



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I – CAMPINA GRANDE  
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AGRÁRIAS  
MESTRADO EM CIÊNCIAS AGRÁRIAS**

**IZAMARA GESIELE BEZERRA DE OLIVEIRA**

**MODELAGEM *IN SILICO* DO GENE *gumD* E SUA REDE DE INTERAÇÃO EM  
*Gluconacetobacter diazotrophicus* PAL5**

**CAMPINA GRANDE-PB  
2023**

IZAMARA GESIELE BEZERRA DE OLIVEIRA

**MODELAGEM *IN SILICO* DO GENE *gumD* E SUA REDE DE INTERAÇÃO EM  
*Gluconacetobacter diazotrophicus* PAL5**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual da Paraíba/ Embrapa Algodão da Paraíba, como requisito parcial à obtenção do título de Mestre em Ciências Agrárias.

**Área de concentração:** Agrobioenergia e Agricultura Familiar. **Linha de pesquisa:** Biotecnologia e Melhoramento Vegetal.

**Orientador:** Prof. Dr. Carlos Henrique Salvino Gadelha Meneses  
**Coorientadora:** Dra. Nathalia Maíra Cabral de Medeiros

**CAMPINA GRANDE-PB  
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

O48m Oliveira, Izamara Gesiele Bezerra de.  
Modelagem *in silico* do gene *gumD* e sua rede de interação em *Gluconacetobacter diazotrophicus* PAL5 [manuscrito] / Izamara Gesiele Bezerra de Oliveira. - 2023.  
115 p. : il. colorido.

Digitado.  
Dissertação (Mestrado em Ciências Agrárias) - Universidade Estadual da Paraíba, Centro de Ciências Biológicas e da Saúde, 2023.  
"Orientação : Prof. Dr. Carlos Henrique Salvino Gadelha Meneses, Coordenação do Curso de Ciências Biológicas - CCBSA. "

1. Interação bactéria-planta. 2. Biotecnologia. 3. Modelagem 3D. 4. STRING. I. Título

21. ed. CDD 660.6

IZAMARA GESIELE BEZERRA DE OLIVEIRA

MODELAGEM *IN SILICO* DO GENE *gumD* E SUA REDE DE INTERAÇÃO EM  
*Gluconacetobacter diazotrophicus* PAL5

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual da Paraíba/ Embrapa Algodão da Paraíba, como requisito parcial à obtenção do título de Mestre em Ciências Agrárias.

**Área de concentração:** Agrobioenergia e Agricultura Familiar. **Linha de pesquisa:** Biotecnologia e Melhoramento Vegetal.

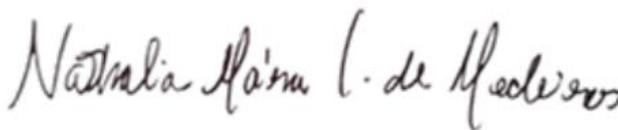
Aprovada em: 04 / 08 / 2023.

**BANCA EXAMINADORA**



---

Prof. Dr. Carlos Henrique Salvino Gadelha Meneses (Orientador)  
Universidade Estadual da Paraíba (PPGCA/UEPB)



---

Dra. Nathalia Maíra Cabral de Medeiros (Coorientadora)  
Universidade Estadual da Paraíba (PPGCA)



---

Prof. Dr. Josemir Moura Maia  
Universidade Estadual da Paraíba (PPGCA/UEPB)



---

Profa. Dra. Márcia Soares Vidal  
Empresa Brasileira de Pesquisa Agropecuária (Embrapa Agrobiologia- RJ)

## **AGRADECIMENTOS**

Agradeço a Deus, por todo discernimento, sabedoria, força, paz e por conduzir cada um dos meus passos.

Ao meu orientador Dr. Carlos Henrique Meneses pela oportunidade de aprendizagem, aprimoramento do conhecimento científico, pela confiança em mim depositada e pelo apoio na continuidade da carreira profissional.

À minha coorientadora Dra. Nathalia Medeiros, pelas valiosas contribuições e ensinamentos, principalmente em bioinformática, por toda dedicação e confiança.

À minha mãe, Erivania Oliveira, ao meu pai, Joselito Alves, aos meus irmãos, Itamara Marié e Joselito Filho, e aos meus sobrinhos, Miguel José, Mariana Oliveira, Marília Maria e Ana Ludmilla por todo amor, incentivo, companheirismo e por alegrarem tanto minha vida. Amo muito vocês!

Ao meu avô Severino de Oliveira (*in memoriam*), conhecido por “Biu”, embora fisicamente ausente, estarás sempre presente em meu coração.

A minha avó Marié Bezerra da Silva (*in memoriam*).

As amigas Thereza Cordeiro e Raiana Santos, pela amizade, apoio e cumplicidade.

Aos componentes da banca, Dr. Josemir Maia e Dra. Márcia Vidal, por aceitarem o convite e pelas contribuições para o aperfeiçoamento deste trabalho.

Ao Programa de Pós-Graduação em Ciências Agrárias (PPGCA-UEPB/Embrapa, em especial o secretário Danilo Cavalcante.

Aos professores do PPGCA, em especial, Dr. Carlos Henrique Meneses (UEPB), Dr. Evandro Mesquita (UEPB), Dr. Francisco José Correia Farias (Embrapa), Dr. Josemir Maia (UEPB), Dra. Nathalia Medeiros (PPGCA) por todos os ensinamentos transmitidos durante as disciplinas.

Aos colegas de pós-graduação Larissa, Jailma, Gabriel, Rogério, Andreza, Camila, Marcelo, Luana e Gaby.

A UEPB e funcionários. E todos que contribuíram direta ou indiretamente para a concretização desse projeto.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

**A Deus seja dada toda Honra e toda Glória!**

## RESUMO

A biologia de sistemas é um ramo da ciência que visa compreender as interações e organização das partes moleculares, celulares e organismos para a produção dos complexos processos biológicos. Para alcançar esse objetivo, são empregados diversos métodos experimentais, incluindo genômica, proteômica e metabolômica, bem como ferramentas computacionais sofisticadas, como modelagem matemática, simulação e análise de redes complexas. Neste contexto, a *Gluconacetobacter diazotrophicus* PAL5 se destaca como uma bactéria diazotrófica endofítica, conhecida por sua tolerância ao ar e por desempenhar um papel fundamental na fixação biológica de nitrogênio (FBN). Pesquisas evidenciaram que o gene *gumD* dessa bactéria é essencial para a produção de Exopolissacarídeos (EPS). O objetivo deste trabalho é compreender a função da proteína GumD em *G. diazotrophicus* PAL5, especialmente relacionada à sua interação com as plantas. Para isso, foi utilizada a modelagem *in silico* e construção de redes de interação dessa proteína. Essa abordagem envolveu a obtenção de informações das proteínas em bancos de dados como UniProt, NCBI, Prosite e Pfam. A modelagem 3D foi realizada utilizando a ferramenta I-TASSER, e os modelos foram refinados e avaliados por meio do ModRefiner e MolProbity, respectivamente. A visualização dos modelos foi feita no programa Chimera 1.16. A análise da rede de interação e Clusters, que foi realizada utilizando o STRING, apontou que a proteína codificada pelo gene *gumD* está envolvida nos processos de transferase de polissacarídeos. Os dados obtidos sugerem que essa proteína está envolvida no transporte de polissacarídeos, o que pode facilitar a formação do biofilme e, possivelmente, influenciar na interação da bactéria *G. diazotrophicus* PAL5 com as plantas. É importante ressaltar que esses resultados são baseados em análises computacionais e preditivas, sendo necessários estudos experimentais adicionais para validar e confirmar as interações e funções da proteína GumD. O aprofundamento dessas investigações contribuirá para uma melhor compreensão do papel dessa proteína na relação bactéria-planta e abrirá caminho para aplicações práticas na agricultura, como o desenvolvimento de ferramentas biotecnológicas para aprimorar o rendimento das culturas.

**Palavras-Chave:** bactéria promotora de crescimento vegetal; interação bactéria-planta; Poliprenil glicosilfosfotransferase; STRING.

## ABSTRACT

Systems biology is a branch of science that aims to understand the interactions and organization of molecular, cellular and organismal parts for the production of complex biological processes. In order to achieve this goal, several experimental methods are employed, including genomics, proteomics and metabolomics, as well as sophisticated computational tools, such as mathematical modeling, simulation and analysis of complex networks. In this context, *Gluconacetobacter diazotrophicus* PAL5 stands out as an endophytic diazotrophic bacterium, known for its tolerance to air and for playing a fundamental role in biological nitrogen fixation (BNF). Recently, research has shown that the *gumD* gene of this bacterium is essential for the production of Exopolysaccharides (EPS). The objective of this work is to understand the function of the GumD protein in *G. diazotrophicus* PAL5, especially related to its interaction with plants. For this, *in silico* modeling and construction of interaction networks of this protein were used. This approach involved obtaining protein information from databases such as UniProt, NCBI, Prosite and Pfam. 3D modeling was performed using the I-TASSER tool, and the models were refined and evaluated using ModRefiner and MolProbity, respectively. Visualization of the models was done in Chimera 1.16 software. The analysis of the interaction network and clusters, which was carried out using STRING, showed that the protein encoded by the *gumD* gene is involved in polysaccharide transferase processes. The data obtained suggest that this protein is involved in the transport of polysaccharides, which may facilitate biofilm formation and possibly influence the interaction of the bacterium *G. diazotrophicus* PAL5 with plants. It is important to highlight that these results are based on computational and predictive analyses, requiring additional experimental studies to validate and confirm the interactions and functions of the GumD protein. Further investigation of these investigations will contribute to a better understanding of the role of this protein in the bacteria-plant relationship and will pave the way for practical applications in agriculture, such as the development of biotechnological tools to improve crop yields.

**Keywords:** bacteria-plant interaction; plant growth promoting bacteria; Polyprenyl glycosylphosphotransferase; STRING.

## LISTA DE FIGURAS

Figura 1– <i>Gluconacetobacter diazotrophicus</i> .....	28
Figura 2 – Escala entre métodos de predição de estrutura tridimensional de proteínas e identidade de sequência com as estruturas existentes .....	32
Figura 3 – Etapas da modelagem comparativa.....	34
Figura 4 – Etapas da modelagem por <i>Threading</i> .....	36
Figura 5 – Proteína GumD com interação na membrana (Psipred - Predict Protein Secondary Structure).....	54
Figura 6 – Resultados do ScanProsite. ....	55
Figura 7 – Gráfico Ramachandran do modelo previsto para a putativa GumD de <i>G. diazotrophicus</i> .....	57
Figura 8 – Modelo previsto para a putativa GumD de <i>G. diazotrophicus</i> .....	67
Figura 9 – Modelo previsto para a putativa GumD de <i>G. diazotrophicus</i> , evidenciando os sítios ativos (ScanProsite).....	67
Figura 10 – Alinhamento da proteína GumD de <i>G. diazotrophicus</i> com a proteína “biossíntese de exopolissacarídeo poliprenil glicosilfosfotransferase” das espécies <i>G. tumulicola</i> , <i>N. vanlangensis</i> , <i>K. baliensis</i> e <i>A. estunensis</i> . ....	70
Figura 11 – Domínios conservados das sequências de biossíntese de exopolissacarídeo poliprenil glicosilfosfotransferase.....	71
Figura 12 – Rede de interação da proteína GumD de <i>G. diazotrophicus</i> PAL 5... ..	73
Figura 13 – Vizinhança da rede de interação da GumD com as demais proteínas. ....	85
Figura 14 – Fusões genéticas da rede de interação da GumD com as demais proteínas. ....	86
Figura 15 – Coocorrência entre genomas da rede de interação da GumD com as demais proteínas.....	88
Figura 16 – Co-expressão da rede de interação da GumD com as demais proteínas. ....	900
Figura 17 – Cluster gerada pelo STRING.....	92
Figura 18 – Clusters que evidenciam os processos biológicos. ....	94
Figura 19 – Clusters que evidenciam as funções moleculares.....	95
Figura 20 – Clusters que evidenciam os agrupamentos de redes locais.....	96
Figura 21 – Clusters que evidenciam as vias metabólicas e de sinalização presentes no banco de dados KEGG.. ....	97
Figura 22 – Clusters que evidenciam as Anotações das palavras-chaves (UniProt).....	99
Figura 23 – Clusters que evidenciam os Domínios de proteínas (Pfam) .....	100

## LISTA DE TABELAS

Tabela 1 – Interação de bactérias promotoras de crescimento de plantas (BPCV) e seus efeitos benéficos às plantas. ....	17
Tabela 2 – Vantagens e desvantagens da modelagem <i>Threading</i> . ....	37
Tabela 3 – Componentes e definição dos elementos envolvidos em uma rede de interação. ....	41
Tabela 4 – Principais bancos de dados de redes de interação. ....	42
Tabela 5 – Informações acessadas no STRING. ....	43
Tabela 6 – Informações do STRING relacionados às interações proteína-proteína. ....	44
Tabela 7 – Aminoácidos discrepantes ou fora do gráfico de Ramachandran. ..	56
Tabela 8 – Espécies e identidade da proteína em relação proteína GumD da <i>G. diazotrophicus</i> . ....	68
Tabela 9 – Processo biológico (Ontologia gênica-GO). ....	72
Tabela 10 – Evidências de ligação funcional entre GumD de <i>G. diazotrophicus</i> PAL5 com as demais proteínas da rede. ....	74
Tabela 11 – Domínios de proteína (Pfam). ....	78
Tabela 12 – Informações gerais das proteínas da rede de interação gerada pelo STRING. ....	80
Tabela 13 – Clusters identificados e relacionados com a rede de interação de GumD. ....	81
Tabela 14 – Informações gerais das proteínas do Clusters. ....	92

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	10
<b>2 REVISÃO DE LITERATURA</b> .....	14
<b>2.1 Bactérias Promotoras de Crescimento Vegetal (BPCV)</b> .....	14
<b>2.1.1 Fixação de nitrogênio</b> .....	15
<b>2.1.2 Liberação de compostos de crescimento vegetal</b> .....	15
<b>2.1.3 Solubilização de nutrientes</b> .....	20
<b>2.1.4 Produção de compostos antimicrobianos</b> .....	21
<b>2.2 Aspectos gerais da Interação bactéria-planta</b> .....	22
<b>2.2.1 Biofilme</b> .....	25
<b>2.3 <i>Gluconacetobacter diazotrophicus</i> PAL5</b> .....	27
<b>2.4 Modelagem de Proteínas</b> .....	30
<b>2.4.1 Métodos dependentes de molde</b> .....	33
<b>2.4.1.1 Modelagem comparativa</b> .....	33
<b>2.4.1.2 Threading</b> .....	34
<b>2.4.1.2.1 I-TASSER</b> .....	35
<b>2.4.1.2.2 Modelagem por Threading no I-TASSER</b> .....	36
<b>2.5 Redes de Interação</b> .....	38
<b>2.5.1 Componentes de uma rede de interação</b> .....	41
<b>2.5.2 STRING</b> .....	42
<b>2.5.2.1 Clusters</b> .....	45
<b>3 MATERIAIS E MÉTODOS</b> .....	47
<b>3.1 Identificação da sequência e sua análise</b> .....	47
<b>3.2 Modelagem da proteína “GumD”</b> .....	47
<b>3.2.1 Refinamento e avaliação do modelo</b> .....	48
<b>3.2.2 Análise dos modelos</b> .....	49
<b>3.3 Busca e alinhamento de sequências</b> .....	49
<b>3.4 Construção de interação</b> .....	50
<b>3.5 Rede de interação e Clustres</b> .....	51
<b>4 RESULTADOS E DISCUSSÕES</b> .....	52
<b>4.1 Caracterização da proteína GumD</b> .....	52
<b>4.2 Modelagem da proteína GumD</b> .....	55
<b>4.2.1 Análise do modelo gerado</b> .....	55

<b>4.3 Modelo de GumD</b>	66
<b>4.4 Análise comparativa da sequência</b>	68
<b>4.5 Rede de interação</b>	71
<b>4.6 Análise das interações entre os componentes do interactoma de GumD</b>	82
<b>4.6.1 Vizinhaça no Genoma</b>	82
<b>4.6.2 Fusões genéticas</b>	84
<b>4.6.3 Coocorrência entre genomas</b>	87
<b>4.6.4 Co-expressão</b>	89
<b>4.6.5 Dados Experimentais/Bioquímicos</b>	90
<b>4.6.6 Associação em bancos de dados selecionados</b>	90
<b>4.7 Clusters</b>	91
<b>5 CONCLUSÃO</b>	101
<b>REFERÊNCIAS</b>	102

## 1 INTRODUÇÃO

O dogma central da biologia molecular, amplamente reconhecido pela comunidade científica, descreve o fluxo de informação genética nas células, seguindo uma direção específica: do DNA para o RNA e, posteriormente, para as proteínas. Esse dogma é composto por três etapas principais: replicação do DNA, transcrição e tradução. Compreender esses mecanismos é fundamental para entender os processos biológicos e seu funcionamento (PATIÑO & PINEDA, 2006). Além disso, esse entendimento é essencial para a pesquisa científica, possibilitando a manipulação genética e a engenharia de proteínas, impulsionando avanços nas áreas da medicina, agricultura e biotecnologia (VERLI, 2014).

As proteínas são macromoléculas essenciais para o funcionamento de todos os organismos vivos. Elas realizam uma variedade de funções vitais, desde a catalisação de reações químicas até a contribuição para a estrutura celular, além de participar em processos de sinalização e transporte. A relevância das proteínas abrange desde processos metabólicos e de sinalização até defesa imunológica, crescimento e reparação (CHU et al., 2021).

Nesse contexto, a bioinformática surge como uma ferramenta fundamental no estudo das proteínas. Ela se baseia na aplicação de técnicas da informática para a análise de informações nas áreas de biologia e campos relacionados (VERLI, 2014). As análises e interpretações de dados biológicos dependem das ferramentas de bioinformática, permitindo aos pesquisadores explorarem e compreenderem melhor a estrutura, função e complexidade dos sistemas biológicos. Essas ferramentas impulsionam avanços na genômica, proteômica e biologia de sistemas, contribuindo para a compreensão dos mecanismos biológicos e o desenvolvimento de aplicações práticas (ESPINDOLA et al., 2010).

Por sua vez, a biologia de sistemas adota uma abordagem interdisciplinar, buscando compreender os sistemas biológicos como um todo, em vez de analisar apenas seus componentes individuais. Essa abordagem tem como objetivo compreender como as partes moleculares, celulares e os organismos interagem e se organizam para produzir os processos biológicos complexos. Para alcançar esse objetivo, são utilizados métodos experimentais variados, como genômica, proteômica e metabolômica, além de recursos computacionais avançados, como modelagem matemática, simulação e análise de redes complexas. A integração de

dados experimentais em diferentes níveis de organização biológica, desde moléculas e células individuais até tecidos e organismos inteiros, é fundamental nessa abordagem (POLONI et al., 2014).

No estudo das proteínas, existem diversas técnicas e abordagens que podem ser utilizadas, tais como expressão e purificação da proteína, caracterização estrutural, análise funcional, estudos de interações proteína-proteína, estudos de expressão gênica e regulação, e modelagem computacional, entre outras. A escolha das técnicas a serem aplicadas depende do objetivo específico da pesquisa e da disponibilidade de recursos e expertise dos pesquisadores. Geralmente, é necessário combinar diferentes métodos para obter uma compreensão completa da proteína e de sua função (EMIDIO et al., 2015).

Dentre as várias técnicas de estudo das proteínas, destaca-se a modelagem de proteínas por homologia. Essa abordagem computacional auxilia na obtenção do modelo tridimensional de proteínas, com base em proteínas homólogas. Através da modelagem de proteínas por homologia, é possível compreender a estrutura e função da proteína, bem como realizar previsões de funções e interações proteicas, entre outros aspectos relevantes (SANTOS FILHO, 2003; CALIXTO, 2013).

A modelagem de proteínas por homologia é uma técnica especialmente útil quando não é possível determinar experimentalmente a estrutura tridimensional da proteína de interesse. Por meio da comparação da sequência de aminoácidos da proteína-alvo com proteínas homólogas já estruturalmente conhecidas, a modelagem por homologia permite inferir a estrutura tridimensional da proteína-alvo com base na estrutura das proteínas homólogas. Essa abordagem oferece uma visão detalhada da proteína, permitindo a análise de sua estrutura e função de maneira virtual, o que é valioso para a compreensão de diversos processos biológicos (SANTOS FILHO, 2003; CALIXTO, 2013).

Vale enfatizar que, na maioria dos casos, as proteínas não atuam isoladamente, mas sim em conjunto com outras proteínas, em uma complexa rede de interações. Essas interações podem ocorrer tanto entre proteínas do mesmo tipo quanto entre proteínas diferentes, e podem ser físicas, envolvendo contatos diretos entre as proteínas, funcionais, baseadas em associações indiretas, ou ainda interações transientes e temporárias (PETREY & HONIG, 2014; RAO et al., 2014).

Dentro deste contexto, torna-se evidente a importância do estudo das proteínas bacterianas. Essas proteínas são fundamentais em diversas áreas, como

a compreensão da patogenicidade bacteriana, a identificação de alvos terapêuticos, o aumento da resistência das plantas, a exploração de mecanismos de biocontrole e a garantia da segurança alimentar. O conhecimento obtido por meio desses estudos é essencial para o desenvolvimento de métodos eficazes de manejo de doenças de plantas causadas por bactérias e para a preservação das culturas agrícolas (BÜTTNER, 2016).

Além disso, é imprescindível destacar a presença das chamadas Bactérias Promotoras do Crescimento Vegetal (BPCV). Esses microorganismos têm a capacidade de interagir de forma benéfica com as plantas, resultando em um impacto positivo em seu crescimento e desenvolvimento. Essa interação pode ser mediada pelas proteínas bacterianas, que atuam de maneira direta ou indireta em benefício do crescimento vegetal. Essas bactérias podem ser naturalmente encontradas no solo, nas raízes das plantas e em outros ambientes relacionados à vegetação (KUMAR et al., 2018).

As interações entre as plantas e os microrganismos têm um papel significativo na rizosfera, influenciando processos como a fixação de carbono, o funcionamento do ecossistema e o ciclo de nutrientes. Essas interações podem ocorrer de forma benéfica ou patogênica, envolvendo tanto associações mutualísticas/comensais quanto organismos de vida livre. Para as plantas, essas interações não são apenas essenciais em termos de nutrição, promoção do crescimento, biocontrole e alívio do estresse, mas também têm impacto na física do solo (VELMOUROUGANE et al., 2017).

Dentro desse contexto, destaca-se a bactéria *Gluconacetobacter diazotrophicus* PAL5, conhecida por ser uma diazotrófica endofítica tolerante ao ar, desempenhando um papel fundamental na fixação biológica de nitrogênio (FBN) em plantas como o arroz (*Oryza sativa* L.). Além da FBN, essas bactérias estão relacionadas com o metabolismo de carboidratos, biossíntese de auxinas e produção de gluconacina, o que as caracteriza como BPCV (SILVA et al., 2020).

Nesse sentido, a proteína selecionada para este trabalho é a "Putative polysaccharide biosynthesis glycosyltransferase (tradução: Biossíntese putativa de polissacarídeo glicosiltransferase)", codificada pelo gene *gumD* do organismo *G. diazotrophicus* da estirpe PAL5, com base no estudo realizado por Meneses et al. (2011). Essa pesquisa forneceu evidências de que o gene *gumD* é essencial para a produção de Exopolissacarídeos (EPS - Extracellular polymeric substances), sendo

fundamental na formação de biofilme e colonização de plantas. A interação planta-bactéria desempenha um papel crucial nesse contexto, uma vez que o estudo das proteínas bacterianas envolvidas nessa interação é fundamental para a compreensão da patogenicidade bacteriana, o desenvolvimento de métodos agrícolas sustentáveis e a melhoria da produtividade das culturas.

Diante desse cenário, o presente estudo tem como objetivo geral compreender a função da proteína GumD em *G. diazotrophicus* PAL5, relacionada à interação com as plantas, por meio da modelagem e construção de redes de interação dessa proteína. Para atingir esse objetivo, primeiramente, obteve-se a estrutura tridimensional da proteína GumD. Em seguida, procedeu-se à construção de uma rede de interação proteína-GumD, identificando as proteínas envolvidas nessa rede e criando clusters para análise funcional das interações.

A busca por esses objetivos visa contribuir significativamente para o entendimento da simbiose entre *G. diazotrophicus* PAL5 e as plantas, resultando em avanços relevantes na área de biotecnologia agrícola e no desenvolvimento de soluções sustentáveis para o cultivo de alimentos. Ademais, a partir da modelagem da proteína e da análise das interações proteicas, busca-se compreender os mecanismos subjacentes a essa interação, contribuindo, assim, para o progresso do conhecimento nessa importante área de pesquisa.

## 2 REVISÃO DE LITERATURA

### 2.1 Bactérias Promotoras de Crescimento Vegetal (BPCV)

As BPCV (Bactérias Promotoras do Crescimento de Plantas) compreendem um grupo de espécies de diversos gêneros, incluindo bactérias endofíticas, organismos de vida livre e bactérias simbióticas (GLICK, 2012). Essas bactérias possuem participação determinante no desenvolvimento das plantas por meio de mecanismos diretos e indiretos. Os mecanismos diretos envolvem a facilitação da absorção de nutrientes e a síntese de fitohormônios que afetam diretamente o crescimento das plantas (NOVO et al., 2018). Já os mecanismos indiretos compreendem uma série de eventos biológicos que atuam indiretamente no controle de fitopatógenos, por meio da produção de compostos antagonistas ou da indução de resistência sistêmica (SARAF et al., 2014).

Essas bactérias podem ser classificadas com base em suas propriedades funcionais, tais como: (1) Biofertilizantes (aumentando a disponibilidade de nutrientes para as plantas); (2) Fitoestimuladores (promovendo o crescimento e desenvolvimento das plantas através da produção de hormônios); (3) Rizorremediadores (atuando na quebra de moléculas de poluentes orgânicos e modulando a solubilização de metais) e (4) Biopesticidas (auxiliando no controle de doenças e patógenos de plantas pela síntese de antibióticos e metabólitos antifúngicos) (SOMERS et al., 2004; NOVO et al., 2018; SANTOS DIAS & SANTOS, 2022).

O uso dessas bactérias pode representar uma alternativa segura e ambientalmente correta a ser adotada no sistema de produção agrícola, contribuindo para a sustentabilidade dos agroecossistemas. Nos últimos anos, sua utilização tem aumentado devido aos seus múltiplos benefícios, que incluem a mitigação dos efeitos estressantes do déficit hídrico e da salinidade, a fitoextração de metais pesados, o equilíbrio hormonal e o biocontrole de fitopatógenos (SANTOS DIAS & SANTOS, 2022).

As BPCV possuem diversos mecanismos de ação que podem auxiliar no crescimento e na saúde das plantas de diversas maneiras. Alguns desses mecanismos incluem: Fixação de nitrogênio; Liberação de compostos de crescimento vegetal; Solubilização de nutrientes; Produção de compostos antimicrobianos, entre outros (SANTOS DIAS & SANTOS, 2022).

Além dos mecanismos já mencionados, é importante destacar a notável capacidade das BPCV de realizar a fixação biológica de nitrogênio, um processo essencial para fornecer esse nutriente vital às plantas. A fixação biológica de nitrogênio será discutida a seguir.

### **2.1.1 Fixação de nitrogênio**

Algumas bactérias promotoras de crescimento possuem a capacidade de realizar a fixação do nitrogênio atmosférico, convertendo-o em formas utilizáveis pelas plantas, como a amônia. Esse processo é fundamental para fornecer uma fonte adicional de nitrogênio, um nutriente essencial para o crescimento vegetal (MARIN et al., 1999).

A fixação biológica do nitrogênio é mediada pela enzima dinitrogenase, presente nas bactérias capazes de realizar essa fixação. Essa enzima tem a capacidade de quebrar a tripla ligação do nitrogênio atmosférico, reduzindo-o a amônia ( $\text{NH}_3$ ), seguindo um processo semelhante ao realizado na indústria. Essas bactérias podem ser classificadas como associativas, endofíticas ou simbióticas, dependendo de sua especificidade em relação a diferentes espécies de plantas (HUNGRIA et al., 2001).

Dentre os diversos gêneros de bactérias promotoras de crescimento, destacam-se: *Azospirillum*, *Azoarcus*, *Bacillus*, *Burkholderia*, *Clostridium*, *Enterobacter*, *Gluconacetobacter*, *Herbaspirillum*, *Pseudomonas* e *Serratia*, entre outros (SOMERS et al., 2004). Esses gêneros demonstram a diversidade e importância das bactérias promotoras de crescimento na agricultura e no desenvolvimento das plantas.

Além da fixação biológica de nitrogênio, as bactérias promotoras de crescimento também desempenham um papel importante na liberação de compostos de crescimento vegetal. Essas substâncias fornecem estímulos essenciais para o desenvolvimento e o crescimento das plantas. No próximo tópico, será explorada a relevância da liberação de compostos de crescimento vegetal pelas BPCV.

### **2.1.2 Liberação de compostos de crescimento vegetal**

Algumas BPCV têm a capacidade de produzir e liberar compostos de crescimento vegetal, como auxinas, citocininas e giberelinas. Essas substâncias

atuam como hormônios vegetais e têm o potencial de melhorar o crescimento das raízes, a formação de folhas e o desenvolvimento geral das plantas (SANTOS DIAS & SANTOS, 2022).

De acordo com Patten et al. (1996), sugerem que a molécula precursora triptofano é necessária para a produção do ácido indol-3-acético (AIA), conhecido como a principal auxina. Quando estabelecem interações com as plantas, várias espécies de microrganismos são capazes de sintetizar essas auxinas. Essa capacidade também é destacada por Solano et al. (2008), que explicam que as bactérias utilizam um metabolismo secundário para a detoxificação de compostos, resultando na produção de triptofano e, conseqüentemente, no aumento da produção de auxinas.

Outros hormônios vegetais, conhecidos como fitormônios, incluem as citocininas e as giberelinas. As citocininas são fundamentais na promoção da citocinese, ou seja, a divisão celular, e no desenvolvimento de brotos laterais. Esses fitohormônios são produzidos nas raízes e transportados para todas as partes da planta através do xilema. Além das raízes, as citocininas também são produzidas em frutos e embriões (DEMASON, 2005).

Por sua vez, o ácido giberélico é uma das giberelinas mais conhecidas. De acordo com Castro et al. (2005), as giberelinas causam alterações fisiológicas significativas nas plantas, incluindo interferência na floração, expressão sexual, senescência, descamação, germinação e superação da dormência.

Na Tabela 1, são apresentados os diversos mecanismos de ação das bactérias promotoras de crescimento de plantas em monocotiledôneas e dicotiledôneas, conforme descrito na literatura.

Além dos efeitos das auxinas, citocininas e giberelinas no crescimento e desenvolvimento das plantas, as bactérias promotoras de crescimento vegetal também possuem um papel-chave na solubilização de nutrientes. Por meio de mecanismos específicos, essas bactérias são capazes de liberar nutrientes presentes no solo, tornando-os disponíveis para as plantas. A solubilização de nutrientes é um processo importante para melhorar a absorção e utilização eficiente dos nutrientes pelas plantas, resultando em um crescimento mais saudável e vigoroso (SANTOS DIAS & SANTOS, 2022).

