

UNIVERSIDADE CATÓLICA DOM BOSCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**Análise Evolutiva e Estrutural das Proteínas PR-4  
*SUGARWINS***

Autora: Lorhenn Bryanda Lemes Maia  
Orientador: Prof. Dr. Márcio de Castro Silva Filho  
Co-orientador: Dr. Humberto D'Muniz Pereira

Campo Grande  
Mato Grosso do Sul  
Setembro – 2021

UNIVERSIDADE CATÓLICA DOM BOSCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**Análise Evolutiva e Estrutural das Proteínas PR-4  
*SUGARWINS***

Autora: Lorhenn Bryanda Lemes Maia  
Orientador: Prof. Dr. Márcio de Castro Silva Filho  
Co-orientador: Dr. Humberto D'Muniz Pereira

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

Campo Grande  
Mato Grosso do Sul  
Setembro – 2021

Dados Internacionais de Catalogação na Publicação (CIP)  
Universidade Católica Dom Bosco  
Bibliotecária Mourãmise de Moura Viana - CRB-1 3360

M217a Maia, Lorhenn Bryanda Lemes  
Análise evolutiva e estrutural das proteínas PR-4  
SUGARWINS Lorhenn Bryanda Lemes Maia; sob orientação  
do Prof. Dr. Márcio de Castro Silva Filho e co-orientação  
do Dr. Humberto D'Muniz Pereira. -- Campo Grande,  
MS : 2021.  
88 p.: il.;

Tese (Doutorado em Biotecnologia) - Universidade Católica  
Dom Bosco, Campo Grande-MS, Ano 2021  
Bibliografia: p. 80 - 86

1. Análise filogenética. 2. Proteínas. 3. Cana-de-açúcar I.  
Márcio de Castro Silva. II. Pereira, Humberto D'Muniz.  
III. Título.

CDD: 660.6

**“ANÁLISE EVOLUTIVA E ESTRUTURAL DAS PROTEÍNAS PR-4  
SUGARWINS”**

**Autora: LORHENN BRYANDA LEMES MAIA**

**Orientador: Prof. Dr. Márcio de Castro Silva Filho**

**TITULAÇÃO: Doutora em Biotecnologia**  
**Área de concentração: Biotecnologia.**

**APROVADA em 28 de setembro de 2021.**

*Marcio W N =*

---

**Dr. Márcio de Castro Silva Filho (orientador) - UCDB**  
**Dr. Octávio Luiz Franco - UCDB**  
**Dr. Ludovico Migliolo – UCDB**  
**Dr<sup>a</sup>. Susana Elisa Moreno - UCDB**  
**Dr. Humberto D’Muniz Pereira – IFSC da USP**

*“Nem todas nascem para ser princesas,  
Algumas meninas,  
Só querem fazer a diferença no mundo,  
Ser resgatada de uma torre, te faz se sentir amada.  
Mas só o reconhecimento que sempre buscou,  
Te ensinará a se amar.”*

*Lorhenn Maia*

Ao meu esposo, André Rodrigues da Silva,  
Que abdicou de sua estabilidade e segurança,  
Para que pudesse me sentir realizada.

## AGRADECIMENTOS

À **Universidade Católica Dom Bosco**, pela oportunidade de uma formação acadêmica de qualidade.

À **Pró-reitoria de Pesquisa e Pós-graduação (PROPP-UCDB)**, por meio do Programa de Pós-graduação em Biotecnologia, pela oportunidade de um bom curso de doutorado que me permitiu crescimento profissional no decurso para a obtenção do título de PhD, aumentando minhas qualificações.

À **CAPES**, por financiar meus estudos e me proporcionar uma bolsa para o desenvolvimento de minha tese.

À **FAPESP e CNPq**, pelo fomento que tornou este belo projeto passível de execução.

Ao **Instituto de Física de São Carlos** da Universidade de São Paulo (IFSC - USP), por me permitir um doutorado-sanduíche, sediando a execução da principal etapa metodológica de meu trabalho.

Ao **Departamento de Genética e Evolução** da Universidade Federal de São Carlos (UFSCar), pela eficaz produção do objeto de estudo deste trabalho, *SUGARWINS* 1 e 2.

Ao **Diamond Light Source**, o Síncrotron Nacional do Reino Unido, pela coleta dos dados cristalográficos que nos permitiram alta resolução na estrutura da proteína.

Ao **Laboratório Nacional de Luz Síncrotron (LNLS)** do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), pela brilhante recepção, suporte e esclarecimentos sobre a coleta de dados num síncrotron.

Ao **Prof. Dr. Cristiano Marcelo Espinola Carvalho**, meu querido chefe, por me abrir as portas da UCDB e me permitir ampliar os horizontes com a criação do grupo de pesquisa *Oncolarning Technology*.

Ao **Prof. Dr. Márcio de Castro Silva Filho**, meu orientador, por me conceder a oportunidade de participar de um projeto tão incrível, fortalecendo meu desenvolvimento profissional e minha formação acadêmica.

Ao **Dr. Humberto D'Muniz Pereira**, meu co-orientador, por me socorrer e por vezes salvar, em meio aos grandes desafios enfrentados na Cristalografia, esta, que é uma metodologia tão complexa.

Ao **Prof. Dr. Richard Charles Garratt**, membro do grupo de Cristalografia do IFSC-USP, que me permitiu acesso ao Laboratório de Biologia Estrutural, onde pude desenvolver as principais etapas metodológicas na execução deste trabalho.

Aos **demais coautores** deste trabalho: Prof. Dr. Flavio Henrique-Silva e Danyelle Toyama da UFSCar, Prof. Dr. José Fernando Ruggiero Bachega e Júlia Vasconcellos Peixoto da UFCSPA, Profa. Dra. Renata O. Dias da UFG e Dr. José Brandao-neto do *Diamond Light Source*. – UK, pelas significativas contribuições para obtenção dos resultados, sem os quais, não seria possível chegar ao final desta tese.

Aos professores **Dr. Octavio Luiz Franco**, **Dr. Ludovico Migliolo** e **Dra. Susana Elisa Moreno** por aceitarem compor a banca de defesa desta tese e pelas



significativas contribuições que fizeram em minha formação acadêmica e neste trabalho.

Ao grupo de pesquisa ***Oncolearning Technology***, pelo apoio, incentivo e confiança. A paixão pela ciência, que tantas vezes vi, nos olhos de meus doces alunos de Iniciação Científica, me deu forças para vencer momentos de cansaço e decepção.

Ao meus pais, **Valdemir e Janete Maia**, que sempre acreditaram em mim, mesmo quando eu duvidava.

Ao meu esposo, **André Rodrigues da Silva**, que não desistiu de mim, mesmo quando eu havia desistido.

À **Deus**, por conduzir meus passos quando a dor e o cansaço me cegaram.

## BIOGRAFIA DO AUTOR

Lorhenn Bryanda Lemes Maia, filha de Valdemir Suriano Maia e Janete Lemes de Souza Maia, nasceu em Campo Grande/MS no dia 28 de agosto de 1990. Recebeu o grau de **Bacharel em Ciências Biológicas** pela Fundação Universidade Federal de Mato Grosso do Sul (UFMS), em 14 dezembro de 2011. E, o título de **Mestra em Genética** pelo Programa de Pós-graduação em Genética da Universidade Federal do Paraná (UFPR), em 24 março de 2014. Enquanto doutoranda pela Universidade Católica Dom Bosco (UCDB), foi bolsista CAPES em projeto científico visando correlacionar estrutura e função de proteínas com importância biotecnológica, a partir da resolução de sua estrutura tridimensional por cristalografia e análises bioinformáticas estruturais complementares, pelo Programa de Pós-graduação em Biotecnologia da Universidade Católica Dom Bosco. Realizou doutorado-sanduíche no Instituto de Física de São Carlos da Universidade de São Paulo (USP), com a coleta dos dados cristalográficos tendo sido realizada no *Diamond Light Source*, localizado no *campus de Harwell Science and Innovation*, Reino Unido.

No doutorado, pertenceu ao grupo de pesquisa *Oncolearning Technology* como pesquisadora principal e contato científico, na linha de pesquisa Biotecnologia aplicada à saúde, visando identificar possíveis marcadores moleculares em carcinomas mamários, a partir de diferentes análises de dados genômicos e transcriptômicos (*in silico* e *in vitro*), na Universidade Católica Dom Bosco. Tendo atuado ativamente em diferentes projetos e orientado alunos de graduação em trabalhos de PIBIC e TCCs. Bem como, contribuído para o desenvolvimento de dissertações de mestrado. Lorhenn possui experiência em diferentes técnicas de biologia molecular e bioinformática para análise de dados biológicos. Ao longo de sua formação acadêmica realizou diversos estágios em diferentes universidades do país, que lhe conferiram conhecimento técnico-científico em análises de genomas, transcriptomas e proteomas.

## SUMÁRIO

	Página
Lista de Tabelas.....	ix
Lista de Figuras.....	x
Figuras do Apêndice.....	xi
Resumo.....	xii
Abstract.....	xiii
Introdução Geral.....	1
1.    Cana-de-açúcar.....	1
2.    Proteínas PR-4.....	4
3.    SUGARWINS.....	5
4.    Biotecnologia e Biologia Estrutural.....	10
5.    Predição de Estrutura.....	13
6.    Justificativa e Aplicabilidade.....	16
Objetivos.....	18
Capítulo I	
Artigo.....	19
Capítulo II	
Padronização Metodológica.....	51
Considerações Finais.....	59
Apêndice.....	60
Referências Bibliográficas .....	68

## LISTA DE TABELAS

	Página
<b>Table 1</b> <i>Data-collection, processing and refinement statistics</i> .....	26
<b>Supplementary Table 1</b> <i>Distribution of sequences among the clades of the phylogenetic tree</i> .....	46

## LISTA DE FIGURAS

	Página
<b>Figura 1.</b> Resumo da atividade enzimática das SUGARWINs.....	6
<b>Figura 2.</b> Estrutura 3D de WHEATWIN1.....	7
<b>Figura 3.</b> Alinhamento de sequência para proteínas homólogas de BARWIN.....	8
<b>Figura 4.</b> Interação esperada entre as SUGARWINs e glicol-quitosano e quitina.....	9
<b>Figure 1</b> <i>Phylogenetic analysis for 436 PR-4 proteins found in the genome annotation from 106 Embryophyta species.....</i>	29
<b>Figure 2</b> <i>Crystallographic structure of SUGARWIN2.....</i>	31
<b>Figure 3</b> <i>The Catalytic Site in (A) Ribonuclease A, (B) CARWIN, and (C) SUGARWIN2.....</i>	32
<b>Figure 4</b> <i>Alignment of SUGARWINs and BARWIN.....</i>	34
<b>Figure 5</b> <i>Electrostatic surface of the SUGARWINs.....</i>	34
<b>Figure 6</b> <i>RMSF of the SUGARWINs.....</i>	36
<b>Figure 7</b> <i>PCA analysis of the SUGARWINs.....</i>	37
<b>Figure 8</b> <i>Structures of SUGARWINs with emphasis on the chitosan binding site.....</i>	38
<b>Supplementary Figure 1</b> <i>12% SDS-PAGE showing the purification of SUGARWIN2 by IMAC.....</i>	45
<b>Supplementary Figure 2</b> <i>DLS of SUGARWIN2.....</i>	45

## FIGURAS DO APÊNDICE

	Página
<b>Figura 1.</b> Gel SDS-PAGE 12% mostrando a purificação de SUGARWIN2 por IMAC (cromatografia de afinidade).....	60
<b>Figura 2.</b> Resultado da segunda purificação de SUGARWIN2.....	61
<b>Figura 3.</b> DLS da SUGARWIN2.....	61
<b>Figura 4.</b> Cristal da proteína SUGARWIN2.....	62
<b>Figura 5.</b> Alinhamento SUGARWIN1 e SUGARWIN2.....	62
<b>Figura 6.</b> Célula unitária da SUGARWIN2.....	63
<b>Figura 7.</b> Estrutura 3D de SUGARWIN2.....	63
<b>Figura 8.</b> Árvore filogenética de máxima verossimilhança para sequências de proteínas homólogas a SUGARWINs.....	64
<b>Figura 9.</b> Análise de similaridade de sequências.....	65
<b>Figura 10.</b> Análise de modos normais.....	65
<b>Figura 11.</b> Análise PCA das SUGARWINs.....	66
<b>Figura 12.</b> Sítio de ligação quitosan das SUGARWINs.....	67

## RESUMO

SUGARWINs são proteínas PR-4 associadas à defesa da cana-de-açúcar contra fitopatógenos. Essas proteínas desempenham um importante papel na defesa das plantas, em particular contra patógenos fúngicos, como *Colletotrichum falcatum* (Went) e *Fusarium verticillioides*. A família das proteínas PR-4 é um grupo equipado com um domínio BARWIN, que pode estar associado a um domínio de ligação à quitina. Várias proteínas PR-4 exibem ambas atividades de quitinase e RNase, sendo esta última associada à presença de dois resíduos de histidina H11 e H113 (BARWIN) [H44 e H146, SUGARWINs]. Na cana-de-açúcar, SUGARWIN1 exibe atividades de ribonuclease, quitosanase e quitinase, enquanto SUGARWIN2 apenas exibe atividade de quitosanase. A fim de decifrar os determinantes estruturais envolvidos nesta gama diversificada de especificidades enzimáticas, determinamos a estrutura 3D da SUGARWIN2, com resolução de 1,55Å por difração de raios-X. Esta é a primeira estrutura de uma proteína PR-4 onde a primeira histidina foi substituída por asparagina. Sendo subsequentemente usada para construir um modelo de homologia para a SUGARWIN1. Simulações de dinâmica molecular de ambas as proteínas revelaram a presença de uma alça flexível apenas em SUGARWIN1 e postulamos que isso, juntamente com a presença da histidina catalítica em posição 42, torna-a competente como uma ribonuclease. Uma análise filogenética de PR-4 obtidas a partir de 106 genomas Embryophyta mostraram que ambas as histidinas catalíticas estão espalhadas entre eles com poucas substituições durante a história evolutiva da família dos genes. A substituição do H11 pelo N11 também está presente em duas outras proteínas PR-4 da cana-de-açúcar: SUGARWIN3 e SUGARWIN4. Assim, propomos que a atividade RNase estava presente nas primeiras proteínas Embryophyta PR-4, mas foi recentemente perdida em membros desta família durante o curso da evolução.

**Palavras-chave:** SUGARWIN, BARWIN, cristalografia, *loop* variável, PR-4, análise filogenética.

## ABSTRACT

SUGARWINs are PR-4 proteins associated with sugarcane defense against phytopathogens. These proteins play an important role in plant defense, in particular against fungal pathogens, such as *Colletotrichum falcatum* (Went) and *Fusarium verticillioides*. The pathogenesis-related protein-4 (PR-4) family is a group of proteins equipped with a BARWIN domain, which may be associated with a chitin-binding domain also known as the hevein-like domain. Several PR-4 proteins exhibit both chitinase and RNase activity, with the latter being associated with the presence of two histidine residues H11 and H113 (BARWIN) [H44 and H146, SUGARWINs]. In sugarcane, SUGARWIN1 exhibits ribonuclease, chitosanase and chitinase activities, whereas SUGARWIN2 only exhibits chitosanase activity. In order to decipher the structural determinants involved in this diverse range of enzyme specificities, we determined the 3-D structure of SUGARWIN2, at 1.55Å by X-ray diffraction. This is the first structure of a PR-4 protein where the first histidine has been replaced by asparagine and was subsequently used to build a homology model for SUGARWIN1. Molecular dynamics simulations of both proteins revealed the presence of a flexible loop only in SUGARWIN1 and we postulate that this, together with the presence of the catalytic histidine at position 42, renders it competent as a ribonuclease. A phylogenetic analysis of PR-4 proteins obtained from 106 Embryophyta genomes showed that both catalytic histidines are widespread among them with few replacements in these amino acid positions during the gene family evolutionary history. We observe that the H11 replacement by N11 is also present in two other sugarcane PR-4 proteins identified and described in this work: SUGARWIN3 and SUGARWIN4. We propose that RNase activity was present in the first Embryophyta PR-4 proteins but was recently lost in members of this family during the course of evolution.

**Keywords:** SUGARWIN, BARWIN, crystallography, variable loop, PR-4, phylogenetic analysis



## INTRODUÇÃO GERAL

### 1. Cana-de-açúcar

A cana-de-açúcar é uma importante cultura de regiões tropicais e subtropicais do mundo, sendo plantada em cerca de 20 milhões de hectares em mais de 90 países (MENOSSI et al., 2008). O Brasil é o maior produtor mundial, com uma área de cerca de 7 milhões de hectares plantados com cana-de-açúcar, direcionada, principalmente, para produção de açúcar e etanol. Na safra 2020/2021, o Brasil produziu aproximadamente 657 milhões de toneladas de cana-de-açúcar, matéria-prima utilizada para a produção de 41,5 milhões de toneladas de açúcar, 32,5 bilhões de litros de etanol e 22,6 TWh para a rede elétrica nacional (UNICA, 2021). O ataque da broca-da-cana *Diatraea saccharalis* (F.) (Lepidoptera: Pyralidae) é um dos principais problemas para a cultura devido aos danos diretos e indiretos que causam (DINARDO-MIRANDA, 2010).

Os danos diretos são causados pelas lagartas que se alimentam do tecido do colmo formando galerias e causando diminuição da massa e de valores nutritivos da planta que, conseqüentemente, geram redução da qualidade e quantidade dos produtos finais (DINARDO-MIRANDA, 2010). Os danos indiretos são causados por fungos patogênicos que se encontram associados a lagartas da broca (FRANCO et al., 2021), como *Colletotrichum falcatum* (Went) e *Fusarium verticillioides* (Sacc.) Nirenberg. Estes fitopatógenos causam as doenças conhecidas por podridão vermelha e podridão de *Fusarium*, respectivamente, que levam a uma redução no teor de açúcar nos colmos devido à inversão da sacarose armazenada na planta e sua transformação em glicose e levulose, que não se cristalizam no processo industrial (DINARDO-MIRANDA, 2010), ocasionando um menor rendimento na produção de açúcar.

O sistema de defesa de plantas está sob constante pressão seletiva para aperfeiçoar a sua resposta aos microrganismos patogênicos e aos insetos de forma

sincronizada (CUI et al., 2002). Durante as interações inseto-planta, elicitores presentes em secreções orais dos insetos ativam ou reprimem uma variedade de vias de sinalização de defesa (WALLING, 2000). As principais vias de sinalização envolvidas na resposta da planta ao ferimento de herbívoros são as vias do ácido jasmônico (AJ) e etileno (ET), que atuam sinergicamente para induzir genes de resposta ao ferimento. O reconhecimento do patógeno pela planta desencadeia inúmeras respostas de defesa, dentre elas a fortificação da parede celular através da síntese de calose e lignina, a produção de metabólitos secundários com efeito antimicrobiano, tais como fitoalexinas, e o acúmulo de proteínas relacionadas à patogênese (proteínas PR - *pathogenesis-related protein*) (PIETERSE et al., 2009). Estudos de expressão gênica mostram que a via do ácido salicílico (AS) é primariamente ativada em resposta a patógenos biotróficos, enquanto as vias de AJ e ET são induzidas em resposta a patógenos necrotróficos (THOMMA, et al., 1998).

Embora seja aceito que, as vias de resposta ao ataque de patógenos ou de herbívoros são distintas, algumas evidências sugerem que o AJ também medeia a resistência induzida por alguns patógenos (SCHWEIZER et al., 1998; TERRAS et al., 1998). Além disso, o AS e as vias de sinalização do AJ podem interagir (BOSTOCK, 1999; STOUT et al., 2006). Neste caso, estas vias podem se reprimir mutuamente (THALER et al., 1999; FELTON et al., 1999) ou podem apresentar um efeito sinérgico entre si (LAWTON et al., 1994; STENNIS et al., 1998). Porém essa interação vai depender da concentração relativa, tempo e sequência de iniciação de cada hormônio (MUR et al., 2006; KOORNNEEF et al., 2008).

Em busca de proteínas relacionadas com a defesa em cana de açúcar, dois homólogos de uma proteína da cevada induzida por ferimento (BARWIN) foram identificadas e designados SUGARWIN1 e 2 (*sugarcane wound-inducible proteins*) (POMPERMAYER, 2004). SUGARWINs são proteínas secretadas cuja indução de transcritos ocorre em resposta ao ferimento mecânico, tratamento com metil jasmonato e, principalmente, ao ataque de *Diatraea saccharalis*, mas não ocorre em resposta à infecção por patógenos (MEDEIROS et al., 2012). Os níveis de expressão de SUGARWIN1 e 2 aumentam com o tempo em plantas tratadas, apresentando expressão local, que cessa com a parada do tratamento (MEDEIROS et al., 2012).

Apesar de sua elevada expressão devido ao ataque de lagartas de *D. saccharalis* a proteína não possui nenhum efeito no desenvolvimento do inseto, mas causa alterações na morfologia das hifas de *F. verticillioides* e *C. falcatum*, que apresentam aumento da vacuolização, vários pontos de fratura, liberação de material intracelular e formação de calos na região dos séptos, seguida de morte por apoptose, quando tratados com SUGARWIN2 recombinante (MEDEIROS et al., 2012; FRANCO et al., 2014). O efeito da SUGARWIN2 é específico para fungos fitopatogênicos, não apresentando efeito em fungos não relacionados a doenças de cana, como *Saccharomyces cerevisiae* e *Aspergillus nidulans* (FRANCO et al., 2014), o que sugere fortemente que a indução dos genes *SUGARWINs* está relacionada com a defesa da planta ao ataque dos patógenos fúngicos associados à broca, visando minimizar seus danos.

Muitos estudos mostram interações entre o ataque de insetos e patógenos em plantas, podendo estas ser favoráveis a ambos (HATCHER e PAUL, 2000; CUI et al., 2002; GATCH e MUNKVOLD, 2002; SCHULTHESS, CARDWELL e GOUNOU, 2002; BURG et al., 2012). Em cana-de-açúcar, alguns isolados de *Fusarium spp.* apresentam efeito positivo na sobrevivência das larvas, e na taxa de desenvolvimento de *Eldana saccharina*, de tal modo que, variedades consideradas com aceitável resistência à broca do caule, mas suscetíveis à infecção por certas espécies de *Fusarium spp.*, mostram aumento nos danos causados por *E. saccharina* (McFARLANE et al., 2009). A broca europeia do milho *Ostrinia nubilalis* (Hübner), se desenvolve 20% mais rápido em tecidos de milho infectado com *Colletotrichum graminicola* (CARRUTHERS et al., 1986). Recentemente, nosso grupo mostrou que o fungo patogênico *Fusarium verticillioides* modula as plantas de cana-de-açúcar e, as larvas e mariposas de *Diatrea saccharalis*, para aumentar sua disseminação no campo (FRANCO et al., 2021). Provavelmente, a expressão de genes induzidos por herbivoria, mas com efeito em fitopatógenos, como é o caso das *SUGARWINs*, é uma tentativa da planta para diminuir danos maiores.

Embora ainda não tenha sido completamente elucidado o mecanismo pelo qual as proteínas *SUGARWINs* afetam *F. verticillioides* e *C. falcatum*, levando estes fitopatógenos à morte por apoptose, acredita-se que as diferentes funções

desempenhadas por estas proteínas estejam fortemente envolvidas em mecanismos específicos da ação fungicida.

## 2. Proteínas PR-4

Um grande número de proteínas PR foi caracterizado e agrupado com base em suas características estruturais e funcionais (CHRISTENSEN et al., 2002) mostrando atividade antifúngica contra patógenos específicos e possuindo atividade enzimática ou inibitória, como glucanases (PR-2), peroxidases (PR-9), quitinases (PR-3, PR-4, PR-8, PR-11), inibidores de proteinase (PR-6) e semelhantes a ribonuclease (PR-10). Seus diversos modos de ação tornam essas moléculas uma ferramenta completa em estratégias de defesa. Nos últimos anos Bertini et al. (2012) estudaram extensivamente as proteínas PR da classe 4 de trigo, uma das famílias menos conhecidas de PRs, denominada WHEATWINS.

De acordo com Medeiros et al. (2012) vários genes que codificam proteínas WINs (do inglês, *wound-inducible proteins*) foram identificados em várias espécies de plantas (CHRISTOPHER et al., 2004; RYAN, 1990; SCHILLMILLER e HOWE, 2005). A família da proteína 4 relacionada à patogênese (PR-4) é um grupo de proteínas equipadas com um domínio do tipo BARWIN. BARWIN é uma proteína PR-4 de cevada induzida por ferimentos ou agentes patogênicos (HEJGAARD et al., 1992; SVENSSON et al., 1992). Embora o domínio BARWIN possa ser encontrado em associação com um domínio de ligação de quitina típico de lectinas (BROEKAERT et al., 1990), foram isoladas proteínas que mostram apenas o domínio BARWIN em várias espécies de plantas: tabaco (FRIEDRICH et al., 1991), tomate (LINTHORST et al., 1991), *Arabidopsis* (POTTER et al., 1993), trigo (CARUSO et al., 1999), *W. japonica* (KIBA et al., 2003), milho (BRAVO et al., 2003), arroz (AGRAWAL et al., 2003; ZHU et al., 2006) e *L. radiata* (LI et al., 2010). Mais recentemente, homólogos de BARWIN foram relatados em maçã (BAI et al., 2013), cacau (MENEZES et al., 2014) e pimenta (KIM e HWANG, 2015).

Os homólogos PR-4 BARWIN de várias espécies de plantas têm sido associados às respostas das plantas à infecção fúngica e ferimentos mecânicos. A atividade

antifúngica foi descrita para proteínas semelhantes a BARWIN em várias espécies de plantas (LINTHORST et al., 1991; HEJGAARD et al., 1992; CARUSO et al., 1999; CARUSO et al., 2001; AGRAWAL et al., 2003; BRAVO et al., 2003; KIBA et al., 2003; ZHU et al., 2006; LI et al., 2010; MEDEIROS et al., 2012; BAI et al., 2013; FRANCO et al., 2014; MENEZES et al., 2014; SOUZA et al., 2017; FRANCO et al., 2019).

Proteínas do grupo PR-4 (proteínas relacionadas à patogênese 4) são classificadas como quitinases (NEUHAUS et al., 1996; VAN LOON e VAN STRIEN, 1999); no entanto, vários estudos também relatam atividade RNase para proteínas semelhantes a BARWIN (BERTINI et al., 2009; 2012; GUEVARA-MORATO et al., 2010; BAI et al., 2013; HUET et al., 2013; MENEZES et al., 2014; FRANCO et al., 2014; KIM e HWANG, 2015). Sítios de ligação com o RNA já foram descritos para WHEATWIN1 (BERTINI et al., 2009) e CARWIN (HUET et al., 2013), mostrando dois importantes resíduos de histidina para a atividade RNase (BERTINI et al., 2009). Uma atividade DNase antifúngica também foi observada, juntamente com a atividade da RNase, para as proteínas PR-4 de *Capsicum chinense* (GUEVARA-MORATO et al., 2010) e TcPR-4b, de *Theobroma cacao* (MENEZES et al., 2014). Estudos anteriores do nosso grupo identificaram dois homólogos PR-4 de BARWIN na cana-de-açúcar: SUGARWIN1 e SUGARWIN2 (FALCO et al., 2001; MEDEIROS et al., 2012).

