



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA

CAMILA TAUANE MONTEIRO DO NASCIMENTO

**IDENTIFICAÇÃO, CARACTERIZAÇÃO E APLICAÇÃO BIOTECNOLÓGICA DE
UMA SERINO-PEPTIDASE DA PLANTA CARNÍVORA *Nepenthes mirabilis* PARA
HIDRÓLISE DAS PROTEÍNAS DO LEITE BOVINO**

FORTALEZA

2020

CAMILA TAUANE MONTEIRO DO NASCIMENTO

IDENTIFICAÇÃO, CARACTERIZAÇÃO E APLICAÇÃO BIOTECNOLÓGICA DE UMA
SERINO-PEPTIDASE DA PLANTA CARNÍVORA *Nepenthes mirabilis* PARA HIDRÓLISE
DAS PROTEÍNAS DO LEITE BOVINO

Tese apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutora em Bioquímica.
Área de concentração: Bioquímica Vegetal.

Orientador: Prof. Dr. Cleverson Diniz Teixeira de Freitas.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- N194i Nascimento, Camila Tauane Monteiro do.
Identificação, caracterização e aplicação biotecnológica de uma serino-peptidase da planta carnívora *Nepenthes mirabilis* para hidrólise das proteínas do leite bovino / Camila Tauane Monteiro do Nascimento. – 2020.
66 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Bioquímica, Fortaleza, 2020.
Orientação: Prof. Dr. Cleverson Diniz Teixeira de Freitas.
1. *Escherichia coli*. 2. Alergia alimentar. 3. Expressão heteróloga. 4. Corpos de inclusão. 5. Peptidase. I. Título.

CDD 572

CAMILA TAUANE MONTEIRO DO NASCIMENTO

IDENTIFICAÇÃO, CARACTERIZAÇÃO E APLICAÇÃO BIOTECNOLÓGICA DE UMA
SERINO-PEPTIDASE DA PLANTA CARNÍVORA *Nepenthes mirabilis* PARA HIDRÓLISE
DAS PROTEÍNAS DO LEITE BOVINO

Tese apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutora em Bioquímica.
Área de concentração: Bioquímica Vegetal.

Aprovada em: 05 / 03 /2020.

BANCA EXAMINADORA

Prof. Dr. Cleverton Diniz Teixeira de Freitas (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Hermógenes David de Oliveira
Universidade Federal do Ceará (UFC)

Prof. Dr. Márcio Viana Ramos
Universidade Federal do Ceará (UFC)

Prof. Dr. Gilvan Pessoa Furtado
Fundação Oswaldo Cruz (Fiocruz-Ceará)

Prof.a Dra. Suelen Carneiro Medeiros
Instituto Tecnológico Federal do Ceará (IFCE)

A Deus.

Aos meus pais, Nascimento e Janemary.

À minha filha e esposo, Maria Clara e Pedro

Igor.

AGRADECIMENTOS

A Deus pelo dom da vida.

À minha família: pai (Nascimento), mãe (Janemary) e irmã (Nicolly Janine), pelo apoio, cuidado e atenção ao longo de toda minha vida.

À minha filha (Maria Clara), por me acolher como mãe-cientista, muitas vezes sem tempo para a função de mãe-mãe e respeitar todas as decisões por mim tomadas.

Ao meu marido (Pedro Igor), por me apoiar sempre e ser esse companheiro maravilhoso.

A todos os professores do Departamento de Bioquímica, pela concessão de seus equipamentos, e por me cumprimentarem, muitas vezes, com sorrisos acolhedores no corredor.

A todos os professores do curso de Biotecnologia por acreditarem nesse curso e por todo dia quererem torná-lo um diferencial no mercado de trabalho. Agradeço também, por terem permitido muitas vezes a presença da minha filha em suas aulas.

Às Profas. Dra. Vânia Maria Maciel Melo e Denise Cavalcante Hissa, por terem sido meu primeiro exemplo de mulheres cientistas, fortes e que lutam por uma ciência de qualidade nesse país.

Ao Prof. Dr. Márcio Viana Ramos, por ter me orientado durante meu mestrado e que continuou como exemplo durante o doutorado.

Ao Prof. Dr. Thalles Barbosa Granjeiro por ter me coorientado no mestrado; com quem aprendi muito do que sei sobre Biologia Molecular.

Ao Prof. Dr. Cleverson Diniz Teixeira de Freitas, pela orientação ao longo desses 4 anos, pelo apoio, paciência e reconhecimento.

Aos colegas e amigos dos Laboratórios de Plantas Laticíferas e o de Biotecnologia de Proteases Vegetais pela convivência diária, pelas risadas e pelos memes.

Aos colegas do curso de Biotecnologia, que embora o curso tenha terminado e nos alocado em laboratórios diferentes, a amizade permaneceu e ultrapassou tais barreiras.

Ao grupo que se formou há uns 10 anos, “*fat friends*”, representado pelas queridas Roberta Laiz, Roberta Cristiane, Marina Gabrielle e pelo querido Acrisio Bastos, que me ampara quando mais preciso.

Aos colegas técnico-administrativos, tanto da chefia de departamento quanto das coordenações de graduação e de pós-graduação, pelas perturbações diárias com declarações, dúvidas diversas, reservas de sala e até configurações de computador.

Às meninas da limpeza Teca e Nega por permitirem condições assépticas de trabalho.

Aos professores doutores participantes da banca examinadora Hermógenes David de Oliveira, Márcio Viana Ramos, Gilvan Pessoa Furtado e Suelen Carneiro Medeiros pelo tempo, pelas valiosas colaborações e sugestões.

Ao CNPq pela concessão da bolsa de estudos da estudante, FUNCAP pelo apoio financeiro que subsidiaram os reagentes e os materiais usados durante a pesquisa da estudante. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

As proteínas do leite bovino são responsáveis pelos tipos mais comuns de alergia alimentar. Pessoas diagnosticadas com alergia ao leite usam fórmulas contendo proteínas do leite parcial ou extensivamente hidrolisadas. Como essas fórmulas são caras, devido ao processo enzimático, a pesquisa por novas proteases, capazes de produzir fórmulas hipoalergênicas mais eficientes, ainda é um assunto interessante. Este trabalho teve como objetivo avaliar o potencial biotecnológico para a hidrólise das proteínas do leite bovino e caracterizar três serino carboxiproteases (SCP) do fluido digestivo da planta carnívora *Nepenthes mirabilis* (SCP3, SCP20 e SCP47). Além disso, uma protease (SCP3) foi escolhida para expressão heteróloga adicional em *Escherichia coli* Shuffle®T7. O SDS-PAGE mostrou que a SCP hidrolisou extensivamente as caseínas do leite bovino, resultando em uma menor antigenicidade dos hidrolisados, medido por ELISA, quando comparado às caseínas do leite bovino não hidrolisadas. As análises *in silico* das três SCP revelaram semelhanças de estruturas primárias, secundárias e tridimensionais com outras SCP de plantas, sendo classificadas como pertencentes ao clã SC das serino proteases. Embora a SCP3 tenha sido obtida na sua forma solúvel usando etanol 1% durante a indução com IPTG 0,5 mM a 16 °C por 18 h, ela não mostrou atividade proteolítica. A obtenção dessa proteína em sua forma ativa pode levar à sua aplicação na indústria de alimentos para a hidrólise de proteínas do leite bovino com a produção de fórmulas hipoalergênicas, pois SCP3 também apresentou baixo potencial alergênico e foi extensamente hidrolisada por proteases digestivas, ambas previstas por ferramentas de bioinformática.

Palavras-chave: *Escherichia coli*. Alergia alimentar. Expressão heteróloga. Corpos de inclusão. Peptidase.

ABSTRACT

Bovine milk proteins are responsible by the most common type of food allergy. People diagnosed with milk allergy use formulas containing partially or extensively hydrolyzed milk proteins. Because these formulas are costly, due to enzymatic process, the research for new proteases, able to more efficiently produce hypoallergenic formulas, is still an exciting subject. This work aimed to evaluate the biotechnological potential for the hydrolysis of bovine milk proteins and to characterize three serine carboxypeptidases (SCP) of the digestive fluid of the *Nepenthes mirabilis* carnivorous plant (SCP3, SCP20, and SCP47). Besides, a protease (SCP3) was chosen for further heterologous expression in *Escherichia coli* Shuffle®T7. SDS-PAGE showed that SCP extensively hydrolyzed bovine caseins, resulting in a lower antigenicity of the hydrolysates, measured by ELISA, when compared to non-hydrolyzed bovine caseins. In silico analyzes of the three SCP revealed similarities of primary, secondary, and three-dimensional structures with other plant SCP, being classified as belonging to the SC clan of the serine proteases. Although SCP3 was obtained in its soluble form using 1% ethanol during induction with 0.5 mM IPTG at 16 °C for 18 h, it did not show proteolytic activity. Obtaining of this protein in its active form can lead to its application in the food industry for the hydrolysis of cow's milk proteins with the production of hypoallergenic formulas, because SCP3 also showed low allergenic potential and it was extensively hydrolyzed by digestive proteases, both predicted using bioinformatics tools.

Keywords: *Escherichia coli*. Food allergy. Heterologous expression. Inclusion bodies. Peptidase.

LISTA DE ILUSTRAÇÕES

Figura 1 – Armadilhas da planta carnívora <i>Dionaea</i>	15
Figura 2 – Armadilhas das plantas carnívoras <i>Drosera</i> e <i>Pinguicula</i>	16
Figura 3 – Armadilhas da planta carnívora <i>Utricularia</i>	16
Figura 4 – Armadilhas da planta carnívora <i>Nepenthes</i>	17
Figura 5 – Plantas carnívoras das Famílias Nepenthaceae, Cephalotaceae, Sarraceniaceae.....	18

LISTA DE ABREVIATURAS E SIGLAS

AP	Peptidases aspártica
APLV	Pacientes com alergia à proteína do leite de vaca
BMPs	Peptídeos bioativos do leite
CI	Corpos de inclusão
CpCP3	Carboxipeptidase de <i>Calotropis procera</i>
DNA	Ácido desoxirribonucleico
<i>E.coli</i>	<i>Escherichia coli</i>
IgE	Imunoglobulina E
MEROPS	Banco de dados de peptidases
N	Nitrogênio
<i>N.mirabilis</i>	<i>Nepenthes mirabilis</i>
pET-SUMO	Plasmídeo utilizado como vetor de expressão
SCP	Serino carboxiproteases
SP	Serino peptidases
SUMO	Proteína modificadora do tipo ubiquitina
Unesp	Universidade Estadual Paulista

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO BIBLIOGRÁFICA.....	15
2.1	Plantas carnívoras.....	15
2.2	Gênero <i>Nepenthes</i>.....	18
2.3	Peptidases.....	20
2.4	Sistema de expressão procarioto (<i>Escherichia coli</i>).....	23
2.5	Alergia Alimentar	25
3	OBJETIVOS.....	27
3.1	Objetivo Geral.....	27
3.2	Objetivos Específicos.....	27
4	BIOTECHNOLOGICAL POTENTIAL FOR THE HYDROLYSIS OF BOVINE MILK PROTEINS AND CHARACTERIZATION OF THREE SERINE CARBOXYPROTEASES FROM THE DIGESTIVE FLUID OF NEPENTHES MIRABILIS CARNIVOROUS PLANT.....	28
5	CONCLUSÃO.....	61
	REFERÊNCIAS.....	62

1 INTRODUÇÃO

Ao longo da evolução, as plantas carnívoras desenvolveram estruturas especializadas responsáveis por atrair e capturar suas presas, como armadilhas de pressão, fluidos viscosos e ventosas (THOROGOOD; BAUER; HISCOCK, 2018). Essas adaptações foram observadas e estudadas desde Darwin (1875), que as consideravam as plantas mais maravilhosas do mundo. Até o momento, cerca de 600 plantas carnívoras foram identificadas em diferentes famílias de angiospermas, sendo o gênero *Nepenthes* o mais numeroso, compreendendo quase 160 espécies (ROTTLOFF et al., 2016).

Após a captura, as presas são digeridas por diferentes enzimas, fornecendo nutrientes essenciais para o crescimento e desenvolvimento dessas plantas, que crescem em solos deficientes em nitrogênio (GIVNISH et al., 1984; GIVNISH et al., 2018). Entre as proteínas digestivas identificadas estão RNases, esterases, fosfatases ácidas e alcalinas, lipases, quitinases e proteases (LEE et al., 2016). Embora os fluidos digestivos de plantas carnívoras representem uma fonte interessante de enzimas, estudos que descrevem a purificação ou a caracterização bioquímica e estrutural dessas proteínas ainda são raros na literatura.

As proteases são enzimas ubíquas capazes de clivar a ligação peptídica de outros peptídeos/proteínas, constituindo as enzimas mais abundantes em qualquer genoma. Eles podem ser classificados com base em seus mecanismos de ação em: aspartato, cisteína, glutamato, metalo, serina e treonina proteases (RAWLINGS, 2013). Essas enzimas são de alto interesse comercial porque são usadas como aditivos nas indústrias de detergente, processamento de couro e alimentos, bem como na indústria terapêutica (PHILIPPS-WIEMANN, 2018). Na indústria de alimentos, as proteases têm sido usadas para reduzir a alergia alimentar.

As alergias alimentares são um importante problema de saúde pública em todo o mundo. Cerca de 8% das crianças e 2% dos adultos têm alguma alergia alimentar (TURNBULL; ADAMSN; GORARD, 2015). Assim, estima-se que 700 milhões de pessoas (considerando uma população mundial atual de cerca de 7 bilhões) tenham algum tipo de alergia alimentar, representando um problema de saúde global substancial. Em particular, a alergia ao leite bovino constitui o principal tipo de alergia alimentar (MANUYAKORN; TANPOWONG, 2019). A alternativa mais comum para pessoas diagnosticadas com alergia ao leite é o uso de fórmulas contendo proteínas do leite hidrolisadas parcial ou extensivamente (MURARO et al., 2012; OLIVEIRA et al., 2019). O alto custo dessas fórmulas pode estar associado às enzimas usadas nesse processo (GOLKAR; MILANI; VASILJEVIC, 2019). Por isso, a busca por proteases

capazes de produzir fórmulas proteicas hidrolisadas, com mais eficiência, é um assunto motivador.

Como as proteases digestivas de plantas carnívoras são extremamente bem adaptadas à hidrólise extensiva de proteínas de origem animal, foi levantada a hipótese de que essas enzimas poderiam ser usadas como novas ferramentas biotecnológicas para digerir proteínas do leite bovino e produzir fórmulas hipoalergênicas. Dessa forma, o extrato proteolítico do fluido digestivo de *Nepenthes mirabilis* foi estudado e testado para hidrolisar as proteínas do leite de vaca. Diante de sua baixa produção pela planta e após caracterizar *in silico* as três principais serina proteases presentes no fluido, uma serina protease foi escolhida e expressa em células de *Escherichia coli* e seu potencial alergênico foi avaliado por bioinformática.

2 REVISÃO BIBLIOGRÁFICA

2.1 Plantas carnívoras

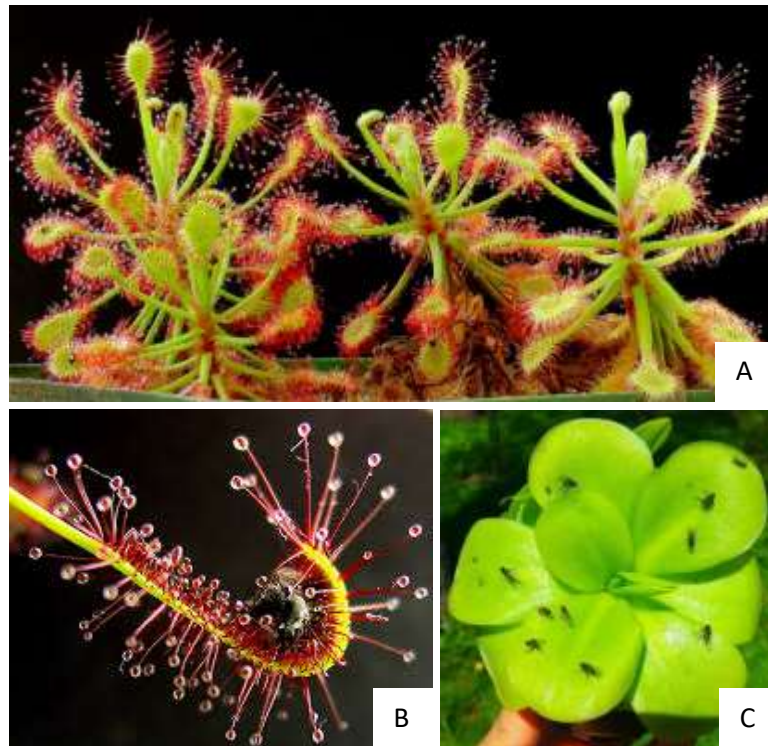
Plantas que possuem estruturas especializadas com a finalidade de atrair, capturar e digerir nutrientes da presa e que se beneficiam dessa captura são chamadas de plantas carnívoras. A carnivoría em plantas é uma adaptação estratégica a ambientes onde a oferta de nutrientes é escassa. Por isso a necessidade de se obter nutrientes minerais adicionais, inclusive nitrogênio, aprisionando e digerindo suas presas (ROTTLOFF et al., 2011; BEHIE; BIDOCHKA, 2013). Existem vários tipos de adaptações morfofisiológicas que funcionam como estruturas de captura: armadilhas de pressão (*Dionaea*) (Figura 1), adesivas (*Drosera* ou *Pinguicula*) (Figura 2), sugadoras (*Utricularia*) (Figura 3) e armadilhas em si (*Nepenthes*) (Figura 4) (KROL et al., 2012). A complexidade de adaptações em diferentes plantas carnívoras, incluindo a presença de atividade proteolítica, levou Darwin (1875) a considerá-las como as plantas mais maravilhosas do mundo. Desde então, aproximadamente 600 espécies de plantas carnívoras foram identificadas em seis subclasses de angiospermas, sendo o gênero *Nepenthes* o mais numeroso, compreendendo quase 160 espécies (ROTTLOFF et al., 2016).