**Tabela 1 – Interação de bactérias promotoras de crescimento de plantas (BPCV) e seus efeitos benéficos às plantas.**

<b>Espécie vegetal</b>	<b>BPCV</b>	<b>Efeito ou mecanismo</b>	<b>Referência</b>
Aveia ( <i>Avena sativa</i> )	<i>Klebsiella</i> sp.	Regula do conteúdo de íons e níveis de prolina	Sapre et al. (2018)
Cevada ( <i>Hordeum vulgare</i> )	<i>Curtobactéria</i> sp.	Regula o conteúdo de prolina	Cardinale et al. (2015)
	<i>Hartmannibacter diazotrophicus</i>	Aumenta a atividade da ACC deaminase	Suarez et al. (2015)
Milho ( <i>Zea mays</i> )	<i>Bacillus atrophaeus</i>	Alivia o estresse salino	Kerbab et al. (2021)
	<i>Azotobacter</i> sp.	Promove a absorção de nutrientes nas plantas	Rojas-Tapias et al. (2012)
	<i>Bacillus amyloliquefaciens</i> SQR9	Melhora atividade da enzima antioxidante e aumenta a expressão de genes de resposta ao estresse salino	Chen et al. (2016)
	<i>Bacillus</i> sp.	Aumento das atividades enzimáticas e dos teores de prolina e açúcar solúvel sob estresse salino; regula a atividade da ACC deaminase	Gupta et al. (2019)
	<i>Pseudomonas</i> sp.	Aprimora o conteúdo de prolina e AIA e a produção de exopolissacarídeos	Nadeem et al. (2009)
	<i>Rhizobium</i> sp.	Regula a biossíntese de pigmentos	Bano et al. (2009)
Arroz ( <i>Oryza sativa</i> )	<i>Bacillus aryabhatai</i> , <i>Achromobacter denitrificans</i> e <i>Ochrobactrum intermedium</i>	Alivia o estresse salino	Sultana et al. (2020)

Espécie vegetal	BPCV	Efeito ou mecanismo	Referência
	<i>Bacillus pumilus</i> e <i>Pseudomonas pseudoalcaligenes</i>	Aumenta a absorção de nutrientes	Jha et al. (2013)
	<i>Glutamicibacter</i> sp. YD01	Regula os níveis de etileno	Ji et al. (2020)
	<i>Micrococcus</i> sp.	Aumentando os níveis de AIA	Sukweenadhi et al. (2015)
	<i>Rhizobacteria pseudomonas</i>	Regula a atividade da ACC deaminase	Jha et al. (2013)
Trigo ( <i>Triticum aestivum</i> )	<i>Aeromonas</i> sp.	Regula o teor de etileno e alivia o estresse salino	Ashraf et al. (2004)
	<i>Bacillus</i> sp.	Regulando a atividade da ACC deaminase	Pourbabaee et al. (2016)
	<i>Pseudomonas fluorescence</i> , <i>Bacillus pumilus</i> e <i>Exiguobacterium aurantiacum</i>	Aumenta a produção de substâncias osmóticas	Nawaz et al. (2020)
	<i>Klebsiella</i> sp.	Regulação do conteúdo de íons, níveis de prolina e atividade de enzimas antioxidantes	Singh & Jha (2017)
Pepino ( <i>Cucumis sativus</i> )	<i>Bacillus</i> sp.	Ajusta a atividade da ACC deaminase	Nadeem et al. (2016)
	<i>Burkholderia</i> sp	Mantém o equilíbrio hídrico e regulando o conteúdo de pigmento fotossintético	Kang et al. (2014)
Morango ( <i>Fragaria ananassa</i> )	<i>Kocuria</i> sp.	Aumenta a solubilização de fosfato	Kore et al. (2013)

<b>Espécie vegetal</b>	<b>BPCV</b>	<b>Efeito ou mecanismo</b>	<b>Referência</b>
Algodão ( <i>Gossypium hirsutum</i> )	<i>Pseudomonas</i> sp.	Aumenta a produção de conteúdo de prolina, AIA e exopolissacarídeos	Yao et al. (2010)
Tomate ( <i>Solanum lycopersicum</i> )	<i>Rizobactérias promotoras de crescimento</i>	Alivia o estresse hídrico e aumenta a absorção de K <sup>+</sup>	Cordero et al. (2018)
	<i>Achromobacter</i> sp.	Regula a produção de etileno	Mayak et al. (2004)
	<i>Enterobacter</i> sp.	Reduz a produção de etileno	
	<i>Sphingomonas</i> sp.	Induz a produção de exopolissacarídeos e prolina	Halo et al. (2015)
Ervilha ( <i>Pisum sativum</i> )	<i>Arthrobacter</i> sp.	Aumenta a absorção de nutrientes	Barnawal et al. (2014)
	<i>Rhizobium</i> sp.	Regula a síntese de pigmentos	
	<i>Variovorax</i> sp.	Aumenta a atividade da ACC deaminase	Wang et al. (2016)

**Fonte:** Adaptado de Gao et al. (2022).

No próximo tópico será explorado a capacidade das bactérias promotoras de crescimento de solubilizar nutrientes, os mecanismos envolvidos nesse processo e os benefícios para o crescimento e nutrição das plantas. A solubilização de nutrientes é mais um aspecto crucial realizado pelas bactérias promotoras de crescimento vegetal na promoção do desenvolvimento sustentável das plantas e na otimização da produção agrícola.

### **2.1.3 Solubilização de nutrientes**

Algumas bactérias promotoras de crescimento possuem a capacidade de solubilizar nutrientes, como o fosfato ( $\text{PO}_4^{3-}$ ), presente no solo, mas que não é facilmente assimilado pelas plantas. Essas bactérias solubilizadoras de fosfato realizam um papel importante ao transformar o fósforo (P) inorgânico em formas solúveis, disponibilizando-o para as plantas como um macronutriente. Esse processo ocorre por meio da liberação de enzimas que convertem os compostos inorgânicos insolúveis em substâncias solúveis que as raízes das plantas podem absorver (AMARAL, 2014).

O fósforo é o segundo mineral mais importante para o crescimento das plantas, logo após o nitrogênio. No entanto, apesar de sua abundância, a maior parte do fósforo no solo está em formas insolúveis, como mencionado anteriormente. O fósforo desempenha um ação fundamental na regulação de vias metabólicas e reações bioquímicas em seu papel fisiológico (BRAMUSSE; BARROSO; INTORNE, 2017).

Devido à baixa disponibilidade de fósforo para as plantas nas formas orgânicas e inorgânicas encontradas no solo, a aplicação de fertilizantes fosfatados em altas concentrações aumenta os custos de produção agrícola (HUNGRIA et al., 2011). Nesse contexto, a aplicação de microrganismos solubilizadores de fosfato surge como uma alternativa viável ao uso de fertilizantes fosfatados em áreas agrícolas. Essa abordagem melhora a eficiência dos sistemas de cultivo, promove o crescimento das plantas e minimiza os impactos negativos para o meio ambiente causados pelo uso excessivo de fertilizantes (RAFI et al., 2019).

Além da solubilização de nutrientes, as bactérias promotoras de crescimento também são importantes na produção de compostos antimicrobianos. Esses compostos são substâncias que ajudam as plantas a combater patógenos, como

fungos, bactérias e vírus, proporcionando um efeito protetor contra doenças (SANTOS DIAS & SANTOS, 2022).

No próximo tópico, aprofundaremos a produção de compostos antimicrobianos pelas BPCVS.

#### **2.1.4 Produção de compostos antimicrobianos**

Os produtos antimicrobianos são substâncias produzidas por certas bactérias promotoras de crescimento dispõe de uma responsabilidade crucial no controle de doenças e pragas de plantas. Diante da preocupação com os impactos negativos dos agrotóxicos no meio ambiente e nos custos de produção de alimentos, os microrganismos endofíticos têm despertado um interesse crescente como uma alternativa promissora para o biocontrole de fitopatógenos em diferentes sistemas agrícolas (SANTOS, 2014).

Esses microrganismos endofíticos, como as espécies de *Pseudomonas spp.* e *Bacillus sp.*, são capazes de produzir uma ampla variedade de compostos fitopatógenos com propriedades antimicrobianas, como antibióticos, enzimas de degradação de patógenos e lipopeptídeos. Esses compostos são capazes de suprimir o crescimento e a proliferação de patógenos de plantas, auxiliando no controle de doenças e no aumento da resistência das plantas (ARAÚJO et al., 2010).

A dependência excessiva de agrotóxicos nas lavouras para o controle fitossanitário resulta em altos custos de produção de alimentos e impactos negativos no meio ambiente quando não são manejados adequadamente (BISOGNIN & AUGUSTIN, 2006). Diante desse cenário, há um crescente interesse no uso de microrganismos endofíticos como uma alternativa promissora para o biocontrole de fitopatógenos em diversos agroecossistemas. Esses microrganismos endofíticos são antagonistas naturais de vários patógenos, representando uma fonte importante e ainda pouco explorada de agentes de biocontrole para doenças e pragas de plantas (SESSITSCH et al., 2004; SCHULZ & BOYLE, 2005).

Diversos experimentos de pesquisa descritos na literatura relatam a supressão de patógenos mediada por antibióticos produzidos por microrganismos, destacando-se principalmente as espécies *Pseudomonas spp.* que produzem compostos como 2,4-diacetil-floroglucinol (KEEL et al., 1990; 1992; DE SOUZA et al., 2003; RAI et al., 2017), fenazina (THOMASHOW et al., 1990; CHIN-A-WOENG et al., 1998; MAZURIER et al., 2009), pioluteorina (HASSAN et al., 2011; RAMETTE

et al., 2011), pirrolnitrina (UPADHYAY et al., 2011; HUANG et al., 2018) e cianeto de hidrogênio (LANTEIGNE et al., 2012). Além disso, algumas espécies do gênero *Bacillus* são conhecidas por produzirem compostos como bacilomicina D (XU et al., 2013), micoss subtilina (MIHALACHE et al., 2018; RADOVANOVIĆ et al., 2018) e outros lipopeptídeos, incluindo iturina A, fengicina e surfactina (GUEVARA-AVENDAÑO et al., 2020).

No próximo tópico, será abordado sobre a interação entre as bactérias promotoras de crescimento e as plantas. Essa interação envolve uma série de processos complexos que vão desde o reconhecimento até a comunicação entre os microrganismos e as células vegetais. As bactérias endofíticas estabelecem uma relação simbiótica com as plantas, promovendo benefícios mútuos que incluem o controle de fitopatógenos e o estímulo do crescimento vegetal. Além disso, será discutido como essas bactérias conseguem colonizar as plantas, os mecanismos envolvidos na supressão de patógenos e como essas interações podem ser exploradas de forma eficiente para aprimorar a agricultura sustentável.

## **2.2 Aspectos gerais da Interação bactéria-planta**

A maioria das bactérias endofíticas está presente nas plantas vasculares, colonizando os tecidos internos de seus hospedeiros (FIRÁKOVÁ et al., 2007; XIE et al., 2020). Elas pertencem a diferentes filos bacterianos, incluindo *Bacteroidetes*, *Actinobacteria*, *Firmicutes* e *Alfa-Proteobacteria* (BACON & HINTON, 2006). Esses microrganismos podem ser encontrados tanto intracelularmente como intercelularmente, sem causar danos visíveis à planta (GOUVEIA et al., 2020; JOO et al., 2021), ou em pelo menos uma fase de seu ciclo de vida (HARDOIM et al., 2008).

A colonização das bactérias endofíticas ocorre em três etapas distintas. Primeiro, a bactéria se move em direção às raízes da planta, seja de forma passiva ou através do fluxo de água no solo. Em alguns casos, elas podem se mover ativamente utilizando seus flagelos, impulsionadas por substâncias liberadas por elas (SPRENT & DEFARIA, 1988). Em seguida, as bactérias aderem às raízes por meio de uma adsorção não específica (SPRENT & DEFARIA, 1988; FIRDOUS et al., 2019). Uma vez conectadas às raízes, as bactérias se multiplicam, formando "microcolônias" (HARDOIM et al., 2008; LIU et al., 2017). Por fim, as bactérias

endofíticas podem alcançar locais específicos nos tecidos vegetais danificados, que podem ocorrer naturalmente durante o crescimento da planta, em junções epidérmicas, pelos radiculares ou por meio de ferimentos (SPRENT & DEFARIA, 1988). A colonização no interior da planta pode ser sistêmica, migrando pelo sistema vascular e apoplasto, ou local, migrando pelos tecidos do córtex e xilema (HALLMANN et al., 1997).

A produção de uma ampla variedade de enzimas é uma característica importante das bactérias endofíticas, incluindo celulases, pectinases, xilanases, liquenases, mananases e outras enzimas hidrolíticas (HALLMANN et al., 1997; CHO et al., 2010; SINGH et al., 2017). Essas enzimas possuem função essencial na colonização endofítica, sendo responsáveis pela degradação de componentes da parede celular vegetal. Essa degradação enzimática ocorre durante a colonização da epiderme radicular, mas não nos espaços intercelulares do córtex. No entanto, ainda há muito a ser descoberto sobre a regulação desses processos. Sabe-se, no entanto, que a liberação constitutiva de hidrolases é indesejável, pois pode levar à patogenicidade da planta (HALLMANN et al., 1997).

A densidade populacional das bactérias endofíticas varia amplamente e depende principalmente da espécie bacteriana, do genótipo da planta, do estágio de desenvolvimento do hospedeiro e das condições ambientais. Essa interação planta-microrganismo tem sido observada em várias partes da planta, como sementes, óvulos, frutos, caules, folhas, raízes, tubérculos, brotos, flores e inflorescências (ZHANG et al., 2006; MIGUEL et al., 2016). A entrada desses microrganismos é provavelmente controlada pela planta e pelas condições ambientais (SILVA et al., 2006; MIGUEL et al., 2019).

No processo conhecido como exsudação, as bactérias liberam substâncias no espaço extracelular, um mecanismo crucial para interações biológicas. Esse processo envolve a biossíntese de compostos como os sideróforos, os quais são posteriormente liberados no ambiente circundante à bactéria. Um sideróforo é um composto orgânico que desempenha um papel essencial ao auxiliar organismos, especialmente bactérias, na captação de ferro do ambiente. São moléculas de baixo peso molecular que se ligam ao ferro, tornando-o solúvel e, assim, facilitando seu transporte para dentro das células (SCAVINO & PEDRAZA, 2013).

No âmbito da interação entre bactérias e plantas, os sideróforos desempenham um papel duplo que pode ser benéfico ou prejudicial, dependendo dos tipos de bactérias e plantas envolvidos.

Algumas bactérias benéficas, como as *Pseudomonas fluorescens*, produzem sideróforos que potencializam o crescimento das plantas. Esses compostos auxiliam no saneamento do solo, removendo substâncias tóxicas, e também fornecem ferro às plantas. Além disso, essas bactérias competem com organismos patogênicos pelo ferro, o que reduz a probabilidade de infecção nas plantas. Em contrapartida, certas bactérias patogênicas, exemplificadas pela *Pseudomonas syringae*, produzem sideróforos que facilitam sua invasão nas plantas. Esses sideróforos removem o ferro de proteínas transportadoras da planta, como a ferritina, tornando a planta mais suscetível à invasão. Adicionalmente, algumas bactérias patogênicas podem utilizar os sideróforos produzidos pelas próprias plantas para obter ferro, aumentando sua virulência (SCAVINO & PEDRAZA, 2013; DIMKIĆ et al., 2022).

Assim, os sideróforos atuam como mediadores em uma complexa dança de interações entre bactérias e plantas, onde o resultado final pode variar de cooperação mutualística a competição prejudicial, dependendo dos agentes envolvidos e das estratégias adaptativas que empregam (SCAVINO & PEDRAZA, 2013).

Existem evidências de que as bactérias endofíticas podem desempenhar uma extrema relevância no controle de nematoides e na promoção do crescimento das plantas, por meio da produção de substâncias como auxinas, citocininas, giberelinas e metabólitos antimicrobianos (REINHOLD-HUREK & HUREK, 2011). Além disso, essa relação simbiótica oferece outras vantagens, como a proteção contra patógenos, através da indução dos mecanismos de defesa da planta (COMPANT et al., 2010; ELJOUNAIDI et al., 2016), aumento na produtividade das culturas, fixação de nitrogênio atmosférico e assimilação de fosfato solubilizado (ROSENBLUETH & MARTINEZ-ROMERO, 2006; CONTI, 2007; PADDA et al., 2019) conforme exposto em seções anteriores desta revisão de literatura.

Muitas bactérias endofíticas dos gêneros *Pseudomonas*, *Burkholderia* e *Bacillus* produzem metabólitos secundários, como antibióticos, compostos voláteis, antifúngicos, antivirais e inseticidas. Esses metabólitos tornam as bactérias endofíticas uma fonte importante de novos produtos naturais com potencial aplicação na biotecnologia. Além disso, essas bactérias são indispensáveis na

biorremediação e na fitorremediação, devido à sua capacidade de degradar uma ampla variedade de substâncias (RYAN et al., 2008; LIU et al., 2017).

Para compreender plenamente a importância das bactérias endofíticas na ecologia, é fundamental estudar essas interações e seu impacto nas plantas hospedeiras. O estudo dessas interações permite a compreensão dos mecanismos moleculares que regulam essas interações, sendo essencial para o desenvolvimento de métodos de controle de doenças de plantas causadas por patógenos bacterianos, o fortalecimento da resistência das plantas, a investigação de métodos de biocontrole e a garantia da segurança alimentar. A criação de estratégias eficazes de manejo de doenças de plantas causadas por bactérias é fundamental para preservar as culturas agrícolas (MIGUEL et al., 2021).

Nesse contexto, a relação entre o biofilme e a interação bactéria-planta são cruciais em diversos ecossistemas. Essas estruturas microbianas aderidas às superfícies têm o potencial de influenciar significativamente o crescimento e o desenvolvimento das plantas, bem como sua resistência a patógenos e estresses ambientais. Na subseqüente seção, será exposto uma análise da presença do biofilme no contexto da interação simbiótica entre bactérias e plantas.

### **2.2.1 Biofilme**

O biofilme (EPS) é uma estrutura complexa composta por comunidades de micro-organismos, principalmente bactérias, que aderem às superfícies e secretam uma matriz de substâncias poliméricas extracelulares. Essa matriz contém uma variedade de materiais, incluindo polissacarídeos, proteínas, ácidos nucleicos e lipídios (BOGINO et al., 2013).

O processo de formação do biofilme inicia quando as bactérias se agarram a superfícies sólidas, como paredes, próteses médicas, dentes ou superfícies industriais, através de mecanismos de adesão específicos. Após a adesão, essas bactérias proliferam e secretam a matriz extracelular de EPS, que atua na proteção da comunidade microbiana (NIAZI et al., 2023).

O biofilme apresenta várias vantagens adaptativas comprovadas por estudos científicos (DANHORN & FUQUA, 2007; VELMOUROUGANE et al., 2017; NIAZI et al., 2023):

1. **Proteção contra agentes antimicrobianos:** As bactérias que compõem o biofilme demonstram maior resistência ao tratamento com antibióticos e

outros compostos antimicrobianos devido à presença da matriz extracelular de substâncias poliméricas (EPS). Essa matriz dificulta a penetração de substâncias antimicrobianas, tornando mais desafiador o combate eficaz às bactérias.

2. **Proteção contra o sistema imunológico:** A matriz EPS também protege as bactérias do ataque do sistema imunológico do hospedeiro. Essa proteção dificulta a eliminação das bactérias pelo sistema de defesa do organismo, permitindo que permaneçam na superfície em que estão aderidas.
3. **Acesso a nutrientes:** O biofilme, concentrando moléculas e nutrientes essenciais para o crescimento bacteriano, proporciona às bactérias um suprimento constante de alimento, evitando escassez nutricional e favorecendo sua sobrevivência e proliferação.
4. **Cooperação e troca de informações:** O biofilme permite que diferentes espécies bacterianas coexistam e cooperem entre si. Essa coexistência favorece a troca de informações genéticas e o compartilhamento de recursos, conferindo à comunidade bacteriana vantagens competitivas em seu ambiente.
5. **Adesão estável:** A matriz EPS desempenha uma tarefa indispensável na adesão do biofilme à superfície em que se forma. Essa adesão estável confere ao biofilme uma maior resistência à remoção, garantindo sua permanência na área onde se desenvolve.

Deste modo, os Biofilmes podem ser úteis para vários organismos vegetais, principalmente para as raízes. Podendo se formar em torno das raízes e outros tecidos vegetais. São essenciais para o crescimento e a saúde das plantas por fornecer nutrientes e protegê-las contra fatores ambientais e patógenos. Essas interações simbióticas representam um papel fundamental na ecologia do solo e na saúde geral dos ecossistemas terrestres (NIAZI et al., 2023).

Apesar dos benefícios que os biofilmes proporcionam tanto para as bactérias quanto para as plantas, é essencial reconhecer que essas estruturas também podem acarretar desafios significativos em diversas situações. A formação de biofilmes em dispositivos médicos, como cateteres e implantes, pode resultar em infecções crônicas de difícil tratamento em ambientes médicos. Além disso, os biofilmes podem causar problemas em sistemas de tratamento de água e esgoto, bem como ocasionar danos a estruturas industriais e de construção.

Diante dessas considerações, fica evidente a importância de se aprofundar os estudos voltados para a análise do biofilme. O conhecimento detalhado dessas estruturas microbianas é fundamental para o desenvolvimento de estratégias eficazes de prevenção e controle de problemas relacionados aos biofilmes. Além disso, a compreensão dos mecanismos envolvidos na formação e estabilidade dos biofilmes pode abrir caminho para inovações na medicina, indústria e meio ambiente. Investir em pesquisas que abordem os aspectos benéficos e prejudiciais dos biofilmes contribuirá para promover aplicações mais seguras e eficientes dessas estruturas na ciência e tecnologia.

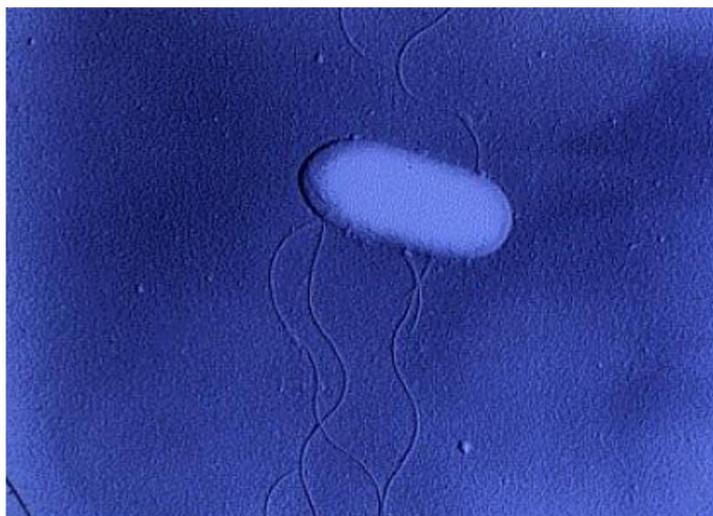
### **2.3 *Gluconacetobacter diazotrophicus* PAL5**

*G. diazotrophicus* PAL5, pertencente ao filo *Proteobacteria*, classe *Alphaproteobacteria*, ordem *Rhodospirillales* e família *Acetobacteraceae*, é uma bactéria endofítica que coloniza os tecidos internos de várias plantas, como cana-de-açúcar (*Saccharum spp.*), batata-doce (*Ipomoea batatas*), abacaxi (*Ananas comosus*), arroz (*Oryza sativa*), café (*Coffea arabica*) e outras (BALDANI & BALDANI, 2005; MUTHUKUMARASAMY et al., 2007). Além disso, também foi encontrada em insetos que infestam a cana-de-açúcar e em esporos de fungos em estado de repouso (MUTHUKUMARASAMY et al., 2002).

Essa bactéria é Gram-negativa, aeróbica obrigatória e possui células em forma de bacilo, com tamanhos variando entre (0,7 a 0,8) x (2 a 4)  $\mu\text{m}$ , como ilustrado na Figura 1. Sob o microscópio, as células podem ser observadas individualmente, em pares ou com estruturas de ligação, mas não apresentam esporos (MUTHUKUMARASAMY et al., 2002). Geralmente, a bactéria é isolada em meio semi-sólido LGI-P (REIS et al., 1994) e, após sete a dez dias, forma uma película alaranjada na superfície do meio, enquanto o meio abaixo da película permanece incolor devido à assimilação do azul de bromotimol (DÖBEREINER et al., 1993).

A bactéria *G. diazotrophicus*, descrita por Cavalcante e Döbereiner (1988), é conhecida por sua capacidade de fixação de nitrogênio em várias plantas. Essa bactéria utiliza o ácido 2-ceto glucônico como fonte de carbono, o que estimula o processo de fixação de nitrogênio (MENESES et al., 2011; SERRATO et al., 2013).

**Figura 1– *Gluconacetobacter diazotrophicus*.** Bactéria endofítica, aeróbia e em formato de bacilos.



Fonte: (SANTOS, 2007).

*G. diazotrophicus* é capaz de crescer em altas concentrações de sacarose (10% sacarose) e em um pH bastante baixo (3,0), além de apresentar a habilidade de fixar nitrogênio em condições microaerofílicas (CAVALCANTE e DOBEREINER, 1988). O pH ideal para o crescimento dessa bactéria está em torno de 5,5, embora os valores de pH e a taxa de respiração possam variar dependendo das fontes de carbono utilizadas (STEPHAN et al., 1991).

É de extrema relevância a realização de estudos que envolvam a bactéria *G. diazotrophicus*, conforme evidenciado pelo trabalho de Filgueiras (2015), que demonstrou os efeitos da inoculação de *G. diazotrophicus* no crescimento das plantas de arroz vermelho, bem como a resposta de tolerância sistêmica induzida (IST) à seca. Além disso, o estudo de Silva et al. (2020), também é relevante, pois seus resultados indicaram que a inoculação bacteriana melhorou a resistência à seca das plantas e regulou positivamente alguns genes relacionados ao desenvolvimento radicular sob condições de escassez de água.

Com base nessas pesquisas, é possível concluir que a inoculação da cepa *G. diazotrophicus* PAL5 desempenha um papel fundamental ao encorajar diversos mecanismos de crescimento e desenvolvimento de raízes, contribuindo para a resistência ao estresse hídrico nas plantas. Essa intervenção bacteriana possibilita o desenvolvimento e aprimoramento da composição bioquímica das plantas, demonstrando um potencial promissor para a melhoria da agricultura em condições adversas.

No estudo conduzido por Grillo-Puertas et al. (2018), foi empreendida uma investigação com o propósito de avaliar os impactos das concentrações de fosfato presentes no meio de cultivo sobre a regulação dos níveis intracelulares de polifosfato em *G. diazotrophicus*. Os resultados obtidos indicaram que a formação de biofilme em *G. diazotrophicus* sofre influência das concentrações de fosfato no ambiente de cultura, constatando-se que a tendência à formação de biofilme é amplificada em situações de maiores concentrações de fosfato. Essa fenomenologia desempenha um papel de relevância na colonização radicular e contribui para as multifáceis repercussões benéficas inerentes as Bactérias Promotoras de Crescimento Vegetal (BPCV).

Outro estudo, que focou na linhagem *G. diazotrophicus* PAL5, se concentrou nas dinâmicas da biossíntese de exopolissacarídeos (EPS) e nas suas implicações aplicadas. A pesquisa identificou que uma enzima bacteriana denominada tirosina quinase (Wzc) assume um papel crucial na síntese dos exopolissacarídeos. A tirosina quinase, sendo uma enzima presente em bactérias, é responsável por catalisar a transferência de grupos fosfato provenientes do trifosfato de adenosina (ATP) para resíduos de tirosina nas proteínas alvo. A autofosforilação da tirosina quinase regula diversas vias metabólicas e processos celulares dentro do contexto bacteriano, dentre os quais se inclui a biossíntese de exopolissacarídeos em *G. diazotrophicus* (WANDERLEY et al., 2021).

Por fim, a *G. diazotrophicus* PAL5 desempenha uma função importante e valiosa tanto na agricultura quanto no meio ambiente. Sendo uma bactéria diazotrófica, possui uma notável capacidade de fixar o nitrogênio atmosférico, essencial para o crescimento das plantas, tornando-se, assim, uma ferramenta poderosa para promover a sustentabilidade agrícola. A redução da dependência de fertilizantes nitrogenados sintéticos contribui para a diminuição da poluição ambiental e dos custos associados à agricultura convencional. Essa bactéria não só é benéfica para a agricultura, mas também para o meio ambiente, contribuindo para a promoção de práticas agrícolas mais ecologicamente equilibradas, que melhoram o bem-estar humano e garantem a preservação dos recursos naturais para as gerações futuras (WANDERLEY et al., 2019).

## 2.4 Modelagem de Proteínas

Cada célula de um organismo vivo pode conter milhares de proteínas, de funções específicas, sendo as macromoléculas mais abundantes. A estrutura tridimensional das proteínas, determinada pela sequência de aminoácidos, é fundamental para sua função. Por exemplo, o arranjo tridimensional pode influenciar a posição dos resíduos catalíticos no sítio ativo da proteína, bem como sua interação com outras moléculas. Portanto, a determinação da estrutura proteica é essencial para melhorar nossa compreensão de como a proteína funciona e oferece *insights* para manipular, controlar ou alterar suas funções (WOLYNES, 2015).

Os modelos de estruturas macromoleculares tridimensionais podem variar entre dados experimentais e previsões computacionais. Técnicas experimentais, como cristalografia de raios-X, ressonância magnética nuclear (RMN) e criomicroscopia eletrônica (cryo-EM), são frequentemente utilizadas para obter coordenadas atômicas precisas de átomos pesados (CARRONI & SAIBIL, 2016; SILVA et al., 2021). No entanto, essas técnicas exigem especialização, habilidades avançadas, recursos financeiros significativos e quantidades substanciais de amostras moleculares expressas e purificadas (SCHWEDE, 2013).

Apesar da confiabilidade dos modelos estruturais gerados por técnicas experimentais, o processo de resolução de estruturas experimentais apresenta desafios devido à velocidade com que novas sequências de proteínas são descobertas (BERMAN et al., 2000). Atualmente, o Banco de Dados de Proteínas (PDB) contém cerca de 207.338 estruturas resolvidas (julho/23), refletindo o crescente número de estruturas tridimensionais disponíveis.

É preciso acentuar que as técnicas computacionais colaboram na interpretação e refinamento dos dados experimentais, contribuindo significativamente para a construção de modelos estruturais. A modelagem por computador e as simulações moleculares são ferramentas valiosas para explorar a dinâmica e a função das proteínas, complementando os dados experimentais e permitindo a investigação de sistemas que ainda não foram resolvidos experimentalmente. A integração de abordagens experimentais e computacionais é essencial para o avanço do conhecimento sobre as estruturas proteicas e suas funções.

Por outro lado, é surpreendente considerar a quantidade de sequências de proteínas disponíveis no UniProt, que abrange o conjunto de proteínas revisadas no

UniProtKB/Swiss-Prot. O UniProtKB é um banco de dados de proteínas que contém informações detalhadas sobre diferentes proteínas de diversas espécies. Cada entrada de proteína no UniProtKB está associada a um resumo das informações funcionais, que podem ter sido verificadas experimentalmente ou previstas computacionalmente. Essas informações incluem dados sobre a função biológica da proteína, sua estrutura, localização celular, interações com outras moléculas, atividades enzimáticas, domínios estruturais e outras características relevantes. Já no UniProtKB/Swiss-Prot, cada entrada de proteína é cuidadosamente revisada por especialistas, resultando em uma descrição precisa e completa das informações funcionais da proteína.

Atualmente, o UniProtKB contém mais de 227 milhões de registros de sequência na versão 2022\_03 do UniProt (UNIPROT, 2023). Essa vasta quantidade de dados proteômicos destaca o desafio de resolver estruturas experimentais para todas as proteínas conhecidas.

Portanto, considerando a realidade dos bancos de dados PDB (Protein Data Bank) e UniProt, é evidente que muitas proteínas ainda não tiveram suas estruturas tridimensionais determinadas experimentalmente. Nesse contexto, a predição computacional (*in silico*) da estrutura tridimensional de proteínas surge como uma alternativa promissora para preencher essa lacuna em constante crescimento (STUDER et al., 2019).

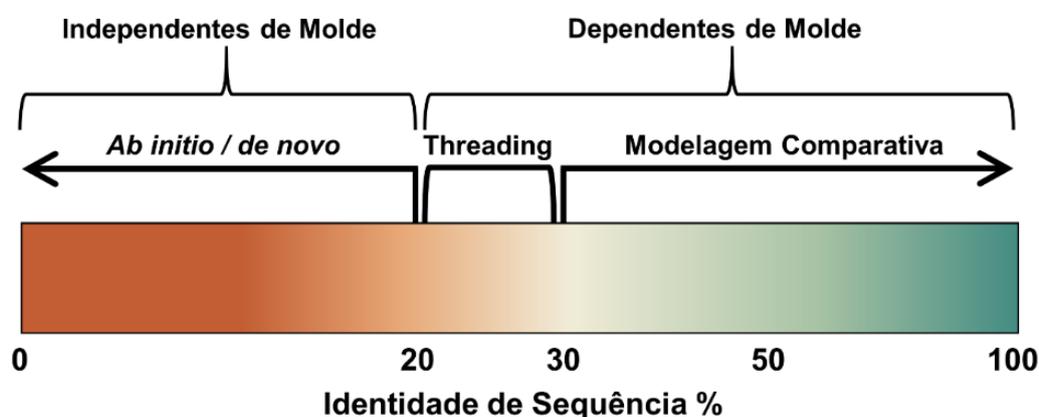
É importante destacar que essa lacuna na determinação experimental de estruturas tridimensionais implica na perda de informações biológicas e biofísicas significativas. Nem todas as novas proteínas identificadas e sequenciadas possuem suas estruturas tridimensionais elucidadas (SCHWEDE, 2013). Como resultado, a compreensão completa das propriedades estruturais e funcionais dessas proteínas fica comprometida, limitando nosso conhecimento sobre sua atividade biológica e seu potencial aplicado.

Através do uso de métodos de predição computacional, é possível obter informações estruturais de uma proteína cuja estrutura tridimensional ainda não foi determinada experimentalmente, baseando-se na sequência de aminoácidos. No passado, esse tipo de predição era considerado desafiador. No entanto, devido ao avanço dos algoritmos computacionais e ao aumento da disponibilidade de dados sobre dobramentos proteicos conhecidos, hoje é possível realizar previsões plausíveis e geralmente precisas (LIU & CHEN, 2016).

Os métodos de predição computacional utilizam algoritmos e modelos baseados em princípios físicos e estatísticos para inferir a estrutura tridimensional de uma proteína a partir de sua sequência de aminoácidos. Essas abordagens se baseiam em correlações estatísticas entre sequências de proteínas conhecidas e suas estruturas correspondentes, bem como em princípios de física e química para modelar as interações entre os átomos na proteína.

Os métodos computacionais de predição de estrutura podem ser divididos em duas categorias principais: métodos baseados em estruturas tridimensionais conhecidas e métodos independentes de estruturas conhecidas (Figura 2).

**Figura 2 – Escala entre métodos de predição de estrutura tridimensional de proteínas e identidade de sequência com as estruturas existentes.**



Fonte: (SILVA; BASTOS; SANTOS, 2021).

Cada método requer um grau específico de similaridade, calculado pela taxa de identidade entre as sequências alvo e as sequências de estruturas conhecidas que serão usadas como moldes.

