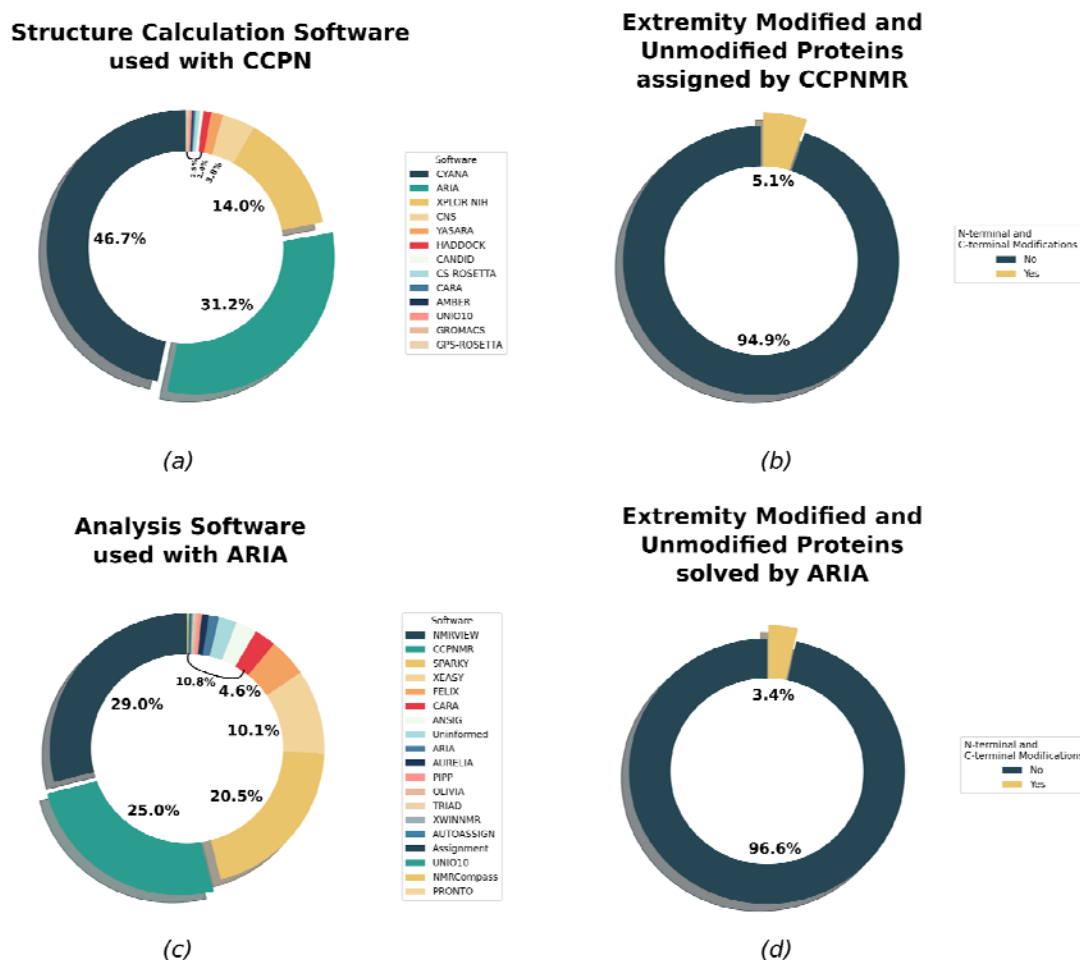
To compare the effect of the termini modifications on the final structures, the L-phenylseptin structures were calculated without N-terminal acetylation and C-terminal amidation. The overall RMSD in the ensemble of 20 lowest energies had no difference between all residues backbone RMSD ( $0.78 \pm 0.26$ ) and all residues heavy atoms RMSD ( $1.669 \pm 0.32$  vs  $1.668 \pm 0.32$ ) (table 1). A slightly difference is observed when no modified structures and N-terminal acetylated structures are compared with C-terminal terminal amidated structures, where all residues backbone RMSD ( $0.81 \pm 0.22$ ) have greater mean value and all residues heavy atoms RMSD ( $1.47 \pm 0.28$ ) have lower mean value. Statistically, we observe that there is no difference between the ensemble of 20-lowest energy structures, however, the lowest structure of the C-terminal amidated peptide seems better folded (Fig 5), as indicated in the tendency of lower heavy atoms RMSD.