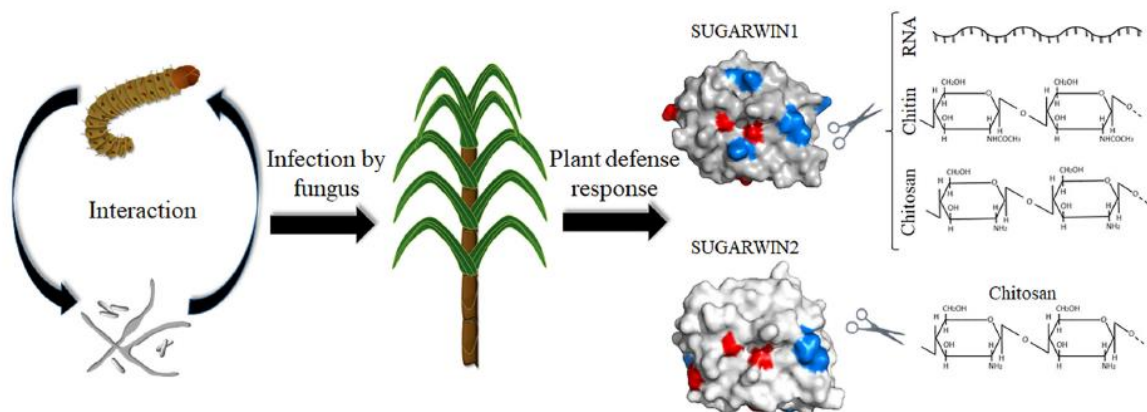
As proteínas PR-4 são agrupadas em classe I e classe II com base na presença ou ausência de um domínio de ligação de quitina (NEUHAUS et al., 1996). Proteínas que compõem o grupo de classe I são classificadas como quitinases, devido à apresentação desta atividade por uma proteína do tabaco (PONSTEIN et al., 1994). Outros trabalhos revelaram que proteínas PR-4 da classe II, apresentam apenas um domínio BARWIN e exibem atividade RNase (BERTINI et al., 2009; BAI et al., 2013; KIM e HWANG, 2015).

### **3. SUGARWINS**

Acredita-se que as proteínas SUGARWINS (WINS de cana-de-açúcar) sejam parte de um mecanismo de defesa contra fungos patogênicos em cana-de-açúcar, por meio das vias dependentes do SA e JA. As SUGARWINS são proteínas secretadas,

fortemente induzidas em resposta aos ferimentos mecânicos provocados pelo ataque de *Diatraea saccharalis* (Fabricius) e após tratamento com metil-jasmonato. Por outro lado, sua indução à infecção por fungos ocorre em níveis significativamente inferiores em relação à herbivoria (MEDEIROS et al., 2012, FRANCO et al., 2019).

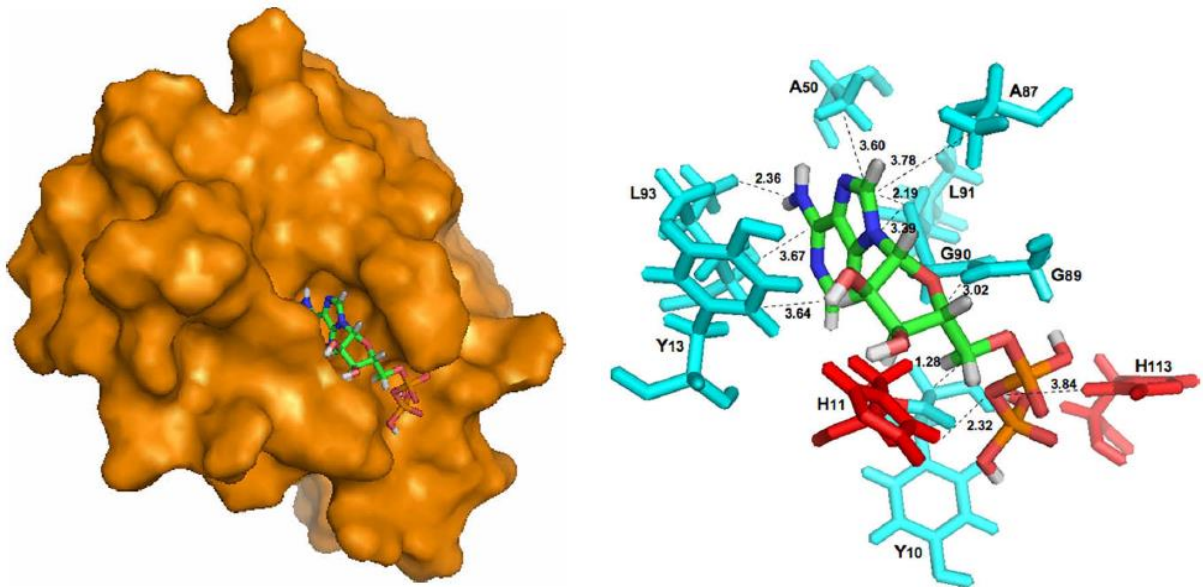
Apesar de seu alto nível de expressão em resposta ao ataque de *D. saccharalis*, a proteína não tem efeito sobre o desenvolvimento ou mortalidade do inseto. No entanto, SUGARWINs tem efeitos antimicrobianos em fungos fitopatogênicos, causando alterações na morfologia das hifas e levando à morte celular por apoptose. Franco et al. (2014) mostraram que as proteínas SUGARWINs desencadeiam mudanças nas hifas de *Fusarium verticillioides* e *Colletotrichum falcatum*, incluindo aumento da vacuolização, múltiplos pontos de fratura na parede celular e extenso vazamento de material intracelular. Buscando caracterizar as funções das PR-4 SUGARWINs e seus papéis na defesa da cana-de-açúcar, Franco e colaboradores (2019) demonstraram que tais proteínas são induzidas por *C. falcatum*, desempenhando um papel importante na defesa das plantas contra fitopatógenos. Quanto aos aspectos funcionais, SUGARWIN1 exibiu atividade ribonuclease, quitosanase e quitinase, enquanto SUGARWIN2 exibiu apenas atividade quitosanase (Figura 1).



**Figura 1.** Resumo da atividade enzimática das SUGARWINs. A broca da cana-de-açúcar *D. saccharalis* pode induzir a expressão dos genes PR-4 SUGARWINs, que combatem a infecção do agente etiológico da podridão vermelha em cana-de-açúcar, *C. falcatum*. Os produtos destes genes atuam como proteínas de defesa que afetam a morfologia do fungo e sua viabilidade devido a atividades enzimáticas. SUGARWIN1 mostra atividade RNase, quitinase e quitosanase, enquanto SUGARWIN2 mostra apenas atividade quitosanase.

**Fonte:** Franco et al. (2019).

Esta especificidade enzimática variável foi sugerida como resultado da composição divergente de aminoácidos no local de ligação ao substrato. Um sítio de ligação a RNA também foi descrito para outras WINS (BERTINI et al., 2009; HUET et al., 2013), evidenciando dois resíduos de histidina evolutivamente conservados. Caporale et al. (2004) demonstraram que a atividade ribonuclease de WHEATWIN1 está relacionada ao mecanismo clássico ácido-base das ribonucleases A e T1 (D’ALESSIO e RIORDAN, 1997). Enquanto Bertini et al. (2009) correlacionaram sua atividade ribonuclease com sua função antifúngica. A Figura 2 mostra os resíduos de His importantes para a atividade RNase, nas posições 11 e 113 (His11 e His113) de WHEATWIN1.



**Figura 2.** Estrutura 3D de WHEATWIN1. A) Mostra a superfície 3D de WHEATWIN1, com a posição de um 50-ADP na cavidade do sítio ativo. B) Mostra as posições espaciais dos resíduos hidrofóbicos (ciano) na cavidade de WHEATWIN1, os quais estabelecem interações com 50-ADP. As posições dos resíduos His11 e His113 (em vermelho) também são mostradas. As respectivas distâncias indicadas, também estão indicadas.

**Fonte:** BERTINI et al., (2009) (modificado).

Para entender as diferentes especificidades de substrato observadas durante a análise das atividades das SUGARWINS, Franco et al. (2019) realizaram a predição da estrutura tridimensional de ambas as proteínas por modelagem molecular. Como pode ser observado na Figura 3, várias posições divergentes de aminoácidos foram

observadas entre a SUGARWIN1 e SUGARWIN2, alguns dos quais se mostraram na superfície das estruturas previstas. O envolvimento dessas posições de superfície como locais de suposta ligação ao substrato, foram posteriormente analisados por estratégias de encaixe (Figura 4).

```

Carwin_Carica_papaya_(4JP6_A)      1 ESASNVRATYHFYNAQQNGWDLR - -KVSAYCATWDADKPYSWRSKYGWTAFCCGPVGPHGRAAC 61
Sugarwin1                          32 QQASNVRATYHFYNPQQNGWNLN - -AVSAYCATWDADKPLSWRQKYGWTAFCCGPAGQKQQAAC 92
Sugarwin2                          27 QQASGVRATTYHFYNPTQNNWDL - -AGTYCATWDAGQPLSWRSKYGWTAFCCGPAGPTGQAAC 85
Triticum_aestivum_1_(O64392.1)    22 QQATNVRATYHYRPAQNNWDLGAPAVSAYCATWDASKPLSWRSKYGWTAFCCGPAGAHGQAAC 84
Triticum_aestivum_2_(O64393.1)    24 QQATNVRATYHYRPAQNNWDLGAPAVSAYCATWDASKPLSWRSKYGWTAFCCGPAGAHGQAAC 86
Barwin_Hordeum_vulgare_(1BW3_A)   1 EQANDVRATYHYRPAQNNWDLGAPAVSAYCATWDASKPLSWRSKYGWTAFCCGPAGPRGQAAC 63
Zea_mays_(NP_001105464.2)         32 QQASNVRATYHFYNPQQNGWDLN - -RVSAYCATWDADKPLSWRQKHGWTAFCCGPAGQKQQAAC 92
Oryza_sativa_(AAL11444.1)        26 QEASNVRATYHFYNPQQNNWDLN - -KVSAYCATWDANKPLSWRQKYGWTAFCCGPAGPRGRDSC 86
Solanum_tuberosum_(P09762.1)     77 GSAQNVRATYHIYNPQNVGWDLN - -AVSAYCSTWDANKPYAWRSKYGWTAFCCGPVGPRGRDSC 137
Nicotiana_tabacum_(AAB29959.2)    78 GAAQNVRATYHIYNPQNVGWDLY - -AVSAYCSTWDGNKPLAWRRKYGWTAFCCGPVGPRGRDSC 138
Hevea_brasiliensis_(AAO63572.1)  70 GSASNVLATYHLYNPQQHGWDLN - -AVSAYCSTWDANKPYSWRSKYGWTAFCCGPVGAGHQQPSC 130
Arabidopsis_thaliana_(P43082.1)   72 ESASNVRATYHFYNPQQNNWDLN - -AVSAYCSTWDADKPYAWRSKYGWTAFCCGPAGPRGQAAC 132
Nicotiana_tabacum_(CAA41437.1)    26 QSATNVRSTYHLYNPQNINWDLR - -AASAFCATWDADKPLAWRSKYGWTAFCCGPAGPRGQVSC 86
Vitis_vinifera_(AAC33732.1)      22 QSASNVRATYHFYNPEQNGWDLN - -AVSAYCSTWDASQPLAWRSKYGWTAFCCGPSGPTGQAAC 82

Carwin_Carica_papaya_(4JP6_A)      62 GKCLRVNTNKTREAETTVRIVDQCSNGGGLDLDW-SVFKKLDTDGSGYLRGHLIVNYQFVNCGN 122
Sugarwin1                          93 GKCLRVNRATGASIVARIVDQCSNGGGLDLDYETVFKKIDTNGQGYQMGHLNVNYQFVAC - - 152
Sugarwin2                          86 GQCLLVNTNTATGASLTVRIVDQCSNGGGLDLDYDTAFKPLDTNGAGIQAGHLTVNYQFVNCGD 147
Triticum_aestivum_1_(O64392.1)    85 GKCLQVTNPATGAQITARIVDQCSNGGGLDLDWDTVFTKIDTNGIGYQQGHLNVNYQFVDCRD 146
Triticum_aestivum_2_(O64393.1)    87 GKCLRVNTPATGAQITARIVDQCSNGGGLDLDWDTVFTKIDTNGIGYQQGHLNVNYQFVDCRD 148
Barwin_Hordeum_vulgare_(1BW3_A)   64 GKCLRVNTPATGAQITARIVDQCSNGGGLDLDWDTVFTKIDTNGIGYQQGHLNVNYQFVDCRD 125
Zea_mays_(NP_001105464.2)         93 GKCLRVNTPATGAQITARIVDQCSNGGGLDLDYETVFKKIDTNGQGYQMGHLNVNYQFVAC - - 152
Oryza_sativa_(AAL11444.1)        87 GKCLQVKNRGTGATIIARIVDQCSNGGGLDLDYETVFKKIDTNGRQGYQMGHLQVDYKFVNC - - 146
Solanum_tuberosum_(P09762.1)     138 GKCLRVNTRTGAQTTVRIVDQCSNGGGLDLDI-NVFQIDDTDGVGNQQGHLIVNYQFVNCGD 198
Nicotiana_tabacum_(AAB29959.2)    139 GKCLRVNTNTGTGAQTTVRIVDQCSNGGGLDLDV-NVFRQLDDTDGRGNQRGHLIVNYEFVNCGD 199
Hevea_brasiliensis_(AAO63572.1)  131 GKCLSVNTNTGTGAKTTVRIVDQCSNGGGLDLDV-NVFRQLDDTDGKGYERGHLIVNYQFVDCGD 191
Arabidopsis_thaliana_(P43082.1)   133 GKCLRVNTRTNAAVTVRIVDQCSNGGGLDLDV-AMFNQIDTDGFGYQQGHLIVDYQFVDCGN 193
Nicotiana_tabacum_(CAA41437.1)    87 GRCLRVNTNTGTGTQTTVRIVDQCSNGGGLDLDV-NVFNQLDDTNGVGYQQGHLIVNYEFVNCND 147
Vitis_vinifera_(AAC33732.1)      83 GKCLSVNTNTATGTQATVRIVDQCSNGGGLDLDS-GVFNKLDDTNGAGYNQQGHLIVNYEFVDCGD 143

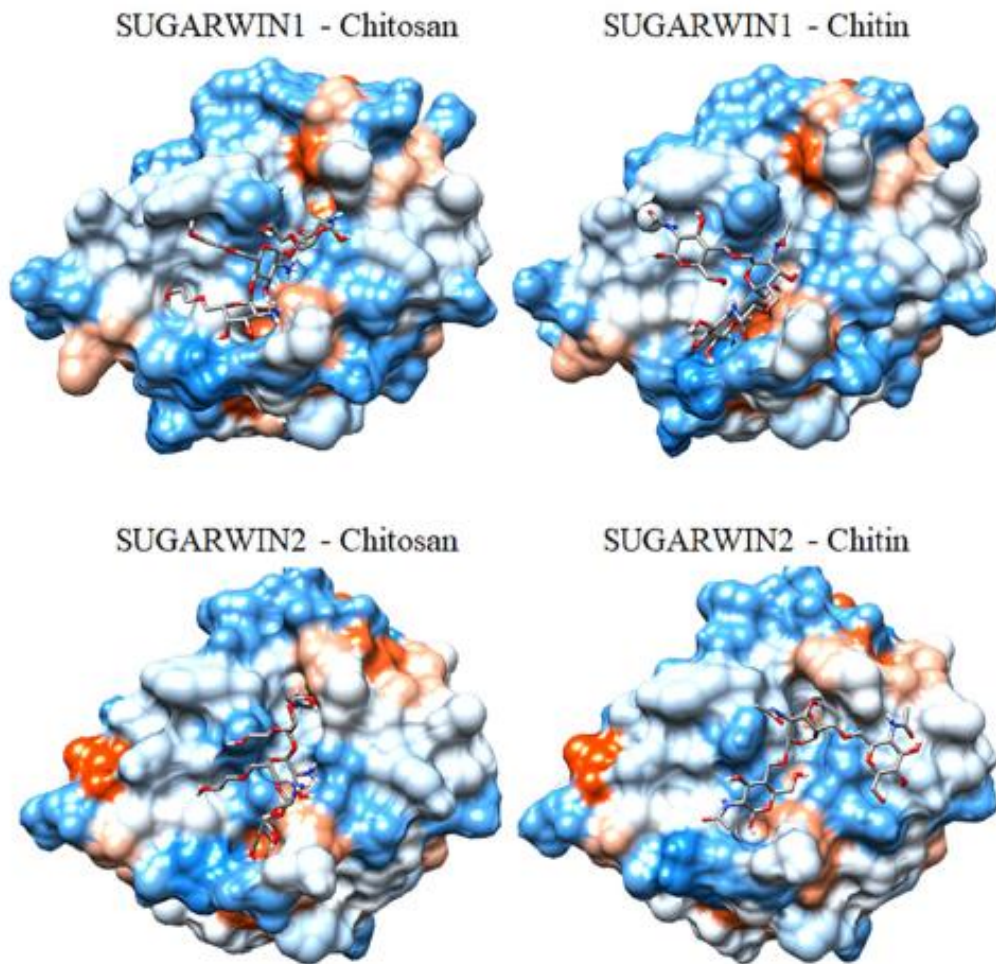
```

**Figura 3.** Alinhamento de sequência para proteínas homólogas de BARWIN. O alinhamento mostra homólogos de diferentes plantas. Os resíduos de sítios de ligação previstos por análises de *docking* para SUGARWIN1 e 2 estão destacados nas caixas. As posições His11 e His111 estão destacadas com um asterisco.

**Fonte:** Franco et al. (2019).

Na predição, dez posições de aminoácidos comuns (enumeradas de acordo com o modelo CARWIN) fizeram contato com glicol-quitosano e/ou quitina em ambas as proteínas, por contatos ligado e não-ligado (Figuras 4 e 5). São elas: 9 (Thr), 11 (His / Asn), 12 (Tyr), 13 (Tyr), 26 (Ser / Gly), 82 (Asp), 83 (Gln), 84 (Cys), 85 (Ser) e 90 (Asp). Estes resultados destacaram a presença de uma região de suposta ligação glicolquitosano/quitina, em ambas as SUGARWINs. No *docking*, algumas diferenças foram observadas, com três posições de aminoácidos contatadas exclusivamente por ligantes em SUGARWIN1 (Tyr10, Ala48 e Asn86), e outros três em SUGARWIN2 (Ala25, Thr27 e Tyr28) (FRANCO et al., 2019).





**Figura 4.** Interação esperada entre as SUGARWINs e glicol-quitosan e quitina. A modelagem molecular acima mostra o complexo previsto entre as proteínas e seus substratos. As superfícies das proteínas estão coloridas de acordo com a hidrofobicidade, azul para resíduos hidrofílico e laranja para resíduos hidrofóbicos.

**Fonte:** Franco et al. (2019).

Nos trópicos, a cana-de-açúcar é uma importante cultura comercial, cultivada por meio de seus colmos, que acumulam altos teores de sacarose. Essa planta contribui com 60% do açúcar bruto produzido no mundo. A podridão vermelha causada por *C. falcatum* é uma das principais doenças da cana-de-açúcar no Brasil e no mundo. Seu principal dano é devido à inversão da sacarose pela invertase induzida por patógenos, resultando em colmos de cana-de-açúcar secos (VISWANATHAN e SAMIYAPPAN, 1999). No Brasil, a presença de *D. saccharalis* e *C. falcatum* estão correlacionadas (Franco et al., 2017; Franco et al., 2021). Nossos trabalhos anteriores mostram que

as proteínas SUGARWINs desempenham um papel importante na defesa da planta contra patógenos fúngico, podendo ser importantes para o controle da podridão vermelha. Embora a indução de transcritos de SUGARWIN ocorra em resposta aos danos de *D. saccharalis* (MEDEIROS et al., 2012), sua importância se deve à especificidade para conter fungos patogênicos da cana-de-açúcar (FRANCO et al., 2014). Ao demonstrar que apenas SUGARWIN1 exibiu atividade ribonuclease e quitinase, Franco et al. (2019) sugeriram que a especificidade enzimática variável mostrada por elas seria resultado da composição de aminoácidos divergente no local de ligação do substrato. Assim, aqui analisamos a biologia estrutural das proteínas SUGARWINs, bem como sua filogenia dentro do grupo das proteínas PR-4, a fim de melhor compreender os mecanismos moleculares que orquestram suas importantes atividades quitosanase, quitinase e RNase na defesa da planta contra patógenos. Com isso, o objetivo do presente trabalho foi avaliar a história evolutiva das proteínas SUGARWINs e suas características estruturais, buscando atribuir uma correlação entre estrutura e função.

#### **4. Biotecnologia e Biologia Estrutural**

Há muito tempo se reconhece que o conhecimento das estruturas 3D de proteínas tem o potencial de acelerar descobertas nas diferentes áreas da biotecnologia. No entanto, desenvolvimentos recentes no sequenciamento do genoma, robótica e bioinformática transformaram radicalmente as oportunidades. Muitas novas proteínas-alvo foram identificadas a partir de análises genômicas e, estudadas por análise de raios-X ou espectroscopia de Ressonância Magnética Nuclear (RMN) (WLODAWER et al., 2013). A primeira difração de raios-X por cristais de proteína foi relatada no início dos anos 30 (BERNAL e CROWFOOT, 1934; CLARK e CORRIGAN, 1932), mas quase 30 anos se passaram antes que a estrutura atômica cristalográfica da mioglobina fosse publicada (KENDREW et al., 1960). A difração de raios-X, seja em material cristalino ou amorfo, ainda é rotineiramente utilizado na indústria farmacêutica (AITIPAMULA, WONG e KANAUIA, 2018; THAKRAL et al., 2018). Mas, pouco explorada na agroindústria.

A ideia de que o conhecimento da estrutura de uma proteína poderia ajudar em projetos de ligantes específicos, apareceu em 1976, alguns anos após o lançamento do *Protein Data Bank* (PDB), em 1971 (BEDDELL et al., 1976; BERNSTEIN et al., 1977). Agora é uma evidência amplamente aceita na literatura, crescente em importância desde que o “ciclo de desenho racional de medicamentos” foi elaborado (HOL, 1986). Onde um medicamento pode ser desenhado racionalmente e otimizado utilizando conhecimentos fornecido pela determinação da estrutura de seu complexo com o alvo macromolecular, que guiará o químico medicinal para a otimização do fármaco (MAVEYRAUD e MOUREY, 2020). Seguindo os moldes do que é feito no desenho racional de medicamentos, busca-se na agricultura fungicidas que favoreçam o manejo, diminuindo os efeitos nocivos no tratamento de pragas e reduzindo o uso de agrotóxicos. Segundo Maveyraud e Mourey (2020), a interação de compostos otimizados com o alvo do fármaco é estruturalmente caracterizada, permitindo o próximo ciclo de otimização química. Assim, mesma estratégia pode ser empregada na agricultura, visando a otimização química de fungicidas.

O número de estruturas macromoleculares depositadas no PDB agora se aproxima de 200 mil (PDB, 2021), com a grande maioria determinada por métodos cristalográficos. Segundo Wlodawer et al. (2013), milhares de artigos que descrevem tais estruturas foram publicados na literatura científica, e 20 prêmios Nobel em química ou medicina foram concedidas por descobertas baseadas em cristalografia macromolecular. Novas ferramentas de *hardware* e *software* fizeram a cristalografia parecer quase uma rotina e muitas estruturas agora estão sendo determinadas por cientistas com pouca experiência nos aspectos práticos do campo. No entanto, esta aparente facilidade é ilusória. Pois procedimentos adequados precisam ser seguidos para manter os altos padrões de qualidade da estrutura. Com isso, percebe-se a necessidade de colaborações com cristalógrafos experientes para que pesquisadores da biotecnologia possam analisar, com segurança e confiabilidade, os resultados obtidos nas análises da biologia estrutural. Pesquisas multidisciplinares que reúnam diferentes pesquisadores com competências distintas tendem a fortalecer o engajamento científico, bem como favorecer o desenvolvimento de soluções rápidas que atendam às necessidades do mercado. Além disso, muitos biotecnologistas não-

cristalógrafos podem ter problemas com a crítica, avaliação e interpretação dos resultados estruturais publicados na literatura científica caso não estejam aliados à cristalógrafos com alta competência técnica.

Com o advento da cristalização automatizada, desenvolvimento de síncrotrons com fontes de raios-X mais potentes, detectores mais rápidos, solução de estrutura automatizada e *pipelines* de refinamento, não é incomum ser capaz de determinar estruturas macromoleculares dentro de alguns poucos dias (DAUTER e WLODAWER, 2016). E, no caso das estruturas de complexos proteína-ligante, a produção de várias por dia pode ser atingível (McINTYRE et al., 2017). Centro Nacionais de Pesquisa, como o Sírius - Laboratório Nacional de Luz Síncrotron (LNLS) do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), tem fornecido suporte técnico completo para pesquisadores brasileiros que recorram à cristalografia de raios-X para compreender as macromoléculas biológicas, visando propor novas soluções para a medicina, o agronegócio e as ciências de materiais. Com a taxa de transferência aumentada, o uso da cristalografia de raios-X progrediu da determinação da estrutura da proteína-alvo, possivelmente na presença de algum ligante pré-identificado, para a determinação da relação estrutura-função por cristalografia, onde várias estruturas de complexos são determinadas a fim de guiar a otimização do ligante (LESUISSE et al., 2002). Os avanços tecnológicos adquiridos na metodologia de cristalografia de raios-X aliado a construção de laboratórios nacionais de pesquisa como o LNLS, tornou mais fácil o acesso de pesquisadores não-cristalógrafos à cristalografia de macromoléculas. De modo que, hoje, pesquisadores do país todo podem enviar suas proteínas ao maior complexo científico do Brasil – o Sírius, para obter com confiabilidade a estrutura 3D de suas proteínas de interesse.

De acordo com Maveyraud e Mourey (2020), a incorporação da biologia estrutural no processo de análise de alvos moleculares, garantiu a oportunidade de pesquisadores usarem informações estruturais detalhadas na agricultura e medicina, a fim de progredir acertos de triagem em *leads* ou alvos, candidatos a ferramentas biotecnológicas. Aqui, a cristalografia de raios-X também provou ser uma ferramenta inestimável, pois é capaz de fornecer informações estruturais extremamente abrangentes sobre a interação de um ligante com um alvo proteico.