Os métodos baseados em estruturas conhecidas exploram a semelhança entre a proteína alvo e proteínas cujas estruturas tridimensionais foram resolvidas experimentalmente. Nesses casos, uma estrutura tridimensional conhecida, chamada de molde ou template, é modificada para se ajustar à sequência da proteína alvo por meio de técnicas de modelagem comparativa e *threading* (SILVA; BASTOS; SANTOS, 2021). Essas abordagens se baseiam na suposição de que proteínas com sequências semelhantes têm estruturas semelhantes.

Já os métodos independentes de um molde, como as abordagens *ab initio* e *de novo*, não dependem diretamente de estruturas tridimensionais conhecidas. Nesses casos, as informações estruturais são obtidas por meio da combinação de vários fragmentos estruturais ou pela previsão da estrutura secundária da proteína

alvo com base em princípios físicos e estatísticos (KC, 2017). Essas abordagens buscam construir uma estrutura tridimensional a partir das características intrínsecas da sequência de aminoácidos, sem utilizar informações externas.

Ambas as categorias de métodos têm suas vantagens e limitações. Os métodos baseados em estruturas conhecidas são mais eficientes quando existe uma estrutura semelhante disponível, enquanto os métodos independentes de um molde podem ser aplicados a proteínas com sequências únicas ou que não apresentam semelhanças com proteínas estruturalmente conhecidas. No entanto, a predição *ab initio* é mais desafiadora devido à complexidade do problema e ao espaço de conformação vasto que precisa ser explorado.

Na próxima seção, será explorado em detalhes as estratégias e ferramentas utilizadas na modelagem dependente de molde, aprofundando-se nas suas aplicações e peculiaridades.

#### **2.4.1 Métodos dependentes de molde**

Os métodos baseados em molde são fundamentados na suposição de que a estrutura tridimensional de uma proteína é conservada ao longo da evolução. Essa conservação estrutural implica que mudanças na sequência de aminoácidos geralmente resultam em alterações mínimas na estrutura tridimensional (PATEL; SINGH; PATEL, 2019). Nessa categoria de métodos, a similaridade estrutural entre proteínas com sequências semelhantes é explorada, levando em consideração que essas proteínas tendem a adotar estruturas quase idênticas. Surpreendentemente, estruturas tridimensionais semelhantes podem ser encontradas mesmo em proteínas com baixa identidade sequencial (até 20%). Portanto, é possível gerar um modelo tridimensional para uma proteína de estrutura desconhecida, desde que exista uma estrutura resolvida experimentalmente que sirva como referência (SILVA; BASTOS; SANTOS, 2021).

Dentro dos métodos dependentes de molde, destacam-se a modelagem comparativa e o *Threading*, os quais serão abordados com maior profundidade nos próximos tópicos.

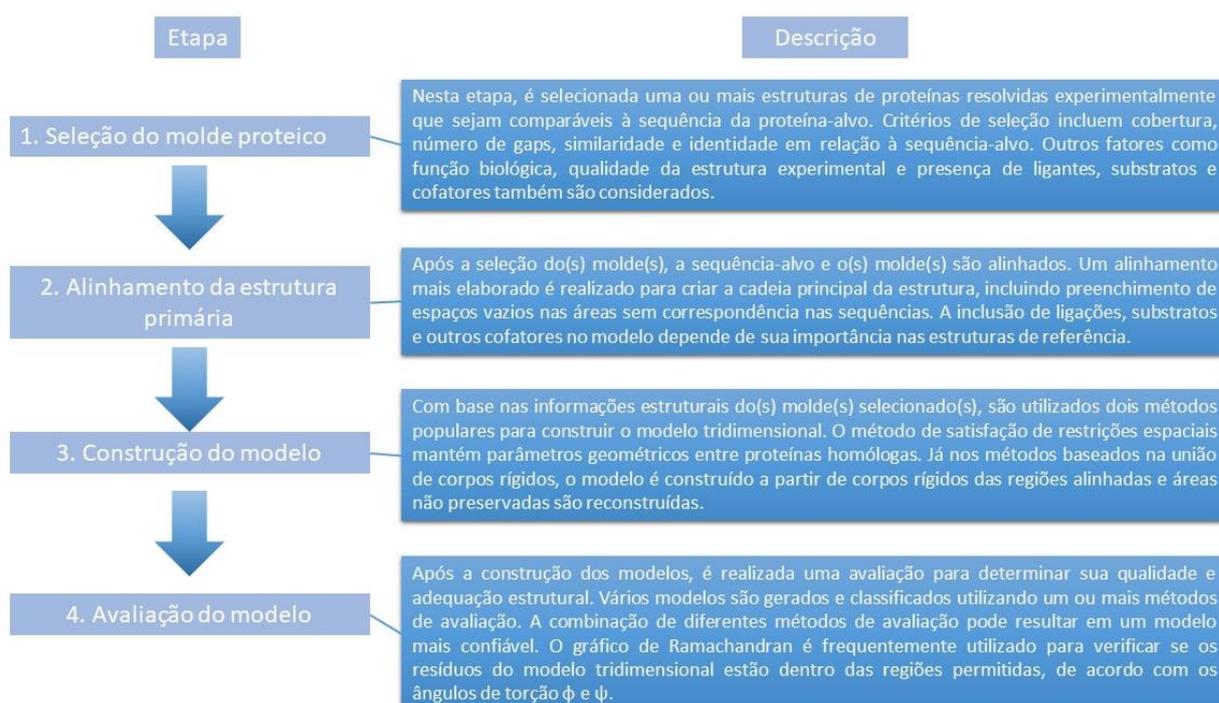
##### **2.4.1.1 Modelagem comparativa**

Também conhecida como modelagem por homologia, a modelagem comparativa é o método mais popular entre as abordagens baseadas em molde.

Essa abordagem requer a disponibilidade de uma estrutura tridimensional resolvida como molde ou referência. Para que seja aplicada, a sequência da proteína-alvo deve ter uma identidade mínima de 25% a 30% em relação à sequência do molde. Com base na estrutura do molde, é possível criar um modelo tridimensional da proteína-alvo, representando sua conformação espacial (ŠALI & BLUNDELL, 1993; GINALSKI, 2006).

O processo de modelagem comparativa envolve quatro etapas principais demonstrados na Figura 3 (BAKER & SALI, 2001) e descritos abaixo:

**Figura 3 – Etapas da modelagem comparativa.**



Fonte: Elaborada pela autora (2023).

Nessa etapa de validação, softwares como MODELLER (ŠALI & BLUNDELL, 1993) e SWISS-MODEL (SCHWEDE et al., 2003) são amplamente utilizados para auxiliar no processo de modelagem comparativa. Esses programas fornecem recursos para avaliar e refinar os modelos gerados, contribuindo para a obtenção de estruturas tridimensionais mais confiáveis.

#### 2.4.1.2 Threading

O *threading* é um método de modelagem que permite criar modelos de estruturas que apresentam enovelamento semelhante a proteínas de estruturas conhecidas, mesmo quando há pouca semelhança entre elas. Nesse método, a

sequência de aminoácidos é fragmentada e utilizada na busca por homólogos estruturais, explorando diversos alinhamentos em vez de alinhar a sequência inteira (ZHANG, 2009). Portanto, o *threading* é aplicado quando modelos com baixa identidade podem abranger diferentes partes da sequência-alvo (VERLI, 2014).

A modelagem por *threading* depende do reconhecimento das características da sequência-alvo, sendo necessário realizar alinhamentos locais para identificar moldes ou estruturas disponíveis na base de dados de proteínas (PDB) que cobrem regiões específicas. Em seguida, para cada molde escolhido, uma abordagem de modelagem comparativa é aplicada, criando estruturas secundárias para cada região. É importante destacar que a qualidade das estruturas selecionadas e a determinação de quais moldes cobrem cada parte da sequência têm um impacto direto na qualidade dos modelos finais gerados (VERLI, 2014).

O *threading* é uma técnica amplamente utilizada pelo programa I-TASSER para a construção de modelos tridimensionais de proteínas. Esse programa será abordado de forma minuciosa no próximo tópico, no qual serão exploradas suas características e aplicações na modelagem de estruturas proteicas.

#### 2.4.1.2.1 I-TASSER

O I-TASSER (Iterative Threading ASSEmbly Refinement) é um dos programas de *threading* mais amplamente utilizados na predição de estruturas tridimensionais de proteínas. Ele recebeu vários prêmios na competição CASP (Critical Assessment of protein Structure Prediction), destacando sua eficácia e precisão. O I-TASSER utiliza uma abordagem combinada de modelagem por homologia e modelagem *ab initio*, conhecida como *threading*, que é empregada quando as sequências de proteínas não têm homólogos conhecidos no banco de dados ou exibem baixa identidade com outras proteínas (ZHANG, 2009; YANG & ZHANG, 2015).

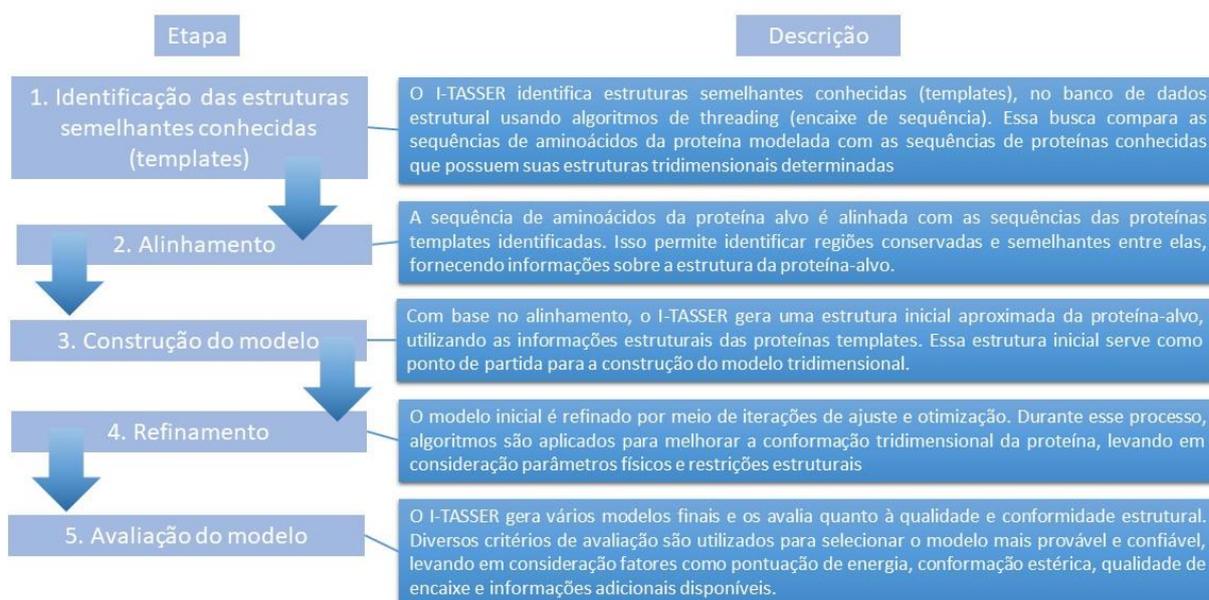
O método de *threading* utilizado pelo I-TASSER envolve a busca de fragmentos estruturais semelhantes em um banco de dados de estruturas conhecidas. Com base nessas informações, são gerados modelos iniciais que são subsequentemente refinados por meio de iterações para melhorar sua precisão e conformação tridimensional. Essa abordagem permite a predição de estruturas de proteínas com baixa similaridade sequencial ou para as quais não há uma estrutura tridimensional conhecida disponível (ZHANG, 2009; YANG & ZHANG, 2015).

O I-TASSER tem sido amplamente aplicado na modelagem de proteínas e na bioinformática devido à sua capacidade de fornecer informações valiosas sobre a estrutura e função de proteínas. Sua combinação de métodos de *threading* e refinamento iterativo torna uma ferramenta poderosa para a predição de estruturas proteicas, contribuindo para avanços significativos na compreensão dos mecanismos moleculares.

#### 2.4.1.2.2 Modelagem por Threading no I-TASSER

Para gerar modelos estruturais de proteínas por Modelagem por "*Threading*", o I-TASSER segue um processo iterativo de várias etapas conforme indicado na imagem abaixo (YANG & ZHANG, 2015; ZHENG et al., 2021; ZHOU et al., 2022):

**Figura 4 – Etapas da modelagem por *Threading*.**



**Fonte:** Elaborada pela autora (2023).

Os cinco modelos gerados pelo I-TASSER com a melhor classificação são apresentados, acompanhados de informações como precisão global (C-score), TM-score e RMSD, que auxiliam na avaliação da qualidade dos modelos. É importante ressaltar que o I-TASSER fornece a previsão de TM-score e RMSD apenas para o primeiro modelo, devido à fraca correlação entre C-score e TM-score em modelos de classificação inferior (YANG & ZHANG, 2015; ZHENG et al., 2021; ZHOU et al., 2022). Essas métricas permitem avaliar a similaridade estrutural entre o modelo gerado e a estrutura nativa da proteína, fornecendo informações valiosas sobre a qualidade e a precisão do modelo.

Para fins de referência, todos os modelos gerados são acompanhados de um valor de C-score, que serve como um indicador de precisão global estimada. O C-score varia de -5 a 2, sendo que valores acima de -1,5 indicam modelos com boa topologia global prevista. O índice TM (ZHANG & SKOLNICK, 2004) é uma escala sugerida para avaliar a similaridade estrutural entre duas estruturas, no caso, a estrutura do molde e a do modelo. Essa escala não depende do comprimento da proteína e não é sensível a diferentes orientações locais e estruturas (YANG & ZHANG, 2015; ZHENG et al., 2021; ZHOU et al., 2022).

Uma pontuação TM acima de 0,50 indica que o modelo possui a topologia correta, enquanto uma pontuação TM abaixo de 0,17 indica uma similaridade aleatória entre as estruturas. O termo RMSD refere-se à sobreposição entre o modelo gerado e o molde. Um valor alto de RMSD indica que certas áreas da proteína apresentam orientação e estrutura divergentes. Além disso, o I-TASSER também pode fornecer informações sobre ligantes, sítios ativos e funções das estruturas modeladas, quando disponíveis (YANG & ZHANG, 2015; ZHENG et al., 2021; ZHOU et al., 2022). Essas métricas e informações adicionais auxiliam na avaliação e interpretação dos modelos gerados pelo I-TASSER.

As vantagens da Modelagem por *Threading* (YANG & ZHANG, 2015; ZHENG et al., 2021; ZHOU et al., 2022), são:

**Tabela 2 –Vantagens e desvantagens da modelagem *Threading*.**

<b>Vantagens da Modelagem por <i>Threading</i></b>	<b>Desvantagens da Modelagem por <i>Threading</i></b>
Fornecer resultados de alta qualidade quando templates estruturais adequados estão disponíveis	Dependência de templates: A qualidade da previsão pode ser comprometida se não houver templates adequados disponíveis
Maior velocidade em comparação com a modelagem Ab initio	Limitação a proteínas com sequências similares às proteínas templates disponíveis
Utilidade na modelagem de proteínas com	Menor precisão na modelagem de proteínas

sequências semelhantes às proteínas já conhecidas	com estruturas únicas ou complexas
---	------------------------------------

Fonte: Elaborada pela autora (2023).

## 2.5 Redes de Interação

O estudo das interações de proteínas ajuda na compreensão dos processos biológicos, uma vez que as proteínas geralmente atuam em conjunto, formando redes complexas de interações. A formação de complexos multiproteicos, a regulação de vias metabólicas e de sinalização, a construção de estruturas celulares, a modulação da expressão gênica e muitos outros processos biológicos dependem dessas interações entre proteínas. Além de fornecer *insights* importantes para o desenvolvimento de novas terapias e tratamentos médicos, o estudo dessas interações nos ajuda a compreender como as proteínas colaboram para desempenhar suas funções específicas, contribuindo para a compreensão global dos sistemas biológicos (WALPORT et al., 2021).

As interações entre proteínas podem ocorrer tanto entre proteínas do mesmo tipo quanto entre proteínas diferentes. Essas interações podem ser classificadas em interações físicas diretas, interações funcionais indiretas e interações transientes.

As interações físicas diretas ocorrem quando as proteínas se ligam entre si por meio de contatos físicos diretos. Esses contatos podem envolver diferentes partes das proteínas, como as cadeias laterais dos aminoácidos ou regiões estruturais específicas. Essas interações físicas diretas podem ser estabelecidas por ligações covalentes ou não covalentes, como ligações de hidrogênio, interações eletrostáticas e interações hidrofóbicas. Geralmente, essas interações são fortes e duradouras, influenciando na estabilidade estrutural das proteínas e em processos funcionais específicos (BRUCKNER, 2012).

Já as interações funcionais indiretas ocorrem quando as proteínas se ligam por meio de componentes intermediários, como outras moléculas ou complexos proteicos. Esses intermediários atuam como pontes de ligação ou reguladores das interações entre as proteínas. Eles podem modular a atividade das proteínas ou fornecer um contexto específico para a ocorrência das interações proteína-proteína. Essas interações funcionais indiretas expressam uma importante atuação na regulação de vias metabólicas, sinalização celular e outros processos biológicos complexos (BRUCKNER, 2012).

Por fim, as interações transientes são caracterizadas por serem temporárias e reversíveis. Nesse caso, as proteínas podem interagir com diferentes parceiros em diferentes momentos e contextos celulares, devido às suas interações mais fracas e dinâmicas. Essas interações transientes são essenciais para processos regulatórios, sinalização celular e resposta a estímulos ambientais. Permitem a formação de complexos multiproteicos temporários e a modulação rápida das atividades das proteínas envolvidas (BRUCKNER, 2012).

As interações entre proteínas exercem contribuição vital na regulação e execução de diversas funções celulares. Essas interações são responsáveis por processos como transmissão de sinais, controle gênico, transporte de moléculas e resposta a estímulos ambientais. As proteínas podem formar complexos multiproteicos que efetuam funções específicas dentro da célula por meio dessas interações (WALPORT et al., 2021).

É importante enfatizar que as interações entre proteínas são altamente seletivas e específicas. Isso significa que uma proteína só pode interagir com proteínas-alvo que possuam regiões de interação complementares. Essa seletividade é essencial para garantir a eficiência e a precisão dos processos biológicos (WALPORT et al., 2021).

A formação de complexos multiproteicos, a regulação de vias metabólicas e de sinalização, a construção de estruturas celulares, a modulação da expressão gênica e uma variedade de outros processos biológicos dependem fundamentalmente das interações entre proteínas. O estudo dessas interações não apenas proporciona insights importantes para o desenvolvimento de novas terapias e tratamentos médicos, mas também contribui para uma compreensão mais profunda de como as proteínas cooperam para desempenhar suas funções específicas (WALPORT et al., 2021).

A investigação das interações de proteínas envolve a aplicação de uma combinação de técnicas experimentais e computacionais. Essas abordagens permitem mapear as redes de interações proteicas e compreender a complexidade dessas interações. Através dessas pesquisas, é possível desvendar os mecanismos moleculares que controlam os processos biológicos e abrir caminho para o desenvolvimento de novos tratamentos (DE ANDRADE, 2017).

A Bioinformática cumpre uma responsabilidade crucial na análise das redes de interação de proteínas, proporcionando uma compreensão abrangente e

sistemática das relações entre as proteínas. Como uma área interdisciplinar que combina biologia, computação e informática, a Bioinformática utiliza técnicas computacionais sofisticadas para analisar e interpretar dados biológicos complexos (VERLI, 2014).

A construção e análise de redes de interação de proteínas envolvem o uso de técnicas bioinformáticas que integram dados de interação proteína-proteína obtidos em experimentos. Além disso, são empregados algoritmos computacionais para fazer previsões de interações e ferramentas especializadas para visualizar os resultados. Essas abordagens bioinformáticas permitem explorar as interações entre proteínas em escala global, identificar proteínas-chave, módulos funcionais e vias de sinalização, bem como compreender a organização e o funcionamento dos sistemas biológicos (DE ANDRADE, 2017).

Ao utilizar a Bioinformática, os pesquisadores podem analisar grandes conjuntos de dados biológicos e extrair informações valiosas sobre as interações de proteínas. Essas informações contribuem não apenas para a pesquisa biomédica, mas também têm um impacto significativo na agricultura, especialmente na relação entre bactérias e plantas.

A Bioinformática desempenha um papel essencial no estudo das interações proteicas entre bactérias e plantas, permitindo uma compreensão mais profunda dos mecanismos envolvidos na simbiose entre esses organismos. Por meio da análise bioinformática, é possível identificar as proteínas envolvidas na interação planta-bactéria, bem como suas funções e mecanismos de ação. Isso proporciona insights valiosos para o desenvolvimento de estratégias de melhoria de plantas e controle de doenças, beneficiando a agricultura de forma sustentável (VERLI, 2014).

Além disso, a Bioinformática auxilia na identificação de proteínas-chave envolvidas em processos agronomicamente relevantes, como a resposta das plantas ao estresse ambiental, a absorção de nutrientes e a resistência a patógenos. Essas informações são fundamentais para o desenvolvimento de culturas mais resistentes, produtivas e adaptadas às condições desafiadoras do ambiente agrícola (VERLI, 2014).

Ao integrar os dados genômicos, transcriptômicos e proteômicos de bactérias e plantas, a Bioinformática oferece uma visão abrangente das interações proteicas envolvidas nos sistemas biológicos. Isso possibilita a identificação de alvos terapêuticos, o desenvolvimento de novos agentes biológicos para o controle de

pragas e doenças, bem como a otimização de estratégias de manejo e produção agrícola.

### **2.5.1 Componentes de uma rede de interação**

Uma rede de proteínas é composta por vários componentes essenciais, cada um dos quais desempenha funções específicas para garantir que a rede seja estruturada e funcionalmente organizada. A seguir a Tabela 3 – indica os componentes principais de uma rede de interação de proteínas, bem como suas definições (SZKLARCZYK et al., 2016; 2019):

**Tabela 3 – Componentes e definição dos elementos envolvidos em uma rede de interação.**

<b>Componente</b>	<b>Definição</b>
Proteínas	Componentes fundamentais das redes de interação proteína-proteína, desempenhando diversas atividades e funções nas células.
Nós	Representam as proteínas individuais em uma rede de interação de proteínas, onde cada proteína é representada por um nó.
Arestas	Conexões que indicam as interações entre as proteínas na rede, mostrando as relações físicas ou funcionais entre elas.
Interações Físicas	Conexões diretas entre domínios estruturais, sítios de ligação ou outras áreas específicas das proteínas, observadas experimentalmente.
Interações Funcionais	Associações indiretas entre proteínas, baseadas em suas funções ou participação em vias metabólicas ou processos biológicos comuns.
Grau	O número de conexões ou arestas que um nó possui em uma rede de interação de proteínas, refletindo a importância relativa do nó na rede.
Módulos	Grupos de proteínas altamente interconectadas que podem ser vias metabólicas, complexos proteicos ou grupos funcionais com uma tarefa específica.
Centralidade	Avaliação da importância relativa de um nó na rede, identificando nós-chave ou proteínas centrais usando métricas como grau, intermediação ou proximidade.

**Fonte:** Elaborada pela autora (2023).

A coleta de dados sobre interações proteína-proteína é um passo crucial na construção da rede de interação proteína-proteína (IPP). Esses dados são obtidos a

partir de experimentos, estudos científicos publicados na literatura e previsões computacionais, e são compilados em bancos de dados e repositórios especializados para o armazenamento dessas informações. Esses recursos fornecem uma ampla gama de dados sobre as interações proteicas em diferentes condições experimentais e organismos.

A tabela abaixo estão alguns dos principais repositórios e bancos de dados amplamente utilizados na pesquisa de interações proteína-proteína (SOARES, 2014; SIMÕES, 2015):

**Tabela 4 – Principais bancos de dados de redes de interação.**

<b>Banco de Dados</b>	<b>Descrição</b>	<b>Site</b>
<b>STRING</b> (Search Tool for the Retrieval of Interacting Genes/Proteins)	Banco de dados que fornece informações sobre as interações entre proteínas, incluindo informações experimentais e preditivas.	<a href="https://string-db.org/">https://string-db.org/</a>
<b>BioGRID</b> (Biological General Repository for Interaction Datasets)	Repositório que contém uma vasta coleção de dados de interações proteína-proteína, abrangendo várias espécies e fontes de dados.	<a href="https://thebiogrid.org/">https://thebiogrid.org/</a>
<b>IntAct</b>	Banco de dados que armazena dados de interações proteína-proteína obtidos a partir de experimentos e literatura científica, oferecendo recursos para análise e visualização dessas interações.	<a href="https://www.ebi.ac.uk/intact/">https://www.ebi.ac.uk/intact/</a>
<b>MINT</b> (Molecular INTERaction Database)	Banco de dados que fornece informações sobre interações moleculares, incluindo interações proteína-proteína, interações proteína-ligante e interações proteína-DNA.	<a href="https://mint.bio.uniroma2.it/">https://mint.bio.uniroma2.it/</a>

**Fonte:** Elaborada pela autora (2023).

Dentro desses bancos de dados, se destaca o STRING, o qual será descrito com mais detalhes na seção seguinte.

### **2.5.2 STRING**

O STRING (<https://string-db.org/>) é um banco de dados abrangente de interações de proteínas, que reúne dados sobre redes de interação de proteínas, associações funcionais e interações proteína-proteína. Ele fornece uma pontuação

de confiança para cada interação, permitindo a identificação das interações mais confiáveis. Além disso, o STRING oferece anotações genômicas, vias metabólicas e funções biológicas (SZKLARCZYK et al., 2016).

O banco de dados contém informações sobre interações físicas diretas e/ou interações funcionais indiretas, que são baseadas em conhecimento prévio, contexto genômico e experimentos de expressão de alta capacidade. Por meio do STRING, é possível identificar se conjuntos específicos de genes ou proteínas estão associados a funções ou processos biológicos de forma mais significativa do que o esperado ao acaso (SZKLARCZYK et al., 2016).

No STRING, é possível acessar várias informações, como Gene Ontology (GO), publicações científicas de referência (PubMed), clusters de redes locais, vias KEGG, domínios proteicos (Pfam) e recursos de domínios e características proteicas (InterPro). A tabela abaixo descreve tais informações bem como suas respectivas definições:

**Tabela 5 – Informações acessadas no STRING.**

<b>Informações acessadas no STRING</b>	<b>Definição</b>
<b>Ontologia gênica (GO)</b>	Ferramenta amplamente utilizada na biologia para categorizar as funções biológicas de genes e proteínas. Fornece um vocabulário organizado e padronizado para descrever funções moleculares, processos biológicos e localização celular.
<b>Publicações de referência</b>	São artigos científicos escritos por pesquisadores e especialistas no campo, que servem de base para os dados apresentados na plataforma STRING. Fornecem detalhes sobre estudos e experimentos, sustentando as informações sobre as interações de proteínas e genes no STRING. Permite verificar a qualidade e confiabilidade dos dados e obter mais informações sobre as pesquisas envolvendo anotações de genes e interações de proteínas.
<b>Cluster de rede local</b>	Refere-se a agrupamentos de proteínas que interagem entre si em uma região específica de uma rede de interações de proteínas. Os clusters indicam a probabilidade de que as proteínas dentro do agrupamento desempenhem funções biológicas semelhantes ou estejam envolvidas em vias metabólicas ou processos celulares específicos. Permite compreender a relação entre as proteínas e a contribuição dos clusters para as funções celulares e os processos biológicos.
	Conjunto de informações sobre vias metabólicas e de sinalização presentes no banco de dados KEGG (Kyoto Encyclopedia of Genes and Genomes). Fornece detalhes sobre

<b>Caminhos KEGG</b>	interações e funções dos genes e proteínas nessas vias. Representações visuais ajudam a entender como diferentes genes e proteínas interagem e se influenciam em uma determinada via. Fornece informações adicionais, como descrições das vias, anotações dos genes e proteínas envolvidas, bem como referências bibliográficas relacionadas.
<b>Domínios de proteína (Pfam)</b>	Refere-se aos domínios de proteínas obtidos do banco de dados Pfam. Esses domínios são unidades estruturais e funcionais que contribuem para as funções biológicas das proteínas. O banco de dados Pfam cataloga e descreve sequências de aminoácidos que ocorrem repetidamente em proteínas, permitindo relacionar sua presença a funções biológicas específicas.
<b>Domínios e características de proteínas (InterPro)</b>	Refere-se aos domínios de proteínas e suas características obtidas do banco de dados InterPro. O InterPro é uma coleção abrangente de anotações e classificações de domínios de proteínas e outras características estruturais e funcionais. Reúne informações de diversos recursos para fornecer uma visão abrangente dos domínios e características das proteínas. Permite determinar os domínios presentes nas proteínas, bem como outras características relevantes, contribuindo para a compreensão dos mecanismos biológicos e da relação entre a estrutura e a função das proteínas.

**Fonte:** Elaborada pela autora (2023).

Em relação à interação proteína-proteína, o STRING oferece uma variedade de informações que mostram as interações em uma rede de proteínas. Essas informações são obtidas por meio de diferentes tipos de evidências e abordagens, incluindo: Vizinhança no Genoma; Fusões genéticas; Coocorrência entre genomas; Co-expressão; Dados Experimentais/Bioquímicos; Associação em bancos de dados selecionados; e Co-Mencionado no Pubmed Abstracts.

A tabela a seguir organiza e descreve as informações sobre os recursos disponíveis no STRING relacionados às interações proteína-proteína:

**Tabela 6 – Informações do STRING relacionados às interações proteína-proteína.**

<b>Recurso</b>	<b>Descrição</b>
<b>Vizinhança no Genoma</b>	Fornecer informações sobre proteínas que interagem diretamente com a proteína de interesse, permitindo a análise de suas relações e compreensão de suas funções biológicas. Também fornece informações sobre interações conhecidas, funções biológicas e processos celulares envolvidos nessas proteínas.
<b>Fusões Genéticas</b>	Refere-se aos casos em que partes de dois genes diferentes se fundem, criando uma sequência híbrida. Essas fusões podem ocorrer como resultado de rearranjos

	cromossômicos e podem ter implicações funcionais para as proteínas resultantes.
<b>Coocorrência entre Genomas</b>	Permite analisar a frequência com que os genes ocorrem juntos em diferentes genomas, sugerindo uma possível associação funcional ou dependência evolutiva entre eles.
<b>Co-expressão</b>	Indica que dois ou mais genes têm padrões de expressão semelhantes em condições experimentais, tecidos ou estágios de desenvolvimento, sugerindo uma possível associação funcional e interação em processos biológicos relacionados.
<b>Dados Experimentais/Bioquímicos</b>	Obtidos por meio de experimentos ou abordagens bioquímicas para investigar interações proteína-proteína e ligações funcionais. Os dados experimentais e bioquímicos são combinados com outras fontes para fornecer uma visão completa das interações e relações funcionais.
<b>Associação em bancos de dados</b>	Incorpora informações sobre interações proteína-proteína e anotações funcionais de bases de dados específicas, enriquecendo a compreensão das proteínas em contextos biológicos.
<b>Co-Mencionado no Pubmed Abstracts</b>	Agrupa proteínas com base em sua menção em resumos de literatura científica indexados no PubMed, fornecendo evidências de possíveis ligações funcionais entre elas.

**Fonte:** Elaborada pela autora (2023).

Esses recursos do STRING auxiliam na investigação das interações proteína-proteína e fornecem informações valiosas sobre a função e o contexto biológico das proteínas (JENSEN et al., 2009).

Dentro das análises efetuadas pelo STRING, os clusters fornecem informações relevantes sobre as funções biológicas e moleculares das proteínas, como mencionado anteriormente. O tópico a seguir descreverá com mais detalhes os clusters no contexto do STRING.

#### *2.5.2.1 Clusters*

Os clusters são utilizados para investigar as conexões entre diferentes proteínas em contextos biológicos mais amplos e a arquitetura funcional das interações proteicas. Esses grupos de proteínas podem fornecer informações importantes sobre o funcionamento da rede de interação de proteínas (SZKLARCZYK et al., 2015).

A partir dessas informações, podemos inferir que proteínas pertencentes ao mesmo cluster geralmente atuam em processos metabólicos semelhantes, diferentes atividades biológicas ou ambos. A análise de clusters pode ajudar a

descobrir grupos de proteínas com funções coordenadas, e os clusters no STRING representam grupos de proteínas que interagem frequentemente e intensamente entre si. Além disso, a análise de clusters pode revelar informações sobre as vias e processos biológicos importantes que ocorrem na rede de interação. Esses grupos correspondem a regiões da rede de interação proteica em que ocorrem muitas interações proteína-proteína (SZKLARCZYK et al., 2015).

Embora o algoritmo K-means seja frequentemente empregado na criação de clusters no contexto do STRING, outros algoritmos, como o DBSCAN (Density-Based Spatial Clustering of Applications with Noise), podem ser mais apropriados em determinadas situações. A escolha do algoritmo depende das características do conjunto de dados e das necessidades específicas do problema em questão (SZKLARCZYK et al., 2015).

O K-means e o DBSCAN são algoritmos utilizados na área de clusterização, que é uma técnica de análise de dados. O objetivo da clusterização é agrupar objetos ou dados similares em conjuntos chamados clusters. Isso facilita a identificação de padrões, estruturas e relações dentro dos dados.

O algoritmo K-means é baseado na ideia de encontrar K centróides (pontos centrais) que representam os clusters. Ele começa selecionando aleatoriamente K pontos como centróides iniciais. Em seguida, ele atribui cada objeto ao cluster cujo centróide está mais próximo. Após essa atribuição, os centróides são recalculados com base nos objetos pertencentes a cada cluster. O processo é repetido até que os centróides não se movam significativamente ou até que um critério pré-definido seja atingido.

