



UNIVERSIDADE FEDERAL DE PERNAMBUCO
LABORATÓRIO DE IMUNOPATOLOGIA KEIZO ASAMI
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA APLICADA À
SAÚDE

ROSILDA CINTRA SOUZA PAES

**IDENTIFICAÇÃO E CARACTERIZAÇÃO ESTRUTURAL DE
PROTEÍNAS COM DOMÍNIO LECTINA EM FUNGOS
PATOGENICOS EM UMA ANÁLISE *In Silico***

RECIFE
2020

ROSILDA CINTRA SOUZA PAES

**IDENTIFICAÇÃO E CARACTERIZAÇÃO ESTRUTURAL DE
PROTEÍNAS COM DOMÍNIO LECTINA EM FUNGOS
PATOGENICOS EM UMA ANÁLISE *In Silico***

Trabalho apresentado como o
cumprimento das exigências para obtenção
do título de Doutor em Biologia Aplicada à
Saúde pela Universidade Federal de
Pernambuco.

Área De Concentração: Bioquímica /
Bioinformática

Orientadora: Prof^a. Dr^a. Cíntia Renata Costa Rocha

Coorientadora: Prof^a. Dr^a. Kyria Santiago Do Nascimento

RECIFE
2020

Catálogo na fonte
Elaine C Barroso
(CRB4 1728)

Paes, Rosilda Cintra Souza

Identificação e caracterização estrutural de proteínas com domínio lectina em fungos patogênicos em uma análise *in silico* / Rosilda Cintra Souza Paes – 2020.

106 f.: il., fig., tab.

Orientadora: Cíntia Renata Costa Rocha

Coorientador: Kíria Santiago do Nascimento

Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-Graduação em Biologia Aplicada à Saúde, 2020.

Inclui referências e apêndices.

1. Lectinas 2. Fungos patogênicos 3. Fenética I. Rocha, Cíntia Renata Costa (orient.) II. Nascimento, Kíria Santiago do (coorient.) III. Título

572.6

CDD (22.ed.)

UFPE/CB – 2020- 255

ROSILDA CINTRA SOUZA PAES

**IDENTIFICAÇÃO E CARACTERIZAÇÃO ESTRUTURAL DE PROTEÍNAS COM
DOMÍNIO LECTINA EM FUNGOS PATOGÊNICOS EM UMA ANÁLISE *In Silico***

Tese apresentada ao Programa de Pós-Graduação em Biologia Aplicada à Saúde, do Centro de Biociências da Universidade Federal de Pernambuco, como requisito para obtenção do título de Doutor em Biologia Aplicada à Saúde.

Aprovada em: ____/____/____

COMISSÃO EXAMINADORA

Prof^a. Dr^a. Cíntia Renata Costa Rocha – Orientadora (LIKA-UFPE)

Prof^a. Dr^a. Kyria Santiago do Nascimento – Examinador externo (BIOMOL-UFC)

Prof^a. Dr^a. Elaine Malosso - Examinador externo (BIOLOGIA DE FUNGOS-UFPE)

Prof^a. Dr^a. Flavia Figueira Aburjaile - Examinador externo (LGBV-UFPE)

Prof^a. Dr^a. Ana Lúcia Figueiredo Porto - Examinador interno (LIKA-UFPE)

*Dedico este trabalho aos meus pais
João Wellington (in memoriam) e Rosilda Valéria,
ao meu esposo Gilmar e a
minha filha Laura.*

AGRADECIMENTOS

Não há nada que eu tenha feito que não tivesse recebido uma mão amiga. Foram quatro anos de intensa busca e plantio, mas também de muito achado e colheita.

Agradeço ao meu Senhor Deus por sua grande misericórdia em me proporcionar anos de concretizações, pois em tudo Ele me amou;

Também agradeço a Deus por me permitir conseguir concluir a promessa feita ao meu pai João Wellington em ser Doutora;

Agradeço aos meus familiares, meu esposo Gilmar, minha mãe Rosilda e meus irmãos. Eles realmente me seguraram e me levantaram quando muitas vezes caí;

Desejo um especial agradecimento a minha florzinha Laura, presente de Deus, chegou na hora certa, quando eu mais precisava e não percebia e caminhou comigo por todo esse tempo, me alegrando e sarando minhas feridas;

À minha querida orientadora Dra. Cíntia Renata Costa Rocha, que me aceitou como sua aluna, me ensinou tanto sobre proteínas, confiou-me um lindo trabalho e pela grande contribuição à minha formação;

À minha coorientadora Dra. Kyria Santiago do Nascimento e toda sua equipe do Laboratório de Bioquímica e Biologia Molecular da UFC, por me acolherem e me ensinarem tanto;

À professora Dra. Ana Maria Benko Iseppon (UFPE-LGBV), pelo imensurável conhecimento e parceria ofertados a mim, pela generosidade em abraçar meu trabalho, permitindo sua conclusão;

Aos pós-doc Roberta Lane e Carlos André e doutoranda Jéssica Barbosa (UFPE-LGBV), pela parceria no meu trabalho, obrigada por me apoiarem, abraçarem e acreditarem nesse projeto, por se dedicarem a ele e por me ensinar tanto sobre Bioinformática;

Aos meus novos, antigos, bons e queridos amigos (Jéssica Frutuoso, Jéssica Barbosa, Cida Seabra, Yasmim Lopes, Ana Rafaela, Roberta, Bárbara), na verdade anjos que entraram na minha vida e percorreram essa trajetória comigo, me ensinado, me fortalecendo, me apoiando, me motivando, acreditando em mim. Nossa! Como vocês são importantes e especiais pra mim;

E é claro, conheci dois anjos especiais, com mesmo nome, cujo significado quer dizer fitar os olhos em alguma coisa ou alguém, duas Jéssicas. Minhas lindas amigas, anjos da minha vida, obrigada por fitarem os olhos em mim e não me abandonarem;

Aos técnicos e funcionários do LIKA, Sandra, Rafael, Fábio, Moisés e a tantos outros que sempre estando presentes e prontos a ajudar;

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo financiamento para o desenvolvimento deste trabalho e ao Programa de Pós-Graduação em Biologia Aplicada à Saúde – PPGBAS.

"Mas aqueles que esperam no Senhor renovam as suas forças.
Voam bem alto como águias; correm e não ficam exaustos, andam e não se
cansam."

Isaías 40:31

RESUMO

Lectinas são proteínas ubíquas de origem não imune que compõem um grupo específico dentro da classe de adesinas, proteínas envolvidas em muitos processos essenciais de reconhecimento celular e molecular. As lectinas diferem das demais proteínas pela capacidade de ligar-se reversivelmente a células, glicolipídios e glicoconjugados sem modificar suas estruturas. Lectinas vegetais e animais têm sido alvo de muitas investigações científicas devido às suas inúmeras aplicações biotecnológicas como agentes anticâncer, imunomodulador, mitogênico, antioxidante e antiviral. No entanto, lectinas fúngicas ainda são pouco caracterizadas e pouco se sabe sobre sua história evolutiva e seu papel nos processos de desenvolvimento de infecções fúngicas humanas. Entender como ocorre o processo de patogenicidade em fungos é importante para uma maior compreensão de sua interação com o hospedeiro bem como para o desenvolvimento de novas drogas antifúngicas. O fornecimento de sequências completas de genomas de fungos patogênicos possibilita a caracterização e exploração de dados *in silico* para prever estruturas e funções de proteínas de diversas espécies. Com o objetivo de identificar e caracterizar estruturalmente domínios lectina conservados nos fungos patogênicos *Candida albicans*, *Cryptococcus neoformans* e *Aspergillus fumigatus*, sequências sonda de 8 famílias proteicas (L-type, R-type, P-type, EPA, PWP, C-type, CBM, FLO) foram alinhadas via tBLASTn contra os genomas de interesse depositado no NCBI. 58 sequências tiveram seus domínios conservados confirmados com o auxílio do CD-search e foram submetidas à predição do ponto isoelétrico (pI) e peso molecular utilizando o JVirGel 2.0, e a localização subcelular através do Cell-PLoc 2.0. As análises de domínio e fenética realizadas com o programa MEGA7, confirmaram o agrupamento das sequências de acordo com a distribuição dos motivos proteicos gerados pelo MEME, sugerindo que as proteínas hipotéticas (PHs) desempenhem funções semelhantes ao grupo lectínico com sequências similares. Três sequências (2180 / 6800 / CLECT) de *Aspergillus fumigatus* tiveram sua análise de expressão transcricional confirmadas por qPCR, e o modelo estrutural da sequência 2180 foi gerado com ajuda do programa Modeller 9.23. A análise estereoquímica dos aminoácidos permitiu selecionar um modelo com $\geq 90\%$ dos aminoácidos em posições favoráveis. Após a simulação de dinâmica, a estrutura teórica da AfLtype1 permaneceu com seu núcleo conservado apresentando homologia com uma lectina animal vesicular integral de 36 kDa (VIP36), que atua no transporte de glicoproteínas entre o Golgi e o retículo endoplasmático além de regular a fagocitose em macrófagos. Este é o primeiro estudo de identificação e caracterização estrutural de lectinas nos fungos patogênicos de importância médica analisados, e o conhecimento adquirido poderá contribuir para a tão necessária elucidação do processo molecular de patogenicidade destes fungos e suas interações com o hospedeiro humano.

Palavras-chave: Fungos patogênicos. Lectinas. Dinâmica. Modelagem. Fenética.

ABSTRACT

Lectins are ubiquitous proteins of non-immune origin that make up a specific group within the adhesin class, which are proteins involved in many essential cellular and molecular recognition processes. These lectins differ from other proteins in their ability to reversibly bind to cells, glycolipids and glycoconjugates without modifying their structures. Plant and animal lectins have been the subject of many scientific investigations due to their numerous biotechnological applications as anticancer, immunomodulating, mitogenic, antioxidant and antiviral. However, fungal lectins are still poorly characterized and little is known about their evolutionary history and their role in the processes of development of human fungal infections. Understanding how the pathogenicity process occurs in fungi is important for a better understanding of its interaction with the host as well as for the development of new antifungal drugs. Providing complete sequences of pathogenic fungal genomes enables the characterization and exploitation of *in silico* data to predict protein structures and functions of various species. In order to identify and structurally characterize conserved lectin domains in the pathogenic fungi *Candida albicans*, *Cryptococcus neoformans* and *Aspergillus fumigatus*, probe sequences of 8 protein families (L-type, R-type, P-type, EPA, PWP, C-type, CBM, FLO) were aligned via tBLASTn against the genomes of interest deposited with the NCBI. 58 sequences had their conserved domains confirmed with the aid of CD-search and were subjected to prediction of isoelectric point (pI) and molecular weight using JVirGel 2.0, and subcellular localization using Cell-PLoc 2.0. The domain and phenetic analysis performed with the MEGA7 program confirmed the sequence grouping according to the distribution of protein motifs generated by MEME, suggesting that the hypothetical proteins (PHs) perform similar functions to the lectinic group with similar sequences. Three *Aspergillus fumigatus* (2180 / 6800 / CLECT) sequences had their transcriptional expression analysis confirmed by qPCR, and the structural model of the 2180 sequence was generated with the help of the Modeller 9.23 program. Stereochemical analysis of amino acids allowed to select a model with $\geq 90\%$ of amino acids in favorable positions. After the dynamics simulation, the AfLtype1 theoretical structure remained with its conserved nucleus presenting homology with a 36 kDa integral vesicular animal lectin (VIP36), which acts in the transport of glycoproteins between the Golgi and the endoplasmic reticulum in addition to regulating phagocytosis in macrophages. This is the first study of identification and structural characterization of lectins in these pathogenic fungi of medical importance, which allows greater knowledge and clarification on the pathogenicity process of human pathogenic fungi and their interactions with their hosts.

Keywords: Pathogenic fungi. Lectins. Dynamic. Modeling. Phenetic.

LISTA DE ILUSTRAÇÕES

Figura 1 - Primeiros eventos na patogênese de fungos na superfície da mucosa	20
Figura 2 - Esquema da via secretora de lectinas L-type como VIP36	22
Figura 3 - Classificação estrutural das lectinas	23
Figura 4 - Diferentes estruturas e associações entre proteínas de membrana na bicamada lipídica	25
Figura 5 - Representação esquemática da membrana e parede celular fúngica	29
Figura 6 - Principais lectinas com estrutura 3D resolvida em fungos e leveduras	30
Figura 7 - Representação de algumas das principais áreas da bioinformática	33
Figura 8 - Ensaios de hemaglutinação e de inibição com açúcares	37
Figura 9 - Estrutura dos domínios lectina em <i>Aspergillus fumigatus</i>	50
Figura 10 - Estrutura dos domínios lectina em <i>Candida albicans</i>	51
Figura 11 - Estrutura dos domínios lectina em <i>Cryptococcus neoformans</i>	52
Figura 12 - Gráficos com as médias preferenciais de p.I. e P.M. por família proteica para <i>A. fumigatus</i> , <i>C. albicans</i> e <i>C. neoformans</i>	52
Figura 13 - Análise <i>in silico</i> da família proteica FLO	57
Figura 14 - Análise <i>in silico</i> da família proteica L-type	58
Figura 15 - Análise <i>in silico</i> da família proteica CBM	60
Figura 16 - Análise <i>in silico</i> da família proteica C-type	64
Figura 17 - Análise <i>in silico</i> da família proteica P-type	66
Figura 18 - Análise <i>in silico</i> da família proteica R-type	67
Figura 19 - Resultados para o mapeamento de oito PHs pelo Blast2GO- em relação à função molecular	72
Figura 20 - Resultados para o mapeamento das nove PHs pelo Blast2GO- em relação à processo biológico	72
Figura 21 - Resultados para o mapeamento das nove PHs pelo Blast2GO- em relação à componente celular	72
Figura 22 - Resultados da anotação após a análise de expansão de anotação	74
Figura 23 - Principais etapas na análise <i>in vitro</i> das sequências com domínio	75
Figura 24 - Modelos tridimensionais de AfLtype1	76
Figura 25 - Gráficos relacionados à simulação em dinâmica molecular do modelo	78

tridimensional de AfLtype1

Figura 26 - Sequências de aminoácidos da proteína putativa AfLtype1 81

LISTA DE TABELAS

Tabela 1 - Principais aplicações biotecnológicas das lectinas como ferramenta nas ciências básicas e na medicina	26
Tabela 2 - Alguns programas e servidores da Internet úteis em análises <i>in silico</i>	34
Tabela 3 - Palavras-chave utilizadas na pesquisa de sondas	42
Tabela 4 - Iniciadores para a reação de PCR em tempo real e genes de referência utilizados no presente estudo	46
Tabela 5 - Número de sequências sonda coletadas no UniProt, CGD e AspGD, candidatos com domínio de interesse coletados via BLAST-NCBI e domínios confirmados nos genomas de interesse no CD-Search	49
Tabela 6 - Principais características físico-químicas das sequências proteicas coletadas nos bancos de dados NCBI, CGD e AspGD	54

LISTA DE ABREVIATURAS E SIGLAS

AAL – *Aleuria aurantia* Lectin

AFL – Lectina de *Aspergillus fumigatus*

AfLtype1 – Lectina como de leguminosas em *Aspergillus fumigatus*

AspGD – Aspergillus Genome Database

BLAST – Basic Local Alignment Search Tool (Ferramenta para Pesquisa Básica de Alinhamento Local)

BSA – Bovine Serum Albumine (Albumina Bovina Sérica)

CBM – Módulos que se ligam a carboidratos

CD-Search – Pesquisa por domínios conservados

CGD – *Candida* Genome Database

CRD – Domínios de Reconhecimento de Carboidratos

C-type – Lectina tipo Concanavalina-A

DNA – Ácido Desoxirribonucleico

EDTA – Ácido etileno-diamino-tetracético

EPA – Adesinas epiteliais de superfície

FASTA – Fast Alignment Search Proteins and Nucleotides sequences

FleA – Lectina com especificidade a fucose de *Aspergillus fumigatus*

FLO – Floculinas

F-type – Lectina com especificidade a fucose

GalNAc – N-Acetilgalactosamina

G-type – Lectina com especificidade a galactose

I-type – Lectina do tipo imunoglobulinas

kDa – kiloDalton

L-type – Lectina de leguminosas

MEGA7 - Molecular Evolutionary Genetics Analysis versão 7.0

MEME – Múltiplos EM (estimativa de máxima verossimilhança) para elicitación de motivos

mRNA – RNA mensageiro

ORF – Open Reading Frame (Quadro Aberto de Leitura)

pb – Pares de bases de nucleotídeos

PBS – Tampão fosfato salina

PCR – Reação em Cadeia da Polimerase

PDB – Protein Data Bank

Pfam – Base de dados de famílias de proteínas

PH – Proteína Hipotética

pI – Ponto Isoelétrico

PM – Peso Molecular

P-type – Lectinas tipo “P”

R-type – Lectina com especificidade a ricina

SDS – Dodecil Sulfato de Sódio

SIGLECS – Proteínas da superfície celular que se ligam ao ácido siálico

UniProt – Universal Protein Data Bank (Banco de dados universal de proteínas)

YPD – Meio de cultura contendo extrato de levedura, peptona e dextrose

SUMÁRIO

1 INTRODUÇÃO	17
2 FUNDAMENTAÇÃO TEÓRICA	20
2.1 Adesinas e lectinas	20
2.2 Classificação das lectinas	21
2.3 Aplicações biotecnológicas das lectinas	25
2.4 Fungos de importância médica	27
2.5 Lectinas de fungos patogênicos	29
2.6 Estudo <i>in silico</i> : Bioinformática	32
2.7 Aplicação da Bioinformática no estudo de proteínas	34
2.8 Técnicas aplicadas ao estudo de lectinas	36
2.9 Modelagem molecular de proteínas	38
3 OBJETIVOS	41
3.1 Objetivo geral	41
3.2 Objetivos específicos	41
4 PROCEDIMENTOS METODOLÓGICOS	42
4.1 Análise bioinformática	42
4.2 Alinhamento, análise de domínios, motivos e fenética	42
4.3 Análises de anotação de ontologia pelo Blast2GO	43
4.4 Material biológico	44
4.5 Extração e quantificação de RNA total e síntese de cDNA	44
4.6 RT-PCR, qPCR e Sequenciamento	44
4.7 Modelagem e dinâmica	46
PIPELINE	48
5 ANÁLISE DOS RESULTADOS	49
5.1 Seleção e caracterização das sequências com domínio lectina	49
5.2 Análise de domínios e fenética	57
5.3 Análises de anotação de ontologia pelo Blast2GO	71
5.4 Análise <i>in vitro</i> das sequências	74
5.5 Modelagem e dinâmica	76
6 DISCUSSÃO	83

7 CONCLUSÃO	90
REFERÊNCIAS	91

1 INTRODUÇÃO

Lectinas são glicoproteínas capazes de se ligarem a carboidratos ou a substâncias que contenham açúcares, tais como glicoconjugados, podendo ser detectadas a partir de ensaios de aglutinação de eritrócitos, nos quais interagem com carboidratos da superfície celular através de seus sítios, formando diversas ligações reversíveis sem modificar seus ligantes. São moléculas ubíquas, presentes em todos os seres vivos e se ligam com alta especificidade a glicoconjugados complexos, por isso possuem um papel fundamental em vários processos de reconhecimento molecular e sinalização celular, apresentando uma variedade de efeitos biológicos, as quais servem de base para a sua aplicação na investigação de atividades químicas e biológicas (AMBROSI et al., 2005).

Várias moléculas que se comportam como adesinas têm sido estudadas em fungos patogênicos, visto que, a adesão de microrganismos às proteínas da membrana extracelular do hospedeiro é provavelmente o primeiro passo para ocorrer o processo de infecção. Contudo, de acordo com os últimos dados reportados, pouco se sabe sobre lectinas de fungos, mas foram descritas como sendo 82% isoladas de cogumelos, 15% de microfungos (mofo) e somente 3% de leveduras (VARROT et al., 2013). Ainda que elas estejam presentes em todas as espécies, as lectinas fúngicas ainda não são objeto de estudo apropriado, e por isto a função exata de muitas delas não é clara (KHAN E KHAN, 2011). Análises *in silico* tem sido possível porque nos últimos anos a bioinformática tem sido uma grande aliada na análise da função de genes e proteínas, e pode contribuir de forma significativa para melhorar a determinação da função dos produtos gênicos (VATANSEVER et al., 2015).

As primeiras lectinas de fungos foram isoladas de corpos frutíferos de duas espécies de cogumelo do gênero *Agaricus campestris* (SAGE e VÁZQUEZ, 1967) e *Agaricus bisporus* (PRESANT e KORNFELD, 1972). Embora um estudo recente indique a caracterização de uma lectina em *Candida glabrata*, ainda não se tem conhecimento de lectinas de *Candida albicans* descritas na literatura (VARROT et al., 2013; GALLEGOS et al., 2014). Algumas lectinas e suas especificidades por certos carboidratos já foram relatadas para algumas espécies de *Aspergillus*, três delas para *A. fumigatus*, uma de ligação ao ácido siálico, uma lectina fucospecífica (FleA) e uma lectina AFL (Lectina de *Aspergillus fumigatus*), descrita como um novo membro da família das lectinas AAL e a

primeira lectina fúngica com estrutura 3D descrita (WIMMEROVA et al., 2003; HOUSER et al., 2015).

De modo geral, as lectinas de fungos oportunistas, podem desempenhar um papel importante em sua interação com o hospedeiro humano, funcionando da mesma forma que em bactérias e vírus, intermediando a interação com o hospedeiro e, assim, possibilitando a invasão. Um estudo estrutural detalhado da AFL, por exemplo, revelou uma alta afinidade em relação aos oligossacarídeos que foram detectados em plantas e também aqueles ligados ao epitélio humano. Estes resultados indicam que a AFL seja membro da família das lectinas e um alvo promissor para o diagnóstico e tratamento da aspergilose e, portanto, útil tanto na biotecnologia quanto na medicina (HOUSER et al., 2015).

Os gêneros responsáveis por mais de 90% de todas as mortes relatadas por fungos são *Cryptococcus*, *Candida*, *Aspergillus* e *Pneumocystis* (BROWN et al., 2012). As espécies de *Candida* e *Aspergillus* são as principais responsáveis por um número significativo de casos de infecções em algumas partes do mundo, e embora não tão frequentes, a incidência de infecção por espécies de *Cryptococcus* aumentou nos últimos anos. A mortalidade por doenças invasivas causadas por estes fungos permanece alta, apesar do aumento no número de antifúngicos disponíveis para terapia no mercado, e isto se deve principalmente a questões de saúde pública e à resistência a certos medicamentos, carecendo de novas abordagens de prevenção e controle a doenças fúngicas (KIBBLER et al., 2018).

Esse aumento da prevalência de infecções fúngicas motivou o estudo das interações hospedeiro-patógeno, a fim de entender os mecanismos responsáveis pelas respostas imunes antifúngicas no corpo humano e como estas leveduras interferem com os mecanismos de defesa do hospedeiro devido à capacidade que têm em evadir-se do sistema imune. Diante disso, a hipótese deste projeto de doutorado é que adesinas do tipo lectinas expressas durante a fase de diferenciação de *C. albicans* (da forma levedura para a forma filamentosa) ou nos fungos *A. fumigatus* e *C. neoformans*, possam ser essenciais para o reconhecimento de receptores do epitélio humano e, portanto, participariam diretamente do processo de virulência e invasão dos tecidos do hospedeiro (ÁLVARES et al., 2007). Diante da hipótese acima, este trabalho de doutorado visa identificar e caracterizar novas proteínas fúngicas com domínio lectina em uma análise *in silico*, verificar suas respectivas expressões gênicas, bem como trabalhar modelagem comparativa em pelo menos uma das

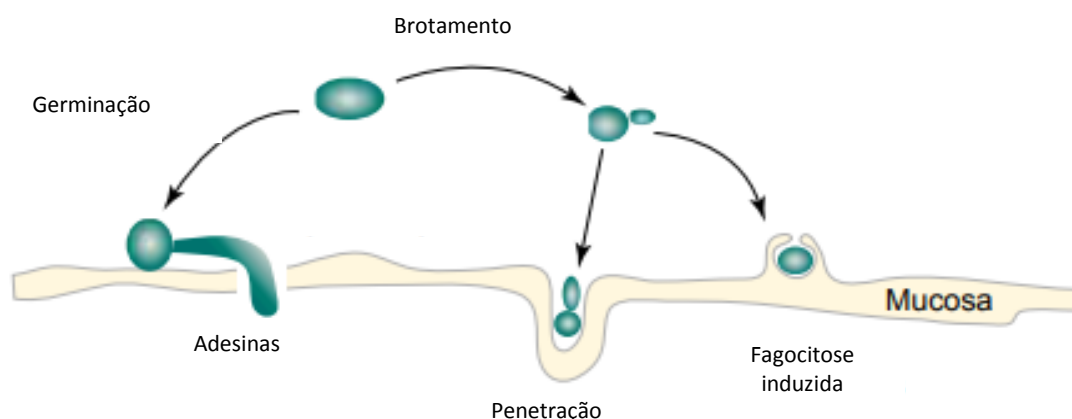
proteínas não caracterizadas encontradas nos fungos analisados, visando uma melhor compreensão do processo de virulência destes patógenos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 ADESINAS E LECTINAS

Por definição, as adesinas são proteínas que reconhecem e se ligam a glicanos específicos da superfície celular do hospedeiro. Em fungos, as adesinas envolvidas no processo de infecção medeiam a ligação da levedura à superfície celular. Este evento é pré-requisito para colonização e infecção de fungos patogênicos, contribuindo para que suas células interajam intimamente com as células do hospedeiro, facilitando a germinação das células de levedura, bem como sua fixação e permitindo a fagocitose induzida por uma célula da mucosa (CALDERONE e FONZI, 2001) (**Figura 1**). Sendo assim, lectinas são adesinas que reconhecem e se ligam a glicanos na superfície das células hospedeiras via domínios de reconhecimento de carboidratos (CRDs) (NATHER e MUNRO, 2008).

Figura 1. Primeiros eventos na patogênese de fungos na superfície da mucosa.



Fonte: Adaptado de CALDERONE e FONZI, 2001.

As lectinas estão envolvidas em muitos processos essenciais de reconhecimento celular e molecular. Como as encontradas em animais, algumas lectinas de microrganismos se ligam aos resíduos de açúcar através dos CRDs, enquanto outras se ligam a sequências internas encontradas em cadeias oligossacarídicas. A interação das lectinas com os glicanos das células do hospedeiro é um importante determinante do tropismo do patógeno (OFEK e SHARON, 1990). Muitos microrganismos exploram glicanos da superfície celular como alvos para a interação com outras células, sendo que proteínas presentes em sua superfície (adesinas ou aglutininas) medeiam a ligação a esses receptores de glicano e um grande número de espécies patogênicas dependem dessas interações para que ocorra a infecção. Estudos detalhados da especificidade dessas lectinas microbianas levaram à

identificação e síntese de inibidores de adesão que pode formar a base para novos agentes terapêuticos para combater doenças infecciosas (OFEK e SHARON, 1990; NIZET et al., 2018).

Por décadas, os estudos eram em sua grande maioria voltados principalmente para aglutininas vegetais, provavelmente por causa de sua disponibilidade. No entanto, ao longo do século 19, vários relatos de atividades de aglutinação bacteriana e de células animais foram publicados (LAM e NG, 2011). Após a descoberta de que estas aglutininas poderiam distinguir especificamente entre vários grupos sanguíneos humanos, o novo termo lectina foi estabelecido e hoje é utilizado para todas as proteínas de ligação de origem não imune (ANGELONI et al., 2005). Embora já se tenha diversos estudos relacionados a lectinas animais, vegetais e bacterianas, ainda são poucos os estudos voltados para a identificação e caracterização de lectinas em fungos, sendo a primeira lectina fúngica isolada do cogumelo comum *Agaricus bisporus* apenas em 1980 (SUEYOSHI et al., 1985).

2.2 CLASSIFICAÇÃO DAS LECTINAS

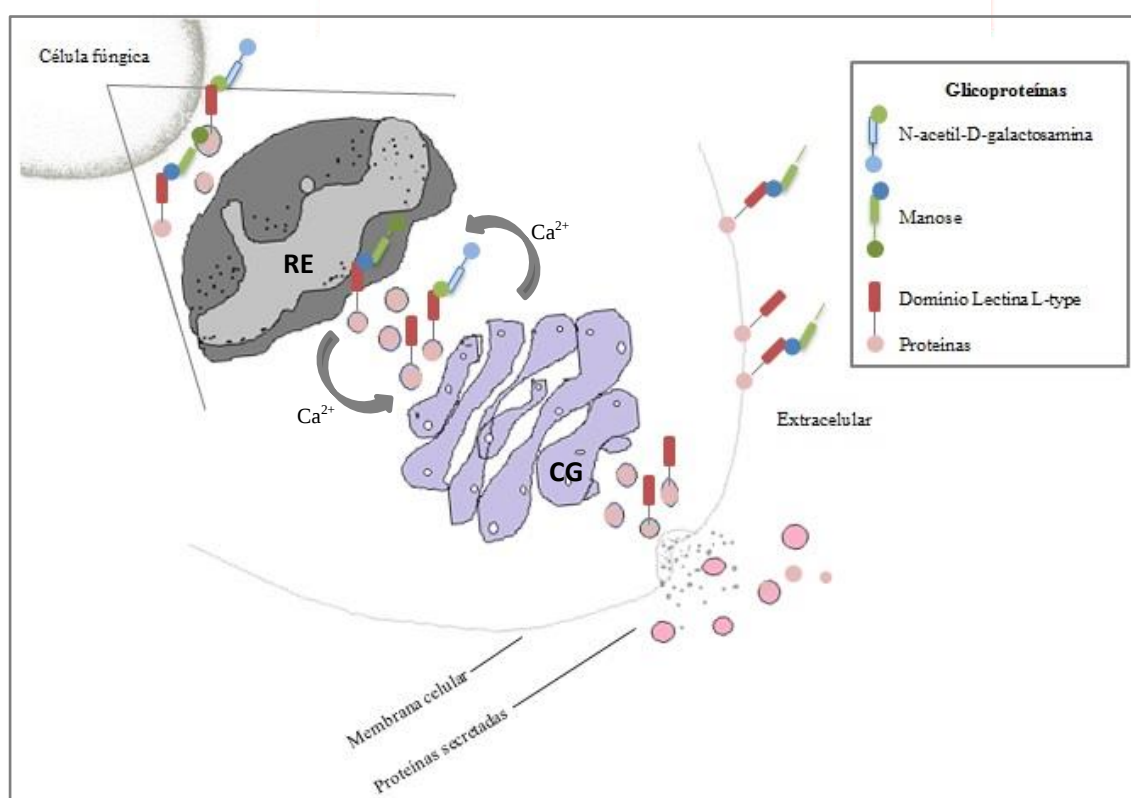
As lectinas podem ser classificadas de acordo com suas propriedades. A forma mais comum e usada até hoje é baseada no organismo de origem, podendo ser derivada de planta, bactérias, animais, fungos e lectinas isoladas de vírus. No entanto, isso não diz nada sobre a estrutura ou função da lectina e por isso costuma-se também classificar as lectinas de acordo com o carboidrato pelo qual a proteína tem afinidade (D-galactose ou N-acetil-galactosamina, D-manose, L-fucose, N-acetil-glucosamina, ácido siálico, entre outros). Mesmo assim, essa classificação não é suficiente, já que existem lectinas que reconhecem mais de um carboidrato em uma cadeia de glicoproteína (LAM e NG, 2011).

Alguns grupos de lectinas foram classificados com base na função proposta e/ou localização da lectina no organismo. Assim, as lectinas do tipo C estão presentes na superfície celular e necessitam de íons Ca^{2+} para se ligarem ao açúcar de afinidade, enquanto que as SIGLECS são proteínas tipo imunoglobulinas (lectinas I-type) localizadas na superfície celular e que interferem na sinalização, inibindo a ativação das células imunológicas (MACAULEY et al., 2014; BORNHOFFT et al., 2018).

Os domínios de reconhecimento de carboidratos (CDRs) são regiões proteicas inseridas na cadeia polipeptídica da lectina e que se enquadra em várias famílias estruturalmente distintas. As lectinas intracelulares (famílias F-type, M-type, L-type e P-

type, entre outras) estão localizadas nos compartimentos luminais da via secretora e funcionam no tráfego, classificação e direcionamento de glicoproteínas (**Figura 2**). Lectinas L-type incluem as VIP36 e ERGIC-53. Essas duas proteínas foram os primeiros membros da família das lectinas animais semelhantes às lectinas das plantas leguminosas relatadas (KAMIYA et al., 2008). As lectinas extracelulares (C-type, R-type, I-type-siglecs, ficolinas, quitinases, intelectinas e galectinas, entre outras) são secretadas na matriz extracelular ou localizadas na membrana plasmática e medeiam uma variedade de funções, incluindo adesão celular, sinalização celular, liberação de glicoproteínas e reconhecimento de patógenos (LAKHTIN et al., 2011; MACAULEY et al., 2014; BORNHOFFT et al., 2018).

Figura 2. Esquema da via secretora de lectinas L-type como VIP36. Proteínas como VIP36 localizam-se no Complexo de Golgi (CG) e Retículo Endoplasmático (RE) e circulam na via secretora transportando carboidratos e glicoproteínas.

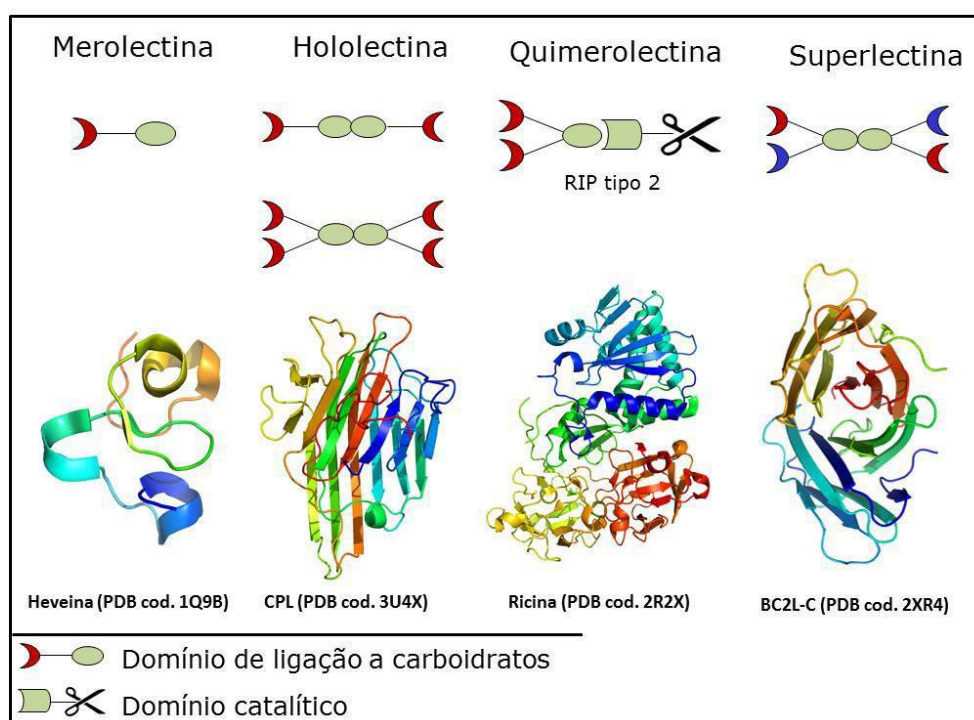


Fonte: Própria autora, 2020.