#### **4. Conclusions**

In this study we improved CcpNmr/ARIA analysis developing a method to implement N-terminal acetylation and C-terminal amidation in proteins for users of CcpNmr Analysis v2.4 and Aria2.3 software. We understand that NMR tools need improvements and we hope to help the scientific academy supporting free aids to the already known software.



**Figure 1** Software used on antimicrobial peptides structure solving. A) Percentage of different assignment software used on PDB deposited AMPs. B) Percentage of different structure calculation software used on PDB deposited AMPs. Detached pie pieces highlight CcpNmr (a) and Aria (b).

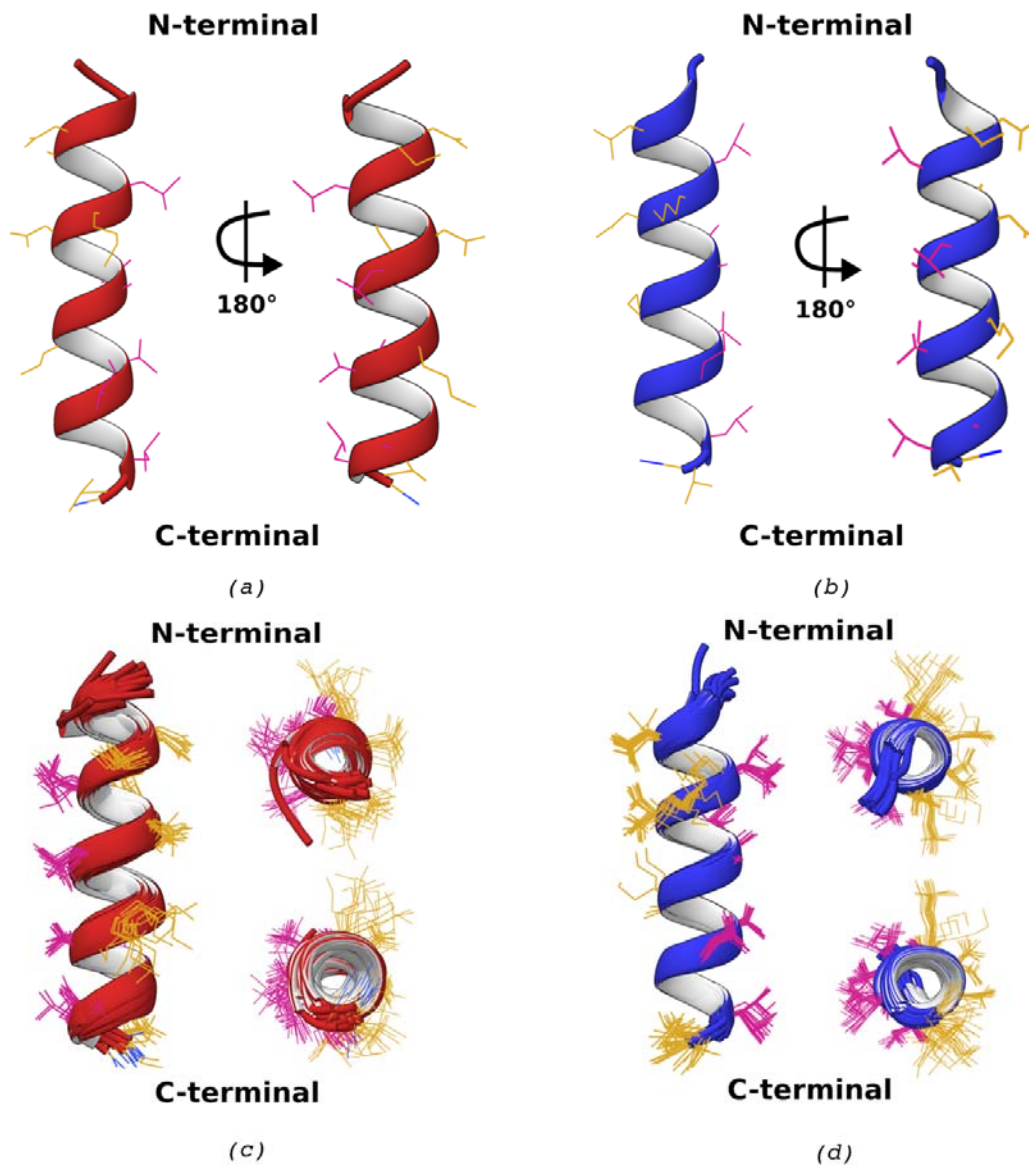


**Figure 2** CcpNmr and ARIA pie-chart. All protein structures deposited in PDB which used CcpNmr (a,b) and ARIA (c,d) was accessed. A) The structure calculation software used with proteins assigned on CcpNmr Analysis. B) Representation of N-terminal acetylated and C-terminal amidated structures against non-N-terminal acetylated and non-C-terminal amidated structures. C) The analysis software used with proteins solved by ARIA. D) Representation of N-terminal acetylated and C-terminal amidated structures against non-N-terminal acetylated and non-C-terminal amidated structures. Detached pie pieces represent Aria (a), CcpNmr (c) and modified proteins (b, d).