Nos anos recentes, emergiu a descoberta de drogas baseadas em fragmentos. Onde a cristalografia de raios-X também se torna uma poderosa tecnologia de triagem, capaz de fornecer informação estrutural sobre complexos envolvendo compostos de baixo peso molecular, apesar da fraca afinidade de ligação. Dado o baixo número de compostos necessários numa biblioteca de fragmentos, em comparação com as centenas de milhares geralmente presentes em bibliotecas de compostos semelhantes, agora se torna viável rastrear uma biblioteca de fragmentos inteiros usando cristalografia de raios-X, fornecendo uma riqueza de detalhes estruturais que vai alimentar o fragmento para o processo biotecnológico (MAVEYRAUD e MOUREY, 2020). Um *screening*, análogo a este empregado na descoberta de drogas, pode ser utilizado na identificação de novas ferramentas biotecnológicas incluindo as aplicações voltadas para a agroindústria.

## 5. Predição de Estrutura

De acordo com Wlodawer et al. (2013), a criação do PDB foi uma das grandes conquistas da biologia estrutural, especialmente tornando este recurso único da ciência moderna disponível gratuitamente para comunidade científica. O PDB, o qual existe há exatos 50 anos (neste ano de 2021, comemoramos os 50 anos da base de dados), oferece total e fácil acesso aos resultados de décadas de investigação de estruturas macromoleculares por cristalografia de raios-X e outros métodos. Quase 200mil estruturas é disponibilizada à comunidade científica. Cada depósito cristalográfico do PDB contém: (a) um cabeçalho com informações sobre o experimento de difração, determinação da estrutura e protocolos de refinamento utilizados, sinalizadores adicionais e outras observações (por exemplo, oligomerização, informações biológicas, etc.); (b) as coordenadas dos átomos que compõem o modelo estrutural da macromolécula biológica; e, em um arquivo fisicamente separado, (c) uma lista de fatores de estrutura experimentais que foram derivados das imagens de difração de raios-X (as estruturas mais antigas eram frequentemente depositado sem os fatores de estrutura) (BERMAN et al., 2000).

Existem *softwares* de computação gráfica útil para renderizar modelos estruturais atômicos em 3D para inspeção e análise junto com os mapas de densidade de elétron correspondente (se as amplitudes do fator de estrutura experimental também forem depositadas). Se necessário, as estruturas podem, portanto, ser reconstruídas e re-refinadas. O PDB, sendo um repositório de estruturas macromoleculares, não pode intervir e modificar os resultados fornecidos pelos investigadores originais, embora conduza a validação das estruturas apresentadas, dando aos investigadores uma chance de corrigir possíveis erros e omissões. No entanto, a responsabilidade final pela veracidade dos depósitos fica exclusivamente com os depositantes.

Várias iniciativas foram iniciadas para inspecionar e validar, promovendo a correção de todos os depósitos (PDB\_REDO) (JOOSTEN et al., 2012), seus ligantes de pequenas moléculas (WEICHENBERGER, POZHARSKI e RUPP, 2013) ou íons de metal (COOPER et al., 2011). Essas ações levaram a identificação de várias estruturas determinadas com baixa qualidade e produziram seus modelos melhorados. A qualidade geral dos modelos depositados e a capacidade dos usuários do PDB de examinar ou mesmo reinterpretar os modelos originais tornam a cristalografia de proteínas uma *'joia da coroa'* da pesquisa experimental. O conteúdo do PDB não deve, no entanto, ser tratado como a última verdade, mas, como sempre na ciência, deve ser visto como um ponto de partida para futuras investigações.

Em dezembro de 2020, a renomada revista científica *Nature* reportou sobre um programa de inteligência artificial (IA) desenvolvido pela *DeepMind* (uma empresa de IA de propriedade da Google) que fez um gigantesco salto na resolução de um dos maiores desafios da biologia: determinar a estrutura 3D de proteínas a partir de sua sequência de aminoácidos (CALLAWAY, 2020). A versão *AlphaFold2* da *DeepMind* apresentada no evento bienal CASP 2020 (Avaliação Crítica de Previsão de Estrutura) superou as outras 100 equipes. Cabe enfatizar que, a capacidade apresentada por esta IA, em prever a estrutura 3D das proteínas com base em sua sequência de aminoácidos, seria um grande benefício para as ciências da vida e medicina. Em alguns casos, as previsões de estrutura do algoritmo *AlphaFold* foram indistinguíveis daquelas determinadas usando os métodos experimentais “padrão-ouro” como a

Cristalografia e a Microscopia Crioelétrica (crio-EM). A primeira interação de *AlphaFold* envolve métodos de *deep learning* em dados genéticos e estruturais para prever a distância entre os pares de aminoácidos na proteína. A segunda etapa envolve uma rede de IA que incorpora informações adicionais sobre restrições físicas e geométricas que determinam como a proteína se dobra. Como tarefa mais difícil, em vez de prever apenas relações entre os aminoácidos, a rede prevê a estrutura final de uma sequência da proteína-alvo em um sistema mais complexo. Embora os custos com os métodos experimentais sejam altos, ferramenta como esta IA não evita a necessidade dos métodos experimentais. No entanto, a IA fará o possível para estudar moléculas vivas de novas maneiras.

Em julho deste ano, a também renomada revista *Science*, noticiou sobre a possibilidade de predição da estrutura de proteínas de forma rápida e fácil. Baek e colaboradores (2021) recentemente publicaram sua nova ferramenta de IA que usa tripla rede neural para prever com acurácia sobre a estrutura das proteínas e suas interações (PENNISI, 2021). Dada, para modelar, uma nova proteína à *RoseTTAFold*, o algoritmo de Baek e Baker prossegue ao longo de três *tracks*. Um compara o aminoácido da sequência da proteína com todas as sequências semelhantes nos bancos de dados de proteínas. O outro, prediz em pares as interações entre os aminoácidos dentro da proteína. E o terceiro, compila a suposta estrutura 3D. Cabe enfatizar que o programa salta entre os *tracks* para refinar o modelo, usando o *output* de cada um, para atualizar os outros (BAEK et al., 2021). Diferentemente do que foi feito pela *DeepMind*, tais autores disponibilizaram amplo acesso gratuito para comunidade científica. De modo que hoje, temos uma abordagem de IA acessível a todos os pesquisadores. O que nos remete a discussão da relevância das comunidades *open source*.

Ambos os métodos funcionam bem. Embora o *AlphaFold2* apresente como resultado, estruturas mais precisas. A abordagem do algoritmo de Baek é melhor. Uma vez que, capta “a essência e particularidades da estrutura da proteína”, como a identificação de cordas de átomos saindo das laterais da proteína, que é um recurso essencial para as interações entre proteínas. É improvável que tanto *AlphaFold2* e *RoseTTAFold* eliminem a necessidade de métodos experimentais para resolver

estruturas de proteínas em laboratórios. No entanto, a análise está convergindo mais no sentido de exigir mais raciocínio e menos pipetagem.

## **6. Justificativa e Aplicabilidade**

As proteínas são os blocos de construção da vida, responsáveis pela maior parte do que acontece dentro das células. Como estas macromoléculas funcionam e o que fazem é determinado por sua estrutura tridimensional. As proteínas tendem a adotar sua forma guiada apenas pelas leis da física. Com isso, para resolver o problema de estrutura, a biologia recorreu a métodos físicos e matemáticos, visando compreender melhor o comportamento das macromoléculas biológicas. A partir de 1950 foram determinadas as primeiras estruturas de proteínas. Por anos, o principal método experimental empregado foi a Cristalografia de raios-X. No entanto, hoje, a crio-EM tornou-se a ferramenta favorita de muitos laboratórios de biologia estrutural (CALLAWAY, 2020).

Neste contexto, a tarefa mais difícil enfrentada pela ciência da vida é compreender como os aminoácidos que constituem uma proteína mapeiam as torções e dobras de sua estrutura final, que ficou conhecida como o “problema de estrutura”. Este problema existe há 50 anos. De acordo com Pennisi (2021), os cientistas sonharam em simplesmente prever a estrutura de proteínas a partir da sua sequência de aminoácidos. Habilidade que abriria um mundo de percepções sobre o funcionamento da vida. Considerando que, a compreensão da estrutura destes importantes blocos de construção das células pode nos ajudar na descoberta de novos medicamentos, bem como, de produtos para a agroindústria, fica evidente a importância dos métodos experimentais e computacionais da biologia estrutural na biotecnologia.

A ciência da cristalografia definida de maneira bastante restrita, tem um caráter excepcionalmente central e interdisciplinar entre as ciências naturais e da vida. Não faz fronteira apenas com a matemática, física, química e biologia, mas também mineralogia e medicina. Ela desempenha um papel fundamental na resolução da estrutura tridimensional de todos os tipos de moléculas químicas e, juntos com suas variantes, foi usada para determinar a estrutura de todos os tipos de macromoléculas



biológicas, de ácidos nucleicos e proteínas a vírus e ribossomos (WLODAWER et al., 2013). No entanto, muitas funções biológicas dependem de interações proteína-proteína que não podem ser completamente entendidas pelos métodos experimentais. De acordo com Srivastava e colaboradores (2018), a biologia estrutural da proteína percorreu um longo caminho desde a determinação da primeira estrutura tridimensional da mioglobina há décadas. Ao longo deste período, a cristalografia de raios-X foi o método experimental mais importante para obter informações de resolução atômica sobre estruturas de proteínas. Porém, à medida que o papel da dinâmica ganhou importância na função das proteínas, as limitações da cristalografia de raios-X vieram à tona. Métodos computacionais provaram ser imensamente bem-sucedidos na compreensão da dinâmica das proteínas em solução, e eles continuam a melhorar em termos de escala e tipos de sistemas em que podem ser estudados.

O fungo *Fusarium verticillioides* é um dos causadores da podridão vermelha, uma das principais doenças da cana-de-açúcar. A cada safra os prejuízos atingem R\$ 5 bilhões somente no Brasil (ARANTES, 2021). Considerando o potencial biotecnológico das SUGARWINS para a agroindústria brasileira, aqui analisamos a biologia estrutural das proteínas SUGARWINS, bem como sua história evolutiva dentro do grupo das proteínas PR-4, a fim de melhor compreender os mecanismos moleculares que orquestram suas importantes atividades quitosanase, quitinase e RNase na defesa da planta contra fitopatógenos. Frente às limitações da cristalografia de raios-X no estudo da dinâmica de proteínas, recorreremos também à diferentes métodos computacionais de dinâmica molecular e filogenia molecular, visando compreender o comportamento funcional dessas macromoléculas.

## OBJETIVOS

### **Geral:**

Compreender o papel das SUGARWINS na interação planta-inseto-patógeno existente entre cana-de-açúcar, *D. saccharalis*, e os fungos fitopatogênicos *F. verticillioides* e *C. falcatum*, a partir da análise comparativa da biologia estrutural das proteínas SUGARWINS.

### **Específicos:**

1. Determinar a estrutura tridimensional das SUGARWINS;
2. Analisar a filogenia molecular destas proteínas;
3. Realizar dinâmica molecular dos modelos estruturais;
4. Realizar uma análise estrutural comparativa destas proteínas;
5. Identificar domínios específicos, correlacionando a estrutura-função das proteínas SUGARWINS;
6. Compreender a história evolutiva dessas macromoléculas.

## CAPÍTULO I

### Artigo

#### ***Structural and Evolutionary Analyses of PR-4 SUGARWINS Points to a Different Pattern of Protein Function.***

**Lorhenn Bryanda Lemes Maia**<sup>1</sup>, Humberto D’Muniz Pereira<sup>2</sup>, Richard Charles Garratt<sup>2</sup>, José Brandão-Neto<sup>3</sup>, Flavio Henrique-Silva<sup>4</sup>, Danyelle Toyama<sup>4</sup>, Renata O. Dias<sup>5</sup>, José Fernando Ruggiero Bachega<sup>6,7</sup>, Julia Vasconcellos Peixoto<sup>6</sup> and Marcio C. Silva-Filho<sup>1,8\*</sup>

#### ***Frontiers in Plant Science***

Fator de Impacto: 5.753

Qualis CAPES (BIOTECNOLOGIA): A1

Recebido: 30 Junho 2021

Aceito: 10 Agosto 2021

Publicado: 09 Setembro 2021

doi: 10.3389/fpls.2021.734248



## Structural and Evolutionary Analyses of PR-4 SUGARWINS Points to a Different Pattern of Protein Function

Lorhenn Bryanda Lemes Maia<sup>1</sup>, Humberto D'Muniz Pereira<sup>2</sup>, Richard Charles Garratt<sup>2</sup>, José Brandão-Neto<sup>3</sup>, Flavio Henrique-Silva<sup>4</sup>, Danyelle Toyama<sup>4</sup>, Renata O. Dias<sup>5</sup>, José Fernando Ruggiero Bachega<sup>6,7</sup>, Julia Vasconcellos Peixoto<sup>7</sup> and Marcio C. Silva-Filho<sup>1,8\*</sup>

<sup>1</sup> Programa de Pós-Graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, Brazil,

<sup>2</sup> Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil,

<sup>3</sup> Diamond Light Source, Harwell Science and Innovation Campus Didcot, Harwell, United Kingdom,

<sup>4</sup> Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, Brazil,

<sup>5</sup> Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil,

<sup>6</sup> Departamento de Farmacociências, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil,

<sup>7</sup> Programa de Pós-Graduação de Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil,

<sup>8</sup> Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, Brazil

### Edited by:

Steffen P. Graether,  
University of Guelph, Canada

### Reviewed by:

Qing-Bo Yu,  
Shanghai Normal University, China  
Rasappa Viswanathan,  
Indian Council of Agricultural  
Research (ICAR), India

### \*Correspondence:

Marcio C. Silva-Filho  
mdcsilva@usp.br

### Specialty section:

This article was submitted to  
Plant Proteomics and Protein  
Structural Biology,  
a section of the journal  
Frontiers in Plant Science

Received: 30 June 2021

Accepted: 10 August 2021

Published: 09 September 2021

### Citation:

Maia LBL, Pereira HD'M, Garratt RC, Brandão-Neto J, Henrique-Silva F, Toyama D, Dias RO, Bachega JFR, Peixoto JV and Silva-Filho MC (2021) Structural and Evolutionary Analyses of PR-4 SUGARWINS Points to a Different Pattern of Protein Function. *Front. Plant Sci.* 12:734248. doi: 10.3389/fpls.2021.734248

SUGARWINs are PR-4 proteins associated with sugarcane defense against phytopathogens. Their expression is induced in response to damage by *Diatraea saccharalis* larvae. These proteins play an important role in plant defense, in particular against fungal pathogens, such as *Colletotrichum falcatum* (Went) and *Fusarium verticillioides*. The pathogenesis-related protein-4 (PR-4) family is a group of proteins equipped with a BARWIN domain, which may be associated with a chitin-binding domain also known as the hevein-like domain. Several PR-4 proteins exhibit both chitinase and RNase activity, with the latter being associated with the presence of two histidine residues H11 and H113 (BARWIN) [H44 and H146, SUGARWINs] in the BARWIN-like domain. In sugarcane, similar to other PR-4 proteins, SUGARWIN1 exhibits ribonuclease, chitosanase and chitinase activities, whereas SUGARWIN2 only exhibits chitosanase activity. In order to decipher the structural determinants involved in this diverse range of enzyme specificities, we determined the 3-D structure of SUGARWIN2, at 1.55Å by X-ray diffraction. This is the first structure of a PR-4 protein where the first histidine has been replaced by asparagine and was subsequently used to build a homology model for SUGARWIN1. Molecular dynamics simulations of both proteins revealed the presence of a flexible loop only in SUGARWIN1 and we postulate that this, together with the presence of the catalytic histidine at position 42, renders it competent as a ribonuclease. The more electropositive surface potential of SUGARWIN1 would also be expected to favor complex formation with RNA. A phylogenetic analysis of PR-4 proteins obtained from 106 Embryophyta genomes showed that both catalytic histidines are widespread among them with few replacements in these amino acid positions during the gene family evolutionary history. We observe that the H11 replacement by N11 is also present in two other sugarcane PR-4 proteins: SUGARWIN3 and SUGARWIN4. We propose that RNase activity was present in the first Embryophyta PR-4 proteins but was recently lost in members of this family during the course of Evolution.

**Keywords:** SUGARWIN, BARWIN, crystallography, flexible loop, PR-4, phylogenetic analysis

## INTRODUCTION

Plant responses to insects and pathogens are complex and modulate the expression of a large number of genes, many of which are believed to have a direct role in plant defense (Xu et al., 1994; Agrawal et al., 2003; Banno et al., 2005; Franco et al., 2017). Pathogen recognition by plants activates the host defense response resulting in the accumulation of pathogenesis-related proteins (PR proteins) (Pieterse and van Loon, 1999). Several genes that encode wound-inducible proteins (WIN) have been identified in various plant species (Ryan, 1990; Christopher et al., 2004; Schillmiller and Howe, 2005; Medeiros et al., 2012).

The pathogenesis-related protein-4 (PR-4) family is a group of proteins composed of a BARWIN domain, which was first identified in the BARWIN protein from barley (Hejgaard et al., 1992; Svensson et al., 1992). Although the BARWIN domain can be found in association with a chitin-binding domain typical of lectins, known as a hevein-like domain (Broekaert et al., 1990), several proteins show the absence of this protein motif. PR-4 proteins composed only by BARWIN domains were identified in several plants, such as: tobacco (Friedrich et al., 1991), tomato (Linthorst et al., 1991), *Arabidopsis* (Potter et al., 1993), wheat (Caruso et al., 1999), *W. japonica* (Kiba et al., 2003), maize (Bravo et al., 2003), rice (Agrawal et al., 2003; Zhu et al., 2006) and *L. radiata* (Li et al., 2010). The PR-4 proteins are grouped into class I and class II based on the presence or absence of this chitin-binding domain (Neuhaus et al., 1996).

The PR-4 BARWIN homologs of several plant species have been associated with the plant responses to fungal infection and mechanical wounding (Linthorst et al., 1991; Hejgaard et al., 1992; Caruso et al., 1999, 2001; Agrawal et al., 2003; Bravo et al., 2003; Kiba et al., 2003; Zhu et al., 2006; Li et al., 2010; Medeiros et al., 2012; Bai et al., 2013; Franco et al., 2014, 2019; Menezes et al., 2014; Souza et al., 2017). PR-4 proteins have been classified as chitinases (Neuhaus et al., 1996; Van Loon and Van Strien, 1999); however, several studies have also reported RNase activity for BARWIN-like proteins (Bertini et al., 2009, 2012; GuevaraMorato et al., 2010; Bai et al., 2013; Huet et al., 2013; Franco et al., 2014; Menezes et al., 2014; Kim and Hwang, 2015). RNA binding sites have been described for WHEATWIN1 (Bai et al., 2013) and

CARWIN (Huet et al., 2013), showing two important histidine residues necessary for RNase activity, one at position 11 and another at position 113, numbered according to the mature BARWIN sequence, which may be used as the reference residue numbering in this text (Bai et al., 2013). Combined DNase and RNase antifungal activities were also observed for the *Capsicum chinense* PR-4 protein (Guevara-Morato et al., 2010) and the *Theobroma cacao* TcPR-4b protein (Menezes et al., 2014).

Our previous studies have identified two homologous genes to BARWIN in sugarcane: *SUGARWIN1* and *SUGARWIN2* (Falco et al., 2001; Medeiros et al., 2012). SUGARWIN proteins (sugarcane wound-inducible proteins) are believed to be part of a defense mechanism against pathogenic fungi in sugarcane plants *via* a salicylic acid (SA)-independent and jasmonic acid (JA)-dependent pathway. These proteins are secreted to the apoplasts, are strongly upregulated in response to mechanical wounding, *Diatraea saccharalis* (Fabricius) attack, methyl jasmonate treatment and to a lesser extent to sugarcane fungus (Medeiros et al., 2012). Despite their high expression levels in response to *D. saccharalis* attack, these proteins have no effect on insect development and mortality. However, SUGARWINs have antimicrobial effects in pathogenic fungus causing changes in hyphae morphology and leading to cell death by apoptosis (Medeiros et al., 2012; Franco et al., 2014). We have recently shown that *Fusarium verticillioides*, a pathogenic fungus, modulates sugarcane plants and *Diatraea saccharalis* moths and larvae to increase its dissemination in the field (Franco et al., 2021). The antifungal activity of SUGARWIN proteins is likely due to its multiple enzyme properties, such as: chitosanase, RNase and chitinase activities. This variable enzyme specificity has been suggested to be a result of a divergent amino acid composition at the substrate-binding site of these two proteins (Franco et al., 2014).

In order to understand how these SUGARWINs acquired these divergent activities, here we analyze their evolutionary history and their three-dimensional structures, looking for a correlation between structural and functional properties. Furthermore, we also describe two new putative SUGARWIN genes: *SUGARWIN3* and *SUGARWIN4*, which may also lack the canonical RNase activity observed in other

PR-4 proteins as both have asparagine instead of histidine in the amino acid position 11 (43 - SUGARWIN3; 38 -SUGARWIN4).

## MATERIALS AND METHODS

### Recombinant Expression and Purification of SUGARWIN2

The heterologous expression of SUGARWIN2 was performed in *Pichia pastoris* as previously described (Medeiros et al., 2012). Basically, a single colony of *P. pastoris* containing the pPICZ $\alpha$ SUG2, that contains the coding region for SUGARWIN2 was inoculated in 10 ml of BMGY medium [1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer (pH 7.0), 1.34% YNB, 10<sup>-5</sup>% biotin, and 1% glycerol], which was incubated at 30°C until reaching an optical density (OD) at 600 nm of about 5. This culture was used to inoculate 500 ml of BMGY and was grown to an OD of 4–5. The cells were harvested by centrifugation at 1,500  $\times$  g for 5 min, resuspended in 100 ml of BMGY with 0.5% methanol instead of glycerol, and incubated at 28°C. For induction, methanol was added to each sample every 24 h to maintain a final concentration of 0.75%. After 96 h, the cells were harvested by centrifugation at 1,500  $\times$  g for 5 min, and the supernatant was filtered with a 0.45 mm membrane (Millipore, Bedford, MA, United States). The recombinant protein in the supernatant was purified by immobilized metal affinity chromatography (IMAC) using NiNTA-agarose (Qiagen) pre-equilibrated with purification buffer (10 mM Tris-HCl, pH 8.0; 50 mM NaH<sub>2</sub>PO<sub>4</sub>; and 100 mM NaCl). After binding, the proteins were eluted with two-column volumes of purification buffer containing increasing imidazole concentrations (10, 25, 50, 75, 100, and 250 mM). The fractions containing the recombinant protein were pooled, dialyzed in phosphate-buffered saline (PBS, pH 7.4) (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>PO<sub>4</sub> and 2 mM KH<sub>2</sub>PO<sub>4</sub>) and sterilized with a 0.22 mm filter (Millipore). The protein concentration was determined using a BCA protein assay kit (Pierce).

The purification of the protein was monitored using 12% SDS-PAGE (**Supplementary Figure 1**). The recombinant SUGARWIN2 sample was further purified in an AKTA Purifier (GE, Healthcare) with a gel filtration column (Superdex 75) in TRIS solution (20 mM; pH 7.4), with 200 mM of NaCl. The final purification of the



recombinant protein SUGARWIN2 was monitored by Dynamic Light Scattering (DLS) (**Supplementary Figure 2**).

### **Crystallization, Data Collection, Structure Determination and Refinement of SUGARWIN2**

Recombinant SUGARWIN2 from gel filtration experiment at 7 mg/mL were crystallized by the sitting drop vapor diffusion method using the Morpheus screening kit (Molecular Dimensions). 200 nL drops were setup employing a Crystal Gryphon robot (Art Robbins). The plate was incubated at 18°C. Several conditions resulted in bipyramidal crystals after 1 week. The crystals were flash-cooled in liquid nitrogen for data collection. The structure presented here is from a crystal grown in 0.1 M MES/imidazole pH 6.5, 10% w/v PEG 4000, 20% v/v glycerol 0.03 M sodium fluoride 0.03 M sodium bromide and 0.3 M sodium iodide.

X-ray diffraction data were collected on beamline I24 of the Diamond Light Source (DLS), United Kingdom, ( $\lambda = 0.96863 \text{ \AA}$ ) using a PILATUS3 6 M detector. The diffraction data were indexed, integrated and scaled using the xia2 package (REF). The data were processed up to 1.55 Å resolution. The structure was solved by molecular replacement using Phaser (McCoy et al., 2007); and papaya BARWIN-like protein (CARWIN, PDB ID: 4JP7) as search model after modification using Chainsaw (Stein, 2008). The structure was refined using Phenix (Adams et al., 2011) and COOT (Emsley and Cowtan, 2004) was used for model building into  $\sigma_a$ -weighted 2Fo-Fc and Fo-Fc electron density maps. R and Rfree were monitored to evaluate the validity of the refinement protocol, and the stereochemistry of the model was assessed using Molprobit (Chen et al., 2010). The coordinates and structure factors have been deposited in Protein Data Bank (PDB ID: 7KSN).

The data-collection, processing and refinement statistics are shown in **Table 1**.

**Table 1.** Data-collection, processing and refinement statistics

SUGARWIN2	
Detector	Pilatus3 6M
Cell parameters (Å) a, b, c, $\alpha$ , $\beta$ , $\gamma$	227.51, 227.51, 227.51, 90.0, 90.0, 90.0
Space Group	F4132
Resolution (Å)	52.20 – 1.51 (1.55 – 1.55)
X ray Source	DLS I24
$\lambda$ (Å)	0.96863
Multiplicity	75.7 (72.2)
R <sub>pim</sub> (%)	1.7 (42.1)
CC (1/2)	0.999 (0.791)
Completeness (%)	100.0 (100.0)
Reflections	5976527 (415898)
Unique Reflections	78971 (5764)
$I/\sigma$	23.2 (2.4)
Reflections used for Refinement	78910
R(%)	16.26
R <sub>free</sub> (%)	18.50
N° of protein atoms	1848
N° of ligant atoms	0
B (Å <sup>2</sup> )	13.88
Coordinate Error (Å)	0.13
Phase error (°)	16.52
<i>Ramachandran Plot</i>	
Favored (%)	96.72
Allowed (%)	3.28
Outliers (%)	0.0
All-atom Clashscore	4.23
<i>RMSD from ideal geometry</i>	
r.m.s. bond lengths (Å)	0.014
r.m.s. bond angles (°)	1.291
PDB ID	7KSN

### Gene Sequences and Phylogenetic Analysis

Protein sequences homologous to SUGARWINs 1 and 2 were searched for in 106 Embryophyta genomes obtained from the NCBI RefSeq Genome database and in the genomes of *Saccharum spontaneum* (Zhang et al., 2018) and the Brazilian hybrid cultivar *Saccharum spp.* SP80-3280 (Souza et al., 2019), using the BLASTp algorithm

(Altschul et al., 1990). For genes with multiple isoforms of splicing, only the longest one was retrieved for further analyses. Protein domains were predicted using the InterProScan software (Jones et al., 2014).

A Maximum Likelihood phylogenetic tree inference was predicted using the IQ-TREE software v.1.6.12 (Nguyen et al., 2015) employing the automatic amino acid substitution model selection. Branch support values were calculated using 100 bootstrap replications. The protein sequence alignment used in the phylogenetic tree analysis was predicted using MAFFT software v7.453 (Kato and Standley, 2013) and filtered using trimAL v.1.4 with the -automated1 method (Capella-Gutierrez et al., 2009).