No entanto, para muitas lectinas a classificação através da função ou localização celular não é possível, já que a função é desconhecida ou apenas prevista, e estas são chamadas proteínas hipotéticas (PHs). Provavelmente a melhor maneira de classificar as

lectinas seja de acordo com sua estrutura 3D e caracterização dos seus motivos estruturais de ligação de açúcar. Esta classificação é possível graças ao desenvolvimento de ferramentas de busca por similaridade de sequências tanto de aminoácidos quanto de nucleotídeos (ALBERTS et al., 2010). De uma forma geral, as lectinas foram subdivididas em quatro classes distintas, com base em suas estruturas globais e mostradas na **figura 3**:

Figura 3. Classificação estrutural das lectinas.



Fonte: SHARON e LIS (2004).

Sendo assim, as merolectinas são proteínas monoméricas e compostas de apenas um sítio de ligação a carboidrato por estrutura. Por isso, não apresentam atividade hemaglutinante ou capacidade de precipitar glicoconjugados, sendo estruturalmente, as mais simples, como podemos encontrar nas heveínas, lectinas presentes em *Hevea brasiliensis* (VAN DAMME et al., 1998).

Já as hololectinas possuem dois ou mais domínios idênticos ou homólogos que reconhecem a carboidratos. Neste caso são geralmente proteínas oligoméricas (dímeros ou tetrâmeros) e possuem a capacidade de aglutinar células e precipitar glicoconjugados. Típico das lectinas de leguminosas, dentre as quais podemos citar as lectinas mais bem conhecidas

desse grupo, presentes em *Canavalia brasiliensis* (ConBr) (SANZ-APARICIO et al., 1997; CAVADA et al., 2001) e *Artocarpus integrifolia* (Jacalina) (VAN DAMME et al., 1998).

Nas quimerolectinas por sua vez, as lectinas possuem pelo menos um único sítio de ligação a carboidratos, bem como outro domínio que exerce uma função distinta de reconhecimento a açúcar, ou seja, um domínio catalítico. Neste grupo geralmente são encontradas proteínas com mais de uma função, complementares entre si, e que podem se comportar como merolectinas ou hololectinas, dependendo do número de sítios ligantes. A ricina, lectina da espécie *Ricinus communis*, é típica desse grupo (VAN DAMME et al., 1998). Outro exemplo são os módulos que se ligam à carboidratos (CBMs) são definidas como uma sequência de aminoácidos dentro de uma enzima ativa e que contém uma dobra com atividade de ligação a carboidratos.

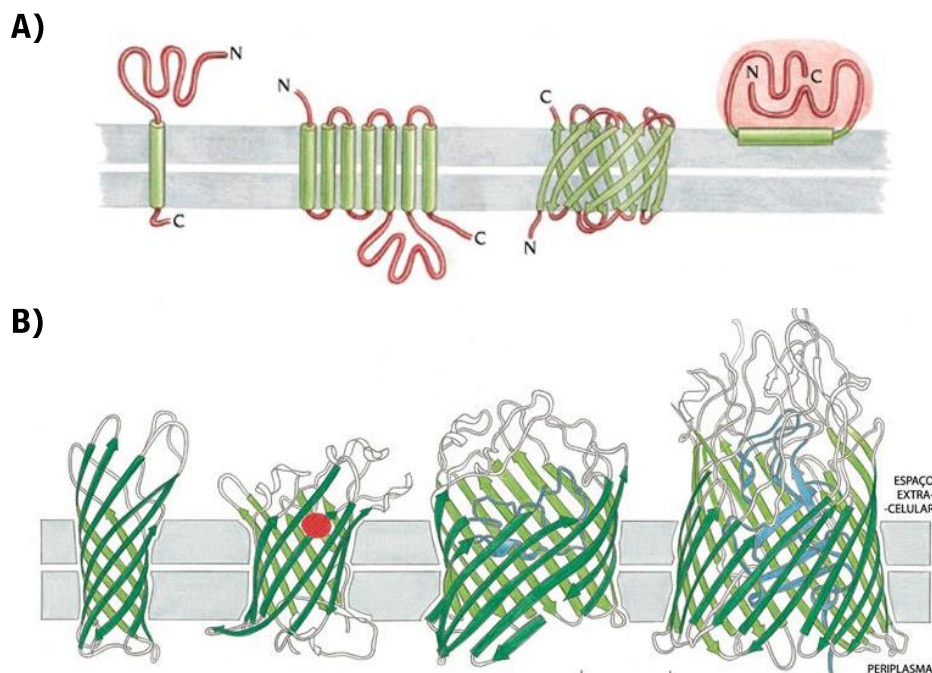
Já para as superlectinas, essas lectinas possuem pelo menos dois sítios de ligação a carboidratos, assim como as hololectinas, porém com especificidade a açúcares distintos. Além disso, são menos encontradas e caracterizadas do que as demais. Podemos citar a BC2L-C, uma lectina oriunda do patógeno oportunista *Burkholderia cenocepacia* e a lectina da espécie de bananeira *Musa acuminata* (VAN DAMME et al., 1998; SULAK et al., 2011).

Entretanto, vale salientar que algumas lectinas de plantas hoje descobertas e estruturalmente resolvidas, não se enquadram especificamente em nenhuma destas classificações. Como exemplo podemos citar a lectina presente na semente de *Parkia platycephala*, homóloga a família das hidrolases, que possui um domínio com capacidade de reagir enzimaticamente com o polissacarídeo composto de monômeros de N-acetyl-D-glucosamina (quitina) e um outro sítio de reconhecimento a carboidrato (CAVADA et al., 2006).

Apesar das similaridades estruturais nos níveis primário, secundário e terciário (motivo beta barril), as lectinas possuem diferenças em suas respectivas estruturas que permitem uma melhor compreensão do papel das sequências e estrutura na oligomerização de proteínas. Por elas terem modos muito diferentes de associações, discutidas acima, estas interações monômero-monômero interferem no modo como estas proteínas se complexam com as superfícies celulares (**Figura 4**) e às matrizes de glicoconjugados, resultando nas

suas respectivas e variadas especificidades e, conseqüentemente, suas diversificadas atividades biológicas (SHARON; LIS, 2002, 2004; BRINDA et al., 2005).

Figura 4. Diferentes estruturas e associações entre proteínas de membrana na bicamada lipídica. Em **A**, as interações monômero-monômero determinam o modo como as proteínas se complexam as superfícies celulares e em **B**, Barris β formados por diferentes números de fitas β .



Fonte: Adaptado de GARLAND, 1999.

2.3 APLICAÇÕES BIOTECNOLÓGICAS DAS LECTINAS

A capacidade das lectinas ligarem-se a carboidratos específicos e induzir processos particulares em células animais são amplamente exploradas em pesquisas biológicas e biomédicas. De uma forma geral as lectinas apresentam uma variedade de efeitos biológicos, as quais servem de base para a sua aplicação na investigação de atividades químicas e biológicas (DAN et al., 2016). As interações carboidrato-lectina permitem o reconhecimento seletivo em vários processos biológicos, os quais estão envolvidos em muitas patologias humanas, como metástase de câncer (YAU et al., 2015), invasão viral (SWANSON et al., 2010; HOPPER et al., 2017) distúrbios patológicos relacionados à microbiota e inflamações (MUSZYNSKA et al., 2018). As diversas atividades biológicas descritas para as lectinas, tais como, antitumoral (YAU et al., 2015; NASCIMENTO et al., 2019), antimicrobiana (SITOHY et al., 2007; HASAN et al., 2014; DIAS et al., 2015), anti-HIV (SWANSON et al., 2010; HOPPER et al., 2017), mitogênica (SINGH et al.,

2015) e anti-inflamatória (SATO et al., 2015; BARI et al., 2016), as tornam instrumentos potenciais na biotecnologia.

Como as lectinas não modificam a estrutura de seus ligantes, elas são uma ferramenta valiosa para separar glicanos ou glicoproteínas de suas misturas para análise ou uso posterior, através da cromatografia de afinidade, uma das etapas fundamentais para purificação de lectinas (GONÇALVES et al., 2017; BOI, 2019). Em muitos estudos, as lectinas têm sido avaliadas como moléculas promissoras para diversas aplicações biotecnológicas, exemplificadas na **Tabela 1**.

Tabela 1. Principais aplicações biotecnológicas das lectinas como ferramenta nas ciências básicas e na medicina

REF.

Bioquímica	Detecção de epítomos definidos para glicanos de glicoconjugados, em blots ou por cromatografia em camada fina;	RUDIGER e GABIUS (2001)
	Purificação de lectinas associadas a glicoconjugados, por cromatografia de afinidade;	
	Caracterização de glicanos por cromatografias de afinidade (lectinas) seriadas;	
	Análise glicômica e glicoproteômica;	
	Quantificação de lectinas associadas a glicoconjugados reativos por ELLA (“enzyme-linked lectin-binding assays”);	
	Quantificação de atividades de glicosiltransferases/glicosidases por lectinas, baseadas na detecção de produtos formados pela reação enzimática.	
Biologia Celular	Análise dos mecanismos envolvidos na glicosilação correta, mediante lectinas resistentes a variantes celulares;	
	Fracionamento de populações de células;	
	Substrato modelo para o estudo da agregação e adesão celular.	
Medicina	Detecção de doenças relacionadas com alterações de síntese dos glicanos;	
	Tipagem dos grupos sanguíneos;	
	Quantificação de aberrações dos glicanos da superfície celular (malignidade);	
	Marcador celular, incluindo agentes infecciosos, com o objetivo de diagnóstico (vírus, bactérias, fungos e parasitas).	

Fonte: adaptado de RUDIGER e GABIUS (2001).

Entre as lectinas de animais e plantas, a função mais atribuída é o seu papel na defesa contra patógenos (BLEULER-MARTÍNEZ et al., 2011). Já as lectinas de microrganismos estão muitas vezes relacionadas à interação inicial do patógeno com o tecido do hospedeiro ou desempenham papel na formação de biofilme. Esta capacidade de se aglomerar e formar biofilme estão estritamente ligadas a patogenicidade de um fungo, tornando-o mais patogênico e protegido contra as defesas do organismo e da ação dos antibióticos (OFEK et al., 2003).

Lectinas específicas têm sido comumente conjugadas a drogas com ação anticâncer e usadas com finalidades antitumorais, sem prejudicar os tecidos saudáveis (LI et al., 2007; TINBERG et al., 2013; HE et al., 2014). Uma lectina específica para GlcNAc, a AAL2, isolada a partir do fungo *Agrocybe aegerita*, demonstrou atividade antitumoral em células de hepatocarcinoma e potencial como ferramenta para o diagnóstico de câncer (HIRABAYASHI et al., 2013).

Lectinas fúngicas de cogumelos e líquens são utilizadas em laboratório de pesquisa para estudos glicômicos, e microarranjos de lectinas são ferramentas utilizadas para criação de perfis de alta sensibilidade de glicomas (JIANG et al., 2012). Portanto, a identificação efetiva da seletividade das lectinas aos carboidratos é significativa não apenas para decifrar as glicoproteínas envolvidas nos processos, mas também possibilita o desenvolvimento de testes úteis para o diagnóstico e a terapia das doenças (VARROT et al., 2013; WANG e HE, 2018).

2.4 FUNGOS DE IMPORTÂNCIA MÉDICA

As infecções causadas por fungos oportunistas têm aumentado significativamente nas últimas décadas, tornando-se importante entender como estas leveduras interferem com os mecanismos de defesa do hospedeiro devido à capacidade que têm em evadir-se do sistema imune. Alguns fungos são considerados oportunistas por afetarem apenas pessoas imunodeprimidas, enquanto que outros são considerados patogênicos por causarem doença independente do estado do sistema imunológico do hospedeiro. A maioria das espécies de *Candida*, por exemplo, são membros da microbiota, mas também causam infecções superficiais ou invasivas, sendo em sua grande maioria causadas por *C. albicans* (ÁLVARES et al., 2007; KHAN et al., 2010; KIBBLER et al., 2018).

C. albicans é polimórfica e pode diferenciar-se da forma unicelular (levedura) para a forma filamentosa (hifas ou pseudo-hifas) (KULETA; KOZIK; KOZIK, 2009). Uma microbiota desequilibrada bem como imunidade prejudicada favorecem infecções superficiais, e com o distúrbio das barreiras das mucosas juntamente com a imunidade comprometida, a *Candida* invade a corrente sanguínea humana causando infecção sistêmica. Mesmo com terapia antifúngica, a candidíase ainda tem uma alta mortalidade (40-50%). Entender o processo de diferenciação celular de *C. albicans* é fundamental para uma maior compreensão da interação do fungo com o hospedeiro, pois no processo de

mudança, as hifas têm maior capacidade de aderir e penetrar nas células epiteliais humanas do que os blastoconídios (SANTANA et al., 2013).

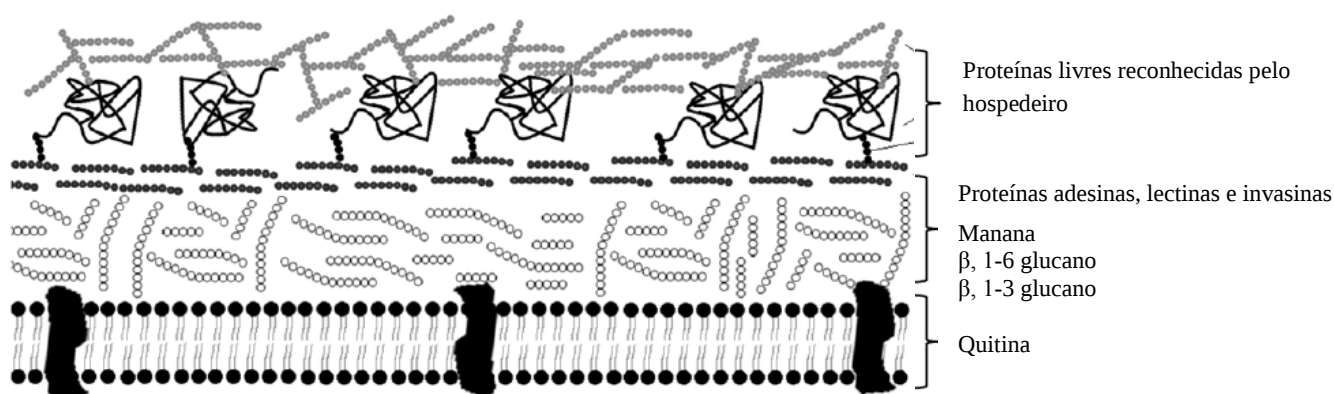
Espécies do gênero *Aspergillus* são também consideradas fungos oportunistas, isso devido ao fato de possuírem conídios bem pequenos, o que facilita serem transportados pelo ar. Se estes conídios forem inalados, podem facilmente entrar no trato respiratório e no caso de um hospedeiro imunossuprimido, causar a germinação e crescimento da hifa, causando então, a aspergilose. Mais de 20% das espécies de *Aspergillus* são conhecidas por serem patogênicas aos humanos. Além disso, algumas espécies também são comercialmente importantes por produzirem um grande número de enzimas extracelulares, como as amilases, muito aplicada na biotecnologia, como é o caso de *Aspergillus nidulans*, (SOARES et al., 2010). As doenças causadas por espécies de *Aspergillus* podem variar amplamente, desde a colonização superficial a doenças invasivas com alta taxa de mortalidade, apesar de um aumento no número de antifúngicos disponíveis para terapia (KIBBLER et al., 2018).

Outra levedura patogênica, *Cryptococcus neoformans*, agente causador da criptococose, é a causa mais comum de doença fúngica fatal em pessoas imunossuprimidas, causando meningoencefalite com risco de morte em aproximadamente 1 milhão de pessoas por ano no mundo todo. No entanto, essa espécie também acomete indivíduos sem problemas de saúde em todo o mundo (KIBBLER et al., 2018). Além disso, possui disponibilidade limitada de antifúngicos eficazes o que contribui para uma taxa de mortalidade alta. A infecção por *C. neoformans* ocorre pela inalação de seus esporos derivado das fezes de aves, principalmente dos pombos (CASADEVALL e PERFECT, 1998; PERFECT, 2005; PARK et al., 2009). *C. neoformans* é um patógeno intracelular facultativo que pode sobreviver dentro de macrófagos, e existe uma correlação positiva entre fagocitose e virulência. Vários estudos demonstram que o conhecimento dos componentes da superfície celular de *C. neoformans* é essencial para entender sua patogênese (PARK et al., 2009; CASADEVALL e PERFECT, 1998; PERFECT, 2005; GHUFRAN et al., 2017; WALSH et al., 2017).

O *C. neoformans* pode também acometer indivíduos sem evidência de imunodepressão aparente, em regiões tropicais e subtropicais, especialmente nas regiões Norte (Amazônia) e Nordeste do Brasil. No entanto, podem ocorrer focos nas demais regiões brasileiras (Disponível em www.saude.gov.br, acessado em 01/01/2020).

A parede celular, e especialmente suas proteínas, tem múltiplas funções na patogênese destes fungos. Em geral, são responsáveis pela relação entre as células fúngicas e as do hospedeiro, contribuindo para adesão ao tecido e para modular a resposta imune do mesmo. Por isso, um melhor entendimento das lectinas da parede celular dos fungos, pode elucidar muitos papéis patológicos que atualmente não são claros no processo de adesão celular, uma vez que elas estão estritamente ligadas com a interação entre a parede celular da levedura e os receptores do tecido do hospedeiro (KLIS et al., 2002; RUIZ-HERRERA et al., 2006; WEINDL, et al, 2010; FRANÇOIS et al, 2013; GOW et al., 2017) (**Figura 5**).

Figura 5. Representação esquemática da membrana e parede celular fúngica.



Fonte: Adaptado de Santos et al. (2012).

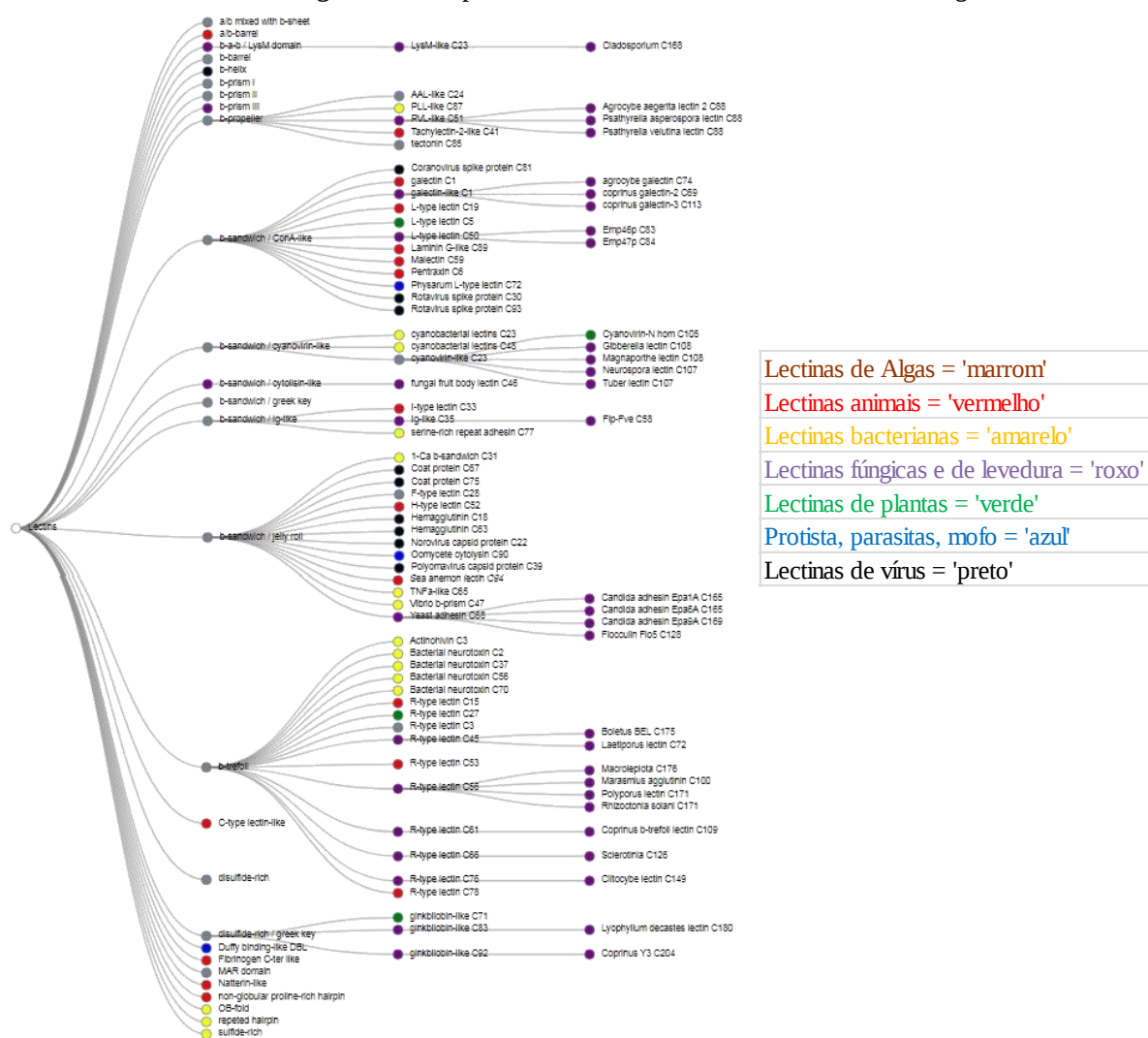
2.5 LECTINAS DE FUNGOS PATOGÊNICOS

As primeiras lectinas de fungos foram isoladas de corpos frutíferos de duas espécies de cogumelo do gênero *Agaricus campestris* (SAGE e VÁZQUEZ, 1967) e *Agaricus bisporus* (PRESANT e KORNFELD, 1972). O aumento do número de genomas de fungos sequenciados também contribuiu para a identificação de muitas outras lectinas fúngicas e desde então, outras lectinas foram isoladas e caracterizadas de outras espécies de fungos. Mesmo assim, as lectinas fúngicas ainda não são adequadamente estudadas, e por isso, a função exata de muitas delas não é clara. Muitos sugerem que as lectinas fúngicas desempenham papéis em processos metabólicos, relacionados com o crescimento e diferenciação do fungo, bem como nas associações entre raízes e fungos de solo chamadas de micorrizas (GUILLOT e KONSKA, 1997; KHAN e KHAN, 2011).

Atualmente existem 220 lectinas fúngicas com suas estruturas 3D resolvidas de acordo com o banco de dados da UniLectin, representando 9 % de todas as lectinas de organismos vivos e são mostradas na **figura 6** (BONNARDEL et al., 2019). A lectina

fúngica atualmente mais bem estudada e amplamente utilizada é a lectina AAL de *Aleuria aurantia* específica para fucose. Muitos estudos foram realizados com esta proteína, alguns deles sugerindo sua utilização promissora como um biomarcador no câncer, devido à sua ligação com a fucose. Ela também foi a primeira lectina fúngica com estrutura 3D descrita (WIMMEROVA et al., 2003).

Figura 6. Principais lectinas com estrutura 3D resolvida em fungos e leveduras.



Fonte: UniLectin (2020).

Muitas das lectinas animais mais pesquisadas reconhecem o β -glucano nas células fúngicas e induzem a resposta inflamatória para eliminar patógenos durante os estágios iniciais da infecção. Por outro lado, muitos patógenos usam lectinas como adesinas, para interagir com as células hospedeiras, como por exemplo, em *Candida glabrata* e *Candida albicans*, que abrigam uma variedade de adesinas epiteliais de superfície (EPA) para a adesão da célula hospedeira como também membros da família de genes ALS (sequência

semelhante à aglutininas) e HWP (McCALL et al., 2019). Recentemente, uma caracterização estrutural e funcional detalhada da família de adesinas EPA1 foi determinada para *C. glabrata*. O domínio de ligação da adesina epitelial EPA1 foi cristalizado, onde foi visto um motivo de ligação dependente de Ca^{2+} provando seu papel no evento de adesão. Já em *Paracoccidioides brasiliensis*, a paracocina liga-se à laminina da célula hospedeira e estimula a liberação do fator de necrose tumoral (TNF α) e óxido nítrico dos macrófagos (WEIS e DRICKAMER, 1996).

Algumas lectinas e suas especificidades por certos carboidratos já foram relatadas para algumas espécies do gênero *Aspergillus*. Para *A. fumigatus*, por exemplo, três lectinas já foram descritas, sendo uma de ligação ao ácido siálico com peso molecular de 32 kDa e que está presente na superfície celular, apresentando grande importância no reconhecimento e fixação dos conídios em células epiteliais humana (ISHIMARU et al., 1996). Outra lectina também específica para *A. fumigatus* é a FleA, que apresenta especificidade por fucose e teve sua estrutura proteica detalhada, bem como sua ação anti-inflamatória (TRONCHIN et al., 2002). Além dessas, uma terceira lectina, a AFL (Lectina de *Aspergillus fumigatus*), foi recentemente descrita como um novo membro da família das lectinas AAL. Como lectina de um patógeno oportunista, desempenha um papel importante na interação do patógeno com o hospedeiro humano. Um estudo estrutural detalhado da AFL revelou uma alta afinidade em relação aos oligossacarídeos que foram detectados em plantas e também aqueles ligados ao epitélio humano. Estes resultados indicam que a AFL seja um membro complexo da família lectina sendo uma meta desafiadora para pesquisas médicas futuras e, devido às suas propriedades, uma ferramenta potencialmente útil em aplicações biotecnológicas específicas (HOUSER et al., 2015; SAKAI et al., 2018).

As adesinas ligadas à parede celular de leveduras estão muitas vezes relacionadas à aderência na matriz extracelular das mucosas e das células endoteliais, patogenicidade e a fatores de virulência (SUNDSTROM, 2002; FU et al., 2008). Exemplos de adesinas tipo lectinas incluem as proteínas da família FLO de *Saccharomyces cerevisiae* (exceto FLO11) (GUO et al., 2000), bem como as da família EPA no patógeno humano *C. glabrata* (CORMACK et al., 1999). Em *S. cerevisiae*, adesinas tipo lectina são divididas em duas subcategorias, FLO1 e NewFLO. O grupo FLO1 (FLO1, FLO5, FLO9 e FLO10) reconhece apenas açúcares do tipo manose, enquanto que o tipo NewFLO liga-se a vários açúcares, incluindo manose, glicose e maltose (SATO et al., 2002).

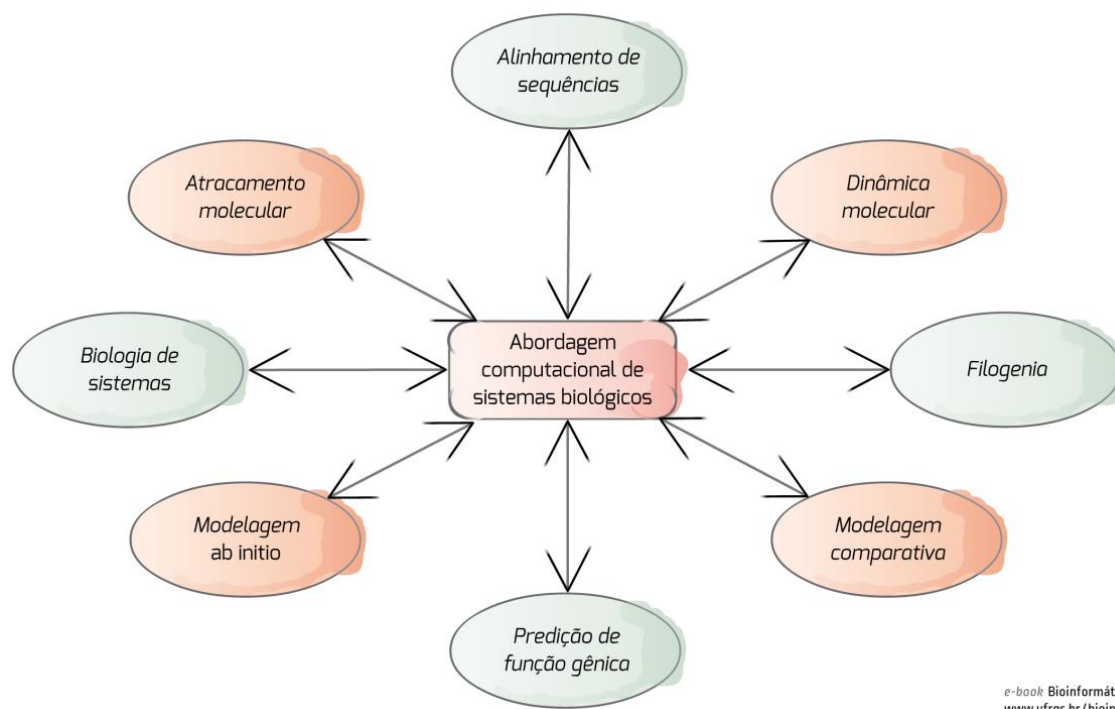
No caso das interações patógeno-hospedeiro, várias lectinas situadas na superfície da célula hospedeira (lectinas C-type, SIGLECS e Galectinas) desempenham papéis importantes no reconhecimento dos carboidratos à superfície de patógenos (SAKAI et al., 2018). Devido às múltiplas funções das diferentes classes de lectinas, o estudo de lectinas é uma estratégia promissora para moldar respostas referentes a infecções sistêmicas (LEPENIES e LANG, 2019).

2.6 ESTUDO *in silico*: BIOINFORMÁTICA

As primeiras tecnologias “ômicas” foram o sequenciador e o sintetizador de DNA desenvolvidos no início dos anos 90 como ferramentas para análise da expressão gênica (por exemplo, transcriptômica), e o sequenciador e sintetizador de proteínas para estudar a expressão de proteínas a nível celular, num processo conhecido como proteômica (HOOD et al., 2002). Além disso, o surgimento em paralelo de estudos metabolômicos completou o entendimento das vias metabólicas da síntese e processamento das informações biológicas a partir da expressão gênica, à síntese proteica e alterações nos metabólitos (OLIVER et al., 1998). Em geral, o objetivo das ciências ômicas têm sido identificar, caracterizar, e quantificar todas as moléculas biológicas envolvidas na estrutura, função e dinâmica de uma célula, tecido ou organismo (VAILATI-RIBONI et al., 2017).

Com o aprimoramento das técnicas de sequenciamento e das tecnologias associadas à leitura e interpretação de dados, o termo bioinformática foi lançado como uma nova área do conhecimento científico, representando a interação da biologia com a informática. Esta nova área passou a fazer parte de projetos nas áreas biológicas como forma de analisar grandes volumes de dados, possibilidade de armazenamento de informações em bancos de dados e apresentação de tais resultados em interfaces acessíveis via web, tornando a pesquisa mais interativa e dinâmica e com baixo custo (**Figura 7**). Uma definição mais ampla da bioinformática seria a aplicação de ferramentas de computação para análise, captura e interpretação de dados biológicos (DNA, RNA, proteínas, lipídios e metabólitos) (OTTO et al., 2007; SHIMIDT et al., 2014).

Figura 7. Representação de algumas das principais áreas da bioinformática



Fonte: Bioinformática da Biologia à flexibilidade molecular (2014).

A abordagem computacional tem sido cada vez mais indispensável no estudo e interpretação de dados nas ômicas. Na genômica ela tem auxiliado no estudo da estrutura, função, contribuindo de forma significativa para melhorar a determinação da função dos produtos gênicos, evolução e mapeamento dos genomas, bem como na caracterização e quantificação de genes (JANICKI et al., 2011).

Na transcriptômica, a bioinformática analisará o perfil de RNA mensageiro de uma célula, tecido ou organismo através de microarranjos e RNA-seq (DJEBAI et al., 2016). Já na proteômica, vai contribuir na caracterização *in silico* das proteínas de uma célula, tecido ou organismo, inferindo funções putativas a proteínas hipotéticas e analisando modificações pós-traducionais por meio de suas respectivas estruturas, bem como trabalhando na recuperação, comparação, alinhamento e análise filogenética de sequências (BAYAT, 2002; SCHMIDT et al., 2014; FENG et al., 2015; HAN et al., 2015; VATANSEVER et al., 2015). Por último a metabolômica que por sua vez complementa o estudo dos produtos finais dos processos celulares (metabólitos) fornecendo anotações putativas de metabólitos de alta qualidade (MEIER et al., 2017).

2.7 APLICAÇÕES DA BIOINFORMÁTICA NO ESTUDO DE PROTEÍNAS

Várias abordagens podem ser utilizadas para auxiliar na caracterização funcional de genes e proteínas. Dentre elas, podemos citar análises de similaridade, busca por domínios já caracterizados, predição da localização celular e análise do enriquecimento gênico em uma lista de genes expressos e mostradas de forma sucinta na **tabela 2** (VATANSEVER et al., 2015).

Tabela 2. Alguns programas e servidores da Internet úteis em análises *in silico*

Nome	Tipo*	Endereço na Internet
Bancos de dados		
GenBank	bd	www.ncbi.nlm.nih.gov/GenBank
AspGD	bd	www.aspgd.org/
CGD	bd	www.candidagenome.org/
PDB	bd	www.rcsb.org/pdb/
SWISS-PROT	bd	www.ebi.ac.uk/swissprot
Análise de sequências		
MOTIF	bd	www.genome.jp/tools/motif/
MEGA	p	www.megasoftware.net/
BLAST	bd	www.ncbi.nlm.nih.gov/BLAST/
FASTA	bd	www2.ebi.ac.uk/fasta3
MEME	bd	meme-suite.org/
CELL-PLOC	bd	www.csbio.sjtu.edu.cn/bioinf/Cell-PLOC/
PROFIT	p	www.bioinf.org.uk/software/
Alinhamento de sequências		
BLAST	bd	www.ncbi.nlm.nih.gov/BLAST/
CLUSTAL	bd	www.ebi.ac.uk/clustalw/
FASTA	bd	www2.ebi.ac.uk/fasta3
Modelagem de proteínas		
MODELLER	p	guitar.rockefeller.edu/modeller/modeller.html
SWISS-MOD	bd	www.expasy.ch/swissmod
SWISS-MODEL	bd	www.expasy.org/swissmod/SWISS-MODEL.html
SwissPdbViewer	p	ca.expasy.org/spdbv/
Phyre2	bd	
Validação de modelos		
BIOTECH	bd	biotech.embl-heidelberg.de:8400/
PROCHECK	p	www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
WHATCHECK	p	www.cmbi.kun.nl/swift/whatcheck/
PROVE	bd	www.ucmb.ulb.ac.be/UCMB/PROVE

* bd: banco de dados; p: programa

Fonte: SOUZA, 2019.

A análise de similaridade, por exemplo, é útil para a descoberta de informações funcionais, estruturais e evolucionárias nas sequências biológicas. Para o alinhamento de sequências existem várias ferramentas desenvolvidas pela bioinformática que podem fornecer informações sobre uma nova sequência comparando-a com outras já previamente conhecidas através do acesso a banco de dados. A ferramenta mais popular de comparação de sequências com os bancos de dados genômicos é o BLAST ou Basic Local Alignment Search (ALTSCHUL et al., 1990). Através deste algoritmo é possível identificar rapidamente no banco de dados, a presença de uma sequência suficientemente parecida com a pesquisada e descartar os resultados não produtivos das regiões de homologia detectadas. Com esta ferramenta podemos comparar uma sequência de DNA ou proteína com todas as sequências depositadas nos bancos de dados públicos. O BLAST pode ser utilizado para inferir relações funcionais e evolutivas entre sequências, bem como ajudar a identificar os membros de famílias gênicas (BARBOSA et al., 2012).