Figura 1 - Armadilhas da planta carnívora *Dionaea*.



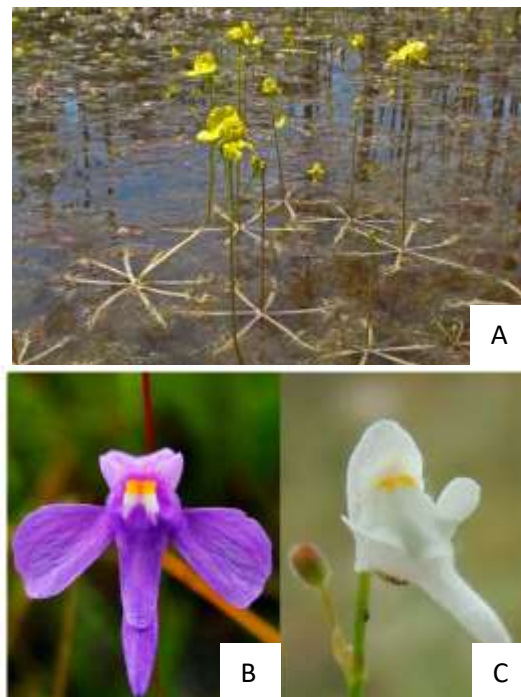
A. Armadilhas de *Dionaea* do tipo pressão. B e C. Zoom dessas armadilhas capturando um inseto. Fonte: <https://www.carnivoras.com.br/>.

Figura 2 - Armadilhas das plantas carnívoras *Drosera* e *Pinguicula*.



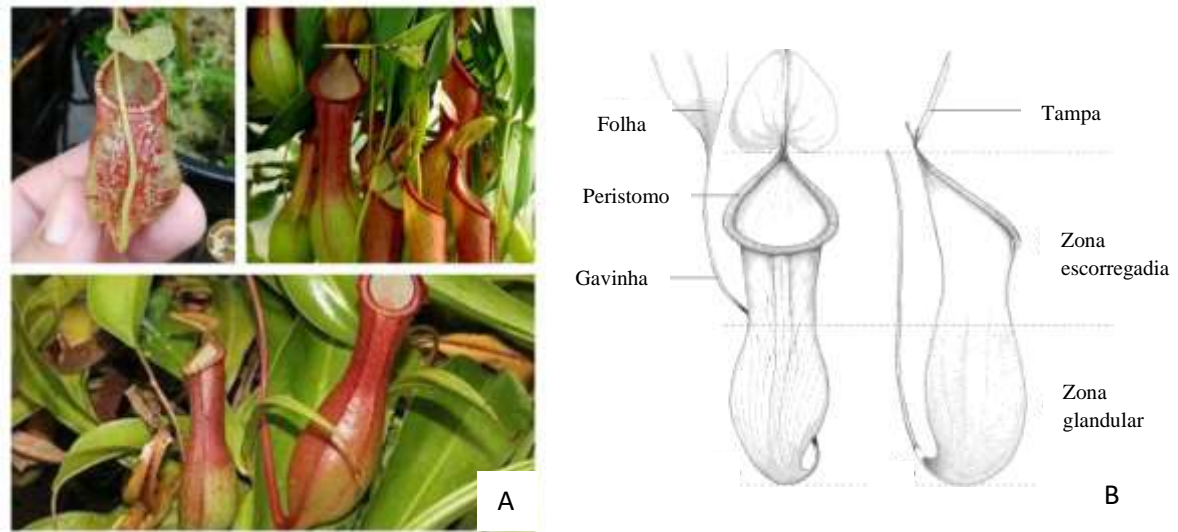
A. Armadilhas de *Drosera* do tipo adesivas ou “pega moscas”. **B.** Zoom dessa armadilha de *Drosera* capturando uma presa. **C.** Armadilhas de *Pinguicula* capturando insetos. Fonte: <https://www.plantascarnivoras.es>.

Figura 3 - Armadilhas da planta carnívora *Utricularia*.



A. Armadilhas de *Utricularia* do tipo adesivas sugadoras. **B e C.** Zoom de dois tipos de *Utricularia* que se assemelham a orquídeas. Fonte: <https://carnivorousplantnursery.com/>

Figura 4 - Armadilhas da planta carnívora *Nepenthes*.



A. Algumas *Nepenthes* do tipo “armadilhas em si” ou “*pitcher plant*”. Fonte: <https://diegooliveiranoticias.home.blog/> **B.** Desenho alusivo às regiões de captura da presa dentro do jarro dessa planta. Fonte: MIGUEL; HEHNB; BOURGAUDA, 2018.

Plantas carnívoras das Famílias Nepenthaceae, Cephalotaceae, Sarraceniaceae são capazes de atrair insetos forrageiros, voadores ou rastejantes e outros pequenos invertebrados em seus poços ou jarros cheios de líquido (Figura 5). Uma vez dentro do poço, o inseto é digerido enzimaticamente (JAFFE et al., 1992). Segundo Ellison e Gotelli (2001), essas plantas são capazes de obter entre 10% e 80% do nitrogênio (N) total do inseto, dependendo do ambiente e do tipo de armadilha empregada. Em geral, as plantas carnívoras apresentam uma grande dependência de nitrogênio, e essa demanda aumenta à medida que as estruturas carnívoras se tornam mais elaboradas (de folhas pegajosas de *Drosera* spp. à jarros com mais de 1m de altura do lírio-cobra *Darlingtonia californica*). As próprias plantas variam ao longo do tempo na contribuição relativa do N derivado de insetos para o conteúdo total de N da planta (SCHULZE et al., 1997; ADLASSNIG et al., 2012).

A evolução múltipla e independente da carnivoría em diversas famílias de plantas sugere que essa é uma adaptação aos habitats com baixo teor de nutrientes e luz, nos quais ocorrem tais espécies (GIVNISH et al., 1984; GRAY et al., 2017). Assim, os benefícios nutricionais da carnivoría têm sido o foco tradicional da pesquisa nas últimas décadas (SKATES et al., 2019).

Figura 5 - Plantas carnívoras das Famílias Nepenthaceae, Cephalotaceae, Sarraceniaceae.



Fonte: <https://www.carnivorousplants.org/>, <http://www.plantecarnivore.fr>,
<https://www.growcarnivorousplants.com/>, <https://www.carnivoras.com.br/>.

2.2 Gênero *Nepenthes*

O grupo monofilético das *Nepenthaceae* (*Caryophyllales*) contém até 100 espécies, e se apresenta como a maior família de plantas-jarro carnívoras (JEBB; CHEEK, 1997). *Nepenthes* é o maior gênero de plantas carnívoras encontrado no sudoeste da Ásia (MORAN; CLARKE, 2010). Elas são restritas aos Paleotrópicos, indo de Madagascar para o leste até a Nova Caledônia e um pequeno número de ilhas ocidentais do Pacífico. O centro da diversidade está no arquipélago indonésio, Filipinas, Sulawesi e nas grandes ilhas de Sonda, em Bornéu e Sumatra (CLARKE, 1997; CLARKE, 2001).

Nepenthes são plantas carnívoras altamente estudadas e desenvolveram uma série de características ecológicas que atraíram a atenção dos botânicos. Várias estratégias adaptativas desenvolvidas por essas plantas as tornam uma fonte de inspiração para muitas aplicações, desde tratamentos terapêuticos até soluções de biocontrole na agricultura. A excelente organização de tecidos do jarro digestivo pode ajudar a criar materiais novos e originais

utilizáveis na vida cotidiana em muitas culturas do mundo, sendo uma ferramenta etnobotânica capaz de curar dores de estômago, febre e até fornecer material para construção de casas (MIGUEL; HEHN; BOURGAUD, 2018).

Ao observarmos a morfologia de plantas carnívoras de *Nepenthes*, encontramos a estrutura em formato de jarro, que inicia no final da folha e forma progressivamente uma cavidade/armadilha que consiste em três seções: uma borda superior chamada peristomo que está envolvida na atração e captura de presas, coberta com uma substância escorregadia semelhante a um muco; quando um inseto pousa, é incapaz de escapar (GAUME; FORTERRE, 2007); uma parede interna escorregadia e cerosa para prender e impedir a fuga; e uma cova inferior preenchida com um fluido viscoelástico ácido usado para digerir a presa (Figura 4b). Diversas variações estruturais e químicas da armadilha parecem ser adaptadas para uma variedade de invertebrados, incluindo detritos de plantas e até mesmo fezes de mamíferos arborícolas (MORAN et al., 2012).

Essas armadilhas, em formato de jarro, como no caso da Família Nepenthaceae, contém uma diversidade de enzimas digestivas como proteases, RNAses, esterases, fosfatases e quitinases (ROTTLOFF et al., 2016; LEE et al., 2016), além de grande densidade de bactérias (TAKEUCHI et al., 2011). O fluido presente nessa armadilha é ácido, apresentando um pH entre 2 a 6 dependendo do tipo de espécie (TAKEUCHI et al., 2011). Adaptações especiais como camadas de cera, anisotropia das glândulas digestivas, peristomo escorregadio e fluido digestivo ácido e viscoelástico estão envolvidos no mecanismo de captura que contribuem com grande porcentagem de exigência nutricional total da planta (MORAN; CLARKE, 2010). A presa geralmente perde a estabilidade em superfícies escorregadias e cai dentro do jarro contendo o fluido digestivo (PAVLOVIČ; SAGANOVÁ, 2015).

Nepenthes mirabilis, conhecida também como planta-jarro, é uma espécie de planta carnívora bastante difundida e diversa, abrangendo o sudeste da Ásia continental e todas as principais ilhas do arquipélago malaio, estendendo-se da China até a Austrália (CLARKE, 2014). A sua estrutura de captura em formato de jarro, que inicia no final da folha, possui comprimento de até 5 m, com o diâmetro de até 5 mm, e pode ser encontrado nas cores verde, vermelho ou verde rosa. Possui folhas finas de forma alongada ou em lanceta, de coloração verde com superfície escorregadia, com comprimento da folha de até 15 cm e com largura de até 8 cm (LESTARININGSIH; SETYANINGSIH, 2017).

Essas plantas são um objeto vantajoso para estudos porque o conteúdo presente no jarro pode ser facilmente coletado, além de ser o próprio fluido da digestão (MITHOFER, 2011). A

identificação de toda a composição de proteínas no fluido digestivo é desafiadora porque sua composição de aminoácidos é incomum. O banco de dados tanto genômico como proteico de plantas carnívoras é limitado. Silva et al. (2016) e Silva et al. (2017), um grupo da Universidade Estadual Paulista (Unesp), sequenciaram o genoma cloroplastidial e mitocondrial de uma planta carnívora brasileira, a *Utricularia reniformis*, endêmica do Brasil. Relataram que trabalhar com genoma nuclear de plantas, por bioinformática, é uma tarefa desafiadora, já que o genoma de planta varia em tamanho e ordem de disposição das bases nos cromossomos. É preciso determinar quais porções do genoma correspondem a genes funcionais.

Dessa forma, plantas carnívoras representam um grande reservatório ainda não explorado de proteases com potenciais atividades biotecnológicas, tanto por apresentarem enzimas estáveis, como pela sua pluralidade (LEE et al, 2016).

2.3 Peptidases

Enzimas proteolíticas (proteases ou peptidases) agem como efetores positivos ou negativos de numerosos processos biológicos, apresentando a habilidade de digerir proteínas usando uma variedade de sítios de clivagem essenciais para a identificação de proteínas e peptídeos a partir de uma mistura complexa (MISHRA et al., 2017). Nesse sentido de enriquecer o “catálogo” da diversidade de proteases, tem-se a degradômica, uma forte abordagem ômica, que permitiu uma ampla visão da biodiversidade de proteases. Um degradoma pode ser definido como toda a diversidade de proteases em um organismo, tecido ou célula em um determinado momento, com o objetivo de elucidar as proteases que estão presentes em um tempo definido, em um meio específico, incluindo informações sobre o substrato enzimático e/ou inibidor enzimático (PEREZ-SILVA et al., 2016). Essa complexidade dos ciclos vitais e suas mudanças entre diferentes organismos fornece uma enorme variedade de proteases com funções diferentes e, conseqüentemente, uma grande variedade de especificidades e estruturas (PUENTE et al., 2003).

Essas enzimas são capazes de hidrolisar as ligações peptídicas próximo às extremidades das cadeias polipeptídicas (exopeptidases) ou dentro delas (endopeptidases) (PALMA et al., 2002). E participam do primeiro lugar no mercado mundial de enzimas, estimado em aproximadamente US\$ 3 bilhões de dólares (LEARY et al., 2009), uma vez que desempenham um papel importante na biotecnologia, já que sua atividade proteolítica altera propriedades químicas, físicas, biológicas e imunológicas das proteínas.

Segundo Choudhary et al. (2004), as enzimas são uma ferramenta tecnológica de interesse da indústria porque são capazes de atingir uma alta taxa de reação em condições variadas de pH, temperatura e pressão, além de uma alta especificidade de reação, biodegradabilidade, de natureza não tóxica e que produzem efluentes não poluentes. Adicionalmente, peptidases de origens vegetal e animal são utilizadas em fármacos para doenças pancreáticas, artrites e intolerância ao glúten (LORKOWSKI, 2012).

Nas últimas décadas, o interesse pelas proteases vegetais aumentou significativamente. Enzimas vegetais, como proteases, são amplamente utilizadas na medicina e na indústria de alimentos. Algumas proteases, como papaína, bromelaína e ficina, são usadas em vários processos, como fabricação de cerveja, amaciamento de carne, coagulação do leite, tratamento de câncer, digestão e distúrbios virais, podendo ser obtidas a partir de sua fonte natural ou através de culturas *in vitro* (GONZÁLEZ-RÁBADE et al., 2011).

Oliveira et al. (2019) avaliaram quatro tipos de fluidos laticíferos como nova fonte de peptidases capazes de hidrolisar proteínas no leite de vaca. Os ensaios *in vivo* mostraram que as proteínas do leite de vaca hidrolisadas por dois fluidos laticíferos do estudo, não exibiram reações imunes em camundongos alérgicos ao leite de vaca, semelhante a uma fórmula comercial parcialmente hidrolisada. Consideradas, portanto, promissoras para o desenvolvimento de novas fórmulas hipoalergênicas para crianças com alergia ao leite.

Diante desse cenário, podemos classificar as enzimas proteolíticas (E. C. 3.4) com base em seu mecanismo catalítico em: aspárticas, cisteínicas, metalo-treoninas e serinas (BARRETT et al., 2003). Com o advento do sequenciamento do genoma, surgiu a necessidade de expandir esse sistema de classificação, abrangendo o repertório catalítico encontrado na natureza. Outrossim, sequências e estruturas de todos os compostos proteolíticos e enzimáticos conhecidos encontram-se no banco de dados MEROPS (<https://www.ebi.ac.uk/merops/>). Esse sistema de classificação divide peptidases em clãs, utilizando o mecanismo catalítico e as famílias na base de ancestralidade comum. Atualmente, mais de 900.000 sequências de proteínas peptidases foram classificadas em 62 clãs e 268 famílias (RAWLINGS et al., 2018).

As peptidases aspárticas (APs; EC: 3.4.23) são uma família de enzimas amplamente distribuídas na natureza, onde desempenham papéis-chave em uma ampla gama de funções biológicas (ZHAO et al., 2013). APs de plantas foram identificadas e purificadas a partir de diferentes espécies (GUO et al., 2013; GUO et al., 2015). Kadek et al. (2014) caracterizaram enzimática e bioquimicamente, e expressaram em organismo heterólogo uma protease aspártica da planta carnívora do gênero *Nepenthes*, nepentesina-1, protease capaz de digerir presas no

jarro de *Nepenthes gracilis*. Embora análises de bioinformática tenham sido realizadas com proteases aspárticas de sementes de plantas para predição de diferentes estruturas e funções, suas funções biológicas não são tão bem conhecidas ou caracterizadas quanto as dos seus correspondentes em mamíferos, microrganismos ou vírus, que demonstraram desempenhar funções diferentes, incluindo o processamento específico de proteínas (DARABI; SEDDIGH, 2015).

Butts et al (2016) identificaram 44 peptidases cisteínicas a partir do DNA genômico de planta carnívora *Drosera capensis*, as quais revelaram homologia com peptidases vegetais já conhecidas, porém, com novas propriedades, confirmando esse grande potencial a ser explorado. Proteases cisteínicas do tipo papaína (E. C. 3. 4. 22. 2) são as mais investigadas dentre as representantes dessa família. A papaína é caracterizada por uma estrutura com dois domínios e com o sítio ativo entre esses domínios. Essas enzimas são sintetizadas como precursores inativos ou menos inativos, que compreendem uma pro-sequência com 38 a 250 resíduos de aminoácidos e a enzima madura com 220 a 260 resíduos de aminoácidos (GRUDKOWSKA; ZAGDAŃSKA, 2004). RISØR et al. (2016) identificaram, também, que proteínas de presas encontradas no líquido digestivo da planta carnívora *Dionaea* são degradadas por endopeptidases cisteínicas em associação com as carboxipeptidases serínicas.

Mais de um terço de todas as enzimas proteolíticas conhecidas são do tipo serino peptidases (SP) e aproximadamente 26.000 peptidases de serina são agrupadas em 13 clãs e 40 famílias, são um dos maiores grupos de enzimas proteolíticas encontradas em eucariotos e procariotos (PAGE; DI CERA, 2008). O nome da família deriva do resíduo nucleofílico do aminoácido serina no sítio ativo da enzima, que ataca a porção carbonila da ligação peptídica do substrato para formar um intermediário acil-enzima (HEDSTROM, 2002). Elas são amplamente distribuídas na natureza, encontradas em todos os reinos de vida celular, bem como muitos genomas virais. Nas plantas, elas são difundidas entre os grupos taxonômicos, de árvores a leguminosas e ervas, e estão presentes em quase todas as partes das plantas, mais abundantes em frutas. SPs de cucurbitáceas, cereais e plantas de grande porte são geralmente classificados como um só grupo (RAWLINGS; BARRETT, 2004).