### **Molecular Modeling, Molecular Dynamics and Homology Modeling of SUGARWINs**

All the SUGARWIN systems were prepared in AmberTool19 (Olsson et al., 2011) using the ff99SB force field for proteins and TIP3P for water molecules. The conversion of the original structures to the amber topology and coordinates, as well as the addition of water molecules and counter ions was performed with the *tLeap* program. A buffer of at least 10 Å between protein and the periodic box wall was used. For the SUGARWIN2 model, the input coordinates were taken from the crystallographic structure presented in this work (PDB ID: 7KSN). For SUGARWIN1, SUGARWIN3 and SUGARWIN4 the coordinates were taken from those obtained from comparative models. Protonation states of the protein were assigned at pH 7.5 using the PROPKA3 (Olsson et al., 2011) web server.

The MD simulations were performed using NaMD, version 2.14 (Olsson et al., 2011). The time step and the temperature were set to 2.0 fs and 300 K, respectively. Simulations were performed in the isothermal-isobaric ensemble (NPT) and periodic boundary conditions were employed, with Van der Waals interactions computed using a 12.0 Å cut-off, and electrostatic contributions calculated *via* a particle mesh Ewald method using a grid with 1.0 Å spacing. Prior to MD, each of the systems was energy minimized with a steepest-descent energy algorithm. Altogether, MD simulations were performed for a duration of 100 ns after equilibration, with snapshots of the system

being saved every 20 ps, giving 5,000 snapshots for subsequent analysis. All the trajectories were analyzed in Bio3D package (Grant et al., 2006).

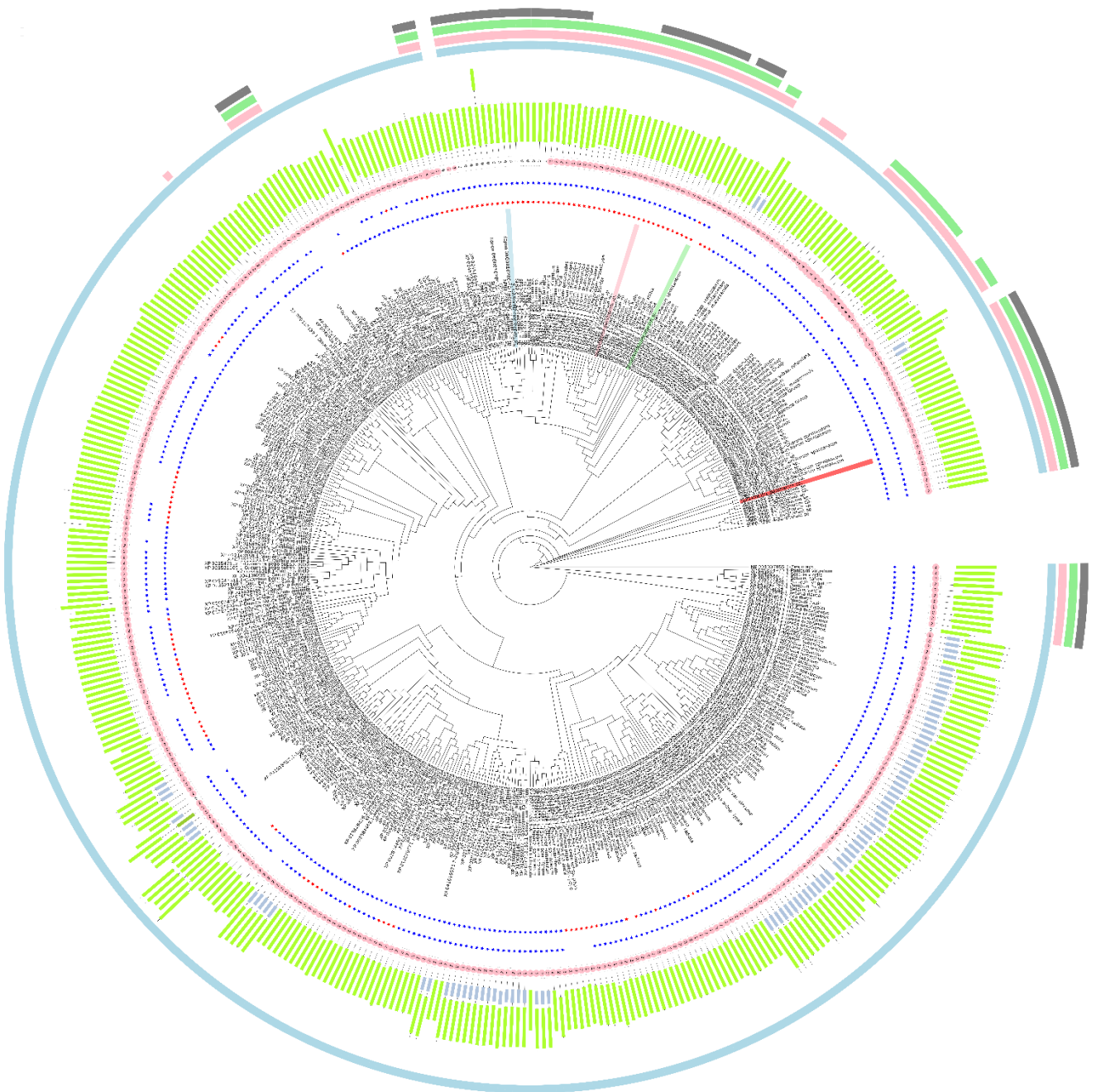
The homology modeling was performed using Modeller v9.24 (Marti-Renom et al., 2000) employing the coordinates of the Sugarwin2 as model. 200 models were built for each SUGARWIN and the best model was selected by using internal Modeller energy score.

## RESULTS

### PR-4 Phylogenetic Proposition

A total of 436 PR-4 protein sequences were identified in our searches of Embryophyta genomes. These sequences were present in 101 out of 106 analyzed genomes, corresponding to 95.3% of the total. PR-4 genes were found to be absent from the five species of Asterales (3) and Lamiales (2), but were present in the form of multiple copies in 94 of the remaining genomes (**Supplementary Table 1**).

The BARWIN domain (Pfam ID: PF00967) was found in all protein sequences analyzed, whereas a “chitin recognition protein” domain (Pfam ID: PF00187) was found in 92, which corresponds to 21.1%. In addition, the chitin recognition domain was absent in all sequences from Liliopsida species. Among the 436 PR-4 protein sequences found in our searches, 312 and 101 of them had histidine or asparagine at position 11 [H42 - SUGARWIN1; N37 - SUGARWIN2; N43 - SUGARWIN3; N38 - SUGARWIN4], respectively, whereas 362 and 18 had histidine or asparagine at position 113 [H142 - SUGARWIN1; H135 - SUGARWIN2; N142 - SUGARWIN3; H137 - SUGARWIN4] (**Figure 1**). These two positions have been associated with the different catalytic activities attributed to PR-4 proteins and a total of 268 of the sequences examined (61.5%) had both histidine residues H11 and H113 in their catalytic site.



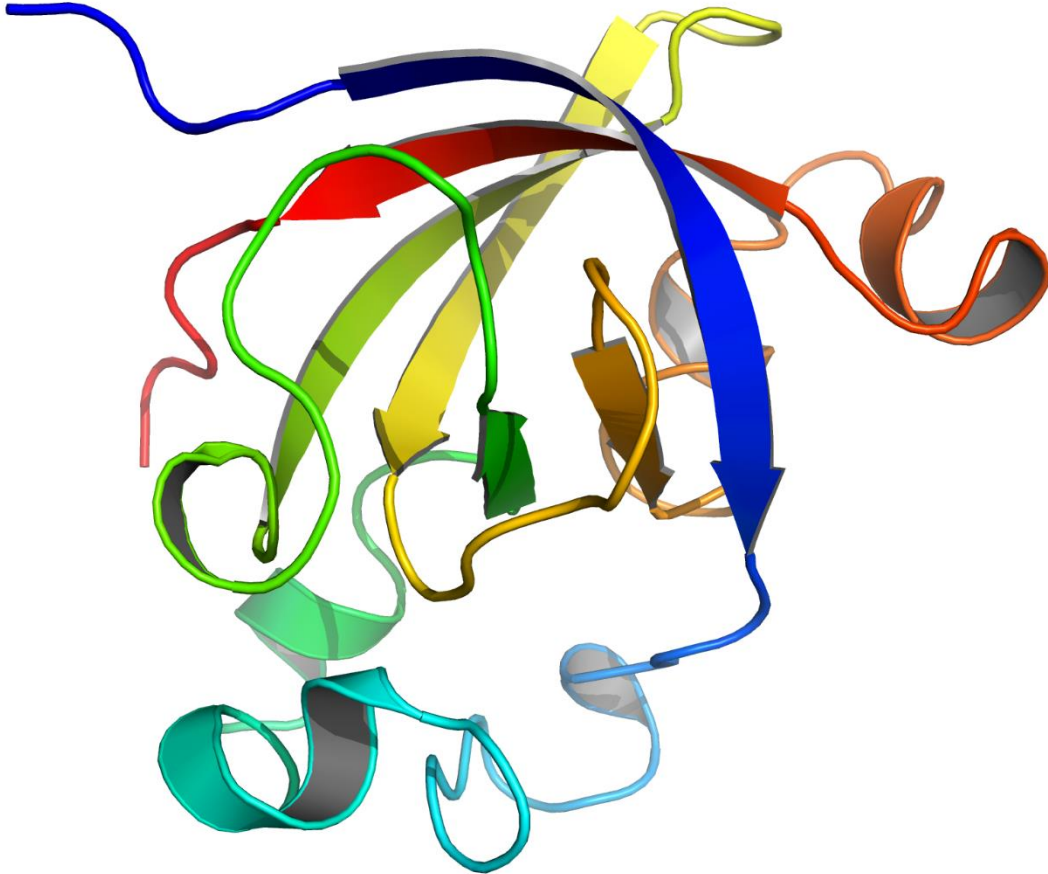
**FIGURE 1** | Phylogenetic analysis for 436 PR-4 proteins found in the genome annotation from 106 Embryophyta species. Labels colored as red, blue, green and pink highlight the positions of SUGARWINS 1, 2, 3, and 4, respectively. The first- and second-star rings represent the amino acid composition in the positions H11 and H111, respectively, with blue and red stars representing histidine and asparagine residues, respectively. The pink dot plots represent the number of additional residues in relation to SUGARWIN2 in the loop corresponding to the SUGARWIN1 variable loop, the larger the circle, the greater the number of additional amino acids. In the next circle, green and blue bars represent the positions of the BARWIN and chitin recognition domains, respectively, according to PFAM results.

The set of four outermost circles shows the sequences from the taxonomic groups: Magnoliopsida (light blue), Liliopsida (pink), Poales (light green), and PACMAD (gray).

### **Crystallography and Molecular Modeling**

PR-4 proteins lacking the His11-His113 pair are less common and do not have its crystal structure, therefore, we solved the crystal structure of SUGARWIN2. Analysis of the protein structure shows an asparagine residue at position 11, which is followed by a loop shorter by two residues when compared either with SUGARWIN1, either with the canonical structures of BARWIN and CARWIN.

SUGARWIN2 crystallizes in cubic space group  $F4_132$  with two molecules per asymmetric unit. The structure was refined to R and  $R_{\text{free}}$  values of 16.26 and 18.50% respectively. The two molecules are in the same conformation with an RMSd of 0.18 Å on  $C\alpha$  positions. When compared with CARWIN from papaya, the RMSd is 0.45 Å, indicating a very well conserved 3D fold. SUGARWIN2 presents the classical BARWIN fold (CATH classification 2.40.40) consisting of a six-stranded  $\beta$ -barrel surrounded by 4  $\alpha$ -helices and several loops (**Figures 2, 3**). Three disulfide bonds (Cys30-Cys62, Cys51-Cys85, Cys65-Cys122) help to maintain the BARWIN domain stability.

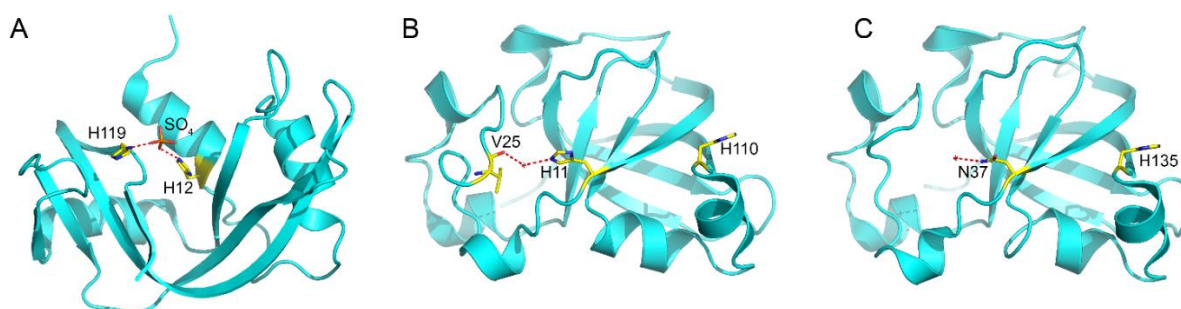


**FIGURE 2** | Crystallographic structure of SUGARWIN2. The structure of SUGARWIN2 at 1.55 Å shown in cartoon-style using the “Chainbow” representation in which the N-terminus is dark blue and the C-terminus red. SUGARWIN2 has the classical BARWIN-like fold. The variable loop between the first and second b-strands (blue and green respectively) is two residues shorter than in previously reported structures.

Classical ribonucleases, such as RNase A, possess two catalytic histidine residues in their active sites. These residues are involved in a general acid-base mechanism leading to substrate hydrolysis (**Figure 3A**). The absence of RNase activity in SUGARWIN2 is believed to be due to the substitution of His11 by an asparagine. SUGARWIN1, BARWIN and CARWIN proteins have both histidines at the same positions and display ribonuclease activity. In both RNase A (PDB ID: 5RSA) and CARWIN (PDB ID: 4JP7) the distance between C $\alpha$ 's of the two histidines is about 10 Å. This is also the case in the inactive SUGARWIN2 structure reported here, albeit that one has been substituted by an asparagine residue. However, different from RNase A, where the histidine sidechains face one to another in a catalytically competent



arrangement, CARWIN structure shows that these residues point to opposite directions. Similar observation is reported for the histidine/asparagine pair in SUGARWIN2. Taking all together, we suggest that SUGARWIN1 and other ribonuclease PR-4 proteins must provide a significant flexibility at the active site to assure the catalytic histidine properties for catalysis (see below). In addition, the histidine sidechains face one another in a catalytically competent arrangement, whereas in CARWIN they point in opposite directions and the same is true of the histidine/asparagine pair in SUGARWIN2 (**Figure 3**). Together, these observations suggest that in the case of SUGARWIN1 and other ribonuclease active PR-4 proteins, there must be significant flexibility to the active site region, in particular the catalytic histidines, to re-orientate into positions appropriate for catalysis (see below).



**FIGURE 3** | The Catalytic Site in (A) Ribonuclease A, (B) CARWIN, and (C) SUGARWIN2. In RNaseA, the catalytic histidines are orientated toward one another and bound to a sulfate ion which mimics the substrate where hydrolytic attack occurs. In CARWIN, although present, the histidines are orientated away from one another and the side chain of His11 is connected to the insertion within the variable loop by water mediated hydrogen bonds. In order to cleave RNA, it is expected that RNase active enzymes (including SUGARWIN1) must undergo a large conformational change to bring the two histidines into the correct orientation for catalysis. The presence of Asn37 in SUGARWIN2, in place of the histidine, makes this protein RNase inactive.

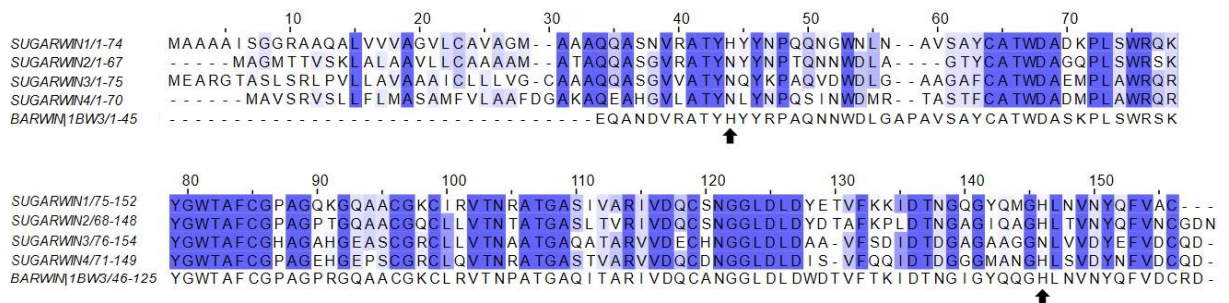
According to our observations there is a variable loop following the first  $\beta$ -strand which is two residues shorter in SUGARWIN2 when compared with the ribonuclease-active SUGARWIN1, BARWIN and CARWIN. In the latter, the inserted residues are Lys-Val which form a type I  $\beta$ -turn, where the valine carbonyl is hydrogen bonded to the catalytic His11 via a water molecule, thereby connecting the active site to the variable loop. The equivalent insertion in SUGARWIN1 is AsnAla and would be expected to behave similarly. In addition, for the SUGARWIN2 the shorter loop forms a type II  $\beta$ -turn (facilitated by the presence of Gly24) and there is no connectivity to the



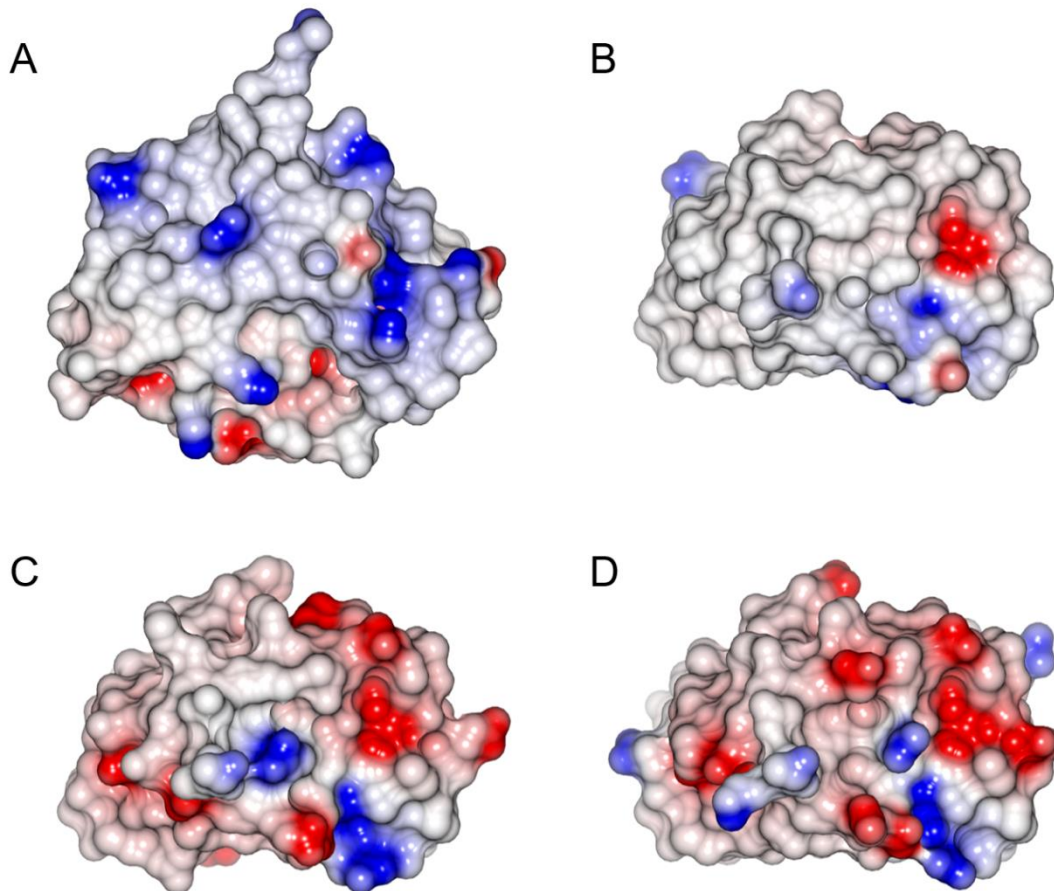
asparagine, which replaces His11 at the active site. Instead, the asparagine side chain appears not to be fixed in a unique conformation but rather, adopts different rotamers in the two monomers of the asymmetric unit. However, in both of them it is structurally disconnected from the variable loop which lies further away due to the lack of the two-residue insertion.

SUGARWIN1 and SUGARWIN2 sequence analysis reveals several substitutions affecting charged amino acid residues. The majority of these alterations correspond to the substitution of basic residues in SUGARWIN1 by neutral residues in SUGARWIN2 (**Figure 4**). These changes have a dramatic consequence for the pI of the two proteins: 8.78 for SUGARWIN1 and 4.82 for SUGARWIN2. By generating a homology model for SUGARWIN1 based on the crystal structure of SUGARWIN2 it was possible to map these residues onto the two molecular surfaces. Figure 5 shows that they give rise to significant differences to the surface electrostatic potential. Such differences would be expected to have consequences for the binding partners of the two proteins and the molecular recognition processes which govern their interaction. For example, the binding of RNA would be expected to be facilitated in the case of SUGARWIN1 compared with SUGARWIN2 based on electrostatic considerations. These changes, together with the absence of the catalytic histidine (His11), help to explain the lack of RNase activity of the latter.

Molecular modeling was also used to build 3D structures for these recently described SUGARWINs and the electrostatic surface analysis shows a mainly negative surface predicting a weak binding to RNA (**Figure 4**). Molecular modeling electrostatic surface analysis were also used to build 3D structures for the newly-identified SUGARWIN3 and SUGARWIN4. Our data shows a mainly negative surface predicting a weak binding to RNA (**Figure 5**). Future research is necessary to functionally characterize these SUGARWINs, however, we suggest that the lack of a complete RNase active site and a less positively charged molecular surface may go hand-in-hand.



**FIGURE 4** | Alignment of SUGARWINs and BARWIN. The alignment (ClustalW) of the full sequences, including the signal peptide, highlights a histidine at alignment position 44 (His11) in BARWIN and SUGARWIN1. SUGARWINs 2, 3 and 4 show the replacement of histidine by asparagine at alignment position 44.

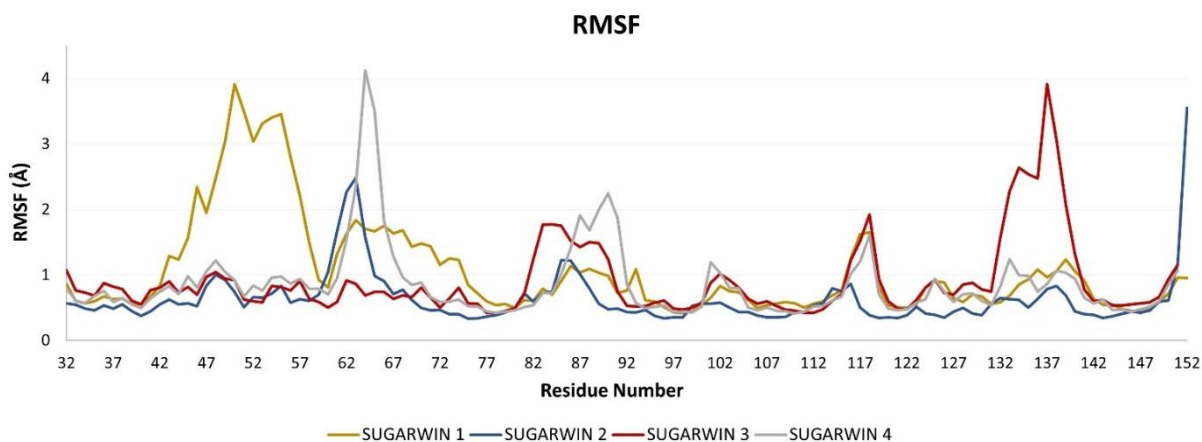


**FIGURE 5** | Electrostatic surface of the SUGARWINs. (A) SUGARWIN1; (B) SUGARWIN2; (C) SUGARWIN3; (D) SUGARWIN4. SUGARWIN1 is more –positive (blue) than the remaining SUGARWINs. The electrostatic surface analysis shows a mainly electronegative surface in SUGARWIN 2, 3 and 4 suggestive of poor binding to RNA, if any.

## Molecular Dynamics

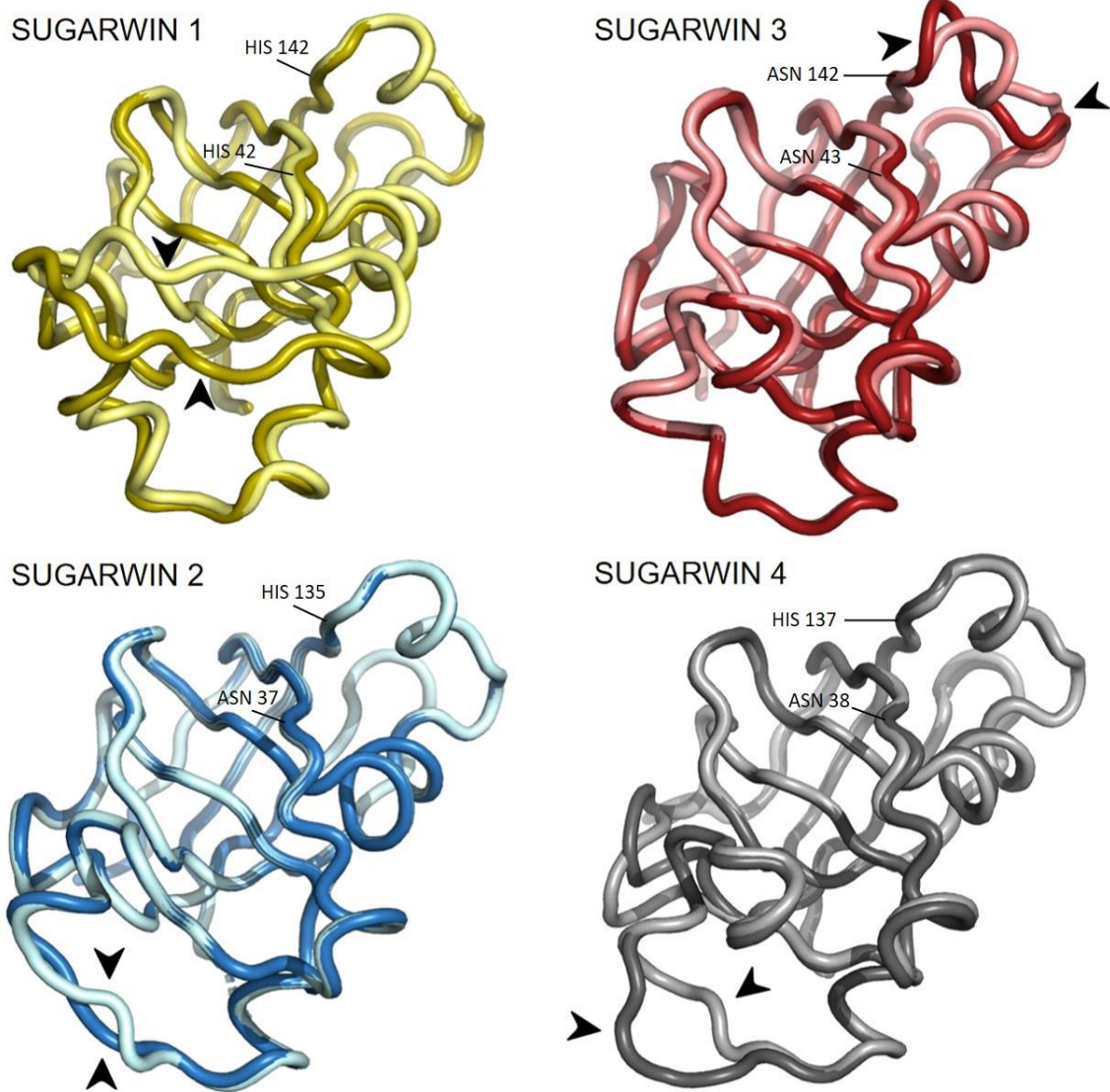
The observation that SUGARWIN2 has a two residues deletion in the variable loop, which is connected by hydrogen bonds to the active site His11 in active RNases, stimulated us to investigate the dynamics of the molecule as a whole. This was further motivated by the observation, that reorientation of the catalytic residues would be expected to occur during catalysis. In order to analyze the consequences of this deletion on structural mobility, a molecular dynamics approach was employed, starting from the crystal structure of SUGARWIN2. Molecular modeling of the remaining 3 SUGARWINs (SUGARWIN1, SUGARWIN3 and SUGARWIN4) was also performed.