Os domínios proteicos são, por sua vez, regiões conservadas de uma dada sequência de proteína que estão relacionadas à sua estrutura e função. Geralmente, as proteínas apresentam um ou mais domínios funcionais. Os domínios são considerados unidades funcionais e correspondem a uma parte da cadeia polipeptídica que é independentemente estável, apresentando capacidade para a realização de uma tarefa física / química específica ou para ligação a um dado substrato, entre outras funções. Diferentes combinações de domínios formam a grande variedade de proteínas que estão presentes na natureza. A identificação dos domínios de uma proteína, portanto, pode fornecer informações sobre sua função (ALBERTS et al., 2010).

O Pfam e o CD-Search são bancos de dados de domínios e famílias de proteínas e constituem importantes ferramentas para predição de função proteica (PUNTA et al., 2012; MARCHLER-BAUER et al., 2017; MARCHLER-BAUER et al., 2015; MARCHLER-BAUER et al., 2011). Neles, as proteínas são analisadas por alinhamentos de sequências múltiplas e modelos ocultos de Markov (HMM) (FINN et al., 2014). Os HMMs são modelos probabilísticos utilizados para a inferência estatística de homologia construída a partir de um conjunto alinhado de sequências de famílias representativas definidas pelo curador. Um alinhamento de alta qualidade é essencial, uma vez que fornece a frequência de bases na posição específica para cada aminoácido (EDDY et al., 1998).

Com o resultado do alinhamento é possível identificar domínios funcionais, visualizar a arquitetura dos domínios proteicos, examinar a distribuição entre espécies e consultar alinhamentos múltiplos (FINN et al., 2014). Outra ferramenta que auxilia na descoberta de novos motivos funcionais é o MEME (disponível em meme-suite.org). Neste servidor é possível descobrir novos motivos não fixados (padrões recorrentes de comprimento fixo) em suas sequências (exemplos de saída de sequências). O MEME divide padrões de comprimento variável em dois ou mais motivos separados (TIMOTHY et al., 1994; TIMOTHY et al., 2009).

Por último, determinar a localização celular de uma proteína é um passo importante para compreensão de sua função. A localização de uma proteína na célula pode ser predita através da identificação de peptídeos sinais presentes na sequência de aminoácidos da proteína por meio de análises *in silico* (KUO-CHEN CHOU e HONG-BIN SHEN, 2008).

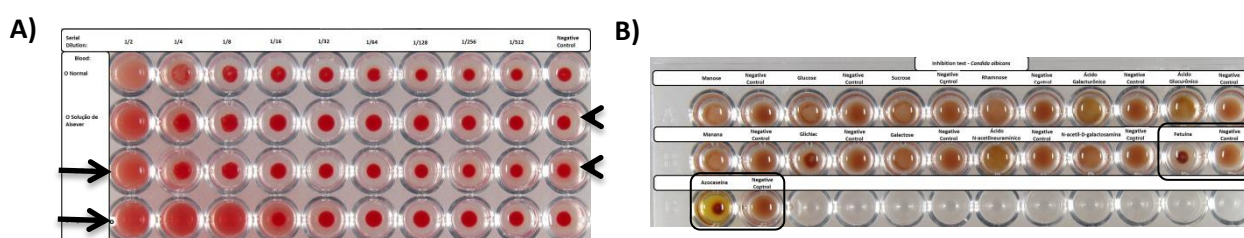
Cada proteína carrega em sua estrutura a informação necessária específica para sua localização na célula. Sequências específicas de aminoácidos chamadas de peptídeos sinais determinam se a proteína será transportada através da membrana para o interior de uma organela, se será exportada para fora da célula ou se será integrada à membrana. Estas sequências podem ser uma pequena porção de aminoácidos geralmente na extremidade amino ou carboxila terminal da proteína; ou podem estar localizadas no meio da proteína (ALBERTS et al., 2010). Vários programas computacionais permitem a predição da localização subcelular de proteínas. O Cell-Ploc (disponível em <http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc/>), por exemplo, prediz localização celular de proteínas através da presença de peptídeos sinal. Este programa identifica proteínas, incluindo aquelas com múltiplos locais, abrangendo 22 localizações subcelulares em diferentes organismos (KUO-CHEN CHOU e HONG-BIN SHEN, 2008).

2.8 TÉCNICAS APLICADAS AO ESTUDO DE LECTINAS

Técnicas específicas foram estabelecidas para identificar, quantificar e caracterizar diferencialmente classes distintas de lectinas. Por serem capazes de ligar-se seletivamente e reversivelmente a diversos tipos de carboidratos da superfície dos eritrócitos sem alterar a estrutura de nenhum resíduo glicosil ligante, ensaios da atividade hemaglutinante detectam a presença de lectinas numa determinada amostra e tem sido o primeiro passo na confirmação do isolamento de lectinas. Neste ensaio, uma amostra do extrato proteico é

incubada com eritrócitos tratados enzimaticamente (tripsina, papaína ou neuroaminidase) ou quimicamente (glutaraldeído ou formaldeído). A ligação de lectinas contidas na solução de incubação com os carboidratos da superfície dos eritrócitos formam uma malha de hemaglutinação, podendo ser visualizado a olho nu. Contudo, a ausência de lectinas na amostra resulta em precipitação dos eritrócitos **Figura 8.A** (COELHO et al., 2012). Confirmada a presença de lectinas por meio da hemaglutinação, outros ensaios são necessários, como os ensaios de inibição da hemaglutinação por açúcares são realizados, a fim de identificar a qual carboidrato (s) a lectina possui afinidade de ligação **Figura 8.B**.

Figura 8. Ensaios de hemaglutinação de proteínas totais de *C. albicans* em **A**, mostrando as diluições seriadas do extrato proteico, contendo tampão e suspensão de eritrócitos de humano (Tipos A, B e O) a 2,5%. Setas em **A** indicam a ocorrência da malha de hemaglutinação enquanto que as pontas de setas indicam a precipitação dos eritrócitos onde não houve hemaglutinação. Em **B**, ensaios de inibição com açúcares (Poços marcados com quadrados mostram respectivamente a inibição da hemaglutinação e a não-inibição da hemaglutinação com os açúcares Fetuína e Azocazeína).



Fonte: Própria autora, 2020.

Posterior aos ensaios de hemaglutinação, técnicas de purificação se fazem necessárias a fim de se obterem produtos finais puros e com boa concentração. Dentre as principais formas de se adquirir um extrato semi-purificado, podemos citar as técnicas de fracionamento com solução de sulfato de amônia ou acetona (CARVALHO et al., 2012), sistema de duas fases aquosas – SDFA (YUZUGULLU; DUMAN, 2015), diálise em membranas semipermeáveis (LAM e NG, 2011) e por fim, cromatografias (COELHO et al., 2012). Essas técnicas gradualmente excluem os produtos indesejáveis do seu extrato, permitindo um produto final puro (LAM e NG, 2011).

Métodos cromatográficos utilizados na purificação de lectinas incluem a cromatografia de troca iônica, afinidade e exclusão molecular (MALUCHE e SANTOS, 2008; LAM e NG, 2011). De forma geral, todas tem como base a separação de acordo com seu peso molecular, solubilidade, diferenças na carga e afinidade de ligação, onde o produto de interesse é ligado em um material sólido poroso com propriedades químicas

específicas (fase estacionária) mantido em uma coluna, pela qual é passada uma solução tampão (fase móvel).

Outra técnica que permite a caracterização e determinação da massa molecular de proteínas, como por exemplo as lectinas, é a espectrometria de massa (MS), onde todos os compostos químicos de uma amostra em fase aquosa são ionizados e as moléculas carregadas resultantes (íons) são analisadas de acordo com sua relação massa/carga (m/z) (AEBERSOLD e MANN 2003).

Métodos de eletroforese em gel de poliacrilamida uni ou bidimensional (1D-PAGE, 2D-PAGE) são utilizados como uma forma mais simples de separação de misturas de proteínas com base em seu ponto isoelétrico e pesos moleculares, e frequentemente usados para caracterizar estruturalmente as lectinas como também para revelar o grau de pureza das mesmas (KUMAR et al., 2015).

2.9 MODELAGEM MOLECULAR DE PROTEÍNAS

Como a função das proteínas está diretamente ligada à sua estrutura, o conhecimento detalhado desta estrutura ajuda a entender o papel de cada proteína. Mesmo depois de desvendar a estrutura das proteínas, é preciso entender suas interações, classificar as proteínas, encontrar relações entre elas, identificar padrões, estudar seu funcionamento. Quanto mais conhecermos sobre as proteínas e suas estruturas, mais dados teremos para saber como ocorrem determinadas doenças e até desenhar novas moléculas que possam atuar como medicamentos (BAKER e SALI, 2001; KLIS et al., 2009).

A visualização de dados e a representação gráfica são de especial interesse para os bioinformatas, pois é quase sempre a maneira mais efetiva de visualizar, descrever, explorar e condensar um grande volume de informações, tornando os gráficos em instrumentos poderosos para a análise da informação quantitativa. Juntos, os dados de sequência, características físico-química, estrutura e função das proteínas, entre outros, fornecem uma multiplicidade de informações cruciais para entender os processos biológicos (BAKER e SALI, 2001).

O conhecimento sobre a estrutura tridimensional de uma proteína é muito importante para entender sua função proteica. A predição da estrutura 3D de proteínas codificadas no genoma é um dos objetivos da genômica estrutural, podendo ser por métodos

experimentais ou teóricos. As principais técnicas experimentais empregadas na predição estrutural de uma proteína são: a cristalografia por difração de raios-X (RX) e a espectroscopia de ressonância magnética nuclear (RMN). Mesmo sendo considerado o melhor processo para a obtenção da estrutura tridimensional de uma proteína, os métodos de predição experimental além de serem financeiramente custosos, podem levar anos e, em alguns casos, a estrutura final pode não chegar a ser obtida (VAILATI-RIBONI et al., 2017). Além disso, um número elevado de sequências proteicas encontram-se depositadas em bancos de dados como o “Swiss-Prot” (BAIROCH et al., 2004), contudo, poucas estruturas de proteínas foram resolvidas experimentalmente e depositadas em bancos de dados como o PDB (Banco de Dados de Proteínas) (WESTBROOK et al., 2002). Diante disso, o emprego de métodos teóricos para a predição da estrutura tridimensional de proteínas seria uma saída para as limitações e entraves encontrados nos métodos de predição experimental (PETSKO e RINGE, 2004). Os principais métodos teóricos de predição de estrutura 3D de proteínas são: modelagem comparativa, predição de enovelamento de proteínas (Threading) e predição por primeiros princípios (*ab initio*).

Diante da necessidade cada vez maior de elucidação da funcionalidade de proteínas no mesmo ritmo que de sequências de genes e de genomas, metodologias baseadas em homologia de sequências têm contribuído muito para predição da função putativa do gene (FILHO e ALENCASTRO, 2003; DESLER et al., 2009; DOERKS et al., 2012). Essa metodologia compara a estrutura da sequência de interesse com sequências de proteínas já conhecidas para prever sua estrutura proteica. Através da similaridade de sequências é possível atribuir função de 60-70% das proteínas em genomas mais estudados (TOUCHON et al., 2009; CHOI et al., 2011; QIAO et al., 2013). As proteínas restantes, sem homólogos ou com homólogos de função desconhecida, são referidas como hipotéticas (KOONIN et al., 2002). A modelagem comparativa, até pouco tempo conhecida como modelagem por homologia tem sido o método mais frequentemente usado e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde (SANTOS-FILHO e ALENCASTRO, 2003).

A predição de enovelamento de proteínas (Threading) e a predição por primeiros princípios (*ab initio*) são utilizadas quando as sequências de proteínas de interesse não possuem uma proteína homóloga no banco de dados de estrutura ou que apresentam uma identidade muito baixa. A metodologia por enovelamento de proteínas tenta ajustar a

estrutura da proteína de interesse aos tipos de enovelamentos de proteínas conhecidas, ou seja, registradas e depositados em bibliotecas de enovelamentos, produzindo modelos precisos de proteínas (ZHANG, 2009).

Já na predição por primeiros princípios (*ab initio*) é possível, em princípio, predizer a estrutura tridimensional de pequenas proteínas e peptídeos a partir de sua estrutura primária. Essa técnica emprega as propriedades físico-químicas conhecidas de cada aminoácido para a construção de funções de energia, sendo que tais funções são minimizadas por algoritmos que realizam buscas no espaço de conformações que a proteína de interesse possa assumir (LEE et al., 2009; PAQUET e VIKTOR, 2018).

3 OBJETIVOS

3.7 Objetivo geral

Identificar e caracterizar *in silico* sequências e estrutura de proteínas com domínio lectina conservados, respectivamente nos fungos patogênicos *Candida albicans*, *Cryptococcus neoformans* e *Aspergillus fumigatus*.

3.8 Objetivos específicos

Os objetivos específicos do trabalho correspondem às ações desenvolvidas para atingir o objetivo geral, sendo eles:

- Identificar sequências características de lectina em *Candida albicans*, *Cryptococcus neoformans* e *Aspergillus fumigatus* utilizando-se ferramentas de bioinformática;
- Caracterizar a estrutura dos domínios lectina conservados nessas proteínas através de uma análise *in silico*;
- Detectar regiões de similaridades entre essas proteínas e outras do mesmo grupo funcional;
- Avaliar o perfil transcricional de sequências codificadoras promissoras através de RT-PCR e qPCR, bem como realizar o sequenciamento *in vitro* de cDNAs amplificados;
- Prever por modelagem comparativa pelo menos uma das proteínas identificadas como hipotéticas.

4 PROCEDIMENTOS METODOLÓGICOS

4.1 ANÁLISE BIOINFORMÁTICA

Inicialmente, foram selecionadas sequências-sonda de 11 famílias proteicas (L-type, R-type, P-type, EPA, PWP, C-type, CBM, FLO, F-type, G-type e I-type) no banco de dados do UniProt (**Tabela 3**). Essas sequências foram alinhadas via tBLASTn contra os genomas de interesse (*C. albicans*, *C. neoformans* e *A. fumigatus*) depositado no NCBI, adotando-se $e\text{-value} \leq 0,05$ e similaridade $\geq 30\%$. Posteriormente, as sequências de nucleotídeos foram coletadas, a estrutura dos genes analisadas e traduzidas para aminoácidos pelo programa Transdecoder. Seus domínios conservados foram identificados com o auxílio do programa e dos servidores CD-search (disponível em <https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) e Pfam (disponível em <https://pfam.xfam.org/search/sequence>), respectivamente, e realizada a predição do ponto isoelétrico (pI) e peso molecular utilizando o programa JVirGel 2.0 bem como a localização subcelular através do servidor Cell-PLoc (disponível em <http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc/>).

Tabela 3. Palavras-chave utilizadas na pesquisa de sondas.

Palavras-chave	Família gênica
lectin cbm taxonomy:fungi	CBM
lectin "c type" taxonomy:fungi	C-TYPE
adhesin epa taxonomy:fungi	EPA
lectin flo taxonomy:fungi	FLO
lectin "f type"	F-TYPE
lectin "g type" taxonomy:fungi	G-TYPE
lectin "I type" taxonomy:fungi	I-TYPE
lectin "l type" taxonomy:fungi	L-TYPE
lectin "p type"	P-TYPE
adhesin pwp	PWP
lectin "r type"	R-TYPE

4.2 ALINHAMENTO, ANÁLISE DE DOMÍNIOS, MOTIVOS E FENÉTICA

As sequências primárias de nucleotídeos e aminoácidos das sequências com domínio lectina foram alinhadas usando o algoritmo ClustalW utilizando a matriz BLOSUM disponível no software MEGA 7.0 para cada família gênica utilizando os parâmetros padrões e editadas pelo programa BioEdit Sequence Alignment Editor. A fim de obter bons resultados de alinhamento, os domínios coletados para cada família foram definidos como

sequências de no mínimo 10 aminoácidos com 30 % ou menos de similaridade, entre regiões de alta similaridade (≥ 70 %) (TAMURA et al., 2011).

A estrutura e distribuição dos motivos conservados foram analisadas usando a ferramenta MEME, disponível em <http://meme-suite.org/tools/meme> (BAILEY et al., 2009) com os seguintes parâmetros: número máximo de motivos para encontrar = 7; mínimo largura do motivo = 5; e largura máxima do motivo = 10. A identificação dos principais motivos dentro das famílias L-type, R-type, P-type, C-type, CBM e FLO foram realizadas em nível de nucleotídeos e aminoácidos, tanto para a sequência do domínio, quanto para a sequência do gene / proteína completa.

As árvores fenéticas foram construídas também pelo software MEGA 7.0 usando o algoritmo Neighbor-joining para 2000 réplicas de bootstrap. Foram construídas árvores em nível de nucleotídeos e aminoácidos, tanto para a sequência do domínio, quanto para a sequência do gene / proteína completa para cada uma das famílias L-type, R-type, P-type, C-type, CBM e FLO. Foram usadas sequências de lectinas já conhecidas dos fungos *Fusarium oxysporum*, *F. graminearum* e *Neurospora crassa* como outgroups.

4.3 ANÁLISES DE ANOTAÇÃO DE ONTOLOGIA PELO BLAST2GO

Os termos utilizados pela Ontologia Gênica (GO), contribuem para a padronização da representação dos genes e seus produtos. Para isso, atribuímos os termos GO para as PHs no genoma de *A. fumigatus*, *C. albicans* e *C. neoformans* usando o programa Blast2GO, a fim de classificar funcionalmente, quando possível, em três categorias para o mesmo produto gênico: (I) Processo Biológico – que se refere à atividade biológica com o qual o gene ou seu produto contribui; (II) Função Molecular – atividade bioquímica do gene ou produto; e (III) Componente Celular – local onde o produto gênico é ativo (CONESA et al., 2005; ASHBURNER et al., 2011). Foi iniciada uma análise específica buscando a caracterização da possível função de cada uma destas proteínas, por meio da mineração de dados (Gene Ontology - GO) (ASHBURNER et al., 2011) e da busca por domínios funcionais pelo Interproscan - IPR (ZDOBNOV; APWEILER, 2001). No Blast2GO, com o objetivo de identificar possíveis homólogos e determinar os níveis de similaridade entre as PHs e de outros organismos, as sequências foram submetidas a uma busca de similaridade utilizando o BLASTp (ALTSCHUL et al., 1997) contra o banco de dados de proteínas não redundantes (NR) do NCBI. Não foram considerados alinhamentos

com valores maiores do que o limiar de $e\text{-value} \leq 1.0E-3$. A pontuação de alinhamento e a cobertura da sequência também foram consideradas na filtragem dos dados. Uma vez que o conjunto de homólogos das proteínas hipotéticas foi criado, mapearam-se os termos GO associados aos hits encontrados com o BLAST. O GO, usado para a representação dos genes e seus produtos, identificou três categorias diferentes de proteínas: Ao fim dessa etapa, foram identificados termos GO de diferentes categorias para cada uma das proteínas hipotéticas analisadas. Os termos do GO encontrados foram associados aos alvos do BLASTp, para uma melhor interpretação dos dados.

4.4. MATERIAL BIOLÓGICO

As linhagens patogênicas dos fungos *Candida albicans* (cepa 6366), *Cryptococcus neoformans* (cepa 6908) e *Aspergillus fumigatus* (cepa 6573) utilizadas neste projeto foram obtidas através da Micoteca URM do Departamento de Micologia do Centro de Biociências da Universidade Federal de Pernambuco, Brasil. Os fungos são linhagens de isolados clínicos e foram mantidos em placas com meio sólido YPD-ágar (extrato de levedura 1% w / v, peptona 2% w / v, dextrose 2% w / v, ágar 2% w / v) e estoques congelados a -80 °C em glicerol 70 % até sua utilização.

4.5 EXTRAÇÃO E QUANTIFICAÇÃO DE RNA TOTAL E SÍNTESE DE cDNA

Para *A. fumigatus*, o micélio foi mantido em meio YPD por sete dias (em fase de diferenciação), sem rotação a 37 °C. Em seguida as culturas foram diluídas em meio YPD até a concentração celular final de 1×10^8 por ml^{-1} . As células foram recolhidas por centrifugação, lavadas com água estéril (4 °C), novamente centrifugadas e congeladas em nitrogênio líquido.

O RNA total foi extraído utilizando o *RNeasy mini kit* (Qiagen) e ~ 500 ng foi analisado em gel de agarose (0,7 % p / v). A síntese do cDNA foi realizada conforme protocolo do kit SMARTScribe Reverse Transcriptase (Clontech Laboratories), realizada a partir do RNA total extraído e quantificado com concentração final de 400 ng / μl em Nanodrop 2000 Spectrophotometer (ThermoScientific).

4.6 RT-PCR, qPCR E SEQUENCIAMENTO

Um total de 16 conjuntos de iniciadores foi projetado conforme mostrado na **tabela 4** através da ferramenta gratuita Primer3Plus (disponível em: <http://www.bioinformatics.nl/cgi->

[bin/primer3plus/primer3plus.cgi](http://www.bmblab.com/bin/primer3plus/primer3plus.cgi)). A PCR foi realizada utilizando a enzima Taq DNA Polymerase (Invitrogen), de acordo com o protocolo estabelecido pelo fabricante. Cada reação de 10 µL continha 10 × tampão de PCR (100 mM de Tris-HCl [pH 9,0], 15 mM MgCl₂, 500 mM KCl e 1,0% Triton X-100), dNTPs (2.5 µM), primers (10 µM cada), polimerase Hot star Taq (Invitrogen), e modelo de DNA genômico (~20 ng) ou cDNA. A amplificação foi realizada com a Mastercycler Pro S (Eppendorf) a 95 ° C durante dois minutos, depois a 95 ° C por 30 segundos, 59 ° C por 30 segundos e 72 ° C por 30 segundos por 30 ciclos, seguido de extensão final a 72 ° C por cinco minutos. Os produtos de PCR foram analisados por eletroforese em gel de agarose 1,8 % a 100V, por 50 min e corado com brometo de etídio (0,5 µg/mL) e os amplicons detectados em transiluminação ultravioleta por um sistema de documentação eletrônica (GelDoc-It Imaging Sistema; UVP). A reação de transcrição reversa (RT-PCR) foi realizada pela enzima GoTaq (Promega) de acordo com o protocolo estabelecido pelo fabricante. Foram utilizados em cada reação 2,5 pmols dos iniciadores Oligo(dT) primer (5'-TTTTT-3') e Random primer (5'-NNNNN-3').

Os experimentos de PCR em tempo real (qPCR) foram realizados a partir do RNA total extraído da linhagem de *A. fumigatus* (cepa 6573) após seu crescimento em meio YPD por sete dias. As análises foram feitas com dois experimentos independentes, cada um sendo em triplicata. A expressão dos três genes (CLECT, 2180 e 6800) foi avaliada utilizando para a detecção o fluorocromo SYBR Green. A reação de qPCR foi composta por 20 ng de cDNA, 4,5 pmols de cada iniciador (**Tabela 4**), 5 uL do PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) em um volume final de 10 uL. A amplificação foi realizada no ViiA 7 Real-Time PCR System (Thermo Fisher Scientific) com as seguintes condições de ciclagem: 50°C por 2 min, 95°C por 2 min e 40 ciclos a 95°C por 15 seg, 55°C por 15 seg e 72°C por 1 min. A expressão relativa foi normalizada pela média geométrica da expressão do gene ACTIN-1.

Posteriormente os amplicons da PCR convencional para CLECT e 2180 foram purificados enzimaticamente utilizando-se Exonuclease 1 e pela fosfatase alcalina de camarão, conforme instruções do fabricante das enzimas (USB). O amplicon 6800 foi purificado a partir de gel de agarose a 2% usando o QIAquick Gel Extraction Kit (Quiagen). Os amplicons foram submetidos à reação de sequenciamento usando a técnica BigDye Terminator v3.1 Cycle Sequence Kit (Applied Biosystems) no sequenciador

automático ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). As sequências nucleotídicas foram analisadas utilizando o software CLC Main Workbench v.6 (Quiagen).

Tabela 4. Iniciadores para a reação de RT-PCR e qPCR e genes de referência utilizados no presente estudo.				
Tamanho do fragmento		Gene-alvo	Sequência dos iniciadores (5'→3')	
<i>C. albicans</i> Housekeeping	148 pb	ACTIN-1	Forward	GCAAGCAGGAGTACGATGAG
			Reverse	CCATGCCAATGTTGTCTCTT
<i>C. neoformans</i> Housekeeping	599 pb	GPD1	Forward	ATGGTCGTCAAGGTTGGAAT
			Reverse	GTATTCGGCACCAGCCTCA
<i>A. fumigatus</i> Housekeeping	77 pb	ACTIN-1	Forward	TGCTCCTCCTGAGCGTAAAT
			Reverse	ACATCTGCTGGAAGGTGGAC
<i>A. fumigatus</i> (C-type)	80 pb	XM_750498.1	Forward	CAACAAGACTGCCGATTGAG
			Reverse	GTTCACCAATAGCCCAGGAC
<i>A. fumigatus</i> (L-type)	155 pb	XM_750924.1	Forward	GCACACTAATGCCGTTCTTG
			Reverse	TAGGTTGCACGACGAACTTG
	121 pb	XM_750457.1	Forward	GGTAACAAGCGCATAACAACG
			Reverse	ATCACCACCAAAATCGAACC
<i>C. neoformans</i> (P-type)	113 bp	XM_012195597.1	Forward	CCTACTCGGGCGGTTATATG
			Reverse	CGGATTGTTACCACATGCAC
	114 bp	XM_769015.1 (CNBG4080)	Forward	ACATGGAGTACCACCCTTGC
			Reverse	TGTCGCTGTATTGGAAGTCG
	140 bp	XM_772876.1	Forward	TATGGGTGCATGACTGAAGC
			Reverse	TGAGTGCTGTATTCCGTTCC
<i>C. neoformans</i> (R-type)	158 bp	XM_012191579.1	Forward	GCTTCCCAATCTGACTCCTG
			Reverse	GTGGCATTCCCCTAAGAATC
	153 bp	XM_012192681.1	Forward	AATACCGCGCCACAATAATC
			Reverse	GGCCAAATAATCAGCTCGTC
<i>C. albicans</i> (L-type)	200 bp	XM_710947.2	Forward	TGGAGAGCATTGACGAATTG
			Reverse	CTTTTCAGCCATTTCGAACC
	141 bp	XM_707997.2	Forward	TGCCCAACCCACACTATTG
			Reverse	ATTTCCCCCAGATTGAACC
<i>C. albicans</i> (FLO)	101 bp	XM_704874.2	Forward	TCCACACAACACTCCTCCTG
			Reverse	TGGCAATTTGAGTCTCTCTC
	94 bp	XM_718358.1	Forward	CGTCGCTGGTTCTTTAGTTG
			Reverse	GTCGGAACAATGAGTGATGG
	187 bp	XM_709369.2	Forward	TCAAGAAACCACCGTTGAAG
			Reverse	TTCAGTTGGAGTGGCAACAG
	86 bp	XM_714904.2	Forward	TCCTGGGTTTGGTCATTCTC
			Reverse	GTAGGGACATCGGTTTGAGC
	88 bp	XM_712681.1	Forward	GCTCCGATCATTAACATTGC
			Reverse	TGGCGACAGAGATTGATTG
	142 bp	XM_715415.2	Forward	TCATCGGCCTTACCTGTTG
			Reverse	CATAACGAGTTTCTGTTGCTTC
	121 bp	XM_709479.2	Forward	AGGCAAAGGTGGCTATCAAG
			Reverse	TGTTGGTTGAATTGGAGTGG
	138 bp	XM_709370.2	Forward	ACCACCAAAGGTTCTTCTGC
			Reverse	CAAACCTGCAGGAACAGTGG

Fonte: SOUZA, 2019.

4.7 MODELAGEM COMPARATIVA E DINÂMICA

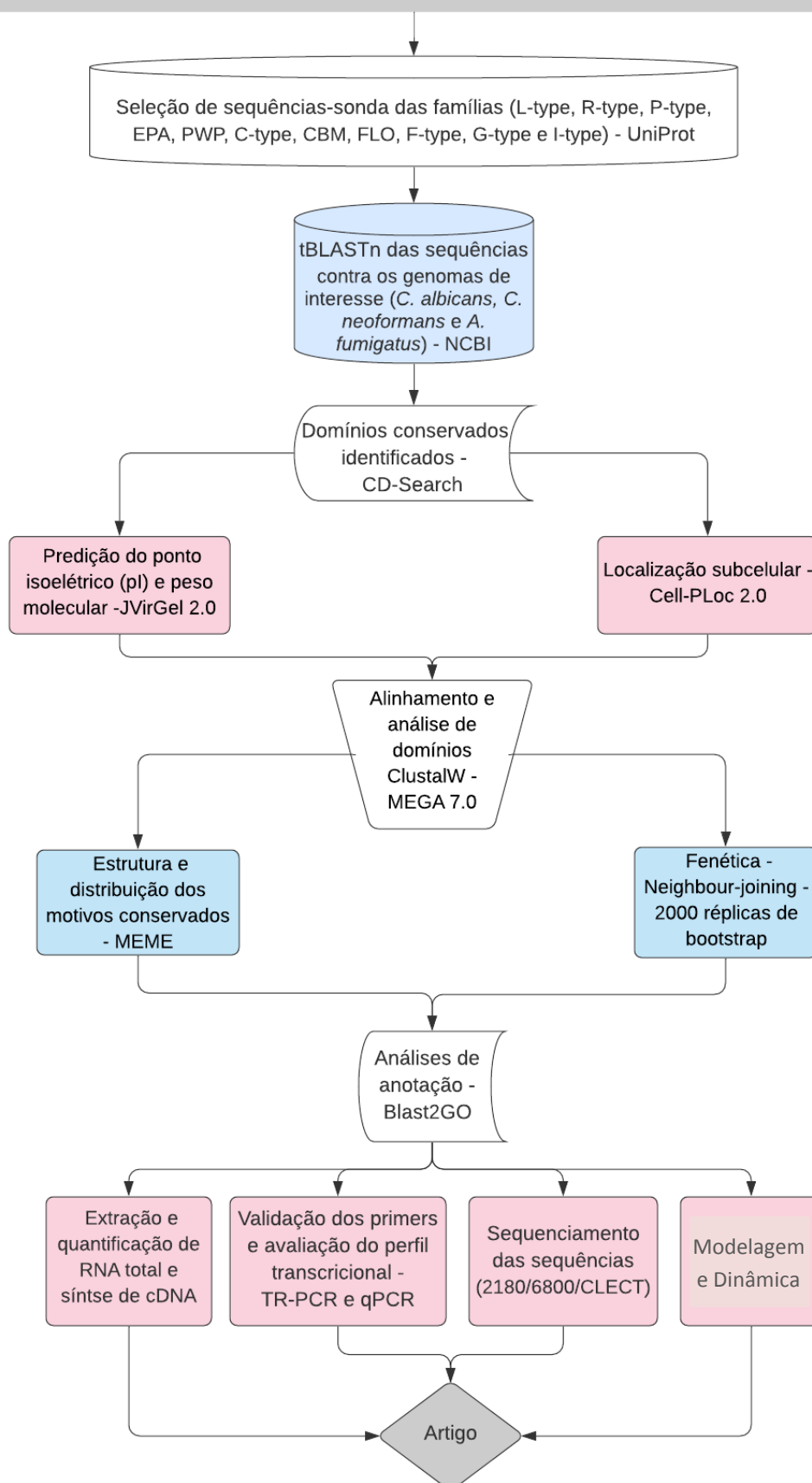
Para modelagem tridimensional, foi utilizado o banco de dados de proteínas PDB, a fim de encontrar o melhor molde para a construção dos modelos comparativos, onde foram

selecionadas lectinas levando em consideração maior identidade de sequência e a cobertura. O modelo estrutural para a sequência 2180 de *A. fumigatus* foi gerado a partir da proteína VIP36 da membrana integral vesicular resolvida por raios-X (PDB id: 2DUP) na resolução de 2,1Å. Então, utilizando o Modeller 9.23 (WEBB e SALI, 2014), mil modelos foram construídos, e o modelo final foi escolhido de acordo com *discrete optimized protein energy* (DOPE) que acessa a energia do modelo e indica a estrutura mais provável baseado em um estado de referência aprimorado que corresponde a átomos que não interagem em uma esfera homogênea com o raio dependente de uma estrutura nativa da amostra. Por fim, para validação dos modelos ProSa II (WIEDERSTEIN e SIPPL, 2007) e PROCHECK (LASKOWSKI et al., 1993) foram utilizados para analisar a qualidade do dobramento e a estereoquímica respectivamente.

As simulações em dinâmica molecular foram conduzidas com o pacote GROMACS 2019.4, as proteínas foram colocadas no centro de uma caixa cúbica de dimensões 5 x 5 x 5 nm e solvatadas usando o modelo de água SPC (BERENDSEN et al. 1981). Os sistemas foram solvatados em solução de NaCl em concentração fisiológica (0,15 M) substituindo moléculas de soluto por íons NA e CL, em seguida a energia foi minimizada. A temperatura foi mantida a 300 K no conjunto NvT com restrição de átomos de soluto na posição inicial. O método LINCS (HESS et al. 1997) foi usado para restringir ligações envolvendo átomos de hidrogênio. A integração foi realizada pelo algoritmo Leapfrog (HOCKNEY, 1970), usando uma etapa de tempo de integração de 2 fs. Os sistemas foram inicialmente otimizados em energia usando 50.000 etapas do algoritmo de descida mais íngreme. A simulação atomística foi realizada usando o campo de força GROMOS 53A6 (OOSTENBRINK et al. 2004) e condições de contorno periódicas nas direções x, y e z usando o conjunto NpT. Em seguida, a dinâmica molecular foi realizada sem restrições com pressão e temperatura constantes de 1 atm e 300 K respectivamente em um tempo de 100 ns. O alinhamento das estruturas foi realizado pelo PyMol 2.3 e A carga residual foi realizada no VMD 1.9.3.

Pipeline

Linhagens patogênicas de *Candida albicans* (6366), *Aspergillus fumigatus* (6573) e *Cryptococcus neoformans* (6908)



5 ANÁLISE DOS RESULTADOS

5.1 SELEÇÃO E CARACTERIZAÇÃO DAS SEQUÊNCIAS COM DOMÍNIO LECTINA

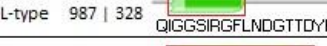
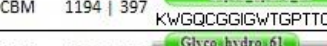
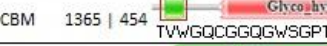
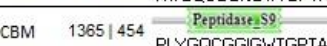
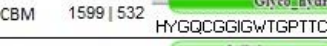
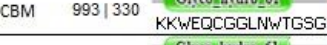
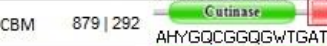
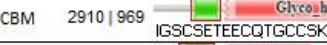
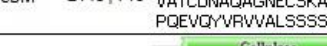
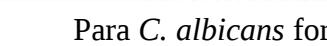
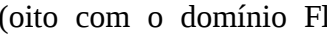
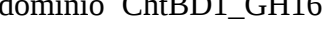
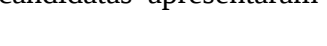
Para o presente estudo foram selecionadas sequências sonda oriundas de proteínas que codificam para as famílias de lectinas de interesse, presentes em outros microrganismos e enumeradas na **tabela 5**. Ao total, obtivemos 11 famílias proteicas (L-type, R-type, P-type, EPA, PWP, C-type, CBM, FLO, F-type, G-type e I-type) nos bancos de dados do UniProt, CGD e AspGD.