SPs de plantas foram encontradas e extraídas de sementes de cevada, soja e arroz; do látex de *Euphorbia supina*, *Taraxacum officinale*, *Synadenium grantii* e *Artocarpus heterophyllus*; de flores, caules, folhas e raízes de *Arabidopsis thaliana*; das raízes de armazenamento de batata-doce e milho; dos brotos de bambu; das folhas de tabaco, feijão e

tomate; e dos frutos de *Cucumis melo*, *Cucurbita ficifolia*, *Maclura pomifera*, *Zehneria japonica* e *Trichosanthes cucumerina*, entre outros (ANTAO; MALCATA, 2005).

Segundo Darabi e Seddigh (2015), devido à baixa quantidade de proteases presentes no jarro de *Nepenthes* e também à acessibilidade limitada à essas plantas, não era possível sua caracterização estrutural mais detalhada, anteriormente. Por isso, escolheram utilizar a estratégia de produção heteróloga em *Escherichia coli* de nepenthesina-1 para obtenção de grandes quantidades de proteína. Desde a descoberta da protease nepenthesina-1 secretada no jarro de *Nepenthes*, enzimas digestivas de plantas carnívoras têm sido o foco de muitos estudos. Recentes abordagens genômicas aceleraram a descoberta de enzimas digestivas, e com o avanço da tecnologia do DNA recombinante e da purificação de proteínas, a identificação e caracterização de enzimas em plantas carnívoras foi facilitada (RAVEE; SALLEH; GOH, 2018).

2.4 Sistema de expressão procarioto (*Escherichia coli*)

O advento da tecnologia do DNA recombinante revolucionou as estratégias para a produção de proteínas heterólogas. Devido ao genoma bem caracterizado e uma variedade de ferramentas bem estabelecidas e disponíveis para manipulação genética, *Escherichia coli* é, ainda, a alternativa mais escolhida para a produção de proteínas recombinantes (BANEYX; MUJACIC, 2004).

Este sistema bacteriano oferece respostas rápidas para questões iniciais acerca da transgenia da proteína desejada, tais como: a estabilidade gênica no sistema heterólogo; a expressão correta e em níveis adequados; fácil manuseio, além de ser um organismo muito bem caracterizado e compreendido geneticamente. No entanto, a produção heteróloga para aplicações industriais ou para fins de pesquisa, na maioria das vezes, apresenta condições de crescimento celular significativamente diferente ao ambiente natural do micro-organismo, resultando numa limitação da produtividade (CHOU, 2007).

Mesmo apresentando desvantagens como incapacidade de realizar modificações pós-traducionais, bem como capacidade limitada de refazer pontes dissulfeto presentes em algumas estruturas proteicas, além de um elevado conteúdo de endotoxinas, comercialmente, a bactéria *E.coli* é classificada como organismo seguro, e provou ser economicamente viável na produção de produtos proteicos, o que não descarta a necessidade de otimização do processo. Essa

otimização se dá pela maior quantidade de proteína recombinante funcional produzida por unidade de volume e unidade de tempo (BROEDEL et al., 2001).

É possível caracterizar corpos de inclusão como agregados proteicos, insolúveis, presentes na maioria das vezes, no citoplasma de um organismo procarioto recombinante, resultante de uma superexpressão do plasmídeo que possui a região codificadora de um gene de interesse, chamados de corpos de inclusão (CI) (GONZÁLEZ-MONTALBÁN et al., 2007). Segundo Prouty, Karnovsky e Goldberg (1975), os CI são grânulos intracelulares densos e amorfos, visíveis por micrografia eletrônica.

Uma solubilização eficiente é o passo inicial na recuperação da proteína. Walther et al. (2013) observaram que após a solubilização, os corpos de inclusão foram penetrados pelo agente de solubilização, diminuindo os núcleos densamente empacotados, bem como a proteína, difundindo-se para fora. Pensando na estratégia de expressão de proteínas solúveis, Lobstein et al. (2012) desenvolveram uma cepa de expressão de proteínas em *E. coli*, chamada SHuffle, com o propósito de produzir proteínas recombinantes ativas, permitindo a correta formação de pontes dissulfeto com altos rendimentos em seu citoplasma com capacidade redutora diminuída. Esta cepa possui integrado ao seu cromossomo um gene que codifica para isomerase DsbC, que permite que ligações dissulfeto sejam formadas no citoplasma. O estado redox de DsbC no citoplasma oxidante, auxilia na formação de correto folding de proteínas com multi pontes dissulfeto.

Panavas, Sanders e Butt (2009) verificaram que uma proteína modificadora do tipo ubiquitina (SUMO), de células eucarióticas, ao ser fusionada à região N-terminal da proteína heteróloga, melhorava a produção dessa proteína em sistemas de expressão tanto procarióticos como eucarióticos. Essa significante melhora era baseada na estabilidade e solubilidade proteica após sua purificação, seguida da remoção da tag SUMO por uma protease específica.

Para ajudar na solubilização e na purificação de proteínas recombinantes, Silva et al. (2020) utilizaram o vetor de expressão pET-SUMO, contendo uma cauda tag e 6 histidinas adicionadas ao N-terminal da protease CpCP3 e expressaram uma protease de *Calotropis procera* em *E. coli* SHuffle (rCpCP3). Eles caracterizaram e avaliaram o potencial biotecnológico dessa protease cisteínica purificada a partir do látex de *C. procera* (CpCP3). Esta enzima mostrou-se altamente estável a diferentes íons metálicos e foi capaz de hidrolisar κ -caseína de maneira semelhante à quimosina bovina, para produção de queijo.

Lacou et al. (2015) relatam que os produtos alimentares são matrizes muito complexas, apresentando diferentes composições, que interagem entre si, resultando em características

específicas do alimento final. Entre eles, temos peptídeos e proteínas contidos, assumindo uma posição de destaque. A hidrólise dessas proteínas afeta as propriedades da matriz alimentar e é realizada por várias razões, que podem produzir efeitos positivos, como modificações na qualidade sensorial (como textura ou sabor), digestibilidade aprimorada, diminuição da alergenicidade ou como a liberação de peptídeos bioativos (GONZÁLEZ-RÁBADE et al., 2011; LACOU et al., 2015).

2.5 Alergia Alimentar

Segundo Prescott e Allen (2011), a alergia alimentar é um problema de saúde pública tão importante, que se coloca em segundo lugar no ranking das principais causas de alergia atrás apenas da asma. A alergia alimentar é definida como uma reação adversa reprodutível aos alimentos, mediada por mecanismos imunológicos. Esse distúrbio é a causa mais comum de reações alérgicas, em crianças ou adultos (CASAÑAL et al., 2013).

Evidências sugerem que 5% dos adultos e 8% das crianças possuam algum tipo de alergia alimentar, com um aumento em prevalência nas últimas décadas. Os fatores que afetam esse aumento progressivo de alergia vão desde o estilo de vida, como o grande consumo de alimentos industrializados, até a genética do indivíduo (SICHERER; SAMPSON, 2014). Os principais alimentos alérgenos identificados pela Codex (2009) são amendoim, nozes, ovo, soja, peixe, crustáceos, leite e trigo.

Hidrólises enzimáticas e processamentos adicionais são comumente utilizados para produzir fórmulas dietéticas hipoalergênicas derivadas de diferentes fontes de proteínas, como o leite de vaca, lentilhas e grão de bico, por exemplo (CUADRADO et al., 2009). A hidrólise da cadeia polipeptídica pode ser a única maneira de impedir a identificação do epítopo pelo sistema imunológico (SATHE; SHARMA, 2009). Uma seleção adequada de uma protease com a especificidade desejada e o protocolo de hidrólise adequado podem diminuir radicalmente a alergenicidade das proteínas. No entanto, existem poucos estudos para avaliar o efeito na reatividade de Imunoglobulina E (IgE) desses produtos de hidrólise com soros de pacientes com alergia a esses alimentos (CUADRADO et al., 2009). E o alto custo dessas fórmulas hipoalergênicas pode estar associado aos tipos de enzimas usados nesse processo (GOLKAR; MILANI; VASILJEVIC, 2019).

Dentre os alimentos de dieta diária, o leite de vaca é considerado uma matriz biológica complexa, fonte de peptídeos bioativos obtidos após fermentação *in vivo* e *in vitro* com

potencial atividade antimicrobiana, antioxidante, antitrombótica, anti-hipertensiva, imunomoduladora, dentre outras (KORHONEN, 2009). A busca pela identificação e caracterização nutricional desses peptídeos com diferentes bioatividades, também chamados de peptídeos bioativos do leite (BMPs), tem sido priorizada devido a essa pluralidade (GIACOMETTI; BURETIĆ-TOMLJANOVIĆ, 2017).

Por outro lado, o leite (ou seus derivados) pode desencadear processos alérgicos. A alergia ao leite envolve o sistema imunológico a partir da reação a proteínas existentes no leite (ou outros produtos lácteos), liberando substâncias que causam sintomas alérgicos (HEINE et al., 2017). Essa reação alérgica pode ser leve (erupções cutâneas) ou muito grave (dificuldade em respirar, perda de consciência). Uma pessoa com alergia ao leite geralmente é alérgica a uma das duas principais proteínas do leite: caseína ou soro de leite. A alergia à caseína ocorre quando o sistema imune desencadeia reações para eliminar a presença da caseína. No caso da proteína do soro de leite, os sintomas alérgicos podem estar associados a alergias ou a intolerância (HUSSAIN, 2018).

Em 2010 foram publicadas diretrizes para classificar, desenvolver e recomendar ações em prol de pacientes com alergia à proteína do leite de vaca (APLV). Dentre essas diretrizes, cita-se a busca por fórmulas alternativas e economicamente acessíveis para a condição clínica de cada paciente (FIOCCHI et al., 2018). A publicação desses documentos visa aprofundar o conhecimento clínico, melhorar a qualidade do diagnóstico e atendimento, reduzindo erros na condução da prática clínica. A aplicação dessas diretrizes pode promover o uso eficiente de recursos, informar e capacitar pacientes e apoiar políticas públicas (ŠTULC et al., 2017). Portanto, a pesquisa por novas proteases capazes de produzir fórmulas proteicas do leite de vaca extensivamente ou parcialmente hidrolisadas, com mais eficiência, ainda é um assunto motivador.

3 OBJETIVOS

3.1 Objetivo Geral

Esse trabalho teve como objetivo caracterizar três serino carboxipeptidase (SCP) do fluido digestivo de *Nepenthes mirabilis*, com o propósito de avaliar seu potencial biotecnológico para a hidrólise das proteínas do leite de vaca.

3.2 Objetivos Específicos

Analisar o perfil proteico e proteolítico do fluido digestivo de *Nepenthes mirabilis*;

Hidrolisar as proteínas do leite de vaca utilizando as proteases do fluido digestivo de *N. mirabilis*;

Caracterizar *in silico* as três serino carboxipeptidase majoritárias do fluido digestivo de *N. mirabilis*;

Construir os modelos tridimensionais das três serino carboxipeptidasas SCP3, SCP20 e SCP47;

Expressar a protease SCP3 em linhagem de *E. coli* como sistema heterólogo;

Estabelecer melhor condição de solubilização da proteína heteróloga;

Avaliar atividade proteolítica da protease recombinante (rSCP3);

Caracterizar *in silico* o potencial alergênico da protease SCP3;

Running title: Study of serine proteases from *Nepenthes mirabilis*

Biotechnological potential for the hydrolysis of bovine milk proteins and characterization of three serine carboxypeptidases from the digestive fluid of *Nepenthes mirabilis* carnivorous plant

Camila T.M.N. Porfírio^a, Márcio V. Ramos^a, Samuel F. Freitas^a, João P.B. Oliveira^a,
Gilvan P. Furtado^b, Pedro F.N. Souza^a, José S.S. Barbosa^a, Thalia L. Frota^a, Celso S. Nagano^c
, Jefferson S. Oliveira^d, Cleverson D.T. Freitas^{a*}

^a Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Ceará, Centro de Ciências, Campus do Pici. Fortaleza, Ceará, Brazil. CEP 60440-900.

^b Fundação Oswaldo Cruz, Fiocruz-Ceará, Brazil.

^c Departamento de Engenharia de Pesca, Universidade Federal do Ceará, Fortaleza, CE, Brazil

^d Departamento de Biomedicina. Universidade Federal do Delta do Parnaíba, Parnaíba, Piauí, Brazil

Corresponding author: Cleverson D.T. Freitas (cleversondiniz@ufc.br or cleversondiniz@hotmail.com). Phone number: +55 85 3366 9816.

ABSTRACT

Bovine milk proteins are responsible by the most common type of food allergy. People diagnosed with milk allergy use formulas containing partially or extensively hydrolyzed milk proteins. Because these formulas are costly, due to enzymatic process, the research for new proteases, able to more efficiently produce hypoallergenic formulas, is still an exciting subject. This work aimed to evaluate the biotechnological potential for the hydrolysis of bovine milk proteins and to characterize three serine carboxypeptidases (SCP) of the digestive fluid of the *Nepenthes mirabilis* carnivorous plant (SCP3, SCP20, and SCP47). Besides, SCP3 was chosen for further heterologous expression in *Escherichia coli* Shuffle®T7. SDS-PAGE showed that SCP extensively hydrolyzed bovine caseins, resulting in a lower antigenicity of the hydrolysates, measured by ELISA, when compared to non-hydrolyzed bovine caseins. *In silico* analyzes of the three SCP revealed similarities of primary, secondary, and three-dimensional structures with other plant SCP, being classified as belonging to the SC clan of the serine proteases. Although SCP3 was obtained in its soluble form using 1% ethanol during induction with 0.5 mM IPTG at 16 °C for 18 h, it did not show proteolytic activity. Obtaining of this protein in its active form can lead to its application in the food industry for the hydrolysis of bovine milk proteins with the production of hypoallergenic formulas, because SCP3 also showed low allergenic potential and it was extensively hydrolyzed by digestive proteases, both predicted using bioinformatics tools.

Keywords: *Escherichia coli*; Food allergy; Heterologous expression; Inclusion bodies; Peptidase;

1 INTRODUCTION

Over evolution, carnivorous plants have developed specialized structures responsible for attracting and capturing their preys, such as pressure traps, viscous fluids, and suckers (Thorogood, Bauer, & Hiscock, 2018). These adaptations have been observed and studied since Darwin (1875), who considered them the most wonderful plants in the world. To date, around 600 carnivorous plants have been identified in different angiosperm families, being *Nepenthes* genera the most numerous, comprising nearly 160 species (Rottloff et al., 2016).

After the capture, the preys are digested by different enzymes, providing essential nutrients for growth and development of these plants, which grow in nitrogen deficient soils (Givnish et al., 1984; Givnish et al., 2018). Among the digestive proteins identified are RNases, esterases, acid and alkaline phosphatases, lipases, chitinases, and proteases (Lee, Zhang, Ozar, Sensen, & Schriemer, 2016). Although digestive fluids from carnivorous plants represent an exciting source of enzymes, studies describing the purification or reporting biochemical and structural characterization of these proteins are still rare in literature.

Proteases are ubiquitous enzymes able to cleave the peptide bond of other peptides/proteins, constituting the most abundant enzymes in any genome. They can be classified based on their mechanistic features into aspartate, cysteine, glutamate, metallo, serine, and threonine proteases (Rawlings, 2013). These enzymes are of high commercial interest because they have been used as additives in detergent, leather processing, and therapeutics as well as in food industry (Philipps-Wiemann, 2018). In the food industry, proteases have been used to reduce food allergy.

Food allergies are important medical and social problems worldwide. Around 8% of infants and 2% of adults have some food allergy (Turnbull, Adamsn, & Gorard, 2015). Thereby, it is estimated that 700 million people (considering a current world population of around 7 billion) have a type of food allergy, representing a substantial global health problem. In

particular, bovine milk allergy constitutes the main type of food allergy (Manuyakorn & Tanpowpong, 2019). The most common alternative for people diagnosed with milk allergy is the use of formulas containing partially or extensively hydrolyzed milk proteins (Muraro et al., 2012; Oliveira et al., 2019). The high cost of these formulas can be associated with the enzymes used in this process (Golkar, Milani, & Vasiljevic, 2018). Therefore, the research for new proteases able to produce extensively or partially hydrolyzed milk protein formulas, more efficiently, is still a motivating subject.

Because carnivorous plant digestive proteases are extremely well adapted to hydrolysis extensively proteins from animal sources, it was hypothesized that these enzymes could be used as new biotechnological tools to digest bovine milk proteins and produce hypoallergenic formulas. In this respect, the proteolytic extract from the *Nepenthes mirabilis* digestive fluid was studied and tested for hydrolyzing the cow's milk proteins. Furthermore, bioinformatics tools were used to characterize its three main serine proteases. Finally, a serine protease was expressed in *E. coli* cells and its allergenic potential was evaluated *in silico*.

2 MATERIAL AND METHODS

2.1 Reagents

Acrylamide (17-1302-02), bis-acrylamide (17-1304-02), sodium dodecyl sulfate (SDS) (17-1313-01), 1-mL Histrap™ HP (Nickel-Sepharose) column (29-0510-21), and molecular mass markers (17-0446-01) were obtained from GE Healthcare Life Sciences (São Paulo, SP, Brazil). β -mercaptoethanol (M6250), kanamycin (60615), isopropyl- β -D-thiogalactopyranoside (IPTG, I5502), 5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside (X-Gal, B1690), Freund's complete (F5881) and incomplete (F5506) adjuvant, alkaline phosphatase-conjugated goat anti-rabbit IgG antibodies (A6066), 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (B6149), and nitro blue tetrazolium (NBT) (N6876) were obtained from Sigma-Aldrich (São Paulo, SP, Brazil). All other chemicals were of analytical grade.