Já o DBSCAN é um algoritmo baseado em densidade, o que significa que ele leva em consideração a proximidade entre os dados. Ele identifica áreas densas de pontos e atribui esses pontos a clusters, enquanto os pontos menos densos são considerados como ruído ou não atribuídos a nenhum cluster. Dessa forma, o DBSCAN é capaz de identificar clusters de diferentes formas e tamanhos, adaptando-se melhor a conjuntos de dados com densidades irregulares.

Em resumo, o K-means divide os dados em K clusters definidos por centróides, enquanto o DBSCAN agrupa os dados com base em sua densidade e proximidade. A escolha entre esses algoritmos depende das características dos dados e do objetivo da análise.

## 3 MATERIAIS E MÉTODOS

### 3.1 Identificação da sequência e sua análise

No contexto deste estudo, a sequência da proteína GumD foi obtida no NCBI (The National Center for Biotechnology Information - <https://www.ncbi.nlm.nih.gov/>) com o número de acesso no GenBank: CAP56490.1. Para obter informações mais abrangentes sobre a proteína, realizou-se uma análise de GumD no UniProt (<https://www.uniprot.org/>) utilizando o número de identificação A9HNR6.

O UniProt, também conhecido como UniProtKB (Universal Protein Knowledgebase), é uma base de dados online que fornece informações detalhadas e atualizadas sobre proteínas. Através do UniProt, é possível acessar detalhes sobre a função, estrutura, interações e anotações das proteínas, bem como outras informações relevantes para a pesquisa científica.

Ao utilizar o UniProt para analisar a proteína GumD (A9HNR6), o objetivo era obter uma descrição mais completa e detalhada de suas características, além de informações relacionadas à sua função, domínios estruturais e possíveis interações com outras proteínas.

Para complementar essa análise, foram verificados domínios conservados e padrões de perfis na sequência de GumD utilizando o banco de dados Prosite (<https://prosite.expasy.org/>) e a ferramenta ScanProsite (<http://prosite.expasy.org/scanprosite/>). Além disso, o PsiPred (<http://bioinf.cs.ucl.ac.uk/psipred/>) foi utilizado para uma análise mais aprofundada da estrutura secundária da sequência de GumD, incluindo a identificação de possíveis elementos de estrutura secundária e a previsão de regiões transmembranares ou a presença de um peptídeo sinal.

Essas análises adicionais contribuem para uma melhor compreensão das características estruturais e funcionais da proteína GumD, fornecendo informações relevantes para o estudo da mesma.

### 3.2 Modelagem da proteína “GumD”

No processo de modelagem de proteínas, utilizou-se o método I-TASSER (<https://zhanggroup.org/I-TASSER/>) (YANG & ZHANG, 2015). Esse método baseia-se em *threading* (associação de segmentos de sequência) e modelagem por

homologia, empregando uma abordagem de montagem hierárquica para gerar modelos tridimensionais de proteínas a partir de suas sequências de aminoácidos (CHENGXIN ZHANG; FREDDOLINO; ZHANG YANG, 2017). Ao final do processo, vários modelos são gerados, e para selecionar o melhor entre eles, são considerados vários parâmetros, como o C-score, Estimated TM-score e Estimated RMSD.

O C-score é uma medida de confiança utilizada para estimar a qualidade dos modelos preditos pelo I-TASSER. Quanto mais próximo de dois, melhor é a qualidade do modelo. Segundo Yang Zhang (2008), o C-score é calculado com base na importância dos alinhamentos de segmentos do modelo e nos parâmetros de convergência da estrutura obtida por meio de simulações de montagem. Por sua vez, o TM-score e o RMSD são padrões amplamente utilizados para medir a similaridade estrutural entre duas estruturas e são empregados para avaliar a precisão da modelagem quando a estrutura nativa é conhecida (YANG & ZHANG, 2015).

Ao final, foi escolhido o modelo com o melhor C-score para posterior análises.

### **3.2.1 Refinamento e avaliação do modelo**

O modelo gerado pelo I-TASSER precisa passar por refinamento utilizando métodos de simulação para aumentar a qualidade, ou seja, é necessário otimizar a estrutura a fim de obter um modelo 3D mais próximo da estrutura nativa da proteína. Para isso, os modelos foram refinados e avaliados estruturalmente. O refinamento dos modelos foi realizado por meio do servidor ModRefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>), que utiliza um algoritmo de refinamento de estrutura de proteína em nível atômico e alta resolução. Durante as simulações de refinamento, os átomos da cadeia lateral e do esqueleto são completamente flexíveis, e a análise conformacional é guiada por conhecimento físico estrutural. O objetivo do ModRefiner é ajustar os modelos iniciais para torná-los mais próximos do estado nativo em termos de ligações de hidrogênio, topologia do esqueleto e posicionamento das cadeias laterais, melhorando, assim, a precisão do modelo tridimensional (XU & ZHANG, 2011).

A avaliação dos modelos gerados pelo I-TASSER e refinados pelo ModRefiner foi realizada utilizando o servidor MolProbity (<http://molprobity.biochem.duke.edu/>). O MolProbity analisa fatores como geometria

molecular, estereoquímica e interações não favoráveis, fornecendo várias métricas e análises para avaliar a precisão e confiabilidade do modelo. Essa ferramenta oferece cálculos e análises para verificar a qualidade do modelo, incluindo informações sobre a qualidade global, como desvio da geometria ideal, ângulos incorretos, sobreposição estérica e interações não favoráveis. Com base nos resultados da avaliação do MolProbity, foi possível identificar áreas que precisam ser aprimoradas ou corrigidas no modelo de proteína. Isso pode orientar ajustes adicionais ou refinamentos para melhorar a qualidade do modelo (CHEN et al., 2010).

Durante a avaliação pelo MolProbity, um dos gráficos gerados é o gráfico de Ramachandran. Esse gráfico exibe a distribuição dos ângulos diédricos phi ( $\Phi$ ) e psi ( $\Psi$ ) dos resíduos de aminoácidos presentes na estrutura da proteína. Esses ângulos são importantes para a conformação das cadeias polipeptídicas. No gráfico de Ramachandran, regiões específicas indicam as conformações permitidas para os resíduos de aminoácidos, considerando as restrições estéricas e energéticas. Essas regiões são classificadas em regiões favoráveis, permitidas e desfavoráveis. A análise do gráfico de Ramachandran ajudou a avaliar a qualidade da estrutura de proteína modelada, identificando possíveis problemas conformacionais ou regiões com desvios significativos.

### **3.2.2 Análise dos modelos**

Após obter o modelo 3D refinado da proteína GumD de *G. diazotrophicus*, os modelos foram comparados com cristais de proteínas depositadas no banco de dados PDB (Protein Data Bank - <http://www.pdb.org/pdb/home/home.do>).

Para a visualização tridimensional dos modelos, foi utilizado o programa UCSF Chimera 1.16 (PETTERSEN et al., 2004). O UCSF Chimera é uma ferramenta amplamente empregada para a visualização e análise de estruturas moleculares. Ele possibilita a manipulação e renderização de modelos 3D, permitindo uma análise detalhada das estruturas proteicas. Por meio do Chimera 1.16, foram exploradas as características estruturais dos modelos obtidos.

### **3.3 Busca e alinhamento de sequências**

Com o objetivo de encontrar regiões de similaridade entre a sequência de GumD e outras sequências biológicas, utilizou-se a ferramenta BLASTP (Basic Local Alignment Search Tool) -

[https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)) do NCBI. Esse programa compara sequências nucleicas ou proteicas com o banco de dados de sequências e calcula a significância estatística. Após o alinhamento, considerou-se principalmente a identidade como critério de escolha para análises posteriores.

A identidade é uma medida que indica a similaridade entre duas sequências de DNA ou proteínas. Ela representa a porcentagem de resíduos idênticos encontrados quando duas sequências são alinhadas. Quanto maior a identidade, maior é a similaridade entre as sequências. Neste estudo, foram selecionados organismos que apresentavam uma variação de identidade entre 50% e 80%, ou seja, sequências que compartilham uma porcentagem significativa de resíduos idênticos com a sequência de GumD.

Dos resultados obtidos, foram selecionadas quatro espécies (*G. tumulisoli*, *Nguyenibacter vanlangensis*, *Kozakia baliensis* e *Acetobacter estunensis*) para serem alinhadas com a proteína GumD de *G. diazotrophicus* PAI 5. O alinhamento múltiplo das sequências de referência extraídas do servidor BLASTP foi realizado utilizando o programa Clustal Omega 1.2.4 (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

### 3.4 Construção de interação

A partir da sequência fasta do gene *gumD*, foi construída uma rede de interação proteína-proteína utilizando a plataforma STRING (<https://string-db.org/>). Essa rede foi analisada para identificar os componentes envolvidos na interação com a proteína GumD, bem como descrever as interações observadas.

A plataforma STRING é uma ferramenta amplamente utilizada para analisar interações entre proteínas e construir redes de interação. Ela integra informações de várias fontes, incluindo interações físicas e funcionais conhecidas, co-expressão gênica, cocorrência em literatura científica, entre outras. Ao analisar a rede de interação proteína-proteína no STRING, foi possível identificar proteínas que interagem com GumD e explorar as relações entre essas proteínas.

As interações encontradas em relação à proteína GumD foram verificadas e descritas, fornecendo informações valiosas sobre as possíveis vias e processos biológicos envolvidos na função da proteína GumD.

### **3.5 Rede de interação e Clustres**

A metodologia empregada neste estudo envolveu a construção de uma rede de interação utilizando a plataforma STRING. Inicialmente, foi criada a estrutura básica da rede, sendo posteriormente adicionados nós extras para enriquecer sua conexões. Esse procedimento foi repetido três vezes, resultando em uma rede mais complexa e abrangente. Em seguida, foi aplicado o algoritmo “Kmeans clustering” disponível na plataforma STRING, com intuito de agrupar os elementos da rede em três clusters distintos. Essa abordagem possibilitou a identificação análise de padrões de interação entre os componentes da rede, bom como a identificação de grupos específicos de nós com características semelhantes.

## 4 RESULTADOS E DISCUSSÕES

### 4.1 Caracterização da proteína GumD

Segundo o banco de dados do UniProt, a proteína GumD (A9HNR6) é uma proteína de membrana composta por 506 aminoácidos (aa), sendo denominada "Exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase" (traduzido para: Poliprenilglicosilfosfotransferase envolvida na biossíntese de exopolissacarídeos). Sua funcionalidade molecular está relacionada com a atividade de transferase e seu papel biológico está associado ao processo de biossíntese de polissacarídeos. Essa proteína apresenta um domínio conservado entre os aminoácidos 312 e 497, conhecido como "Bacterial sugar transferase domain" (tradução para: Domínio de transferase de açúcar bacteriano).

O domínio "Domínio de transferase de açúcar bacteriano" (Pfam: PF02397) representa uma região conservada presente em várias enzimas de transferência de açúcares bacterianas, envolvidas em diversas vias de biossíntese. Alguns exemplos notáveis incluem a galactosil-PP-undecaprenol sintetase (2.7.8.6), responsável pela transferência de galactose-1-fosfato para o precursor lipídico undecaprenol fosfato nas etapas iniciais da biossíntese do O-polissacarídeo. Outros exemplos são a UDP-galactose-lípido transportador transferase, envolvida na biossíntese do amylovoran, e a galactosil transferase CpsD, essencial para a montagem do polissacarídeo capsular tipo III em *Streptococci* do grupo B (GBS) (YOSHIDA et al., 1998).

Esse domínio é fundamental para as vias de biossíntese que envolvem transferência de açúcares nas bactérias, e sua conservação sugere uma função importante na manutenção da estrutura e função dessas vias.

A Figura 5 apresenta uma representação gráfica das prováveis regiões de inserção da proteína GumD na membrana. Essas imagens foram geradas pelo Psipred (Predict Protein Secondary Structure), um servidor que utiliza diversas técnicas para prever os componentes da estrutura secundária presentes em uma sequência de proteína. Na Figura 5, é possível observar as sequências de aminoácidos que se encontram tanto na região extracelular da célula bacteriana quanto aquelas que interagem com a membrana e estão expostas ao citoplasma.

O Psipred utiliza o método MEMSAT-SVM para prever regiões transmembranares em proteínas, conforme indicado na Figura 5. Ele especifica a

quantidade de vezes que a sequência da proteína estudada atravessa a membrana plasmática. No caso da proteína GumD, foram identificadas cinco regiões transmembranares (S1, S2, S3, S4 e S5), como representado na Figura 5 (A).

A identificação dessas regiões transmembranares é de grande importância, pois indica que a proteína GumD está inserida na membrana plasmática e atravessa a bicamada lipídica. Isso sugere seu envolvimento direto nas interações com o ambiente extracelular e a possibilidade de desempenhar funções como transporte de substâncias a qual está associada a função da GumD.

Segundo a ferramenta ScanProsite, foram identificados dois sítios ativos na proteína GumD, localizados nos aminoácidos 474 e 478, associados aos termos "Nucleophile" e "Proton donor", respectivamente. Na Figura 6, podem ser vistos esses resultados:

O termo Nucleophile, ou nucleófilo, refere-se a uma espécie química que possui elétrons livres ou alta densidade eletrônica, capaz de doar esses elétrons para formar novas ligações químicas com outras espécies. No contexto do sítio ativo, um resíduo de aminoácido ou grupo funcional da enzima/proteína pode atuar como um nucleófilo ao interagir com um substrato. Já o termo Proton donor, ou doador de próton, refere-se a uma espécie capaz de doar um próton (íon  $H^+$ ) para outra espécie. No sítio ativo, um resíduo de aminoácido ou grupo funcional pode atuar como um doador de prótons, fornecendo íons  $H^+$  para facilitar uma reação química.

No contexto do sítio ativo da proteína GumD, os resíduos de aminoácidos ou grupos funcionais desenvolvem papéis fundamentais como nucleófilos e doadores de prótons. Esses grupos funcionais estão diretamente envolvidos na transferência de grupos glicosilfosfato entre poliprenóis e substratos, como ocorre em reações de biossíntese de exopolissacarídeos.

A proteína GumD é classificada como uma poliprenilglicosilfosfotransferase, sendo responsável por catalisar a transferência de resíduos de glicose do nucleotídeo glicosil difosfato para um poliprenol-1-fosfato, formando assim o poliprenilglicosilfosfato. Esse processo é essencial para a síntese de exopolissacarídeos, importantes componentes da matriz extracelular de algumas bactérias, que são fundamentais na proteção celular, adesão a superfícies e interações com o ambiente.

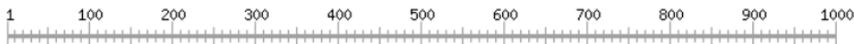


GumD, apresentando as previsões feitas pelo Psipred para os segmentos localizados extracelularmente, no citoplasma e com interação com a membrana.

## Figura 6 – Resultados do ScanProsite.

tr-A9HNR6-A9HNR6\_GLUDA (506 aa)

(A) MPDGDLRMSEWYGQLQDSIGLSTDPITYKTPSARIAEPGTYPPYHSLVLAGMITSVDTLAILAATFI  
 GASVEPAIIHQDWRVAQPLADLMSALGFLMLPKNRRLLHVPTVQDFSAQVRYLTPPLLALGALLHVI  
 VLYMLHHPVPSFELALTWLAFASTGVLAVVRGTETVLLSRPAIANQLARNVAVVGSDDAAIRLAAR  
 IGEEAGSTYRMIGVFDHDTALDPTATTGTLDDLIERSRETPLHAILAIIPPSTDPIDHVAEINWR  
 LRSVLADVYVLPNIVHGIDVLLPIERLGFALLVLQRRPLSDWQIVKKTVLDVFLGAIALVMLAPL  
 MAAVAITIKATSPGPVFFRQPRLGFNNRTFMVFKFRMTFDTKSDMMAARQTARDDPRVTPIGKWLR  
 KLSIDELPQLLNVLRGEMSLVGPRPHAPHTRAAGMLLDDALA EYVIRHQVKGITGWAQVNGARGQ  
 LVTLDDLRRRV **E**LDLE**Y**MQR**W**S LRFDLKILMLTVVREVF SRHAF

ruler: 

(B) tr-A9HNR6-  
 A9HNR6\_GLUDA  
 (tr-A9HNR6-A9H  
 NR6\_GLUDA )  (506 aa)

(C) PS01034 GH16\_1 Glicosil hidrolases família 16 sítios ativos :  
 474 - 484: [ nível de confiança: (-1) ] ELDL.EymQRwS  
**Características previstas:**

ACT_SITE	474	nucleófilo
ACT_SITE	478	doador de prótons

**Fonte:** Elaborada pela autora, baseado nos dados do ScanProsite.

**A-** Sequência FASTA, de verde e amarelo são indicados alguns possíveis sítios ativos; **B-** representação da sequência, com símbolos coloridos indicando os sítios ativos E474 e E478; e **C-** Indicativo dos sítios ativos dessa sequência.

O sítio ativo da proteína GumD é responsável por coordenar as reações químicas que ocorrem durante a transferência desses grupos funcionais específicos, garantindo a formação adequada do poliprenilglicosilfosfato.

## 4. 2 Modelagem da proteína GumD

### 4.2.1 Análise do modelo gerado

O modelo gerado pelo I-TASSER apresentou os seguintes resultados: C-score=-1.35; Pontuação TM-score = 0.55±0.15; e RMSD estimado = 10.5±4.6Å.

Esses parâmetros indicam que o modelo é considerado razoável, mas ainda há margem para melhoria, a fim de se aproximar da estrutura da proteína nativa. Portanto, foi necessário realizar um processo de refinamento utilizando o ModRefiner, seguido de uma avaliação no MolProbity.

Após o refinamento no ModRefiner, observou-se uma melhora significativa no modelo. Isso foi confirmado pelo gráfico de Ramachandran, que indica a conformação dos resíduos na estrutura da proteína (Figura 7). De acordo com o gráfico, 90.9% de todos os resíduos estão em regiões favoráveis e 97.8% estão em regiões permitidas. Apenas onze aminoácidos estão localizados fora dessas regiões (conforme mostrado na Tabela 7).

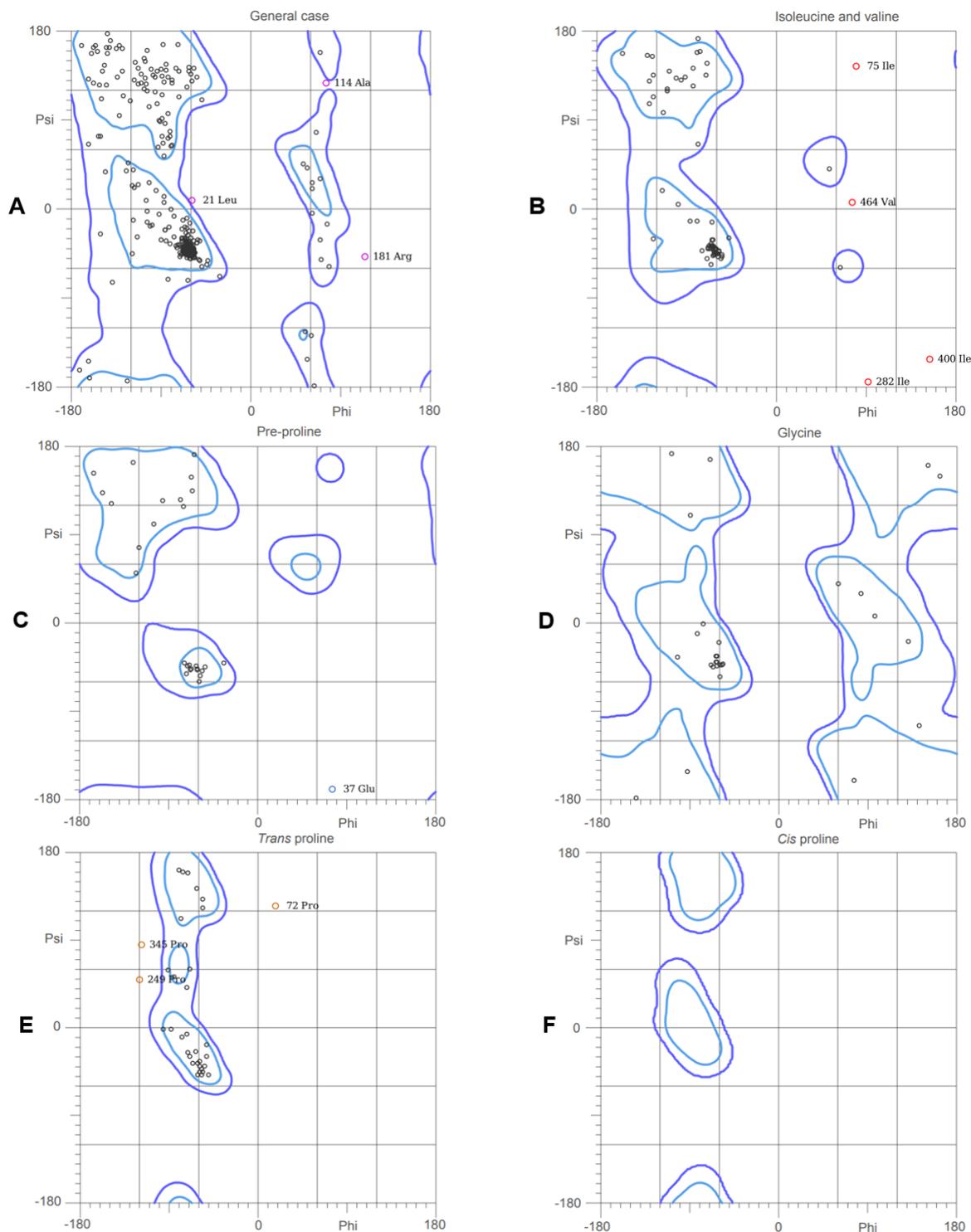
**Tabela 7 – Aminoácidos discrepantes ou fora do gráfico de Ramachandran.**

Posição	Aminoácido	Abreviação	$\phi$ ( $\Phi$ )	$\psi$ ( $\Psi$ )
21	Leucina	Leu	-59.4	9.6
37	Glutamato ou Ácido Glutâmico	Glu	76.6	-170.3
72	Prolina	Pro	18.8	126.2
75	Isoleucina	Ile	80.2	145.2
114	Alanina	Ala	76.8	128.2
181	Arginina	Arg	115.1	-48.1
249	Prolina	Pro	-120.1	50.2
282	Isoleucina	Ile	92.1	-175.1
345	Prolina	Pro	-118.8	86.1
400	Isoleucina	Ile	154.5	-153.0
464	Valina	Val	77.0	7.7

**Fonte:** Elaborada pela autora (2023).

Focando nos aminoácidos que estavam fora das regiões permitidas no gráfico de Ramachandran (Figura 7), é necessário realizar uma análise mais detalhada para compreender melhor as implicações dessas discrepâncias na estrutura tridimensional da proteína modelada.

**Figura 7 – Gráfico Ramachandran do modelo previsto para a putativa GumD de *G. diazotrophicus*.**



**Fonte:** Elaborada pela autora, com base nos dados do MolProbity (2023).

**A-** Gráfico Geral; **B-** Isoleucina e Valina; **C-** Pré-Prolina; **D-** Glicina; **E-** Trans Prolina; **e F-** Cis Prolina, que estão das regiões permitidas de acordo com o gráfico de Ramachandran.

### **Leucina**

A leucina é um aminoácido hidrofóbico que desempenha um papel fundamental na estrutura de proteínas. Sua função está relacionada principalmente à estabilidade estrutural e à formação de regiões hidrofóbicas no núcleo da proteína.

A presença da leucina nas sequências de aminoácidos de uma proteína contribui para a formação de regiões helicoidais, como alfa hélices. Essas alfa hélices são estabilizadas pela interação entre os grupos metila da leucina e outros resíduos hidrofóbicos adjacentes. A leucina também pode estar envolvida na formação de outras estruturas secundárias, como folhas beta (PINILLA et al., 2001).

Além disso, a leucina age na interação entre diferentes regiões da proteína. Pode estar envolvida na formação de interfaces entre subunidades de proteínas multímeras, contribuindo para sua montagem e estabilidade. A presença de leucina em certas posições da sequência de aminoácidos também pode influenciar a interação com outras proteínas ou moléculas, desempenhando um papel na função biológica da proteína (RICHARDSON & RICHARDSON, 1988).

A presença da leucina em uma conformação desfavorável nessa posição específica da estrutura proteica em análise pode indicar possíveis distorções estruturais. A ocorrência de uma conformação fora das regiões permitidas pode ser observada em casos em que a estrutura proteica é flexível ou sofre algum tipo de perturbação, como mutações ou interações com outros ligantes. Nessas situações, é importante avaliar se essa conformação não esperada está afetando significativamente a estrutura e função da proteína. Estudos anteriores destacaram a importância de analisar as conformações fora das regiões permitidas e compreender seu impacto na estrutura e função das proteínas (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

### **Glutamato**

O glutamato é um aminoácido que desempenha várias funções na estrutura de uma proteína. Ele pode estar envolvido em interações e ligações químicas importantes para a estabilidade e a estrutura tridimensional da proteína. Além disso, o glutamato pode participar de interações eletrostáticas, formando pontes de hidrogênio e interações iônicas com outros aminoácidos ou moléculas, contribuindo para a conformação correta da proteína. Devido à sua natureza ácida, o grupo carboxila do glutamato pode atuar como um doador ou receptor de prótons em reações químicas. Isso pode ser relevante para a regulação do pH local na estrutura

da proteína ou para a catálise de reações enzimáticas específicas. Além disso, o glutamato pode estar envolvido na formação de sítios de ligação para íons metálicos ou outras moléculas, permitindo interações específicas e cumprindo a sua função biológica da proteína (HASSUNUMA, 2023).

Quando o glutamato ou ácido glutâmico está fora do gráfico de Ramachandran, isso pode indicar que a conformação desses aminoácidos na estrutura proteica é desfavorável. No entanto, é importante levar em consideração outros fatores e o contexto em que a proteína está inserida ao interpretar essa conformação específica. Em algumas situações, os aminoácidos podem assumir conformações atípicas devido a interações específicas na estrutura ou a fatores ambientais (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Se o glutamato estiver ligeiramente fora da região permitida, mas não causar um impacto significativo na estrutura global da proteína ou em suas interações funcionais, não há motivo imediato de preocupação. Porém, se essa conformação fora do esperado afetar a estabilidade estrutural da proteína, suas interações com outras moléculas ou sua função biológica, isso pode ser motivo de preocupação e exigir uma análise mais detalhada (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

É notável ressaltar que a conformação dos aminoácidos dentro da proteína desempenha um papel crucial em sua função biológica adequada. Desvios significativos em sua estrutura tridimensional podem comprometer sua capacidade de realizar as interações necessárias para desempenhar funções específicas. Portanto, compreender e avaliar as conformações atípicas é fundamental para a compreensão da estrutura e função das proteínas em estudo (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

### **Prolina**

A prolina é um aminoácido especial devido à sua estrutura cíclica, que difere dos outros aminoácidos. Sua função principal na estrutura de uma proteína está relacionada à estabilidade estrutural e à conformação de estruturas secundárias, como as alfas hélices. A prolina possui uma cadeia lateral que se dobra de volta em direção ao grupo amina, formando uma ligação covalente com esse grupo, o que resulta em uma estrutura em anel. Essa estrutura em anel da prolina cria uma restrição estérica, o que significa que a prolina possui uma conformação rígida e

inflexível. Essa característica única da prolina é essencial para a formação de curvas ou voltas na estrutura proteica (MALSCHITZKY et al., 2021).

A presença de prolina em determinadas posições da sequência de aminoácidos pode influenciar a forma como a proteína se dobra e se organiza no espaço tridimensional. A prolina é frequentemente encontrada em regiões onde ocorrem curvas ou voltas na estrutura proteica, onde ajuda a estabilizar essas estruturas e a facilitar a mudança de direção da cadeia polipeptídica. Além disso, a prolina também pode desempenhar importância na interação entre proteínas, pois pode estar envolvida em sítios de reconhecimento e ligação de proteínas (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Devido à estrutura característica da prolina, conforme descrito nos parágrafos anteriores, é comum encontrar a prolina localizada em regiões específicas do gráfico de Ramachandran, fora das regiões permitidas para outros aminoácidos. Essas conformações são consideradas características da prolina e são amplamente toleradas em muitas estruturas proteicas (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

No entanto, é fundamental considerar o contexto e avaliar o impacto da conformação atípica da prolina na estabilidade estrutural, nas interações com outras moléculas e na função da proteína. Em alguns casos, as conformações peculiares da prolina podem desempenhar papéis funcionais importantes, como na formação de curvas estruturais ou na interação com parceiros moleculares específicos.

### **Isoleucina**

A isoleucina é fundamental na estrutura das proteínas, como um dos vinte aminoácidos essenciais, a isoleucina desempenha várias funções importantes na conformação tridimensional e estabilidade das proteínas (AZEVEDO, 1999).

Uma das principais funções da isoleucina é contribuir para a formação de estruturas secundárias, como as hélices alfa. Ela ajuda a estabilizar as interações de empacotamento entre as hélices adjacentes, promovendo a estruturação correta da proteína. A isoleucina também desempenha um papel essencial no dobramento da proteína, contribuindo para a formação de interações hidrofóbicas dentro do núcleo hidrofóbico da proteína, o que é crucial para sua estabilidade estrutural (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Além disso, a isoleucina também está envolvida em sítios ativos de enzimas, onde desempenha um papel importante na interação com substratos ou cofatores.

Sua presença em sítios ativos pode influenciar a atividade catalítica e a especificidade de ligação da enzima, afetando diretamente sua função biológica.

Outra função da isoleucina é sua participação em proteínas transmembranares, onde operam no transporte de moléculas através da membrana celular. A isoleucina pode estar envolvida na formação de sítios de ligação que facilitam o transporte seletivo de íons ou outras substâncias através da membrana, contribuindo para a função de transporte da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

A conformação da isoleucina no gráfico de Ramachandran está geralmente localizada na região das hélices  $\alpha$ , onde sua estrutura específica é acomodada. No entanto, em certas circunstâncias, podem ocorrer variações na conformação da isoleucina devido à flexibilidade estrutural ou interações específicas na proteína. Caso a conformação da isoleucina esteja fora da região permitida no gráfico de Ramachandran e tenha um impacto significativo na estrutura global da proteína, nas interações com outras moléculas ou na função da proteína, isso pode ser motivo de preocupação e exigir uma análise mais aprofundada (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

A avaliação precisa da relevância da conformação atípica da isoleucina fora da região permitida requer uma análise detalhada da estrutura, como simulações moleculares ou experimentos complementares. Essas análises são essenciais para determinar se a conformação não usual da isoleucina afeta adversamente a estrutura e a função da proteína. É importante destacar que a conformação da isoleucina fora da região permitida não necessariamente indica um problema estrutural imediato, mas é necessário compreender como essa conformação afeta a proteína como um todo (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

### **Alanina**

A alanina desempenha notoriedade importante na estrutura de proteínas, especialmente devido às suas características estruturais e químicas. A alanina é um aminoácido não polar, o que significa que sua cadeia lateral é constituída apenas por átomos de carbono e hidrogênio, sem grupos funcionais carregados ou polares. Essa natureza não polar permite que a alanina interaja com outras partes da proteína de forma hidrofóbica, contribuindo para a estabilidade e a estrutura tridimensional da proteína (MEDEIROS, 2022).

Devido à sua simplicidade estrutural, a alanina é frequentemente encontrada em locais estruturais importantes das proteínas. Ela pode ocupar posições estratégicas nas regiões de núcleo hidrofóbico, onde atua como um componente essencial para estabilizar a estrutura compacta (MEDEIROS, 2022).