Tabela 5. Número de sequências sonda coletadas no UniProt, CGD e AspGD, candidatos com domínio de interesse coletados via BLAST-NCBI e domínios confirmados nos genomas de interesse no CD-Search.		
Famílias gênicas	Sequências sonda do Uniprot, CGD e AspGD	Sequências com domínio lectina (CD-search / Pfam)
CBM	133	26
C-TYPE	120	1
EPA	9	0
FLO	202	9
F-TYPE	31	0
G-TYPE	7	0
I-TYPE	29	0
L-TYPE	1299	6
P-TYPE	30	3
PWP	7	0
R-TYPE	2	3
11	1869	48

Fonte: Resultados do próprio autor, 2019.

Para *A. fumigatus*, foram encontradas 26 sequências de aminoácidos com domínio lectina mostradas na **figura 9** (17 com o domínio fCBD, dois com o domínio ChtBD1_1, um com o domínio F5_F8_type_C (Discoidina), um CBM_X2, um CBM_1, um domínio CLECT e três Lectin_leg-like) com pI variando entre 4 a 10.24 e PM de 16.75 a 184.80 KDa. Além disso, esses candidatos apresentaram localização extracelular, de membrana e citoplasma.

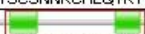
Figura 9. Estrutura dos domínios lectina pelo Pfam (quadrados e retângulos em vermelho) e suas respectivas seqüências de aminoácidos pelo CD-Search (seqüência abaixo à estrutura do domínio) em *Aspergillus fumigatus*. Domínio C-type aparece no CD-Search como incompleto na região N-terminal.

Família	nt/AA	Arquitetura do domínio - Pfam / Sequência do domínio de interesse - CD-Search
C-type	1794 597	 S/WAIGEPDPSPpGIDEDsqhierCAVMOLSlgGHWRSTNCSQERHAAC
L-type	987 328	 QIGGSIRGFLNDGTTDYKSHRzVDSLAFGHCDYAYRNLGPPSVVRLKHTNAVLEVTVDDKLCFSTNKVGLPTGNNGITAATPENPDSEIFKFVV
L-type	984 327	 PLRTHSLAPPYLDSDQSRWDFGDDTIIRaDKYIRLTSORPSQQGWFSRVLPTATNWEIEVEFNHGGQGNLHGGFAMWLTQKQATQGPVFGSTDN FEGLGIFFDYKNNRpGTSFPYVMAMMGDGKTSYDQAHGKANEVAGCSARGLRGASIPTKARLYFQDKsLTLDLQYKSEDTWTNCFNLNapeNIAIP AVSYLGFSaETGELSDNHDIISVN
CBM	1194 397	 KWGQCGGIGWTGPTTCVSGTTCQKLNWYSQCL
CBM	1029 342	 VYGQCGGSGWSGPTACVTGATCTSYNSYSQCI
CBM	714 237	 LYGQCGGSGWTGPTACASGATCKVLNPYSQCL
CBM	1365 454	 TWGQCGGQGWGSGPTSCVAGAACSTLNPYYAQCI
CBM	1434 477	 KHwYQCGGINWTGPTTCESGYTCVEQNPYYHQCv
CBM	1116 371	 HWGQCGGNGWTGPTACASGYTCTVINEWYSQCL
CBM	1365 454	 PLYGQCGGIGWTGPTACATG-VCTAYSPYYAQCL
CBM	1599 532	 HYGQCGGIGWTGPTTCASPYTCQKLNWYSQCL
CBM	1194 397	 HWGQCGGIGWTGPTACASPYTCQKLNWYSQCL
CBM	1383 460	 THYGQCGGQGWGPTACASPYTCQVLNPWYSQCL
CBM	1503 500	 AWGQCGGISYTGSTCCSGNTCAEINSYYYQCI
CBM	993 330	 KKWEQCGGLNWTGSGSCEARTTCHQNPYYQCI
CBM	1050 349	 YGQCGGQWYSGPTSCEAGLICKEWNPYYHQCv
CBM	1191 396	 PLYAQCGGNGWTGSTQCIAGACSSINDWYYQCL
CBM	879 292	 AHYGQCGGQGWGATSCQSPYTCTVISQWYSQCL
CBM	915 304	 GPWQCGGRSYTGETSCVSGWSCVLFNEWYSQCG
CBM	1401 466	 GPWQCGGRSYTGETSCVSGWSCVLFNEWYSQCG
CBM	528 175	 GPWQCGGRSYTGETSCVSGWSCVLFNEWYSQCG
CBM	2910 969	 IGSCSETEECQTGCCSKAGYCGFGPDFCGDDVCISNCD
CBM	1464 487	 RCGPGVG--TCPNGKCCSTAGYCGTTQAHCRSPdCQ
CBM	2148 715	 VATCDNAQAGNECSKAIDGDNSTYwHTDSNSPIPHSITINLGTAKVSLAVwPRKIEDGWIAAHVYLSPDGNSWGPVAGHTwWPDSTVKMAIFE PQEVQYVRVVALSSSSGNNAISIADL
CBM	1728 575	 YIFHQVGTPTVTSQTLPIFNGNLTLSIRDSSgTTLKQGTQDYSVSGSNLVFTASYLSKHfsastkPGIADLTLVFSSGASPVVQ
CBM	1158 385	 LYAQCGGEGYSGPTSCVAGAVCTVWNP

Para *C. albicans* foram identificadas também 12 seqüências com domínios completos (oito com o domínio Flocculin-t3, duas com o domínio Lectin_leg-like, uma com o domínio ChtBD1_GH16 e uma com o domínio Flo11), mostrados na **figura 10**. As candidatas apresentaram pI entre 3.38 e 9.31 e PM variando entre 22 e 111 kDa. Os

resultados mostraram ainda que todas apresentam localização na membrana celular e / ou extracelular.

Figura 10. Estrutura dos domínios lectina em *Candida albicans* pelo Pfam (retângulos em vermelho) e suas respectivas seqüências de aminoácidos do domínio de interesse pelo CD-Search (seqüência abaixo à estrutura do domínio).

Família	nt/AA	Arquitetura do domínio - Pfam / Seqüência do domínio de interesse - CD-Search
L-type	1311 436	 Lectin _{leg} -like PYLDESQFHKNWDSQGNLLVRNhdFIRLTANAPHQVSNMFSKWPIHAESFEMELTFHIHNPDVKHGLwDGLAIWFLDKPSDIGDVFQGNKFNGLGIML DTFKNGKRQqFPYVNLMLGDGNAMYNKATDGYETRLAGCIAQLLNPeaKETKMRLVYIKsGYLSIDFNYYGHHeQWQNCVTLTDVKLPETKYLGLSA ETGQLVENVDIENRIYAL
L-type	1398 465	 Lectin _{leg} -like PNPHYSLPNLLSIESKEDINNFDSNNIQFDNGRLLLGELGSIWgKYKIPILNKPWTIELIFRSTGTKEdkYEDNSLNVWLLNEDNNLFPDSFDGFSEINNE GQIPGVKLYNNDGLQNIHDASHVLGTCKFYLDSDVPFTLRISYDANSWFKIQLDNNLCFTNQISIPFQELKLGITSKINqQSNEKFEILSLKTWELL
FLO	660 219	 TTEITITHCSDNKCATSVAPAVQSVNTVTIEGVVTEYTTYCPLT
FLO	1602 533	 TKVITITSCSENKCQETTVEATPSTATTVSEGWVTEYVTCYCPVS
FLO	2262 753	 TSVVTITSCSKGGCGTQSYATGVTKTKTVHEGVTEYTTYCPLT
FLO	1536 511	 Hyphal _{reg} -CWP TKVVTITSCSNACHPTTAPTGLTVVTLTTSQVKTWVTTYCPLT
FLO	1548 515	 TAVITITSCSNKCHEQTKTTGKIVYTSEETVYTYCPLS
FLO	1962 653	 TTIVITSSCYEDKCSVSSVTTGVVTTISSEETNTTYCPIT
FLO	1233 410	 TKIITITSCSNnnnnHVCPTTVLATPSTMTKTIDNTITEYTYCPLT
FLO	2166 721	 NAGFNAAFVVSNAKKLSDGSYGIDCNFKSDssVQLNLAFGKKVKQLSITGTGYSDISLLGNvaNPFEWSASLKVKAEIVKGKCLPSgFRIVTDF
FLO	1905 634	 TSCSNSVCTESEVTTGVVITSKDTNTTYCPLT
CBM	1413 470	 TTCSSSKHCPEDKPCCSQFGICGTGAYCLGGCDIRYSYNLTACMPMP

Estas análises identificaram 12 seqüências com domínios de lectina conservados para *C. neoformans* (dois com o domínio ChtBD1_GH16, três com o domínio WSC, quatro com o domínio tipo Lectin_{leg} e três domínios Ricin_B-lectina), mostrados na **figura 11**, apresentando pI entre 3,985 e 12,247, PM variando de 11 a 102 KDa e localização subcelular na membrana celular, parede celular e espaço extracelular para todas as seqüências.

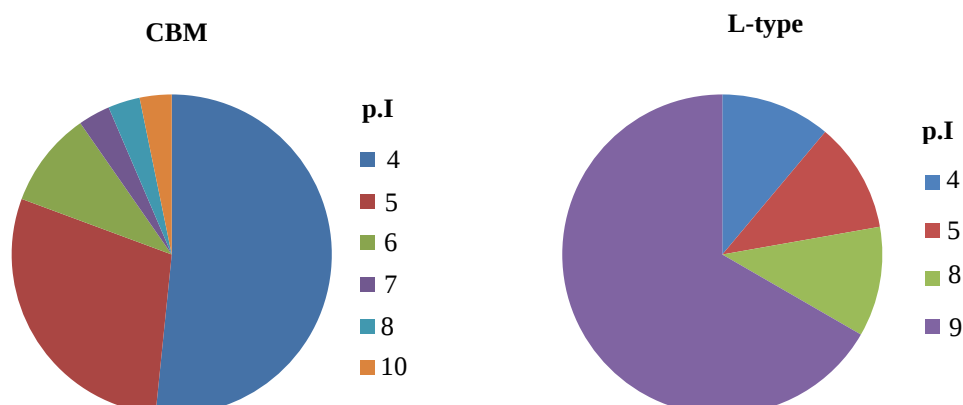
Figura 11. Estrutura dos domínios lectina pelo Pfam (retângulos em vermelho) e suas respectivas seqüências de aminoácidos de interesse pelo CD-Search (seqüência abaixo à estrutura do domínio) em *Cryptococcus neoformans*.

Família	nt/AA	Arquitetura do domínio - Pfam / Sequência do domínio de interesse - CD-Search
R-type	2370 789	 NQHPSMCLAPAHGWEGETDVLKDCDEDDTTWLTGQSFGNTATNFCIDIRDS _{Gaw} SGNKAQVWGCFSYNTNQQFSVEESMIHWNGFCWDLTDGSS SAGIKLQIWSCYSYNDNQW
R-type	1716 571	 STRIVRLHPNgdksKCVLLGNTIQDQGPVQICDCDGTAPQDWVLNaGRGQkVQLAGTSFCLDATHPy _a ADGNTMKIWKCLDVQQQDWYWTSDNRV LRQGGKCLD _w ATGDRSDFNQLQVWRCSTDNNNQVW
R-type	1734 577	 TVRLhPnGDKSKCVLLGNDHQDGPVQVYDCNGTPAQDWVLNaGRGQkVQLAGTNFCLDATHPy _a ADGIKMKIWKCLDVRRQQDWYWTSDNKIAL RDQGGKCLD _w ANGDRSDFNQLQVWQCSTGNNNQVW
P-type	5337 1778	 YLGCYKDGAGSNMAFNMYTYHITSMTVETCKAACVELGYLAGLTDGFECRCGDNYPQTTQMVSOLYCTKPCYGNASETCGA
P-type	4119 1372	 YVGCFKDqGSNAFTGNWYTHRSNSMDTEMCKQICLFGYTYAGLKANECRCGNHPASQRMVSSLYCTASCSSSQTCGA
P-type	1824 607	 YGCMTEATNARALAKYSFTSSMTPQLCMSTCQAEGYMSGTEYSTQCICYGNTFSAGSVPADESSCSMACGGDKLQTCGG
L-type	1035 344	 PLRTHLSAAPYVDTDLQNRWVDFGDDTInTNKHVRLTQDKPSQSGSWLWARMPLSVSNWQIDVEFKVGDGAHNIFGDGWAFWLTSRAKQGPVFG SVDWFKIGGIFFDYANSKhaYTTPRVSAMLDGDKTSYDHDORNEANEIGGCSENFRRGDvPTKARLTyVKGaLQLKLQTKKSDEWKICFETNVDL ESPYIGFSAATGDVSDDDHIVAVNTYSL
CBM	1131 376	 YDGNATAYDwVVNTGSLvaSSDNDGARLILTNNTGSKISSTRYIHGYTIDFTLQTSKWAGVVTAAITMSDVKDEIDewWPGSNTDMVQTNyWFLGVA NY _{sa} TEGASSVSSDTATNFHTYIDWQESYISWIDGLLVRTVRKIDTisedGTQYKFPSTPSRVQISWIPAGTDONAQGTIDWAGGYIDW _{Tnsdyl} INGYF WNTIKSVNISCAY
CBM	1131 376	 YDGNATAYDwVVNTGSLvaSSDNDGARLILTNNTGSKISSTRYIHGYTIDFTLQTSKWAGVVTAAITMSDVKDEIDewWPGSNTDMVQTNyWFLGVA NY _{sa} TEGASSVSSDAAFNHTYIDWQESYISWIDGLLVRTVRKIDTisedGTQYKFPSTPSRVQISWIPAGTDONAQGTIDWAGGYIDW _{Tnsdyl} ISINGYF WNTIKSVNISCAY

Candidatas com o domínio FLO podem ocorrer na parede celular, citoplasma e núcleo. Enquanto aquelas com domínios lectin_leg-like e Laminin_G (L-type) podem ser encontradas também no retículo endoplasmático e complexo de golgi. Foram gerados gráficos que mostram a média dos pIs e PMs para cada família proteica (**Figura 12**) e esses resultados foram confrontados com os dados da literatura e corroboram com as características para cada família de proteínas em outros microrganismos e são mostradas na **Tabela 6**. Esses resultados permitem entender melhor as características das lectinas nestes fungos, como saber o pH ótimo em que cada uma apresentará sua melhor atividade, como também para uma futura metodologia de extração otimizada.

Figura 12. Gráficos mostrando as médias preferenciais de ponto isoelétrico (p.I.) (A) e peso molecular (P.M.) (B) por família proteica para *A. fumigatus*, *C. albicans* e *C. neoformans*.

A)



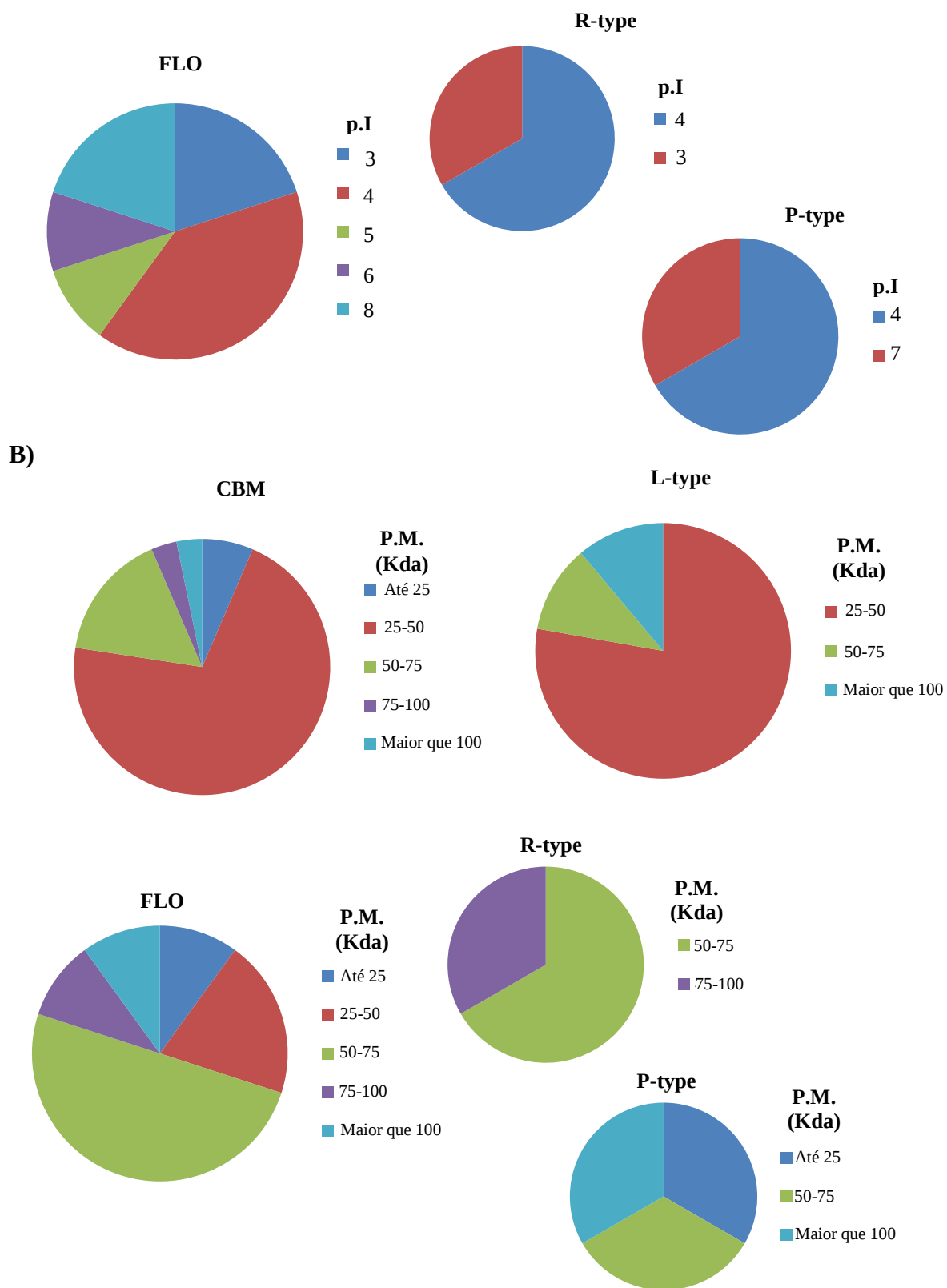


Tabela 6. Principais características físico-químicas das sequências proteicas coletadas nos bancos de dados NCBI, CGD e AspGD.

Família	Proteína	ID	Domínios	pI	p.M (Kda)	Localização subcelular	Loci	Ref.
<i>Aspergillus fumigatus</i>								
C-type	Lectina (C-type)	XM_750498.1	CLECT (D.I.)	4,593	65,097	MC / PC / EC / C / N	chr2	ITO et al., 2011
L-type	Lectina Putativa (AFUA_2G16800)	XM_750924.1	lectin_L-type (D.I.)	9,233	36,066	MC	chr2	ZHANG et al., 2012
L-type	Lectina Putativa (AFUA_2G12180)	XM_750457.1	lectin_leg-like	8,748	36,193	EC	chr2	
L-type	PH (Lectina como Concanavalin A)	EDP49519.1	Laminin_G_3	4,884	184,802	EC		SUMNER et al., 1938
CBM	endo-1,4-beta-xylanase	XM_746144.1	fCBD	5,33	42,13	EC	chr6	SHOSEYOV et al., 2006
CBM	endo-1,4-beta-glucanase	XM_743614.1	fCBD	5,13	35,52	EC	chr3	
CBM	endoglucanase_A	XM_743744.1	fCBD	4.15	24.64	MC / EC	chr7	
CBM	cellobiohydrolase	XM_743418.1	fCBD	4.83	47.79	EC	chr3	
CBM	endoglucanase_B	KU933946.1	fCBD	4.39	50.81	EC		
CBM	endoglucanase/celullase_A	HQ836477.1	fCBD	4.39	50.81	EC		
CBM	endoglucanase/celullase_B	XM_750193.1	fCBD	4.39	50.81	EC	chr2	
CBM	acetyl xylan esterase	XM_742365.1	fCBD	5.78	39.25	EC	chr8	
CBM	polyhydroxybutyrate depolymerase	XM_741450.1	fCBD	7.96	47.59	EC	chr4	
CBM	1,4-beta-D-glucan-cellobiohydrolyase	XM_745951.1	fCBD	4.49	56.45	EC	chr6	
CBM	endoglucanase_C	XM_745950.1	fCBD	4.59	42.64	EC	chr6	
CBM	endoglucanase_D	XM_742804.1	fCBD	4.56	48.20	EC	chr6	
CBM	endoglucanase_E	JQ796068.1	fCBD	4.48	48.20	EC		
CBM	swollenin	XM_742655.1	fCBD	4.66	52.71	EC	chr6	
CBM	beta-1,4-endoglucanase	KF154781.1	fCBD	4.87	48.33	EC		
CBM	endoglucanase_F	XM_745750.1	fCBD	5.57	35.37	EC	chr6	

CBM	endoglucanase_G	XM_742389.1	fCBD	5.86	35.73	EC	chr8	
CBM	endoglucanase_H	JQ796069.1	fCBD	8.46	35.30	EC		
CBM	glycosyl hydrolase/cellulase	XM_744136.1	fCBD	5.90	42.92	EC	chr2	
CBM	AXE1	XM_743269.1	fCBD	5.36	29.27	EC	chr3	
CBM	endo-beta-mannanase_A	EU925595.1	fCBD	4.74	45.65	MC / EC		
CBM	endo-1,4-beta-mannosidase	XM_742408.1	fCBD	5.36	50.29	MC / EC	chr8	
CBM	endo-beta-mannanase_B	EU925594.1	fCBD	10,24	16,75	EC		
CBM	chitinase	XM_745774.2	ChtBD1	4,003	103,662	EC	chr6	
CBM	chitin deacetylase	XM_745838.1	ChtBD1_1	6,512	52,736	C / EC	chr6	
CBM	galactose oxidase	XM_743736.1	F5_F8_type_C (Discoidin)	5,063	77,265	EC	chr7	
CBM	endoglucanase_I	XM_743066.1	CBM_X2	6,864	62,487	EC	chr5	
CBM	cellulose-binding GDSL lipase/acylhydrolase	XM_744094.1	CBM_1	6,098	41,795	EC	chr2	
<i>Candida albicans</i>								
L-type	PH (CAALFM_C207810WA)	XM_710947.2	lectin_leg-like	9,313	50,604	EC	chr2	ZHANG et al., 2012
L-type	EMP46	XM_707997.2	lectin_L-type	5,078	53,306	EC / N	chrR	
FLO	PGA6	XM_718358.1	Flocculin_t3	4,204	22,098	EC	chr1	STRAVER et al., 1994
FLO	YWP1	XM_709369.2	Flocculin_t3	4,526	54,271	MC	chr2	
FLO	PGA18	XM_714904.2	Flocculin_t3	5,229	75,167	EC	chr3	
FLO	IFF11	XM_712681.1	Flocculin_t3	8,221	53,352	MC	chr3	
FLO	PGA38	XM_715415.2	Flocculin_t3	4,286	53,532	EC	chr4	
FLO	EAP1	XM_709479.2	Flocculin_t3	3,38	111,136	EC	chr2	
FLO	PH (CAALFM_C208580WA)	XM_709370.2	Flocculin_t3	6,43	44,761	MC / PC / EC / C / N	chr2	
FLO	RBT1	XM_704874.2	Flo11	4,041	73,635	EC	chr4	

FLO	HWP1_B	XM_704869.2	Flocculin_t3	3,323	65,305	EC	chr4	
CBM	UTR2	XM_716655.1	ChtBD1	4,226	51,7	MC / EC	chr3	SHOSEYOV et al., 2006
<i>Crptococcus neoformans</i>								
R-type	Celulase	XM_012191579.1	Ricin_B_lectin	3,985	85,291	EC	chr1	CUMMINGS et al., 2009
R-type	PH (CNK01820)	XM_567661.1	Ricin_B_lectin	4,75	61,872	EC	chr11	
R-type	PH (CNAG_02694)	XM_012192681.1	Ricin_B_lectin	4,872	62,427	EC	chr3	
P-type	Receptor de transmembrana	XM_012195597.1	WSC	4,091	102,695	MC	chr8	DAHMS E HANCOCK, 2002
P-type	PH - (CNBG4080)	XM_012191009.1	WSC	7,075	67,437	EC	chr1	
P-type	PH -(CNBA4380)	XM_772876.1	WSC	4,257	19,417	MC / EC	chr1	
L-type	Lectina (Leg-like)_A	XM_572915.1	lectin_leg-like	9,498	38,285	EC / C	chr9	
CBM	Proteína de Parede Celular	XM_566898.1	ChtBD1	4,21	41,684	MC / EC	chr1	SHOSEYOV et al., 2006
CBM	PH (CNBA5600)	XM_772770.1	ChtBD1	4,281	40,528	MC / EC	chr1	

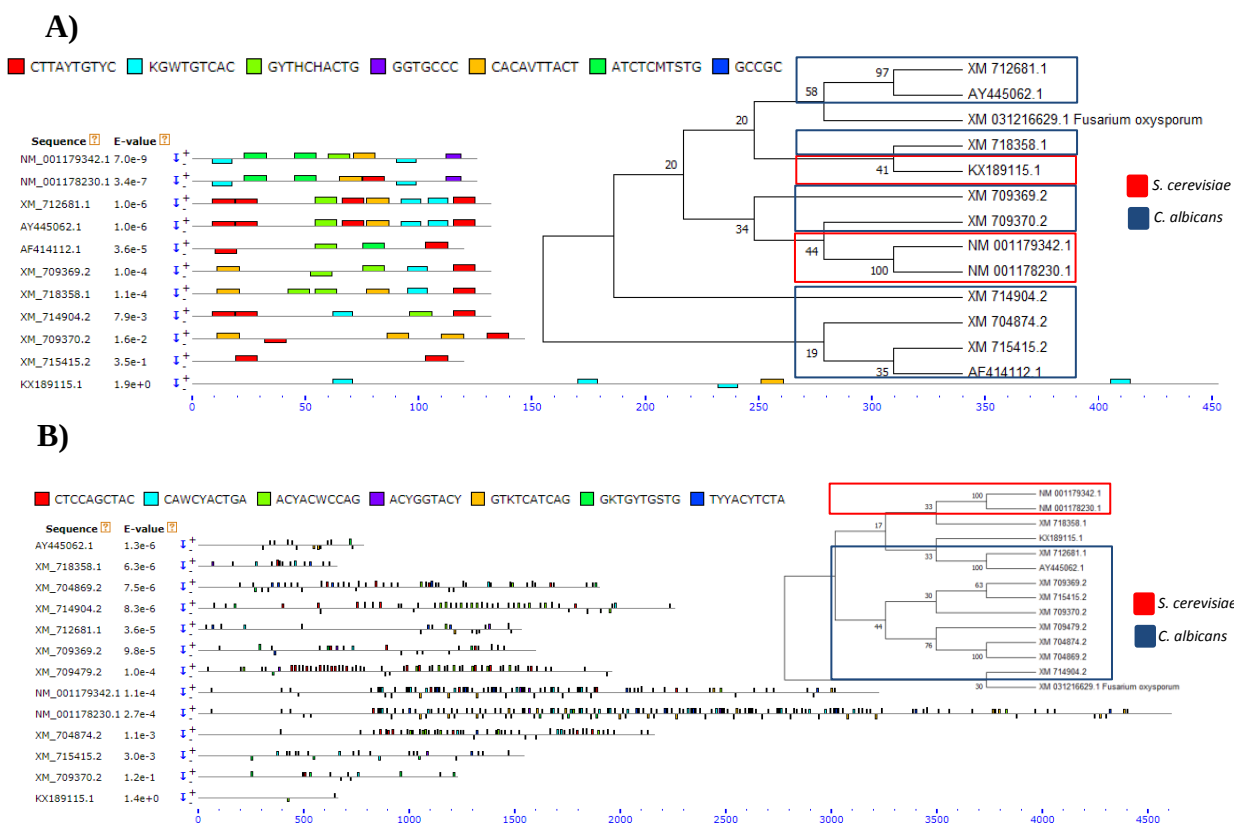
MC = Membrana Celular PC = Parede Celular EC = Extracelular C = Citoplasma N = Núcleo RE = Retículo Endoplasmático pI = Ponto Isoelétrico PM = Peso Molecular

Fonte: Resultados do próprio autor, 2019.

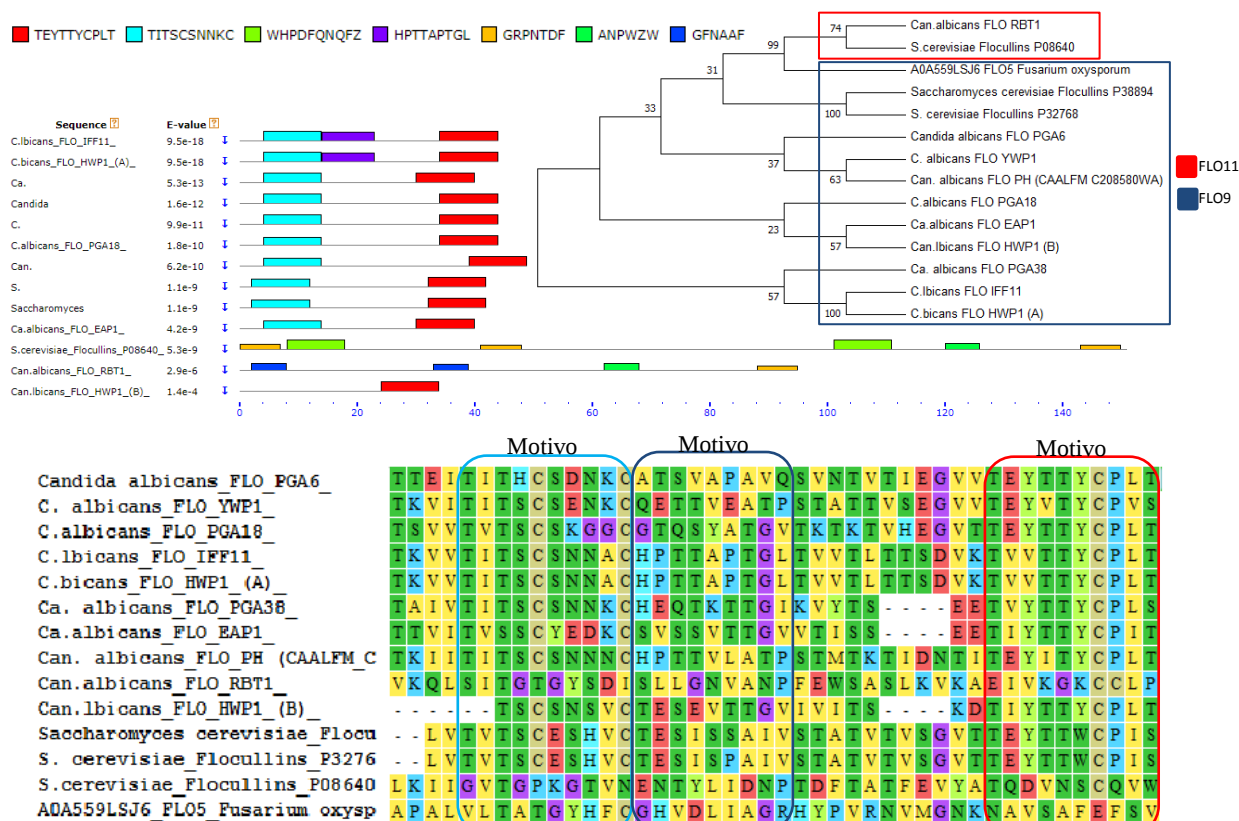
5.2 ANÁLISE DE DOMÍNIOS E FENÉTICA

Sequências nucleotídicas e de aminoácidos dos domínios-lectinas do banco de dados de proteínas da NCBI, CGD e AspGD foram coletadas e analisadas para entender sua homologia evolutiva e encontrar a presença de domínios conservados. As análises foram feitas para cada família em separado (CBM, FLO, L-type, R-type, P-type e C-type), juntamente com sequências conhecidas entre espécies próximas para cada família e são mostrados nas **figuras 13, 14, 15, 16, 17 e 18**.

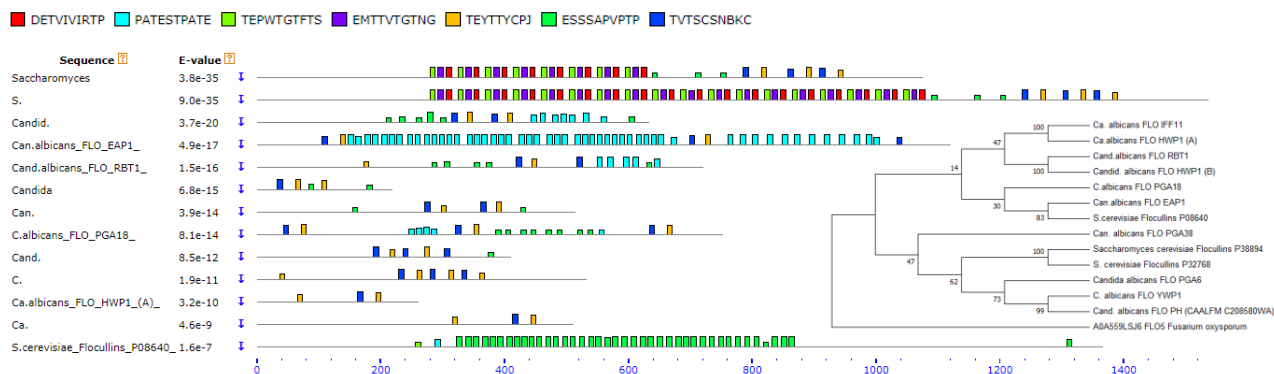
Figura 13. Análise *in silico* da família proteica FLO. Alinhamento múltiplo entre *C. albicans* e espécies próximas (já descritas na literatura), mostrando as regiões mais conservadas entre elas, distribuição dos motivos encontrados pelo MEME, bem como a árvore fenética, mostrando o agrupamento por espécies e por classificação, confirmando e justificando o resultado dos motivos gerados pelo MEME. Em (A) gerados com a sequência nucleotídica dos domínios, em (B) gerados com a sequência nucleotídica dos genes, em (C) gerados com a sequência de aminoácidos dos domínios e em (D) gerados com a sequência de aminoácidos das proteínas. Sequência de *F. oxysporum* foi usada como outgroup.