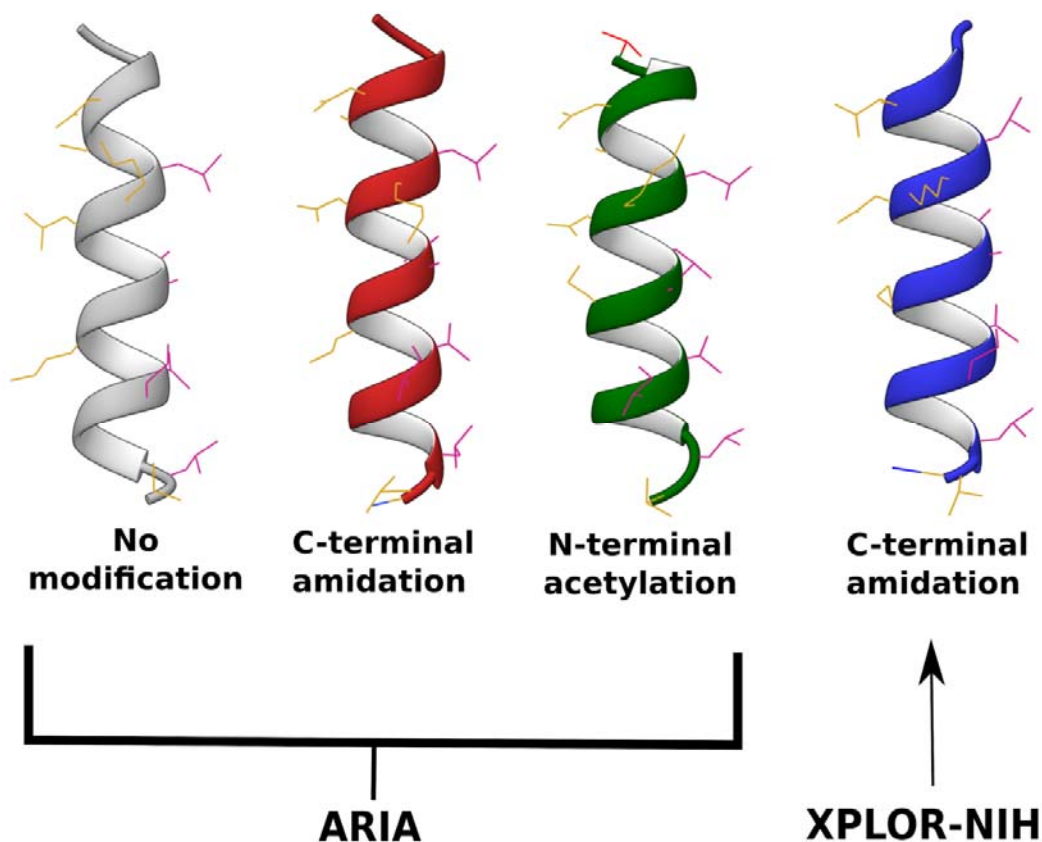




**Figure 3** Experimental restraints for L-Phenylseptin. Hydrogen atoms correlations:  $d_{\alpha N}$  (hydrogen-alpha and amidic hydrogen),  $d_{NN}$  (neighbours amidic hydrogens),  $d_{\beta N}$  (hydrogen-beta and amidic hydrogen),  $d_{\alpha N} (i, i+3)$  (hydrogen-alpha in position  $i$  and amidic hydrogen in position  $i+3$ ),  $d_{\alpha\beta} (i, i+3)$  (hydrogen alpha in position  $i$  and hydrogen-beta in position  $i+3$ ),  $d_{\alpha N} (i, i+4)$  (hydrogen-alpha in position  $i$  and amidic hydrogen in position  $i+4$ ),  $d_{NN} (i, i+2)$  (amidic hydrogen in position  $i$ , amidic hydrogen in position  $i+2$ ),  $d_{\alpha N} (i, i+2)$  (hydrogen-alpha in position  $i$  and amidic hydrogen in position  $i+2$ ),  $\Delta\delta(^1H^\alpha)$  (hydrogen-alpha secondary shifts). The primary structure is shown in the top. Sequential N-N and  $\alpha$ -N indicated by back bars: the thicker the bar, the stronger the connection.



**Figure 4** Comparison of D-Phenylseptin structures calculated in ARIA and XPLOR-NIH softwares. A) Peptides structures calculated by Aria2.3, in red. Vertical views from N-terminal until C-terminal. The peptide is represented in two views, after a rotation of 180°. B) Peptides structures calculated by XPLOR-NIH, in blue. Vertical views from N-terminal until C-terminal. The peptide is represented in two views, after a rotation of 180°. Both structures have C-terminal amide modification. C) 20 lowest energy structures of Aria2.3 peptide. Visions: vertical, C-terminal transversal and N-terminal transversal. C) 20 lowest energy structures of Xplor-NIH peptide. Visions: vertical, C-terminal transversal and N-terminal transversal.