Além disso, a alanina são fundamentais na formação de hélices  $\alpha$ . A alanina tem uma tendência intrínseca de adotar uma conformação em hélice  $\alpha$  devido à sua capacidade de formar ligações de hidrogênio internas. Essa característica facilita a formação e a estabilização de hélices  $\alpha$  em regiões específicas da proteína, contribuindo para sua estrutura tridimensional (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Outra função importante da alanina está relacionada à sua capacidade de ser um resíduo flexível na estrutura proteica. Devido à sua cadeia lateral simples e compacta, a alanina pode atuar como um "espaçador" ou um elo flexível entre outras regiões estruturais da proteína. Essa flexibilidade permite que a alanina exerça uma tarefa indispensável na acomodação de curvas e dobras na estrutura proteica, contribuindo para sua conformação tridimensional complexa (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

A alanina é um aminoácido amplamente tolerado em diferentes conformações dentro de uma estrutura proteica. Pequenas variações na conformação da alanina geralmente não têm um impacto significativo na estrutura global da proteína ou em suas interações funcionais. No entanto, se a conformação da alanina estiver fora da região permitida no gráfico de Ramachandran e afetar a estabilidade estrutural, as interações com outras moléculas ou a função da proteína, isso pode levantar preocupações (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

É importante destacar que a conformação da alanina fora da região permitida não é automaticamente um indicativo de um problema estrutural (RICHARDSON & RICHARDSON, 1988; TING et al., 2010). Embora a presença de alanina fora da região permitida no gráfico de Ramachandran possa chamar a atenção, é necessário avaliar cuidadosamente o impacto dessa conformação em termos de estabilidade estrutural e função da proteína. Uma abordagem integrada, combinando análises estruturais e funcionais, é essencial para compreender o significado biológico dessas conformações atípicas de alanina (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

## Arginina

A arginina desempenha várias funções importantes na estrutura de uma proteína. Ela possui uma cadeia lateral longa e carregada positivamente, o que a torna especial em termos de sua interação com outras moléculas e grupos funcionais. A presença da arginina em uma sequência de aminoácidos pode influenciar a estabilidade estrutural e as interações da proteína com outras moléculas (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Uma das principais funções da arginina é a capacidade de formar ligações de hidrogênio com outros resíduos de aminoácidos ou grupos funcionais dentro da proteína. Essas interações de hidrogênio são cruciais para a estabilização da estrutura tridimensional da proteína e para manter a sua conformação correta. Além disso, a arginina pode estar envolvida na formação de interações iônicas e pontes de sal, que contribuem para a estabilidade e funcionalidade da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Outra função importante da arginina é sua participação em sítios de ligação e reconhecimento de proteínas. Devido à sua carga positiva, a arginina pode interagir com regiões negativamente carregadas em outras proteínas, formando ligações eletrostáticas. Essas interações são cruciais para a formação de complexos proteicos, regulação de atividades enzimáticas e transdução de sinais intracelulares. Além disso, a arginina também está envolvida, na estrutura de proteínas, na compactação do DNA, como histonas, facilitando a ligação e organização do DNA nas células (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

A conformação da arginina no gráfico de Ramachandran geralmente é observada na região associada às hélices  $\alpha$  ou folhas  $\beta$ . Essas são as conformações preferidas para a arginina devido às suas características estruturais. No entanto, pequenas variações na conformação da arginina podem ocorrer sem causar um impacto significativo na estrutura global da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

É importante considerar se a conformação fora da região permitida no gráfico de Ramachandran afeta a estabilidade estrutural, as interações com outras moléculas ou a função da proteína. Caso a arginina esteja em uma região onde suas interações importantes, como ligações de hidrogênio ou interações iônicas, são comprometidas, isso pode resultar em consequências negativas para a estrutura e função da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Portanto, a presença de arginina fora da região permitida no gráfico de Ramachandran pode ser motivo de preocupação, dependendo do contexto estrutural e funcional da proteína em questão. A conformação da arginina é importante para a estabilidade e funcionalidade da proteína, e desvios significativos dessa conformação podem ter implicações na estrutura e função. A compreensão das características estruturais e do contexto da proteína em estudo é fundamental para a interpretação da relevância dessas conformações atípicas (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

### **Valina**

A valina possui importância na estrutura das proteínas devido às suas propriedades estruturais e hidrofóbicas. Como um dos aminoácidos apolares, a valina é frequentemente encontrada em regiões do núcleo proteico, onde contribui para a estabilidade da estrutura tridimensional. Sua estrutura ramificada e cadeia lateral hidrofóbica permite que a valina se acomode em ambientes hidrofóbicos, como o interior da proteína, onde ajuda a promover interações favoráveis com outras moléculas de valina e outros aminoácidos hidrofóbicos (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Além disso, a valina pode estar envolvida em interações de empacotamento, onde suas cadeias laterais se encaixam em bolsos ou cavidades formadas por outras regiões da proteína. Essas interações de empacotamento contribuem para a formação de uma estrutura compacta e estável, aumentando a rigidez da proteína. A valina também está relacionada a ancoragem de hélices  $\alpha$ , fornecendo pontos de contato hidrofóbicos que estabilizam a conformação helicoidal (RICHARDSON, 1988; TING et al., 2010).

É importante destacar que a valina, assim como outros aminoácidos, não possui uma função estritamente específica na estrutura proteica, mas sim contribui para a estabilidade, empacotamento e estrutura geral da proteína. Sua presença e posicionamento adequados são essenciais para garantir uma estrutura proteica corretamente dobrada e funcional. Variações na sequência de aminoácidos e, portanto, na presença de valina em diferentes posições podem influenciar as propriedades estruturais e funcionais da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

A conformação da valina no gráfico de Ramachandran geralmente está associada à região das hélices, onde ocorre uma torção específica para acomodar

sua estrutura. No entanto, pequenas variações na conformação da valina podem ocorrer sem causar impactos significativos na estrutura global da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

A conformação da valina fora da região permitida no gráfico de Ramachandran pode ser motivo de preocupação se afetar a estabilidade estrutural, as interações com outras moléculas ou a função da proteína. Por exemplo, se a valina estiver localizada em uma região crítica onde suas interações importantes, como ligações de hidrogênio ou interações hidrofóbicas, são comprometidas, isso pode levar a consequências negativas na estrutura e na função da proteína (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Portanto, a presença de valina fora da região permitida no gráfico de Ramachandran pode ser preocupante, dependendo do contexto específico da proteína em estudo. A conformação da valina desempenha uma função essencial na estrutura das proteínas, especialmente nas hélices. Pequenas variações nessa conformação podem ser toleradas, desde que não comprometam a estabilidade estrutural e as interações funcionais. No entanto, quando essas variações afetam negativamente a estrutura e a função da proteína, uma investigação mais aprofundada é necessária para compreender seu impacto específico (Figura 7-B) (RICHARDSON & RICHARDSON, 1988; TING et al., 2010).

Quando esses aminoácidos estão posicionados fora da região permitida no gráfico de Ramachandran, isso pode indicar uma conformação desfavorável. No entanto, é fundamental investigar se essas conformações afetam negativamente a estabilidade estrutural, as interações com outras moléculas ou a função da proteína. Cada aminoácido possui características estruturais e propriedades químicas distintas, que podem influenciar sua capacidade de adotar conformações atípicas (RICHARDSON, 1988; TING et al., 2010).

É importante destacar que o gráfico de Ramachandran fornece uma visão estatística das conformações preferenciais dos aminoácidos, porém, não descarta completamente a ocorrência de conformações atípicas. Em alguns casos, a presença de aminoácidos fora da região permitida no gráfico de Ramachandran pode ser tolerada e não ter impacto negativo na estrutura ou função da proteína. A flexibilidade estrutural, interações com outras moléculas ou adaptações específicas

da proteína podem permitir a existência de conformações não convencionais para certos aminoácidos (RAMACHANDRAN & SASISEKHARAN, 1968).

Vale ressaltar que a Prolina possui uma estrutura especial que a torna mais restrita em termos de conformações possíveis. Portanto, conformações fora da região permitida no gráfico de Ramachandran são consideradas características normais da Prolina e são bem toleradas em muitas estruturas proteicas (RICHARDSON, 1988; TING et al., 2010).

Assim sendo, embora os aminoácidos Leucina, Glutamato ou Ácido Glutâmico, Prolina, Isoleucina, Alanina, Arginina e Valina possam estar localizados fora das regiões permitidas ou favoráveis no gráfico de Ramachandran, isso não implica necessariamente em preocupações para a estrutura proteica. É crucial considerar outros fatores e o contexto específico da proteína em análise (RICHARDSON, 1988; TING et al., 2010).

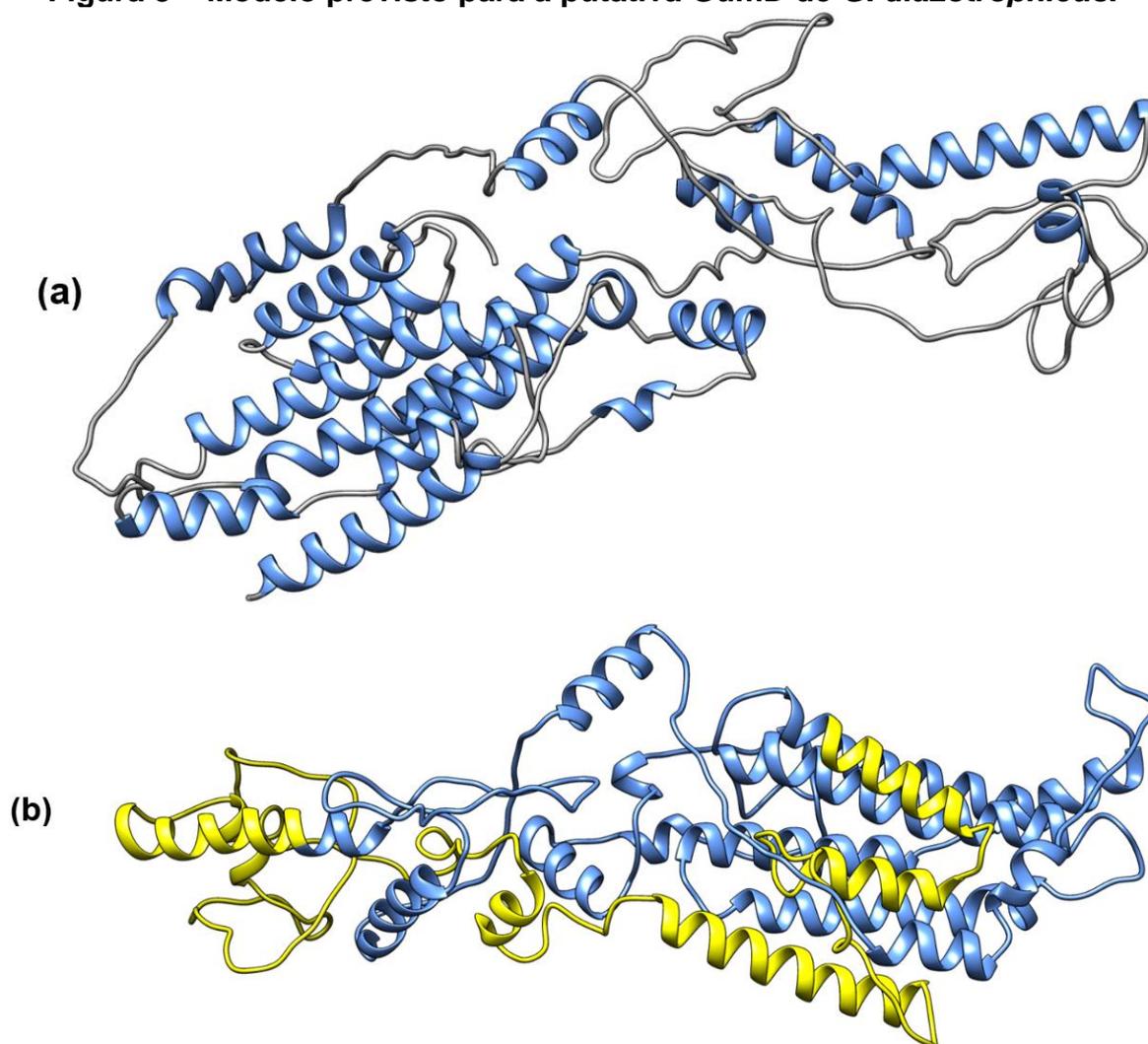
Logo se faz é necessário realizar uma análise mais aprofundada, utilizando técnicas como simulações moleculares ou experimentos complementares, para avaliar o impacto dessas conformações atípicas desses aminoácidos. Essas abordagens nos permitirão determinar se a conformação desses aminoácidos citados anteriormente, afeta adversamente a estrutura e a função da proteína em questão.

#### **4. 3 Modelo de GumD**

A Figura 8 apresenta o modelo proposto para a proteína putativa GumD de *G. diazotrophicus* PAL5. Nessa representação, é possível observar a organização da estrutura secundária da proteína em alfa-hélices ( $\alpha$ -hélice), representadas na cor azul. Além disso, é identificável o domínio da proteína, destacado na cor amarelo (Domínio de transferase de açúcar bacteriano- UniProt). Todas as imagens da estrutura tridimensional da proteína foram geradas utilizando o software Chimera 1.16.

Na Figura 9, são apresentados os dois sítios ativos da proteína GumD, relacionados aos aminoácidos 474 (Nucleophile) e 478 (Proton donor), como discutido anteriormente na seção de análise estrutural da sequência.

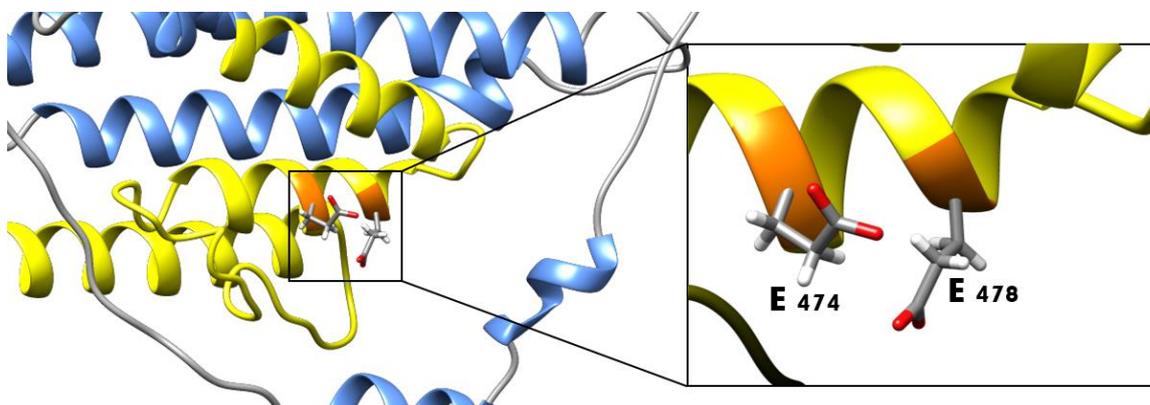
**Figura 8 – Modelo previsto para a putativa GumD de *G. diazotrophicus*.**



Fonte: Elaborada pela autora (2023).

(a) Estrutura secundária (azul); (b) Domínio da proteína (amarelo).

**Figura 9 – Modelo previsto para a putativa GumD de *G. diazotrophicus*, evidenciando os sítios ativos (ScanProsite).**



Fonte: Elaborada pela autora (2023).

A estrutura secundária é representada na cor azul e domínio na cor amarelo. Este domínio caracteriza uma região conservada de várias diferentes açúcares transferases bacterianas, envolvidas em diversas vias de biossíntese.

#### 4. 4 Análise comparativa da sequência de GumD

O resultado obtido no BLASTP permitiu a seleção de quatro espécies para o alinhamento com a sequência da proteína GumD de *G. diazotrophicus*. A escolha dessas espécies levou em consideração principalmente a porcentagem de identidade de suas proteínas com a sequência-alvo. As sequências selecionadas são referentes as espécies *G. tumulisoli*, *Nguyenibacter vanlangensis*, *Kozakia baliensis* e *Acetobacter estunensis*, apresentando identidades de 73,75%, 72,92%, 52,83% e 52,56%, respectivamente (Tabela 8).

**Tabela 8 – Espécies e identidade da proteína em relação proteína GumD da *G. diazotrophicus*.**

Espécie	Nome (NCBI)	Porcentagem de Identidade em relação a GumD de <i>G. diazotrophicus</i>
<i>Nguyenibacter vanlangensis</i>	WP_176639921.1	73.75%
<i>Gluconacetobacter tumulisoli</i>	WP_182955420.1	72.91%
<i>Kozakia baliensis</i>	WP_227003205.1	52.83%
<i>Acetobacter estunensis</i>	WP_254454693.1	52.56%

Fonte: Elaborada pela autora (2023).

No banco de dados, todas essas proteínas foram nomeadas como exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase. Essas informações são relevantes para a análise comparativa das sequências e para a inferência de possíveis funções e características da proteína GumD em *G. diazotrophicus*, com base nas proteínas homólogas encontradas em outras espécies.

O resultado do alinhamento múltiplo da proteína "exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase (tradução: biossíntese de exopolissacarídeo poliprenil glicosilfosfotransferase)" das espécies mencionadas na Tabela 8 foi realizado utilizando o Clustal Omega 1.2.4. O programa demonstrou um alinhamento satisfatório, com a presença de pequenos gaps, que são regiões onde

não houve alinhamento entre as sequências. Além disso, o alinhamento revelou a conservação de um sítio ativo no aminoácido 478 da proteína GumD, conforme representado na Figura 10.

Esses resultados são importantes para compreender a similaridade e a conservação da sequência e estrutura da proteína entre as diferentes espécies analisadas. A preservação do sítio ativo na posição do referido aminoácido indica que essa região desempenha um papel funcional importante na biossíntese de exopolissacarídeo poliprenil glicosilfosfotransferase em todas as espécies investigadas. Isso pode fornecer *insights* valiosos sobre a função e a importância dessa proteína no contexto da síntese de exopolissacarídeos.

De acordo com o banco de dados do NCBI, as proteínas listadas na Tabela 8 e representadas na Figura 11 possuem o mesmo domínio denominado "WcaJ\_sugtrans", indicando que são proteínas que compartilham uma região conservada. Esse domínio foi visualizado e representado na Figura 11.

O domínio WcaJ\_sugtrans pertence a uma superfamília que engloba diversas proteínas, incluindo a WcaJ da *Escherichia coli*, envolvida na biossíntese do ácido colânico, a EpsB do *Methylobacillus*, relacionada à biossíntese do metanolan, e a GumD, responsável pela biossíntese da xantana. Além disso, essas proteínas estão intimamente relacionadas à proteína WbaP (anteriormente conhecida como RfbP), é fundamental na biossíntese do antígeno O em *Salmonella typhimurium*. A enzima WbaP transfere galactose de UDP-galactose para um transportador de poliprenil, utilizando o domínio altamente conservado de transferase de açúcar C-terminal (pfam02397), em uma reação que ocorre na face citoplasmática da membrana interna. Acredita-se que o domínio hidrofóbico N-terminal facilite a função de "flippase", permitindo a transferência da unidade lipossacarídica da face citoplasmática para a face periplasmática da membrana interna. Esses genes estão frequentemente agrupados em operons que estão envolvidos na produção de exopolissacarídeos complexos, como o antígeno O encontrado em bactérias do gênero *Enterobacteriaceae*.

**Figura 10 – Alinhamento da proteína GumD de *G. diazotrophicus* com a proteína “biossíntese de exopolissacarídeo poliprenil glicosilfosfotransferase” das espécies *G. tumulisoli*, *N. vanlangensis*, *K. baliensis* e *A. estunensis*.**

**(A)**

<i>K. baliensis</i>	-----MSKPMPSPLTKAAHILHAFDAEM---PPAEASPDSAAGRQLVDPYPYFNPV	50
<i>A. estunensis</i>	-----MIDRSG-----A-----TGHTEELLAESGQMLMQPVRRLQLQEPYPYLHTL	44
<i>G. tumulisoli</i>	-----MSESYGQLQ-GT----ISLSNESLLEPRP-----SGTADTGPYRSPYYPHPV	43
<i>G. diazotrophicus</i>	MPDGLRMSSEWYQLQ-DS----IGLSDPTITKTPS-----ARLAE-----PGTYPYHPV	47
<i>N. vanlangensis</i>	-----MSESYSLQ-ST----IDFPDPLRQVESV-----TKPAE-----SGVYPYHPV	40
	* . : : : : : * * * * :	
<i>K. baliensis</i>	LSKIFSFCDLLSVLASSVMYLEVMWSPRPVETVAVSGMVAVITFVLLPRKRNLMNYP	110
<i>A. estunensis</i>	LAATIAMMDLAGVALGSGFAHLECSGLTRQMAEAMGAADLVAVLVFLFPKRRLLAFA	104
<i>G. tumulisoli</i>	MAGVITTSMDAFVGLGATVLCVHVEPTILMENVHVALPLADLMSALGFLMIPKRRLLIYP	103
<i>G. diazotrophicus</i>	LAGMITSDVTLAI LAATF IGASVEPAIHQDWRVAQPLADLMSALGFLMIPKRRLLHVP	107
<i>N. vanlangensis</i>	LSGMITSIDGLTVMGATLLTAYVGPVVHQQWMSQSVANLMSVLFGLMAPKRRLLIYV	100
	: : : * : : : . : . : : : : : : : : * : * : * : *	
<i>K. baliensis</i>	RVQRISAQLRYISPALFLAVLAQFIARVLDMSNTRAFDMALFWLALCIGALVLRGTET	170
<i>A. estunensis</i>	KVRRLLKACRYLVPPVLAACAHLVLCVLEMPVIQSVRLTGIWLVAVVPLIARGGLT	164
<i>G. tumulisoli</i>	DFLQVSAQVRYLARPLVIGALLHFIVLYMLHRPTAQAIELSLWVIGCAATLAVLRGGEI	163
<i>G. diazotrophicus</i>	TVQDFSQVRYLTPPLLAGALLHVIVLYMLHHPVPSFELATWLAFTGVLAVVRGTET	167
<i>N. vanlangensis</i>	TLLQSAQRYLAPPLLIGALHAVVLSMLRHPVAPSAELAAWVLFCSALAGVGRTEI	160
	: . * * * : : . . : : * : : : * : * * *	
<i>K. baliensis</i>	ILLHTRGIQNHLTRKIAIIGTGDVADQLSDRIDNDAGHTYQLGRFNDAPKSSRNSKITG	230
<i>A. estunensis</i>	FFLYQPSVIAGLRKRIAVIGSDGLASELAERISHDAGTYGLFGVDFDSEASAD-DAVKG	223
<i>G. tumulisoli</i>	LILHHSVTRHLRNLAVIGSDEVALKVAARVTEEAGPYRMVGVDFDHDALDPTVTDG	223
<i>G. diazotrophicus</i>	VLLSRPAIANQLARNVAVGSDDAIIRLAARIGEEAGSYRMIGVDFDHDALDPTATTG	227
<i>N. vanlangensis</i>	ALLGSSAISQLARNVAIVGSDETAIKFAARIGAGEAGPYRMIGVDFDHDALPNVAVDG	220
	: * . : * * : * * : * . : * . : * : * * * : * * * : *	
<i>K. baliensis</i>	TINDLVALSRHSIHAVIIALAPGSTQDEHEINHLANRLRTVLSDIYIAPYLLHGVDNL	290
<i>A. estunensis</i>	SLQALVQRSRHSLHAIILAFPEATVQ-EDVRRRTCFALRSVACDIYVTPMVTGIDQAL	282
<i>G. tumulisoli</i>	TDDLIVRSRESRLHAIILAIPPSSDP-VDHVAEISWRLRSVLSVVMPNLVHGIDVLL	282
<i>G. diazotrophicus</i>	TLDLIIERSRETPHAIILAIPTSDP-IDHVAEINWRLRSVLDVVVLPNIVHGIDVLL	286
<i>N. vanlangensis</i>	TLEDLIVRSRETPHAIILAIPTSHYDP-HDHVAEISWRLRSVLSVVLPNIVHGIDVLL	279
	: : * : * * : * * : * * : * . : : * * * : * * : * : * * *	
<i>K. baliensis</i>	PVENLGPHSLFVLQRRPLSELQTIQSAFDFVFLIVLAFVLPVLLAVAIKIDSKGPV	350
<i>A. estunensis</i>	PQFELGPNTLVLQRRPLTETQNEKRVLDIVLGFIFLFLSPLILVALLIRLDSRGPV	342
<i>G. tumulisoli</i>	PIERLGPFFALVLQRRPLSDWQIVRKTMLDVVAGAIALALTPLMCMIAIAIKATSPGPV	342
<i>G. diazotrophicus</i>	PIERLGPFFALVLQRRPLSDWQIVKKTVDVLFVGAIALVMLAPLMAAVAITIKATSPGPV	346
<i>N. vanlangensis</i>	PIEQLGPFFALVLQRRPLSDWQAVKKTMLDMVLGAIALVILAPVLAITVAIAIKATSPGPV	339
	* * * * : * * * * * : * . * : * . * : * : * * : : * : * : * * *	
<i>K. baliensis</i>	FFRQPRIGFNNRPTVFKFRSMYDMSDLGAAKQTSRDDPRVTRIGKWLKLSIDELPQI	410
<i>A. estunensis</i>	FFRQPRIGKNGREFLVYKFRSMYANQADLGAAKQTSRDDPRVTRVGGWLRKLSIDELPQL	402
<i>G. tumulisoli</i>	FFRQPRLGYNNRPFVFKFRTHMNSDMMAAKQTSRDDPRVTRIGKWLKLSIDELPQL	402
<i>G. diazotrophicus</i>	FFRQPRLGFNNRPFVFKFRTHMNSDMMAARQTSRDDPRVTRIGKWLKLSIDELPQL	406
<i>N. vanlangensis</i>	FFKQPRLGFNNRPFVYKFRSMYTNMSDMMAARQTSRDDPRVTVGRWLRKLSIDELPQL	399
	* * * * : * * * * * : * . : * : * * * * * * : * * * * * * * * * *	
<i>K. baliensis</i>	FNVLTCGMSLVGPRPHAPHTTAGGILLHDALAEVYIRHVKPGITGNAQINGSRGELITV	470
<i>A. estunensis</i>	FNVLKGEHSLVGRPHAPHTRAGGLLDALAEVLYRYHVKPGITGNAQINGARGELVTL	462
<i>G. tumulisoli</i>	LNVLKGEHSLVGRPHAPHTRAGGLLDALAEVYIRHVKPGITGNAQINGARGELVTL	462
<i>G. diazotrophicus</i>	LNVLKGEHSLVGRPHAPHTRAAGHLLDDALAEVYIRHVKPGITGNAQINGARGQLVTL	466
<i>N. vanlangensis</i>	LNVLKGEHSLVGRPHAPHTRAGGLLDALAEVYIRHVKPGITGNAQINGARGELVTL	459
	: * * * * * * * * * * : * * * * * * * * * * : * * * * * * * * * *	
<i>K. baliensis</i>	DDLKRRVALDLEYIQKWSLRFDIKIMALTVVREVF SRNAF	510
<i>A. estunensis</i>	EDLEKRVNLDLEYIRRWISILFDLKIMILTVVREVF SRHAF	502
<i>G. tumulisoli</i>	DDLRRRVEYDLEYMQKWSLRFDLKILALTVLREVF SKHAF	502
<i>G. diazotrophicus</i>	DDLRRRVEYDLEYMQKWSLRFDLKILMLTVVREVF SRHAF	506
<i>N. vanlangensis</i>	DDLRRRVTYDLEYIQKWSLRFDIKIMALTVVREVF SRHAF	499
	: * * . : * * * * * * * * * * : * * * * * * * * * * : * * * * * * * * * *	

**(B)**

<i>K. baliensis</i>	DDLKRRVALDLEYIQKWSLRFDIKIMALTVVREVF SRNAF	510
<i>A. estunensis</i>	EDLEKRVNLDLEYIRRWISILFDLKIMILTVVREVF SRHAF	502
<i>G. tumulisoli</i>	DDLRRRVEYDLEYMQKWSLRFDLKILALTVLREVF SKHAF	502
<i>G. diazotrophicus</i>	DDLRRRVEYDLEYMQKWSLRFDLKILMLTVVREVF SRHAF	506
<i>N. vanlangensis</i>	DDLRRRVTYDLEYIQKWSLRFDIKIMALTVVREVF SRHAF	499
	: * * . : * * * * * * * * * * : * * * * * * * * * * : * * * * * * * * * *	

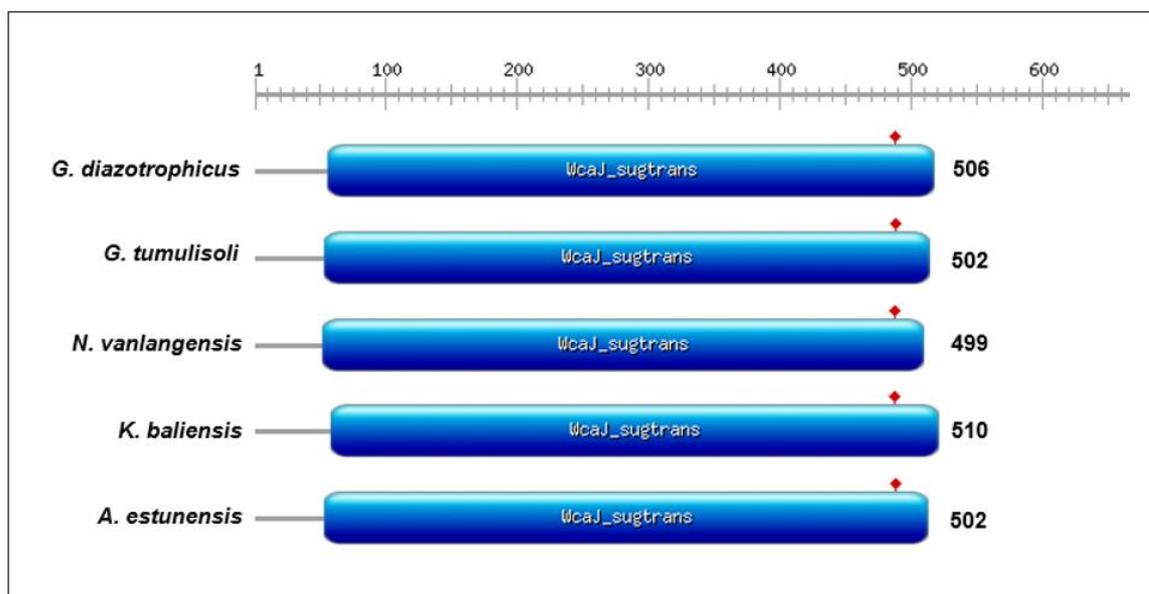
474 ←      → 478

Fonte: Elaborada pela autora (2023).

(A) Alinhamento geral; (B) Fragmento que evidencia a conservação do sítio ativo “E-478 aa).

Os asteriscos (\*) indicam o alinhamento de aa idênticos entre as sequências, e há vários. Dois pontos (: ) significa que são do mesmo grupo ou natureza de aa (Polar, Apolar, Ácidos e Básicos), e um ponto (.) significa pelo menos um grupo de natureza diferente.

**Figura 11 – Domínios conservados das sequências de biossíntese de exopolissacarídeo poliprenil glicosilfosfotransferase.**



**Fonte:** Elaborada pela autora (2023).

Domínio conservado WcaJ\_sugtrans foi destacado em azul. O losango em vermelho destaca o sítio conservado no 478 aa.

Além disso, a biossíntese do ácido colânico utiliza um transportador de glicose-undecaprenil, enquanto o *knockout* da proteína EpsB resulta na ausência da incorporação de UDP-glicose na fase lipídica. A porção C-terminal da proteína GumD demonstrou ser responsável pela atividade da glicosil-1-transferase, sendo fundamental nesse processo (LU et al., 2020).

Essas informações destacam a importância e a relação entre essas proteínas que possuem o domínio WcaJ\_sugtrans, sugerindo sua participação em vias metabólicas relacionadas à síntese de exopolissacarídeos complexos e antígenos O.

#### 4. 5 Rede de interação

A rede de interações proteicas gerada pelo STRING (Figura 12) mostra que a proteína GumD (GDI2547) interage com outras dez proteínas: exoP (GDI2549), GDI2493, GumM (GDI2548), ManC (GDI0835), GumB (GDI2552), GDI2551, GumH (GDI2550), ExoF (GDI2490), GDI2380 e GumJ (GDI2535). Essa rede possui um total de 11 nós e 41 arestas, representando as interações entre essas proteínas. Na Tabela 10, são apresentadas as proteínas que fazem parte da rede de interações

com a proteína GumD, juntamente com o tipo de interação que elas compartilham. Além disso, o score de confiabilidade dos resultados gerado pelo programa é fornecido, juntamente com a pontuação combinada. O score gerado pelo STRING é uma medida que indica a probabilidade de interação entre as proteínas. Quanto maior o score, maior é a confiabilidade da interação predita. Essa pontuação é calculada com base em diversos critérios, incluindo a presença de evidências experimentais, associação funcional em experimentos anteriores, coexpressão gênica, informações de bancos de dados e conhecimento de interações em espécies relacionadas.

É relevante enfatizar que, embora o score forneça uma estimativa da confiabilidade das interações, é necessário validar experimentalmente essas interações para confirmar sua existência e natureza. No caso das proteínas listadas na Tabela 10, é importante mencionar que não foram realizados experimentos relacionados à sua estrutura tridimensional, nem mesmo utilizando modelos computacionais. Isso significa que a conformação espacial dessas proteínas ainda não foi elucidada e pode ser objeto de futuras investigações.

A Tabela 9 fornece informações sobre os processos biológicos (Ontologia Gênica-GO) nos quais essas proteínas estão envolvidas, incluindo a biossíntese de polissacarídeos, macromoléculas, polissacarídeos celulares, transporte de polissacarídeos e polissacarídeos extracelulares. Esses processos são essenciais para a interação entre bactérias e plantas, desempenhando funções importantes na colonização bacteriana, formação de biofilmes, proteção contra estresses ambientais e nas interações benéficas ou patogênicas com as plantas.