C)



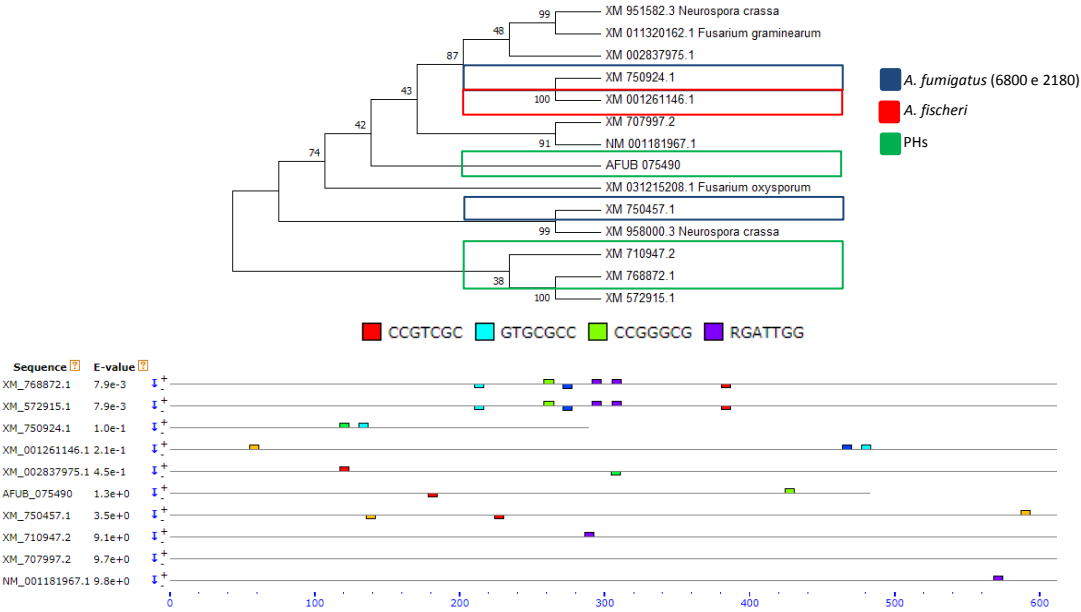
D)



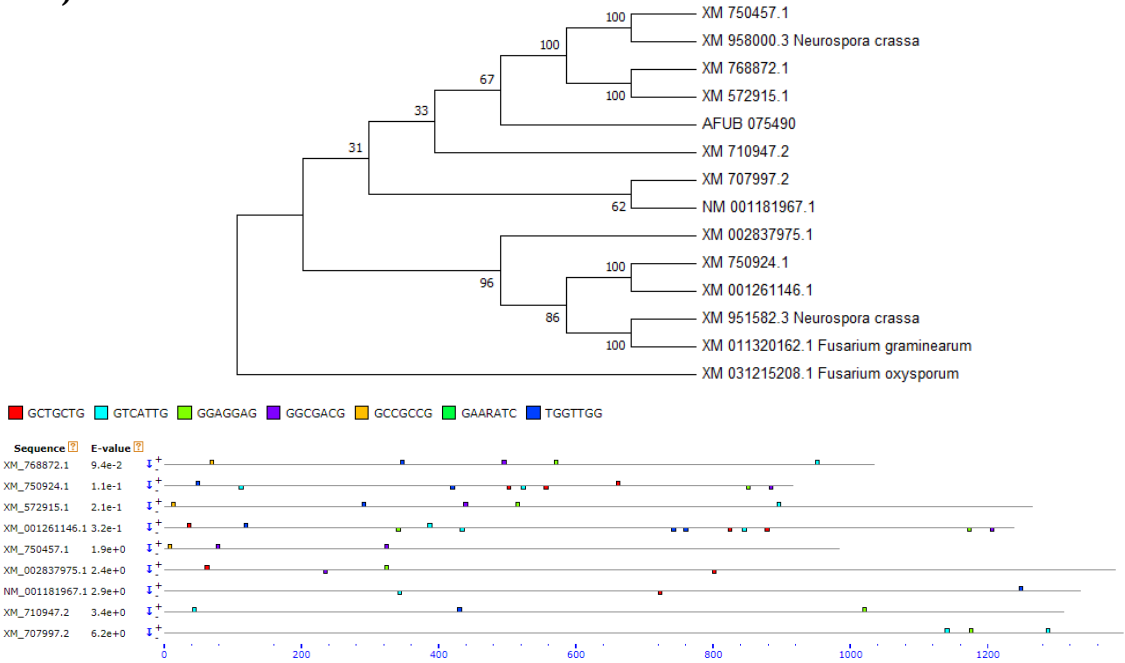
Fonte: Resultados do próprio autor, 2019.

Figura 14. Análise *in silico* da família proteica L-type. Alinhamento múltiplo entre *C. albicans*, *A. fumigatus*, *C. neoformans* e espécies próximas (já descritas na literatura), mostrando as regiões mais conservadas entre elas, distribuição dos motivos encontrados pelo MEME, bem como a árvore fenética, mostrando o agrupamento por espécies e por classificação de proteína, confirmando e justificando o resultado dos motivos gerados pelo MEME. Em (A) gerados com a sequência nucleotídica dos domínios, em (B) gerados com a sequência nucleotídica dos genes, em (C) gerados com a sequência de aminoácidos dos domínios e em (D) gerados com a sequência de aminoácidos das proteínas. Sequências de *F. oxysporum*, *F. graminearum* e *N. crassa* foram usadas como outgroup.

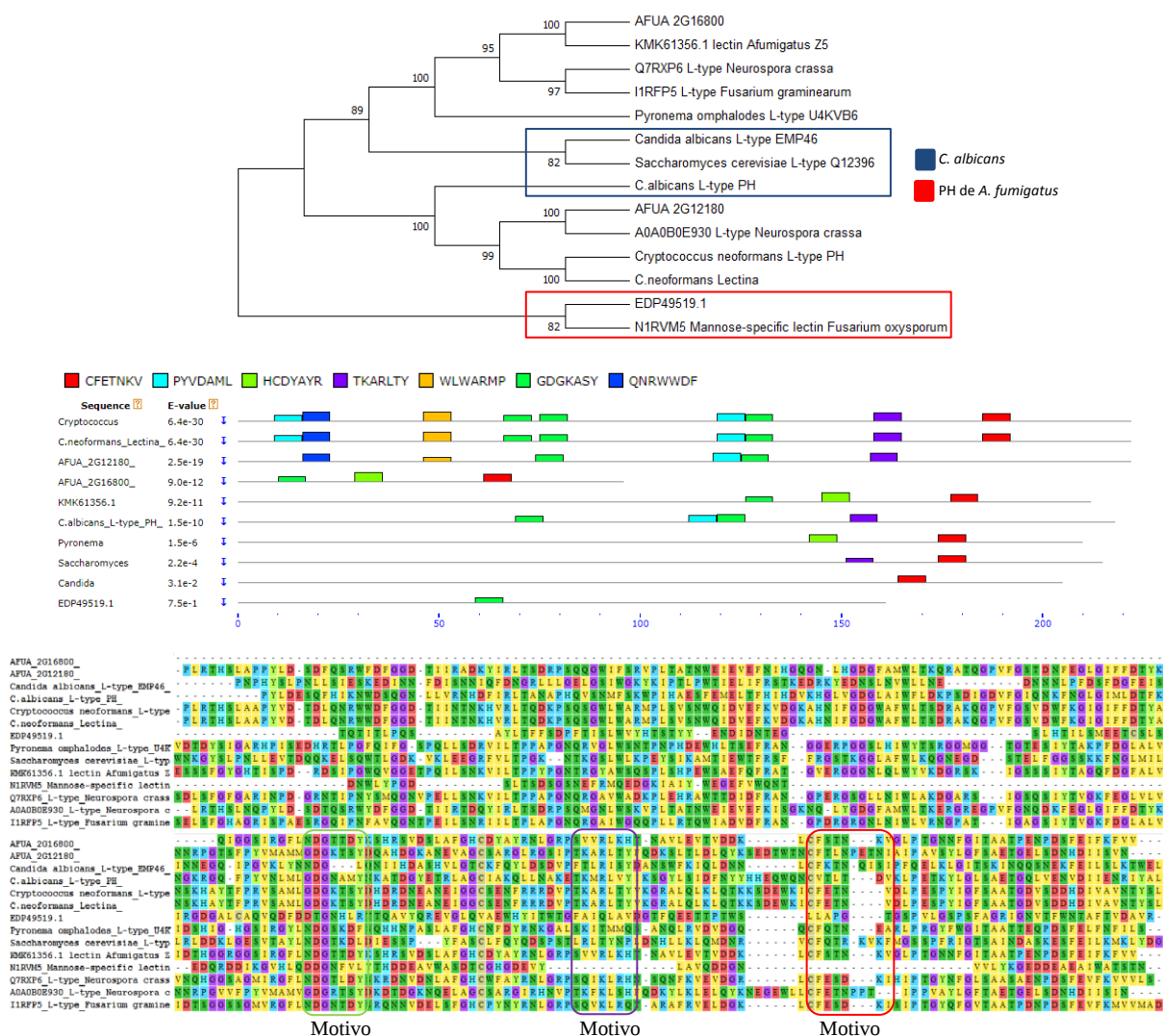
A)



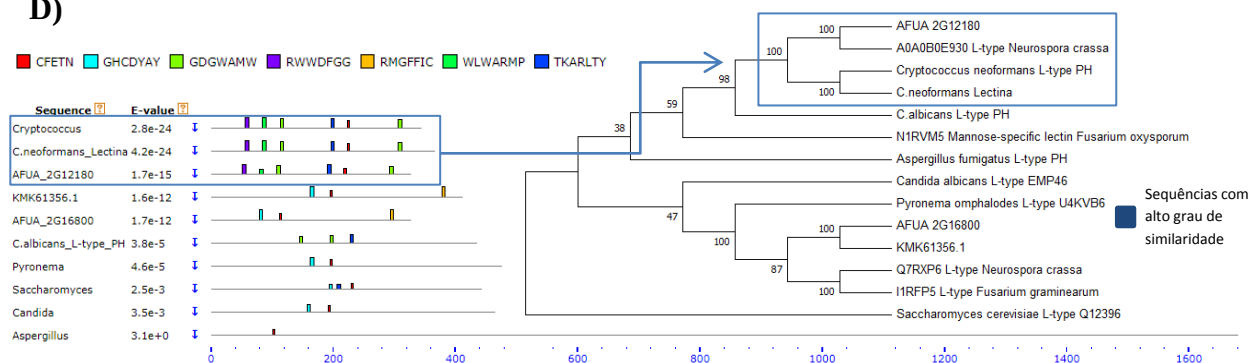
B)



C)



D)

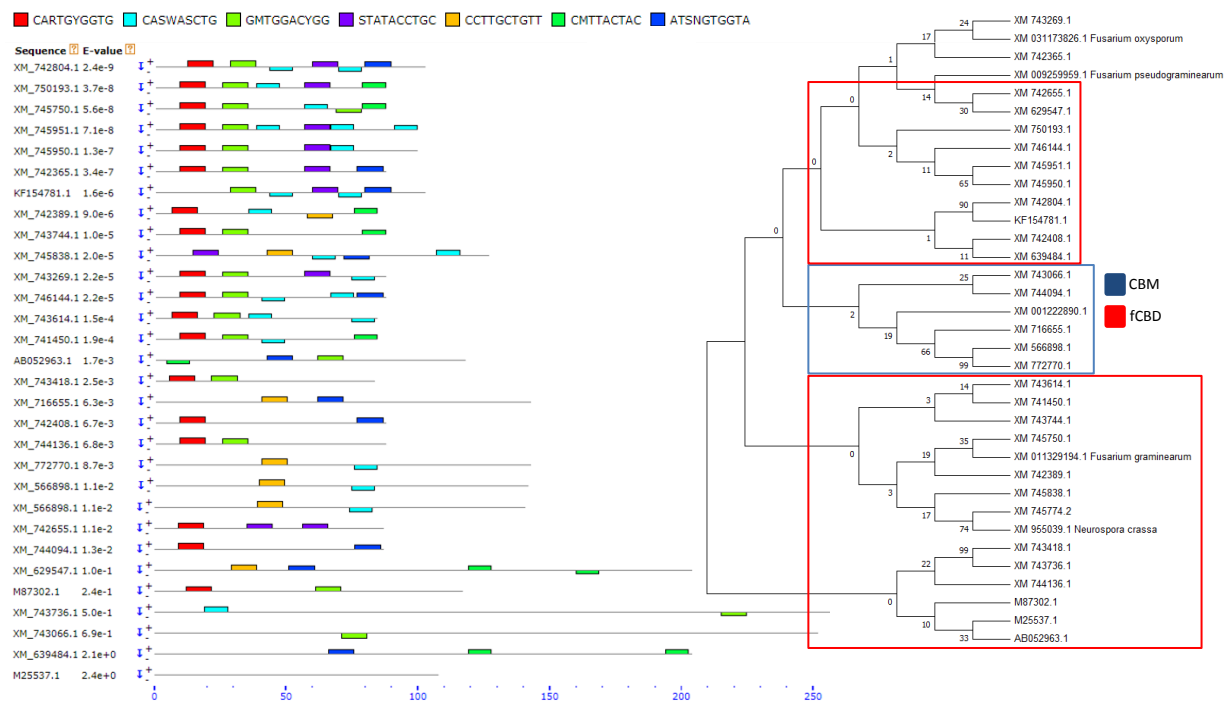


Fonte: Resultados do próprio autor, 2019.

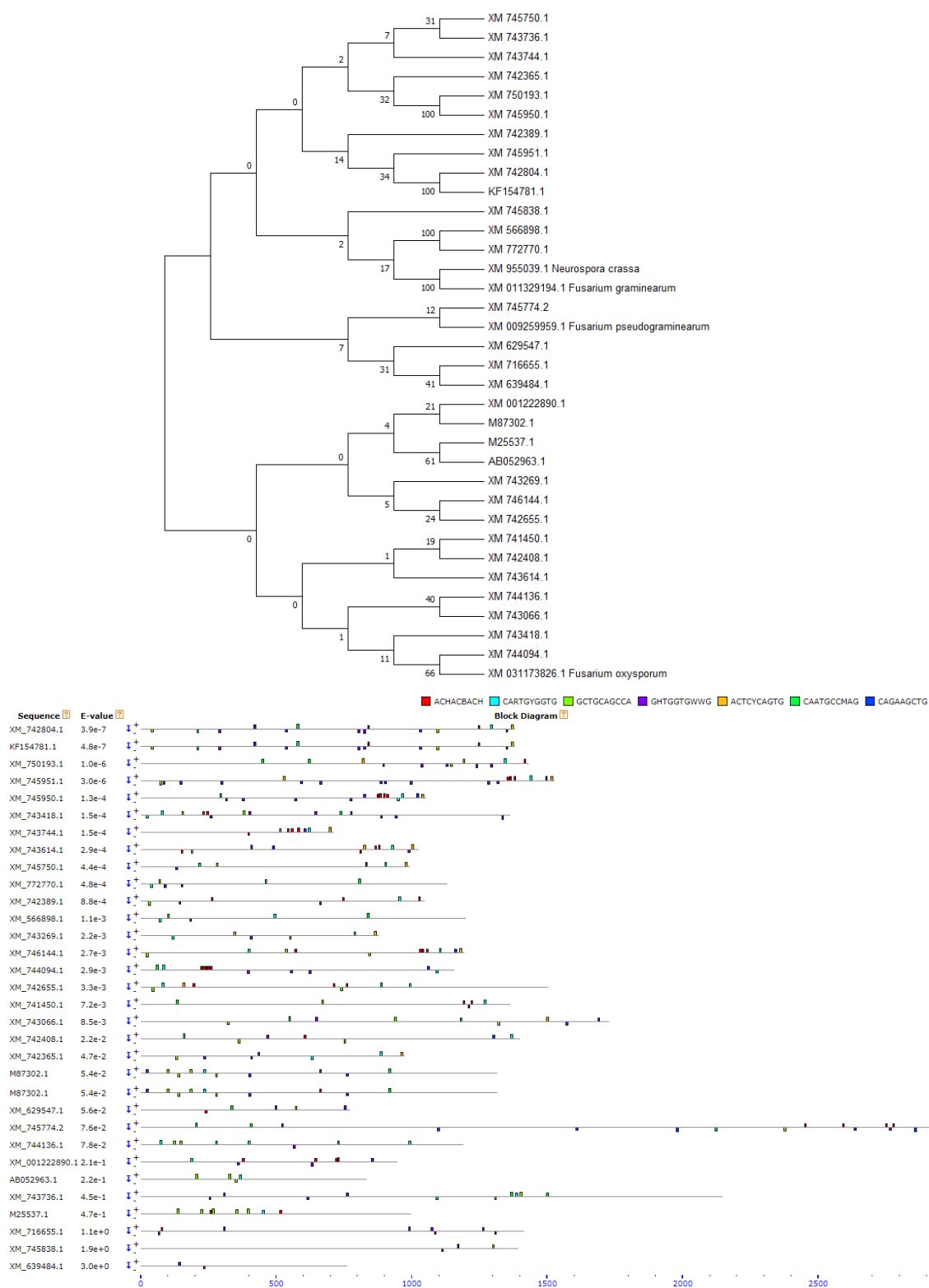
Figura 18. Análise *in silico* da família proteica CBM. Alinhamento múltiplo entre *C. albicans*, *A. fumigatus*, *C. neoformans* e espécies próximas (já descritas na literatura), mostrando as regiões mais conservadas entre elas, distribuição dos motivos encontrados pelo MEME, bem como a árvore fenética, mostrando o agrupamento por espécies e por classificação de proteína, confirmando e justificando o resultado dos motivos gerados pelo MEME. Em (A) gerados com a sequência nucleotídica dos domínios, em (B) gerados com a sequência nucleotídica dos genes, em (C) gerados com a sequência de aminoácidos

dos domínios e em (D) gerados com a sequência de aminoácidos das proteínas. Sequências de *F. oxysporum*, *F. graminearum* e *N. crassa* foram usadas como outgroup.

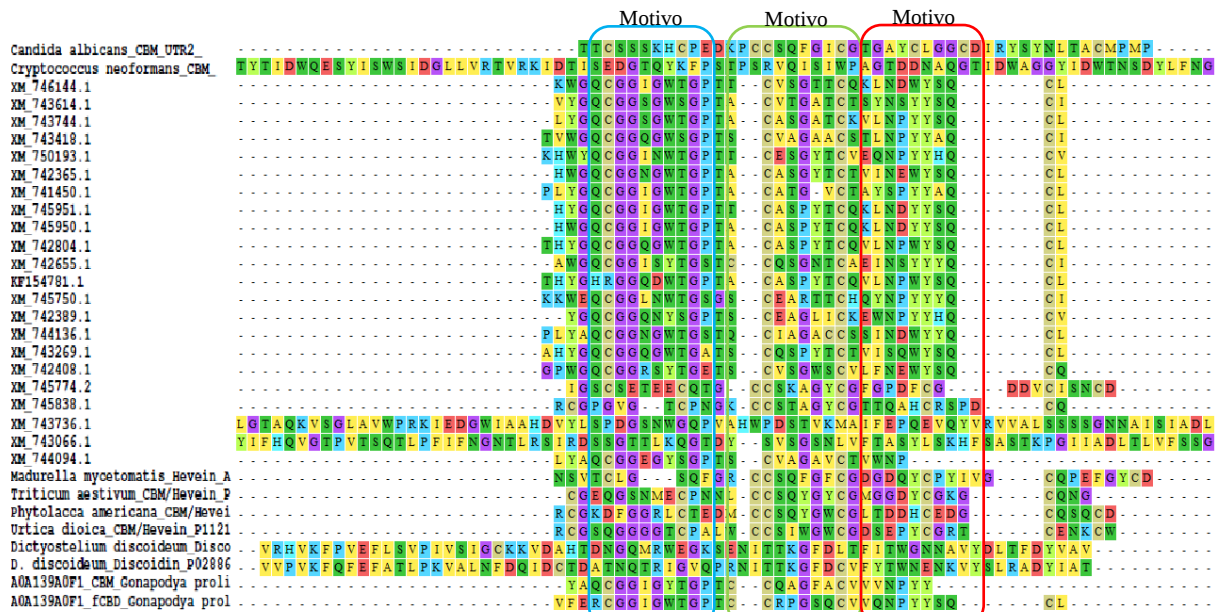
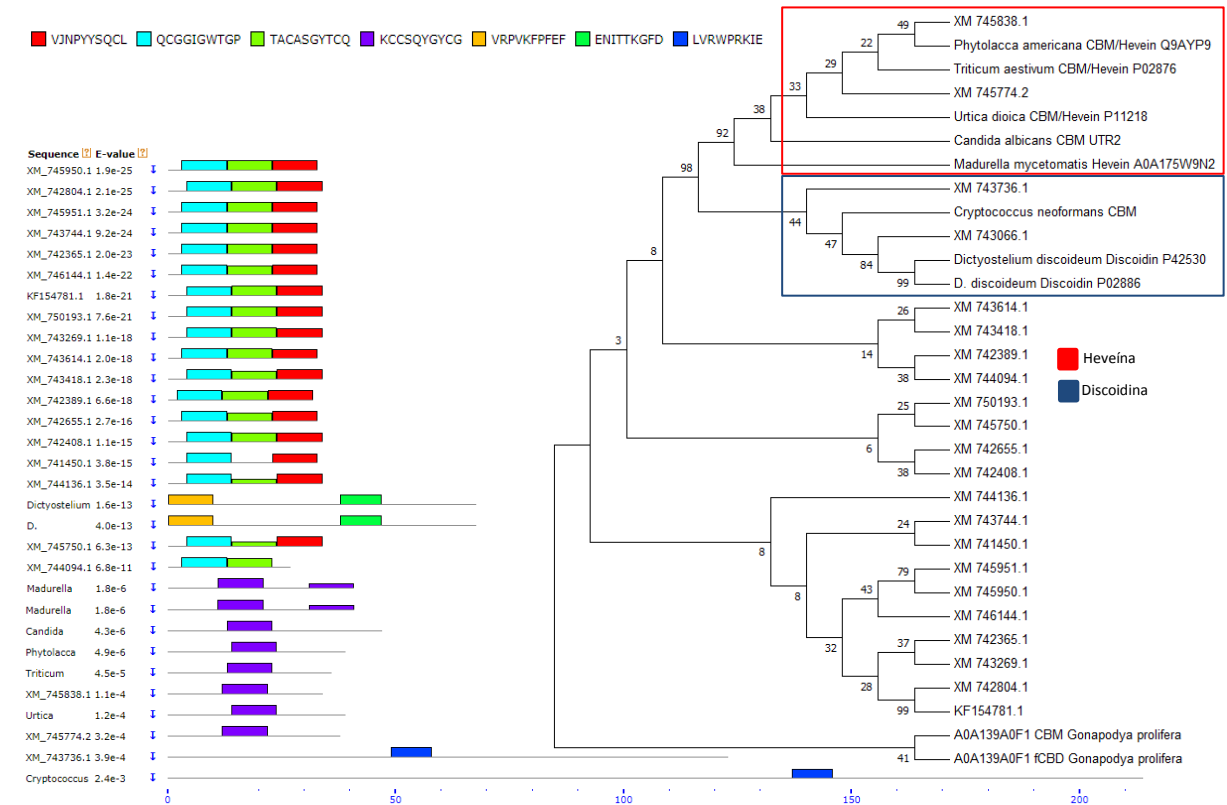
A)



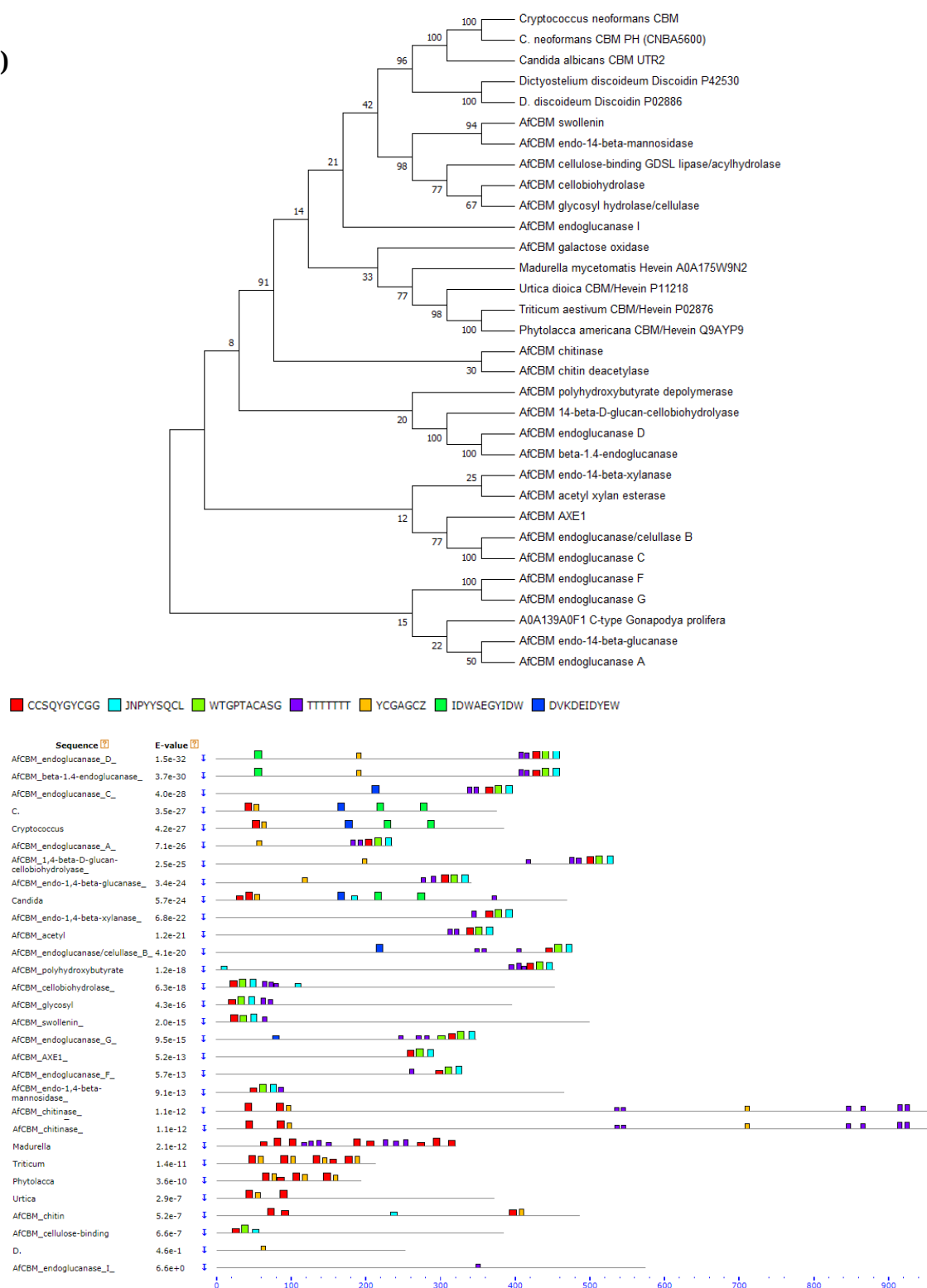
B)



C)



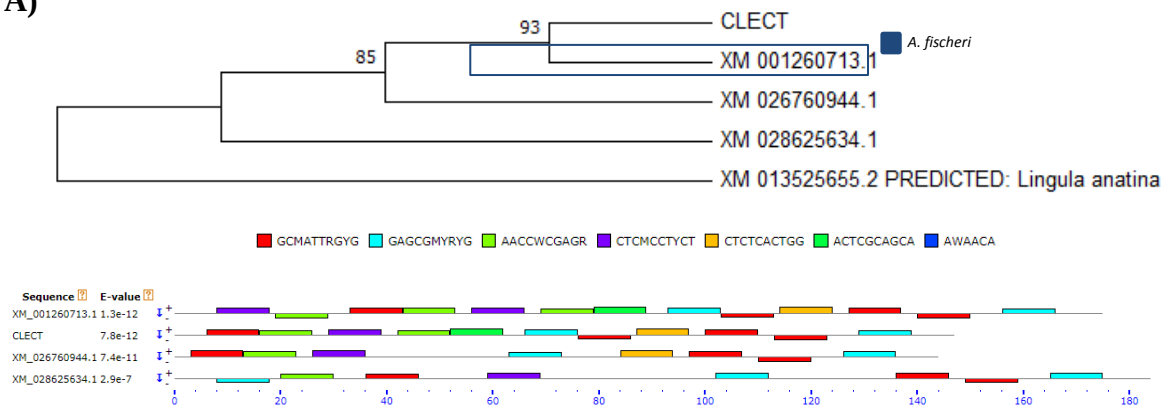
D)



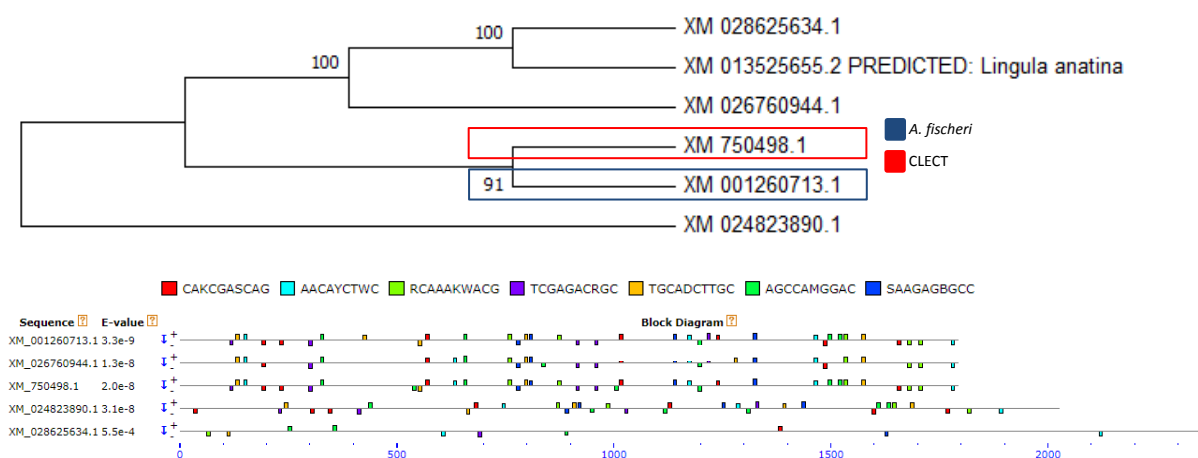
Fonte: Resultados do próprio autor, 2019.

Figura 16. Análise *in silico* da família proteica C-type. Alinhamento múltiplo entre *A. fumigatus* e espécies próximas (já descritas na literatura), mostrando as regiões mais conservadas entre elas, distribuição dos motivos encontrados pelo MEME, bem como a árvore fenética, mostrando o agrupamento por espécies e por classificação de proteína, confirmando e justificando o resultado dos motivos gerados pelo MEME. Em (A) gerados com a sequência nucleotídica dos domínios, em (B) gerados com a sequência nucleotídica dos genes, em (C) gerados com a sequência de aminoácidos dos domínios e em (D) gerados com a sequência de aminoácidos das proteínas.

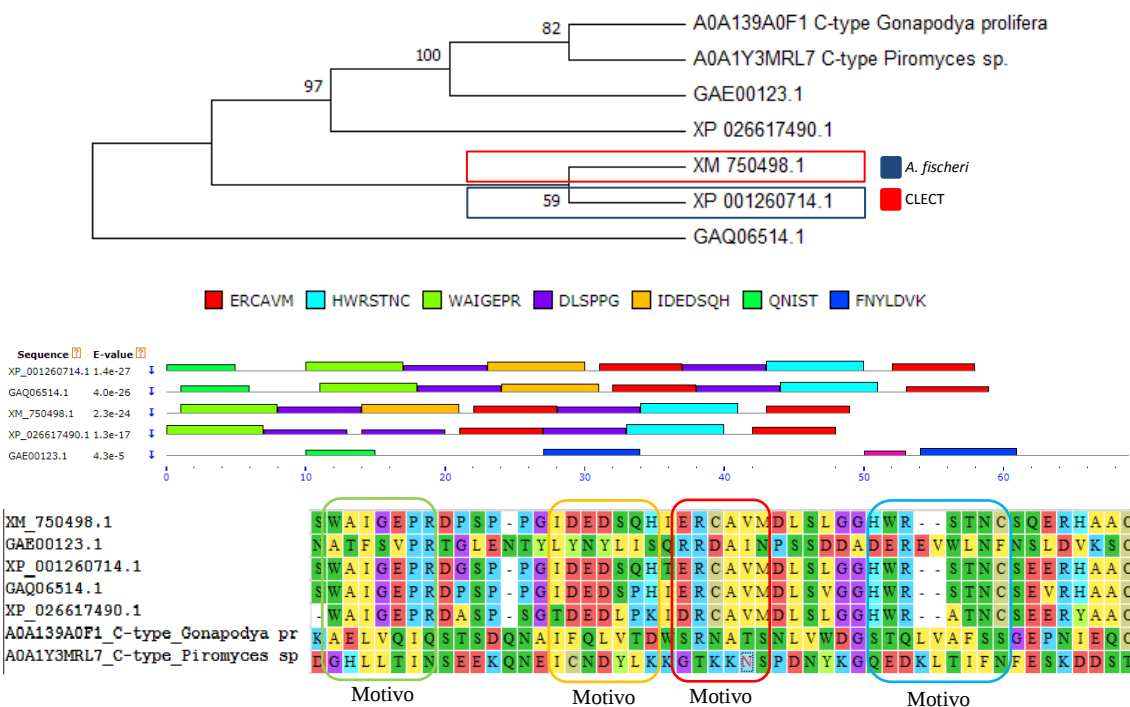
A)



B)



C)



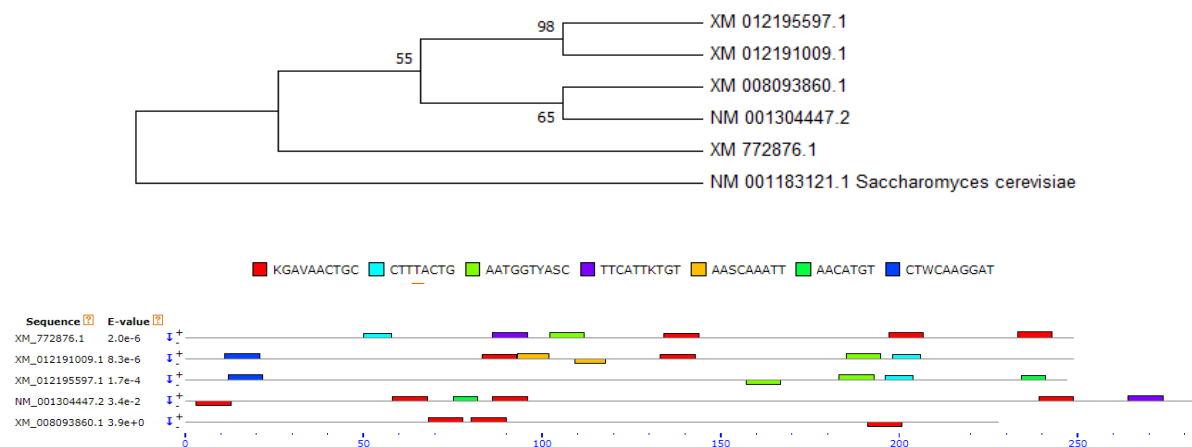
D)



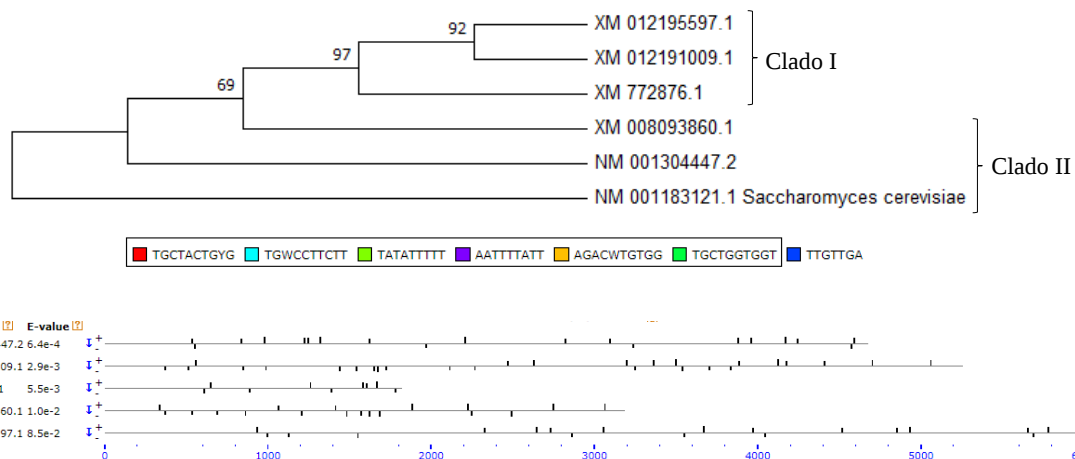
Fonte: Resultados do próprio autor, 2019.