2.2. Digestive fluid harvest and analysis of its protein and proteolytic profile

Nepenthes mirabilis species were purchased from a Brazilian website specialist in carnivorous plants (<https://www.plantascarnivoras.com.br/>), and then cultivated under natural light (12 h) at 28 ± 3 °C, relative humidity of 70%, using a substrate poor in nutrients. The plants were daily watered with distilled water. Digestive fluid samples (3 -5 mL) were collected from newly opened pitchers, lyophilized, and then resuspended in 50 mM sodium acetate buffer, pH 5.0 to a final concentration of 20 mg/mL. The soluble protein content was estimated according to Bradford's method, using bovine serum albumin as protein standard (Bradford, 1976).

The protein profile was assessed by 15% polyacrylamide gel electrophoresis in the presence of SDS (SDS-PAGE) as described by Laemmli (1970). Protein samples were dissolved in sample buffer [0.0625 M Tris buffer (pH 6.8) containing 2% SDS] and runs were

performed at 25 mA per gel at 25 °C for 2 h. Gels were stained with coomassie brilliant blue (R-350) solution.

The proteolytic activity was determined by zymogram containing 0.1% gelatin, following Freitas et al. (2007). After SDS-PAGE, the gels were incubated in 2.5% Triton X-100 for 30 min to remove all SDS and permit the protein renaturation. After, the gels were incubated in 50 mM sodium acetate buffer, pH 5.0, for 12 h at 37 °C. Proteolytic activity was detected as transparent bands in blue gels after staining with 0.1% Coomassie brilliant blue R-350.

2.3 Hydrolysis of bovine milk proteins

The *N. mirabilis* digestive fluid proteases were used to hydrolyze the bovine milk proteins as described by Oliveira et al. (2018). Briefly, aliquots of 50 µL of proteolytic extract (0.05 mg/mL of protein) were mixed with 450 µL of cow's milk proteins (whole proteins, purified caseins or whey proteins, at 10 mg/mL in 50 mM sodium acetate buffer, pH 5.0) at 37 °C for 24 h. The hydrolysis was monitored by 15% SDS-PAGE and the residual antigenicity of the hydrolysates was assessed by ELISA assay using anti-casein and anti-whey proteins polyclonal antibodies produced in rabbits, according to Oliveira et al. (2019). The reaction was detected using p-nitrophenyl phosphate disodium (150 µL, 5 mg/mL) as the substrate, and the absorbance was measured at 405 nm. Enfamil Gentlease commercial formula (Mead Johnson Nutrition) was used as the control for partially hydrolyzed bovine milk proteins.

2.4. Analysis of protein sequences

The amino acid sequences of three *N. mirabilis* digestive fluid serine proteases (named SCP3, SCP20, and SCP47) were obtained from Rottloff et al. (2016). The three protein sequences were freely accessed from the NCBI protein database with the following accession numbers: KT209999 (SCP3), KT210000 (SCP20), and KT210001 (SCP47). The BLASTP

bioinformatics tool was used to compare the three protein sequences with other plant proteases (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>), as well as the molecular masses and isoelectric points were calculated using the EMBOSS Pepstats server (<http://www.bioinformatics.nl/emboss-explorer/>). All bioinformatics analysis were performed after the removal of signal peptides using SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP>). The DiANNA 1.1 webserver was used to predict the disulfide bonds (<http://clavius.bc.edu/~clotelab/DiANNA/>).

2.5. Phylogenetic tree

Different protein sequences representing all serine protease clans were used to classify SCP3, SCP20, and SCP47 (Supplementary Table 1). The sequences were obtained from the MEROPS database and the phylogenetic analysis was performed using MEGA X package (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). Phylogenetic reconstructions were inferred using the Maximum Parsimony with full deletion option for gap treatment. The Bootstrap analysis consisted of 500 replicates (Kumar, Stecher, Li, Knyaz, & Tamura, 2018).

2.6. Molecular modeling

The three-dimensional (3D) models of SCP3, SCP20, and SCP47 were obtained by the Swiss-Model server (<http://swissmodel.expasy.org>), using two serine proteases from yeast (PDB 1YSC; PDB 1WPX) as templates, because they showed the highest identity with the input sequence. The structural and stereochemical quality of all models was checked by programs PROCHECK version 3.5 (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) and Molprobit version 4.4 (<http://molprobit.biochem.duke.edu/>) (Chen et al., 2010). The 3D structures were visualized using PyMOL software (<https://pymol.org/2/>).

2.7. Heterologous expression, purification, and proteolytic activity of rSCP3

Because the *N. mirabilis* digestive fluid presented a minimal amount of proteases, further assays were developed for the heterologous expression of one of these proteases. The protease chosen was SCP3 because some *in silico* analyses showed that it presented the best results for the prediction of protein solubility upon overexpression, when compared with SCP20 and SCP47 [SCRATCH, <http://scratch.proteomics.ics.uci.edu/>, Protein-Sol, <https://protein-sol.manchester.ac.uk/>, and EMBOSS Pepstats, https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/] (Supplementary Table 2).

For heterologous expression, firstly, the nucleotide sequence of SCP3 was reviewed to optimize the codons more frequently used by *E. coli*. After, its nucleotide sequence was synthesized by GenOne Biotechnologies (São Paulo, Brazil, <http://www.geneone.com.br/>), and then cloned in the pET-His₆-SUMO expression vector, which was used to transform *E. coli* Shuffle@T7 competent cells (New England Biolabs) by electroporation. Thus, the resultant plasmid pET-His₆-SUMO-SCP3 encoded a sequence containing six histidine residues in frame with SUMO tag and the synthetic gene of SCP3. His₆-SUMO tag was used to help in the solubilization and purification of recombinant protein (Silva et al., 2020).

The transformed *E. coli* cells were cultivated in liquid Luria-Bertani (LB) medium containing kanamycin (50 µg/mL) at 37 °C, as described by Silva et al. (2020). Briefly, after culture reaches the OD_{600nm} = 0.5, at 200 rpm, the expression of the recombinant protein was induced by different IPTG concentrations (0.1, 0.3 and 0.5 mM). Different additives were also added during the cell growth to help to express the protein in its soluble form. The first experiment used ethanol (1 or 2%, v/v), according to Chhetri, Kalita, & Tripathi (2015), whereas second protocol used 75 mM trehalose, 100 mM proline, and 50 mM mannitol as Leibly et al. (2012). After 18 h at 16 °C, at 200 rpm, the cells were treated with 0.5% Triton X-100 and lysozyme (100 µg/mL), followed by sonication at 50 kHz, 30 s pulse-on, 30 s pulse-

off (7 times). Soluble and insoluble materials were separated by centrifugation (12.000x g, 30 min, at 4 °C) and analyzed by 15% SDS-PAGE, as described by Laemmli (1970). The identification of recombinant SCP3 (rSCP3) in gels was performed by western blotting using polyclonal antibodies raised in rabbit against the purified SUMO tag, as described by Freitas et al. (2015).

rSCP3 was purified using a 1 mL HisTrap HP affinity column coupled to an AKTA chromatography system (São Paulo, Brazil, GE Healthcare), according to Silva et al. (2020). After washes with 50 mM Tris-HCl buffer (pH 8.0) containing 150 mM NaCl and 6 mM imidazole, the proteins were eluted with 50 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl and 50 mM or 250 mM imidazole. After dialysis against distilled water and lyophilization, the protein profile of each peak was determined by 15% SDS-PAGE. Furthermore, purified rSCP3 was incubated with ULP1 (Ubiquitin-Like-specific Protease 1) (1:20, enzyme:substrate ratio) for 16 h at 4 °C. ULP1, also known as SUMO protease, is a protease from *Saccharomyces cerevisiae* that cleaves in a highly specific manner between the SUMO tag and the heterologous protein (Müller, Hoegge, Pyrowolakis, & Jentsch, 2001). The ULP1 protein was subcloned into the pET-28a-SUMO vector and also transformed into the *E. coli* Shuffle®T7 competent cells. The ULP1 was expressed at a temperature of 22 °C for 5 h and it was obtained in its soluble form and active. The purity of the ULP1 protein was checked by SDS-PAGE as a single band with an apparent molecular mass of 32 kDa (data not shown).

Finally, rSCP3, without SUMO tag, was purified using a HisTrap HP affinity column, dialyzed against distilled water and used in further experiments. The *in gel* and *in vitro* proteolytic activity of the rSCP3 were investigated using 0.1% gelatin and 1% azocasein, respectively, as described by Freitas et al. (2007).

2.8. *In silico* evaluation of allergenic epitopes in SCP3

In silico tools were used to estimate the allergenicity of SCP3 comparing it to other known allergens present on Structural Database of Allergenic Proteins (SDAP) (<https://fermi.utmb.edu/>). The searches were performed according to the recommendations of FAO/WHO (2001), matching six contiguous amino acid residues with all the sequences of allergenic proteins deposited in the SDAP database (Ivanciuc, Schein, & Braun, 2003). For comparative purposes, the same study was performed using human trypsin. The predicted allergenic epitopes were labeled in the three-dimensional structures of SCP3 and trypsin using the PyMOL program (<http://pymol.org/>). Moreover, the cleavage sites of SCP3 by digestive enzymes pepsin, trypsin and chymotrypsin were evaluated using the ExPASy PeptideCutter tool (https://web.expasy.org/peptide_cutter/).

3 RESULTS AND DISCUSSION

3.1. Protein and proteolytic profile

The soluble protein content in the digestive fluid from *N. mirabilis* was negligible (0.05 mg/mL). Similarly, Athauda et al. (2004) described that *N. distillatoria* digestive fluid had around 0.06 mg/mL of protein. In contrast, Miguel et al. (2019) reported that different plants from the *Nepenthes* genera were able to produce up to 0.2 mg/mL of protein in their digestive fluids, among them aspartic proteases, cysteine proteases, serine carboxyproteases and prolyl-endoproteases. Likely, the low protein content found for *N. mirabilis* was due to the absence of an inductor as the chitin (Hatano & Hamada, 2012) or because the digestive fluid was collected from recently opened pitchers (Rottloff et al., 2016). We used the digestive fluids from freshly opened pitchers to avoid any contamination with microbial proteases.

SDS-PAGE showed an unique protein band of around 28 kDa in the *N. mirabilis* digestive fluid (Fig. 1A). In contrast, Rottloff et al. (2016) showed that *N. alata*, *N. sanguinea*, *N. albomarginata* and *N. bicalcarata* digestive fluids are rich sources of proteins with apparent molecular masses from 10 to 250 kDa, including carboxyproteases, peroxidases, galactosidases, phosphatases, chitinases, lipid transfer protein, glucosidases, among others. In addition, Lee et al. (2016) identified 36 proteins in *N. ventrata* digestive fluid, which had molecular masses ranging from 25 to 70 kDa, including acid phosphatases, serine proteases, chitinases, and β -1,3-glucanases.

Two bands with proteolytic activity (28 kDa and 52 kDa) were detected in gel containing gelatin at pH 5.0. Protease of 28 kDa has not been described in literature yet, while the band with higher proteolytic activity (52 kDa) can be the result of synergism among the three serine proteases SCP3, SCP20 and SCP47, which have theoretical molecular masses of around 52-55 kDa (Rottloff et al., 2016). Because they were not seen in SDS-PAGE, we assume that SCP3,

SCP20 and SCP47 are present in very low quantity in the digestive fluid. Even though the concentration of these enzymes is not high, their proteolytic activity are enough to efficiently digest the proteins of their preys. Interestingly, to overcome this issue of low concentration, proteases are synthetized in several isoforms in carnivorous plant digestive fluids. According to Stephenson & Hogan (2006), at least six different protease isoforms (124, 65, 61, 49, 36, and 33 kDa) were found in the *N. ventricosa* digestive fluid. Since these enzymes are present in very low concentration, little is also known about their synthesis, regulation, induction during development stages of the pitcher or response to catching and digestion of prey (An et al 2002; Buch et al., 2015).

3.2. Hydrolysis of bovine milk proteins

As shown in Fig. 1B, even at very small concentration (0.05 mg/mL), the *N. mirabilis* digestive fluid proteases were able to extensively hydrolysis the caseins, resulting in a very low residual antigenicity of the hydrolysates (12%) when compared to nonhydrolyzed milk proteins (100%). This value (12%) was very similar to that from a partially hydrolyzed commercial formula (8%) used as standard (Fig. 1B). The extensive hydrolysis of the caseins is required to reduce milk allergenicity, since these proteins are reported as the most abundant and major allergens in bovine milk (Docena et al., 1996). In contrast, the whey proteins were only partially hydrolyzed (Fig. 1B). Although immunoreactive peptides were still detected in the hydrolysates (65% of residual allergenicity), the result was very similar to commercial hypoallergenic formula (Enfamil Gentlease, Mead Johnson Nutrition) (66%). Likewise, proteases from other plant extracts were unable to eliminate the full allergenicity of whey proteins (Oliveira et al., 2019). The resistance to proteolysis of whey proteins (α -lactalbumin and β -lactoglobulin) is reflected by their compact structures, which are stabilized by disulfide bonds. Therefore, additional processing, such as heat pretreatment or pressurization, has allowed a significant enhance of hydrolyzing of whey proteins (Bu et al., 2009; Chicón et al., 2009).

3.3. Analysis of protein sequences

Some theoretical biochemical characteristics of the three serine carboxypeptidases (SCP3, SCP20 and SP47) are shown in Table 1. The three proteases have at least eight conserved cysteine residues, which are involved in the formation of four disulfide bonds, as reported by Rottloff et al. (2016). The theoretical pI values and molecular masses of SCP3, SCP20 and SP47 were 4.10 and 52.51 kDa, 5.78 and 52.41kDa, and 5.02 and 55.30 kDa, respectively (Table 1). Similarly, Lee et al. (2016) reported that the majority of the digestive proteins from *Nepenthes* species are acidic and more than 60% of the identified proteins had pI < 6.0, which is coherent with the acidic nature of the digestive fluid. These values of pI and molecular mass are also in concordance with those of most carnivorous plant proteases (Hatano & Hamada, 2008; Rottloff et al., 2016). These biochemical characteristics were also quite similar to other digestive fluid proteases, such as those from *N. mirabilis*, *N. alata* and *N. sanguinea*, in which the mature amino acid sequences ranged from 417 to 496 amino acids, molecular masses from 45.7 to 55.3 kDa and pI from 4.3 to 5.6 (Rottloff et al., 2016).

The multiple alignment of the open reading frames of the three serine carboxypeptidase sequences from *N. mirabilis* (SCP3, SCP20 and SP47) with other plant serine proteases (such as serine carboxypeptidase from *Herrania umbratica*, *Morus notabilis*, *Spinacia oleracea*, and *Durio zibethinus*) shows some conserved sequences in serine carboxypeptidases belonging to S10 family, which are represented by four boxes (box 1: GGPGCSS, box 2: GESYAG, box 3: GDXDXXV/C, box 4: DEXH) (Fig. 2). These conserved amino acid residues are also crucial to the enzymatic activity of these proteins, and thus they have been used as markers to categorize serine proteases in the group of chymotrypsin-like, subtilisin-like and α - and β -hydrolases (Krem & Di Cera, 2001).

The alignment also revealed that SCP3 was more similar to SCP47 than to SCP20. The change of amino acid class was the main responsible for these differences. For example, in

SCP3 and SCP47 there is a positively charged amino acid (Arg37), while in SCP20, at the same position, there is a negatively charged amino acid (Glu37) (Fig. 2). Despite differences, the spatial arrangement of the catalytic triad composed by serine (box 2), aspartic acid (box 3), and histidine (box 4) (Barrett, 1999) remained without alteration in SCP3, SCP47 and SCP20, similar to other plant serine proteases (Rawling et al, 2018). The sequence analyses also showed that the three proteases have a conserved glutamic acid residue prior to the catalytic serine, which interacts with histidine (present in the active site) via carboxylic group substrates and may be involved in the catalytic action (Jung, Ueno, Hayashi, & Liao, 1995).

3.4. Phylogenetic tree

The MEROPS database groups the proteases into clans, based on the similarity of primary sequence and tertiary structures as well as similarity of adjacent residues to the catalytic site (Rawlings & Barrett, 1999). Each clan is represented by more than one family and it is identified with two capital letters. The first letter represents the catalytic type of the protease family (i.e. serine, cysteine, aspartic, metallo) followed by an arbitrary second. The serine proteases are grouped in twelve clans (SB, SC, SE, SF, SH, SJ, SK, SO, SP, SR, SS, ST) and 54 families. The main differences among them are in the catalytic mechanism, proteolytic activity or protein fold (Rawlings et al., 2018). Therefore, in order to classify SCP3, SCP20 and SCP47, a phylogenetic analysis was performed using other proteases belonging to all twelve clans. Based on that analysis, SCP3, SCP20 and SP47 were classified as serine proteases belonging to Clan SC. The clan SC includes both endoproteases and exoproteases, and it is represented by seven families with different catalytic functions (S9-prolyl oligoprotease, S10-carboxyprotease Y, S15-Xaa-Pro dipeptidyl-protease, S28-lysosomal Pro-Xaa carboxyprotease, S33-prolyl aminoprotease, S37-PS-10 protease, S82-autocrine proliferation repressor protein A) (Rawlings et al., 2018). These families have representatives of both

prokaryotic and eukaryotic taxonomic domains, such as: *Sus scrofa*, *Saccharomyces cerevisiae*, *Lactococcus lactis*, *Homo sapiens*, *Neisseria gonorrhoeae*, *Streptomyces lividans*, and *Dictyostelium discoideum*. SCP3, SCP20 and SCP47 were included in S10 family, which differ from most other serine proteases families because these enzymes are only active at acidic pH (Page & Di Cera, 2008). The specific site of hydrolysis is also used to classify these enzymes. The carboxyproteases present in family S10 show preference for hydrophobic residues and basic amino acids on either side of the scissile bond, although exhibit broad substrate specificity such as peptides, esters, amides and anilides (Hayashi, 1976; Rawlings et al., 2018).