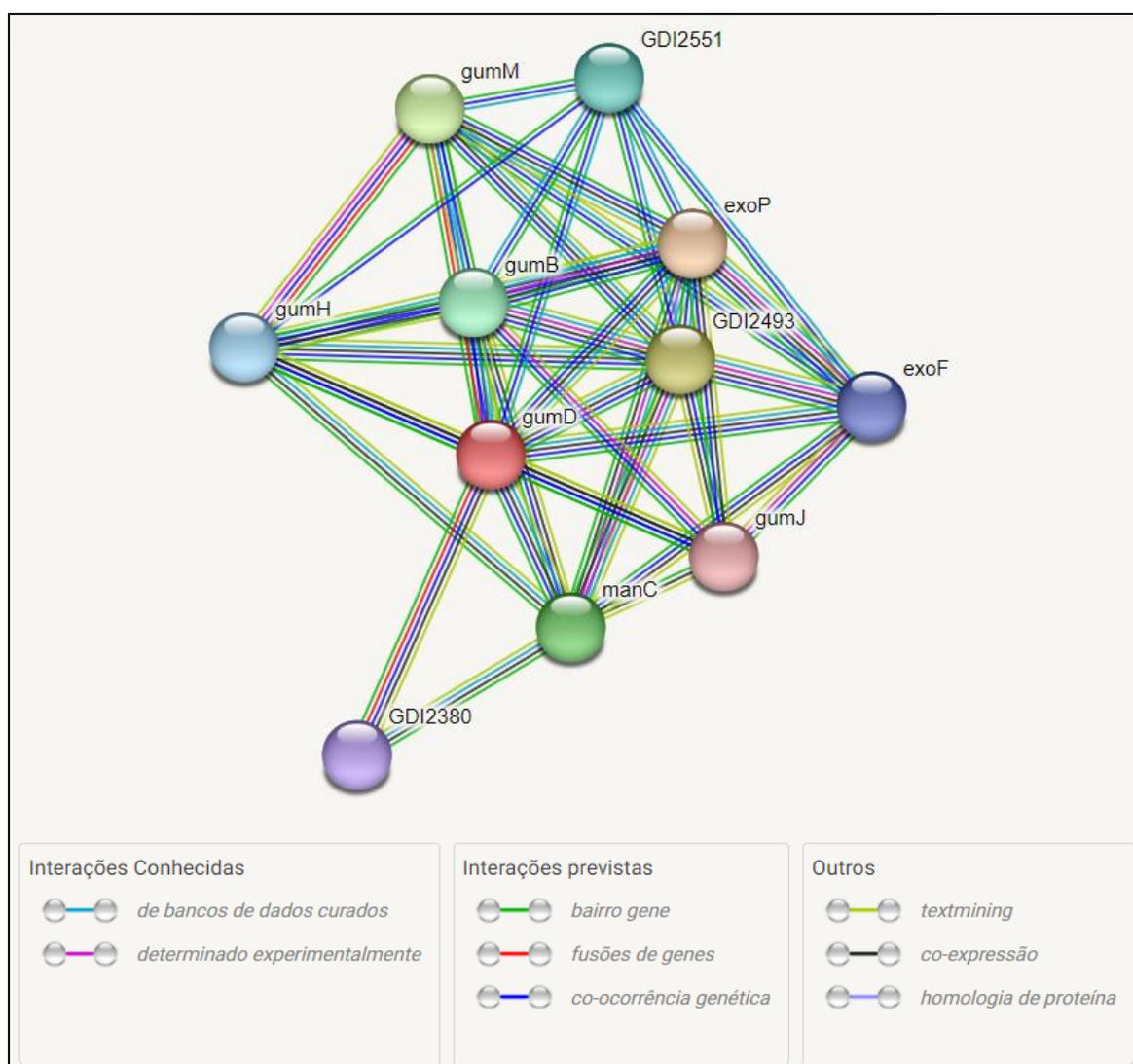
**Tabela 9– Processo biológico (Ontologia gênica-GO).**

<b>Go-term</b>	<b>Descrição</b>	<b>Taxa de descoberta falsa</b>
GO:0000271	Processo de biossíntese de polissacarídeo	8.19e-08
GO:0009059	Processo biossintético de macromoléculas	3.74e-05
GO:0033692	Processo biossintético de polissacarídeo celular	0,0289
GO:0015774	Transporte de polissacarídeos	0,0392

GO:0045226	Processo biossintético de polissacarídeo extracelular	0,0457
------------	---	--------

Fonte: Elaborada pela autora (2023).

**Figura 12 – Rede de interação da proteína GumD de *G. diazotrophicus* PAL 5.**



Fonte: Elaborada pela autora, a partir dos dados do STRING (2023).

Cada esfera representa um nó, ou seja, uma proteína. A esfera vermelha é a proteína alvo deste estudo (GumD). Já as outras esferas coloridas indicadas as proteínas na qual GumD realiza interações.

Tabela 10 – Evidências de ligação funcional entre GumD de *G. diazotrophicus* PAL5 com as demais proteínas da rede.

Proteína	Vizinhança no Genoma (score)	Fusões genéticas (score)	Coocorrência entre genomas (score)	Co-expressão (score)	Dados Experimentais/ Bioquímicos (score)	Associação em bancos de dados selecionados (score)	Co-Mencionado no Pubmed Abstracts (score)	Pontuação combinada
<b>exoP</b>	sim, distância intergênica 831 pb (0,774). Além disso, genes homólogos são vizinhos em outros genomas (0,267)	nenhum/ insignificante	Sim (0,714)	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,075)	nenhum/ insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos (0,750)	sim (0,110). Além disso, homólogos putativos são mencionados juntos em outros organismos (0,581)	<b>0,994</b>
<b>GDI2493</b>	Nenhum, mas genes homólogos são vizinhos em outros genomas (0,267)	nenhum / insignificante	sim (0,691).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,075).	nenhum / insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos (0,750).	nenhum, mas homólogos putativos são mencionados juntos em outros organismos (0,581).	<b>0,974</b>
<b>gumM</b>	sim, distância intergênica 0 pb (0,845). Além disso, genes	sim (0,419)	sim (0,360).	nenhum/ insignificante	nenhum/ insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos	sim (0,110). Além disso, homólogos putativos são mencionados juntos em outros organismos (0,184).	<b>0,957</b>

	homólogos são vizinhos em outros genomas (0,088)					(0,123)		
<b>ManC</b>	Nenhum, mas genes homólogos são vizinhos em outros genomas (0,274).	nenhum / insignificante	sim (0,201).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,225).	nenhum / insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos (0,694).	nenhum, mas homólogos putativos são mencionados juntos em outros organismos (0,583).	<b>0,932</b>
<b>GumB</b>	sim, distância intergênica 5593 pb (0,696). Além disso, genes homólogos são vizinhos em outros genomas (0,097).	nenhum / insignificante	sim (0,579).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,056).	nenhum / insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos (0,247).	sim (0,148). Além disso, homólogos putativos são mencionados juntos em outros organismos (0,123).	<b>0,921</b>
<b>GDI2551</b>	sim, distância intergênica 4239 pb (0,759). Além disso, genes homólogos	nenhum / insignificante	sim (0,402).	nenhum / insignificante	nenhum / insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos (0,247).	nenhum / insignificante	<b>0,887</b>

	são vizinhos em outros genomas (0,088).							
<b>GumH</b>	sim, distância intergênica 3058 pb (0,761). Além disso, genes homólogos são vizinhos em outros genomas (0,101)	nenhum / insignificante	sim (0,283).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,084).	nenhum / insignificante	nenhum / insignificante	sim (0,148). Além disso, homólogos putativos são mencionados juntos em outros organismos (0,085)	<b>0,864</b>
<b>ExoF</b>	Nenhum, mas genes homólogos são vizinhos em outros genomas (0,097).	nenhum / insignificante	sim (0,606).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,056)	nenhum / insignificante	nenhum, mas homólogos putativos são relatados para interagir em outros organismos (0,247)	nenhum, mas homólogos putativos são mencionados juntos em outros organismos (0,123)	<b>0,737</b>
<b>GDI2380</b>	Nenhum, mas genes homólogos são vizinhos em outros genomas (0,089).	sim (0,492).	sim (0,179).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,084).	nenhum / insignificante	nenhum / insignificante	nenhum, mas homólogos putativos são mencionados juntos em outros organismos (0,209).	<b>0,675</b>

<b>GumJ</b>	Nenhum, mas genes homólogos são vizinhos em outros genomas (0,089).	nenhum / insignificante	sim (escore 0,478).	nenhum, mas homólogos putativos são coexpressos em outros organismos (0,111).	nenhum / insignificante	nenhum / insignificante	sim (0,136). Além disso, homólogos putativos são mencionados juntos em outros organismos (0,211).	<b>0,659</b>
-------------	---	-------------------------	---------------------	---	-------------------------	-------------------------	---	--------------

**Fonte:** Elaborada pela autora, a partir dos dados do STRING (2023).

Sabe-se que a biossíntese de polissacarídeos é fundamental para a formação da matriz extracelular bacteriana, contribuindo na colonização bacteriana de tecidos vegetais. Essa matriz fornece uma estrutura física que permite a adesão bacteriana às superfícies das plantas e a formação de biofilmes, que são comunidades microbianas complexas. Além disso, os polissacarídeos extracelulares possuem importância na proteção das bactérias contra estresses ambientais, como a dessecação e a resposta a agentes antimicrobianos microbiana (NIAZI et al., 2023).

O transporte de polissacarídeos é essencial para o tráfego eficiente dessas moléculas dentro e fora das células bacterianas, a qual GumD parece estar associada. Isso é importante para a síntese, exportação e montagem adequada de polissacarídeos na superfície celular, contribuindo para a formação da matriz extracelular e a interação com as plantas.

Esses processos biológicos estabelece um papel crítico nas interações entre bactérias e plantas, permitindo a adaptação bacteriana aos nichos ambientais presentes nas plantas, a colonização dos tecidos vegetais e a modulação das respostas do sistema imunológico das plantas. Essas interações podem ser benéficas, como no caso de simbiose entre bactérias fixadoras de nitrogênio e plantas leguminosas, ou patogênicas, como em infecções bacterianas que causam doenças nas plantas (BOGINO et al., 2013; VELMOUROUGANE et al., 2017).

Na Tabela 11, são sugeridos três domínios nos quais as proteínas que fazem parte da rede estão envolvidas: o Chain Length Determinant Protein (Proteína Determinante do Comprimento da Cadeia), a Polysaccharide Biosynthesis/Export Protein (Proteína de Biossíntese/Exportação de Polissacarídeos) e o SLBB Domain (Domínio SLBB).

**Tabela 11 – Domínios de proteína (Pfam).**

Domínio	Descrição	Taxa de descoberta falsa
PF02706	Proteína determinante do comprimento da cadeia	0,0396
PF02563	Biossíntese de polissacarídeos/proteína de exportação	0,0396
PF10531	Domínio SLBB	0,0396

**Fonte:** Elaborada pela autora (2023).

A família de proteínas "Proteína Determinante do Comprimento da Cadeia" engloba proteínas relacionadas à biossíntese de lipopolissacarídeos (LPS). Essas

proteínas possuem uma participação determinante no comprimento da cadeia do componente O-antígeno do LPS, resultando em uma distribuição modal do comprimento da cadeia. Além disso, essa região também é encontrada em tirosina quinases, como a proteína P38134, que desempenham funções em outros processos biológicos (FRANCO; LIU; REEVES, 1998).

Por outro lado, a família "Biossíntese de polissacarídeos/proteína de exportação" é composta por proteínas periplasmáticas envolvidas na biossíntese e/ou exportação de polissacarídeos. Essas proteínas são indispensáveis na síntese e transporte adequados de polissacarídeos na célula, contribuindo para a formação de matrizes extracelulares e para a interação com o ambiente (STEVENSON et al., 1996).

O Domínio SLBB, é um domínio de ligação solúvel  $\beta$ -grasp. Esse domínio é encontrado em uma variedade de proteínas, incluindo aquelas envolvidas na captação de vitamina B12 em animais, como a transcobalamina e o fator intrínseco. Além disso, algumas proteínas desse domínio podem estar envolvidas no transporte de elétrons em complexos de membrana ou na exportação do polissacarídeo extracelular ácido colânico da célula (BURROUGHS et al., 2007).

Esses domínios fornecem informações importantes sobre as funções potenciais das proteínas presentes na rede. Eles estão relacionados à biossíntese de polissacarídeos, exportação de polissacarídeos e interações com moléculas solúveis. Essas funções refletem papéis cruciais nas interações entre bactérias e plantas, incluindo a formação de biofilmes, a colonização bacteriana e a proteção contra estresses ambientais (SENGUPTA & DEY, 2019; NGUYEN et al., 2020; SAEED et al., 2021; TRIVEDI et al., 2020; AJIJAH et al., 2023).

As proteínas da rede desempenham funções que foram identificadas através da análise da ontologia gênica (GO), conforme apresentado na Tabela 12. A maioria dessas proteínas está associada à membrana, assim como a proteína GumD. Elas ainda têm grande valor nos processos biossintéticos de polissacarídeos e possuem atividade de transferase em nível molecular.

**Tabela 12 – Informações gerais das proteínas da rede de interação gerada pelo STRING.**

Proteína	UniProt ID	Componente Molecular	Função Molecular	Processo biológico	Nome	Comprimento (aa)
<b>GumD</b>	<b>A9HNR6</b>	membrana	atividade de transferase	Processo biossintético de polissacarídeo	Exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase	506
<b>ExoP</b>	<b>A9HNS0</b>	membrana	-	-	Putative succinoglycan biosynthesis transport protein ExoP	738
<b>GDI2493</b>	<b>A9HNF3</b>	membrana	-	-	Putative succinoglycan biosynthesis transport protein	740
<b>GumM</b>	<b>A9HNR9</b>	-	atividade de transferase	Processo biossintético	Putative glycosyl transferase	272
<b>ManC</b>	<b>A9HB78</b>	-	atividade de guaniltransferase (GTP) de manose-1-fosfato	-Processo biossintético da metionina - Processo biossintético de polissacarídeo	mannose-1-phosphate guanylyltransferase	487
<b>GumB</b>	<b>A9HNS8</b>	membrana	Atividade de transporte transmembrana de polissacarídeo	-	Putative exopolysaccharide production protein	217
<b>GDI2551</b>	<b>A9HNS6</b>	membrana	-	-	Outer membrane protein	443
<b>GumH</b>	<b>A9HNS3</b>	-	atividade da glicosiltransferase	-	Putative glycosyl transferase protein	390
<b>ExoF</b>	<b>A9HNE5</b>	membrana	Atividade de transporte transmembr	-	Putative exopolysaccharide production protein ExoF	200

			ana de polissacarídeo			
<b>GDI2380</b>	<b>A9HMM4</b>	-	atividade da glicosiltransferase	-	Glycosyl transferase group 1	379
<b>GumJ</b>	<b>A9HNP1</b>	membrana	-	-	Putative lipopolysaccharide biosynthesis protein	518

**Fonte:** Elaborada pela autora, a partir dos dados do STRING (2023).  
(-) Não há informação no banco de dados.

As proteínas mencionadas estão presentes nos clusters identificados pelo String, desempenham papéis essenciais nas interações entre bactérias e plantas. Cada cluster representa um conjunto de proteínas que compartilham características e processos biológicos semelhantes. Abaixo, (Tabela 13) será explicado sucintamente cada um desses clusters identificados e correlacionados a rede de interação de GumD:

**Tabela 13 – Clusters identificados e relacionados com a rede de interação de GumD.**

<b>Nome do Cluster</b>	<b>Explicação</b>
Processo metabólico de polissacarídeos e Glicosiltransferases grupo 1	Este cluster envolve proteínas relacionadas ao processo metabólico de polissacarídeos, bem como às enzimas conhecidas como glicosil transferases do grupo 1. Essas enzimas são responsáveis pela transferência de grupos glicosil de um substrato para outro, sendo fundamental na biossíntese de polissacarídeos.
Processo metabólico de polissacarídeos e Glicosiltransferase como família 2	Neste cluster, encontramos proteínas relacionadas ao processo metabólico de polissacarídeos, juntamente com membros da família de glicosiltransferases semelhantes à glicosiltransferase. Essas proteínas possuem semelhanças estruturais e funcionais com as glicosiltransferases, mas podem desempenhar papéis específicos na biossíntese e modificação de polissacarídeos.
Domínio de ligação Co-A e glicosil transferase WecB/TagA/CpsF	Esse cluster inclui proteínas que possuem o domínio de ligação a coenzima A (Co-A) e pertencem à família de glicosiltransferases WecB/TagA/CpsF. Essas proteínas exercem na transferência de grupos glicosil e estão envolvidas na biossíntese de polissacarídeos específicos.
Proteína de biossíntese de polissacarídeos e putativa	Neste cluster, encontramos proteínas relacionadas à biossíntese de polissacarídeos e a uma proteína putativa de porina do tipo barril-beta. Essas proteínas podem

beta-barril porina 2	desempenhar na formação de complexos de transporte e exportação de polissacarídeos, contribuindo para a integridade e funcionalidade da matriz extracelular de polissacarídeos.
----------------------	---

**Fonte:** Elaborada pela autora (2023).

Esses clusters fornecem uma organização funcional das proteínas na rede, agrupando-as com base em características e processos biológicos semelhantes. Cada cluster representa um conjunto de proteínas que cumprem papéis específicos na biossíntese, modificação e transporte de polissacarídeos, sendo essenciais para as interações entre bactérias e plantas. Afinal, essas proteínas, juntamente com outros fatores, contribuem para a capacidade das bactérias de interagir com as plantas de diferentes maneiras. Elas podem facilitar a adesão inicial das bactérias às superfícies das plantas, permitir a formação de estruturas protetoras, modular as respostas imunológicas das plantas e estabelecer relações simbióticas ou patogênicas com as plantas (DAS, 2022; ATES, 2015; SZKLARCZYK et al., 2015).

#### **4. 6 Análise das interações entre os componentes do interactoma de GumD**

Em relação à rede de interação proteína-proteína do STRING, as informações fornecidas nos resultados específicos possibilitaram a obtenção de inferências importantes, que serão apresentadas e discutidas a seguir:

##### **4.6.1 Vizinhaça no Genoma**

A vizinhaça gênica (Figura 13) refere-se à expressão de genes localizados em regiões próximas no genoma. Esse aspecto é especialmente interessante quando se trata de bactérias, devido ao tamanho reduzido do genoma bacteriano. No caso específico, é importante destacar a presença do operon, que consiste em um grupo de genes procarióticos (e alguns eucarióticos) que estão funcionalmente relacionados, contíguos e coordenadamente controlados, sendo todos expressos a partir de um único RNA mensageiro (IGARASHI et al., 2001).

Ao analisar o score fornecido pelo STRING, é possível determinar a qualidade da interação entre as proteínas. Quanto mais próximo de 1, melhor é a pontuação (score). Nesse contexto, as proteínas GumM (score 0,845), ExoP (score 0,774) e

GDI2551 (score 0,759) apresentaram os melhores resultados, indicando uma alta probabilidade de interação, estando próximas um do outro no genoma (Tabela 10).

Além disso, essas proteínas possuem homólogos nos três domínios da vida: Bacteria, Eukarya e Archaea. Entre os exemplos de homólogos estão *Natronolimnobius innermongolicus*, *Halogramum salarium*, *Natronococcus amylolyticus*, *Halobiforma lacisalsi* e *Natronorubrum bangense*.

Essas informações fornecem *insights* valiosos sobre a presença e interação dessas proteínas em diferentes organismos, destacando sua relevância biológica e potenciais funções conservadas ao longo da evolução.

A literatura científica menciona diversas bactérias conhecidas como "Bactérias promotoras de crescimento de plantas", que estabelecem associações benéficas com várias espécies vegetais. Algumas dessas bactérias incluem: *Klebsiella* sp., *Curtobactéria* sp., *Hartmannibacter diazotrophicus*, *Bacillus* sp., *Azotobacter* sp., *Pseudomonas* sp., *Rhizobium* sp., *Ochrobactrum intermedium*, *Glutamicibacter* sp. YD01, *Micrococcus* sp., *Rhizobacteria pseudomonas*, *Aeromonas* sp., *Exiguobacterium aurantiacum*, *Burkholderia* sp., *Kocuria* sp., *Achromobacter* sp., *Enterobacter* sp., *Sphingomonas* sp., *Arthrobacter* sp., e *Variovorax* sp., entre outras (SANTOS DIAS & SANTOS, 2022).

Dentre esses organismos, destaca-se a vizinhança de *Bacillus* sp., *Azotobacter vinelandii* sp., *Pseudomonas* sp., *Rhizobium* sp., *Aeromonas* sp., *Klebsiella* sp., *Burkholderia* sp., *Achromobacter* sp., *Enterobacter* sp., *Sphingomonas* sp., *Rhizobium* sp. e *Variovorax* sp. Essa vizinhança gênica sugere a conservação de genes envolvidos na interação planta-bactéria. No caso específico do gene *gumD* de *G. diazotrophicus*, é possível inferir uma possível associação com a planta ao comparar os genes homólogos em outros organismos bacterianos conhecidos por estabelecer essa interação benéfica com organismos vegetais (MIGUEL et al., 2021).

Essa descoberta aponta para a importância do gene *gumD* na interação entre *G. diazotrophicus* e plantas. A preservação desse gene em organismos que interagem positivamente com plantas sugere que ele desempenha uma relevância na promoção do crescimento e no estabelecimento de uma relação benéfica entre a bactéria e as plantas hospedeiras.

Essas informações fornecem uma base sólida para futuras pesquisas sobre as funções específicas do gene *gumD* e seus mecanismos de interação com as

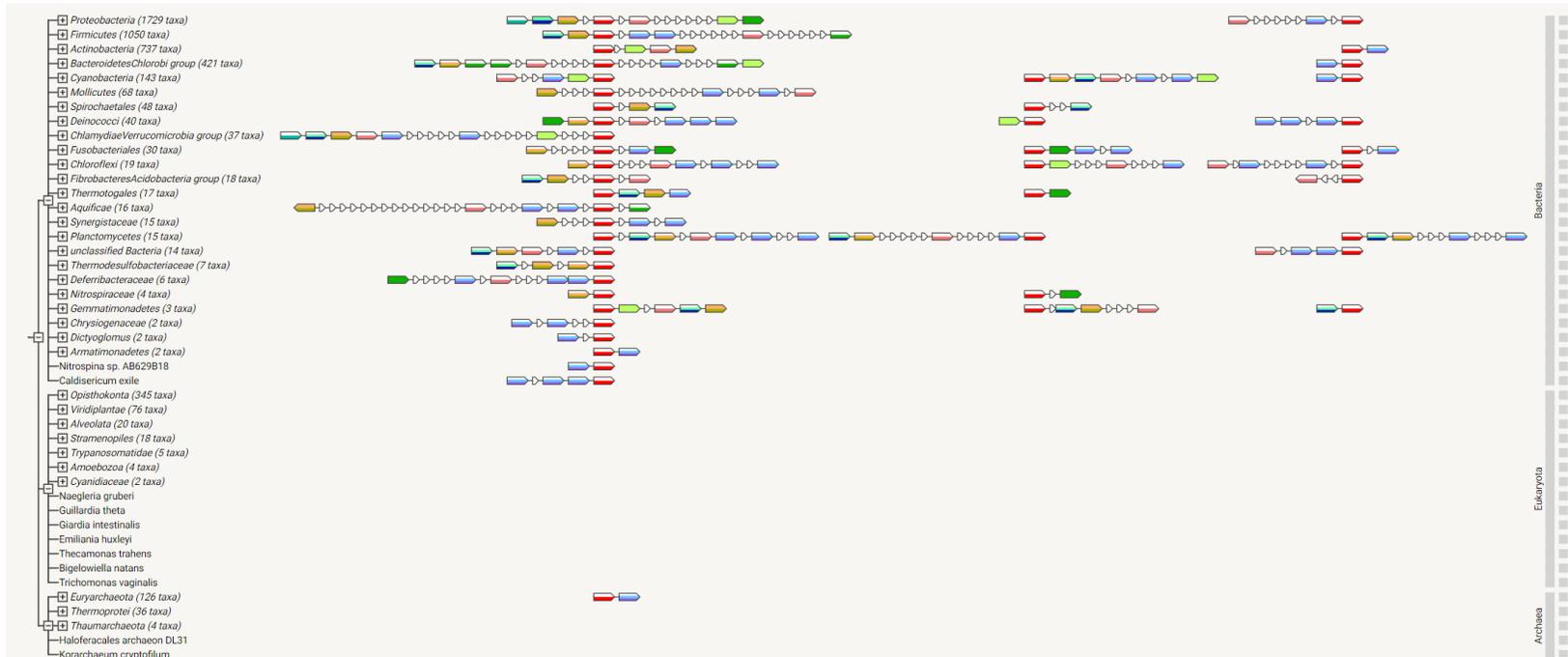
plantas. Compreender os detalhes moleculares dessa interação pode abrir caminhos para a exploração de estratégias agrícolas sustentáveis e o desenvolvimento de produtos biotecnológicos voltados para a melhoria do crescimento e saúde das plantas.

#### **4. 6.2 Fusões genéticas**

A fusão gênica ocorre quando genes de diferentes regiões do genoma se combinam ao longo do tempo. Neste estudo, foram identificadas duas proteínas que apresentaram fusão nuclear: GumM e GDI2380, com scores de 0,419 e 0,492, respectivamente. Essas fusões foram observadas apenas no domínio Bacteria, conforme representado na Figura 14.

No caso da fusão entre GumD e GumM, ela ocorreu em *Oscillatoriales cyanobacterium*, *Desulfobulbus japonicus* e *Rhodobacteraceae* sp. Já a fusão entre GumD e GDI2380 foi identificada em *Sphingobium* sp. YL23, *Parvularcula oceani*, *Paenibacillus* sp. JDR2, *Paenibacillus* sp. MSt1, *Paenibacillus elgin*, *Dorea longicatena* DSM13814, *Deinococcus wulumuqiensis* e *Acidobacteriaceae bacterium* KBS96.

Figura 13 – Vizinhança da rede de interação da GumD com as demais proteínas.



**Legenda:**

**Referência de cores**

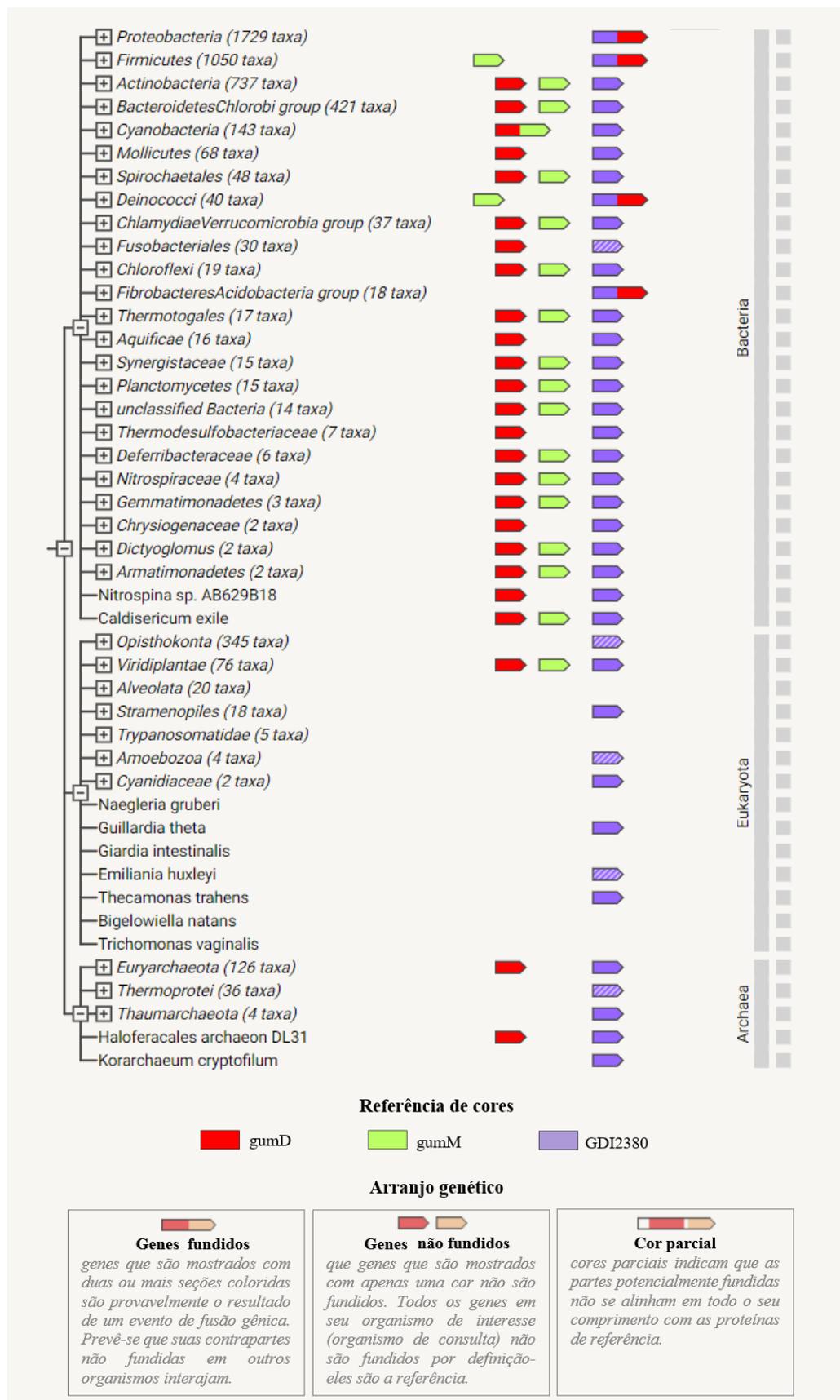
<span style="color: red;">■</span> gumD	<span style="color: yellowgreen;">■</span> gumM	<span style="color: teal;">■</span> GDI2551	<span style="color: purple;">■</span> GDI2380
<span style="color: orange;">■</span> exoP	<span style="color: green;">■</span> manC	<span style="color: lightblue;">■</span> gumH	<span style="color: pink;">■</span> gumJ
<span style="color: olive;">■</span> GDI2493	<span style="color: lightgreen;">■</span> gumB	<span style="color: darkblue;">■</span> exoF	

**Arranjo genético**

<p><b>Gene Runs consecutivos</b></p> <p><i>os genes são mostrados em execuções consecutivas, se estiverem localizados próximos no mesmo cromossomo. Eles são orientados na mesma direção ou em uma orientação frente a frente (provavelmente compartilhando promotores).</i></p>	<p><b>Seções horizontais</b></p> <p><i>seções horizontais indicam que as relações de ortologia do gene são complexas. Isso ocorre devido a eventos de duplicação de genes (paralogia) ou devido a problemas técnicos ao atribuir a ortologia.</i></p>
--	---

Fonte: Elaborada pela autora, a partir dos dados do STRING (2023).

Figura 14 – Fusões genéticas da rede de interação da GumD com as demais proteínas.



Fonte: Elaborada pela autora, a partir dos dados do STRING (2023).

#### 4.6.3 Coocorrência entre genomas

Na análise de co-ocorrência, ocorre o alinhamento de fragmentos dos genes, indicando uma sobreposição entre eles. Na Figura 15, quanto mais intenso for o sinal, maior será a similaridade entre os genes e, conseqüentemente, a probabilidade de co-ocorrência nos genomas. Quando um "quadrado" está faltando a parte superior, indica que houve alinhamento em uma parte do gene e não na outra.

A Figura 15 também mostra que a co-ocorrência ocorreu nos três domínios. No entanto, observou-se uma expressão predominante em *Proteobacteria*, uma vez que esse filo é o qual a *G. diazotrophicus* faz parte.

Dentre as proteínas analisadas, aquelas que obtiveram os melhores scores foram: ExoP (0,714), GDI2493 (0,691), ExoF (0,606) e GumB (0,579). Esses scores indicam uma alta probabilidade de co-ocorrência dessas proteínas nos genomas estudados, sugerindo uma possível relação funcional entre elas.

É possível estabelecer uma hipótese de co-ocorrência desses genes/proteínas relacionados à biossíntese e transporte de exopolissacarídeos (EPS) na membrana plasmática da bactéria.

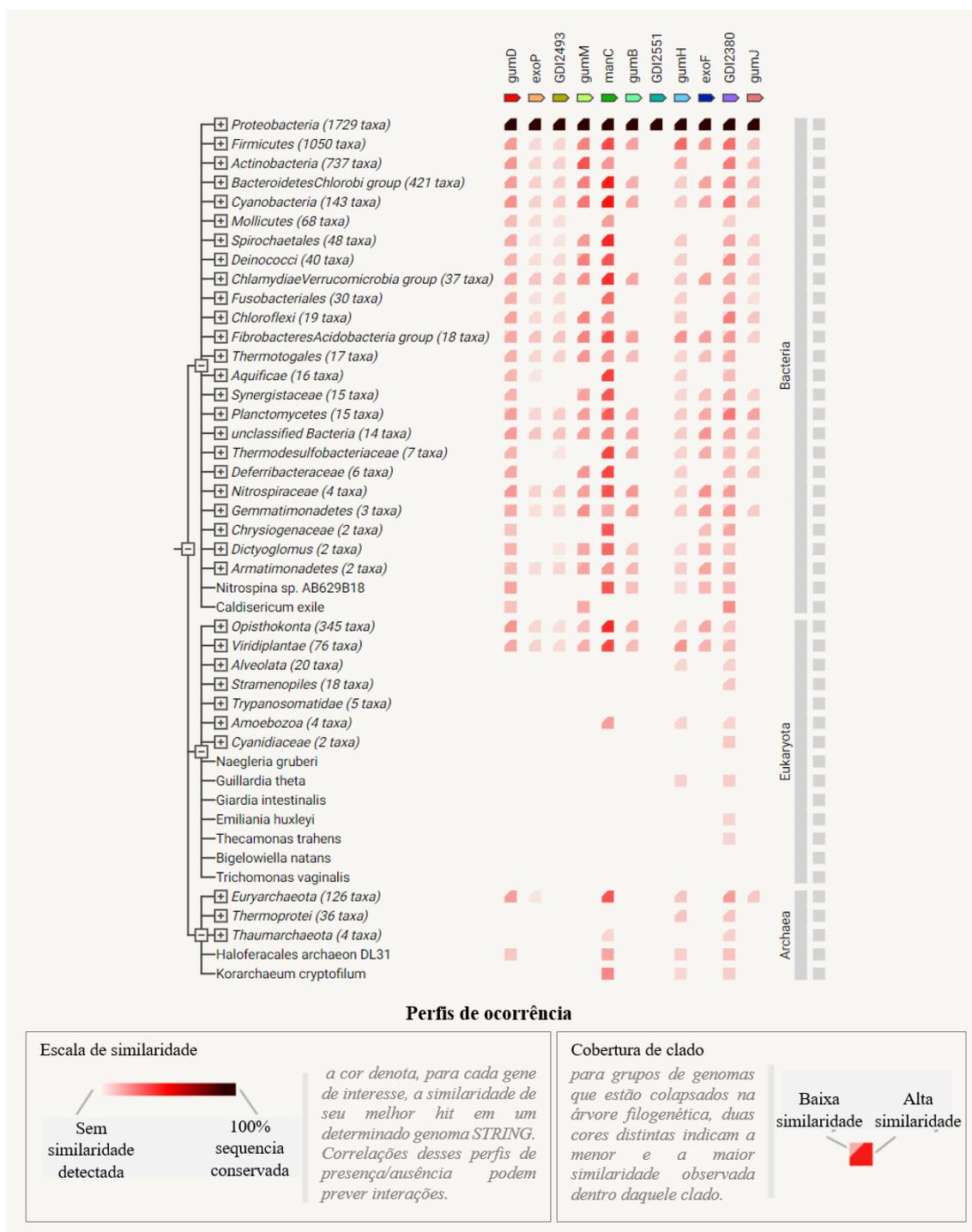
As proteínas Exop, GDI2493, ExoF e GumB possuem nomes que indicam sua possível função na biossíntese e transporte de exopolissacarídeos. A presença dessas proteínas na membrana plasmática sugere que elas podem estar envolvidas no processo de produção e exportação dos exopolissacarídeos para o ambiente externo.