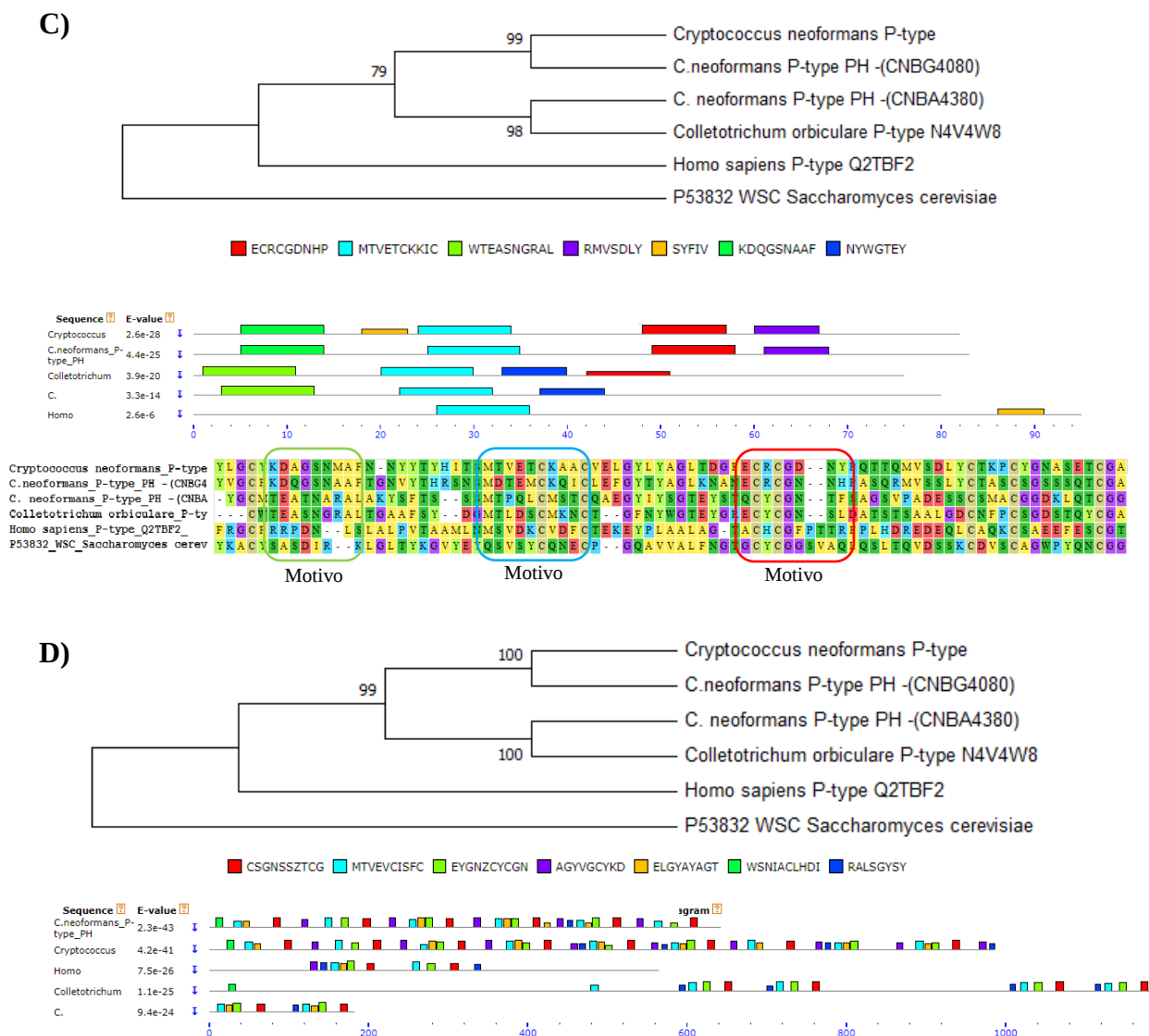
Figura 17. Análise *in silico* da família proteica P-type. Alinhamento múltiplo entre *C. neoformans* e espécies próximas (já descritas na literatura), mostrando as regiões mais conservadas entre elas, distribuição dos motivos encontrados pelo MEME, bem como a árvore fenética, mostrando o agrupamento por espécies e por classificação de proteína, confirmando e justificando o resultado dos motivos gerados pelo MEME. Em (A) gerados com a sequência nucleotídica dos domínios, em (B) gerados com a sequência nucleotídica dos genes, em (C) gerados com a sequência de aminoácidos dos domínios e em (D) gerados com a sequência de aminoácidos das proteínas. Sequências de *S. cerevisiae* foi usada como outgroup.

A)



B)

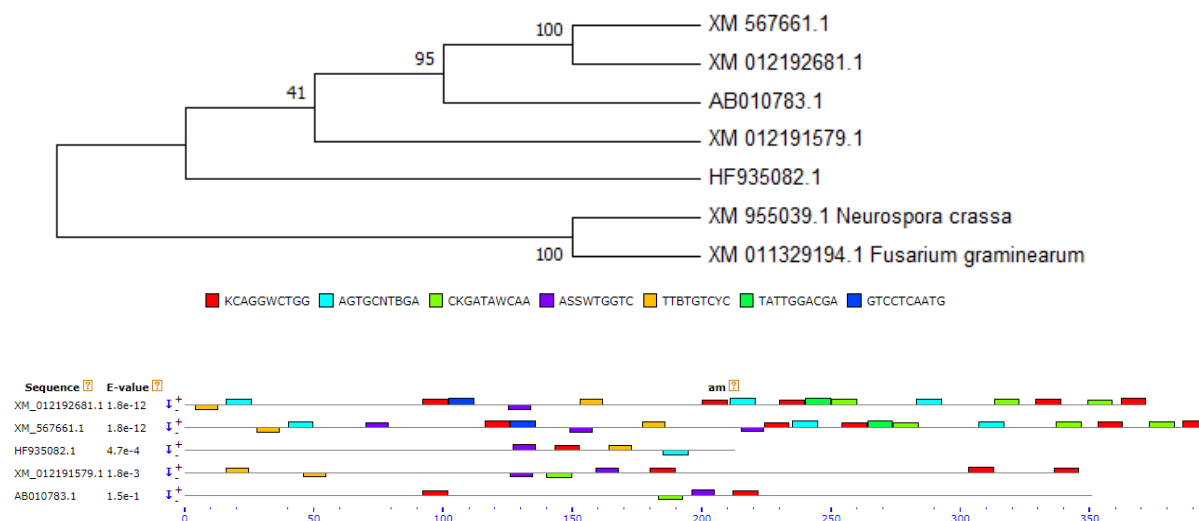




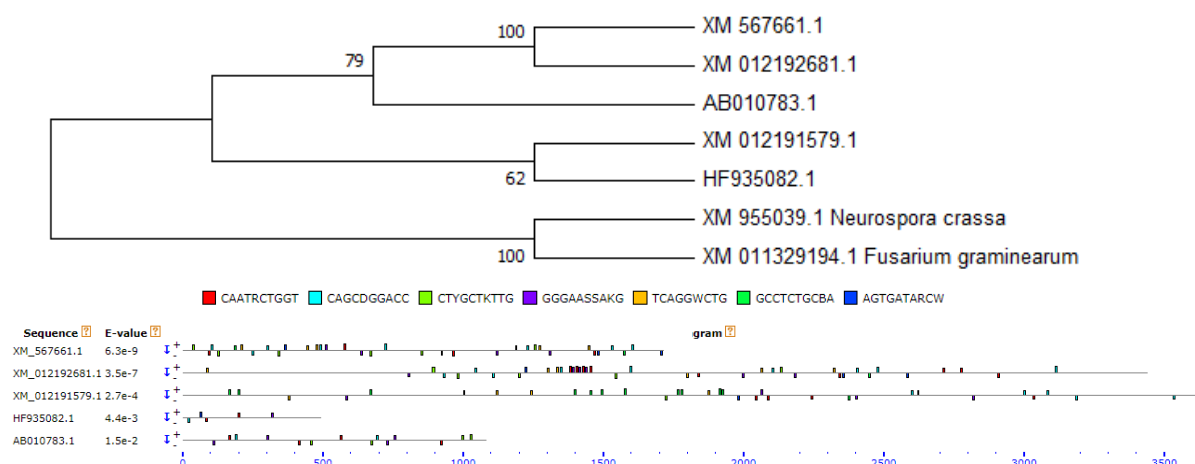
Fonte: Resultados do próprio autor, 2019.

Figura 18. Análise *in silico* da família proteica R-type. Alinhamento múltiplo entre *C. neoformans* e espécies próximas (já descritas na literatura), mostrando as regiões mais conservadas entre elas, distribuição dos motivos encontrados pelo MEME, bem como a árvore fenética, mostrando o agrupamento por espécies e por classificação de proteína, confirmando e justificando o resultado dos motivos gerados pelo MEME. Em (A) gerados com a sequência nucleotídica dos domínios, em (B) gerados com a sequência nucleotídica dos genes, em (C) gerados com a sequência de aminoácidos dos domínios e em (D) gerados com a sequência de aminoácidos das proteínas. Sequências de *F. graminearum* e *N. crassa* foram usadas como outgroup.

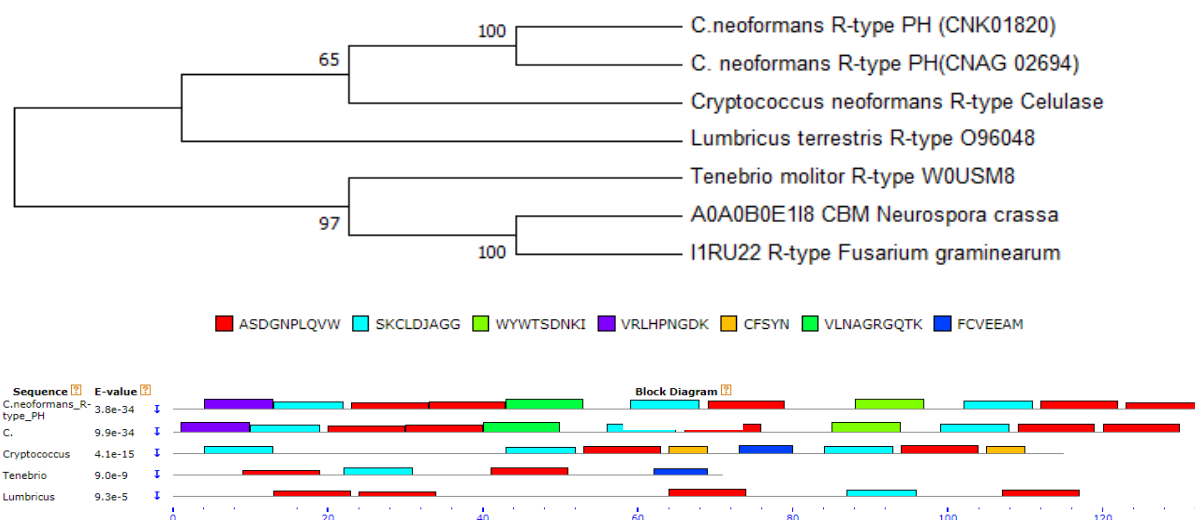
A)

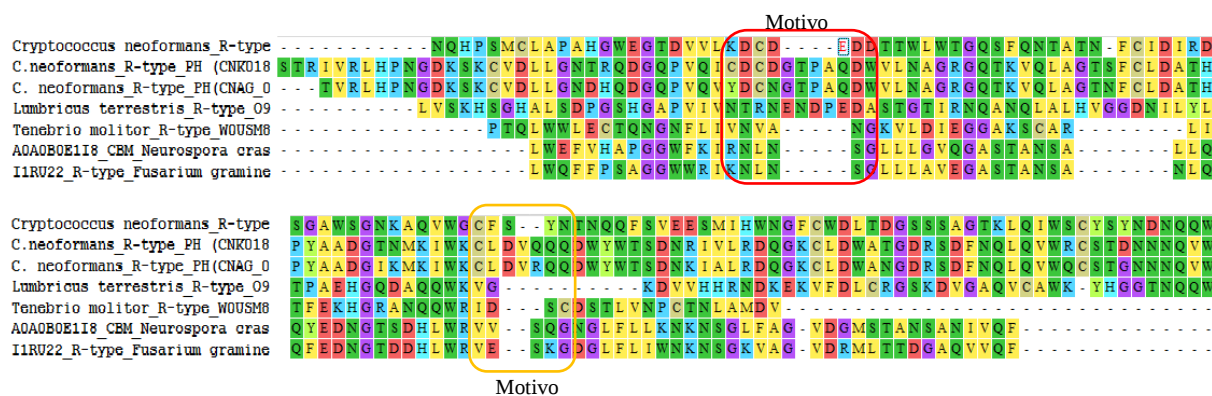


B)

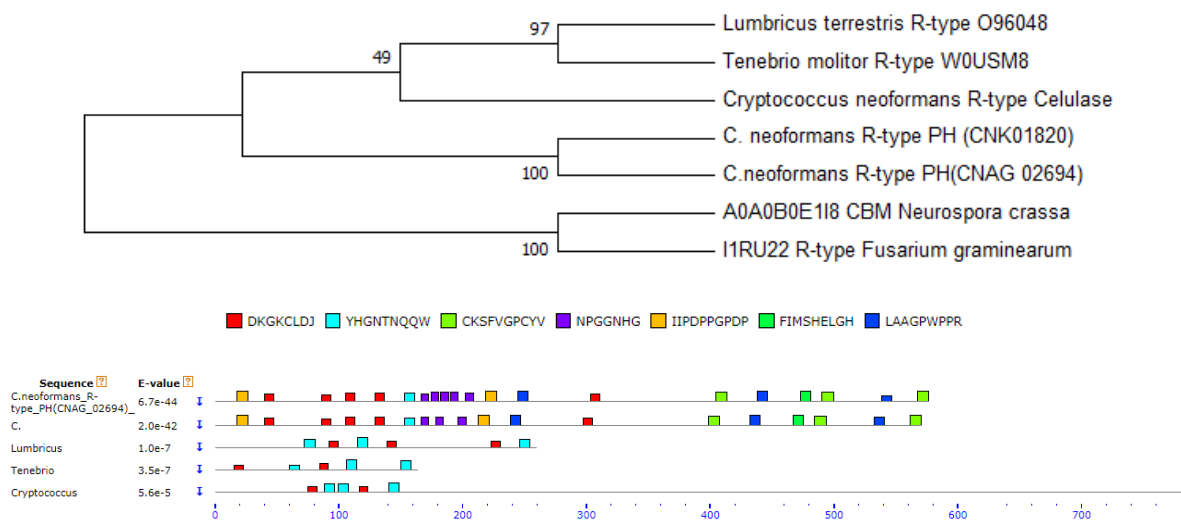


C)





D)



Fonte: Resultados do próprio autor, 2019.

As análises de domínio e fenética da família FLO, específica para *C. albicans* (**Figura 13**), mostrou que suas sequências se agruparam próximas entre si, evidenciando dois clados separados entre as sequências de *C. albicans* e as de *S. cerevisiae*. Estes clados são mais facilmente visualizados a nível genômico, e clados por classificação, quando analisadas a nível de sequência de aminoácidos, em proteínas do tipo FLO11 e proteínas do tipo FLO9, confirmando e justificando o resultado dos motivos gerados pelo MEME. Pode-se observar que a PH de *C. albicans* (XM_709370.2) apresentou domínio e motivos FLO conservados, e se agrupou com as sequências YWP1 de *C. albicans* (XM_709369.2), uma floculina que apresenta domínio FLO9, sugerindo que a PH seja também uma FLO9. Proteínas do tipo floculinas, foram caracterizadas como sendo de membrana celular, parede celular, extracelular, citoplasma e núcleo e que participam diretamente de interações célula-célula adesiva durante floculação de *S. cerevisiae* e da *C. albicans*, um processo que é reversível e dependente de Ca^{2+} no qual as células aderem para formar agregados.

Para a família L-type, o alinhamento das sequências gerado pelo MEGA7 mostrou uma grande diversidade entre elas, tanto entre os microrganismos estudados quanto entre espécies do mesmo gênero (**Figura 14**). Se observarmos os resultados da fenética, veremos o agrupamento das sequências desta família por espécies e por semelhança de domínios. Com isso, olhando as análises da PH de *C. albicans* (XM_710947.2) de forma geral, podemos sugerir que ela seja uma Concanavalin-A-like lectin/glucanase com domínios conservados bastante semelhantes aos fungos *A. fumigatus*, *C. neoformans* e *C. tenuis*. A PH de *A. fumigatus* (AFUB_075490) por sua vez se agrupou próxima à sequência de *S. cerevisiae* (NM_001181967.1), que compartilham o mesmo domínio Laminin_G. Já a PH de *C. neoformans* (XM_768872.1) se agrupou com outra L-type do mesmo fungo (XM_572915.1) e próxima também da PH de *C. albicans* (XM_710947.2), ambas com alta similaridade. As sequências 2180 e 6800 se agruparam sempre próximas às sequências de *N. crassa* e *F. oxysporum*, respectivamente.

O alinhamento múltiplo dos domínios conservados em CBM entre espécies próximas mostrou as regiões mais conservadas entre elas. Os motivos encontrados pelo MEME entre essas sequências mostrou uma alta similaridade entre a PH de *C. neoformans* (XM_772770.1) e outra sequência CBM da mesma levedura (XM_566898.1), sugerindo que elas compartilham a mesma função. Na fenética as sequências se agruparam conforme o tipo específico de domínio: domínios tipo CBM, fCBD, Discoidinas e Heveínas, confirmando o agrupamento dos motivos gerados pelo MEME (**Figura 15**).

A sequência de *A. fumigatus* que codifica para proteína com domínio CLECT (XM_750498.1) se agrupou próximo à sequência de *A. fischeri* em todas as análises, sugerindo que elas possam desempenhar funções semelhantes (**Figura 16**). Já as sequências de *C. neoformans* que codificam para proteínas com domínio WSC (XM_012195597.1, XM_012191009.1 e XM_772876.1) se agruparam próximas umas das outras (clado I), enquanto que as sequências de espécies próximas permaneceram agrupadas em um segundo clado (**Figura 17**). A PH XM_772876.1 agrupou-se sempre em proximidade com a sequência do fungo *Colletotrichum orbiculare*.

As sequências de *C. neoformans* (XM_012191579.1, PH XM_567661.1 e PH XM_012192681.1) agruparam-se entre si, compartilhando poucas regiões similares com

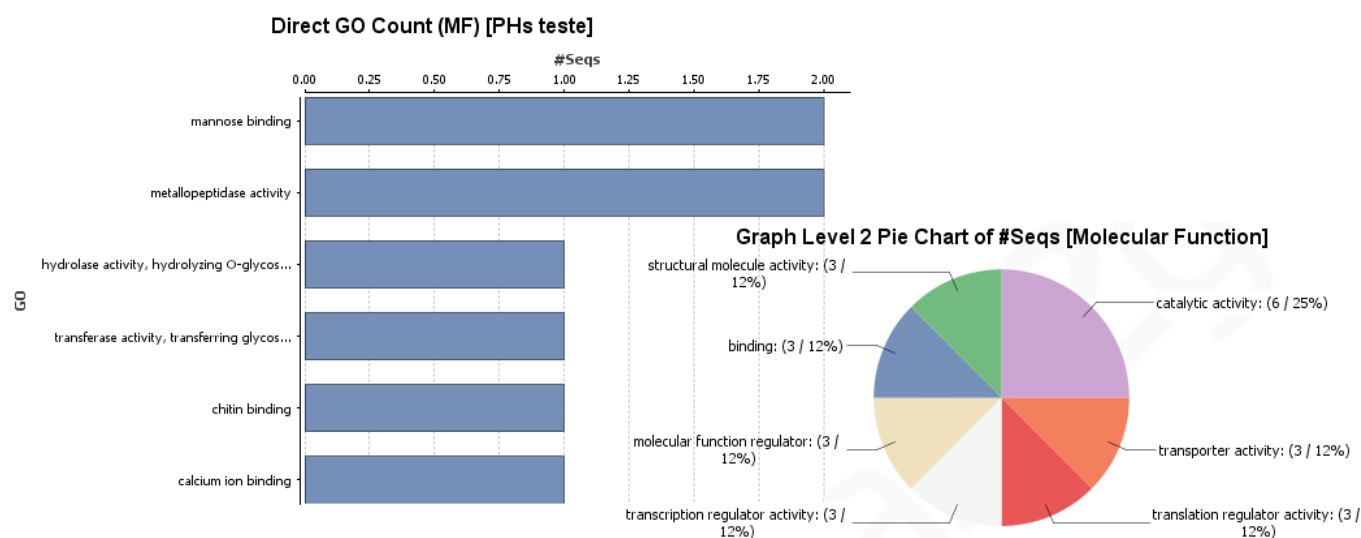
as sequências de espécies próximas, sugerindo que elas sejam mais conservadas entre espécies do mesmo gênero (**Figura 18**).

Ao final da análise de domínios, foram identificadas uma PH conservada em *A. fumigatus* (L-type), duas em *C. albicans* (FLO e L-type) e seis em *C. neoformans* com domínio lectina (uma L-types, duas P-type, duas R-type e uma CBM), a partir de dados de sequência disponíveis no banco de dados da Uniprot, NCBI, CGD e AspGD e graças às ferramentas bioinformáticas, tais como BLAST, Batch CD-Search, Cell-Ploc 2.0, JVirGel, MEME, MEGA7 e JalView, foi possível identificar características estruturais, suficientes para nos permitir sugerir a função de muitas das proteínas hipotéticas estudadas através do Blast2GO, vide **Tabela 6**.

5.3 ANÁLISES DE ANOTAÇÃO DE ONTOLOGIA PELO BLAST2GO

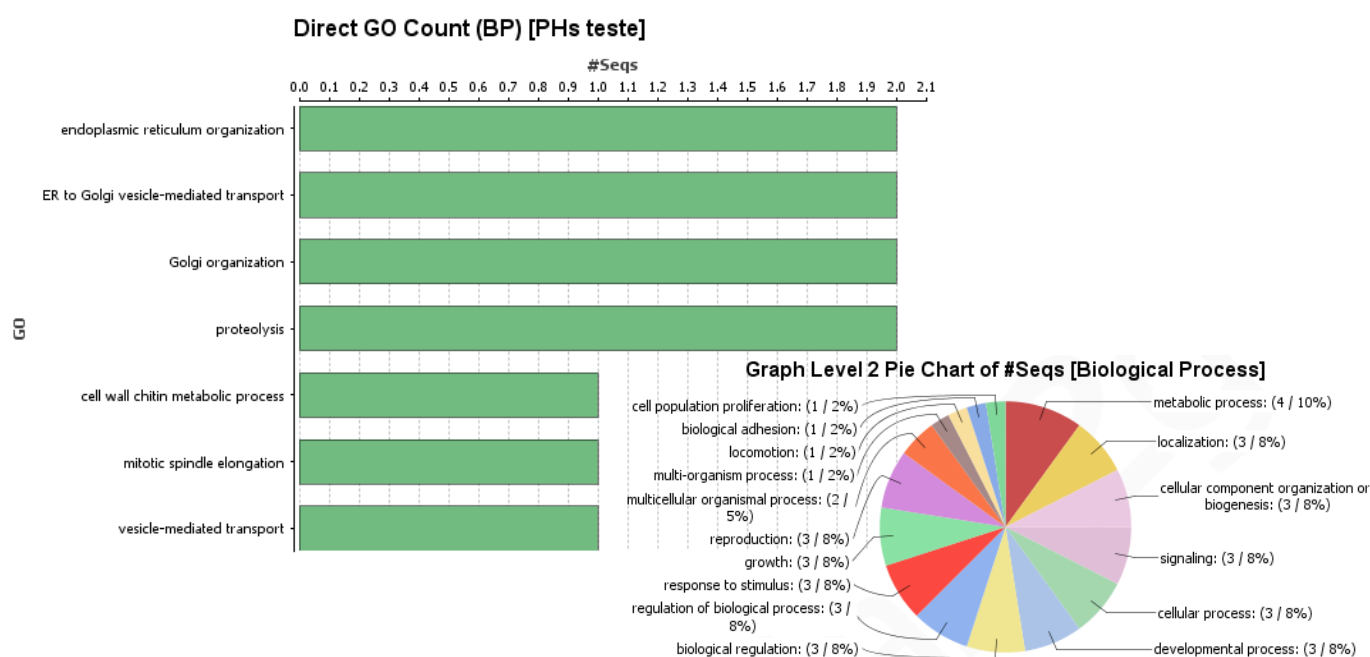
O arquivo FASTA, contendo as nove sequências das PHs conservadas identificadas no presente estudo, foi submetido ao programa Blast2GO (versão gratuita), que utiliza diversas ferramentas a fim de se caracterizar e anotar sequências de genes e proteínas. A primeira etapa da análise consistiu na busca de sequências similares utilizando o programa BLASTp. Essa pesquisa por homologia foi feita no banco de dados público (NCBI). Encontradas as sequências similares às proteínas hipotéticas conservadas do presente estudo, os dados foram submetidos ao mapeamento dos termos do GO. Das 9 PHs, oito foram classificadas com respectivas funções moleculares e os resultados mostrados na **figura 19**. Os resultados do processo biológico para todas as nove PHs foram mostrados na **figura 20** e os de componente celular mostrados na **figura 21**, mostrando ainda que algumas delas podem estar presentes em mais de um local na célula. As PHs foram classificadas em uma ampla gama de processos, indicando que elas participam de vários processos celulares conhecidos e também podem participar de outros processos descritos.

Figura 19. Resultados para o mapeamento de oito PHs pelo Blast2GO- em relação à função molecular.



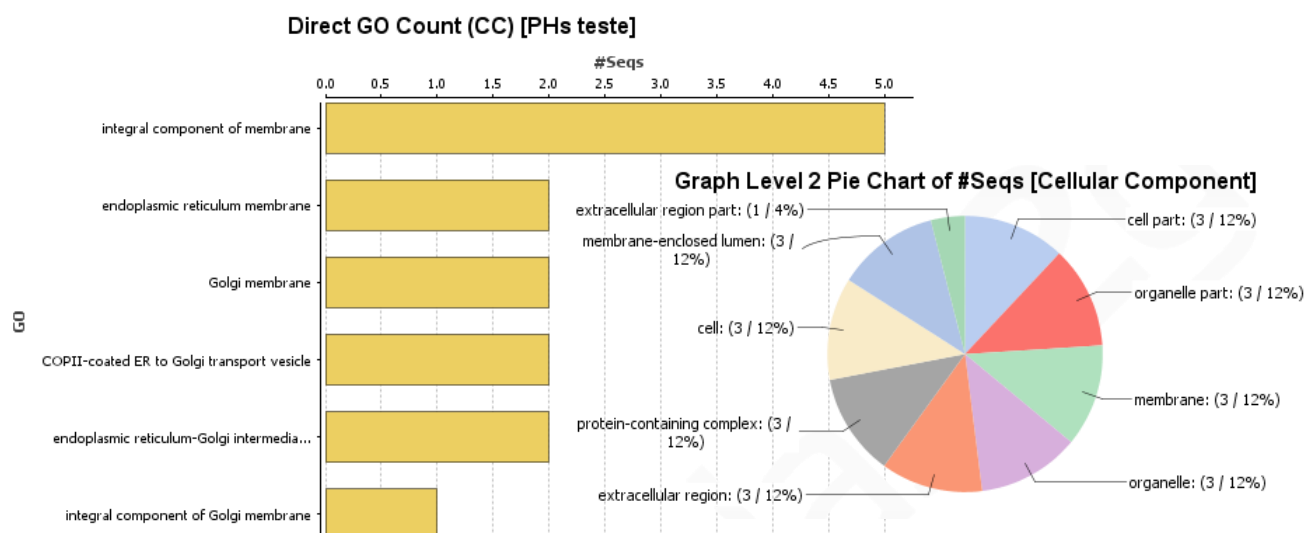
Fonte: Resultados do próprio autor, 2019.

Figura 20. Resultados para o mapeamento das 9 PHs pelo Blast2GO- em relação à processo biológico.



Fonte: Resultados do próprio autor, 2019.

Figura 21. Resultados para o mapeamento das 9 PHs pelo Blast2GO- em relação à componente celular, mostrando que algumas delas podem estar presentes em mais de um local na célula.

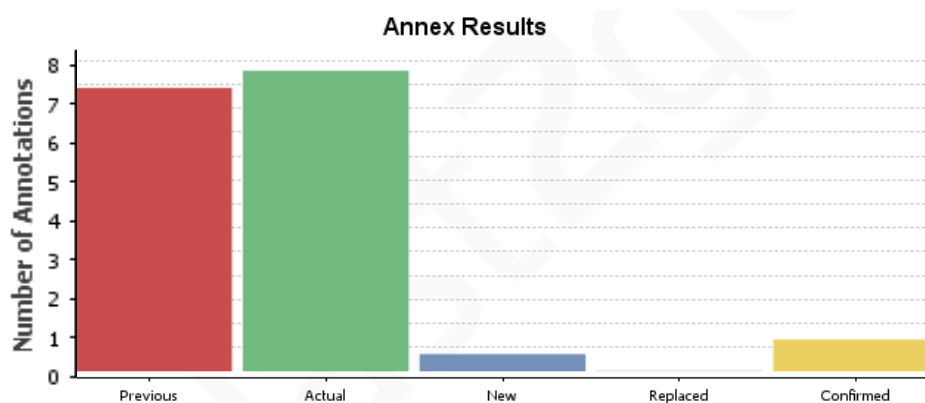


Fonte: Resultados do próprio autor, 2019.

Para cada proteína foram obtidos mais de um termo em cada uma das categorias, desde a mais abrangente, na qual ela está inserida até a mais específica que determina de fato sua função. Nesta análise, o nível de classificação de cada proteína foi escolhido de forma que houvesse equilíbrio entre especificidade e sensibilidade. Assim escolheu-se fazer as comparações no nível dois, já que, para alguns casos, níveis mais altos torna a análise muito específica, reduzindo a quantidade de proteínas representadas em cada gráfico. Em um gráfico, algumas proteínas podem ser representadas mais de uma vez por obter mais de um termo, portanto a quantidade de termos e proteínas visualizados nos gráficos não está diretamente relacionada à quantidade de proteínas analisadas neste trabalho. Dentre as 9 PHs conservadas analisadas neste trabalho, oito foram classificadas dentro da categoria Função Molecular, e três com função molecular de ligação à carboidratos, como pode ser observado na **figura 19**. Dentro da categoria de Processos Biológicos, mostrados na **figura 20**, foram classificadas todas as nove PHs, muitos deles relacionados com o processo de adesão e invasão destas leveduras nas células do hospedeiro. Outra categoria que foi detectada foi a de componentes celulares visualizado na **figura 21** e que conseguiu classificar todas as PHs.

Realizadas as etapas de busca de similaridade, mapeamento dos termos do GO e busca de domínios, o Blast2GO por meio da ferramenta ANNEX infere, a partir dos resultados obtidos na categoria função molecular, anotações para as categorias Processos Biológicos e Componentes Celulares, com a finalidade de expandir e melhorar as anotações por proteína. Utilizando o programa ANNEX, inferimos uma nova anotação, como pode ser observado na **figura 22**.

Figura 22. Resultados da anotação após a análise de expansão de anotação.



Fonte: Resultados do próprio autor, 2019.

5.4. ANÁLISE *IN VITRO* DAS SEQUÊNCIAS

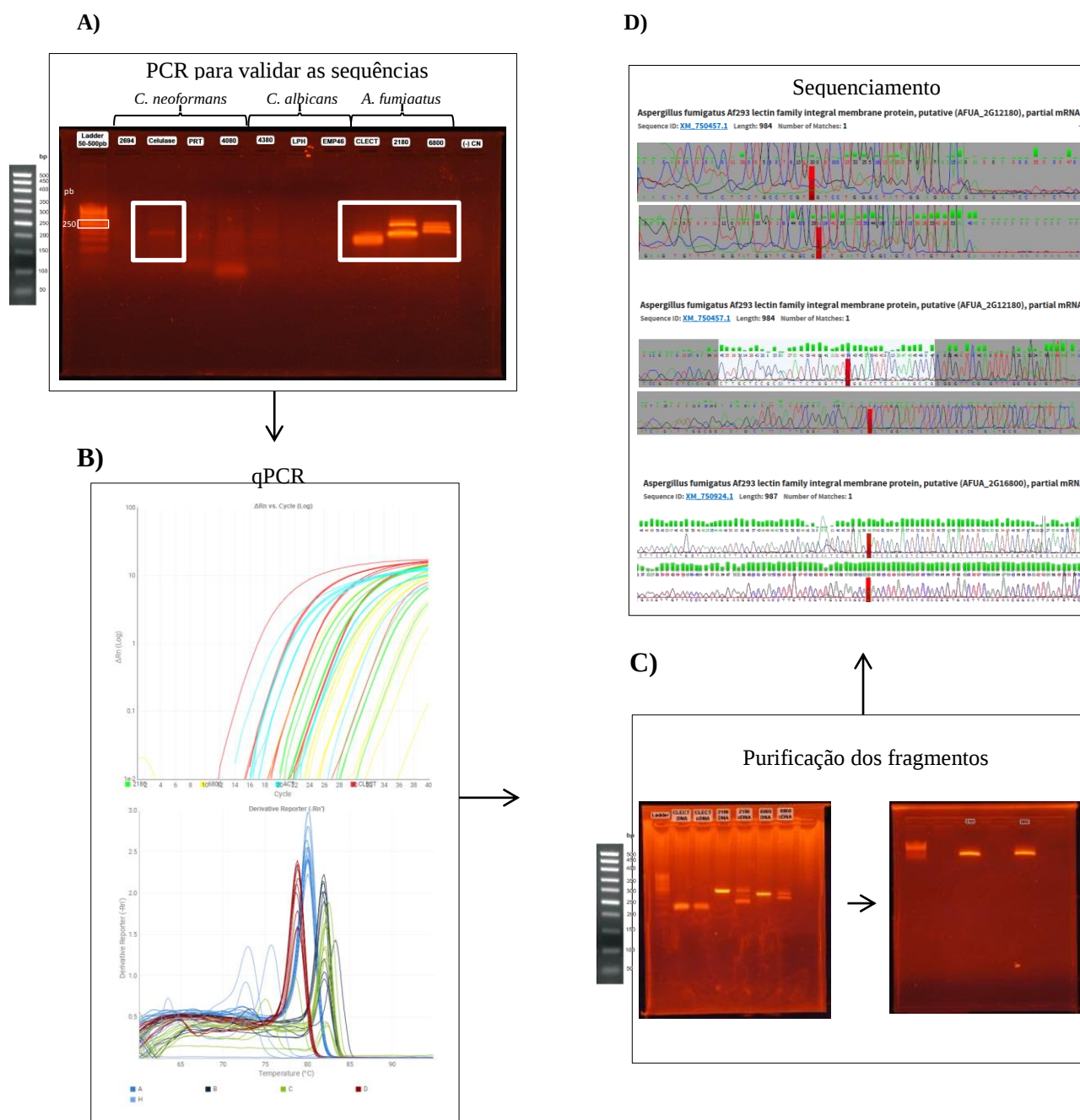
Com o intuito de avaliar o perfil transcricional das sequências coletadas das etapas anteriores e selecionar aquela que prosseguirá para a etapa de modelagem e dinâmica, 16 conjuntos de iniciadores de reação de PCR para os três fungos estudados foram postos à validação via PCR convencional utilizando-se DNA genômico total, onde quatro genes mostraram amplificação, dentre eles um codificador de uma PH em *C. neoformans* e outros três genes que codificam proteínas putativas não caracterizadas em *A. fumigatus*. A amplificação foi confirmada através da visualização dos fragmentos de 80 pb (CLECT), 77 pb (ACTIN-1), 121 pb (2180) e 154 pb (6800) após eletroforese (**Figura 23.A**).