3.5. Three-dimensional (3D) model

Fig. 4 shows the 3D models of the three serine carboxyproteases generated by homology modeling. As pointed out, the catalytic triad (Asp, His and Ser) was conserved between the three proteases (SCP3, SCP20 and SP47), similar to other plant serine proteases, as follows: SCP3 (Asp³⁹³, His⁴⁵⁰, Ser²⁰⁵), SCP20 (Asp³⁹⁶, His⁴⁴⁹, Ser¹⁵⁹), and SCP47 (Asp⁴⁰⁴, His⁴⁵⁷, Ser²¹⁶) (Fig. 4D). Serine carboxyproteases are proteolytic enzymes that cleave the C-terminal peptide bond of polypeptides. They have been found in higher plants, fungi and animals (Breddam, 1986) and act through the nucleophilic attack of the hydroxyl group present in the reactive serine residue to the carbon of the carbonyl group of the substrate. The histidyl residue acts as a base, generating a tetrahedral intermediate and an imidazolium ion. The intermediate decomposes, through general acid catalysis, into an acyl-enzyme, an imidazole base and alcohol or amine, depending on the nature of the substrate (Bender & Kezdy, 1965). The main difference among the action mechanism of serine carboxyproteases and endoproteases is that the carboxyproteases release C-terminal amino acids from peptides, and the pKa of the catalytically essential histidyl residue is somewhat lower in the carboxyproteases than in the endoproteases (Breddam, 1986).

All three carboxypeptidases exhibited secondary structures common to those of Clan SC, such as ten parallel β -sheets surrounded by ten α -helices (Fig. 4). Substrate specificity is usually shaped by the presence and the length of loops between the third and the eighth core β -strands (Lazniewski et al., 2011). The members of clan SC are included in the 'alpha-beta hydrolase' fold family, which also includes proteases, lipases, esterases, dehalogenases, peroxidases and epoxide hydrolases, making it one of the most versatile and widespread protein folds known (Nardini & Dijkstra, 1999). According to the MEROPS database, the family S10 proteases contain 14 alpha helices and 11 beta sheets. Besides, the six central beta-strands are surrounded on either side by the 14 alpha-helices.

3.6. Heterologous expression, purification and proteolytic activity of rSCP3

The heterologous expression strategy was used to overcome the problem of low amount of the proteases. Similarly, most proteins of relevant industrial value are found in small amounts in their natural sources. Hence, prokaryotic systems have been used in order to express high levels of these proteins (Villaverde & Carrió, 2003). Once the expression system has been obtained, the heterologous production can be improved, increasing the protein production per cell and per unit of time (Sørensen & Mortensen, 2005). In this respect, SCP3 was chosen because some *in silico* analyzes predicted that it was a unique soluble protein upon overexpression in *E. coli*, when compared with SCP20 and SCP47 (Supplementary Table 2).

Fig. 5A shows that IPTG, at all concentrations tested, induced the overexpression of a 70 kDa protein band, which was expected to be the combined molecular mass for SCP3 (52 kDa) and SUMO tag containing His_{6x} (18 kDa). The expression of rSCP3 was confirmed by western blotting using polyclonal antibodies produced against SUMO tag (Fig. 5B). The optimal condition for the induction of rSCP3 was 0.5 mM IPTG at 16 °C for 18 h. Although the protein was expressed in high levels, it was obtained as an insoluble form (inclusion bodies)

(Fig. 5A). High-level expression of recombinant proteins in the prokaryotic systems can result in protein aggregation, and then inclusion bodies. A reduced environment of bacterial cytosol and a lack of post-translational machinery have been reported to contribute to this problem (Singh et al., 2015). The formation of inclusion bodies imposes a bottleneck in the production and purification of recombinant proteins. Therefore, additional steps are needed to solubilize or renature these proteins from inclusion bodies (Upadhyay et al., 2016; Santos & Souza, 2018). Different protocols were tested to obtain soluble rSCP3. According to Chhetri, Kalita & Tripathi (2015), the usage of ethanol during *E. coli* cultivation enhances the expression and solubilization of recombinant proteins. They proposed that ethanol can affect the cell environment by making changes in the cell membrane transport and fluidity. Ethanol is an amphipathic molecule and thus can modify lipid composition and assembly of membrane proteins, providing greater fold of poorly or less expressed recombinant proteins. Using this procedure, rSCP3 was successfully produced in its soluble form adding 1% or 2% ethanol during the induction with 0.5 mM IPTG (Fig. 5C). Another protocol to obtain the soluble rSCP3 was performed by adding different additives during cell growth, such as 75 mM trehalose, 100 mM proline, and 50 mM mannitol, as described by Leibly et al. (2012). According to them, the environmental conditions promoted by these additives can avoid the unfolding and aggregation of proteins, stabilizing it. Many of these additives are osmolytes, which are protein stabilizing molecules and chemical folders in nature. Its general stabilization mechanism is by excluding the protein hydration layer, increasing the energy needed to denature the proteins during cell lysis or by increasing the thermal stability of proteins, changing the melting temperature of these proteins when present in the protein buffer. Fig. 5C shows that among the three additives used, mannitol was the best one for the induction of the soluble rSCP3. However, when the expression experiment was scaled up, using 2 L of cultivation medium, the usage of mannitol was impracticable due to its high cost when compared to ethanol. Therefore, 1% ethanol was

used to further assays for expression, and then purification and characterization of rSCP3. The purification of rSCP3 was performed using a HisTrap HP affinity column containing 50 mM Tris-HCl buffer (pH 8.0), 150 mM NaCl and 6 mM imidazole. The bound proteins were eluted with 50 mM Tris-HCl buffer (pH 8.0), 300 mM NaCl, containing 50 mM or 250 mM imidazole (Fig. 5 E, F). After, purified rSCP3 (eluted with 50 mM imidazole) was incubated with ULP1 (Protease SUMO), for cleavage of SUMO tag. SDS-PAGE showed a band of 52 kDa, representing the purified rSCP3 (rSCP3*), without SUMO tag (Fig. 5F), but a protein band of 70 kDa was also displayed, because the cleavage of SUMO tag by SUMO protease was not complete (Fig. 5F). Finally, the proteolytic activity of the rSCP3, without SUMO tag (rSCP3*), was evaluated by testing its ability to digest the gelatin and azocasein (both non-specific substrates for proteases). The zymogram showed that rSCP3* did not exhibit proteolytic activity (Fig. 5G) as well as no activity was detected *in vitro* using azocasein as substrate (Fig. 5H).

When massively expressed in bacteria, recombinant proteins often tend to misfold and accumulate as soluble and insoluble nonfunctional aggregates (De Marco et al., 2005). Creighton (1986) explained that in a reducing environment as in the *E. coli* cytoplasm, the disulfide bonds are not correctly formed. Therefore, for proteins in which the disulfide bonds are critical for correct folding and then activity, this result in denatured inactive protein. rSCP3 has four disulfide bonds and because of that, the *E. coli* Shuffle®T7 was chosen to generate an oxidative cytoplasm besides disulfide bond isomerase capacity (Lobstein et al., 2012). In addition, a SUMO tag was fused to synthetic gene SCP3 resulting in plasmid pET-His6-SUMO-SCP3. SUMO fusion technology usually promotes correct folding and structural stability of the fusion protein (Panavas, Sanders, & Butt, 2009). Lee et al. (2008) reported that the SUMO system is widely used to facilitate the recombinant expression of difficult-to-express proteins. They expressed the RecA protein family in pSUMO-RecA vector in *E. coli* prokaryotic system,

and purified a soluble and catalytically active protein after cleavage by an engineered SUMO protease, with final yield ~10 mg of protein per liter of cell culture. In contrast, Silva et al. (2020) expressed an inactive protease from *Calotropis procera* latex in *E. coli* (rCpCP3). They attributed the inactivity to protein misfolding due to high amounts of disulfide bridges present in the protein. The production of recombinant proteins containing disulfide bond is a challenging task in heterologous expression. Otherwise, Kadek et al. (2014) expressed high amounts of nepenthesin-1, which is an aspartic protease from *Nepenthes gracilis* carnivorous plant, using *E. coli*. That recombinant protease showed optimal activity around pH 2.5. Similar results were obtained by Ishisaki et al. (2012), who reported a soluble and active class IV chitinase from *Nepenthes alata* (NaCHIT1) after the digestion of the fusion protein by HRV3C protease, even with three disulfide bonds in the tridimensional structure. Therefore, other protocols should be tested to express SCP3 in its soluble and active form.

3.7. *In silico* evaluation of allergenic epitopes in SCP3

In silico analysis of the complete amino acid sequence of SCP3 revealed no significant similarity with sequences of other toxic or antinutritional proteins deposited in SDAP protein databases. The SDAP database predicted only six peptides (PVVLWL; AIGNGLT; NAGHMVP; LTGGPG; YITGES; GIPALL) from SCP3, which were similar to allergens of insects, fungi or foods (wheat, lentil) (Api m 9, Len c 3, Asp f 12, and Tri a glutenin) (Supplementary Table 3). Interestingly, five peptides (PYQVSL; HFCGGS; DSCQGDSGGPVV; GGDSCQGDSGGPVV; GVVSWG) from human trypsin matched with eleven allergens (Blo t 3, Der f 3, Tyr p 3, Der p 3, Bom p 4, Eur m 3, Can f 5, Der p 9, Blo t 6, Der f 6, Api m 7). Allergenic epitopes on the surface of SCP3 and trypsin structures can be seen in Fig. 6A. According to FAO / WHO (2001), research with eight contiguous amino acids needs to be carried out so that the allergenic potential of proteins is also evaluated through research

with small segments of amino acids. It was observed that the length of the allergenic peptides in SCP3 was lesser than eight amino acids, making it difficult to be recognized by antibodies. Therefore, SCP3 could be used to hydrolyze food proteins with a very low or no residual allergy. On the other hand, trypsin had one possible allergenic peptide with 12 and another with 15 amino acids. Although trypsin has an allergenic potential higher than SCP3, there is no report about trypsin allergenicity. Yu & Ahmedna (2012) report the several applications of trypsin in food processing and food science research. Besides, this protease has been widely used in protein chemistry, proteomics and nutrition research. Accordingly, Bu et al. (2013) reported that the most used enzymes to hydrolyze milk proteins are commercial preparations of digestive enzymes (trypsin, chymotrypsin, pepsin) (Peñas et al., 2006), proteases from vegetable origin (papain) or microbiological (preparations Neutrase, Alcalase, Corolase) (Nakamura, Sado, Syukunobe, & Hirota, 1993).

The *in silico* digestion of SCP3 was performed using different digestive enzymes to simulate its digestion. The ExPASy PeptideCutter tool was used to predict cleavage sites under the action of enzymes pepsin, trypsin and chymotrypsin. Fig. 6B shows several catalytic sites for each enzyme in SCP3 sequence. This high susceptibility to these digestive enzymes supports the hypothesis of the low toxic potential of SCP3 by via oral (Codex, 2009). In the context of safety, Delaney et al. (2008) showed that International Food Biotechnology Committee recommends that the potential risk of new proteins should be assessed comparing its amino acid sequence with other toxic and allergenic proteins as well as its digestibility evaluated by different digestive enzymes.

4 CONCLUSION

This study suggests that proteases from *N. mirabilis* digestive fluid can be used as source of new biomolecules to hydrolysis bovine milk proteins, producing a hypoallergenic milk formula. The three major serine carboxypeptidases from *N. mirabilis* digestive fluid showed structural and functional characteristics similar to other plant serine proteases belonging to SC clan and S10 family. Moreover, the heterologous expression in *E. coli* of a serine carboxypeptidase (SCP3) was best achieved using 1% ethanol, 0.5 mM IPTG at 16 °C for 18 h. Even though the protein was successfully hydrolyzed from its SUMO tag and purified, it did not show proteolytic activity under the tested conditions. Shortly, more studies are need to reach the correct formation of disulfide bridges of SCP3 and then its correct refolding and proteolytic activity. Because SCP3 has low allergenic and toxic potentials, predicted by bioinformatics, the achievement of high expression of this protein in its active form can lead to its application in food industry to hydrolysis bovine milk proteins producing hypoallergenic formulas.

CONFLICT OF INTEREST

The authors confirm that the contents of this article pose no conflicts of interest.

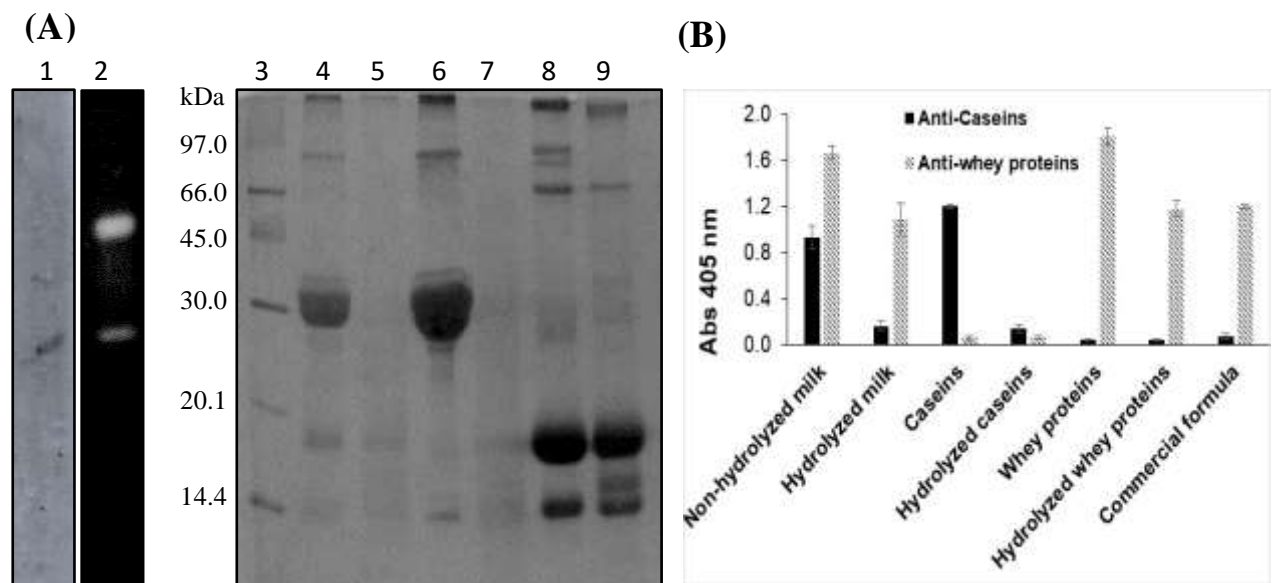
CONTRIBUTIONS

CTMNP, SFF, TLF, JSSB and CDTF performed the main research work, including digestive fluid protease purification, gel electrophoresis, proteolytic activity assays and hydrolysis assays. JPBO, MVR, and JSO produced polyclonal antibodies and performed ELISA. GPF and CSN helped in heterologous expression of rSCP3. PFNS supported with *in silico* evaluation of allergenic epitopes. All authors contributed to data analysis, discussion and writing of the manuscript.

ACKNOWLEDGMENTS

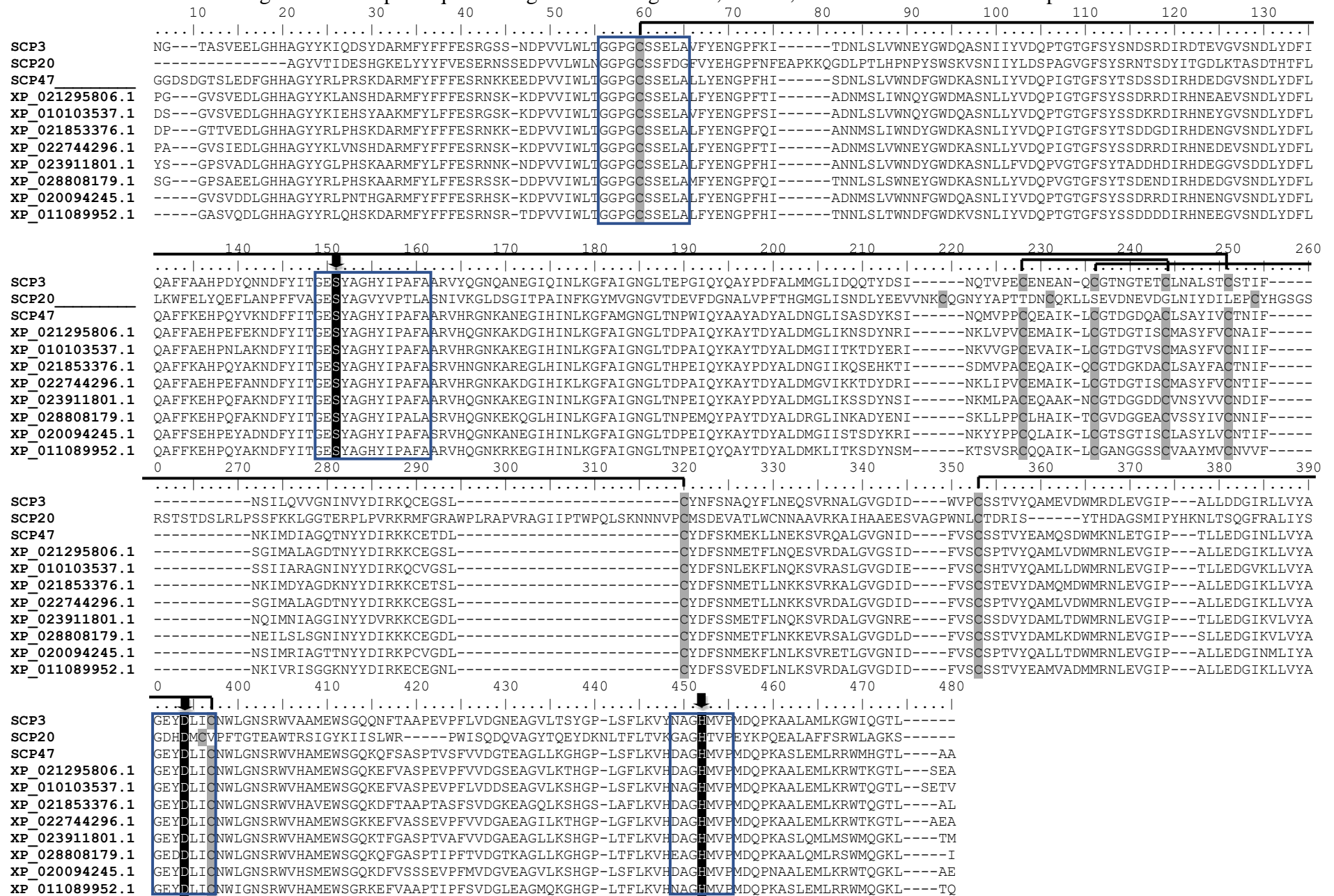
This work was supported by grants from the following Brazilian agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP). This study is part of the consortium “Molecular Biotechnology of Plant Latex”.