Uma possível hipótese é que a proteína Exop, denominada "Putative succinoglycan biosynthesis transport protein Exop (tradução: Proteína de transporte de biossíntese putativa de succinoglicano Exop)", esteja envolvida no transporte dos precursores ou subunidades necessários para a síntese de succinoglicanos, que são um tipo específico de exopolissacarídeo. A proteína GDI2493, denominada "Putative succinoglycan biosynthesis transport protein (tradução: Proteína de transporte da biossíntese putativa de succinoglicano)", também pode desempenhar um papel semelhante na exportação dos exopolissacarídeos.

Por outro lado, as proteínas ExoF, denominada "Putative exopolysaccharide production protein ExoF (tradução: Proteína de produção de exopolissacarídeo ExoF putativa)", e GumB, denominada "Putative exopolysaccharide production protein

(tradução: Proteína de produção de exopolissacarídeo putativa)", podem estar envolvidas diretamente na síntese ou modificação dos exopolissacarídeos na membrana plasmática. Elas podem ser responsáveis por catalisar reações específicas de biossíntese dos exopolissacarídeos ou estar envolvidas na montagem dos componentes necessários para a sua produção.

**Figura 15 – Coocorrência entre genomas da rede de interação da GumD com as demais proteínas.**



Fonte: Elaborada pela autora, a partir dos dados do STRING (2023).

A co-ocorrência dessas proteínas em genomas e estarem localizados na membrana plasmática sugere uma possível interação funcional entre elas, onde cada uma tem função específica na síntese, modificação e transporte dos exopolissacarídeos. Esses exopolissacarídeos podem desempenhar funções importantes para a bactéria, como proteção contra estresses ambientais, formação de biofilmes, adesão a superfícies e interações com outros organismos, incluindo plantas hospedeiras.

#### **4.6.4 Co-expressão**

A co-expressão apresentou uma associação mais significativa entre GumD e ManC, o que sugere uma possível interação funcional entre essas duas proteínas. Essa associação é baseada na observação da coexpressão de homólogos dessas proteínas em outras espécies (Figura 16).

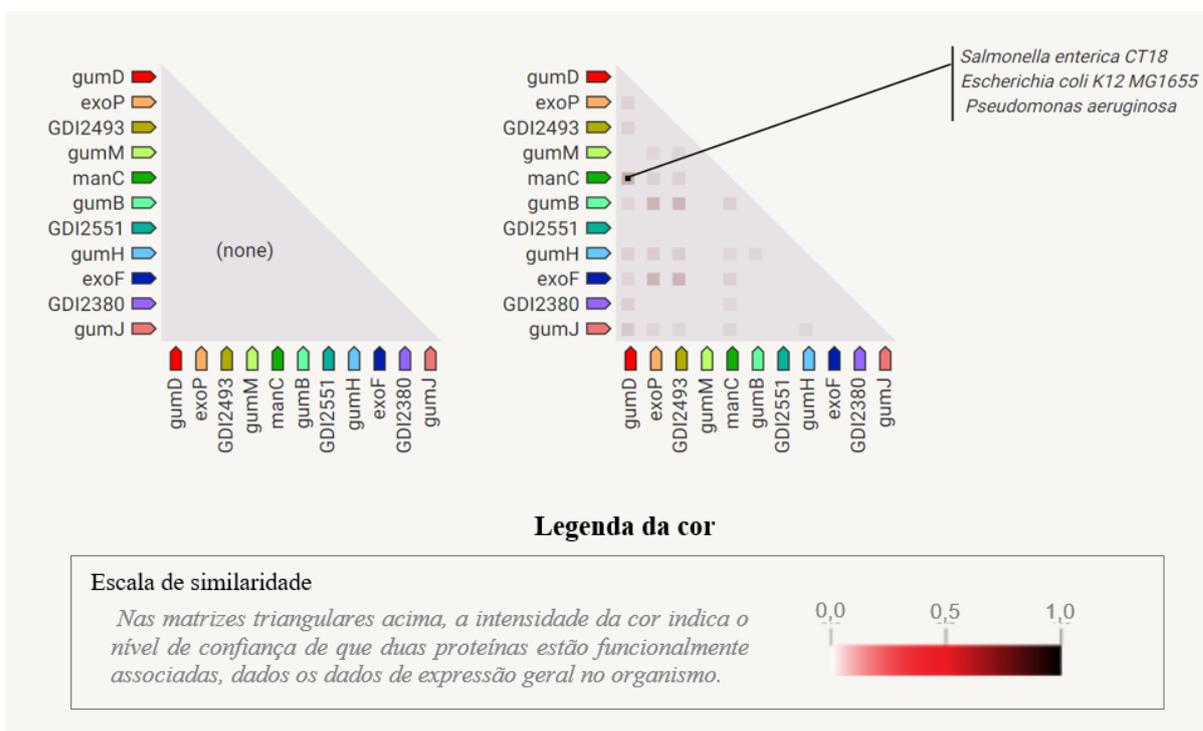
É interessante notar que a proteína ManC obteve o melhor score de homologia putativa (0,225), no entanto, esse valor é relativamente baixo. Isso pode indicar que, embora haja uma tendência de coexpressão entre GumD e ManC com base em outros organismos, a força dessa associação pode ser moderada ou variar em diferentes contextos.

Sabe-se que GumD e ManC estão envolvidos em vias metabólicas relacionadas à biossíntese de polissacarídeos. A enzima ManC, denominada "mannose-1-phosphate guanylyltransferase (tradução: manose-1-fosfato guanililtransferase)", está envolvida no processo de biossíntese da manose-1-fosfato, que é um intermediário na síntese de polissacarídeos. Essa enzima desempenha um papel importante na incorporação de manose em cadeias de polissacarídeos (BERTALAN et al., 2009). Por outro lado, GumD, denominada "Biossíntese de exopolissacarídeos poliprenil glicosilfosfotransferase", é uma enzima com função de transferase e está associada ao processo biossintético de polissacarídeos. Essa proteína está localizada na membrana, indicando sua participação na síntese e transporte de polissacarídeos na célula bacteriana.

Com base nessa informação, uma hipótese é que GumD e ManC podem estar envolvidos em etapas complementares da via de biossíntese de polissacarídeos. A enzima ManC seria responsável pela produção do substrato necessário para a síntese de polissacarídeos, enquanto GumD estaria envolvida na transferência e modificação desses polissacarídeos na membrana.

A coexpressão observada entre GumD e ManC pode indicar uma regulação coordenada dessas enzimas, permitindo uma produção eficiente e sincronizada de polissacarídeos.

**Figura 16 – Co-expressão da rede de interação da GumD com as demais proteínas.**



**Fonte:** Elaborada pela autora, a partir dos dados do STRING (2023).

#### 4.6.5 Dados Experimentais/Bioquímicos

Não foram realizados experimentos diretamente com *G. diazotrophicus*, apenas com homólogos, como *Escherichia coli* K12 MG1655 (NESPER et al., 2003; COLLINS et al., 2007). Portanto, é crucial a realização de experimentos envolvendo as proteínas encontradas no interactoma específico de *G. diazotrophicus* para impulsionar o avanço científico.

#### 4.6.6 Associação em bancos de dados selecionados

Também não foram encontrados experimentos específicos com *G. diazotrophicus*, apenas com homólogos relacionados a vias metabólicas de formação de biofilme em *Vibrio cholerae*, biossíntese de xantana e *Pseudomonas aeruginosa*.

As proteínas que apresentaram os melhores scores em homólogos putativos, indicando interações em outros organismos, foram: ExoP (0,750), GDI2493 (0,750) e ManC (0,694).

#### **4.7 Clusters**

Após a aplicação do algoritmo de clustering, foram identificados três grupos distintos na rede de interação estudada. O primeiro grupo, denominado Cluster 1, apresentou um total de 13 genes. O segundo grupo, chamado Cluster 2, foi composto por 22 genes, enquanto o terceiro grupo, Cluster 3, incluiu 6 genes (Tabela 14 e Figura 17).

Uma análise geral da rede revelou que esses clusters representam conjuntos de nós que possuem interações mais frequentes entre si do que com os nós fora de seus respectivos grupos. Essa organização em clusters evidencia uma estrutura modular na rede, onde os genes em cada grupo tendem a desempenhar funções ou estar envolvidos em processos biológicos semelhantes.

A análise de clustering aplicada à rede de interação resultou nos seguintes resultados: A rede era composta por 41 nós e 223 arestas, indicando uma densidade de conexões considerável. A média de grau dos nós foi de 10.9, o que sugere um nível razoável de interação entre os elementos da rede. Além disso, o coeficiente médio de clustering local foi de 0.698, o que indica uma forte tendência de formação de grupos ou sub-redes densamente conectadas dentro da rede global.

O número esperado de arestas, calculado com base na densidade média da rede, foi de 63, revelando que a rede observada possui um número de arestas significativamente maior do que o esperado aleatoriamente. Esse resultado é reforçado pelo valor extremamente baixo ( $p$ -valor  $< 1.0e-16$ ) obtido no teste de enriquecimento de interações proteína-proteína (PPI enrichment), indicando que as conexões observadas na rede são estatisticamente significantes e não ocorreriam ao acaso.

Esses resultados sugerem que a rede de interação analisada apresenta uma estrutura organizada, com grupos de nós altamente conectados entre si e uma quantidade de conexões superior à esperada. Essas descobertas são relevantes para a compreensão dos padrões de interação e do funcionamento dos componentes da rede em estudo.



O STRING revelou que esses grupos estão envolvidos em várias categorias, incluindo processos biológicos, funções moleculares, agrupamentos locais de rede, vias KeGG, palavras-chave anotadas (uniProt) e domínios proteicos (Pfam).

O GO (Processos biológicos) mostrou que estas proteínas estão envolvidas nos seguintes processos (processo/taxa falsa de descoberta- Quanto mais próximo de zero for a taxa de descoberta falsa, menor é a probabilidade de ser um resultado falso): Processo biossintético de UDP-glucuronato (0.0410); Transporte de polissacarídeos (0.0100); Processo biossintético de polissacarídeo extracelular (0.0164); Processo metabólico nucleotídeo-açúcar (0,0027); Processo de biossíntese de polissacarídeo ( $5.52e-13$ ); Processo metabólico de polissacarídeos ( $1.75e-13$ ); Processo biossintético de nucleotídeo-açúcar (0,0305); Transporte transmembrana de carboidratos (0,0054); Processo de biossíntese de carboidratos ( $3.46e-13$ ); Processo metabólico de polissacarídeos celulares (0,00082); Processo biossintético de polissacarídeo celular (0,0038); e Processo metabólico de carboidratos ( $2.05e-17$ ). A Figura 18 apresenta uma visualização esquemática desses processos.

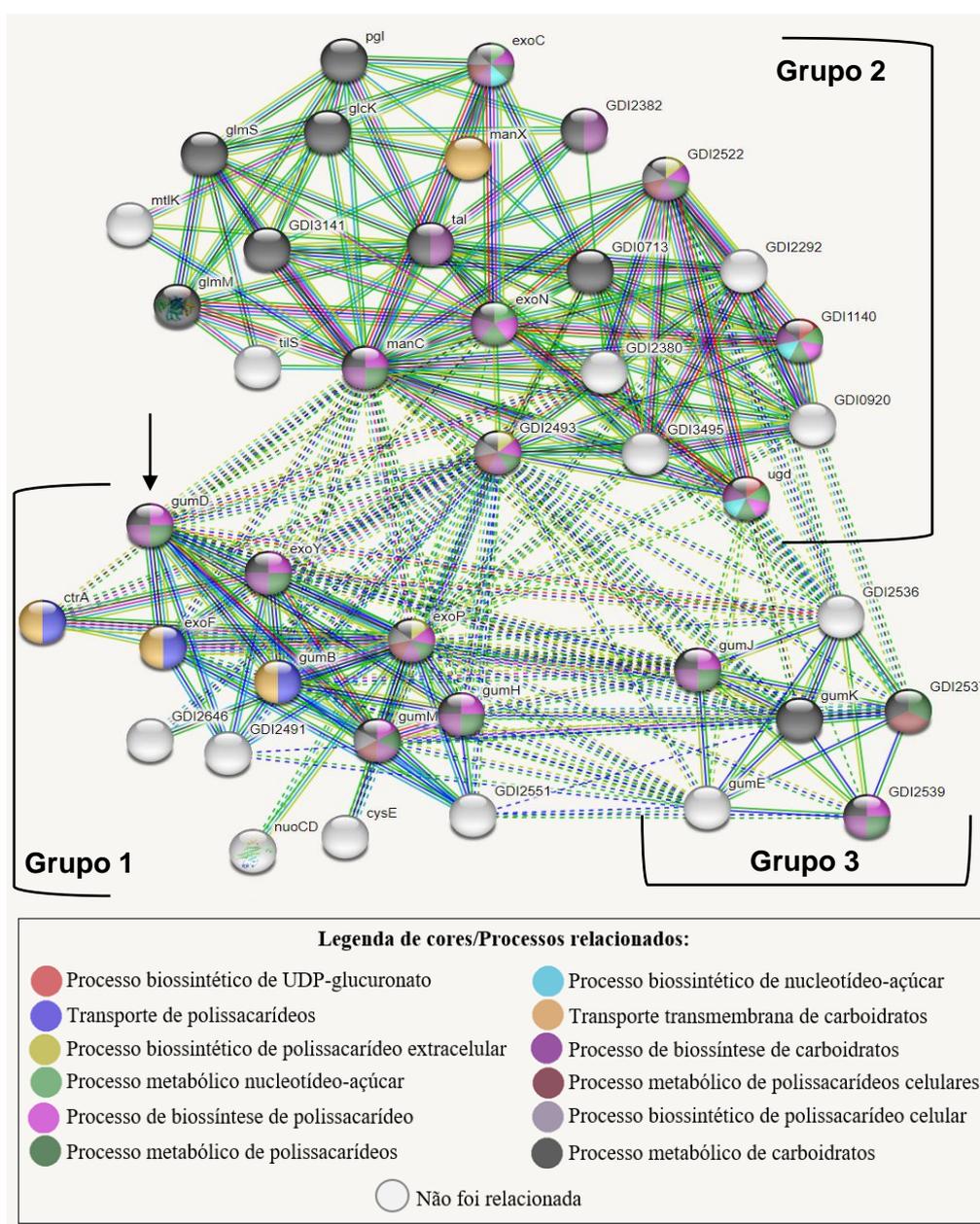
Em uma visão geral, o Grupo 1 e 3 estão principalmente envolvidos em processos metabólicos e biossintéticos de carboidratos. Por outro lado, o Grupo 2 está associado ao transporte de polissacarídeos, processo de biossíntese de polissacarídeos e transporte transmembranar de carboidratos.

As funções moleculares (GO) elencadas foram (processo/taxa falsa de descoberta): Atividade de transferase intramolecular, fosfotransferases (0,0135); Atividade do transportador transmembranar de macromoléculas (0,0478); Atividade do transportador transmembrana de carboidratos (0,0150); Atividade de isomerase (0,0135); Atividade de transferase, transferindo grupos glicosil (0,0014); e Atividade da transferase (0,00036) (Figura 19).

Os Clusters de rede local apontou (processo/taxa falsa de descoberta): Misto, incluindo Domínio de ligação a CoA e glicosil transferase WecB/TagA/CpsF ( $5.94e-08$ ); Misto, incluindo atividade da celulase e ligação à superfície da superfície da célula hospedeira ( $1.51e-06$ ); Biossíntese de açúcar nucleotídeo O-Antígeno e Glicosiltransferase como família 2 ( $6.15e-10$ ); Misto, incluindo processo metabólico de polissacarídeos e Glicosiltransferase como família 2 ( $3.51e-15$ ); Misto, incluindo Glicosiltransferase como família 2 e atividade de celulase ( $4.63e-07$ ). Essas associações podem ser visualizadas na Figura 20, onde o Grupo 1 está mais

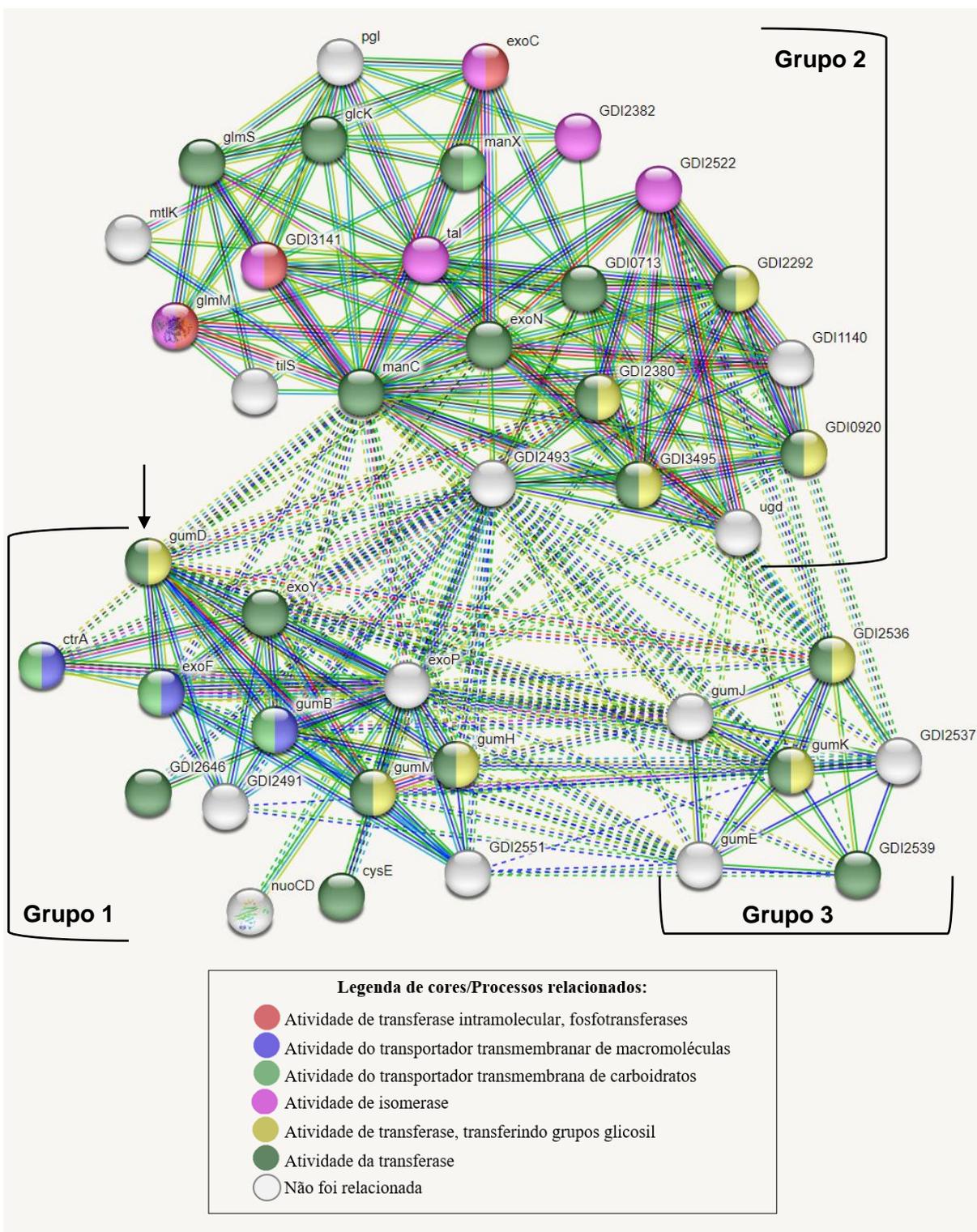
envolvido com "Misto, incluindo Domínio de ligação a CoA e glicosil transferase WecB/TagA/CpsF" e "Misto, incluindo processo metabólico de polissacarídeos e Glicosiltransferase como família 2". O Grupo 2 está associado à "Biossíntese de açúcar nucleotídeo O-Antígeno e Glicosiltransferase como família 2". O Grupo 3 está relacionado a "Misto, incluindo atividade da celulase e ligação à superfície da célula hospedeira", "Misto, incluindo processo metabólico de polissacarídeos e Glicosiltransferase como família 2" e "Misto, incluindo Glicosiltransferase como família 2 e atividade de celulase".

**Figura 18 – Clusters que evidenciam os processos biológicos.**



**Fonte:** Elaborada pela autora, segundo os dados do STRING (2023).  
A seta evidencia a proteína GumD.

**Figura 19 – Clusters que evidenciam as funções moleculares.**

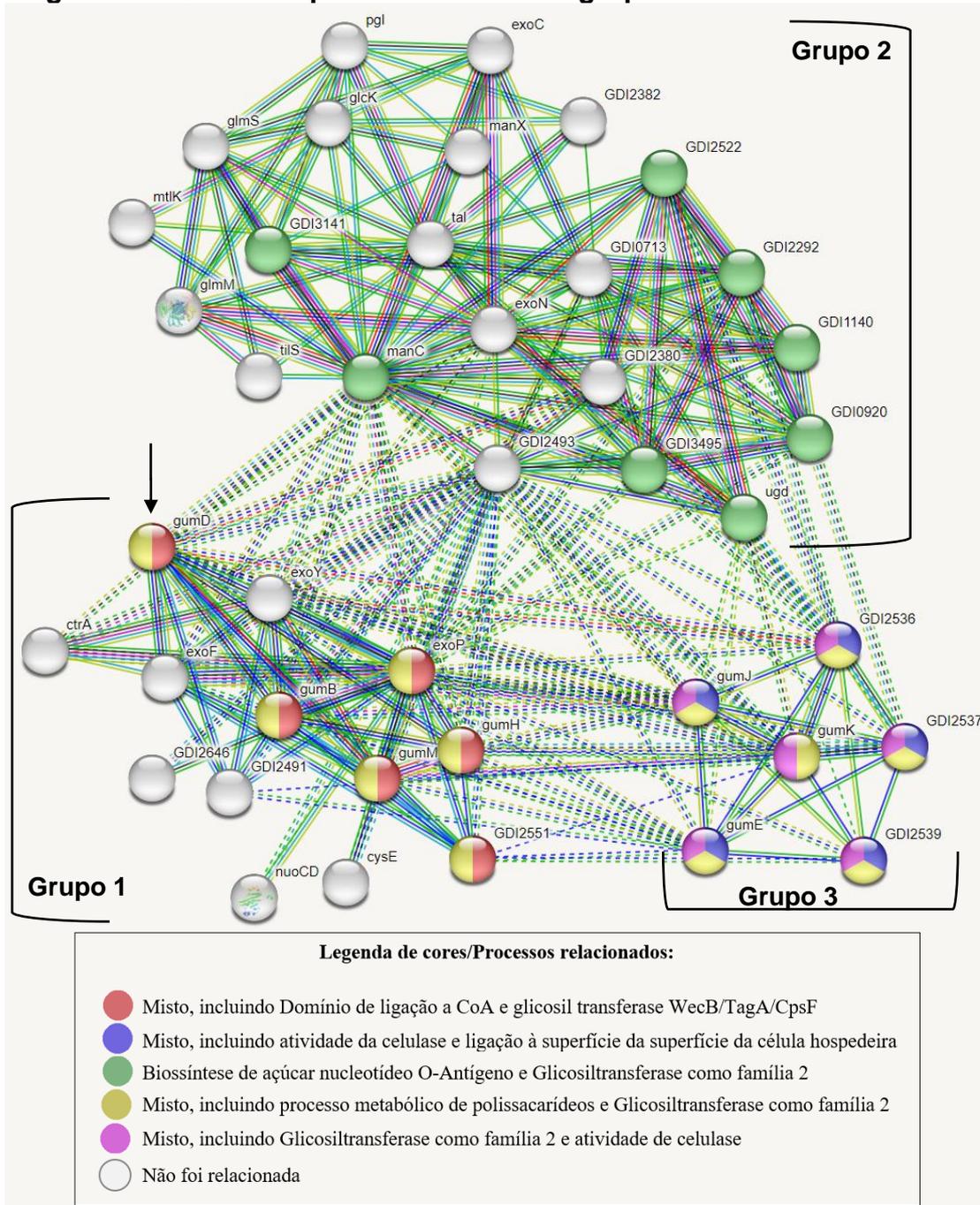


**Fonte:** Elaborada pela autora, segundo os dados do STRING (2023).  
A seta evidencia a proteína GumD.

As vias metabólicas e de sinalização presentes no banco de dados KEGG, conhecidas como vias KeGG, resultaram nas seguintes associações (taxa de descoberta falsa): Metabolismo do Ascorbato e do Aldarato (0.0390); Biossíntese de

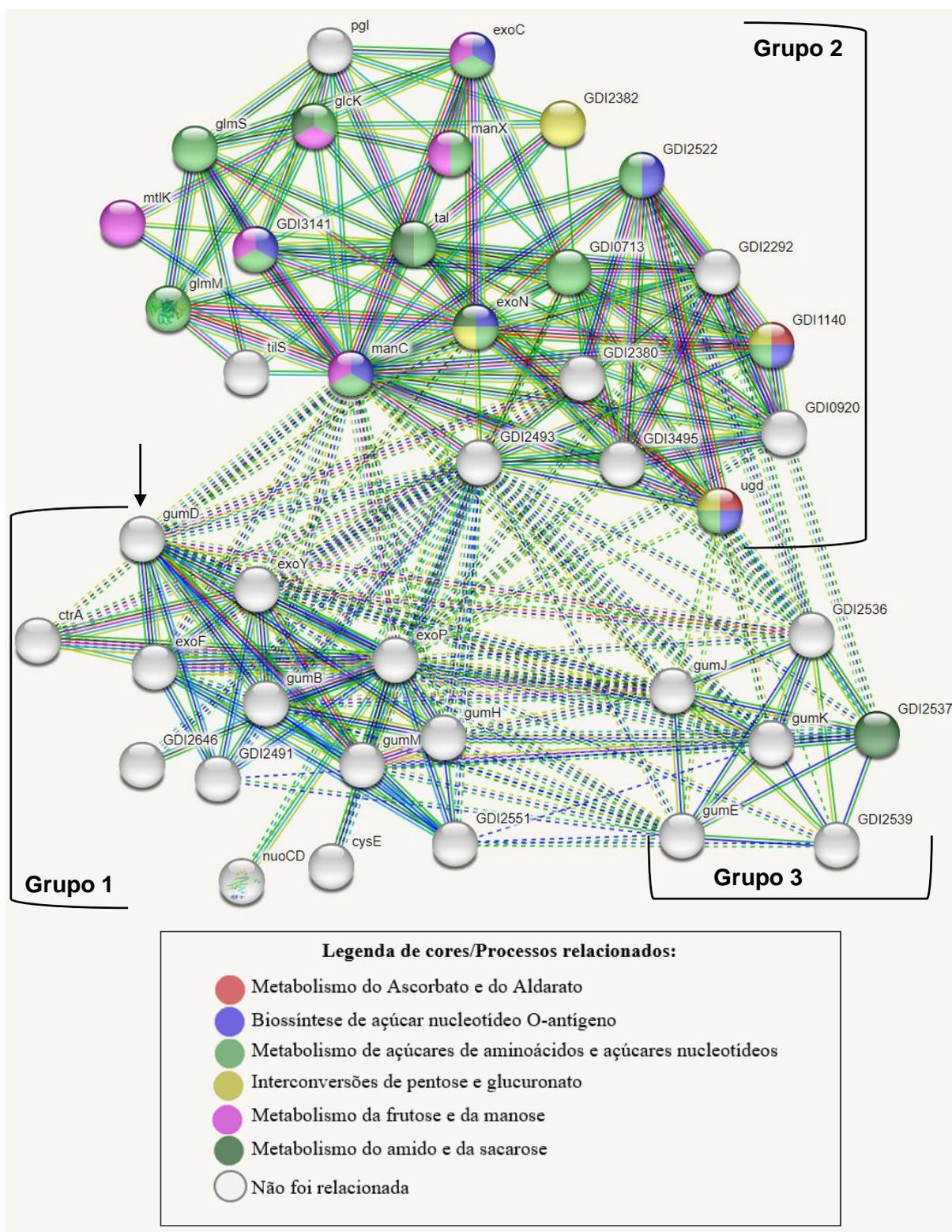
açúcar nucleotídeo O-antígeno (9.49e-07); Metabolismo de açúcares de aminoácidos e açúcares nucleotídeos (1.01e-12); Interconversões de pentose e glucuronato (0.0012); Metabolismo da frutose e da manose; e Metabolismo do amido e da sacarose (0.0074). Conforme ilustrado na Figura 21, o Grupo 1 não apresentou nenhuma relação com esses tópicos das vias KeGG. No Grupo 2, apenas a proteína GDI2537 foi associada ao Metabolismo do amido e da sacarose.

**Figura 20 – Clusters que evidenciam os agrupamentos de redes locais.**



**Fonte:** Elaborada pela autora, segundo os dados do STRING (2023).  
A seta evidencia a proteína GumD.

**Figura 21 – Clusters que evidenciam as vias metabólicas e de sinalização presentes no banco de dados KEGG.**



**Fonte:** Elaborada pela autora, segundo os dados do STRING (2023).  
A seta evidencia a proteína GumD.

As anotações das palavras-chave disponíveis na plataforma Uniprot, desempenhando um papel importante na classificação e descrição das proteínas.

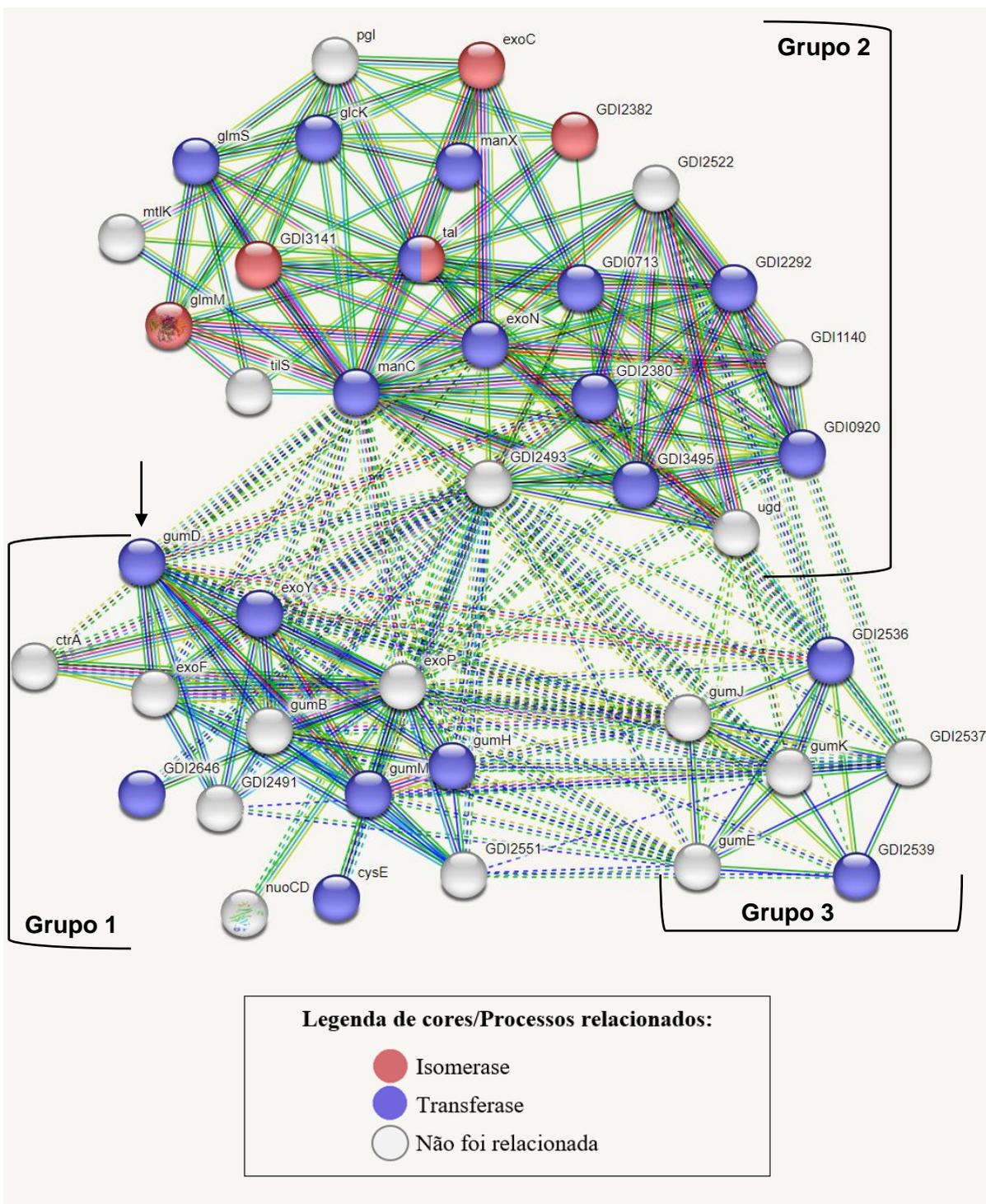
Essas anotações são atribuídas com base em características específicas das proteínas e em informações experimentais disponíveis.