A fim de confirmar a expressão gênica, os experimentos de RT-PCR foram realizados a partir do RNA total, extraído do crescimento da linhagem fúngica de *A. fumigatus* (cepa 6573) após sete dias de inoculo em meio YPD. O cDNA foi sintetizado pela enzima transcriptase reversa, utilizando primers específicos para a sequência de interesse. As sequências de *A. fumigatus* (2180, 6800 e CLECT) puderam ser confirmadas em gel de agarose 1,8 %, indicando que os três genes estão sendo expressos e originando RNAm para tais proteínas (**Figura 23.C**).

Os resultados da PCR em tempo real (qPCR) corroboraram com os dados anteriores, mostrando as expressões das triplicatas para os quatro genes de *A. fumigatus* (ACTIN-1 em azul, CLECT em vermelho, 2180 em verde e 6800 em amarelo),

confirmando a expressão destas proteínas. (**Figura 23.B**). Os fragmentos amplificados por qPCR foram purificados à partir de gel de agarose, e posteriormente os mesmos foram sequenciados. As sequências nucleotídicas obtidas foram analisadas utilizando o software CLC Main Workbench v.6 (Quiagen). Os resultados gerados pelo sequenciamento (CLECT / 2180 / 6800) foram posteriormente alinhados via BLASTx (NCBI) a fim de identificar um ou mais trechos destas sequências de entrada que fossem semelhantes às sequências presentes no banco de dados em questão (NCBI) e estão resumidos na **Figura 23**.

Figura 23. Principais etapas da análise *in vitro* das sequências. Em **A**, PCR com DNA genômico amplificando um fragmento para *C. neoformans* (PM 158pb) e três para *A. fumigatus* (PM 80pb, 121pb, 154pb, respectivamente). **B**, PCR em tempo real, mostrando as expressões das triplicatas para os quatro genes de *A. fumigatus* (ACTIN-1 em azul, CLECT em vermelho, 2180 em verde e 6800 em amarelo). **C**, Bandas purificadas do cDNA de 2180 e 6800. **D**, Sequenciamento dos fragmentos de *A. fumigatus* (CLECT, 2180 e 6800).



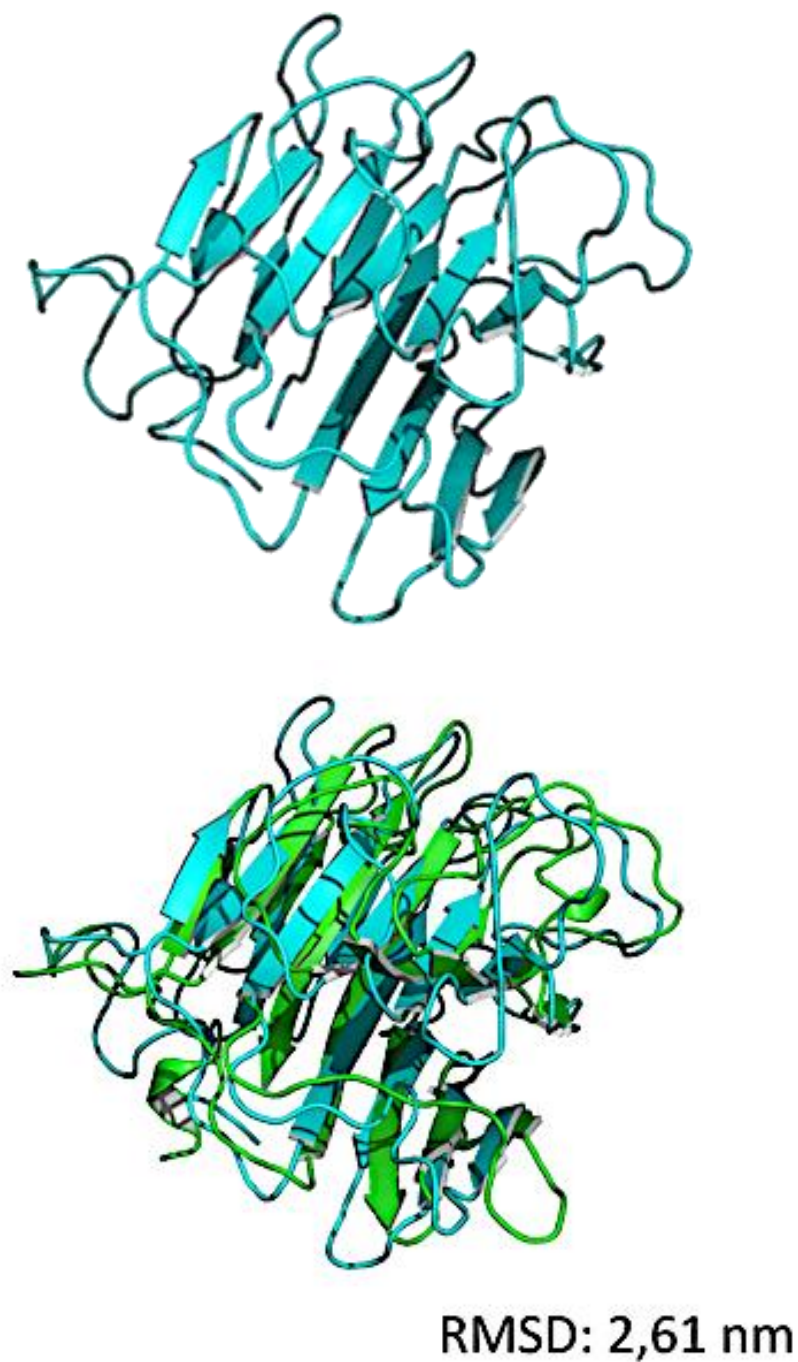
5.5. MODELAGEM E DINÂMICA DE LECTINA DE *A. fumigatus*

A sequência gênica (XM_750457.1) contida no cromossomo 2 de *A. fumigatus* codifica para uma proteína de 327 resíduos do tipo L-type e apresenta localização extracelular, no complexo de golgi e retículo endoplasmático. Seu domínio conservado foi escolhido para as etapas seguintes. A modelagem das sequências foi baseada em estruturas resolvidas disponíveis no *Protein Data Bank* (PDB) usando o programa Modeller 9.23 (**Figura 24**). Para AfLtype1, o melhor modelo selecionado foi uma lectina de *Canis lupus familiaris* (código PDB 2DUP), compartilhando 46,67% da identidade. Após análise estereoquímica dos aminoácidos, foi possível selecionar um modelo que apresentava $\geq 90\%$ dos aminoácidos em posições favoráveis e o *Z-score* indicou que todo o modelo estava dentro da faixa de escores normalmente encontrado para proteínas de tamanho semelhante (**Figura 25**).

Figura 24. Modelos tridimensionais de AfLtype1. **A.** Modelo estrutural antes (em verde) e depois (em azul) da simulação em dinâmica molecular, bem como o alinhamento das duas estruturas. **B.** modelo de AfLtype1 colorido pela carga residual. Resíduos não polares (branco), resíduos básicos (azul), resíduos ácidos (vermelho) e resíduos polares (verde), sendo a primeira antes e a segunda depois. Fator de temperatura (*b-factor*) do modelo de AfLtype1 mostrando a mobilidade dinâmica dos átomos em **C**.

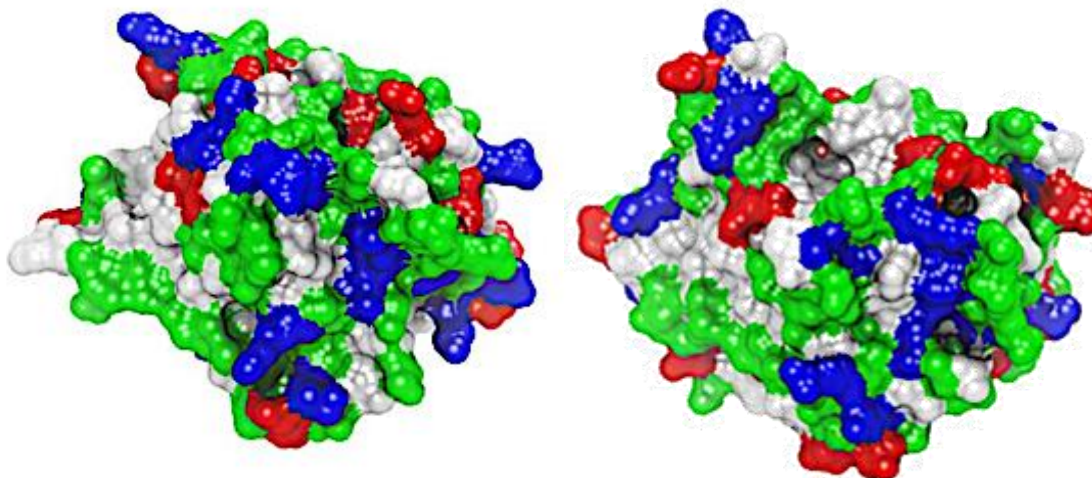
A



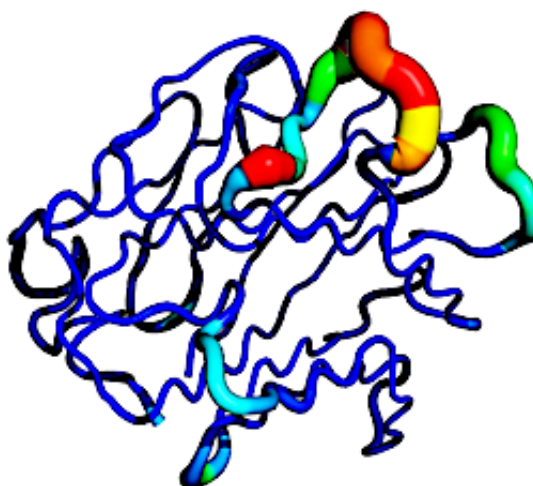


A **Figura 24.A** mostra que o núcleo (*backbone*) da estrutura tridimensional permaneceu conservado após a simulação em dinâmica molecular feita com NaCl 0,15 M. Ao observar a carga residual (**Figura 24.B-C**), foi possível perceber que o modelo de AfLtype1 possui regiões estruturais anfipáticas intercaladas por resíduos catiônicos e aniônicos.

B



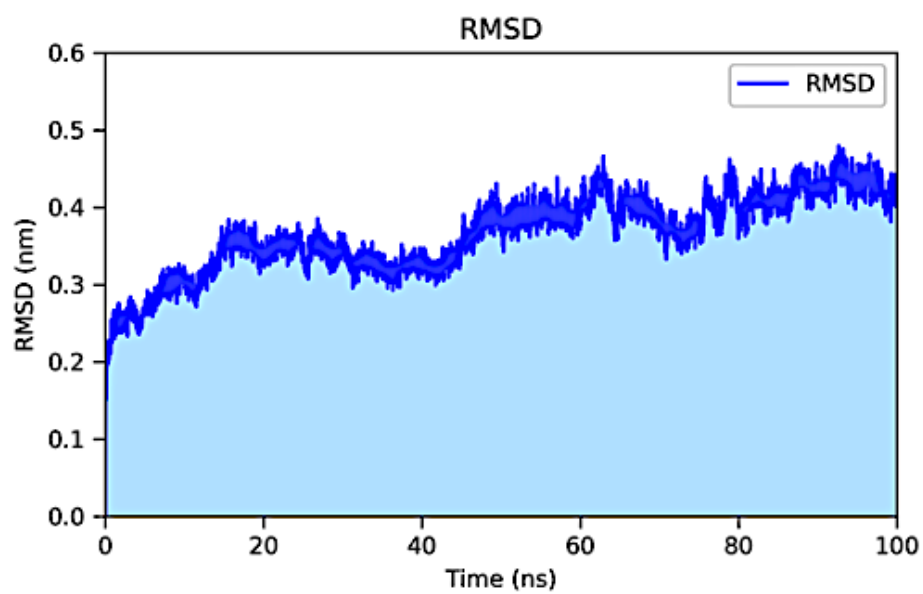
C



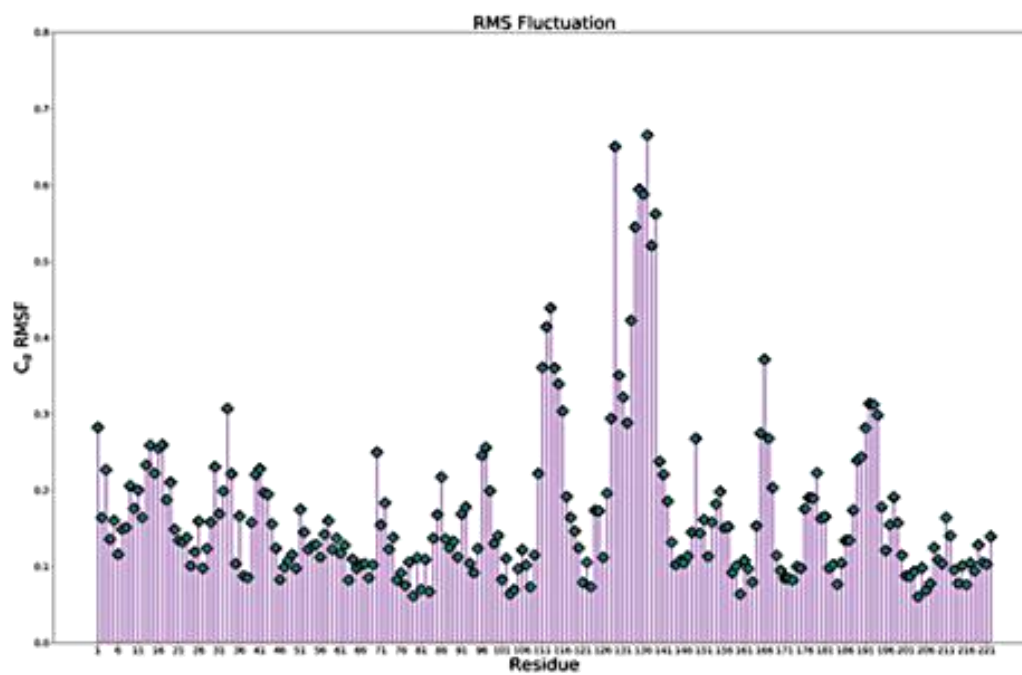
Na simulação em dinâmica molecular, a fim de observar as possíveis modificações que o modelo poderia sofrer, foram avaliados os gráficos de *root mean square deviation* (RMSD), *root mean square fluctuation* (RMSF), o raio de giro e a estrutura secundária (**Figura 25**).

Figura 25. Gráficos relacionados à simulação em dinâmica molecular do modelo tridimensional de AfLtype1. Em **A** o RMSD do modelo com o tempo de 100 ns. Em **B** o RMSF do modelo representado pela flutuação de aminoácidos pelo tempo em 100 ns. Em **C** o raio de giro do modelo em relação ao tempo de simulação. Em **D** o DSSP do modelo mostrando em vermelho a conservação da estrutura secundária (folha- β) durante a simulação em dinâmica molecular em 100 ns.

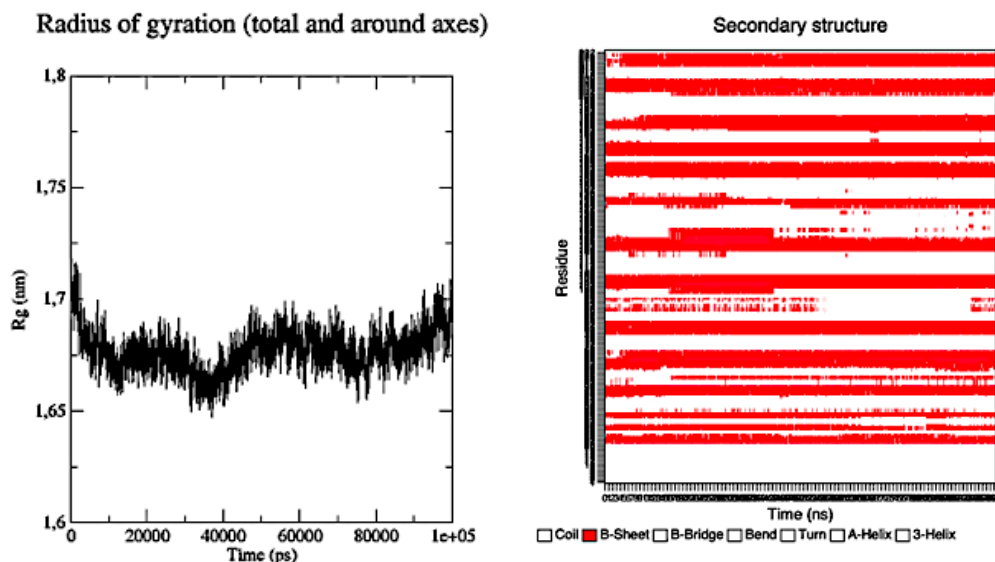
A



B



C



Analisando os gráficos é possível observar que o modelo saiu da conformação original, contudo se mostrou estável na maior parte da simulação em dinâmica molecular. Quando foi analisada a flutuação dos aminoácidos (**Figura 25.B**), foi possível verificar que os aminoácidos que mais flutuaram estavam entre os aminoácidos 110 – 140, de fato, o b-factor, mostrado na **Figura 25.B**, revelou que essas regiões as quais tiveram maior mobilidade dinâmica dos átomos (mais próximo do vermelho) estavam localizadas normalmente em regiões sem estrutura secundária (loop), que normalmente possui um grau de flexibilidade maior para flutuação e que quanto mais próximo do azul, menor mobilidade dinâmica dos átomos.

O gráfico do raio de giro (**Figura 25.C**), comprova que a estrutura manteve-se compacta corroborando a hipótese de que as ligações entre as regiões de folha- β em forma de concha mantiveram-se rígidas enquanto que as outras regiões sem estruturas secundárias foram mais flexíveis. Além disso, o gráfico do DSSP (**Figura 25.C**) mostra que as regiões de folha- β mantiveram-se estáveis com a manutenção de sua estrutura, havendo apenas diminuição no tamanho de algumas regiões nas extremidades da estrutura.

O servidor Philius (disponível em <http://www.yeastrc.org/philius/runPhilius.do>) determinou a posição do sítio de clivagem entre os aminoácidos 23 e 24 (**Figura 26.A**). O índice de hidropatia de um aminoácido (GRAVY) da AfLtype1 revelou que a maior quantidade de seus aminoácidos apresentam números hidropáticos negativos (**Figura 26.B**), indicando que a maior parte da proteína seja hidrofílica. Através do alinhamento entre as sequências de AfLtype1 e do modelo referência (PDB: 2DUP), confirmou-se a presença das quatro alças (loop) com seus respectivos aminoácidos invariantes, bem como a região residual de ligação à carboidratos (**Figura 26.C**)

Figura 26. Sequências de aminoácidos da proteína putativa AfLtype1, produzida a partir de dados gerados no Philius em **A**. Em **B**, Domínio funcional pelo Interproscan, local de clivagem de peptídeos sinalizadores (PS) e nível de índice de aminoácidos hidropáticos (GRAVY) pelo ProtoParam. Em **C**, locais de ligação do glicano composto por quatro voltas (loop) com quatro aminoácidos invariantes essenciais para a ligação de carboidratos. Volta evidenciada em verde (contém um aspartato invariante), em amarelo (contém uma glicina ou arginina), em azul (contém uma asparagina invariante e um aminoácido hidrofóbico invariante) e em laranja (cujas diferenças nas especificidades são provocadas por sua variabilidade na conformação e tamanho).

A)



B)

Família de Lectina	ID	Domínio	GRAVY	PS	Tamanho da proteína (aa)
L-type	XM_750457.1	IPR005052	-0.355	A23-A24	327

C)

20
 2180 MLPPRLSSLLYLAGLATLPVTSAYNGDENIKS I PLRT
 40 60
 2180 HSLAPPYLDSDFQSRWFDFGGDT I IRADKY I RLTS DR
 80 100
 2180 PSQQGWIFSRVPLTATNWE I EVEFN I HGQGNLHG DGF Loop 1
 120 140
 2180 AMWL TKQRATQGPVFGSTDNFEGLG I FFDTKNNRPG Loop 2
 160 180
 2180 TSFPYVMAMMGDGKTSYDQAHDGKANEVAGCSARGLR
 200 220
 2180 GAS I PTKARLTYFQDKSLTDLQYKSEDTWTNCFTLN
 240
 2180 APETN I A I PAVSYLGFSAETGELSDNHD I I SVN AKNL Loop 4
 260 280
 2180 YS I GGNAASSSGRGPANNGRQDQAHVKLPKQKGGWGW
 300 320
 2180 FLFKF I LFLGAI VGAYVGFTMYRTKQRYSRF

6 DISCUSSÃO

Grande parcela dos genomas das leveduras em estudo codifica proteínas de função desconhecida ($\geq 70\%$), de acordo com os bancos de dados específicos (CGD e AspGD). Diante destes dados, tornam-se importantes trabalhos de caracterização de sequências, para que cada vez mais as perguntas elaboradas referentes ao processo de patogênese possam ser respondidas, e para que no futuro os dados obtidos possam ser utilizados para a elaboração de métodos eficazes de controle e de tratamento contra as doenças.

Os motivos conservados encontrados para as famílias gênicas estudadas foram poucos descritos na literatura, sugerindo que muitas dessas regiões conservadas entre elas sejam novos motivos não descritos para esses microrganismos, com funções desconhecidas, o que carece de estudos para que sejam elucidados os fatos a respeito de suas vias metabólicas, dando uma maior confiabilidade na existência funcional dessas proteínas. Neste trabalho foi observada a ocorrência de 39 genes (26 para *A. fumigatus*, 10 para *C. albicans* e 3 para *C. neoformans*) altamente conservados que codificam para proteínas não caracterizadas, e 9 genes (1 para *A. fumigatus*, 2 para *C. albicans* e 6 para *C. neoformans*) também conservados, que codificam para proteínas hipotéticas com domínio lectina conservados, ou seja, para proteínas cujas funções ainda não foram totalmente elucidadas e que serão discutidas a seguir.

Os resultados da caracterização *in silico* da PH de *C. albicans* (XM_709370.2), que se agrupou semelhantemente a uma floculina do tipo FLO9 (domínio Flocculin_t3/Hyphal_reg_CWP), corrobora com os dados encontrados para essas proteínas em estudos desenvolvidos por Willaert em 2018, apresentando o mesmo padrão de motivos e conferindo propriedades semelhantes à lectina para esta proteína, visto que repetições do tipo Flocculin_t3 são encontradas em proteínas tipo FLO9 em *S. cerevisiae* e está associada ao crescimento e virulência das hifas, a resistência à morte de neutrófilos e ligação à carboidratos (KLIS et al., 2009). Por elas apresentarem diversas localizações celulares, se destacam da parede celular das células floculentas e seletivamente ligam-se aos resíduos de manose das paredes celulares das células adjacentes. O número de repetições dos domínios FLO está relacionado com a extensão da adesão e varia bastante entre diferentes cepas de *S. cerevisiae*, quanto mais repetições, mais fortes as propriedades de adesão e maior a fração de células floculantes. Nossas análises mostraram sequências proteicas com até quatro repetições dos domínios, e a PH

apresentou duas repetições, indicando uma diversidade nos níveis de virulência em *C. albicans*. As repetições ricas em Ser/Thr nas sequências também são importantes para o direcionamento adequado da proteína na parede celular e já foi estudado em espécies de *Candida* mostrando grandes rearranjos de elementos repetidos para ALS7 (ZHANG et al., 2003; VERSTREPEN et al., 2006; GOOSSENS e WILLAERT, 2010; XIE e LIPKE, 2010). A repetição do domínio Flocculin_t3 é encontrada em proteínas do tipo FLO9, perto da região C-terminal e em várias outras proteínas de *Saccharomyces*, enquanto que seu domínio Hyphal_reg_CWP, apresenta-se na porção N-terminal da cadeia.

As três PHs com domínio L-type (lectin_leg-like e Laminin_G_3) encontradas nas leveduras *C. albicans* (XM_710947.2), *C. neoformans* (XM_768872.1) e *A. fumigatus* (AFUB_075490) foram caracterizadas e a função putativa a elas atribuídas foi determinada como semelhante a de uma proteína integral vesicular transmembranar de 36 kDa (VIP36), que atua como um receptor de carga transportando glicoproteínas do tipo manose entre o Golgi e o retículo endoplasmático (RE) (KAMIYA et al., 2005; KAMIYA et al., 2007; SATOH et al., 2007). As seis sequências das proteínas com domínio do tipo L-type em quase sua totalidade apresentaram uma faixa de peso molecular entre 30 a 55 kDa e ponto isoelétrico de 8 a 9,5, o que são características predominantes em proteínas deste tipo (NAGASUMA et al., 2001; LAGARDA-DIAZ et al., 2017). Além disso, podem ser encontradas em ampla localização celular, o que condiz com os resultados encontrados pelo Cell-Ploc para essas proteínas.

O domínio WSC presente em grandes repetições em proteínas tipo P-type, é um domínio putativo de ligação a carboidratos. O domínio contém até oito resíduos de cisteína conservados que podem estar envolvidos em pontes dissulfeto. A exoglucanase beta-1,3 de *Trichoderma harzianum* contém duas cópias do domínio WSC, enquanto que a proteína SLG1 de levedura contém apenas uma (PONTING et al., 1999; COHEN-KUPIEC et al., 1999; WAWRA et al., 2019). Esse domínio também está presente em conjunto em uma β -1,3-exoglucanase fúngica e na parede celular de *Saccharomyces cerevisiae* integridade e resposta ao estresse proteínas componentes (VERNA et al., 1997; DAHMS e HANCOCK, 2002).

As três sequências de proteínas do tipo R-type que codificam um domínio Ricin_B_Lectin em comum (XM_012191579.1, PH XM_567661.1 e PH

XM_012192681.1), são típicas das sementes da leguminosa mamona (*Ricinus communis*). A análise da estrutura primária dessas sequências mostrou a presença de um domínio similar com muitas proteínas de reconhecimento de carboidratos como toxinas-AB, glicosidases ou proteases de plantas e bactérias. Este domínio pode estar presente em uma ou mais cópias e foi mostrado, em alguns casos, que se liga a açúcares simples, como galactose ou lactose. Além disso, são proteínas de peso molecular entre 60 e 85 kDa e ponto isoeletrico entre 4 e 5, o que confirma o padrão dessas características para as duas PHs (XM_567661.1 e XM_012192681.1).

A presença de motivos e domínios de proteínas podem indicar papéis funcionais para a proteína. Embora sua ocorrência também possa estar presente em pseudogenes como resultado de duplicação gênica, a presença destes elementos na sequência proteica sugere que a proteína pode, de fato, ser expressa (DESLER et al., 2009). De forma geral, os domínios lectina exibem enorme diversidade em suas sequências (WANG e WANG, 2013). Uma das características mais importantes para proteínas com esse domínio, embora não completamente conservada, é a presença de um padrão Q-W. Consequentemente, o domínio Ricin_B_Lectin foi também referido como o domínio (QxW) e encontrado nas análises feitas pelo MEME para as três proteínas de *C. neoformans* (DGxxxQxWxC) (HAZES, 1996; HIRABAYASHI et al., 1998).

Semelhante a módulos catalíticos de glicosil hidrolases, os CBMs são divididos em 39 famílias com base na similaridade de sua sequência de aminoácidos. Em nosso estudo, as proteínas contêm os domínios CBM (ChtBD1, GH16, GH17, GH18, fCBD, F5_F8_type_C (Discoidina), CBM_X2 e CBM_1) que são proteínas que podem se ligar à N-acetilglucosamina (por exemplo, lectinas de tomate e batata, endocitinas vegetais e heveínas) e a celulose (lectinas com domínio catalítico). Esses domínios podem ocorrer em uma ou mais cópias podendo estar envolvido também no reconhecimento ou ligação de subunidades de quitina. Em algumas famílias CBM, a especificidade do ligante é única, enquanto que em outras famílias contêm proteínas que se ligam a uma variedade de carboidratos diferentes. Assim, os CBMs são excelentes modelos para estudar o mecanismo de reconhecimento de proteínas e carboidratos sendo úteis em inúmeras aplicações biotecnológicas (BORASTON et al., 2004).

As duas sequências que codificam para proteína do tipo EPA, uma em *C. albicans* (XM_714005.2) e uma em *A. fumigatus* (XM_746738.1) não apresentaram domínios

que confirmassem sua ligação a carboidratos, ao invés disso, essas sequências apresentaram um domínio TVP38 e um DUF1682, sendo este último presente em proteínas hipotéticas de função desconhecida. Uma dessas proteínas é descrita como sendo específica do adipócito em ratos, mas nenhuma evidência disso foi encontrada. O domínio TVP38 é composto de proteínas do Complexo de Golgi que estão em associação com proteínas SNARE e está envolvido no transporte vesicular em eventos de fusão das membranas celulares em leveduras (INADOME et al., 2007; IELASI et al., 2012; MORITA et al., 2018).

A sequência XM_750498.1 em *A. fumigatus* apresenta um domínio incompleto CLECT que codifica proteína lectina do tipo C-type. Um número de diferentes famílias destas proteínas compartilham um domínio conservado que foi caracterizado pela primeira vez em algumas lectinas animais e que parecem funcionar como um domínio de reconhecimento de carboidrato dependente de cálcio. As lectinas são um grupo diverso de proteínas, tanto em termos de estrutura quanto de atividade. A capacidade de ligação a carboidratos pode ter evoluído de forma independente e esporádica em numerosas famílias não relacionadas, onde cada uma delas desenvolveu uma estrutura que foi conservada para cumprir alguma outra atividade e função (CAMBI e FIGDOR, 2003; ZELENSKY e GREADY, 2005).

É interessante destacar que domínios envolvidos na regulação da expressão gênica estão presentes em todas as condições, mesmo quando submetidos a estresses bióticos e abióticos. Estes dados sugerem que as PHs, bem como as putativas estão envolvidas na reprogramação da expressão gênica em resposta a mudanças na temperatura, que desencadeia a diferenciação e também em resposta à presença de elementos hospedeiros. Os dados biológicos e estruturais relatados acima abrem as portas para os estudos de base molecular entre estas proteínas e seus respectivos ligantes, podendo assim mostrar uma possível seletividade para estas famílias de lectinas fúngicas. Este é o primeiro trabalho de caracterização *in silico* de proteínas com domínio lectina em leveduras de importância médica.

A associação dos alvos encontrados com o BLAST aos termos do GO usando a ferramenta Blast2GO, bem como a busca de domínios, utilizando os programas CD-Search, IPR e Pfam, que combina diferentes métodos de busca em vários bancos de dados de domínios e ou de famílias de proteínas, permitiu encontrar domínios

funcionais em nove proteínas hipotéticas pesquisadas. Ao compararmos as ferramentas IPR, CDD e Pfam, observamos que o IPR, ferramenta associada ao Blast2GO, em estudos de avaliação de preditores foi descrito como o melhor preditor de função de proteínas juntamente com o Blast2GO (RODRIGUES, 2013). A utilização de outros programas de predição, como os CDD e Pfam, nos permitiu confirmar os resultados obtidos, já que os domínios encontrados nos três preditores foram praticamente os mesmos. Exceções foram observadas para algumas proteínas. Este resultado nos mostra que, apesar dessas ferramentas serem muito utilizadas e consideradas os melhores preditores, ainda existe um considerável número de resultados negativos. Todavia é necessária a utilização de metodologias que combinem vários preditores, a fim de se obter resultados positivos para o maior número de sequências e confirmar os resultados positivos obtidos.

Entre as 48 sequências caracterizadas *in silico*, 39 são proteínas sem estrutura 3D resolvida, codificando para as seis famílias de lectinas, relacionadas com a resposta imune protetora ou envolvidas na patologia imunológica, durante infecções (KOTECHA e PODUVAL, 2019; SINGH et al., 2019). A compreensão do mecanismo de função de uma proteína geralmente requer o conhecimento de sua estrutura tridimensional (WEBER, 1990), que pode ser determinada por sua sequência de aminoácidos (ANFINSEN, 1973). O método de modelagem comparativa é aquele em que a estrutura tridimensional de uma sequência proteica alvo baseia-se principalmente em sua similaridade com uma ou mais proteínas de estruturas resolvidas (KOLINSKI et al., 1999). Ao buscar sequências consenso para AfLtype1 contra o banco de estruturas do PDB, a molécula de sequência mais similar foi a 2DUP (*Canis lupus familiaris*) com 46,67 % de similaridade, resolvida pelo método de difração em Raio-X. De acordo com Sanchez & Sali (1997), quando a identidade de sequência é superior a 40%, o alinhamento ocorre de maneira satisfatória.

É interessante ressaltar que a busca por estruturas com maior similaridade e com boa porcentagem de identidade nem sempre resulta em sequências base derivadas do mesmo grupo taxonômico. Isto ocorre porque o banco de dados do SWISS-MODEL compara sua sequência alvo com todas as sequências conhecidas (GUEx e PEITSCH, 1997; ARNOLD et al., 2006). Por isso, a molécula de referência 2DUP é resolvida como uma proteína de transporte do tipo lectina animal com domínio VIP36 exoplasmático / lumenal, livre de metais.

Quando comparada com sequências homólogas, a AfLtype1 é putativa para lectina do tipo L-type, com domínio e motivos altamente conservados (**Figura 14**). Resíduos de glicina (G) em lectinas dessa classe estão correlacionados ao local de ligação a carboidratos, enquanto que resíduos de Aspartato (D) associam-se à ligação de metais (CUMMINGS et al., 2017). A AfLtype1 é o primeiro modelo teórico proposto por modelagem comparativa de um domínio lectina tipo de leguminosas (L-type) em *A. fumigatus*.

A estrutura secundária foi composta por 2 folhas- β pregueadas antiparalelas, uma folha contendo 8 fitas e uma segunda folha contendo 7 fitas. Além disso, apresentou duas pequenas α –hélices e uma única ponte dissulfeto, formada entre os aminoácidos 1 e 2. Essa arquitetura de concha (estrutura β -sanduíche) forma o domínio e é característico de lectinas animais semelhantes às lectinas das plantas leguminosas (VIP36 e ERGIC-53) (FIEDLER e SIMONS, 1994; SATOH et al., 2007). O local de ligação aos carboidratos geralmente está localizado no ápice dessa concha (CUMMINGS et al., 2017). Proteínas como VIP36 (proteína integral vesicular de 36 kDa) são transmembrânicas e atuam como receptores de carga transportando glicoproteínas ligadas à manose entre o complexo de golgi (CG) e o retículo endoplasmático (ER) de forma seletiva (**Figura 2**) (KAMIYA et al., 2005; KAMIYA et al., 2008). Além disso, estudos também mostraram a relação dessas proteínas na regulação da fagocitose em macrófagos, contribuindo para a ativação ou inibição da inflamação / infecção (CALDERONE e FONZI, 2001; SHIRAKABE et al., 2011).

A AfLtype1 apresentou pequenas modificações comparada com sua conformação antes da dinâmica, avaliadas pelo gráfico de *root mean square deviation* (RMSD). Contudo, a estrutura característica permaneceu intacta após os testes de dinâmica molecular, evidenciando a estabilidade e rigidez das fitas (**Figura 24.A**). Quando as regiões que mais sofreram com o aquecimento foram analisadas, observou-se que as regiões de alças (loops), por serem bem mais flexíveis, vibraram mais quando submetidos aos testes de flutuação (**Figura 25**). Como foi visto, a dinâmica molecular tende a avaliar a conformação da proteína em condições nativas. As modificações pós-dinâmica mostraram que a AfLtype1 perdeu as duas pequenas α –hélices, fato que corrobora com as características estruturais L-types conhecidas, as quais são compostas de folhas β antiparalelas conectadas por loops e β -curvas, e geralmente não possuem α -hélices (CUMMINGS et al., 2017).