The molecular dynamics of the crystal structure of SUGARWIN2 together with the derived models for SUGARWINs 1, 3 and 4 shows a very similar pattern of intrinsic movements in all cases. The most notable features of this analysis are the large peaks associated with the variable loop in the model for SUGARWIN1 (residues 42 - 61) which are absent from the remaining structures. The RMSF data from the molecular dynamics simulations (**Figure 6**) highlight the mobility of the variable loop, as shows a major peak for SUGARWIN1 along residues 44–58. The loop begins at residue H11 [42 - SUGARWIN1] which is believed to be involved in RNase activity and which, in the structure of CARWIN (H11), is coupled to an insertion within the loop itself by water mediated hydrogen bonds. SUGARWIN2 was the least flexible structure found in this analysis, with a minor peak between residues 60–65. SUGARWIN3 showed a major peak between residues 132 and 140, a region which precedes the second catalytic residue assumed to be essential for RNA hydrolysis (His113). SUGARWIN3 is the only structure to show this mobility and is also unique in having an asparagine at this position. Together with the behavior observed for SUGARWIN1 this suggests that structural mobility close to the catalytically implicated residues, H11 and H113, may be correlated with the type of amino acid which occupies these positions. SUGARWIN4 shows a single major peak involving residues 62 - 68, whereas other less prominent peaks are observed in all structures.



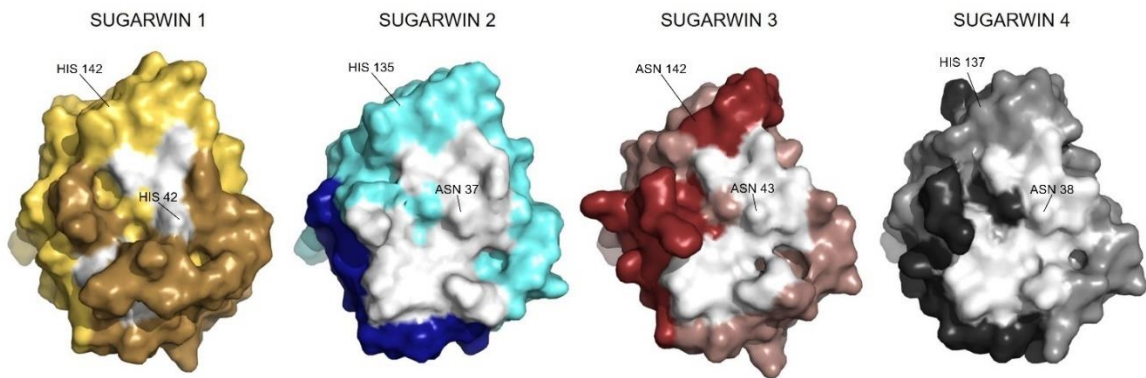
**FIGURE 6** | RMSF of the SUGARWINs. Root mean square fluctuation (RMSF) analysis for SUGARWINs 1, 2, 3 and 4 as derived from molecular dynamics simulations. Residues are numbered in reference to SUGARWIN1 complete sequence.

PCA analysis for each simulation was performed in order to identify putative sequence movements, which were then grouped and graphically represented. This data allowed the observation of the most flexible regions of each structure which correspond to the most representative group for each simulation (PC1), and are shown in **Figure 7**. Regarding the active site of the chitosanase function, only SUGARWIN1 showed significant movement in this region due to the flexibility of the variable loop described above (**Figure 8**). The mobility of this region may therefore be related to the ability for SUGARWIN1 to act as both a chitosanase and a ribonuclease, whereby different catalytic activities would be related to different loop conformations. We suggest that it is the dynamic nature of SUGARWIN1 (particularly the variable loop), which allows functional diversity implying that different structural conformations are a necessary requirement for different catalytic activities (**Figure 8**).



**FIGURE 7** | PCA analysis of the SUGARWINs. Structures of SUGARWINs 1, 2, 3 and 4 generated by Bio3D PCA analysis indicating the main movements of each protein. Brighter and darker colors indicate extreme points of movements, extracted from PCA generated frames. Black arrowheads highlight regions of major flexibility. Residues in equivalent positions to the RNase active site histidines are indicated.





**FIGURE 8** | Structures of SUGARWINs with emphasis on the chitinase binding site. Structures of SUGARWINs 1, 2, 3 and 4 in surface representation showing, in white, the regions previously identified as the binding site for chitinase activity in SUGARWIN2. Darker colors indicate regions of major flexibility according to the RMSF analysis. These superpose with the binding site regions in a few cases. Residues in equivalent positions to the RNase active site histidines are indicated.

## DISCUSSION

Plant response to biotic stresses such as herbivore insects and fungal pathogens involves a wide set of complex interactions, which cause profound physiological, chemical and morphological adaptations (Silva-Filho and Vivanco, 2017). Previous studies using transcriptomic data showed only two genes encoding for PR-4 proteins in sugarcane: SUGARWIN1 and SUGARWIN2 (Medeiros et al., 2012; Franco et al., 2014, 2019). However, in this work, we expand this observation by two additional protein sequences with distinct structural characteristics, named SUGARWIN3 and SUGARWIN4.

SUGARWIN1 exhibits RNase, chitinase and chitinase activity, whereas SUGARWIN2 has only chitinase activity (Franco et al., 2019). SUGARWIN3 and SUGARWIN4 have yet to be functionally characterized. Therefore, in order to shed light on the known differences among this protein family, we determined the three-dimensional structure of SUGARWIN2, which was crystallized and diffracted to 1.55 Å. Although SUGARWIN2 does not have RNase activity, its crystal structure shows the classic BARWIN fold (a six stranded  $\beta$ -sheet surrounded by four  $\alpha$ -helices and several loops). In addition, we were able to build homology models for the remaining SUGARWINs for structural comparisons.

Examination of the structures for the four variants revealed that SUGARWIN1 presented at least three unique properties. Firstly, it is the only SUGARWIN to possess

the catalytic residue, His11, associated with RNase activity; secondly, it is the most flexible of the four structures and finally, it has the most positive electrostatic surface potential. We argue that this unique combination of features renders SUGARWIN1 an active ribonuclease.

Several studies have correlated the RNase activity of BARWIN with the presence of two highly conserved histidine residues: at positions 11 and 113 (in relation to the BARWIN mature protein) [CARWIN (H11 and H110), WHEATWIN (H11 and H113) and SUGARWINs as referred above] (Caporale et al., 1999; 2004; Bertini et al., 2009; Huet et al., 2013; Franco et al., 2019). In addition, Bertini et al. (2009) also evaluated the relationship between the ribonuclease activity of SUGARWIN counterparts and their antifungal properties. It has been demonstrated that two different mutations to His11 (H11G and H11L) of WHEATWIN partially inhibited RNase activity, pointing out the importance of this residue for catalysis. Furthermore, these authors also shown that the presence of the catalytic residues (His11 and His113, in WHEATWIN1) is also fundamental for the antifungal activity, at least in WHEATWIN1, thereby correlating the two phenomena. Huet et al. (2013) also reported that the two homologous histidine residues are important for RNase activity in CARWIN.

Although SUGARWINs 1, 2 and 4 all have His146, only SUGARWIN1 possesses both histidine residues. Franco et al. (2019) demonstrated the loss of SUGARWIN1's RNase activity by mutating His11 to asparagine (H42N - SUGARWIN1). This strongly suggests that RNA hydrolysis by SUGARWIN1 occurs according to the classic acid-base mechanism involving both histidines, as seen in RNase A, RNase T1, BARWIN, CARWIN and WHEATWINs (Silva-Filho and Vivanco, 2017).

Although SUGARWIN1 has the catalytic machinery necessary for the hydrolysis of RNA, the SUGARWIN2 structure, similar to CARWIN's, shows that the His residues at positions 11 and 113 (11 and 111 - CARWIN; 37 and 135 - SUGARWIN2) point to opposite directions. In the case of SUGARWIN1 the two catalytic histidines would be expected to be oriented similarly, as observed in our homology-based model. However, this arrangement is incompatible with the acid-base mechanism, which suggests that SUGARWIN1 must have sufficient intrinsic flexibility in order to reposition the catalytic diad in such a way as to be competent for catalysis. This hypothesis is compatible with

the separation observed between the C $\alpha$  atoms of Asn11 and His113 in SUGARWIN2 (10.6 Å), which is similar to that observed in bovine RNase A (8.9 Å), suggesting a putative rearrangement of the protein structure. For this reason, we investigated the dynamics of all four SUGARWIN structures with particular interest in SUGARWIN1.

SUGARWIN1 displayed significant flexibility throughout the entire region of the variable loop (residues 44 - 56), but particularly in its N-terminal portion (residues 44 - 51). This loop starts at the catalytic His11, which, as mentioned above, is only present in SUGARWIN1. Furthermore, this loop includes a two-residue insertion in SUGARWIN1 and this insertion is coupled to His11 via water mediated hydrogen bonds (Figure 3). It seems reasonable to assume that the altered dynamic behavior of the variable loop in SUGARWIN1 is due to these notable structural differences. A direct association between loop flexibility with RNase function has already been suggested, whereby the loop may be important for the binding and correct positioning of RNA in the active site as well as for RNA release after catalysis (Doucet et al., 2009; Menezes et al., 2014; Li and Hammes-Schiffer, 2019).

In contrast, SUGARWINs 2 and 4 showed significant structural mobility only within the C-terminal portion of the variable loop (residues 60 - 68, **Figure 6**), whilst SUGARWIN3 showed a markedly different profile altogether with a peak toward the C-terminus (residues 132 - 141), which precedes the latter catalytic histidine. It is worth emphasizing that SUGARWIN3, which is unique in presenting increased mobility in this region, is the only SUGARWIN that has an asparagine residue in place of histidine at this position (H142N - SUGARWIN3). Overall, it would therefore appear that differences in chain mobility would be related to the nature of the amino acids occupying the catalytic positions 11 and 113 as well as to the variable loop insertion in the case of SUGARWIN1. These results are in agreement with previous data (Ludvigsen and Poulsen, 1992; Bertini et al., 2009, 2012; Huet et al., 2013; Franco et al., 2014, 2019).

All known PR-4 ribonucleases, such as BARWIN, CARWIN and WHEATWIN, possess the two-residue insertion in the variable loop (Doucet et al., 2009; Menezes et al., 2014; Li and Hammes-Schiffer, 2019). Movement of this loop may be critical to give His11 the necessary structural liberty to move from the position seen in our model for



SUGARWIN1 (and CARWIN) in order to form the RNase active site. This seems particularly evident on examining the CARWIN structure in which one of the two inserted residues of the loop holds the histidine sidechain to its mainchain carbonyl via a water-mediated hydrogen bond. The additional flexibility associated with the variable loop in SUGARWIN1 may be sufficient to decouple His11 from the loop insertion, thereby allowing it to adopt an alternative conformation—one compatible with RNase hydrolysis. However, we should point out that during our simulations we did not observe His11 pointing toward His113 as might be expected. This may suggest that the generation of a mature RNase active site may depend not only on the flexibility of the variable loop but also on association with the substrate itself, RNA, through an induced-fit like mechanism.

Evidence for the reorientation of His11 comes from the NMR structure of BARWIN, where it is partially rotated in the direction of His113, its partner in the RNase active site. This is facilitated by the variable loop adopting a different conformation in BARWIN, such that His11 is not coupled to the insertion within the variable loop. This is direct evidence that changes to the conformation of the variable loop could be related to liberating His11 in order to allow the formation of a mature RNase active site. Amongst the four different variants only SUGARWIN1 has the histidine at position 11 and the necessary structural plasticity.

For the ribonuclease activity of SUGARWIN1 to occur it is not sufficient to have the catalytic machinery and the expected flexibility allowing the formation of the active site with the histidines residues properly oriented. In addition to these two features, it is also necessary for the enzyme to attract and bind the highly negatively charged substrate, RNA. Examination of the electrostatic potential of the four SUGARWINs revealed that SUGARWIN1 is, by far, the most basic protein. This is probably due to a series of amino acid substitutions leading to an accumulation of lysines and arginines residues on the molecular surface (**Figure 3**) and a significant increase in the pI in comparison to the remaining SUGARWINs. This alteration to the electrostatic properties of the SUGARWIN1 surface would be expected to back the formation of the enzyme-substrate complex thereby permitting RNA hydrolysis to occur.

Different from the RNase activity, both SUGARWIN1 and SUGARWIN2 are known to be active chitosanases. However, the major movement to the variable loop in SUGARWIN1 leads to a significant occlusion of the known chitosan binding site (**Figure 8**). Therefore, the loop movement which we consider to be critical for the RNase active site, simultaneously covers the chitosan binding site. This suggests that SUGARWIN1 (and by analogy BARWIN, CARWIN and WHEATWIN) would be unable to simultaneously perform both catalytic functions. If the variable loop is positioned to expose the chitosan-binding site, then the RNase active site is not formed, as implied by the structure for SUGARWIN2. On the other hand, it seems likely that when the latter is correctly formed, then the chitosanbinding site will be occluded by the loop. Therefore, we propose that SUGARWIN1 might be considered as an enzyme with two different catalytic functions which are mutually exclusive.

No function has been described yet for the newly-described SUGARWINs (SUGARWIN3 and 4). It seems clear that no nuclease function might be associated with these proteins, since both lack the catalytic histidine at position 11. Additionally, SUGARWIN3 also has an asparagine residue replacing the histidine at position 113. Nevertheless, according to the structural analyses, it might be expected that both proteins display chitosanase activities. This is consistent with the fact that it is only in SUGARWIN1 that we observe the variable loop crosses over the chitosan binding site. For the other isoforms, the binding site for the chitosan substrate remains exposed during the simulations. We have previously shown that for SUGARWIN1 both the RNase and chitinase activities do not impair its chitosanase function. As mentioned above this is probably associated with the structural flexibility in the region of the variable loop which is most notable in SUGARWIN1, the only variant with a known nuclease function (SUGARWIN1).

Our phylogenetic analysis showed that both histidines related to RNase activity and a variable loop of the same size to that found in the variable region of SUGARWIN1 are present in the majority of the Embryophyta PR-4 protein sequences and are widespread in the phylogenetic tree. In contrast, proteins with an asparagine residue at position 11 were found only in a few genomes and do not form a separate clade. These results show that both histidine and the flexible loop observed in SUGARWIN1

may have been present in the PR-4 protein from the Embryophyta ancestral species and that independent replacements by asparagine may have happened in the evolutionary history of this group. One of these replacements may have occurred in the common ancestor of Poales originating the sequence that gave rise to SUGARWIN2, SUGARWIN3 and SUGARWIN4 (**Figure 1**).

## CONCLUSION

Thus far only two PR-4 proteins have had their tridimensional structures solved: BARWIN from Barley [PDB ID: 1BW3 (Ludvigsen and Poulsen, 1992)] and CARWIN from papaya [PDB ID: 4JP6 (Huet et al., 2013)], both of which contain H11 and H113 (H11 and H110 in CARWIN) and with chitinase and RNase activity. Here, we described the first tridimensional structure of a PR-4 protein with no RNase activity and with the H44N replacement: SUGARWIN2. Comparison of the structures, dynamics and surface properties of the two proteins allows us to propose that (1) the more positive surface potential of SUGARWIN1 aids in binding the substrate RNA, (2) flexibility of the variable loop is essential for the formation of the RNase binding site which otherwise would have the catalytic histidines pointing to opposite directions as observed in the crystal structures, and (3) the flexibility of the loop also occludes the chitosan binding site suggesting that the two catalytic activities are mutually exclusive. Unexpectedly, despite its lack of RNase activity, SUGARWIN2 seems to be the more relevant for sugarcane defense against pathogenic fungi, since its expression is strongly upregulated after 10 days of *C. falcatum* treatment, whereas SUGARWIN1 showed a moderate increase in gene expression under the same conditions. This may imply that a more specialized enzyme is required for this purpose. We also describe here two new SUGARWIN proteins, SUGARWIN3 and SUGARWIN4, which are predicted, based on the above analyses, to lack RNase activity. Further studies are being conducted to better understand these new proteins. The exact details of the catalytic capabilities of each of these enzymes appears to be a complex function of several factors including the residues which comprise the active site, the intrinsic flexibility of the structures themselves and the surface electrostatics necessary for substrate recognition.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

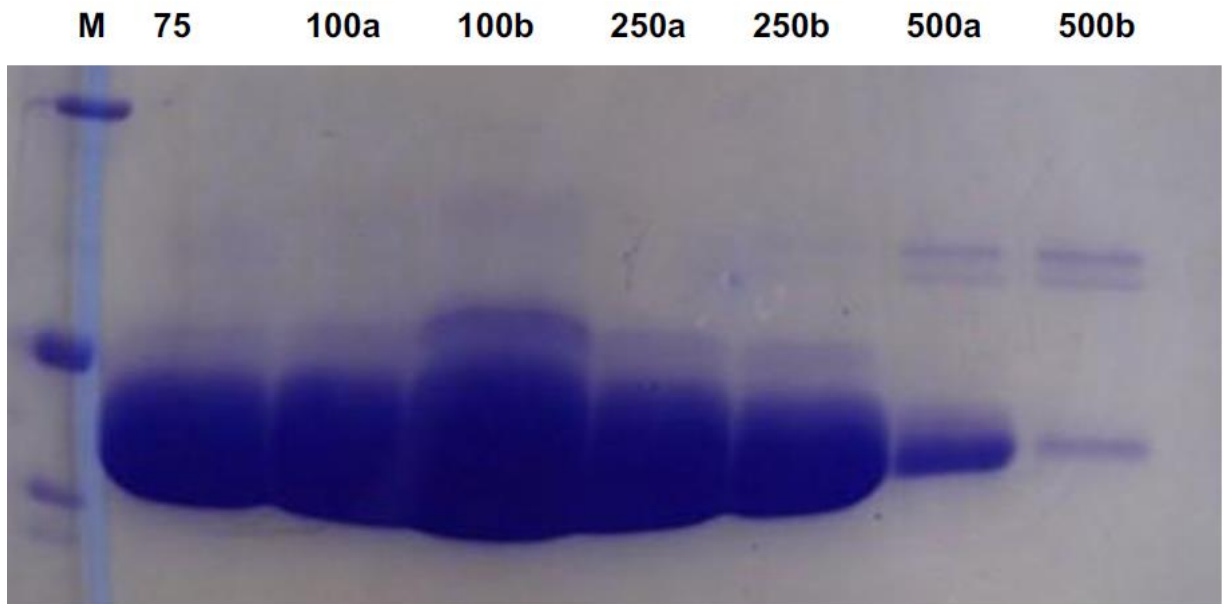
MS-F: design of the experiments, lab infrastructure, providing research funds, discussions, and writing. All authors contributed to the article and approved the submitted version.

## FUNDING

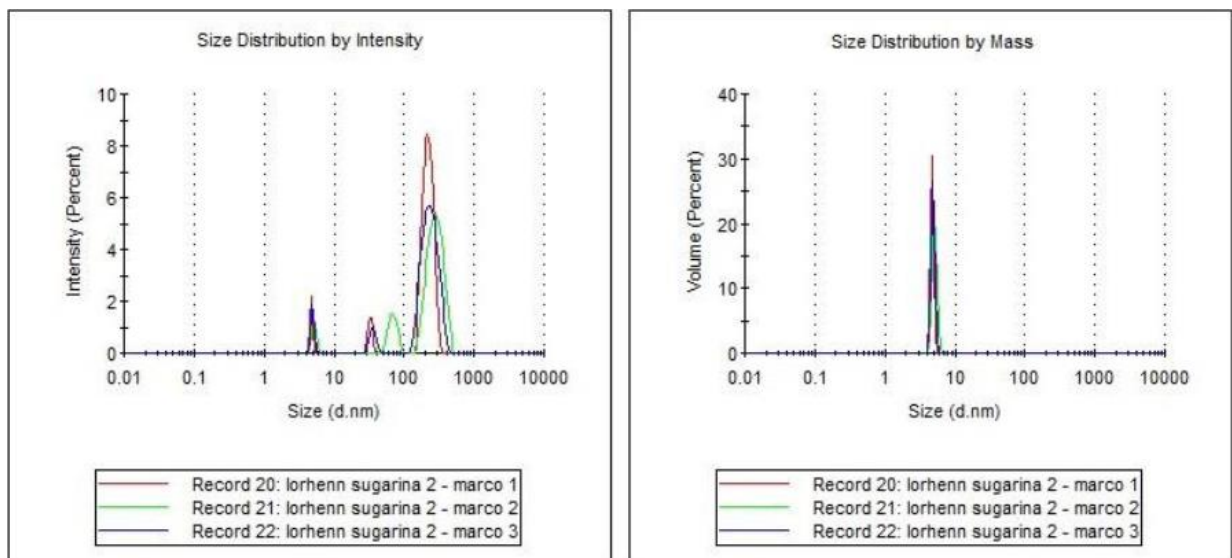
This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant 2014/50275-9, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant 482737/2012-3 to MS-F and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Finance Code 001. LM is the recipient of CAPESPROSUC fellowship. MS-F, RG, and FH-S are also research fellows of CNPq.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.734248/full#supplementary-material>.



**SUPPLEMENTARY FIGURE 1 | 12% SDS-PAGE showing the purification of SUGARWIN2 by IMAC.** Lane M Molecular Weight Marker; Lane 1-7 aliquots of fractions eluted with crescent concentrations of imidazole (75 to 500 mM). The arrow indicates the band representing purified SUGARWIN2.



**SUPPLEMENTARY FIGURE 2 | DLS of SUGARWIN2.** The DLS of SUGARWIN2 showed a monodisperse peak that corresponded to 99.7% of the mass consistent with the mass of the protein.

**SUPPLEMENTARY TABLE 1.** Distribution of sequences among the clades of the phylogenetic tree

	Clades			
	C1	C2	C3	C4
<b>Sequences with H or N residues in position 11</b>				
	Percentage (%)			
<b>H11</b>	15	97	6	100
<b>N11</b>	75	3	94	0
<b>Number of genes</b>				

Subclass	Clade	Order	Clade	Species	C1	C2	C3	C4
Petrosaviidae		Asparag.		<i>Asparagus officinalis</i> (Aoff)	1	0	0	0
				<i>Dendrobium catenatum</i> (Dcat)	0	1	0	0
				<i>Phalaenopsis equestris</i> (Pequ)	0	1	0	0
		Arec.		<i>Elaeis guineensis</i> (Egui)	2	0	0	0
				<i>Phoenix dactylifera</i> (Pdac)	2	0	0	0
				<i>Ensete ventricosum</i> (Even)	1	0	0	0
	Commelinids	Zingib.		<i>Musa acuminata</i> subsp. <i>malaccensis</i> (Macu)	3	0	0	0
				<i>Musa balbisiana</i> (Mbal)	2	0	0	0
				<i>Ananas comosus</i> (Acom)	2	0	0	0
			Poales	<i>Carex littledalei</i> (Clit)	1	0	0	0

BOP	<i>Aegilops tauschii</i> subsp. <i>Tauschii</i> (Atau)	3	3	1	0	
	<i>Triticum dicoccoides</i> (Tdic)	6	6	2	0	
	<i>Triticum turgidum</i> subsp. <i>durum</i> (Ttur)	1	6	0	0	
	<i>Brachypodium distachyon</i> (Adis)	1	1	0	0	
	<i>Oryza brachyantha</i> (Obra)	0	2	0	1	
	<i>Oryza meyeriana</i> var. <i>granulata</i> (Omey)	0	2	0	1	
	<i>Oryza sativa Japonica</i> Group (Osat)	0	3	0	1	
	PACMA D	<i>Eragrostis curvula</i> (Ecur)	3	1	2	6
		<i>Dichanthelium oligosanthes</i> (Doli)	2	0	0	2
		<i>Panicum hallii</i> (Phal)	2	0	1	2
		<i>Setaria italica</i> (Sita)	3	0	1	2
		<i>Setaria viridis</i> (Svir)	3	0	1	2
		<i>Sorghum bicolor</i> (Sbic)	1	0	1	2
		<i>Zea mays</i> (Zmay)	1	0	2	2
<i>Miscanthus lutarioriparius</i> (Mlut)		3	0	4	9	
<i>Saccharum</i> spp. (SP80-3280)		8	0	14*	18*	

				<i>Saccharum spontaneum</i> (Sspo)	1	0	2	5
		Al.		<i>Zostera marina</i> (Zmar)	0	0	2	0

\*Includes SUGARWIN2

\*\*includes SUGARWIN1.

\*\*\*Al., Alismatales; Zigib., Zingiberales; Arec., Arecales; and Asparag., Asparagales.