**Figure 5** Comparison of peptides C-terminal and N-terminal modification. First structure represents the D-phenylseptin structure without C-terminal or N-terminal modifications. Second structure represents the D-phenylseptin peptide with a C-terminal amidation (dark blue). Third structure represents the peptide calculated with a N-terminal acetylation (red). All three cited structures were calculated using ARIA. Fourth structure contains a C-terminal amidation and was calculated in XPLOR-NIH. All structures represent the lower energy structure between 20 refined in water.

**Table 1** Summary of 20 solution NMR structural statistics

Values for the outer shell are given in parentheses.

	No modifications	N-terminal Acetylated	C-terminal Amidated
<b>NOEs</b>			
Intraresidue	129	129	130
Sequential	49	49	50
Short range	27	27	27

---

Medium range	8	8	8
Longe range	0	0	0
Total unambiguous	217	217	218
Ambiguous	0	0	0
Total NOE-derived	217	217	218
Dihedral angles ( <i>psi</i> + <i>phi</i> )	16	18	18
<b>Structural statistics</b>			
Ramachandran Plot PROCHECK			
(%)			
Most favored	98.1	100	99.6
Additionally allowed	1.9	0	0.4
Generously allowed	0	0	0
Ramachandran Plot PROCHECK			
(%)			
Most favored	100	100	99.7
Additionally allowed	0	0	0.3
Generously allowed	0	0	0
Quality Factors (z score)			
Procheck G-Factor (psi/phi only)	2.16	2.56	2.87
Procheck G-Factor (all dihedral angles)	0.83	1.18	1.54
Molprobity clashcore	0.36	0.5	1.54
<b>NOE violations</b>			
>0.5 Å	0	0	0
>0.3 Å	0	0	0
<b>Structural precision</b>			
RMSD from average structure (Å)			
Backbone (all residues)	0.783 ± 0.265	0.782 ± 0.265	0.817 ± 0.228

---

Heavy atom (all residues)	$1.669 \pm 0.328$	$1.668 \pm 0.327$	$1.472 \pm 0.283$
Backbone (secondary structure)	$0.220 \pm 0.071$	$0.220 \pm 0.071$	$0.223 \pm 0.082$
Heavy atom (secondary structure)	$0.804 \pm 0.169$	$0.804 \pm 0.169$	$0.826 \pm 0.111$

---

**Table 2** Chemical Shift Table of L-Phenylseptin

This is a table headnote (style: IUCr table headnote)