Figure 1 - Protein and proteolytic profile and bovine milk protein hydrolysis by serine carboxyproteases from *N. mirabilis* digestive fluid.



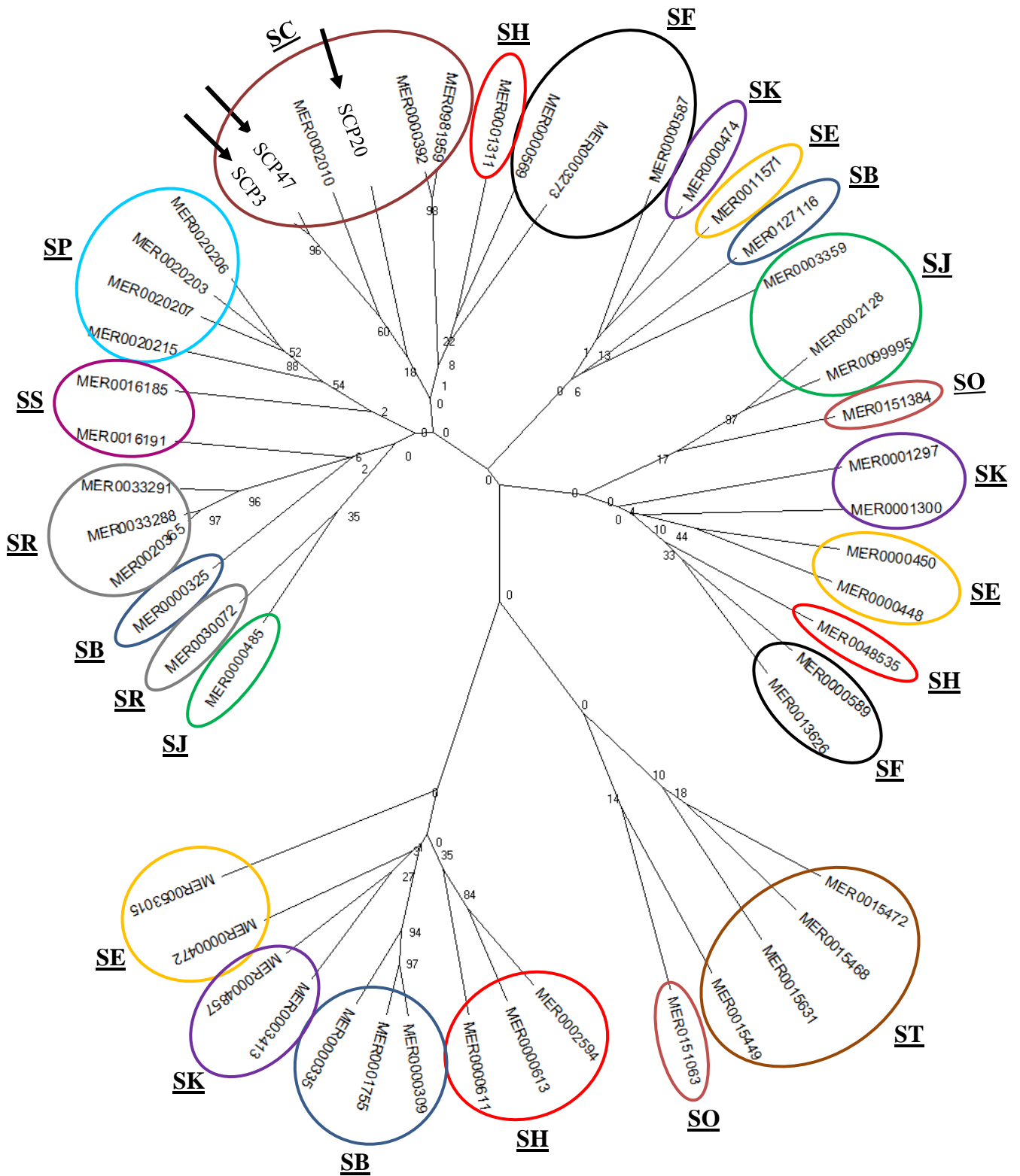
(A) Legend: (1) protein profile and (2) proteolytic activity of *N. mirabilis* digestive fluid by zymogram; (3) molecular mass markers; (4) non-hydrolyzed and (5) hydrolyzed bovine milk whole proteins; (6) non-hydrolyzed and (7) hydrolyzed purified caseins; (8) non-hydrolyzed and (9) hydrolyzed purified whey proteins. Assays were performed at 37 °C (pH 5.0) for 24 h. **(B)** Residual antigenicity of cow's milk proteins hydrolyzed by serine carboxyproteases from *N. mirabilis* digestive fluid measured by ELISA using polyclonal antibodies against caseins and whey proteins. Enfamil® Gentlease formula (Commercial formula) was used as the control for partially hydrolyzed cow's milk.

Figure 2 - Multiple sequence alignment among SCP3, SCP20, SCP47 and different serine proteases.



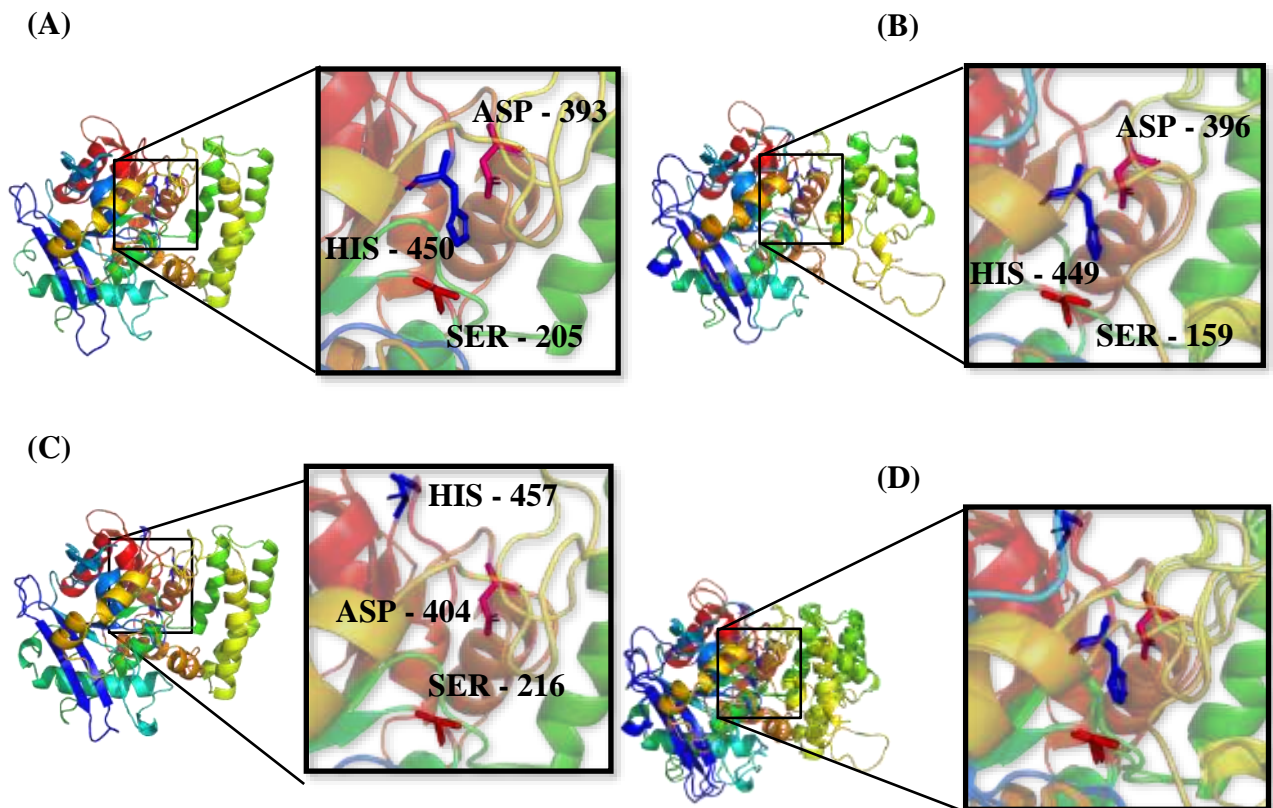
The catalytic triad are highlighted in black and by arrows. The cysteine residues involved in disulfide bonds are highlighted in gray.

Figure 3 - Phylogenetic relationship of SCP3, SCP20 and SCP47 with other serine proteases belonging to twelve different clans.



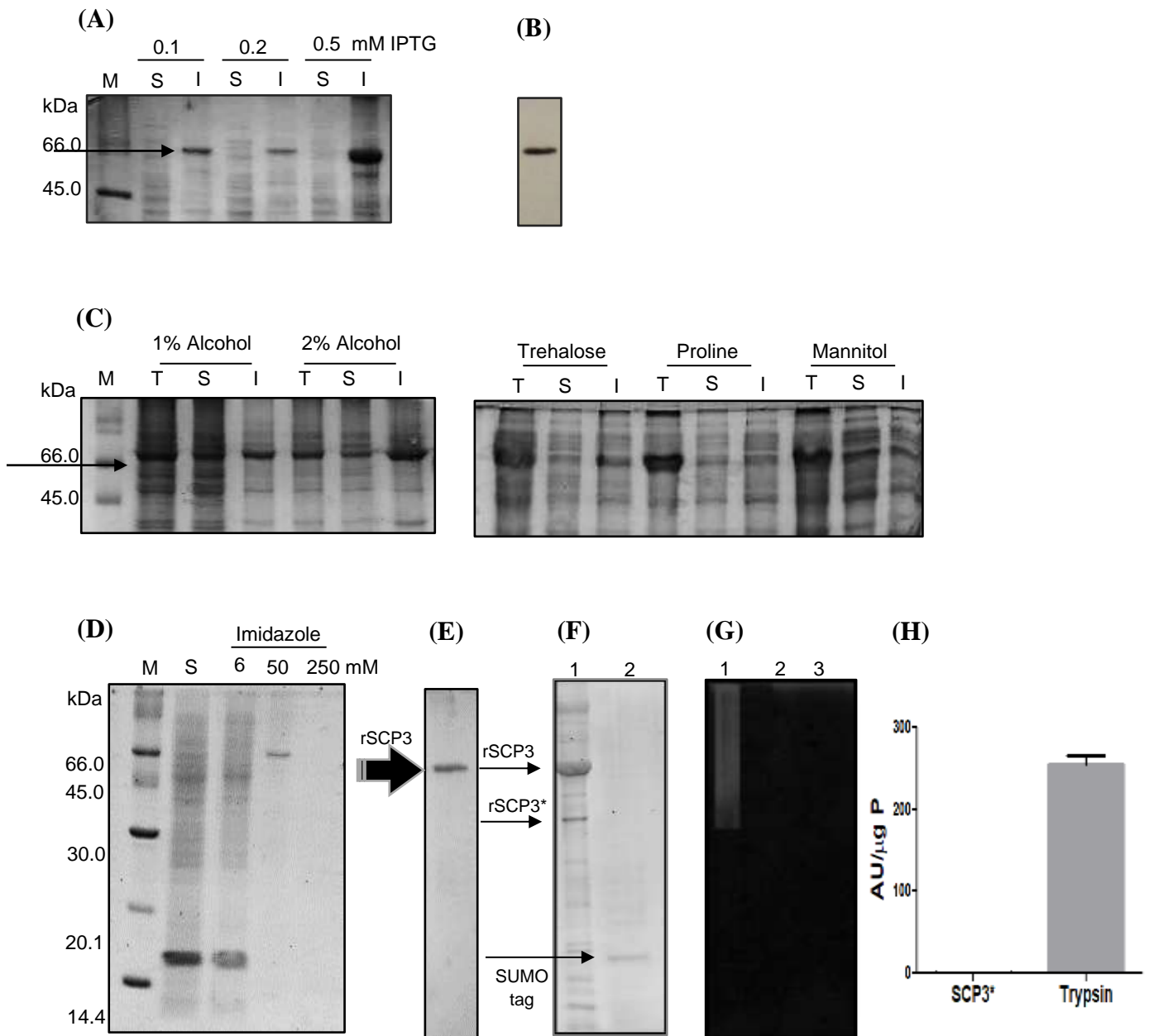
The proteins are represented by their accession numbers and each clan is displayed by circles of different colors (see Supplementary Table 1). SCP3, SCP20 and SCP47 are indicated by arrows.

Figure 4 - Three-dimensional models of *Nepenthes mirabilis* digestive fluid carboxypeptidases showing their active sites.



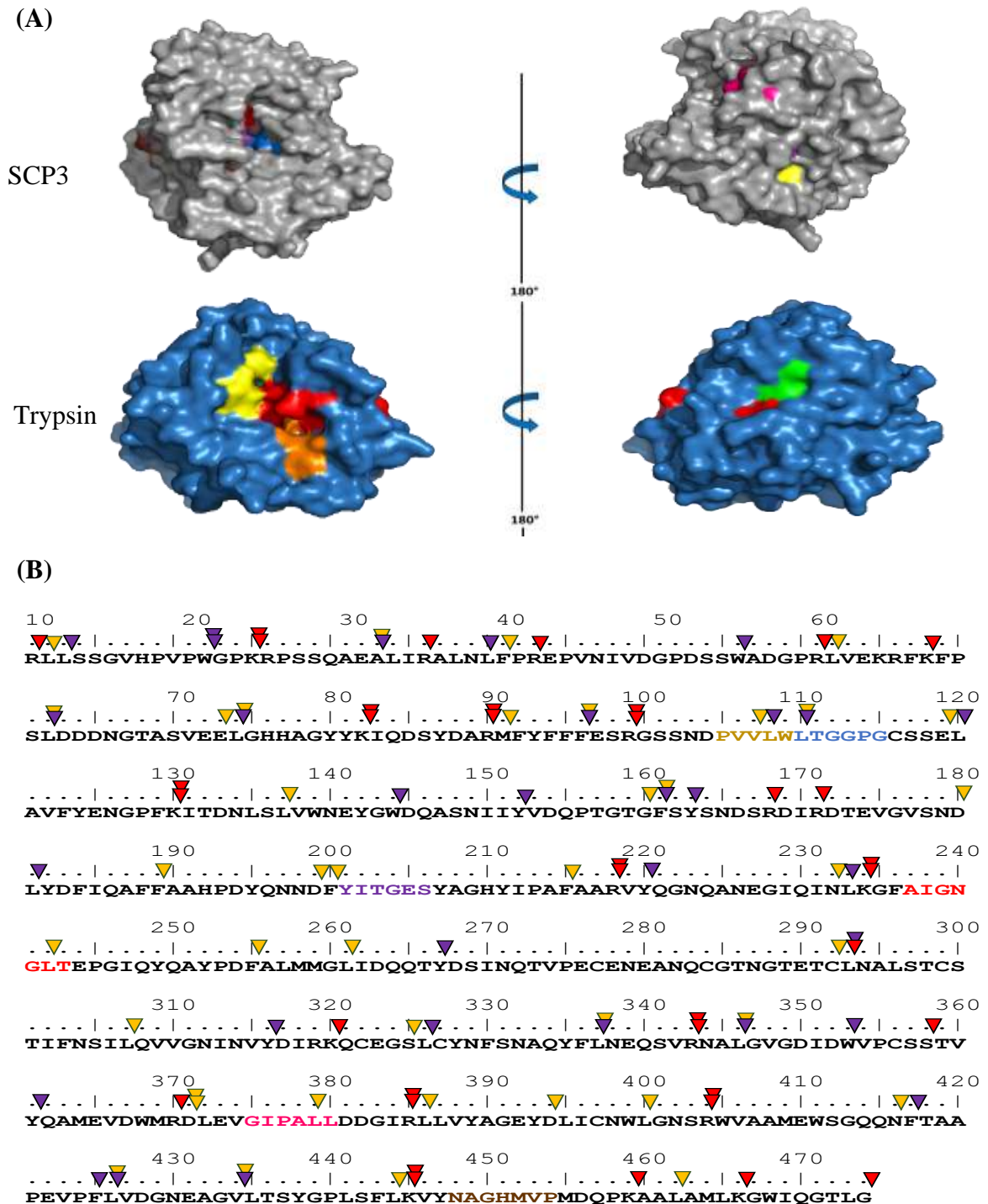
(A) SCP3, (B) SCP20, (C) SCP47 and (D) overlap of SCP3, SCP20 and SCP47.

Figure 5 - Expression, purification and proteolytic activity of recombinant SCP3 (rSCP3).



(A) Effect of different concentration of IPTG; (B) Western Blot analyses using polyclonal antibodies against SUMO tag. (C) Effect of different additives on rSCP3 expression. (D) Purification of rSCP3 by affinity chromatography. (E) Purified rSCP3. (F) 1- Purified rSCP3 digested by SUMO protease, without SUMO tag (rSCP3*). 2- SUMO tag used as control. (G) Proteolytic activity of rSCP3 by zymogram containing gelatin. 1- Trypsin (positive control). 2: Purified rSCP3 with and (3) without SUMO tag. (H) In vitro proteolytic activity using azocasein as substrate and purified rSCP3 without SUMO tag (rSCP3*). Legend: M Molecular mass markers. T: total protein fraction. S: soluble fraction; I: insoluble fraction.