## REFERENCES

- Adams, P. D., Afonine, P. V., Bunkóczy, G., Chen, V. B., Echols, N., Headd, J. J., et al. (2011). The Phenix software for automated determination of macromolecular structures. *Methods* 55, 94–106. doi: 10.1016/j.ymeth.2011.07.005
- Agrawal, G. K., Jwa, N. S., Han, K. S., Agrawal, V. P., and Rakwal, R. (2003). Isolation of a novel rice PR-4 type gene whose mRNA expression is modulated by blast pathogen attack and signaling components. *Plant Physiol. Biochem.* 41, 81–90. doi: 10.1016/s0981-9428(02)00012-8
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Bai, S., Dong, C., Li, B., and Dai, H. (2013). A PR-4 gene identified from *Malus domestica* is involved in the defense responses against *Botryosphaeria dothidea*. *Plant Physiol. Biochem.* 62, 23–32. doi: 10.1016/j.plaphy.2012.10.016
- Banno, S., Ochiai, N., Noguchi, R., Kimura, M., Yamaguchi, I., Kanzaki, S., et al. (2005). A catalytic subunit of cyclic AMP-dependent protein kinase, PKAC1, regulates asexual differentiation in *Neurospora crassa*. *Genes Gen. Syst.* 80, 25–34. doi: 10.1266/ggs.80.25
- Bertini, L., Caporale, C., Testa, M., Proietti, S., and Caruso, C. (2009). Structural basis of the antifungal activity of wheat PR-4 proteins. *FEBS Lett.* 583, 2865–2871. doi: 10.1016/j.febslet.2009.07.045
- Bertini, L., Proietti, S., Aleandri, M. P., Mondello, F., Sandini, S., Caporale, C., et al. (2012). Modular structure of HEL protein from *Arabidopsis* reveals new potential functions for PR-4 proteins. *Biol. Chem.* 393, 1533–1546. doi: 10.1515/hsz-2012-0225
- Bravo, J. M., Campo, S., Murillo, I., Coca, M., and Segundo, B. S. (2003). Fungus and wound-induced accumulation of mRNA containing a class II chitinase of the pathogenesis-related protein 4 (PR-4) family of maize. *Plant Mol. Biol.* 52, 745–759.
- Broekaert, I., Lee, H. I., Kush, A., Chua, N. H., and Raikhel, N. (1990). Wound-induced accumulation of mRNA containing a hevein sequence in laticifers of rubber tree (*Hevea brasiliensis*). *Proc. Natl. Acad. Sci. U.S.A.* 87, 7633–7637. doi: 10.1073/pnas.87.19.7633
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Caporale, C., Di Berardino, I., Leonardi, L., Bertini, L., Cascone, A., Buonocore, V., et al. (2004). Wheat pathogenesis-related proteins of class 4 have ribonuclease activity. *FEBS Lett.* 575, 71–76. doi: 10.1016/j.febslet.2004.07.091
- Caruso, C., Bertini, L., Tucci, M., Caporale, C., Leonardi, L., Saccardo, F., et al. (1999). Isolation and characterisation of wheat cDNA clones encoding PR-4 proteins. *DNA Seq.* 10, 301–307. doi: 10.3109/10425179909033956
- Caruso, C., Nobile, M., Leonardi, L., Bertini, L., Buonocore, V., and Caporale, C. (2001). Isolation and amino acid sequence of two new PR-4 proteins from wheat. *J. Protein Chem.* 20, 327–335.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 1), 12–21. doi: 10.1107/s0907444909042073
- Christopher, M. E., Miranda, M., Major, I. T., and Constabel, C. P. (2004). Gene expression profiling of systemically wound-induced defenses in hybrid poplar. *Planta* 219, 936–947. doi: 10.1007/s00425-004-1297-3
- Doucet, N., Watt, E. D., and Loria, J. P. (2009). The flexibility of a distant loop modulates active site motion and product release in ribonuclease A. *Biochemistry* 48, 7160–7168. doi: 10.1021/bi900830g
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2126–2132. doi: 10.1107/s0907444904019158
- Falco, M. C., Marbach, P. A. S., Pompermayer, P., Lopes, F. C. C., and Silva-Filho, M. C. (2001). Mechanisms of sugarcane response to herbivory. *Genet. Mol. Biol.* 24, 113–122. doi: 10.1590/s1415-4757200100010 0016
- Franco, F. P., Dias, R. O., Toyama, D., Henrique-Silva, F., Moura, D. S., and Silva-Filho, M. C. (2019). Structural and functional characterization of PR-4 SUGARWINs from sugarcane and their role in plant defense. *Front. Plant Sci.* 9:1916. doi: 10.3389/fpls.2018.01916
- Franco, F. P., Moura, D. S., Vivanco, J. M., and Silva-Filho, M. C. (2017). Plant-insect-pathogen interactions: a naturally complex ménage à trois. *Curr. Opin. Microbiol.* 37, 54–60. doi: 10.1016/j.mib.2017.04.007
- Franco, F. P., Santiago, A. C., Henrique-Silva, F., de Castro, P. A., Goldman, G. H., Moura, D. S., et al. (2014). The sugarcane defense protein SUGARWIN2 causes cell death in *Colletotrichum falcatum* but not in non-pathogenic fungi. The sugarcane defense protein SUGARWIN2 causes cell death in *Colletotrichum falcatum* but not in non-pathogenic fungi. *PLoS One* 9:e91159. doi: 10.1371/journal.pone.0091159
- Franco, F. P., Túler, A. C., Gallan, D. Z., Gonçalves, F. G., Favaris, A. P., Peñaflo, M. F. G. V., et al. (2021). Fungal phytopathogen modulates plant and insect responses to promote its dissemination. *ISME J.* 15:10. doi: 10.1038/s41396-021-01010-z
- Friedrich, L., Moyer, M., Ward, E., and Ryals, J. (1991). Pathogenesis-related protein-4 is structurally homologous to the carboxy-terminal domains of hevein, win-1 and win-2. *Mol. Gen. Genet.* 230, 113–119. doi: 10.1007/bf00290658



- Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006). Bio3D: an R package for the comparative analysis of protein structures. *Bioinformatics* 22, 2695–2696. doi: 10.1093/bioinformatics/btl461
- Guevara-Morato, M., de Lacoba, M. G., Garcia-Luque, I., and Serra, M. T. (2010). Characterization of a pathogenesis-related protein 4 (PR-4) induced in *Capsicum chinense* L-3 plants with dual RNase and DNase activities. *J. Exp. Bot.* 61, 3259–3271. doi: 10.1093/jxb/erq148
- Hejgaard, J., Jacobsen, S., Bjorn, S. E., and Kragh, K. M. (1992). Antifungal activity of chitin-binding PR-4 type proteins from barley-grain and stressed leaf. *FEBS Lett.* 307, 389–392. doi: 10.1016/0014-5793(92)80720-2
- Huet, J., Mbooso, E. J. T., Soror, S., Meyer, F., Looze, Y., Wintjens, R., et al. (2013). High resolution structure of a papaya plant-defence barwin-like protein solved by in-house sulfur-SAD phasing. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 69, 2017–2026. doi: 10.1107/s0907444913018015
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 3, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kiba, A., Saitoh, H., Nishihara, M., Omiya, K., and Yamamura, S. (2003). C-terminal domain of a hevein-like protein from *Wasabia japonica* has potent antimicrobial activity. *Plant Cell Physiol.* 44, 296–303. doi: 10.1093/pcp/pcg035
- Kim, N. H., and Hwang, B. K. (2015). Pepper pathogenesis-related protein 4c is a plasma membrane-localized cysteine protease inhibitor that is required for plant cell death and defense signaling. *Plant J.* 81, 81–94. doi: 10.1111/tpj.12709
- Li, P., and Hammes-Schiffer, S. (2019). Substrate-to-product conversion facilitates active site loop opening in yeast enolase: a molecular dynamics study 9. *ACS Catal.* 9, 8985–8990. doi: 10.1021/acscatal.9b03249
- Li, X. D., Xia, B., Jiang, Y. M., Wu, Q. S., Wang, C. Y., He, L. S., et al. (2010). A new pathogenesis-related protein, LrPR-4, from *Lycoris radiata*, and its antifungal activity against *Magnaporthe grisea*. *Mol. Biol. Rep.* 3, 995–1001. doi: 10.1007/s11033-009-9783-0
- Linthorst, H. J. M., Danhash, N., Brederode, F. T., Vankan, J. A. L., Dewit, P., and Bol, J. F. (1991). Tobacco and tomato pr proteins homologous to win and pro-hevein lack the hevein domain. *Mol. Plant Microbe Interact.* 4, 586–592. doi: 10.1094/mpmi-4-586
- Ludvigsen, S., and Poulsen, F. M. (1992). Positive  $\phi$ -angles in proteins by nuclear magnetic resonance spectroscopy. *J. Biomol. NMR* 2, 227–233. doi: 10.1007/bf01875318
- Marti-Renom, M. A., Stuart, A., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325. doi: 10.1146/annurev.biophys.29.1.291
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.
- Medeiros, A. H., Franco, F. P., Matos, J. L., de Castro, P. A., Santos-Silva, L. K., Henrique-Silva, F., et al. (2012). Sugarwin: a sugarcane insect-induced gene with antipathogenic activity. *Mol. Plant Microbe Interact.* 25, 613–624. doi: 10.1094/mpmi-09-11-0254
- Menezes, S. P., de Andrade Silva, E. M., Matos Lima, E., Oliveira de Sousa, A., Andrade, B. S., Lemos, L. S. L., et al. (2014). The pathogenesis-related protein PR-4b from *Theobroma cacao* presents RNase activity, Ca<sup>2+</sup> and Mg<sup>2+</sup> dependent DNase activity and antifungal action on *Moniliophthora perniciosa*. *BMC Plant Biol.* 14:161. doi: 10.1186/1471-2229-14-161
- Neuhaus, J. M., Fritig, B., Linthorst, H. J. M., Meins, F., Mikkelsen, J. D., and Ryals, J. (1996). A revised nomenclature for chitinase genes. *Plant Mol. Biol. Rep.* 14, 102–104. doi: 10.1007/bf02684897
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. I. (2015). A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Olsson, M. H., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). Propka3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* 7, 525–537. doi: 10.1021/ct100578z
- Pieterse, C. M. J., and van Loon, L. C. (1999). Salicylic acid-independent plant defence pathways. *Trends Plant Sci.* 4, 52–58. doi: 10.1016/s1360-1385(98)01364-8
- Potter, S., Uknes, S., Lawton, K., Winter, A. M., Chandler, D., Dimaio, J., et al. (1993). Regulation of a hevein-like gene in *Arabidopsis*. *Mol. Plant Microbe Interact.* 6, 680–685. doi: 10.1094/mpmi-6-680
- Ryan, C. A. (1990). Protease Inhibitors in plants—genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* 28, 425–449. doi: 10.1146/annurev.py.28.090190.002233
- Schillmiller, A. L., and Howe, G. A. (2005). Systemic signaling in the wound response. *Curr. Opin. Plant Biol.* 8, 369–377. doi: 10.1016/j.pbi.2005.05.008
- Silva-Filho, M. C., and Vivanco, J. M. (2017). Guest editorial: plants and their surrounding microorganisms: a dynamic world of interactions. *Curr. Opin. Microbiol.* 37, 5–6.
- Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., et al. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *Gigascience* 8:giz129. doi: 10.1093/gigascience/giz129
- Souza, T. P., Dias, R. O., and Silva-Filho, M. C. (2017). Defense-related proteins involved in sugarcane responses to biotic stress. *Genet. Mol. Biol.* 40, 360–372. doi: 10.1590/1678-4685-gmb-2016-0057
- Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* 41, 641–643. doi: 10.1107/s0021889808006985
- Svensson, B., Svendsen, I., Hojrup, P., Roepstorff, P., Ludvigsen, S., and Poulsen, F. M. (1992). Primary structure of BARWIN - a barley seed protein closely related to the c-terminal domain of proteins encoded by wound-induced plant genes. *Biochemistry* 31, 8767–8770. doi: 10.1021/bi00152a012
- Van Loon, L. C., and Van Strien, E. A. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* 55, 85–97. doi: 10.1006/pmpp.1999.0213
- Xu, Y., Chang, P. F. L., Liu, D., Narasimhan, M. L., Raghothama, K. G., Hasegawa, P. M., et al. (1994). Plant defense genes are synergistically induced by ethylene and methyl jasmonate. *Plant Cell* 6, 1077–1085. doi: 10.2307/3869886
- Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573.

Zhu, T., Song, F., and Zheng, Z. (2006). Molecular characterization of the rice pathogenesis-related protein, OsPR-4b, and its antifungal activity against *Rhizoctonia solani*. *J. Phytopathol.* 154, 378–384. doi: 10.1111/j.1439-0434.2006.01110.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright ©2021 Maia, Pereira, Garratt, Brandão-Neto, Henrique-Silva, Toyama, Dias, Bachega, Peixoto and Silva-Filho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms*

## CAPÍTULO II

### Padronização metodológica

Os resultados mais relevantes das análises realizadas, neste trabalho, com as proteínas SUGARWINs foram apresentados no artigo publicado. No entanto, aqui enfatizaremos separadamente os resultados obtidos em cada etapa metodológica realizada, bem como os desafios enfrentados na bancada, para a análise estrutural e evolutiva das proteínas PR-4 SUGARWINs. Até a obtenção dos resultados destas robustas análises, algumas etapas metodológicas precisaram ser padronizadas, visando otimizar as condições experimentais que nos permitiram gerar os dados que foram analisados.

#### 1. Produção

Para execução deste trabalho, ambas as proteínas SUGARWIN1 e SUGARWIN2 foram produzidas por expressão heteróloga em *Pichia pastoris*. Seus respectivos genes foram clonados e recombinados usando vetores específicos (pPICZ e pPICZ $\alpha$ ), em trabalhos anteriores do grupo de pesquisa (MEDEIROS et al., 2012). Para atender a alta demanda dos produtos proteicos requerida nos ensaios de biologia estrutural realizados neste trabalho, utilizou-se desses sistemas de expressão heteróloga já existentes, confeccionados por meio da tecnologia do DNA recombinante. Os estoques das colônias de *Pichia* recombinante, com SUGARWIN1 e SUGARWIN2, foram acessados e induzidos para a produção em larga escala das proteínas de interesse, a fim de que os ensaios para as análises da biologia estrutural fossem realizados. Na sequência, purificou-se a cultura por cromatografia de afinidade para obtenção da proteína em solução e dializou-se para a troca do tampão da proteína. A qualidade do produto obtido foi averiguada por meio de corrida eletroforética em gel de SDS-PAGE 12% (Figura 1 – APÊNDICE). Esta etapa de expressão das proteínas recombinantes foi realizada no Laboratório de Biologia Molecular do Departamento de Genética e Evolução da

Universidade Federal de São Carlos (UFSCar), em São Carlos/SP, sob a supervisão do Prof. Flávio Henrique-Silva. A etapa de produção de ambas as proteínas foi realizada com sucesso. Cabe enfatizar que, na produção, o rendimento da expressão de SUGARWIN1 recombinante (1.5 - 3.0 mg/mL) em sistema heterólogo foi inferior à expressão da proteína recombinante SUGARWIN2 (7.0 - 8.0mg/mL).

Na etapa de produção das proteínas preparou-se 2L de cultura da *Pichia* recombinante para indução da superexpressão visando atingir uma concentração final de 5 – 15mg/mL de ambas as proteínas puras, para um volume final de 300µL, que seriam utilizados nos ensaios de cristalização. Cabe ressaltar que, a produção de SUGARWIN2 *in vitro* foi muito superior a SUGARWIN1, tal característica também é vista *in vivo*, uma vez que com estímulo a produção de SUGARWIN2 aumenta cerca de 600x.

## **2. Purificação**

Uma segunda purificação foi realizada afim de garantir a pureza da proteína produzida. Os resultados obtidos na purificação em coluna de filtração em gel (resina Superdex – para exclusão molecular) realizada pelo AKTA, pode ser vista na Figura 2 - APÊNDICE. A qualidade da proteína purificada foi averiguada quanto à agregamento e pureza da amostra, por DLS (Figura 3 - APÊNDICE) e a quantidade mensurada por meio do espectrofotômetro nanodrop®. Após a segunda purificação, ambas as proteínas foram estocadas em solução final 20mM TrisHCl/ 200mM NaCl - pH 7.4 (tampão adequado para a cristalização).

Os resultados apresentados na purificação de ambas as proteínas apontaram para um estado oligomérico de monômero. Ambas as proteínas recombinantes apresentaram bons parâmetros de pureza nas soluções finais, conduzidas para os ensaios de cristalização.

## **3. Cristalização**

Quanto às características moleculares e físico-químicas, a proteína SUGARWIN1 (*Genebank*: CA145787) é um pequeno polipeptídeo com 121 aminoácidos, sendo 4 triptofanos. Sua massa molecular é de 13,4kDa e o pI teórico de 8.78. Já a proteína SUGARWIN2 (*Genebank*: CA138924) também é um pequeno polipeptídeo com 122 aminoácidos, sendo também 4 triptofanos. A sua massa molecular corresponde à 13kDa

e o  $pI$  teórico é de 4.54. Ambas as proteínas são pequena e apresentaram bons parâmetros de pureza na purificação.

A SUGARWIN2 formou cristais em diferentes condições de cristalização. Numa concentração de 7mg/mL, SUGARWIN2 foi cristalizada pelo método *sitting drop vapor diffusion* usando o kit de *screening Morpheus* (Molecular Dimensions). Os cristais mais adequados foram separados e encaminhados para a coleta dos dados cristalográficos (Figura 4 – APÊNDICE). Já a SUGARWIN1 ficou instável nas soluções finais estabelecidas. Testamos diferentes condições de soluções e por vezes, houve precipitação da proteína. Diante disso, realizamos o termofluor (DSF) para encontrar um tampão que fosse ideal para ela. Com os resultados do termofluor, preparamos a solução 20mM TrisHCl/250mM NaCl /Glicerol 2.5 - pH 8.0 (tampão favorável para SUGARWIN1). Um *screening* com vários testes de diferentes kits comerciais de cristalização foram empregados na única solução em que não houve precipitação de SUGARWIN1, mas apenas formações instáveis de minúsculos cristais foram obtidas. Num segundo momento, “abriu-se na mão” 24 condições em placa de cristalização, variando quantidades de sulfato de lítio (1.5 a 0.75M) e pH (7.75 a 9.0), num intervalo próximo à condição em que houve formações instáveis de cristais, no *screening* inicial. O desenho experimental de condições exclusivamente especificadas para SUGARWIN1, foi o recurso final utilizado por nós buscando otimizar condições em que a proteína fosse favorável à formação de cristais.

Diante da expressiva diferença de comportamento molecular nos experimentos de bancada da cristalização, realizamos uma análise detalhada da estrutura primária de ambas, buscando explicações que justificasse o comportamento instável da SUGARWIN1 nas diferentes soluções de cristalização testadas. Quanto às características físico-químicas, cabe enfatizar que a SUGARWIN1 apresentou um alto  $pI$  teórico (8.78), se mostrando bem mais positiva que a SUGARWIN2 (4.54). O fato da SUGARWIN1 ser muito básica se deve a predominância de resíduos básicos em sua cadeia primária. Quando realizamos o alinhamento de sequência no BLAST, notamos que, embora houvesse um alto percentual de identidade entre elas (73,73%), algumas trocas de aminoácidos se mostravam relevantes para a grande diferença nos valores de  $pI$ . A Figura 5 (APÊNDICE), mostra o alinhamento de ambas as proteínas, com os

resíduos básicos da SUGARWIN1 destacados. Essas poucas modificações na composição de resíduos parecem ter conferido significativas diferenças físico-química que poderiam estar impactando nas dificuldades que enfrentamos para cristalizar a SUGARWIN1. Aqui, levantamos que essas diferenças poderiam estar associadas também à mobilidade das estruturas, dificultando a formação de cristais de SUGARWIN1. Uma possível flexibilidade na estrutura da SUGARWIN1 poderia ser responsável pelo comportamento instável na solução de cristalização. E ao mesmo tempo, seria essa mesma característica a lhe conferir as funções quitinase e RNase?

Não obtivemos bons cristais de SUGARWIN1 e por isso realizamos a modelagem molecular desta proteína, a partir da estrutura cristalografada de sua homóloga, SUGARWIN2. Os ensaios de cristalização foram realizados no Instituto de Física de São Carlos da Universidade de São Paulo (IFSC/USP), em São Carlos.

#### **4. Coleta dos dados cristalográficos**

A fonte de raios-X utilizada para a coleta dos dados cristalográficos da proteína SUGARWIN2 foi o síncrotron britânico *Diamond Light Source*, que fica no *campus Science and Innovation*, Harwell, no Reino Unido.

#### **5. Processamento dos dados, determinação da estrutura e refinamento**

O processamento dos dados, bem como a determinação da estrutura e refinamento da proteína, que são etapas extremamente técnicas, foi realizado juntamente com cristalógrafos experientes do IFSC/USP e *Diamond Light Source*.

A Figura 6 (APÊNDICE) ilustra o modelo refinado da célula unitária do cristal de SUGARWIN2. Cabe enfatizar que há duas moléculas da proteína na cela. No entanto, a célula unitária (cela) em nada tem relação com o estado oligomérico, pois ela diz respeito apenas ao empacotamento da proteína no cristal. É um artefato da cristalização. A probabilidade de que as moléculas formem dímeros para desempenharem funções biológicas juntas é baixa, pois os resultados da análise do DLS e da purificação mostram que ela é um monômero.

As coordenadas e fatores de estrutura do modelo final (Figura 7 - APÊNDICE) foram depositados no *Protein Data Bank* (PDB ID: 7KSN).

#### **6. Análise da filogenia molecular**

Uma primeira análise filogenética das SUGARWINs foi realizada, buscando compreender pela história evolutiva, as diferenças que haviam entre elas. Usando o algoritmo BLASTp, pesquisamos sequências de proteínas homólogas as SUGARWINs em 26 genomas de Liliopsida, (NCBI), nos genomas de *Saccharum spontaneum* (Zhang et al., 2018) e o genoma do híbrido brasileiro de *Saccharum spp.* SP80-3280 (Souza et al., 2019). E selecionamos as sequências selecionadas com base na presença dos seis resíduos de cisteína envolvidos nas ligações dissulfeto estruturais da proteína e um peptídeo sinal.

A Figura 8 (APÊNDICE) nos mostrou os primeiros resultados obtidos pela inferência de árvore filogenética de máxima verossimilhança. Cabe ressaltar que em cana-de-açúcar há um complexo de vários conjuntos de genes com 20 genes de SUGARWIN1 e outros 20 genes de SUGARWIN2. Quando analisamos genomas, apareceu 4 cladogramas diferentes de PR-4 dentro de monocotiledôneas, enquanto a análise transcriptômica apontava apenas para existência de duas isoformas (SUGARWIN1 e SUGARWIN2). É relevante apontar que ambas não faziam parte do mesmo clado, com a SUGARWIN2 no clado C3 e a SUGARWIN1 no clado C4. Com estes resultados, dois outros tipos de SUGARWINs foram identificados, os quais chamamos de SUGARWIN3 e SUGARWIN4. Nesta primeira análise, que nos apontou para a existência de outras duas novas SUGARWINs, a presença da SUGARWIN1 no clado C4 sugeriu que esta talvez fosse a proteína mais recente. No entanto, análises de filogenia molecular posteriores foram realizadas fortalecendo a evidência de mais duas isoformas de SUGARWINs, porém refutando que SUGARWIN1 tenha sido a isoforma mais recente. Pelo contrário, a história evolutiva dessas proteínas sugere que, a perda da função RNase, característica da SUGARWIN1, tenha sido a característica mais recente (MAIA et al. 2021).

A análise da filogenia molecular das proteínas SUGARWINs foi realizada em parceria com o Instituto de Ciências Biológicas da Universidade Federal de Goiás, de Goiânia/GO, sob a supervisão da Profa. Renata Oliveira Dias.

## **7. Análise estrutural comparativa**

Como não conseguimos obter cristais da SUGARWIN1, realizamos uma modelagem molecular por homologia, usando a estrutura obtida por cristalografia da SUGARWIN2 para construir um modelo de estrutura 3D da SUGARWIN1.

Embora muito tenha se avançado a respeito das SUGARWINs até aqui, algumas perguntas permaneciam sem resposta. Quais os mecanismos de catálise da função ribonuclease/quitinase/quitosanase da SUGARWIN1? Seria o mesmo sítio ativo para ambas as funções? Com isso, primariamente, alguns pontos sobre a análise estrutural das SUGARWINs1 e 2 foram levantados. 1) A partir dos resultados estruturais obtidos, fica claro que o estado oligomérico das SUGARWINs é monomérico. 2) Embora o mecanismo de catálise das ribonucleases envolvesse apenas uma His catalítica, tanto a SUGARWIN1 quanto outras WINs com atividade RNase, apresentavam duas His conservadas. 3) A análise estrutural apontava que as diferentes funções estavam desvinculadas, portanto deveria haver diferentes motivos associados às respectivas funções. 4) A presença de resíduos básicos, que aumentam o pI da SUGARWIN1, provavelmente está relacionada com a função ribonuclease. Um alto pI pode favorecer a ligação com ácidos nucleicos. 5) Ambas as proteínas possuem um *loop* variável em sua estrutura 3D. No entanto, o *loop* da SUGARWIN1 parece apresentar maior mobilidade que o da SUGARWIN2. Esta flexibilidade na SUGARWIN1 pode favorecer o sinergismo.

Partindo desses pontos da análise estrutural comparativa entre SUGARWIN1 e 2, desenvolvemos modelos de estrutura para as proteínas SUGARWINs recém-descobertas, por modelagem molecular, também a partir da estrutura obtida por cristalografia da SUGARWIN2. Em seguida, foram iniciadas comparações entre as 4 diferentes SUGARWINs.

Quando realizamos testes de similaridade de sequência, observamos a similaridade na região do *loop* variável, especialmente na parte mais expressiva da movimentação do *loop* (entre os resíduos de posição 15 – 30). A conservação dos resíduos foi considerada o parâmetro mais importante nesta análise. Como pode ser visto na Figura 9 (APÊNDICE), a baixa conservação pode ser observada entre as proteínas SUGARWINs na região do *loop*, entre os resíduos 15 e 30. Uma deleção nos resíduos 26 e 27 também pode ser observada em SUGARWIN2, que apresenta grande diferença na região do *loop* com a SUGARWIN1. Quanto aos percentuais de similaridade, SUGARWIN1 e SUGARWIN2 apresentam 75,4% de similaridade, enquanto SUGARWIN3 e SUGARWIN4, apresentam 70,7% de similaridade entre si.

## 8. Dinâmica molecular das SUGARWINs



As análises da dinâmica molecular revelaram características individuais interessantes. Em todas as análises, observamos se as regiões de grande movimentação estavam próximas de possíveis sítios catalíticos. Ressalta-se que, embora tenhamos acrescentado nas nossas análises o modelo de estrutura das novas SUGARWINs (3 e 4), ainda não há nenhuma informação sobre suas funções.

O RSMD (Figura 10 – APÊNDICE), através da análise de frames específicos sem o estímulo, revelou um padrão de conservação na região C-terminal em todas as proteínas. O qual abre precedentes sobre esta região ter alguma importância ainda não explorada. No gráfico, dois picos da SUGARWIN1 na região N-terminal próximo ao *loop* variável, são destacados. O RSMF avaliou as mudanças de conformação e flutuação/variação de resíduos mediante estímulo (energia), mostrando um padrão de mudanças bem parecido com o RSMD.