	H	HA	HB	HG	HD	HE
1 Phe	-	4.25	3.12, 2.95	-	7.41	7.30
2 Phe	7.87	4.53	2.85	-	7.33	7.15
3 Phe	8.50	4.50	3.06, 2.99	-	7.30	7.06
4 Asp	7.28	4.30	3.02, 3.01	-	-	-
5 Thr	7.78	4.23	4.05	1.26	-	-
6 Leu	7.81	4.11	1.64	-	0.89, 0.92	-
7 Lys	7.87	3.97	1.84	1.43	1.53	-
8 Asn	7.89	4.53	2.94, 2.83	-	7.41, 6.77	-
9 Leu	8.13	4.17	1.79, 1.75	1.72	0.90	-
10 Ala	8.52	4.03	1.46	-	-	-
11 Gly	8.02	3.82, 3.93	-	-	-	-
12 Lys	7.81	4.19	2.14, 2.00	1.53	1.63, 1.71	-
13 Val	8.15	3.68	2.22	1.04, 0.95	-	-
14 Ile	8.66	3.76	1.91	0.92, 1.21, 1.74	0.83	-
15 Gly	8.20	3.78, 3.89	-	-	-	-
16 Ala	7.94	4.24	1.57	-	-	-
17 Leu	8.30	4.24	1.92	1.58	0.87	-
18 Thr	7.84	4.35	4.34	1.29	-	-

---

This is a table footnote (style: IUCr table footnote)

## Acknowledgements

BPSO, BSM and MTQM acknowledge grants from CNPq, FAPEMIG and CAPES. This work is a collaboration research project of members of the Rede Mineira de Imunobiológicos (MG) supported by FAPEMIG.

## References

- Bardiaux, B., Malliavin, T., & Nilges, M. (2012). *Methods Mol. Biol.* **831**, 453–483.
- Brand, G. D., Ramada, M. H. S., Manickchand, J. R., Correa, R., Ribeiro, D. J. S., Santos, M. A., Vasconcelos, A. G., Abrão, F. Y., Prates, M. V., Murad, A. M., et al. (2019). *PLoS ONE*. **14**, e0220656.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998). *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.
- Dennison, S. R., Morton, L. H. G., & Phoenix, D. A. (2012). *Biochim. Biophys. Acta.* **1818**, 2094–2102.
- Dennison, S. R. & Phoenix, D. A. (2011). *Biochemistry.* **50**, 10898–10909.
- Dockray, G. J. (1987). *Am. Rev. Respir. Dis.* **136**, S9-15.
- Driessen, H. P., de Jong, W. W., Tesser, G. I., & Bloemendal, H. (1985). *CRC Crit. Rev. Biochem.* **18**, 281–325.
- Eipper, B. A., Stoffers, D. A., & Mains, R. E. (1992). *Annu. Rev. Neurosci.* **15**, 57–85.
- Howard, M. J. (1998). *Curr. Biol.* **8**, R331-3.
- de Magalhães, M. T. Q., Barbosa, E. A., Prates, M. V., Verly, R. M., Munhoz, V. H. O., de Araújo, I. E., & Bloch, C. (2013). *PLoS ONE*. **8**, e59255.
- Mura, M., Wang, J., Zhou, Y., Pinna, M., Zvelindovsky, A. V., Dennison, S. R., & Phoenix, D. A. (2016). *Eur. Biophys. J.* **45**, 195–207.
- Papanastasiou, E. A., Hua, Q., Sandouk, A., Son, U. H., Christenson, A. J., Van Hoek, M. L., & Bishop, B. M. (2009). *APMIS.* **117**, 492–499.
- Prigge, S. T., Mains, R. E., Eipper, B. A., & Amzel, L. M. (2000). *Cell. Mol. Life Sci.* **57**, 1236–1259.
- Skinner, S. P., Goult, B. T., Fogh, R. H., Boucher, W., Stevens, T. J., Laue, E. D., & Vuister, G. W. (2015). *Acta Crystallogr. D Biol. Crystallogr.* **71**, 154–161.
- Zhang, Z.-Z., Yang, S.-S., Dou, H., Mao, J.-F., & Li, K.-S. (2004). *Protein Expr. Purif.* **36**, 292–299.