Figure 6 - Allergenic epitope superimposition on the surface of the SCP3 structure.



(A) Predicted allergenic peptides on three-dimensional model of SCP3 and trypsin. SCP3: putative allergenic peptides were highlighted in yellow (PVVLW), blue (LTGGPG), purple (YITGES), red (AIGNGLT), pink (GIPALL), and brown (NAGHMVP). Trypsin: putative allergenic peptides were highlighted in green (PYQVSL), yellow (HFCGGS), red (GGKDSCQGDSGPPVV) and orange (GVVSWG). The colors grey (SCP3) and blue (trypsin) represent the non-allergenic amino acid sequences. (B) Predicted cleavage sites in SCP3 by Pepsin (pH1.3) (▼), Chymotrypsin-low specificity (not before P) (▼) and Trypsin (▼) using the PeptideCutter program.

Table 1 - Some biochemical characteristics of three serine carboxyl peptidases from *Nepenthes mirabilis*.

Peptidase	Protein length	Molecular mass*	Isoelectric point*	Cysteine residues*	Disulphide bonds (amino acids)**	BLASTP best match*** (Protein/Species/ Accession number)	Identity (E-Value)
NmSCP3 (AMN148 54.1)	474	52518.47	4.10	9	4 (116–299;277–292; 284–356; 327–397)	Serine carboxypeptidase-like/ <i>Herrania umbratica</i> / XP_021295806.1	71% (0)
NmSCP20 (AMN148 55.1)	473	52414.83	5.78	8	4 (64–362;228–240;263–329;339–399)	Serine carboxypeptidase-like 20/ <i>Chenopodium quinoa</i> / XP_021733488.1	75% (0)
NmSCP47 (AMN148 56.1)	496	55309.10	5.02	9	4 (127–288;295–303;310–367;338–408)	Serine carboxypeptidase-like / <i>Spinacia oleracea</i> / XP_021853376.1	76% (0)

*The EMBOSS Pepstats server was used to compute the theoretical molecular weight, isoelectric point, and cysteine residues (<http://www.bioinformatics.nl/emboss-explorer/>).

** The DiANNA 1.1 webserver was used to predict the disulfide bonds (<http://clavius.bc.edu/~clotelab/DiANNA/>).

*** The BLASTP was used to compare the protein query sequence against a protein sequence database in NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>).

Supplementary Table 1 - List of all serine peptidases used to build the phylogenetic tree of the Fig. 3.

CLAN SB			
Accession number	Protein name	Family	Organism
MER0000309	subtilisin Carlsberg	S08	<i>Bacillus licheniformis</i>
MER0001755	mesentericopeptidase	S08	<i>Bacillus pumilus</i>
MER0000325	aqualysin 1	S08	<i>Thermus aquaticus</i>
MER0000335	endopeptidase K	S08	<i>Engyodontium album</i>
MER0003688	mycosin-1	S08	<i>Mycobacterium tuberculosis</i>
MER0020159	CspC peptidase	S08	<i>Clostridium perfringens</i>
MER0127116	scytalidolisin	S53	<i>Scytalidium lignicolum</i>
CLAN SC			
Accession number	Protein name	Family	Organism
MER0981959	peptide cyclase 1	S09	<i>Vaccaria hispanica</i>
MER0002010	carboxypeptidase Y	S10	<i>Saccharomyces cerevisiae</i>
AMN14854.1	putative serine carboxypeptidase type 3	S10	<i>Nepenthes mirabilis</i>
AMN14855.1	serine carboxypeptidase 20-like protein	S10	<i>Nepenthes mirabilis</i>
AMN14856.1	serine carboxypeptidase 47-like protein	S10	<i>Nepenthes mirabilis</i>
MER0000446	lysosomal Pro-Xaa carboxypeptidase	S28	<i>Homo sapiens</i>
MER0000431	prolyl aminopeptidase	S33	<i>Neisseria gonorrhoeae</i>
MER0001350	PS-10 peptidase	S37	<i>Streptomyces lividans</i>
CLAN SE			
Accession number	Protein name	Family	Organism
MER0000448	D-Ala-D-Ala carboxypeptidase A	S11	<i>Geobacillus stearothermophilus</i>
MER0000455	murein-DD-endopeptidase	S11	<i>Escherichia coli</i>
MER0000450	penicillin-binding protein 6	S11	<i>Escherichia coli</i>

MER0000472	D-Ala-D-Ala peptidase C	S13	<i>Escherichia coli</i>
MER0011571	D-Ala-D-Ala carboxypeptidase PBP3	S13	<i>Neisseria meningitidis</i>
MER0053015	penicillin-binding protein 4	S13	<i>Corynebacterium jeikeium</i>
CLAN SF			
Accession number	Protein name	Family	Organism
MER0000569	repressor LexA	S24	<i>Escherichia coli</i>
MER0000587	phage lambda repressor protein	S24	<i>bacteriophage lambda</i>
MER0000589	signal peptidase I	S26	<i>Escherichia coli</i>
MER0000597	mitochondrial inner membrane peptidase 1	S26	<i>Saccharomyces cerevisiae</i>
MER0003273	SpsB signal peptidase	S26	<i>Staphylococcus aureus</i>
MER0013626	PA1303 peptidase	S26	<i>Pseudomonas aeruginosa</i>
CLAN SH			
Accession number	Protein name	Family	Organism
MER0000613	cytomegalovirus assemblin	S21	<i>human herpesvirus 5</i>
MER0002594	herpesvirus 6-type assemblin	S21	<i>human herpesvirus 6</i>
MER0000611	Varicella zoster assemblin	S21	<i>human herpesvirus 3</i>
MER0150756	gpO peptidase	S73	<i>Enterobacteria phage P2</i>
MER0001311	prohead peptidase gp21	S77	<i>Enterobacteria phage T4</i>
MER0048535	prohead peptidase	S78	<i>bacteriophage T5</i>
CLAN SJ			
Accession number	Protein name	Family	Organism
MER0000485	Lon-A peptidase	S16	<i>Escherichia coli</i>
MER0041647	LonC peptidase	S16	<i>Thermus thermophilus</i>
MER0003359	protease La	S16	<i>Methanocaldococcus jannaschii</i>
MER0002128	avian infectious bursal disease birnavirus Vp4 peptidase	S50	<i>avian infectious bursal disease virus</i>
MER0099995	Tellina virus 1 VP4 peptidase	S69	<i>Tellina virus 1</i>

CLAN SK			
Accession number	Protein name	Family	Organism
MER0000474	peptidase Clp	S14	<i>Escherichia coli</i>
MER0004857	ClpP1 peptidase	S14	<i>Streptomyces coelicolor</i>
MER0003413	C-terminal processing peptidase-2	S41	<i>Scenedesmus obliquus</i>
MER0001297	CtpA peptidase	S41	<i>Synechocystis sp. PCC 6803</i>
MER0001300	sohB peptidase	S49	<i>Escherichia coli</i>
MER0001303	protein C	S49	<i>bacteriophage lambda</i>
CLAN SO			
Accession number	Protein name	Family	Organism
MER0151384	endosialidase CIMCD self-cleaving protein	S74	<i>Enterobacteria phage K1F</i>
MER0151063	neck appendage CIMCD self-cleaving protein	S74	<i>Bacillus phage GA-1</i>
CLAN SP			
Accession number	Protein name	Family	Organism
MER0020203	nucleoporin 145	S59	<i>Homo sapiens</i>
MER0020215	nuclear pore complex component Nup145	S59	<i>Saccharomyces cerevisiae</i>
MER0020212	At1g80680	S59	<i>Arabidopsis thaliana</i>
MER0020206	nucleoporin 98	S59	<i>Drosophila melanogaster</i>
MER0020207	NPP-10	S59	<i>Caenorhabditis elegans</i>
CLAN SR			
Accession number	Protein name	Family	Organism
MER0020365	lactoferrin	S60	<i>Homo sapiens</i>
MER0030072	hemiferrin	S60	<i>Rattus norvegicus</i>
MER0033288	serotransferrin precursor	S60	<i>Homo sapiens</i>
MER0033291	melanotransferrin domain 1	S60	<i>Homo sapiens</i>
CLAN SS			
Accession number	Protein name	Family	Organism

MER0016191	murein tetrapeptidase LD-carboxypeptidase	S66	<i>Pseudomonas aeruginosa</i>
MER0016185	microcin C7 immunity protein	S66	<i>Escherichia coli</i>
CLAN ST			
Accession number	Protein name	Family	Organism
MER0015449	rhomboid-1	S54	<i>Drosophila melanogaster</i>
MER0015472	Pcp1 peptidase	S54	<i>Saccharomyces cerevisiae</i>
MER0015468	GlpG peptidase	S54	<i>Escherichia coli</i>
MER0015631	At1g52580	S54	<i>Arabidopsis thaliana</i>

Supplementary Table 2 - Prediction of solubility of three *Nepenthes mirabilis* serine peptidases using different bioinformatic programs.

	Programs	Predicted Solubility upon Overexpression	
		Soluble Probability	Insoluble Probability
SCP3	Protein-Sol*	0.448	
	EMBOSS Pepstats**		0.530
	SCRATCH***	0.506	
SCP20	Protein-Sol		0.316
	EMBOSS Pepstats		0.218
	SCRATCH		0.549
SCP47	Protein-Sol		0.442
	EMBOSS Pepstats		0.424
	SCRATCH		0.812

*Protein-Sol: <https://protein-sol.manchester.ac.uk/>

**EMBOSS Pepstats: https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/

***<http://scratch.proteomics.ics.uci.edu/>

Supplementary Table 3 - Prediction of allergenic peptides in the SCP3 sequence using Structural Database of Allergenic Proteins program.

Allergens	Species	Peptides	Source
Api m 9	<i>Apis mellifera</i>	PVVLWL AIGNGLT NAGHMVP	Insect
Len c 3	<i>Lens culinaris</i>	LTGGPG	Foods
Asp f 12	<i>Ascomycota</i> <i>Eurotiales</i> ; <i>Aspergillus</i> <i>fumigatus</i>	YITGES	fungi (moulds)
Tri a glutenin	<i>Triticum aestivum</i>	GIPALL	Foods

Supplementary Table 4 - Prediction of allergenic peptides in the trypsin sequence using Structural Database of Allergenic Proteins program.

Allergens	Species	Peptides	Source
Blo t 3	<i>Blomia tropicalis</i>	PYQVSL CQGDSGGPVV	Mites
Der f 3	<i>Dermatophagoides</i> <i>farinae</i>	HFCGGS DSCQGDSGGPVV	Mites
Tyr p 3	<i>Tyrophagus</i> <i>putrescentiae</i>	HFCGGS DSCQGDSGG	Mites
Der p 3	<i>Dermatophagoides</i> <i>pteronyssinus</i>	GGKDSCQGDSGGPVV	Mites
Bom p 4	<i>Bombus</i> <i>pennsylvanicus</i>	KDSCQGDSGGP	Insect
Eur m 3	<i>Euroglyphus maynei</i>	DSCQGDSGGPV	Mites
Can f 5	<i>Canis familiaris</i>	GDSGGP	Animal
Der p 9	<i>Dermatophagoides</i> <i>pteronyssinus</i>	GDSGGP GVVSWG	Mites
Blo t 6	<i>Blomia tropicalis</i>	GDSGGP	Mites
Der f 6	<i>Dermatophagoides</i> <i>farinae</i>	GDSGGP	Mites
Api m 7	<i>Apis mellifera</i>	DSGGPV	Insect

5 CONCLUSÃO

Este estudo sugere que proteases do fluido digestivo de *N. mirabilis* podem ser usadas como fonte de novas biomoléculas para hidrólise de proteínas do leite bovino, produzindo uma fórmula de leite hipoalergênica. As três principais serino carboxiproteases do fluido digestivo de *N. mirabilis* apresentaram características estruturais e funcionais semelhantes a outras serino proteases vegetais pertencentes ao clã SC e à família S10. Além disso, a expressão heteróloga em *E. coli* de uma serino carboxiprotease (SCP3) foi melhor obtida com etanol a 1%, IPTG 0,5 mM a 16 °C por 18 h. Embora a proteína tenha sido hidrolisada com sucesso a partir da cauda SUMO e purificada, ela não mostrou atividade proteolítica nas condições testadas. Dessa forma, são necessários mais estudos para alcançar a formação correta de pontes dissulfeto em SCP3 e, em seguida, sua correta redobragem e atividade proteolítica. Como a SCP3 possui baixos potenciais alergênicos e tóxicos, preditos por bioinformática, a obtenção de alta expressão dessa proteína em sua forma ativa pode levar à sua aplicação na indústria de alimentos, para a hidrólise de proteínas do leite bovino produzindo fórmulas hipoalergênicas.

REFERÊNCIAS

- ADLASSNIG, W.; KOLLER-PEROUTKA, M.; BAUER, S.; LENDL, E.; LICHTSCHEIDL, I. Endocytotic uptake of nutrients in carnivorous plants. **The plant journal**, [s.l.], v. 71, n.2, p. 303-313, 2012.
- ANTAO, C. M.; MALCATA, F.X. Plant serine proteases: biochemical, physiological and molecular features. **Plant physiology and biochemistry**, [s.l.], v. 43, n. 7, p. 637–650, 2005.
- BANEYX, F.; MUJACIC, M. Recombinant protein folding and misfolding in Escherichia coli. **Nature biotechnology**, [s.l.], v. 22, n. 11, p. 1399-1408, 2004.
- BARRETT, A. J.; TOLLE, D. P.; RAWLINGS, N. D. Managing peptidases in the genomic era. **Biological chemistry**, [s.l.], v. 384, n. 6, p. 873–882, 2003.
- BEHIE, S. W.; BIDOCHKA, M.J. Insects as a Nitrogen Source for Plants. **Insects**, [s.l.], v. 4, n. 3, p. 413-424, 2013.
- BROEDEL, S. E.; PAPCIAK, S. M.; JONES, W. R. The selection of optimum media formulations for improved expression of recombinant proteins in E. coli. **Athena enzyme systems technical bulletin**, [s.l.], v. 2, n. 1, p. 1-8, 2001.
- BUTTS, C. T.; ZHANG, X.; KELLY, J. E.; ROSKAMP, K. W.; UNHELKAR, M. H.; FREITES, J. A.; MARTIN, R. W. Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, *Drosera capensis*. **Computational and structural biotechnology journal**, [s.l.], v. 14, p. 271–282, 2016.
- CUADRADO, C.; CABANILLAS, B.; PEDROSA, M. M.; VARELA, A.; GUILLAMÓN, E.; MUZQUIZ, M.; CRESPO, F. J.; RODRIGUEZ, J.; BURBANO, C. Influence of thermal processing on IgE reactivity to lentil and chickpea proteins. **Molecular nutrition & food research**, [s.l.], v. 53, n. 11, p. 1462-1468, 2009.
- CASAÑAL, A.; ZANDER, U.; DUPEUX, F.; VALPUESTA, V.; MARQUEZ, J. Purification, crystallization and preliminary X-ray analysis of the strawberry allergens Fra a 1E and Fra a 3 in the presence of catechin. **Acta crystallographica section F: structural biology and crystallization communications**, [s.l.], v. 69, n. 5, p. 510-514, 2013.
- CHOU, C. P. Engineering cell physiology to enhance recombinant protein production in Escherichia coli. **Applied microbiology and biotechnology**, [s.l.], v. 76, n. 3, p. 521-532, 2007.
- CHOUDHARY, R.B.; JANA, A.K.; JHA, M.K. Enzyme technology applications in leather processing. **Indian journal of chemical technology**, [s.l.], v. 11, n. 5, p. 659–671, 2004.
- CLARKE, C. M. *Nepenthes of Borneo*. **Kota Kinabalu: Natural history publications**, 1997.

CLARKE, C. M. *Nepenthes* of Sumatra & Peninsular Malaysia. **Kota Kinabalu: Natural history publications**, 2001.

CLARKE, C.M. *Nepenthes mirabilis*. **The IUCN red list of threatened species**: e.T49122515A21844202, 2014.

CODEX ALIMENTARIUS COMMISSION. Joint FAO/WHO Food Standards Programme, **Food and Agriculture Organization of the United Nations. Foods derived from modern biotechnology**, Roma, Italia, p. 1–33, 2009.

DARABI, M.; SEDDIGH, S. Bioinformatic characterization of aspartic protease (AP) enzyme in seed plants. **Plant systematics and evolution**, [s.I.], v. 301, n. 10, p. 2399-2417, 2015.

DARWIN, C. *Insectivorous plants*. London: Murray, 1875.

ELLISON A. M.; GOTELLI N. J. Evolutionary ecology of carnivorous plants. **Trends in ecology & evolution**, [s.I.], v. 16, n. 11, p. 623–629, 2001.

FIOCCHI, A.; SCHUNEMANN, H.; ANSOTEGUI, I.; ASSA'AD, A.; BAHNA, S.; CANANI, R. B.; BOZZOLA, M.; DAHDAH, L.; DUPONT, C.; EBISAWA, M.; GALLI, E.; LI, H.; KAMENWA, R.; LACK, G.; MARTELLI, A.; PAWANKAR, R.; SAID, M.; SÁNCHEZ-BORGES, M.; SAMPSON, H.; SHAMIR, R.; SPERGEL, J.; TERRACCIANO, L.; VANDENPLAS, Y.; VENTER, C.; WASERMAN, S.; WONG, G.; BROZEK, J. The global impact of the DRACMA guidelines cow's milk allergy clinical practice. **World allergy organization journal**, [s.I.], v. 11, n. 1, p. 2, 2018.