O índice de hidropatia de um aminoácido (GRAVY) de uma proteína mede as propriedades hidrofóbicas ou hidrofílicas da proteína, quanto maior o número, mais hidrofóbico é o aminoácido (KYTE e DOOLITTLE, 1982). Isso é importante para visualizarmos a estrutura de uma proteína, onde os aminoácidos hidrofóbicos tendem a ser internos, enquanto que os aminoácidos hidrofílicos são mais comumente encontrados na superfície da proteína. A disposição dos resíduos de aminoácidos de acordo com a polaridade e carga, mostrou que a AfLtype1 é bastante catiônica, com um arranjo de resíduos básicos que estão agrupados na base e no topo da estrutura, com predominância nas regiões de loop, conectando as folhas β -pregueadas e a α -hélice. Como foi observado, a proteína apresenta um caráter hidrofílico, que justifica sua polaridade, apesar da presença de numerosos resíduos hidrofóbicos (**Figura 24.B**).

De forma geral, as lectinas podem ser classificadas em grupos com base em suas especificidades de carboidratos. Estruturalmente falando, essas diferenças nas especificidades são provocadas em sua maioria, pela variabilidade na conformação e tamanho em regiões de voltas (loop). Estudos realizados em lectinas L-type mostraram a presença de quatro voltas com quatro aminoácidos invariantes essenciais para a ligação de carboidratos e que foram evidenciados em AfLtype1 (CUMMINGS et al., 2017). Estudos recentes realizados por Cummings e colaboradores em lectinas como L-type mostraram a presença de quatro voltas com quatro aminoácidos invariantes essenciais para a ligação de carboidratos e que foram evidenciados em AfLtype1 e mostrados na **Figura 26.C**. Em nosso estudo, a AfLtype1 apresentou uma estrutura como das lectinas VIP36/ERGIC-53, indicando que essa proteína seja a primeira lectina da família L-type como VIP36/ERGIC-53 caracterizada em *A. fumigatus*.

7 CONCLUSÃO

- Caracterizou-se *in silico* as proteínas que contém domínio lectina em *Aspergillus fumigatus*, *Candida albicans* e *Cryptococcus neoformans*, anotadas como proteínas hipotéticas, putativas ou preditas;
- Identificou-se elementos funcionais de ligação à carboidratos, em oito das nove PHs para estes fungos;
- Este foi o primeiro estudo em larga escala em busca de sequências com domínios lectina nos respectivos genomas dos fungos patogênicos *A. fumigatus*, *C. albicans* e *C. neoformans*, bem como da caracterização dessas sequências *in silico*;
- Este é o primeiro estudo de modelagem comparativa de uma nova lectina como L-type de *A. fumigatus* (AfLtype1), fornecendo informações valiosas que têm o potencial de contribuir para a compreensão do processo de patogênese destes fungos mediada pelas lectinas;
- Os dados biológicos e estruturais estudados abrem as portas para estudos de base molecular entre a AfLtype1 e seus respectivos ligantes, podendo assim mostrar uma possível seletividade para esta família de lectinas fúngicas.

REFERÊNCIAS

- AEBERSOLD R., MANN M. Mass spectrometry-based proteomics. **Nature** 422(6928):198–207, 2003.
- ALBERTS, B., LEWIS, J., RAFF, M.C. Biologia molecular da célula. 5ª ed. Porto Alegre: **Artmed**, 2010.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., LIPMAN, D. J.. Basic local alignment search tool. **J Mol Biol.**, 215(3):403-10, 1990.
- ÁLVARES, C. A.; SVIDZINSKI, T.I.E.; CONSOLARO, M.E.L. Candidíase vulvovaginal: fatores predisponentes do hospedeiro e virulência das leveduras. **J. Bras. Patol. Med. Lab.**, Rio de Janeiro, v. 43, n. 5, p. 319- 327, 2007.
- AMBROSI, M., CAMERON, N. R., AND DAVIS, B. G. Lectins: tools for the molecular understanding of the glycode. **Org. Biomol. Chem.** V. 3, 1593-1608, 2005.
- ANFINSEN, C.B. Principles that govern the folding of protein chains. **Science**, 181: 223-230p, 1973.
- ANGELONI, S., RIDET, J.L., KUSY, N., GAO, H., CREVOISIER, F., GUINCHARD, S., KOCHHAR, S., SIGRIST, H., SPRENGER, N.; Glycoprofiling with micro-arrays of glycoconjugates and lectins, **Glycobiology**, Volume 15, Issue 1, Pages 31–41, 2005.
- ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. **Bioinformatics** 22:195-201p, 2006.
- BAILEY T.L., BODÉN M., BUSKE F.A., FRITH M., GRANT C.E., CLEMENTI L., REN J., LI W.W., NOBLE W.S. "MEME SUITE: tools for motif discovery and searching", **Nucleic Acids Research**, 37:W202-W208, 2009.
- BAKER, D.; ŠALI, A. Protein structure prediction and structural genomics. **Nature**, v.294, p.93-96, 2001.

BARBOSA, E. B. et al. Proteômica: metodologias e aplicações no estudo de doenças humanas. **Revista da Associação Médica Brasileira**, v. 58, n. 3, p. 366–375, 2012.

BARI A.U., SANTIAGO M.Q., OSTERNE V.J.S., PINTO-JUNIOR V.R., PEREIRA L.P., SILVA-FILHO J.C., DEBRAY H., ROCHA B.A.M., DELATORRE P., TEIXEIRA C.S., NETO C.C., ASSREUY A.M.S., NASCIMENTO K.S., CAVADA B.S., Lectins from *Parkia biglobosa* and *Parkia platycephala*: A comparative study of structure and biological effects, **Int. J. Biol. Macromol.** 92, 2016.

BERENDSEN, H. J., POSTMA, J. P., VAN GUNSTEREN, W. F., AND HERMANS, J., Interaction Models for Water in Relation to Protein Hydration, In *Intermolecular Forces*, edited by Bernard Pullman, **Dordrecht: Springer Netherlands**. 14, 331–42. 1981.

BLEULER-MARTÍNEZ S., BUTSCHI A., GARBANI M., WÄLTI M.A., WOHLSCHLAGER T., POTTHOFF E., SABOTIĆ J., POHLEVEN J., LÜTHY P., HENGARTNER M.O., AEBI M., KÜNZLER M. A lectin mediated resistance of higher fungi against predators and parasites. **Mol. Ecol.**, v. 20, n. 14, p. 3056-3070, 2011.

BOI C. Membrane Chromatography for Biomolecule Purification. **Current Trends and Future Developments on (Bio-) Membranes**, 151–166. 2019.

BONNARDEL F., MARIETHOZ J., SALENTIN S., ROBIN X., SCHROEDER M., PEREZ S., LISACEK F., IMBERTY A.; UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. **Nucleic Acids Research**, Volume 47, Issue D1, 8 January, 2019.

BORASTON AB, BOLAM DN, GILBERT HJ, DAVIES GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. **Biochem J.** 382 (Pt 3):769–781, 2004.

BORNHOFFT KF, GOLDAMMER T, REBL A, GALUSKA SP. Siglecs: a journey through the evolution of sialic acid-binding immunoglobulin-type lectins. **Dev Comp Immunol** 86:219–231, 2018.

BRINDA KV, SUROLIA A, VISHVESHWARA S. Insights into the quaternary association of proteins through structure graphs: a case study of lectins. **Biochem J.** 391(Pt 1):1-15, 2005.

BROWN G.D., DENNING D.W., GOW N.A., LEVITZ S.M., NETEA M.G., WHITE T.C. Hidden killers: human fungal infections. **Sci Transl Med.** 4(165):165rv13. 2012.

CALDERONE, R A; FONZI, W A. Virulence factors of *Candida albicans*. **Trends in microbiology**, v. 9, n. 7, p. 327–35, 2001.

CAMBI A. AND FIGDOR C. G. Dual function of C-type lectin-like receptors in the immune system. **Current Opinion in Cell Biology**, 15:539–546, 2003.

CARVALHO E.V.M.M., BEZERRA R.F., BEZERRA R.S., ARAÚJO J.M., SANTOS A.J.G., CORREIA M.T.S., COELHO L.C.B.B. Detection of the first lectin with antimicrobial activity present in Serum of the Amazonian Fish Tambaqui (*Colossoma macropomum*). *Fisheries Science*, 78(1):01-09, 2012.

CASADEVALL A, PERFECT JR. *Cryptococcus neoformans*. Washington, DC: **ASM Press**; 1998.

CAVADA, B.S., BARBOSA, T., ARRUDA, S., GRANGEIRO, T.B., BARRAL-NETTO, M. Revisiting proteus: do minor changes in lectin structure matter in biological activity? Lessons from and potential biotechnological uses of the diocleinae subtribe lectins. **Curr. Protein Pept. Sci.** 2, 1–13. 2001.

CAVADA B.S., MORENO F.B., DA ROCHA B.A., DE AZEVEDO W.F.JR., CASTELLON R.E., GOERSCH G.V., NAGANO C.S., DE SOUZA E.P., NASCIMENTO K.S., RADIS-BAPTISTA G., DELATORRE P., LEROY Y., TOYAMA M.H., PINTO V.P., SAMPAIO A.H., BARETTINO D., DEBRAY H., CALVETE J.J., SANZ L. cDNA cloning and 1.75 Å crystal structure determination of PPL2, an endochitinase and N-acetylglucosamine-binding hemagglutinin from *Parkia platycephala* seeds. **FEBS J.** Sep;273(17):3962-74, 2006.

COELHO, L.C.B.B., SANTOS, A.F.S., NAPOLEÃO, T.H., CORREIA, M.T.S., PAIVA, P.M.G. Protein Purification by Affinity Chromatography. In: Ahmad, R. (Org.). Protein Purification. **INTECH Open Access Publisher**, Rijeka, Croatia 1, 224p, 2012.

COHEN-KUPIEC R., BROGLIE K.E., FRIESEM D., BROGLIE R.M., CHET I. Molecular characterization of a novel β -1,3- exoglucanase related to mycoparasitism of *Trichoderma harzianum*. **Gene**, 226:147-154, 1999.

CUMMINGS R.D., ETZLER M.E. R-TYPE LECTINS. IN: VARKI A, CUMMINGS RD, ESKO JD, et al., editors. **Essentials of Glycobiology**. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; Chapter 28. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1940/>. 2009.

CUMMINGS R.D., ETZLER M.E., SUROLIA A. L-Type Lectins. **Essentials of Glycobiology**, 3rd edition. Chapter 32. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK453092/>. 2017.

DAN, X., LIU, W., NG, T. B., Development and Applications of Lectins as Biological Tools in Biomedical Research. **Med. Res. Rev.**, 36: 221-247, 2016.

DAHMS N. M., HANCOCK M. K. P-type lectins. **Biochimica et Biophysica Acta**, 1572, 317 – 340, 2002.

DESLER, C., SURAVAJHALA, P., SANDERHOFF, M., RASMUSSEN, M., RASMUSSEN, L. J. *In Silico* screening for functional candidates amongst hypothetical proteins. **BMC Bioinformatics**, 10:289, 2009.

DIAS R.O., MACHADO L.S., MIGLIOLO L., FRANCO O.L. Insights into animal and plant lectins with antimicrobial activities. **Molecules** 20, 519– 541. 2015.

DJEBALI, S., WUCHER, V., FOISSAC, S., HITTE, C., CORRE, E., & DERRIEN, T. Bioinformatics Pipeline for Transcriptome Sequencing Analysis. **Enhancer RNAs**, 201–219, 2016.

FIEDLER K, SIMONS K. A putative novel class of animal lectins in the secretory pathway homologous to leguminous lectins. **Cell** 77, 625-6, 1994.

FILHO O. A. S., ALENCASTRO R. B. Modelagem de proteínas por homologia. **Quim. Nova**, Vol. 26, No. 2, 253-259, 2003.

FINN, R. D., BATEMAN, A., CLEMENTS, J., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., HEGER, A., HETHERINGTON, K., HOLM, L., MISTRY, J., SONNHAMMER, E. L., TATE, J., PUNTA, M.. Pfam: the protein families database. **Nucleic Acids Res.** 1;42(1):222-230, 2014.

FRANÇOIS L. MAYER, DUNCAN WILSON & BERNHARD HUBE. *Candidaalbicans* pathogenicity mechanisms, **Virulence**, 4:2, 119-128, DOI: 10.4161/viru.22913, 2013.

FU, Y.; LUO, G.; SPELBERG, B.J.; EDWARD, J.E.; IBRAHIM, B.C. Gene overexpression/Suppression Analysis of candidate virulence factores of *Candida albicans*. **Eukaryotic cell**. V.7, p.-483-492, 2008.

GALLEGOS B., MARTÍNEZ R., PÉREZ L., PINA M. S., PEREZ E., HERNÁNDEZ P. Lectins in human pathogenic fungi. **Rev Iberoam Micol.**;31(1):72–75, 2014.

GHUFRAN MS., KRISHNA GHOSH, SANTOSH R. KANADE; A fucose specific lectin from *Aspergillus flavus* induced interleukin-8 expression is mediated by mitogen activated protein kinase, p38, **Medical Mycology**, Volume 55, Issue 3, 1 April 2017.

GONÇALVES G.R.F., GANDOLFI O.R.R., SANTOS L.S., BONOMO R.C.F., VELOSO C.M., VERÍSSIMO L.A.A., FONTAN R.C.I. Immobilization of sugars in supermacroporous cryogels for the purification of lectins by affinity chromatography. **Journal of Chromatography B**, 1068-1069, 71–77. 2017.

GOOSSENS K., WILLAERT R. Flocculation protein structure and cell–cell adhesion mechanism in *Saccharomyces cerevisiae*. **Biotechnol Lett**, 32:1571–1585, 2010.

GOW N.A.R., LATGE J.P., MUNRO C.A. The fungal cell wall: structure, biosynthesis, and function. **Microbiol Spectr**. 5(3). 2017.

GUERX, N.; PEITSCH, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. **Electrophoresis**, 18: 2714-2723p, 1997.

HASAN I., OZEKI Y., KABIR S.R. Purification of a novel chitin-binding lectin with antimicrobial and antibiofilm activities from a Bangladeshi cultivar of potato (*Solanum tuberosum*). **Indian J Biochem Biophys**, 51, 142– 148. 2014.

HAZES B. The (QxW)₃ domain: A flexible lectin scaffold. **Protein Science**, 5:1490-1501, 1996.

HE, X., LIU, F., LIU, L., DUAN, T., ZHANG, H., WANG, Z. Lectin-conjugated Fe₂O₃@ Au core@ shell nanoparticles as dual mode contrast agents for in vivo detection of tumor. **Mol. Pharm.** 11 (3), 738745, 2014.

HESS, B., BEKKER, H., BERENDSEN, H. J., AND FRAAIJE, J. G., LINCS: A Linear Constraint Solver for Molecular Simulations. **Journal of Computational Chemistry**, 18: 1463–72. 1997.

HIRABAYASHI J., DUTTA S. K. AND KASAI K. Novel Galactose-binding Proteins in Annelida. Characterization of 29-kDa tandem repeat-type lectins from the earthworm *Lumbricus terrestris*. Vol. 273, No. 23, **Issue of June 5**, pp. 14450–14460, 1998.

HIRABAYASHI J, YAMADA M, KUNO A, TATENO H: Lectin microarrays: concept, principle and applications. **Chem Soc Rev**, 2013.

HOCKNEY, R. W., Methods in Computational Physics. **Alder, B**, 136–211. 1970.

HOPPER J.T.S., AMBROSE S., GRANT O.C., KRUMM S.A., ALLISON T.M., DEGIACOMI M.T., TULLY M.D., PRITCHARD L.K. et al. The tetrameric plant lectin BanLec neutralizes HIV through bidentate binding to specific viral glycans. **Structure** 25, 773– 782.e5. 2017.

HOUSER, J., KOMAREK, J., CIOCI, G., VARROT, A., IMBERTY, A., & WIMMEROVA, M. Structural insights into *Aspergillus fumigatus* lectin specificity: AFL binding sites are functionally non-equivalent. **Acta Crystallographica Section D Biological Crystallography**, 71(3), 442–453, 2015.

IELASI F. S., DECANNIERE K. AND WILLAERT R. G. The epithelial adhesin 1 (Epa1p) from the humanpathogenic yeast *Candida glabrata*: structural and functional study of the carbohydrate-binding domain. **Acta Cryst.** D68, 210–217, 2012.

INADOME H., NODA Y., KAMIMURA Y., ADACHI H., YODA K. Tvp38, Tvp23, Tvp18 and Tvp15: Novel membrane proteins in the Tlg2-containing Golgi/endosome compartments of *Saccharomyces cerevisiae*. **Experimental Cell Research** 313, 688 – 697, 2007.

ISHIMARU T., BERNARD E. M., TAMADA S., AND ARMSTRONG D., Letter, **Clin. Infect. Dis.** 23:898, 1996.

JANICKI, M., ROOKE, R., & YANG, G. Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. **Chromosome Research**, 19(6), 787–808, 2011.

JIANG S, CHEN Y, WANG M, YIN Y, PAN Y, GU B, YU G, LI Y, WONG BH, LIANG Y et al.: A novel lectin from *Agrocybe aegerita* shows high binding selectivity for terminal N-acetylglucosamine. **Biochem J**, 2012.

KAMIYA, Y., YAMAGUCHI, Y., TAKAHASHI, N., ARATA, Y., KASAI, K., IHARA, Y., ... KATO, K. Sugar-binding Properties of VIP36, an Intracellular Animal Lectin Operating as a Cargo Receptor. **Journal of Biological Chemistry**, 280(44), 37178–37182, 2005.

KAMIYA, Y., KAMIYA, D., YAMAMOTO, K., NYFELER, B., HAURI, H.-P., & KATO, K. Molecular Basis of Sugar Recognition by the Human L-type Lectins ERGIC-53, VIPL, and VIP36. **Journal of Biological Chemistry**, 283(4), 1857–1861, 2007.

KAMIYA Y, KAMIYA D, YAMAMOTO K, NYFELER B, HAURI HP, KATO K. J. Molecular basis of sugar recognition by the human L-type lectins ERGIC-53, VIPL, and VIP36. **Biol. Chem.** 283, 1857-61, 2008.

KHAN, M. S. A, et al. Virulence and Pathogenicity of Fungal Pathogens with Special Reference to *Candida albicans*. Combating Fungal Infections. In: AHMAD, I., ET AL. *Combating Fungal Infection: problems and remedy*. Berlin: **Springer**, 2010.

KLIS, F.M.; MOL, P.; HELLINGWERF, K.; BRUL, S. Dynamics of cell wall structure in *Saccharomyces cerevisiae*. **FEMS Microbiol. Rev.** v.26, p.239-256, 2002.

KLIS, F. M., BRUL, S., DE GROOT, P. W. J. Covalently linked wall proteins in ascomycetous fungi. **Yeast**, 27(8), 489–493, 2009.

KOTECHA H., PODUVAL P.B. Microbial lectins: roles and applications. **Advances in Biological Science Research**, Chapter 9, 135-139. 2019.

KULETA, J. K.; KOZIK, M. R.; KOZIK, A. Fungi pathogenic to humans: molecular bases of virulence of *Candida albicans*, *Cryptococcus neoformans* and *Aspergillus fumigatus*. **Acta Bioch. Pol.**, Warsaw, v. 56, n. 2, p. 211-224, 2009.

KUO-CHEN CHOU AND HONG-BIN SHEN, "Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms", **Nature Protocols**, 2008.

KUMAR, R., SINGH, S., & DUBEY, V. K. Bioinformatics Tools to Analyze Proteome and Genome Data. **Advances in the Understanding of Biological Sciences Using Next Generation Sequencing (NGS) Approaches**, 179–194, 2015.

LAM S.K., NG T.B. Lectins: Production and practical applications. **Appl Microbiol Biotechnol.** 89(1):45–55. 2011.

LAGARDA-DIAZ I., GUZMAN-PARTIDA A. M., VAZQUEZ-MORENO L. Legume Lectins: Proteins with Diverse Applications. **Int. J. Mol. Sci.**, 18, 1242, 2017.

LASKOWSKI R.A., MACARTHUR M.W., MOSS D.S., THORNTON J.M., PROCHECK: a program to check the stereochemical quality of protein structures, **J. Appl. Crystallogr.** 26, 283–291. 1993.

LEE J., SITAO W., ZHANG Y. *Ab Initio* Protein Structure Prediction. D.J. Rigden (ed.) From Protein Structure to Function with Bioinformatics, **Springer Science**, 2009.

LEPENIES B., LANG R. Lectins and Their Ligands in Shaping Immune Responses. **Front. Immunol.** 10:2379. 2019.

LI, Y. R. et al. A novel lectin with potent antitumor, mitogenic and HIV-1 reverse transcriptase inhibitory activities from the edible mushroom *Pleurotus citrinopileatus*. **Biochimica et Biophysica Acta**, *in press*, 2007.

MACAULEY, M. S., CROCKER, P. R., & PAULSON, J. C. *Siglec-mediated regulation of immune cell function in disease*. **Nature Reviews Immunology**, 14(10), 653–666, 2014.

MALUCHE, M.E.; SANTOS, J.I. *Candida* sp. e infecções hospitalares: aspectos epidemiológicos e laboratoriais. **Rev. Bras. Anal. Clin.**, Rio de Janeiro, v. 40, n. 1, p. 65-67, 2008.

MARCHLER-BAUER A et al., "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures", **Nucleic Acids Res.** 45(D)200–3, 2017.

MARCHLER-BAUER A et al., "CDD: NCBI's conserved domain database", **Nucleic Acids Res.**43(D)222-6, 2015.

MARCHLER-BAUER A et al., "CDD: a Conserved Domain Database for the functional annotation of proteins", **Nucleic Acids Res.**, 39(D)225-9, 2011.

MCCALL, A.D., PATHIRANA, R.U., PRABHAKAR, A. et al. *Candida albicans* biofilm development is governed by cooperative attachment and adhesion maintenance proteins. **npj Biofilms Microbiomes**, 5, 21, 2019.

MEIER R., RUTTKIES C., TREUTLER H., NEUMANN S. Bioinformatics can boost metabolomics research, **J. Biotechnol.** 261, 137e141, 2017.

MORITA K., HAMA Y., IZUME T., TAMURA N., UENO T., YAMASHITA Y., SAKAMAKI Y., MIMURA K., MORISHITA H., SHIHOYA W., NUREKI O., MANO H., MIZUSHIMA N. Genome-wide CRISPR screen identifies TMEM41B as a gene required for autophagosome formation. **J Cell Biol Nov**, 217 (11) 3817-3828; 2018.

MUSZYŃSKA B., ROJOWSKI J., ŁAZARZ M., KAŁA K., DOBOSZ K., OPOKA W. The accumulation and release of Cd and Pb from edible mushrooms and their biomass. **Polish Journal of Environmental Studies.** 27, 1-8. 2018.

NAGASUMA R. CHANDRA, M.M. PRABU, K. SUGUNA, M. VIJAYAN; Structural similarity and functional diversity in proteins containing the legume lectin fold. **Protein Engineering, Design and Selection**, Volume 14, Issue 11, Pages 857–866, 2001.

NASCIMENTO A.P.M., WOLIN I.A.V., WELTER P.G., HEINRICH I.A., ZANOTTO-FILHO A., OSTERNE V.J.S., LEAL R.B. Lectin from *Dioclea violacea* induces autophagy in U87 glioma cells. **International Journal of Biological Macromolecules.** 134 (2019) 660–672, 2019.

NIZET, V.; VARKI, A.; AEBI, M. Lectins : Hemagglutinins , Adhesins , and Toxins. **National Library of Medicine**, n. Chapter 37, Microbial, p. 1–7, 2018.

OFEK I., SHARON N. Adhesins as lectins: specificity and role in infection. **Current Topics in Microbiology and Immunology**, Volume: 151. Pages: 91-113. 1990.

OSTENBRINK, C., VILLA, A., MARK, A. E., & VAN GUNSTEREN, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. **Journal of computational chemistry**, 25, 1656-1676. 2004.

OTTO, T. D. et al. A plataforma PDTIS de bioinformática: da sequência à função. **Reciis**, v. 1, n. 2, 2007.

PAQUET E. AND VIKTOR H.L. Computational Methods for *Ab Initio* Molecular Dynamics. **Hindawi, Advances in Chemistry**, Article ID 9839641, 14 pages, 2018.

PARK BJ, WANNEMUEHLER KA, MARSTON BJ, GOVENDER N, PAPPAS PG, CHILLER TM. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. **AIDS Lond Engl.**, 2009.

PERFECT JR. *Cryptococcus neoformans*: A sugar-coated killer with designer genes. **FEMS Immunol Med Microbiol.**, 2005.

PONTING, C. P., HOFMANN, K., & BORK, P. A latrophilin/CL-1-like GPS domain in polycystin-1. **Current Biology**, 9(16), R585–R588, 1999.

PUNTA., M., COGGILL, P. C., EBERHARDT, R. Y., MISTRY, J., TATE J., BOURSNEILL, C., PANG, N., FORSLUND, K., CERIC, G., CLEMENTS, J., HEGER, A., HOLM, L., SONNHAMMER, E. L., EDDY, S. R., BATEMAN, A., FINN., R. D.. The Pfam protein families database. **Nucleic Acids Res.**, 40:D290-301, 2012.

RUIZ-HERRERA, J.; ELORZA, M.V.; VALENTIN,E.; SENTANDREU, R. Molecular organization of the cell wall of *Candida albicans* and its relation to pathogenicity. **FEMS Yeast Res**, v.6, p.14-29, 2006.

SAKAI K., HIEMORI K., TATENO H., HIRABAYASHI J., GONOI T. Fucose-specific lectin of *Aspergillus fumigatus*: binding properties and effects on immune response stimulation. **Medical Mycology**, Volume 57, Issue 1, 2018.

SANTANA D.P.; RIBEIRO E.L.; MENEZES A.C.S.; NAVES P.L.F. Novas abordagens sobre os fatores de virulência de *Candida albicans*. **Rev. Ciênc. Méd. Biol.**, Salvador, v.12, n.2, p.229-233, mai./ago. 2013.

SANTOS, A. F. S., NAPOLEÃO, T. H., PAIVA, P. M. G. AND COELHO, L. C. B. B. Lectins: Important Tools for Biocontrol of *Fusarium* Species. T. F. Rios and E. R. Ortega (Ed.), **Nova Science Publishers**, Inc., New York, 161, 2012.

SANTOS-FILHO O.A., ALENCASTRO R.B. Modelagem de proteínas por homologia. **Quím. Nova**, São Paulo , v. 26, n. 2, p. 253-259, Mar. 2003.

SANZ-APARICIO, J., HERMOSO, J., GRANGEIRO, T.B., CALVETE, J.J., CAVADA, B.S. The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from Concanavalin A. **FEBS Lett.** 405, 114–118. 1997.

SATO Y., MORIMOTO K., KUBO T., SAKAGUCHI T., NISHIZONO A., HIRAYAMA M., HORI K. Entry inhibition of influenza viruses with high mannose binding lectin ESA-2 from the red alga *Eucheuma serra* through the recognition of viral hemagglutinin. **Mar Drugs**, 13, 3454– 3465. 2015.

SATOH, T., COWIESON, N. P., HAKAMATA, W., IDEO, H., FUKUSHIMA, K., KURIHARA, M., WAKATSUKI, S. Structural Basis for Recognition of High Mannose Type Glycoproteins by Mammalian Transport Lectin VIP36. **Journal of Biological Chemistry**, 282(38), 28246–28255, 2007.

SHARON, N.; LIS, H. How proteins bind carbohydrates: Lessons from legume lectins. **Journal of Agricultural and Food Chemistry**, v. 50, n. 22, p. 6586– 6591, 2002.

SHARON, N.; LIS, H. History of lectins: from hemagglutinins to biological recognition molecules. **Glycobiology**, v. 14, n. 11, p. 53–62, 2004.

SHIMIDT A., FORNE I., IMHOF A. Bioinformatic analysis of proteomics data. **BMC Systems Biology**, 8(Suppl 2):S3, 2014.

SHIRAKABE K., HATTORI S., SEIKI M., et al. VIP36 protein is a target of ectodomain shedding and regulates phagocytosis in macrophage raw 264.7 cells. **Journal of Biological Chemistry**, 286:43154–43163. 2011.

SHOSEYOV O. SHANI Z. AND LEVY I. Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. **Microbiology And Molecular Biology Reviews**, p. 283–295, 2006.

SINGH R.S., KAUR H.P., SINGH J. Purification and characterization of a mycelial mucin specific lectin from *Aspergillus panamensis* with potent mitogenic and antibacterial activity. **Process Biochem** 50, 2251– 2258. 2015.

SINGH R.S., WALIA A.K., KENNEDY J.F., Mushroom lectins in biomedical research and development. **International Journal of Biological Macromolecules**. 2019.

SITOHY, M.; DOHEIM, M.; BADR, H. Isolation and characterization of a lectin with antifungal activity from Egyptian *Pisum sativum* seeds. **Food Chemistry**, v. 104, n. 3, p. 971-979, 2007.

SOARES I.A., FLORES A.C., ZANETTIN L., PIN H.K., MENDONÇA M.M., BARCELOS R.P., TREVISOL L.R., CARVALHO R.D., SCHAUREN D., ROCHA C.L.M.S.C., BARONI S. Identificação do potencial amilolítico de linhagens mutantes do fungo filamentoso *Aspergillus nidulans*. **Ciênc. Tecnol. Aliment.**, Campinas, 30(3): 700-705, jul.-set. 2010.

STRAVER M. H., GERRIT S. AND KIJNE J. W. Purification and Partial Characterization of a Flocculin from Brewer's Yeast. **Applied And Environmental Microbiology**. p. 2754-2758, 1994.

SUEYOSHI S, TSUJI T, OSAWA T. Purification and characterization of four isolectins of mushroom (*Agaricus bisporus*). **Biol Chem Hoppe Seyler** 366:213–221, 1985.

SULAK O, CIOCI G, LAMEIGNERE E, BALLOY V, ROUND A, et al. *Burkholderia cenocepacia* BC2L-C Is a Super Lectin with Dual Specificity and Proinflammatory Activity. **PLoS Pathog** 7(9): e1002238, 2011.

SUMNER J. B., GRALEN N., ERIKSSON-QUENS I. B. The Molecular Weights Of Urease, Canavalin, Concanavalin A and Concanavalin B. **Science**, Vol. 87, 1938.

SUNDSTROM, P. Adhesion in *Candida spp*. **Cell Microbiol**. V.4, p.461-469, 2002.

SWANSON M.D., WINTER H.C., GOLDSTEIN I.J., MARKOVITZ D.M. A lectin isolated from bananas is a potent inhibitor of HIV replication. **J Biol Chem** 285(12):8646–8655. 2010.

TINBERG C.E., KHARE, S.D., DOU, J., DOYLE, L., NELSON, J.W., SCHENA, A., et al. Computational design of ligand-binding proteins with high affinity and selectivity. **Nature** 501 (7466), 212, 2013.

TRONCHIN G, ESNAULT K, SANCHEZ M et al. Purification and partial characterization of a 32-kilodalton sialic acid-specific lectin from *Aspergillus fumigatus*. *Infect Immun.* 70: 6891–6895, 2002.

VAILATI-RIBONI, M., PALOMBO, V., & LOOR, J. J. What Are Omics Sciences? **Periparturient Diseases of Dairy Cows**, 1–7. 2017.

VAN-DAMME E.J.M., PEUMANS W.J., BARRE A., ROUGE P. Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. **Crit Rev Plant Sci** 17, 575– 692, 1998.

VARROT, A., BASHEER, S.M. AND IMBERTY, A.. Fungal lectins: structure, function and potential applications. **Current Opinion in Structural Biology**, 23, 678–685, 2013.

VERNA J., LODDER A., LEE K., VAGTS A., BALLESTER R. A family of genes required for maintenance of cell wall integrity and for the stress response in *Saccharomyces cerevisiae*. **Proc Natl Acad Sci USA**, 94:13804-13809, 1997.

VERSTREPEN, K. J. AND KLIS, F. M. Flocculation, adhesion and biofilm formation in yeasts. **Molecular Microbiology**, 60: 5-15, 2006.

WALSH NM, WUTHRICH M, WANG H, KLEIN B, HULL CM. Characterization of C-type lectins reveals an unexpectedly limited interaction between *Cryptococcus neoformans* spores and Dectin-1. **PLoS ONE** 12(3): e0173866, 2017.

WANG X. W., WANG J. X. Diversity and multiple functions of lectins in shrimp immunity. **Developmental and Comparative Immunology**, 39, 27–38, 2013.

WANG, C.; HE, X. Supramolecular glycorhodamine-polymer dot ensembles for the homogeneous , fluorogenic analysis of lectins. **Carbohydrate Research**, v. 455, p. 1–4, 2018.

WAWRA, S., FESEL, P., WIDMER, H., NEUMANN, U., LAHRMANN, U., BECKER, S., HEHEMANN, J., LANGEN, G., ZUCCARO, A. FGB1 and WSC3 are *in*

planta-induced β -glucan-binding fungal lectins with different functions. **New Phytol**, 222: 1493-1506. 2019.

WEBB B., SALI A., Comparative Protein Structure Modeling Using MODELLER, **Curr. Protoc. Bioinforma.** 47, 5.6.1-5.6.32. 2014.

WEBER, I.T. Evaluation of homology modeling of HIV Protease. **Proteins**, 7 (2): 172-184p, 1990.

WEIS, W. I., AND DRICKAMER, K. Structural basis of lectincarbohydrate recognition, **Annu. ReV. Biochem.** 65, 441-473. 1996.

WIEDERSTEIN M., SIPPL M.J., ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, **Nucleic Acids Res.** 35, 407–410. 2007.

WILLAERT, R. G. Adhesins of Yeasts: Protein Structure and Interactions. **Journal of Fungi**, 4(4), 119. 2018.

XIE X., and LIPKE P. N. On the evolution of fungal and yeast cell walls. **Yeast**. August, 27(8): 479–488. 2010.

YAU T., DAN X., NG C.C., NG T.B. Lectins with potential for anti-cancer therapy. **Molecules**, 20(3):3791–3810. 2015.

YUZUGULLU, Y.; DUMAN, Y. A. Aqueous Two Phase (PEG4000/Na₂SO₄) Extraction and Characterization of an Acid Invertase from Potato Tuber (*Solanum Tuberosum*). **Preparative Biochemistry and Biotechnology**, v. 7, p. 696-711. 2015.

ZELENSKY A. N. AND GREASY J. E. The C-type lectin-like domain superfamily. **FEBS Journal**, 272, 6179–6217, 2005.

ZHANG N, HARREX AL, HOLLAND BR, et al. Sixty alleles of the ALS7 open reading frame in *Candida albicans*: ALS7 is a hypermutable contingency locus. **Genome Res.** 13:2005–2017, 2003.

ZHANG Y. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**, 9:40, 2009.

ZHANG H., PEATMAN E., LIU H., FENG T., CHEN L., LIU Z. Molecular characterization of three L-type lectin genes from channel catfish, *Ictalurus punctatus* and their responses to *Edwardsiella ictaluri* challenge. **Fish & Shellfish Immunology** 32, 598-608, 2012.