A análise de componentes principais (PCA) avaliou grupos de diferentes movimentações, revelando informações relevantes para cada uma das proteínas. Como pode ser visto na Figura 11 (APÊNDICE), a SUGARWIN1 apresentou alta mobilidade em regiões bem características (*loop* variável), com movimentações bem expressivas. De modo que suas regiões mais móveis podem estar relacionadas às diferentes funções que desempenham. Já a SUGARWIN2 é a estrutura com menor mobilidade, restringindo-se a um pequeno *loop* secundário que fica logo após o *loop* variável. A SUGARWIN4 apresenta padrão de movimentação parecido com a SUGARWIN2, inclusive na mesma região, com movimentação pouco mais expressiva. Além disso, ela conta com uma segunda movimentação leve, que aparentemente não tem função conhecida. A SUGARWIN3 conta com um padrão de movimentação completamente diferente das demais. Embora pouco expressiva, sua movimentação é observada próxima da porção C-terminal.

Complementando os dados já obtidos, a Figura 12 (APÊNDICE), ilustra a região das SUGARWINs que possivelmente faz ligação com o quitosano (a priori, o único substrato comum às quatro). Aqui, cabe ressaltar que o *loop* móvel da SUGARWIN1 se movimenta sobre esta região, passando por cima do sítio de ligação ao quitosano e ainda assim, não interfere na função quitosanase. Isto evidencia a importância da flexibilidade desta

proteína para desempenhar suas variadas funções. As demais proteínas não possuem movimentações sobre este sítio ativo, mantendo-o livre permanentemente.

As análises de dinâmica molecular das SUGARWINS, bem como a modelagem das recém-descobertas SUGARWINS 3 e 4, foram realizadas em parceria com o Departamento de Farmacociências, da Universidade Federal de Ciências da Saúde de Porto Alegre e o Programa de Pós-Graduação de Biologia Celular e Molecular, da Universidade Federal do Rio Grande do Sul, em Porto Alegre/RS, sob a supervisão do Prof. José Fernando Ruggiero Bachega.

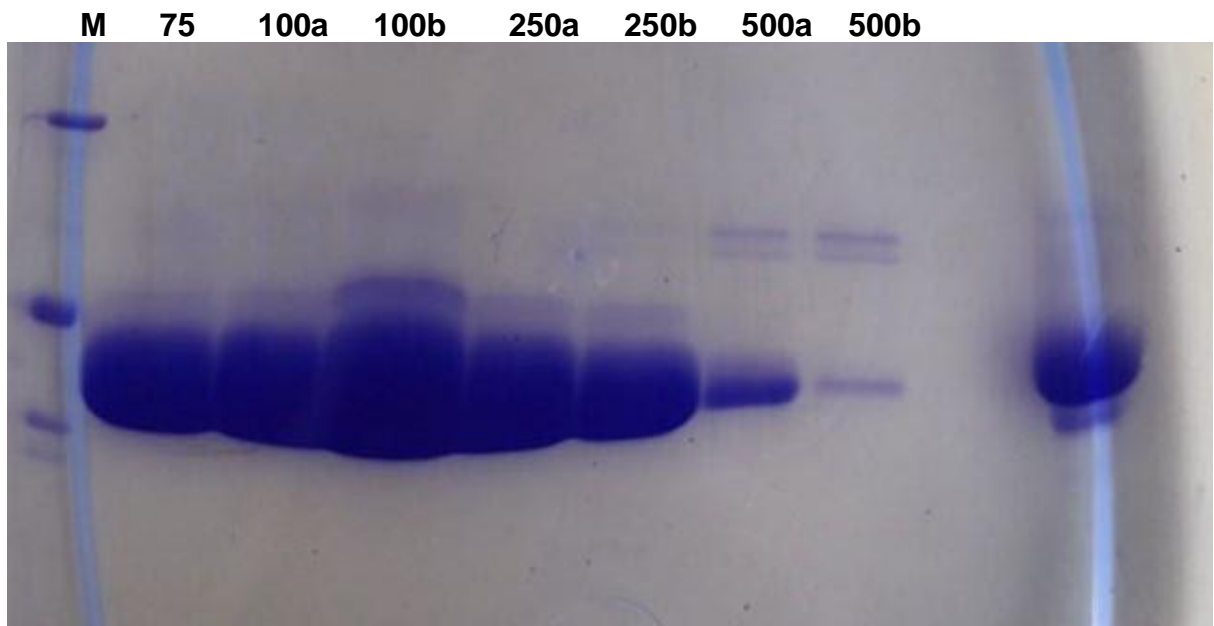
## CONSIDERAÇÕES FINAIS

Neste trabalho, pretendemos conectar estrutura e evolução das proteínas SUGARWINs, alinhando dados da biologia estrutural e história evolutiva, na busca por compreender o comportamento funcional de proteínas com relevante potencial biotecnológico para a agricultura da cana-de-açúcar.

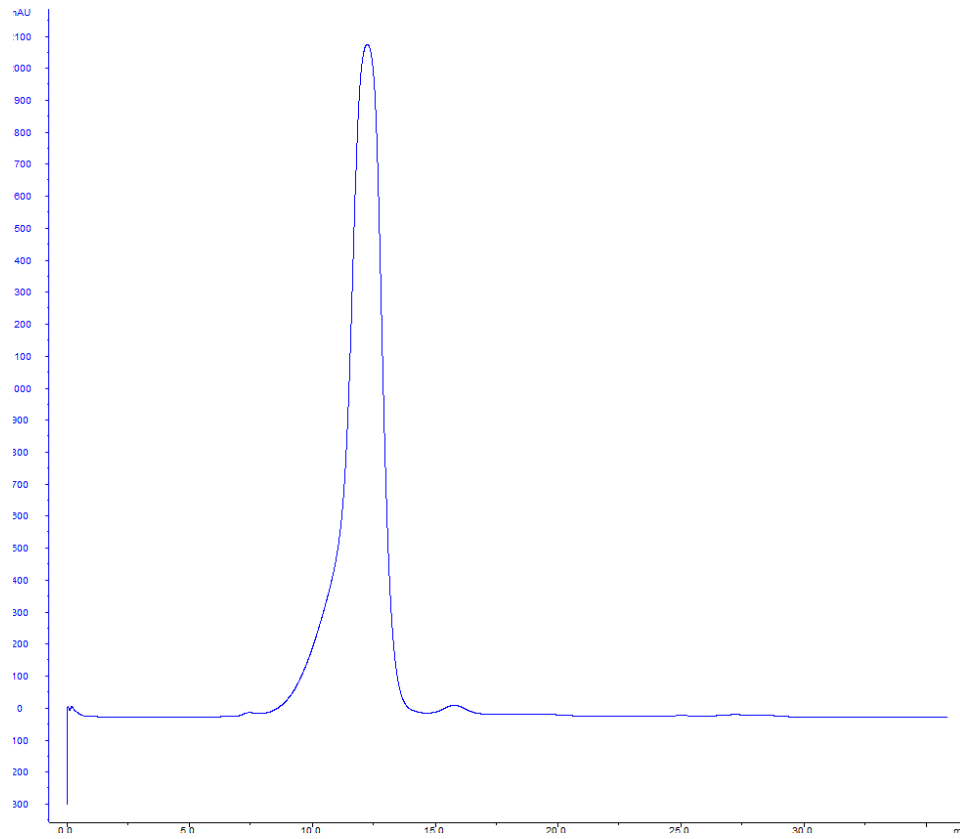
A presença de um *loop* flexível é determinante para o desempenho da função ribonuclease. A natureza dinâmica da SUGARWIN1, particularmente a flexibilidade do *loop* variável, permite também a diversidade de suas funções. A análise comparativa da biologia estrutural nos revelou que a região do *loop* variável apresentou significativas variações entre as SUGARWINs. A análise filogenética aponta para um padrão evolutivo de conservação na região C-terminal (extremidade oposta ao *loop* variável tão relevante para a atividade nuclease). A perda da função RNase neste grupo de proteínas parece ser o evento mais recente. As alterações em SUGARWIN2, desde a deleção de dois resíduos no *loop* variável, a substituição de uma histidina por asparagina (H11N) e alteração no potencial eletrostático de superfície pela substituição de resíduos básicos, tornaram sua estrutura mais rígida e estável, conferindo-lhe vantagem evolutiva.

As diferentes análises realizadas neste trabalho nos permitiram melhor compreensão acerca do comportamento funcional das proteínas SUGARWINs na defesa da planta. Diante disso, um leque de novas estratégias de combate aos patógenos causadores de doenças, que acometem a plantação da cana-de-açúcar, se apresenta. Assim, mais uma vez, a análise da biologia estrutural de macromoléculas com potencial biotecnológico se mostra crucial para o desenvolvimento de novas soluções. Aqui, demonstramos o potencial da biotecnologia na solução de problemas enfrentados pela agroindústria no cultivo deste importante produto agrícola no Brasil.

## APÊNDICES

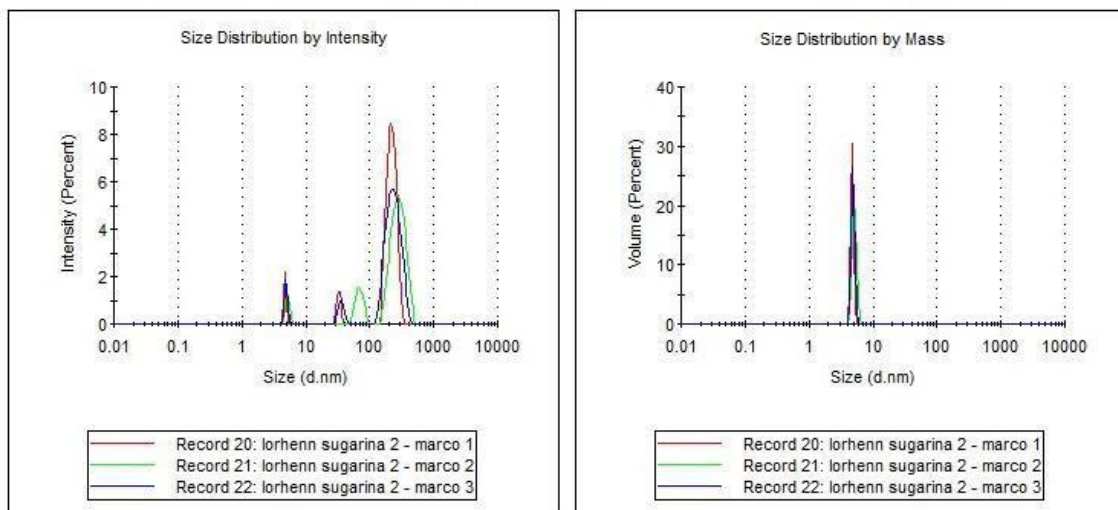


**Figura 1.** Gel SDS-PAGE 12% mostrando a purificação de SUGARWIN2 por IMAC (cromatografia de afinidade). Marcador de peso molecular Lane M; Pista 1-7 alíquotas de frações eluidas com concentrações crescentes de imidazol (75 a 500 mM). A seta indica a banda que representa a proteína recombinante SUGARWIN2 purificada.

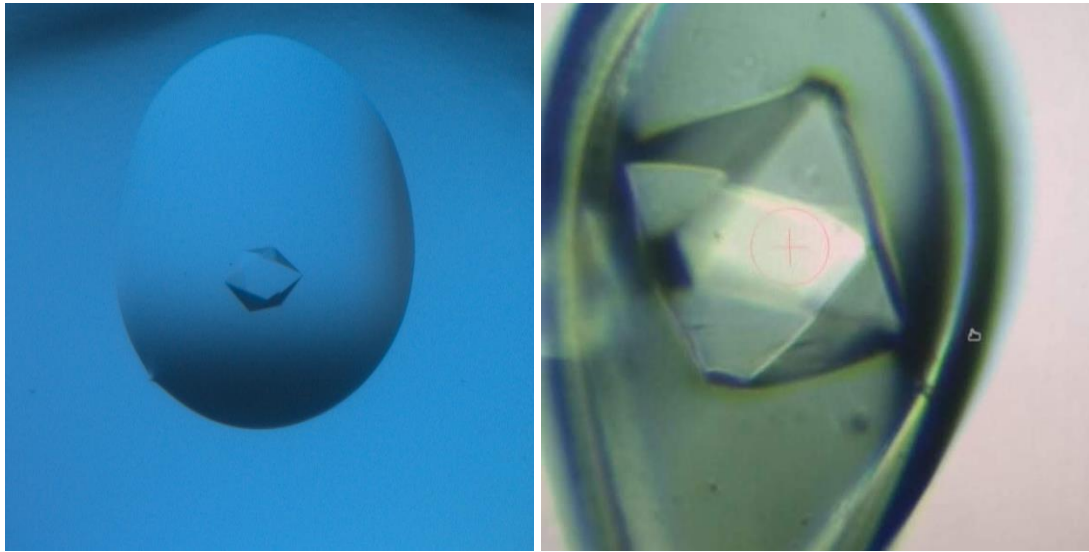


**Figura 2.** Resultado da segunda purificação de SUGARWIN2. O pico ilustrado, resulta da etapa de purificação em coluna de gel filtração, realizado no AKTA, após a obtenção da proteína por expressão heteróloga. No processo de purificação a proteína foi solubilizada em solução estabilizadora para posteriormente seguir para os ensaios de cristalização. Os dados da purificação apontam para uma configuração monomérica da proteína.

**Fonte:** Imagem obtida por meio do AKTA.



**Figura 3.** DLS da SUGARWIN2. O DLS da proteína SUGARWIN2 apresentou um pico monodisperso que correspondeu a 99,7% da massa consistente com a massa da proteína (13kDa).



**Figura 4.** Cristal da proteína SUGARWIN2. A) Cristal obtido nos ensaios de cristalização da proteína SUGARWIN2 na gota com solução precipitadora em placa do ensaio automatizado de cristalização. B) Cristal em *loop* pronto para ser difratado.

**Fonte:** Imagem obtida por microscopia

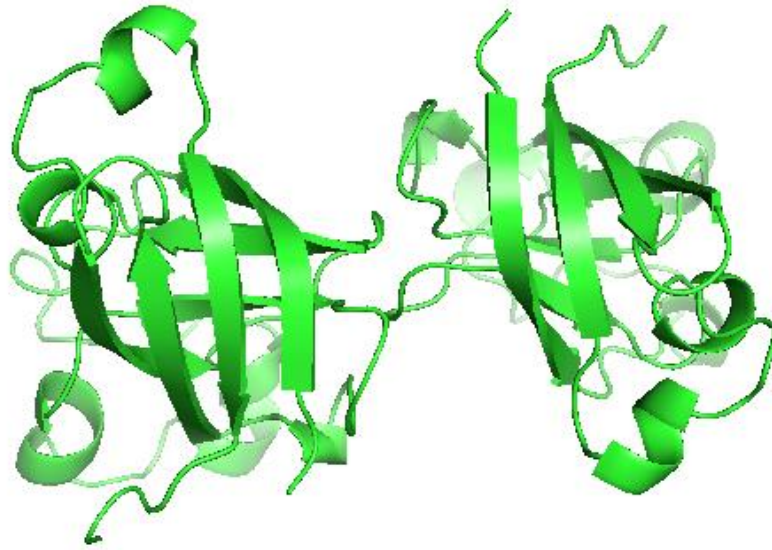
```

SUGARWIN1  MAAAAISGGRAACALVVVAGVLCAVAGMAAAQQASNVRATYHYYNPQNGWNLNAVSAYC
SUGARWIN2  -----MAGMTTVSKLALAAVLLCAAAMATAQQASGVRATYHYYNPTQNNWDLA--GTYC
          ::*  :..  *..*  :***.*:**:*****.*****:****  **.*:*  .:**

SUGARWIN1  ATWDADKKPLSWRQKYGWTAFCGPAGQKGQAACGKCIRVTNRATGASIVARIVDQCSNGGL
SUGARWIN2  ATWDAGQPLSWRSKYGWTAFCGPAGPTGQAACGQCLLVTNTATGASLTVRIVDQCSNGGL
          *****.:*****.*****.*****.*****:*.  ***  *****:..*****

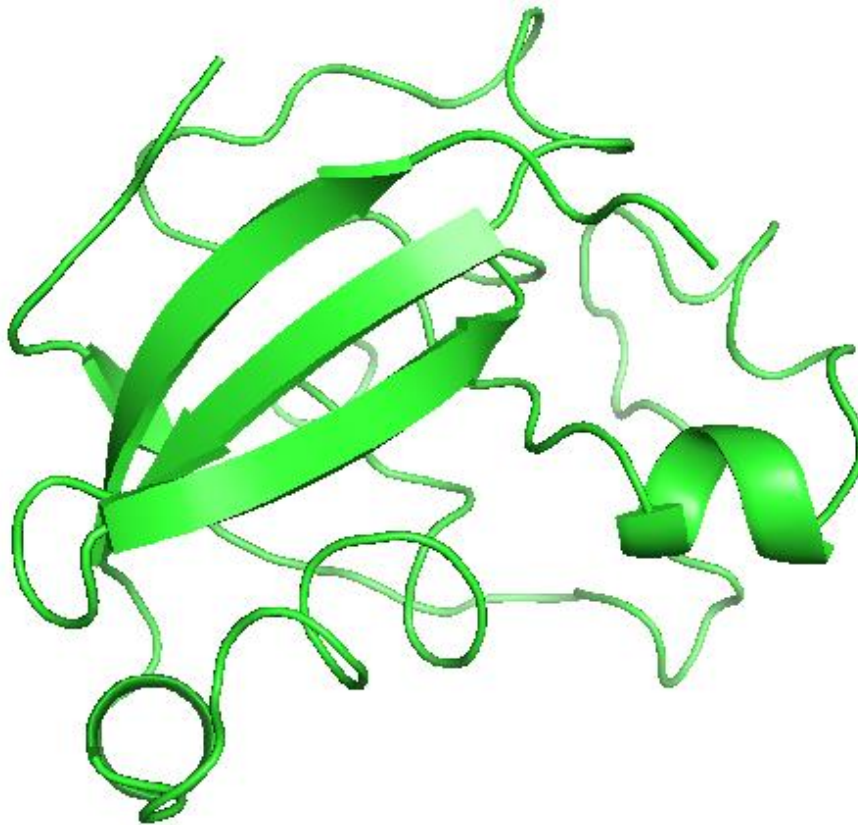
SUGARWIN1  DLDYETVFKKIDTNGGGYQMGHLNVNYQFVAC---
SUGARWIN2  DLDYDTAFKPLDTNGAGIQAGHLTVNYQFVNCGDN
          ****.*..**  :****  *  *  ***.*****  *
  
```

**Figura 5.** Alinhamento da SUGARWIN1 e SUGARWIN2. O alinhamento destaca em azul os aminoácidos básicos (arginina e lisina) presentes na SUGARWIN1. Em vermelho, as histidinas (também básicas) nas posições 11 e 111. SUGARWIN2 mostra a substituição de histidina por asparagina na posição 11, descrita como histidina catalítica na função nuclease de outras WINS.



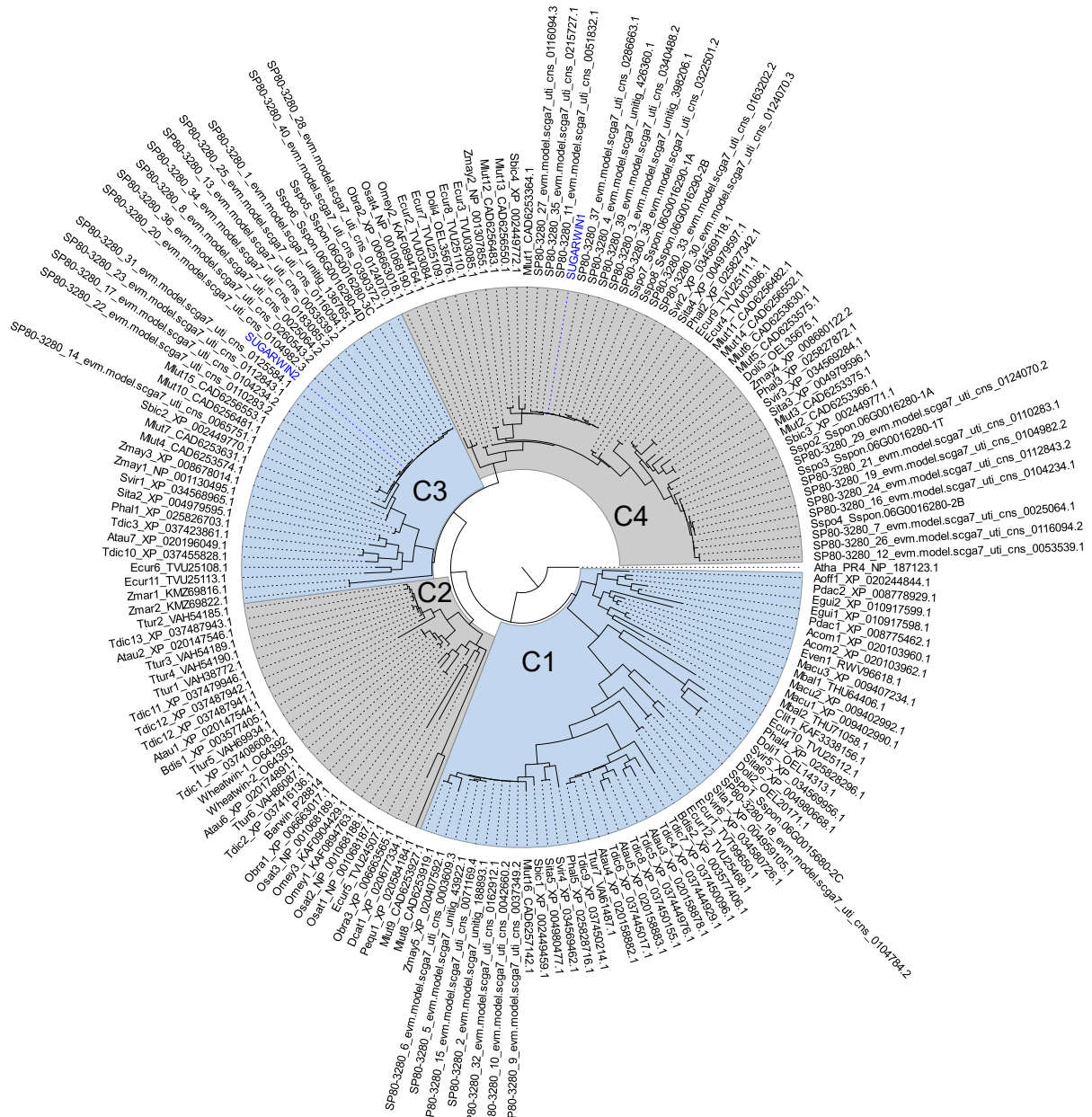
**Figura 6.** Célula unitária da SUGARWIN2. A célula unitária na resolução dos dados cristalográficos da SUGARWIN2 continha duas unidades da proteína.

**Fonte:** Imagem obtida por meio do software de modelagem molecular *pymol*.



**Figura 7.** Estrutura 3D de SUGARWIN2. Estrutura final da proteína SUGARWIN2 resolvida e refinada, como foi depositada no PDB.

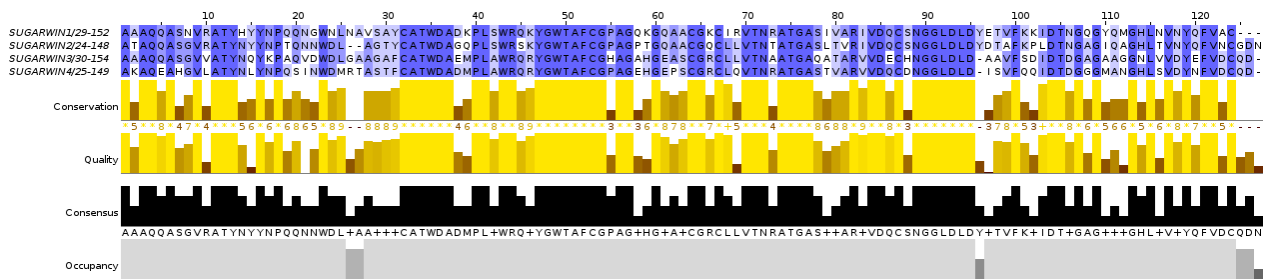
**Fonte:** Imagem obtida por meio do software de modelagem molecular *pymol*.



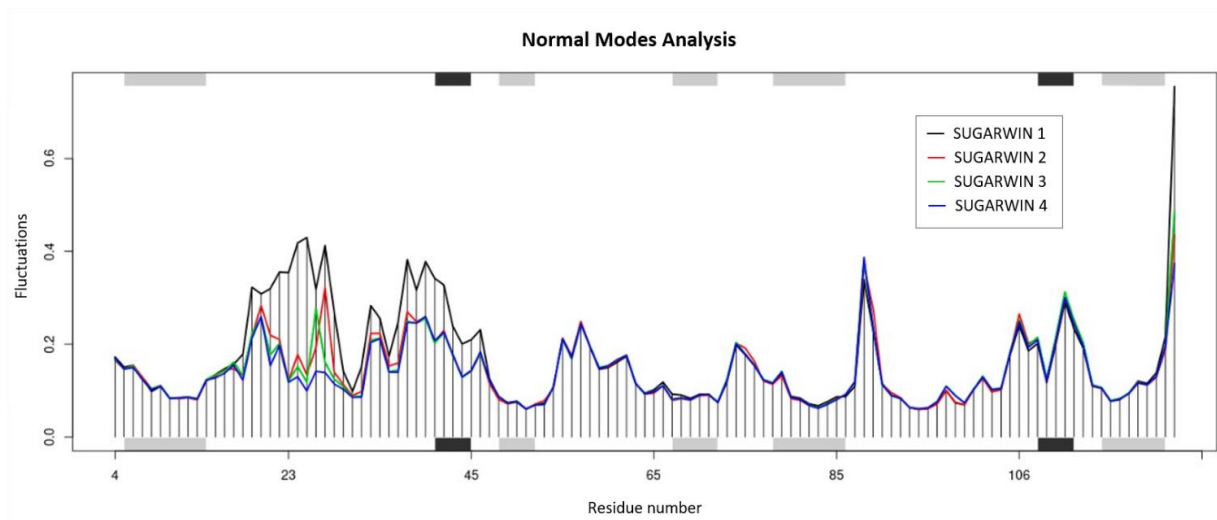
0.2

**Figura 8.** Árvore filogenética de máxima verossimilhança para seqüências de proteínas homólogas a SUGARWINs. Os números nos nós representam os valores de *bootstrap* com base em 1000 replicações. Um código de quatro letras é usado para identificar os rótulos de seqüência na árvore.