GAUME, L.; FORTERRE, Y. A viscoelastic deadly fluid in carnivorous pitcher plants. **PLoS One**, [s.I.], v. 2, n. 11, p. e1185, 2007.

GIACOMETTI, J.; BURETIĆ-TOMLJANOVIĆ, A. Peptidomics as a tool for characterizing bioactive milk peptides. **Food chemistry**, [s.I.], v. 230, p. 91-98, 2017.

GIVNISH, T.J.; BURKHARDT, E. L.; HAPPEL, R. E.; WEINTRAUB, J. D. Carnivory in the bromeliad *Brocchinia reducta*, with a cost/benefit model for the general restriction of carnivorous plants to sunny, moist nutrient-poor habitats. **The american naturalist**, [s.I.], v. 124, n. 4, p. 479-497, 1984.

GIVNISH, T.; SPARKS, W.; HUNTER, S.; PAVLOVIC, A. Why are plants carnivorous? Cost/benefit analysis, whole plant growth, and the context-specific advantages of botanical carnivory. **Carnivorous plants: physiology, ecology and evolution**, New York, chapter 18, Oxford University Press, p. 232-255, 2018.

GOLKAR, A.; MILANI, J. M.; VASILJEVIC, T. Altering allergenicity of cow's milk by food processing for applications in infant formula. **Critical reviews in food science and nutrition**, [s.I.], v. 59, n. 1, p. 159-172, 2019.

GONZÁLEZ-RÁBADE, N., BADILLO-CORONA, J. A., ARANDA-BARRADAS, J. S., & DEL CARMEN OLIVER-SALVADOR, M. Production of plant proteases in vivo and in vitro—a review. **Biotechnology advances**, [s.I.], v. 29, n. 6, p. 983-996, 2011.

GONZÁLEZ-MONTALBÁN, N.; GARCÍA-FRUITÓS, E.; VILLAVERDE, A. Recombinant protein solubility—does more mean better? **Nature biotechnology**, [s.I.], v. 25, n. 7, p. 718–720, 2007.

GRAY, L.; CLARKE, C.; WINT, G.; MORAN, J. Potential effects of climate change on members of the Palaeotropical pitcher plant family Nepenthaceae. **PLoS One**, [s.I.], v. 12, n. 8, p. 1–17, 2017.

GRUDKOWSKA, M.; ZAGDAŃSKA, B. Multifunctional role of plant cysteine proteinases. **Acta biochimica polonica**, [s.I.], v. 51, n. 3, p. 609–624, 2004.

GUO, R.; XU, X.; CAROLE, B.; LI, X.; GAO, M.; ZHENG, Y.; WANG, X. Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. **BMC genomics**, [s.I.], v. 14, n. 1, p. 554, 2013.

GUO, R.; ZHAO, J.; WANG, X.; GUO, C.; LI, Z.; WANG, Y.; WANG X. Constitutive expression of a grape aspartic protease gene in transgenic Arabidopsis confers osmotic stress tolerance. **Plant Cell, Tissue and Organ Culture (PCTOC)**, [s.I.], v. 121, n. 2, p. 275–287, 2015.

HEDSTROM, L. Serine protease mechanism and specificity. **Chemical reviews**, [s.I.], v. 102, n. 12, p. 4501–4524, 2002.

HEINE, R. G.; ALREFAEE, F.; BACHINA, P.; DE LEON, J. C.; GENG, L.; GONG, S.; MADRAZO, J. A.; NGAMPHAIBOON, J.; ONG, C.; ROGACION, J. M. Lactose intolerance and gastrointestinal cow's milk allergy in infants and children—common misconceptions revisited. **World allergy organization journal**, [s.I.], v. 10, n. 1, p. 41, 2017.

HUSSAIN, Z. A review on milk allergy. **Journal of nutritional health & food engineering**, [s.I.], v. 8, n. 1, p. 84–86, 2018.

JAFFE, K.; MICHELANGELI, F.; GONZALEZ, J.M.; MIRAS, B.; RUIZ, M.C. Carnivory in pitcher plants of the genus *Heliamphora* (Sarraceniaceae). **New phytologist**, [s.I.], v. 122, n. 4, p. 733–744, 1992.

JEBB, M.H.P.; CHEEK, M. A skeletal revision of *Nepenthes* (Nepenthaceae). **Blumea—biodiversity, evolution and biogeography of plants**, [s.I.], v. 42, n. 1, p. 1–106, 1997.

KADEK, A.; TRETYACHENKO, V.; MRAZEK, H.; IVANOVA, L.; HALADA, P.; REY, M.; SCHRIEMER, D. C.; MAN, P. Expression and characterization of plant aspartic protease nepenthesin-1 from *Nepenthes gracilis*. **Protein expression and purification**, [s.I.], v. 95, p. 121–128, 2014.

KORHONEN, H. Milk-derived bioactive peptides: From science to applications. **Journal of functional foods**, [s.I.], v. 1, n. 2, p. 177–187, 2009.

KROL, E.; PLACHNO, B. J.; ADAMEC, L.; STOLARZ, M.; DZIUBINSKA, H.; TREBACZ, K. Quite a few reasons for calling carnivores 'the most wonderful plants in the world'. **Annals of botany**, [s.I.], v. 109, n. 1, p. 47–64, 2012.

LACOU, L.; LE, S.; PEZENNEC, S.; GAGNAIRE, V. An in silico approach to highlight relationships between a techno-functional property of a dairy matrix and a peptide profile. **Colloids and surfaces A: physicochemical and engineering aspects**, [s.I.], v. 475, p. 44–54, 2015.

LEARY, D.; VIERROS, M.; HAMON, G.; ARICO, S.; MONAGLE, C. Marine genetic resources: a review of scientific and commercial interest. **Marine policy**, [s.I.], v. 33, n. 2, p. 183–194, 2009.

LESTARININGSIH, N.; SETYANINGSIH, D. Explorative study of tropical pitcher plants (*Nepenthes* sp.) types and insects that trapped inside in Sebangau National Park Palangka Raya Central Kalimantan. **Journal of physics: conference series**, [s.I.], v. 795, n. 1, IOP Publishing, 2017.

LEE, L.; ZHANG, Y.; OZAR, B.; SENSEN, C.W.; SCHRIEMER, D.C. Carnivorous Nutrition in Pitcher Plants (*Nepenthes* spp.) via an Unusual Complement of Endogenous Enzymes. **Journal of proteome research**, [s.I.], v. 15, n. 9, p. 3108-3117, 2016.

LOBSTEIN, J.; EMRICH, C.A.; JEANS, C.; FAULKNER, M.; RIGGS, P.; BERKMEN, M. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. **Microbial cell factories**, [s.I.], v. 11, n. 1, p. 753, 2012.

LORKOWSKI, G. Gastrointestinal absorption and biological activities of serine and cysteine proteases of animal and plant origin: review on absorption of serine and cysteine proteases. **International journal of physiology, pathophysiology and pharmacology**, [s.I.], v. 4, n. 1, p. 10–27, 2012.

MANUYAKORN, W.; TANPOWONG, P. Cow milk protein allergy and other common food allergies and intolerances. **Paediatrics and international child health**, [s.I.], v. 39, n. 1, p. 32-40, 2019.

MIGUEL, S.; HEHN, A.; BOURGAUD, F. *Nepenthes*: State of the art of an inspiring plant for biotechnologists. **Journal of biotechnology**, [s.I.], v. 265, p. 109-115, 2018.

MISHRA, S.S.; RAY, R.C.; ROSELL, C.M.; PANDA, D. Microbial enzymes in food. applications: history of progress. **Microbial enzyme technology in food applications**, CRC Press, [s.I.], p. 17–32, 2017.

MITHOFER, A. Carnivorous pitcher plants: insights in an old topic. **Phytochemistry**, [s.I.], v. 72, n. 13, p. 1678-1682, 2011.

MORAN, J. A.; CLARKE, C.; GREENWOOD, M.; CHIN, L. Tuning of color contrast signals to visual sensitivity maxima of tree shrews by three Bornean highland *Nepenthes* species. **Plant signaling & behavior**, [s.I.], v. 7, n. 10, p. 1267–1270, 2012.

MORAN, J. A.; CLARKE, C. M. The carnivorous syndrome in *Nepenthes* pitcher plants: current state of knowledge and potential future directions. **Plant signaling & behavior**, [s.I.], v. 5, n. 6, p. 644–648, 2010.

MURARO, A.; HOEKSTRA, M. O.; MEIJER, Y.; LIFSCHITZ, C.; WAMPLER, J. L.; HARRIS, C.; SCALABRIN, D. M. F. Extensively hydrolysed casein formula supplemented with *Lactobacillus rhamnosus* GG maintains hypoallergenic status: randomised double-blind, placebo-controlled crossover trial. **BMJ open**, [s.I.], v. 2, n. 2, p. e000637, 2012.

OLIVEIRA, J. P. B.; CANDREVA, A. M.; RIZZO, G.; RAMOS, M. V.; OLIVEIRA, J. S.; OLIVEIRA, H. D.; ARY, M. B.; DOCENA, G.; FREITAS, C. D. T. Allergenicity reduction of cow's milk proteins using latex peptidases. **Food chemistry**, [s.I.], v. 284, p. 245-253, 2019.

PAGE, M. J.; DI CERA, E. Serine peptidases: Classification, structure and function. **Cellular and molecular life sciences**, [s.I.], v. 65, n. 7-8, p. 1220-1236, 2008.

PALMA, J.M.; SANDALIO, L.M.; CORPAS, F.J.; ROMERO-PUERTAS, M.C.; MCCARTHY, I.; DEL RIO, L. A. Plant proteases, protein degradation, and oxidative stress: role of peroxisomes. **Plant physiology and biochemistry**, [s.I.], v. 40, n. 6-8, p. 521–530, 2002.

PANAVAS, T.; SANDERS, C.; BUTT, T. R. SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. **In: SUMO protocols**. Humana Press, Totowa, NJ, p. 303-317, 2009.

PAVLOVIČ, A.; SAGANOVÁ, M. A novel insight into the cost–benefit model for the evolution of botanical carnivory. **Annals of botany**, [s.I.], v. 115, n. 7, p. 1075-1092, 2015.

PEREZ-SILVA, J.G.; ESPANOL, Y.; VELASCO, G.; QUESADA, V. The Degradome database: expanding roles of mammalian proteases in life and disease. **Nucleic acids research**, [s.I.], v. 44, n. D1, p. D351–355, 2016.

PHILIPPS-WIEMANN, P. Proteases—general aspects. **Enzymes in human and animal nutrition**, Academic Press. [s.I.], chapter 12, p. 257-266, 2018.

PRESCOTT, S.; ALLEN, K. J. Food allergy: Riding the second wave of the allergy epidemic. **Pediatric allergy and immunology**, [s.I.], v. 22, n. 2, p. 155–160, 2011.

PROUTY, W. F.; KARNOVSKY, M. J.; GOLDBERG, A. L. Degradation of abnormal proteins in *Escherichia coli*. Formation of protein inclusions in cells exposed to amino acid analogs. **Journal of biological chemistry**, [s.I.], v. 250, n. 3, p. 1112-1122, 1975.

PUENTE, X.S.; S´ANCHEZ, L.M.; OVERALL, C.M.; L´OPEZ-OT´IN, C. Human and Mouse Proteases: A Comparative Genomic Approach. **Nature reviews genetics**, [s.I.], v. 4, n. 7, p. 544–558, 2003.

RAVEE, R.; SALLEH, F. I. M.; GOH, H. Discovery of digestive enzymes in carnivorous plants with focus on proteases. **PeerJ**, [s.I.], v. 6, p. e4914, 2018.

RAWLINGS, N. D.; BARRETT, A. J. Families of serine peptidases. **Methods in enzymology**, Academic Press, [s.I.], v. 244, p. 19–61, 2004.

RAWLINGS, N. D.; BARRETT, A. J.; THOMAS, P. D.; HUANG, X.; BATEMAN, A.; FINN, R. D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. **Nucleic acids research**, v. 46, n. D1, p. D624-D632, 2018.

RAWLINGS, N. D. Protease families, evolution and mechanism of action. **Proteases: structure and function**. Springer, Vienna, p. 1-36, 2013.

RISØR, M. W.; THOMSEN, L. R.; SANGGAARD, K. W.; NIELSEN, T. A.; THØGERSEN, I. B.; LUKASSEN, M. V.; ROSSEN, L.; GARCIA-FERRER, I.; GUEVARA, T.; SCAVENIUS, C.; MEINJOHANN, E.; GOMIS-RÜTH, F. X.; ENGHILD, J. J. Enzymatic and structural characterization of the major endopeptidase in the Venus flytrap digestion fluid. **Journal of biological chemistry**, [s.I.], v. 291, n. 5, p. 2271-2287, 2016.

ROTTLOFF, S.; STIEBER, R.; MAISCHAK, H.; TURINI, F. G.; HEUBL, G.; MITHOFER, A. Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, *Nepenthes*. **Journal of experimental botany**, [s.I.], v. 62, n. 13, p. 4639–4647, 2011.

ROTTLOFF, S.; MIGUEL, S.; BITEAU, F.; NISSE, E.; HAMMANN, P.; KUHN, L.; CHICHER, J.; BAZILE, V.; GAUME, L.; MIGNARD, B.; HEHN, A.; BOURGAUD, F. Proteome analysis of digestive fluids in *Nepenthes* pitchers. **Annals of botany**, [s.I.], v. 117, n. 3, p. 479–495, 2016.

SATHE, S.K.; SHARMA, G. M. Effects of food processing on food allergens. **Molecular nutrition & food research**, [s.I.], v. 53, n. 8, p. 970–978, 2009.

SCHULZE, W.; SCHULZE, E. D.; PATE, J. S.; GILLISON, A. N. The nitrogen supply from soils and insects during growth of the pitcher plants *Nepenthes mirabilis*, *Cephalotus follicularis* and *Darlingtonia californica*. **Oecologia**, [s.I.], v. 112, n. 4, p. 464-471, 1997.

SICHERER, S. H.; SAMPSON, H. A. Food allergy: epidemiology, pathogenesis, diagnosis, and treatment. **Journal of allergy and clinical immunology**, [s.I.], v. 133, n. 2, p. 291-307. e5, 2014.

SILVA, M. Z.; OLIVEIRA, J. P.; RAMOS, M. V.; FARIAS, D. F.; DE SÁ, C. A.; RIBEIRO, J. A.; SILVA, A. F. B.; SOUSA, J. S.; ZAMBELLI, R. A.; SILVA, A. C.; FURTADO, G. P.; GRANGEIRO, T. B.; VASCONCELOS, M. S.; SILVEIRA, R. S.; FREITAS, C. T. D. Biotechnological potential of a cysteine protease (CpCP3) from *Calotropis procera* latex for cheesemaking. **Food chemistry**, [s.I.], v. 307, p. 125574, 2020.

SILVA, S. R.; DIAZ, Y. C. A.; PENHA, H.A.; PINHEIRO, D. G.; FERNANDES, C.C.; MIRANDA, V.F.O; MICHAEL, T. P.; VARANI, A. M. The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial

carnivorous plants from the Lentibulariaceae family. **PLoS One**, [s.I.], v. 11, n. 10, p. 1-29, 2016.

SILVA, S. R.; ALVARENGA, D. O.; ARANGUREN, Y.; PENHA, H. A.; FERNANDES, C. C.; PINHEIRO, D. G.; OLIVEIRA, M.T.; MICHAEL, T. P.; MIRANDA, V. F. O.; VARANI, A. M. The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. **PLoS One**, [s.I.], v. 12, n. 7, p. 1-26, 2017.

SKATES, L.M.; PANIW, M.; CROSS, A.T.; OJEDA, F.; DIXON, K.W.; STEVENS, J.C.; GEBAUER, G. An ecological perspective on ‘plant carnivory beyond bogs’: nutritional benefits of prey capture for the Mediterranean carnivorous plant *Drosophyllum lusitanicum*. **Annals of botany**, [s.I.], v. 124, n. 1, p. 65–76, 2019.

ŠTULC, T.; LÁNSKÁ, V.; ŠNEJDRLOVÁ, M.; VRABLÍK, M.; PRUSÍKOVÁ, M.; ČEŠKA, R. A comprehensive guidelines-based approach reduces cardiovascular risk in everyday practice: the VARO study. **Archives of medical science: AMS**, [s.I.], v. 13, n. 4, p. 705–710, 2017.

TAKEUCHI, Y.; SALCHER, M. M.; USHIO, M.; SHIMIZU-INATSUGI, R.; KOBAYASHI, M. J.; DIWAY, B.; VON MERING, C.; PERNTHALER, J.; SHIMIZU, K. K. In situ enzyme activity in the dissolved and particulate fraction of the fluid from four pitcher plant species of the genus *Nepenthes*. **PLoS One**, [s.I.], v. 6, n. 9, p. 1-9, 2011.

THOROGOOD, C. J.; BAUER, U.; HISCOCK, S. J. Convergent and divergent evolution in carnivorous pitcher plant traps. **New phytologist**, [s.I.], v. 217, n. 3, p. 1035-1041, 2018.

TURNBULL, J.; ADAMSN, H.; GORARD, D. Review article: the diagnosis and management of food allergy and food intolerances. **Alimentary pharmacology & therapeutics**, [s.I.], v. 41, n. 1, p. 3-25, 2015.

WALTHER, C.; MAYER, S.; SEKOT, G.; ANTOS, D.; HAHN, R.; JUNGBAUER, A.; DÜRAUER, A. Mechanism and model for solubilization of inclusion bodies. **Chemical engineering science**, [s.I.], v. 101, p. 631-641, 2013.

ZHAO, G.; ZHOU, A.; LU, G.; MENG, M.; SUN, M.; BAI, Y.; HAN, Y.; WANG, L.; ZHOU, H.; CONG, H. Identification and characterization of *Toxoplasma gondii* aspartic protease 1 as a novel vaccine candidate against toxoplasmosis. **Parasites & vectors**, [s.I.], v. 6, n. 1, p. 175, 2013.