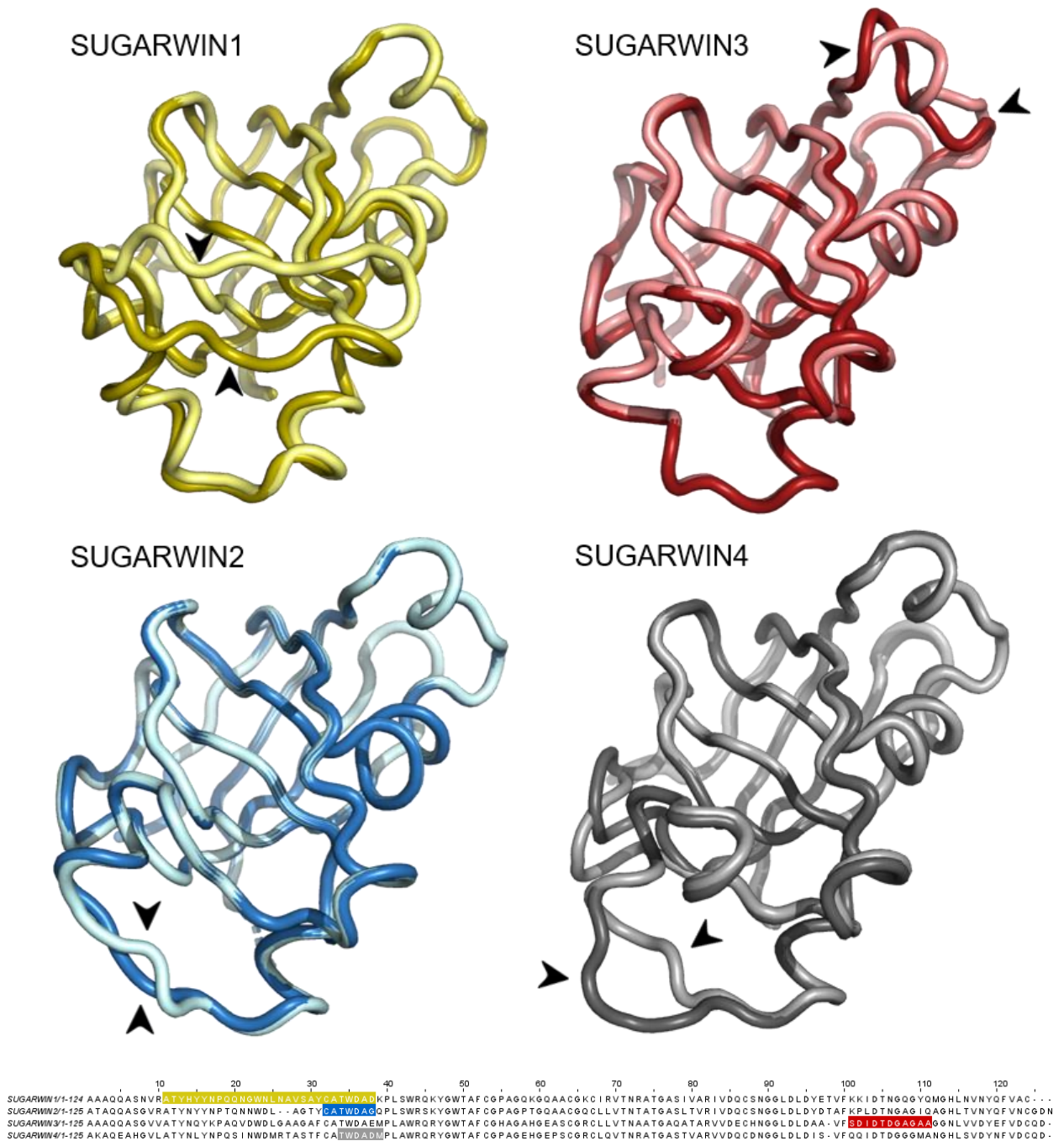




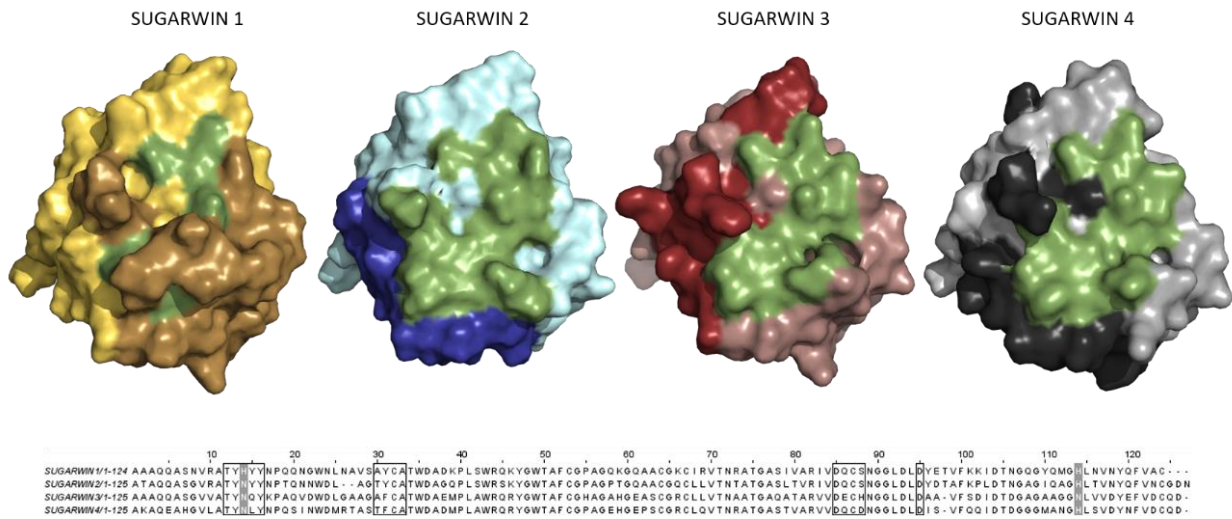
**Figura 9.** Análise de similaridade de seqüências. A conservação dos resíduos foi considerada o parâmetro mais importante na análise de similaridade de seqüência. Baixa conservação pode ser observada entre as proteínas SUGARWINs na região do loop, entre os resíduos 15 e 30. Uma deleção nos resíduos 26 e 27 pode ser observada em SUGARWIN2.



**Figura 10.** Análise de modos normais. Análise de modos normais para as estruturas das SUGARWINs 1, 2, 3 e 4. A barra preta indica a estrutura secundária de uma hélice e a barra cinza indica as folhas.



**Figura 11.** Análise PCA das SUGARWINs. Estruturas de SUGARWINs 1, 2, 3 e 4 geradas pela análise de Bio3D PCA indicando os principais movimentos de cada proteína. Pontas de flechas pretas destacam regiões de grande flexibilidade. As regiões flexíveis são destacadas em sequência.



**Figura 12.** Sítio de ligação quitosana das SUGARWINs. Estruturas de SUGARWINs 1, 2, 3 e 4 em representação de superfície mostrando em verde as regiões previamente identificadas como o local de ligação para a atividade de quitosanasase de SUGARWIN 2. Cores mais escuras indicam regiões de maior flexibilidade de acordo com a análise de RMSF, que se sobrepõem às sequências do local de ligação em alguns casos. Na sequência, as regiões de sítio de ligação são destacadas em caixas e as regiões H11 / N11 e H111 / N111 são destacadas em cinza.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Agrawal, G. K.; Jwa, N.S.; Han, K.S.; Agrawal, V.P.; Rakwal, R. Isolation of a novel rice PR-4 type gene whose mRNA expression is modulated by blast pathogen attack and signaling components. **Plant Physiol. Biochem.** 41, 81-90, 2003.
- Aitipamula, S.; Wong, A.B.H.; Kanaujia, P. Evaluating suspension formulations of theophylline cocrystals with artificial sweeteners. **J. Pharm. Sci.** 107, 604–611, 2018.
- Arantes, J. T. Fungo causador da podridão vermelha ‘manipula’ insetos e plantas para se disseminar pelos canaviais. **Agência FAPESP**, 2021. Disponível em: <https://agencia.fapesp.br/fungo-causador-da-podridao-vermelha-manipula-insetos-e-plantas-para-se-disseminar-pelos-canaviais/36100/> Acesso em: 28 de outubro de 2021.
- Bai, S.; Dong, C.; Li, B.; Dai, H. A PR-4 gene identified from *Malus domestica* is involved in the defense responses against *Botryosphaeria dothidea*. **Plant Physiol. Biochem.** 62, 23–32, 2013.
- Baek, M.; DiMaio, F.; Anishchenko, I. et al. Accurate prediction of protein structures and interactions using a three-track neural network. **Science**, 373, 871–876, 2021.
- Beddell, C.R.; Goodford, P.J.; Norrington, F.E.; Wilkinson, S.; Wootton, R. Compounds designed to fit a site of known structure in human haemoglobin. **Br. J. Pharmacol.** 57, 201–209, 1976.
- Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. **The Protein Data Bank**, 2000.
- Bernal, J.D.; Crowfoot, D. X-ray photographs of crystalline pepsin. **Nature**, 133, 794–795, 1934.
- Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. **J. Mol. Biol.** 112, 535–542, 1977.
- Bertini, L.; Caporale, C.; Testa, M.; Proietti, S.; Caruso C. Structural basis of the antifungal activity of wheat PR-4 proteins. **FEBS Letters**, 583, 2865–2871, 2009.
- Bertini, L.; Proietti, S.; Aleandri, M.P.; Mondello, F.; Sandini, S.; Caporale, C.; Caruso, C. Modular structure of HEL protein from *Arabidopsis* reveals new potential functions for PR-4 proteins. **Biol. Chem.** 393, 1533–1546, 2012.

Bostock, R.M. Signal conflicts and synergies in induced resistances to multiple attackers. **Physiol Mol Plant Pathol**, 55, 99–109, 1999.

Bravo, J.M.; Campo, S.; Murillo, I.; Coca, M.; Segundo, B.S. Fungus- and wound-induced accumulation of mRNA containing a class II chitinase of the pathogenesis-related protein 4 (PR-4) family of maize. **Plant Mol. Biol.** 52, 745-759, 2003.

Broekaert, I.; Lee, H.I.; Kush, A.; Chua, N.H.; Raikhel, N. Wound-induced accumulation of mRNA containing a hevein sequence in laticifers of rubber tree (*Hevea brasiliensis*). **Proc. Natl. Acad. Sci. U.S.A.** 87, 7633-7637, 1990.

Burg S.V.; Alvarez-Alfageme, F.; Romeis, J. Indirect effect of a transgenic wheat on aphids through enhanced powdery mildew resistance. **PloS One**, 7: e46333, 2012.

Callaway, E. 'It will change everything': AI makes gigantic leap in solving protein structures DeepMind's program for determining the 3D shapes of proteins stands to transform biology, say scientists. **Springer Nature**, 588, 203-204, 2020.

Caporale, C.; Caruso, C.; Facchiano, A.; Nobile, M.; Leonardi, L.; Bertini, L.; Colonna, G.; Buonocore, V. Probing the modelled structure of wheatwin1 by controlled proteolysis and sequence analysis of unfractionated digestion mixtures by controlled. **PROTEINS: Structure, Function, and Genetics**: 36, 192–204, 1999.

Caporale, C.; Di Berardino, I.; Leonardi, L.; Bertini, L.; Cascone, A.; Buonocore, V.; Caruso, C. Pathogenesis-related proteins of class 4 have ribonuclease activity. **FEBS Lett.** 575 (1–3), 71–76, 2004.

Caruso, C.; Bertini, L.; Tucci, M.; Caporale, C.; Leonardi, L.; Saccardo, F.; Bressan, R.A.; Veronese, P.; Buonocore, V. Isolation and characterisation of wheat cDNA clones encoding PR-4 proteins. **DNA Seq.** 10, 301-307, 1999.

Caruso, C.; Bertini, L.; Tucci, M.; Caporale, C.; Nobile, M.; Leonardi, L.; Buonocore, V. Recombinant wheat antifungal PR-4 proteins expressed in *Escherichia coli*. **Protein Expr. Purif.** 23, 380-388, 2001.

Carruthers, R.I.; Bergstrom, G.C.; Haynes, P.A. Accelerated development of the European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae), induced by interactions with *Colletotrichum graminicola* (Melanconiales: Melanconiaceae), the causal fungus of maize anthracnose. **Ann Entomol Soc Am**, 79, 385–389, 1986.

Christensen, A.B.; Cho, B.H.; Naesby, M.; Gregersen, P.L.; Brandt, J.; Madriz-Ordenana, K.; Collinge, D.B.; Thordal-Christensen, H. The molecular characterization of two barley proteins establishes the novel PR-17 family of pathogenesis-related proteins. **Mol. Plant Pathol.** 3, 135–144, 2002.

Christopher, M.E.; Miranda, M.; Major, I.T.; Constabel, C.P. Gene expression profiling of systemically wound-induced defenses in hybrid poplar. **Planta** 219, 936-947, 2004.

Clark, G.L.; Corrigan, K.E. The crystal structure of insulin. **Phys. Rev.** 40, 639, 1932.

Cooper, D.R.; Porebski, P.J.; Chruszcz, M.; Minor, W. X-ray crystallography: assessment and validation of protein-small molecule complexes for drug discovery. **Expert Opin Drug Discov.** 6, 771–782, 2011.

Cui, J.; Jander, G.; Racki, L.R.; Kim, P.D.; Pierce, N.E.; Ausubel, F.M. Signals involved in *Arabidopsis* resistance to *Trichoplusia ni* caterpillars induced by virulent and avirulent strains of the phytopathogen *Pseudomonas syringae*. **Plant Physiol**, 129, 551–564, 2002.

D'Alessio, G.; Riordan, F. Ribonucleases: Structures and Functions, **Academic Press**, New York, 1997.

Dauter, Z.; Wlodawer, A. Progress in protein crystallography. **Protein Pept. Lett.** 23, 201–210, 2016.

Dinardo-Miranda, L.L. Pragas. In: Dinardo-Miranda LL, Vasconcelos ACM, Landell MGA. **Cana-de-açúcar**, Instituto Agronômico, 394-404, 2010.

Falco, M.C.; Marbach, P.A.S.; Pompermayer, P.; Lopes, F.C.C.; Silva-Filho, M.C. Mechanisms of sugarcane response to herbivory. **Genet. Mol. Biol.** 24, 113-122, 2001.

Felton, G.W.; Korth, K.L.; Bi, J.L.; Wesley, S.V.; Huhman, D.V.; Mathews, M.C.; Murphy, J.B.; Lamb, C.; Dixon, R.A. Inverse relationship between systemic resistance of plants to microorganisms and to insect herbivory. **Curr Biol**, 9, 317–320, 1999.

Franco, F.P.; Santiago, A.C.; Henrique-Silva, F.; de Castro, P.A.; Goldman, G.H.; Moura, D.S.; Silva-Filho, M.C. The sugarcane defense protein SUGARWIN2 causes cell death in *Colletotrichum falcatum* but not in non-pathogenic fungi. **PLoS ONE**, 9, e91159, 2014.

Franco, F.P.; Moura, D.S.; Vivanco, J.M.; Silva-Filho, M.C. Plantinsect- pathogen interactions: a naturally complex menage à trois. **Curr. Opin. Microbiol.** 37, 54–60, 2017.

Franco, F.P., Dias, R.O., Toyama, D., Henrique-Silva, F., Moura, D. S., Silva-Filho, M.C. Structural and functional characterization of PR-4 SUGARWINs from sugarcane and their role in plant defense. **Front. Plant Sci.** 9, 1916, 2019.

Franco, F.P.; Túler, A.C.; Gallan, D.Z.; Gonçalves, F.G.; Favaris, A.P.; Peñafior, M.F.G.V.; Leal, W.S.; Moura, D.S.; Bento, J.M.S.; Silva-Filho, M.C. Fungal phytopathogen modulates plant and insect responses to promote its dissemination. **ISME J.** 15, 10, 2021.

Friedrich, L.; Moyer, M.; Ward, E.; Ryals, J. Pathogenesis-related protein-4 is structurally homologous to the carboxy-terminal domains of hevein, win-1 and win-2. **Mol. Gen. Genet.** 230, 113-119, 1991.

- Gatch, E.W.; Munkvold, G.P. Fungal species composition in maize stalks in relation to european corn borer injury and transgenic insect protection. **Plant Disease**, 86, 1156–1162, 2002.
- Guevara-Morato, M.; de Lacoba, M.G.; Garcia-Luque, I.; Serra, M.T. Characterization of a pathogenesis-related protein 4 (PR-4) induced in *Capsicum chinense* L-3 plants with dual RNase and DNase activities. **J. Exp. Bot.** 61, 3259–3271, 2010.
- Hatcher, P.E.; Paul, N.D. On integrating molecular and ecological studies of plant resistance: variety of mechanisms and breadth of antagonists. **J Ecol**, 88, 702–706, 2000.
- Hejgaard, J.; Jacobsen, S.; Bjorn, S.E.; Kragh, K.M. Antifungal activity of chitin-binding PR-4 type proteins from barley grain and stressed leaf. **FEBS (Feb. Eur. Biochem. Soc.) Lett.** 307, 389-392, 1992.
- Hol, W.G.J. Protein crystallography and computer graphics - Toward rational drug design. **Angew. Chem.Int. Ed. Eng.** 25, 767–778, 1986.
- Huet, J.; Jean, E.; Mbosso, T.; Soror, S.; Meyer, F.; Looze, Y.; Wintjensa, R.; Wohlkonig, A.; High-resolution structure of a papaya plant-defence barwin-like protein solved by in-house sulfur-SAD phasing. **Biol Crystal**, 69, 2017–2026, 2013.
- Joosten, R.P.; Joosten, K.; Murshudov, G.N.; Perrakis, A. PDB\_REDO: constructive validation, more than just looking for errors. **Acta Crystallogr.** 68, 484–496, 2012.
- Kendrew, J.C.; Dickerson, R.E.; Strandberg, B.E.; Hart, R.G.; Davies, D.R.; Phillips, D.C.; Shore, V.C. Structure of myoglobin: A three-dimensional fourier synthesis at 2 Å. Resolution. **Nature**, 185, 422–427, 1960.
- Kiba, A.; Saitoh, H.; Nishihara, M.; Omiya, K.; Yamamura, S. C-terminal domain of a hevein-like protein from *Wasabia japonica* has potent antimicrobial activity. **Plant Cell Physiol.** 44, 296-303, 2003.
- Kim, N. H.; Hwang, B.K. Pepper pathogenesis-related protein 4c is a plasma membrane-localized cysteine protease inhibitor that is required for plant cell death and defense signaling. **Plant J.** 81, 81–94, 2015.
- Koorneef, A.; Leon-Reyes, A.; Ritsema, T.; Verhage, A.; den Otter, F.C.; Van Loon, L.C.; Pieterse, C.M.J. Kinetics of salicylate-mediated suppression of jasmonate signaling reveal a role for redox modulation. **Plant Physiol**, 147, 1358–1368, 2008.
- Lawton, K.A.; Potter, S.L.; Uknes, S.; Ryals, J. Acquired resistance signal transduction in *Arabidopsis* is ethylene independent. **Plant Cell**, 6: 581–588, 1994.
- Lesuisse, D.; Lange, G.; Deprez, P.; Benard, D.; Schoot, B.; Delettre, G.; Marquette, J.P.; Broto, P.; Jean-Baptiste, V.; Bichet, P.; Sarubbi, E.; Mandine, E. SAR and X-ray. A new

approach combining fragment-based screening and rational drug design: Application to the discovery of nanomolar inhibitors of Src SH2. **J. Med. Chem.** 45, 2379–2387, 2002.

Li, X.; Xia, B.; Jiang, Y.; Wu, Q.; Wang, C.; He, L.; Peng, F.; Wang, R. A new pathogenesis-related protein, LrPR-4, from *Lycoris radiata*, and its antifungal activity against *Magnaporthe grisea*. **Mol. Biol. Rep.** 37, 995-1001, 2010.

Linthorst, H.J.M.; Danhash, N.; Brederode, F.T.; Vankan, J.A.L.; Dewit, P.J.G.M.; Bol, J.F. Tobacco and tomato PR proteins homologous to win and pro-hevein lack the hevein domain. **Mol. Plant-Microbe Interact.** 4, 586-592, 1991.

Maveyraud, L.; Mourey, L. Protein X-ray Crystallography and Drug Discovery. **Molecules**, 25, 1030, 1-18, 2020.

McFarlane, S.A.; Govender, P.; Rutherford, R.S. Interactions between *Fusarium* species from sugarcane and the stalk borer, *Eldana saccharina* (Lepidoptera: Pyralidae). **Ann Appl Biol**, 155, 349–359, 2009.

McIntyre, P.J.; Collins, P.M.; Vrzal, L.; Birchall, K.; Arnold, L.H.; Mpamhanga, C.; Coombs, P.J.; Burgess, S.G.; Richards, M.W.; Winter, A.; Veverka, V.; von Delft, F.; Merritt, A.; Bayliss, R. Characterization of three druggable hot-spots in the Aurora-A/TPX2 interaction using biochemical, biophysical, and fragment-based approaches. **ACS Chem. Biol.** 12, 2906–2914, 2017.

Medeiros, A.H.; Franco, F.P.; Matos, J.L.; de Castro P.A.; Santos-Silva, L.K.; Henrique-Silva, F.; Goldman, G.H.; Moura, D.S.; Silva-Filho, M.C. Sugarwin: A sugarcane insect-induced gene with antipathogenic activity. **Mol Plant-Microbe Int**, 25, 613–624, 2012.

Menezes, S.P.; de Andrade Silva, E.M.; Lima, E.M.; de Sousa, A.O.; Andrade, B.S.; Lima Lemos, L.S.L.; Gramacho, K.P.; Gesteira, A.S.; Pirovani, C.P.; Micheli, F. The pathogenesis-related protein PR-4b from *Theobroma cacao* presents RNase activity, Ca<sup>2+</sup>C and Mg<sup>2+</sup>C dependent- DNase activity and antifungal action on *Moniliophthora perniciosa*. **BMC Plant Biol.** 14, 161, 2014.

Menossi, M.; Silva-Filho, M.C.; Vincentz, M.; Van-Sluys, M.A.; Souza, G.M. Sugarcane functional genomics: gene discovery for agronomic trait development. **Int J Plant Genomics**, 1, 1-11, 2008.

Mur, L.A.J.; Kenton, P.; Atzorn, R.; Miersch, O.; Wasternack, C. The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. **Plant Physiol**, 140, 249–262, 2006.

Neuhaus, J.M.; Fritig, B.; Linthorst, H.J.M.; Meins, F.; Mikkelsen, J.D.; Ryals, J. A revised nomenclature for chitinase genes. **Plant Mol. Biol. Rep.** 14, 102–104, 1996.



Pennisi, E. Protein structure prediction now easier, faster. AI approach is accessible to all structural biology, drug discovery researchers. **Science**, 373, 6552, 262-263, 2021.

Pompermayer, P. Identificação e caracterização de um promotor de cana-de-açúcar induzido por herbivoria. 142 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2004.

Ponstein, A. S., Bresvloemans, S. A., Selabuurlage, M. B., Vandeneizen, P. J. M., Melchers, L. S., and Cornelissen, B. J. C. A novel pathogen-inducible and wound-inducible tobacco (*Nicotiana tabacum*) protein with antifungal activity. **Plant Physiol.** 104, 109–118, 1994.

Pieterse, C.M.J.; Leon-Reuyes, A.; Ent. S.; Van der Ent, S.; Van Wees, S.C.M. Networking by small-molecule hormones in plant immunity. **Nature Chem Biol**, 5, 308–316, 2009.

Potter, S.; Uknes, S.; Lawton, K.; Winter, A.M.; Chandler, D.; Dimaio, J.; Novitzky, R.; Ward, E.; Ryals, J. Regulation of a Hevein-like gene in Arabidopsis. *Mol. Plant Microbe Interact.* 6, 680–685, 1993.

Ryan, C.A. Protease Inhibitors in plants—genes for improving defenses against insects and pathogens. **Annu. Rev. Phytopathol.** 28, 425- 449, 1990.

Schillmiller, A.L.; Howe, G.A. Systemic signaling in the wound response. **Curr. Opin. Plant Biol.** 8, 369-377, 2005.

Schulthess, F.; Cardwell, K.F.; Gounou, S. The effect of endophytic *Fusarium verticillioides* on infestation of two maize varieties by lepidopterous stemborers and coleopteran grain feeders. **Phytopathol.** 92, 120–128, 2002.

Schweizer P, Buchala A, Dudler R, Métraux JP. Induced systemic resistance in wounded rice plants. **Plant J**, 14, 475–481, 1998.

Souza, T.P., Dias, R.O., Silva-Filho, M.C. Defense-related proteins involved in sugarcane responses to biotic stress. **Genet. Mol. Biol.** 40, 360–372, 2017.

Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-ray Crystallography to Explore Protein Structure and Dynamics. **Int. J. Mol. Sci.** 19, 3401, 1–23, 2018.

Stennis, M.; Chandra, S.; Ryan, C.A.; Low, P. Systemin potentiates the oxidative burst in cultured tomato cells. **Plant Physiol.** 117, 1031–1036, 1998.

Stout, M.J.; Thaler, J.S.; Thomma, B.P.H.J. Plant-mediated interactions between pathogenic microorganisms and herbivorous arthropods. **Annu Rev Entomol**, 51, 663–689, 2006.

- Svensson, B.; Svendsen, I.; Hojrup, P.; Roepstorff, P.; Ludvigsen, S.; Poulsen, F.M. Primary structure of barwin: a barley seed protein closely related to the C-terminal domain of proteins encoded by wound-induced plant genes. **Biochemistry**, 31, 8767-8770, 1992.
- Terras, F.R.; Penninckx, I.A.; Goderis, I.J.; Broekaert, W.F. Evidence that the role of plant defensins in radish defense responses is independent of salicylic acid. **Planta**, 206, 117–124, 1998.
- Thakral, N.K.; Zanon, R.L.; Kelly, R.C.; Thakral, S. Applications of powder X-ray diffraction in small molecule pharmaceuticals: Achievements and aspirations. **J. Pharm. Sci.** 107, 2969–2982, 2018.
- Thaler, J.; Fidantsef, A.; Duffey, S.; Bostock, R. Trade-offs in plant defense against pathogens and herbivores: a field demonstration of chemical elicitors of induced resistance. **J Chem Ecol**, 25, 1597–1609, 1999.
- Thomma, B.P.; Eggermont, K.; Penninckx; Mauch-Mani, B.; Vogelsang, R.; Cammue, B.P.; Broekaert, W.F. Separate jasmonate-dependent and salicylate-dependent defense-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. **Proc Natl Acad Sci USA**, 95, 15107–15111, 1998.
- UNICA. União da Agroindústria de cana-de-açúcar. Disponível em: <<http://www.portalunica.com.br/>>. Acesso em 20 de setembro de 2021.
- Van Loon, L.C.; Van Strien, E.A. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. **Physiol. Mol. Plant Pathol.** 55, 85–97, 1999.
- Viswanathan, R.; Samiyappan, R. Red rot disease in sugarcane: a major constraint for the Indian Sugar Industry. **Sugar Cane**, 9–15, 1999.
- Walling, L.L. The myriad plant responses to herbivores. **J Plant Growth Regul**, 19, 195–216, 2000.
- Weichenberger, C.X.; Pozharski, E.; Rupp, B. Visualizing ligand molecules in twilight electron density. **Acta Crystallogr.** 69, 195–200, 2013.
- Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. **FEBS Journal**, 280, 5705–5736, 2013.
- Zhu, T.; Song, F.; Zheng, Z. Molecular characterization of the rice pathogenesis-related protein, OsPR-4b, and its antifungal activity against *Rhizoctonia solani*. **J. Phytopathol.** 154, 378-384, 2006.