No contexto da análise realizada no STRING, as anotações das palavras-chave (UniProt) revelaram os seguintes resultados, acompanhados dos respectivos valores estatísticos de taxa de descoberta: Isomerase (0.0426) e Transferase (2.39e-06) (Figura 22).

Os domínios das proteínas (Pfam) são componentes estruturais ou funcionais de uma proteína que representam um papel importante em sua atividade biológica. Eles são regiões conservadas que podem ser identificadas através de análises bioinformáticas.

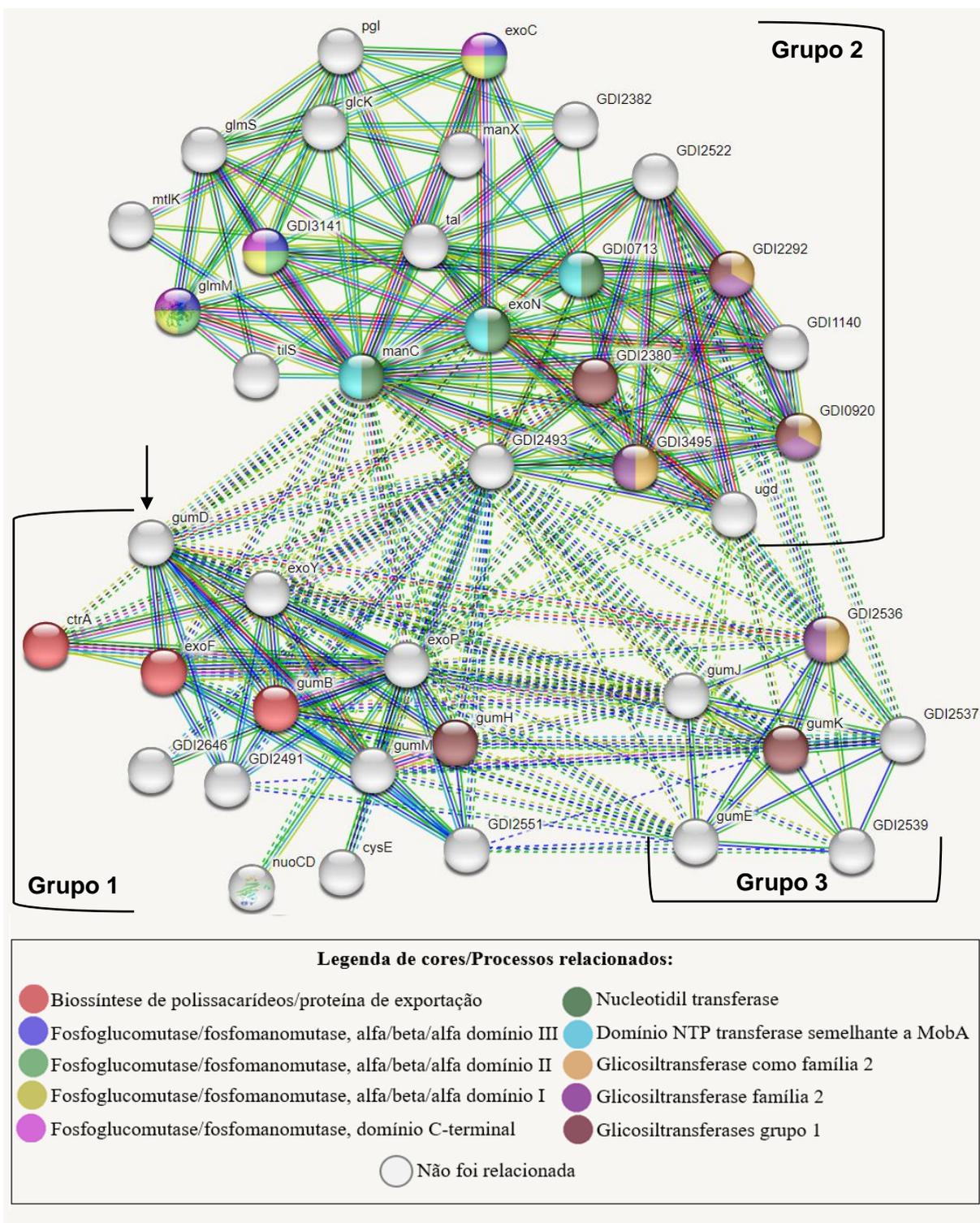
Na análise realizada, foram identificados os seguintes domínios das proteínas, acompanhados dos respectivos valores estatísticos de taxa de descoberta falsa, que indicam o grau de significância do enriquecimento. Quanto mais próximo de zero for a taxa de descoberta falsa, menor é a probabilidade de ser um resultado falso: Biossíntese de polissacarídeos/proteína de exportação (0.0110); Fosfoglucumutase/fosfomanomutase, alfa/beta/alfa domínio III (0.0112); Fosfoglucumutase/fosfomanomutase, alfa/beta/alfa domínio II (0.0112); Fosfoglucumutase/fosfomanomutase, alfa/beta/alfa domínio I (0.0112); Fosfoglucumutase/fosfomanomutase, domínio C-terminal (0.0112); Nucleotidil transferase (0.0125); Domínio NTP transferase semelhante a MobA (0.0331); Glicosiltransferase como família 2 (0.0112); Glicosiltransferase família 2 (0.0112); e Glicosiltransferases grupo 1 (0.0110) (Figura 23).

**Figura 22 – Clusters que evidenciam as Anotações das palavras-chaves (UniProt).**



**Fonte:** Elaborada pela autora, segundo os dados do STRING (2023).  
A seta evidencia a proteína GumD.

**Figura 23 – Clusters que evidenciam os Domínios de proteínas (Pfam).**



**Fonte:** Elaborada pela autora, segundo os dados do STRING (2023).  
A seta evidencia a proteína GumD.

## 5 CONCLUSÃO

A partir das informações obtidas, pode-se inferir que a proteína resultante do gene *gumD* possui uma função de significativa relevância, desempenhando um papel crucial tanto no processo de biossíntese de polissacarídeos quanto no transporte através da membrana celular. Essa dupla função contribui, por sua vez, para a formação do biofilme, reforçando e corroborando com achados previamente estabelecidos em estudos anteriores. A solidez dessa conclusão é fortalecida por meio das análises minuciosas das redes de interações e dos clusters formados, os quais apontam inequivocamente para a associação desta proteína com essas importantes funções biológicas.

Ademais, foi possível identificar evidências da participação do GumD na interação entre a bactéria *G. diazotrophicus* e as plantas. Esse indício se encontra correlacionado com a formação de clusters na rede de interações que estão ligados à atividade celulásica e à aderência à superfície da célula hospedeira vegetal. Com uma compreensão mais aprofundada do papel da proteína GumD na interação planta-bactéria poderia potencialmente abrir caminho para o desenvolvimento de estratégias agrícolas mais eficazes e sustentáveis. Ao decifrar os processos envolvidos nessa interação, pode-se direcionar esforços para otimizar a agricultura, contribuindo para a promoção do crescimento das plantas e aumentando sua resistência, o que, por sua vez, pode ter implicações significativas na segurança alimentar e na sustentabilidade agrícola.

No entanto, é importante ressaltar que as análises *in silico* realizadas fornecem uma visão preliminar e computacionalmente preditiva. Portanto, estudos experimentais adicionais são necessários para validar e confirmar as interações e funções da proteína GumD. A integração de abordagens experimentais e análises bioinformáticas aprofundará nossa compreensão da atuação dessa proteína na relação bactéria-planta e proporcionará uma base mais sólida para aplicações práticas na agricultura.

## REFERÊNCIAS

- AJIJAH, N., FIODOR, A., PANDEY, A. K., RANA, A., & PRANAW, K. Plant Growth-Promoting Bacteria (PGPB) with biofilm-forming ability: A multifaceted agent for sustainable agriculture. **Diversity**, v. 15, n. 1, p. 112, 2023.
- AMARAL, M. B. Inoculação de bactérias solubilizadoras de fosfatos e fixadoras de nitrogênio dos gêneros *Herbaspirillum* e *Azospirillum* em plantas de arroz. 2014.
- ARAÚJO, Ka. S. et al. Avaliação e identificação de *Pseudomonas* sp. e *Bacillus* sp. dois isolados de rizobactérias antagônicas a *Radopholus similis*. **Jornada Científica-Embrapa Mandioca e Fruticultura**, 2010.
- ASHRAF, M. et al. Inoculating wheat seedlings with exopolysaccharide-producing bacteria restricts sodium uptake and stimulates plant growth under salt stress. **Biology and Fertility of soils**, v. 40, p. 157-162, 2004.
- ATES, O. Systems biology of microbial exopolysaccharides production. **Frontiers in bioengineering and biotechnology**, v. 3, p. 200, 2015.
- AZEVEDO, R. A. de. O metabolismo de Lisina e Treonina em arroz e milho. 1999.
- BACON, C.; HINTON, D. Bacterial endophytes: the endophytic niche, its occupants, and its utility. **Plant-associated bacteria**, p. 155-194, 2006.
- BAKER, D.; SALI, A. Protein structure prediction and structural genomics. **Science**, v. 294, n. 5540, p. 93-96, 2001.
- BALDANI, J. I.; BALDANI, V. L. D. History on the biological nitrogen fixation research in graminaceous plants: Special emphasis on the Brazilian experience. **Anais Da Academia Brasileira De Ciências**, v. 77, n. 3, p. 549-579, 2005.
- BANO, A.; FATIMA, M. Salt tolerance in *Zea mays* (L). following inoculation with *Rhizobium* and *Pseudomonas*. **Biology and Fertility of Soils**, v. 45, p. 405-413, 2009.
- BARNAWAL, D. et al. ACC deaminase-containing *Arthrobacter protophormiae* induces NaCl stress tolerance through reduced ACC oxidase activity and ethylene production resulting in improved nodulation and mycorrhization in *Pisum sativum*. **Journal of plant physiology**, v. 171, n. 11, p. 884-894, 2014.
- BERMAN, H. M. et al. The protein data bank. **Nucleic acids research**, v. 28, n. 1, p. 235-242, 2000.
- BERTALAN, M. et al. Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5. **BMC genomics**, v. 10, p. 1-17, 2009.

BISOGNIN, D. A.; AUGUSTIN, Li. IX REUNIÃO TÉCNICA DE PESQUISA E EXTENSÃO DA CULTURA DA BATATA DA REGIÃO SUL. **Santa Maria: Universidade Federal de Santa Maria**, 2006.

BOGINO, P. C. et al. The role of bacterial biofilms and surface components in plant-bacterial associations. **International journal of molecular sciences**, v. 14, n. 8, p. 15838-15859, 2013.

BRAMUSSE, L. B.; BARROSO, L. M.; INTORNE, A. C. Caracterização de bactérias solubilizadoras de fosfato associadas à planta aquática *Typha domingensis*. **Confict**, 2017.

BRUCKNER, F. P. Aspectos da interação entre a proteína TCTP e o potyvírus PepYMV na infecção de tomateiro e *Nicotiana benthamiana*. 2012.

BURROUGHS, A. M. et al. A novel superfamily containing the  $\beta$ -grasp fold involved in binding diverse soluble ligands. **Biology direct**, v. 2, p. 1-11, 2007.

BÜTTNER, D. Behind the lines—actions of bacterial type III effector proteins in plant cells. **FEMS microbiology reviews**, v. 40, n. 6, p. 894-937, 2016.

CALIXTO, P. H. M. Aspectos gerais sobre a modelagem comparativa de proteínas. **Ciência Equatorial**, v. 3, n. 1, 2013.

CARDINALE, M. et al. Paradox of plant growth promotion potential of rhizobacteria and their actual promotion effect on growth of barley (*Hordeum vulgare* L.) under salt stress. **Microbiological research**, v. 181, p. 22-32, 2015.

CARRONI, M.; SAIBIL, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. **Methods**, v. 95, p. 78-85, 2016.

CASTRO, P. RC; KLUGE, R. A. PERES, L. E. P. Manual de fisiologia vegetal: teoria e prática. 2005.

CAVALCANTE, V. A.; DOBEREINER, J. A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. **Plant and Soil**, Rio de Janeiro, v. 108, n. 1, p. 23-31, 1988.

CHEN, L. et al. Induced maize salt tolerance by rhizosphere inoculation of *Bacillus amyloliquefaciens* SQR9. **Physiologia plantarum**, v. 158, n. 1, p. 34-44, 2016.

CHEN, V. B.; ARENDALL, W. B.; HEADD, J. J.; KEEDY, D. A.; IMMORMINO, R. M.; KAPRAL, G. J.; MURRAY, L. W.; RICHARDSON, J. S.; RICHARDSON, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D: Biological Crystallography**, v. 66, n. 1, p. 12–21, 2010.

CHENGXIN ZHANG, P. L. FREDDOLINO, and YANG ZHANG. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. **Nucleic Acids Research**, 45: W291-299, 2017.

CHIN-A-WOENG, T. F. C. et al. Biocontrol by phenazine-1-carboxamide-producing *Pseudomonas chlororaphis* PCL1391 of tomato root rot caused by *Fusarium oxysporum* f. sp. *radicis-lycopersici*. **Molecular plant-microbe interactions**, v. 11, n. 11, p. 1069-1077, 1998.

CHO, S. H. et al. Four new species of *Chryseobacterium* from the rhizosphere of coastal sand dune plants, *Chryseobacterium elymi* sp. nov., *Chryseobacterium hagamense* sp. nov., *Chryseobacterium lathyri* sp. nov. and *Chryseobacterium rhizosphaerae* sp. nov. **Systematic and applied microbiology**, v. 33, n. 3, p. 122-127, 2010.

CHU, S. et al. Protein based biomaterials for therapeutic and diagnostic applications. **Progress in Biomedical Engineering**, v. 4, n. 1, p. 012003, 2021.

COLLINS, R. F. et al. The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. **Proceedings of the National Academy of Sciences**, v. 104, n. 7, p. 2390-2395, 2007.

COMPANT, S.; CLÉMENT, C.; SESSITSCH, A. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization. **Soil Biology and Biochemistry**, v. 42, n. 5, p. 669-678, 2010.

CONTI, R. **Diversidade e atividade antimicrobiana de microrganismos endofíticos da planta medicinal *Borreria verticillata* (L.) GFW Meyer**. 2007. Dissertação de Mestrado. Universidade Federal de Pernambuco.

CORDERO, I. et al. Inoculation of tomato plants with selected PGPR represents a feasible alternative to chemical fertilization under salt stress. **Journal of Plant Nutrition and Soil Science**, v. 181, n. 5, p. 694-703, 2018.

DANHORN, T.; FUQUA, C. Biofilm formation by plant-associated bacteria. **Annu. Rev. Microbiol.**, v. 61, p. 401-422, 2007.

DAS, S. et al. Genetic regulation, biosynthesis and applications of extracellular polysaccharides of the biofilm matrix of bacteria. **Carbohydrate Polymers**, v. 291, p. 119536, 2022.

DE ANDRADE, R. S. **Análise de bioinformática em redes de interações funcionais e metabólicas de pacientes com fibromatose gengival hereditária**. 2017. Tese de Doutorado. [sn].

DE SOUZA, J. T. et al. Effect of 2, 4-diacetylphloroglucinol on *Pythium*: cellular responses and variation in sensitivity among propagules and species. **Phytopathology**, v. 93, n. 8, p. 966-975, 2003.

DEMASON, D. A. Auxin–cytokinin and auxin–gibberellin interactions during morphogenesis of the compound leaves of pea (*Pisum sativum*). **Planta**, v. 222, p. 151-166, 2005.

DIMKIĆ, I. et al. Plant-associated *Bacillus* and *Pseudomonas* antimicrobial activities in plant disease suppression via biological control mechanisms-A review. **Physiological and Molecular Plant Pathology**, v. 117, p. 101754, 2022.

DÖBEREINER; V.L.D., B.; F., O.; REIS, V. M. Endophytic diazotrophs: the key to BNF in gramineous plants. In: HEGAZI, N. A.; FAYEZ, M.; MONIB, M. (Ed.). **Nitrogen Fixation with Non-Legumes**. Giza, Egypt: Faculty of Agriculture, Cairo University, 1993. p. 395-408.

ELJOUNAIDI, K.; LEE, Seung Kyu; BAE, Hanhong. Bacterial endophytes as potential biocontrol agents of vascular wilt diseases—review and future prospects. **Biological control**, v. 103, p. 62-68, 2016.

EMIDIO, N. B. et al. Proteômica: uma introdução aos métodos e aplicações. **HU Revista**, v. 41, n. 3 e 4, 2015.

ESPINDOLA, F. S. et al. Recursos de bioinformática aplicados às ciências ômicas como genômica, transcriptômica, proteômica, interatômica e metabolômica. **Biosci. j.(Online)**, p. 463-477, 2010.

FILGUEIRAS, L. M. B. et al. Eficiência de *Gluconacetobacter diazotrophicus* no aumento da tolerância de arroz vermelho à deficiência hídrica durante a fase reprodutiva. 2015.

FIRÁKOVÁ, S.; ŠTURDÍKOVÁ, M.; MÚČKOVÁ, M. Bioactive secondary metabolites produced by microorganisms associated with plants. **Biologia**, v. 62, p. 251-257, 2007.

FIRDOUS, J. et al. Endophytic bacteria and their potential application in agriculture: a review. **Indian Journal of Agricultural Research**, v. 53, n. 1, p. 1-7, 2019.

FRANCO, A. V.; LIU, D.; REEVES, P. R. The wzz (cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. **Journal of bacteriology**, v. 180, n. 10, p. 2670-2675, 1998.

GAO, H. et al. Nicking mechanism underlying the DNA phosphorothioate-sensing antiphage defense by SspE. **Nature Communications**, v. 13, n. 1, p. 6773, 2022.

GINALSKI, K. Comparative modeling for protein structure prediction. **Current opinion in structural biology**, v. 16, n. 2, p. 172-177, 2006.

GLICK, B. R. (2012). Plant growth-promoting bacteria: mechanisms and applications. **Scientifica**, 1–15. DOI: 10.6064/2012/963401.

GOUVEIA, M. J. et al. Isolamento e avaliação qualitativa de bactérias endofíticas e epífíticas quanto à habilidade de utilizar ácido tânico. **Brazilian Journal of Development**, v. 6, n. 12, p. 95524-95533, 2020.

GRILLO-PUERTAS, M. et al. Intracellular polyphosphate levels in *Gluconacetobacter diazotrophicus* affect tolerance to abiotic stressors and biofilm formation. **Microbes and environments**, v. 33, n. 4, p. 440-445, 2018.

GUEVARA-AVENDAÑO, E. et al. Diffusible and volatile organic compounds produced by avocado rhizobacteria exhibit antifungal effects against *Fusarium kuroshium*. **Brazilian Journal of Microbiology**, v. 51, p. 861-873, 2020.

GUPTA, S.; PANDEY, S. ACC deaminase producing bacteria with multifarious plant growth promoting traits alleviates salinity stress in French bean (*Phaseolus vulgaris*) plants. **Frontiers in Microbiology**, v. 10, p. 1506, 2019.

HALLMANN, J. et al. Bacterial endophytes in agricultural crops. **Canadian journal of microbiology**, v. 43, n. 10, p. 895-914, 1997.

HALO, B. A. et al. Endophytic bacteria (*Sphingomonas* sp. LK11) and gibberellin can improve *Solanum lycopersicum* growth and oxidative stress under salinity. **Journal of Plant Interactions**, v. 10, n. 1, p. 117-125, 2015.

HARDOIM, P. R.; VAN OVERBEEK, L. S.; VAN ELSAS, J. D. Properties of bacterial endophytes and their proposed role in plant growth. **Trends in microbiology**, v. 16, n. 10, p. 463-471, 2008.

HASSAN, M. N.; AFGHAN, S.; HAFEEZ, F. Y. Biological control of red rot in sugarcane by native pyoluteorin-producing *Pseudomonas putida* strain NH-50 under field conditions and its potential modes of action. **Pest management science**, v. 67, n. 9, p. 1147-1154, 2011.

HASSUNUMA, R. M. **Um Desafio no Ensino de Bioquímica: o Mapa de Ramachandran pela Aprendizagem Baseada em Jogos**. 2023. Tese de Doutorado. Universidade Paulista.

HUANG, R. et al. Pyrrolnitrin is more essential than phenazines for *Pseudomonas chlororaphis* G05 in its suppression of *Fusarium graminearum*. **Microbiological research**, v. 215, p. 55-64, 2018.

HUNGRIA, M. Inoculação com *Azospirillum brasilense*: inovação em rendimento a baixo custo. 2011.

HUNGRIA, M.; CAMPO, Rubens José; MENDES, I. de C. Fixação biológica do nitrogênio na cultura da soja. **Embrapa Soja-Circular Técnica (INFOTECA-E)**, 2001.

IGARASHI, N. et al. Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. **Journal of molecular evolution**, v. 52, p. 333-341, 2001.

JENSEN, L. J. et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. **Nucleic acids research**, v. 37, n. suppl\_1, p. D412-D416, 2009.

- JHA, Y.; SUBRAMANIAN, R. B. Paddy plants inoculated with PGPR show better growth physiology and nutrient content under saline condition. **Chilean journal of agricultural research**, v. 73, n. 3, p. 213-219, 2013.
- JI, J. et al. Enhancement of growth and salt tolerance of rice seedlings (*Oryza sativa* L.) by regulating ethylene production with a novel halotolerant PGPR strain *Glutamicibacter* sp. YD01 containing ACC deaminase activity. **Acta Physiologiae Plantarum**, v. 42, p. 1-17, 2020.
- JOO, H. S.; DEYRUP, S. T.; SHIM, S. H. Endophyte-produced antimicrobials: a review of potential lead compounds with a focus on quorum-sensing disruptors. **Phytochemistry Reviews**, v. 20, p. 543-568, 2021.
- KANG, S. M. et al. Plant growth-promoting rhizobacteria reduce adverse effects of salinity and osmotic stress by regulating phytohormones and antioxidants in *Cucumis sativus*. **Journal of Plant Interactions**, v. 9, n. 1, p. 673-682, 2014.
- KC, Dukka B. Recent advances in sequence-based protein structure prediction. **Briefings in bioinformatics**, v. 18, n. 6, p. 1021-1032, 2017.
- KEEL, C. et al. Pseudomonads as antagonists of plant pathogens in the rhizosphere: role of the antibiotic 2, 4-diacetylphloroglucinol in the suppression of black root rot of tobacco. **Symbiosis**, 1990.
- KEEL, C. et al. Suppression of root diseases by *Pseudomonas fluorescens* CHA0- importance of the bacterial secondary metabolite 2, 4-diacetylphloroglucinol. **Molecular Plant-Microbe Interactions**, v. 5, n. 1, p. 4-13, 1992.
- KERBAB, S. et al. Mitigation of NaCl stress in wheat by rhizosphere engineering using salt habitat adapted PGPR halotolerant bacteria. **Applied Sciences**, v. 11, n. 3, p. 1034, 2021.
- KORE, V. T.; DEVI, H. L.; KABIR, J. Packaging, storage and value addition of aonla, an underutilized fruit, in India. **Fruits**, v. 68, n. 3, p. 255-266, 2013.
- KUMAR, A. et al. Plant growth-promoting rhizobacteria (PGPR): perspective in agriculture under biotic and abiotic stress. In: **Crop improvement through microbial biotechnology**. Elsevier, 2018. p. 333-342.
- LANTEIGNE, C. et al. Production of DAPG and HCN by *Pseudomonas* sp. LBUM<sup>12000</sup> contributes to the biological control of bacterial canker of tomato. **Phytopathology** 108 102, n. 10, p. 967-973, 2012.
- LIU, H.; CHEN, Q. Computational protein design for given backbone: recent progresses in general method-related aspects. **Current opinion in structural biology**, v. 39, p. 89-95, 2016.
- LIU, H. et al. Inner plant values: diversity, colonization and benefits from endophytic bacteria. **Frontiers in microbiology**, v. 8, p. 2552, 2017.

LU, S. et al. CDD/SPARCLE: the conserved domain database in 2020. **Nucleic acids research**, v. 48, n. D1, p. D265-D268, 2020.

MALSCHITZKY, S. A. et al. Teor de prolina em acessos de *Saccharum* spp sob estresse hídrico. 2021.

MARIN, V. A. et al. Fixação biológica de nitrogênio: Bactérias fixadoras de nitrogênio de importância para a agricultura tropical. 1999.

MAYAK, S.; TIROSH, T.; GLICK, B. R. Plant growth-promoting bacteria confer resistance in tomato plants to salt stress. **Plant physiology and Biochemistry**, v. 42, n. 6, p. 565-572, 2004.

MAZURIER, S. et al. Phenazine antibiotics produced by fluorescent pseudomonads contribute to natural soil suppressiveness to *Fusarium wilt*. **The ISME journal**, v. 3, n. 8, p. 977-991, 2009.

MEDEIROS, M. de A. **Análise estrutural e eletrônica de pequenos peptídeos de Alanina: um estudo de métodos in silico**. 2022. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Norte.

MENESES, C. H. S. G.; ROUWS, L. F. M.; ARAÚJO, J. L. S.; VIDAL, M. S.; BALDANI, J. I. Exopolysaccharide Production Is Required for Biofilm Formation and Plant Colonization by the Nitrogen-Fixing Endophyte *Gluconacetobacter diazotrophicus*. **Molecular Plant-Microbe Interactions**. Rio de Janeiro, v. 24, n. 12, p. 1448–1458, 2011.

MIGUEL, P. S. B.; MIGUEL, F. B.; MOREIRA, B. C.; OLIVEIRA, M. N. V.; DELVAUX, J. C.; FREITAS, F. S.; COSTA, M. D. Diversity of the endophytic filamentous fungal leaf community at different development stages of eucalyptus. *Journal of Forestry Research*, v. 30, n. 3, p. 1093-1103, 2019.

MIGUEL, P. S. B. et al. Bactérias endofíticas: Colonização, benefícios e identificação. **Brazilian Journal of Development**, v. 7, n. 1, p. 8777-8791, 2021.

MIGUEL, P. S. B. et al. Diversity and distribution of the endophytic bacterial community at different stages of Eucalyptus growth. **Antonie Van Leeuwenhoek**, v. 109, p. 755-771, 2016.

MIHALACHE, G. et al. Lipopeptides produced by *Bacillus subtilis* as new biocontrol products against fusariosis in ornamental plants. **Environmental Science and Pollution Research**, v. 25, p. 29784-29793, 2018.

MUTHUKUMARASAMY, R.; KANG, U. G.; PARK, K. D.; JEON, W. T.; PARK, C. Y.; CHO, Y. S.; KWON, S. W.; SONG, J.; ROH, D. H.; REVATHI, G. Enumeration, isolation and identification of diazotrophs from Korean wetland rice varieties grown with long-term application of N and compost and their short-term inoculation effect on rice plants. **Journal of Applied Microbiology**, v. 102, n. 4, p. 981-991, 2007.

- MUTHUKUMARASAMY, R.; REVATHI, G.; SESHADRI, S.; LAKSHMINARASIMHAN, C. *Gluconacetobacter diazotrophicus* (syn. *Acetobacter diazotrophicus*), a promising diazotrophic endophyte in tropics. **Current Science**, India, v. 83, n. 2, p. 137-145, 2002.
- NADEEM, S. M. et al. Relationship between in vitro characterization and comparative efficacy of plant growth-promoting rhizobacteria for improving cucumber salt tolerance. **Archives of microbiology**, v. 198, p. 379-387, 2016.
- NADEEM, S. M. et al. Rhizobacteria containing ACC-deaminase confer salt tolerance in maize grown on salt-affected fields. **Canadian journal of microbiology**, v. 55, n. 11, p. 1302-1309, 2009.
- NAWAZ, A. et al. Potential of salt tolerant PGPR in growth and yield augmentation of wheat (*Triticum aestivum* L.) under saline conditions. **Frontiers in Microbiology**, v. 11, p. 2019, 2020.
- NESPER, J. et al. Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30: structural and functional analysis of the outer membrane lipoprotein Wza. **Journal of Biological Chemistry**, v. 278, n. 50, p. 49763-49772, 2003.
- NGUYEN, P. T., NGUYEN, T. T., BUI, D. C., HONG, P. T., HOANG, Q. K., & NGUYEN, H. T. Exopolysaccharide production by lactic acid bacteria: The manipulation of environmental stresses for industrial applications. **AIMS microbiology**, v. 6, n. 4, p. 451, 2020.
- NIAZI, P. et al. Review on Surface Elements and Bacterial Biofilms in Plant-Bacterial Associations. **Journal for Research in Applied Sciences and Biotechnology**, v. 2, n. 1, p. 204-214, 2023.
- NOVO, L. AB et al. Plant growth-promoting rhizobacteria-assisted phytoremediation of mine soils. In: **Bio-geotechnologies for mine site rehabilitation**. Elsevier, 2018. p. 281-295.
- PADDA, K. P.; PURI, A.; CHANWAY, C. Endophytic nitrogen fixation—a possible 'hidden' source of nitrogen for lodgepole pine trees growing at unreclaimed gravel mining sites. **FEMS microbiology ecology**, v. 95, n. 11, p. f172, 2019.
- PATEL, B.; SINGH, V.; PATEL, D. Structural bioinformatics. **Essentials of Bioinformatics, Volume I: Understanding Bioinformatics: Genes to Proteins**, p. 169-199, 2019.
- PATIÑO, P. J.; PINEDA, J. R. R. El dogma central de la Biología molecular. **Fondo Editorial Biogénesis**, p. 107-118, 2006.
- PATTEN, C. L.; GLICK, B R. Bacterial biosynthesis of indole-3-acetic acid. **Canadian journal of microbiology**, v. 42, n. 3, p. 207-220, 1996.

PETREY, D.; HONIG, B. Structural bioinformatics of the interactome. **Annual review of biophysics**, v. 43, p. 193-210, 2014.

PETTERSEN, E. F.; GODDARD, T. D.; HUANG, C. C.; COUCH, G. S.; GREENBLATT, D. M.; MENG, E. C.; FERRIN, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. **J Comput Chem**, v. 25, p. 1605–1612, 2004.

PINILLA, J. M. S. **Importancia de la biosíntesis bacteriana de leucina para el establecimiento de la simbiosis rhizobium-leguminosa**. 2001. Tese de Doutorado. Universidad de Granada.

POLONI, J. de F. et al. Biología de sistemas. **Bioinformática: da biologia à flexibilidade molecular. Cap. 6, p. 116-146**, 2014.

POURBABAEE, A. A. et al. Promotion of wheat growth under salt stress by halotolerant bacteria containing ACC deaminase. 2016.

RADOVANOVIĆ, N. et al. Biocontrol and plant stimulating potential of novel strain *Bacillus* sp. PPM3 isolated from marine sediment. **Microbial pathogenesis**, v. 120, p. 71-78, 2018.

RAFI, M. M.; KRISHNAVENI, M. S.; CHARYULU, P. B. B. N. Phosphate-solubilizing microorganisms and their emerging role in sustainable agriculture. **Recent developments in applied microbiology and biochemistry**, p. 223-233, 2019.

RAI, R. et al. Isolation, characterization and evaluation of the biocontrol potential of *Pseudomonas protegens* RS-9 against *Ralstonia solanacearum* in Tomato. 2017.

RAMACHANDRAN, GN t; SASISEKHARAN, V. Conformation of polypeptides and proteins. **Advances in protein chemistry**, v. 23, p. 283-437, 1968.

RAMETTE, A. et al. *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol compounds 2, 4-diacetylphloroglucinol and pyoluteorin. **Systematic and applied microbiology**, v. 34, n. 3, p. 180-188, 2011.

RAO, V. S. et al. Protein-protein interaction detection: methods and analysis. **International journal of proteomics**, v. 2014, 2014.

REINHOLD-HUREK, B.; HUREK, T. Living inside plants: bacterial endophytes. **Current opinion in plant biology**, v. 14, n. 4, p. 435-443, 2011.

REIS, V. M.; OLIVARES, F. L.; DOBEREINER, J. Improved methodology for isolation of *Acetobacter diazotrophicus* and confirmation of its endophytic habitat. **World Journal of Microbiology & Biotechnology**, v. 10, n. 4, p. 401-405, 1994.

RICHARDSON, J. S.; RICHARDSON, D. C. Amino acid preferences for specific locations at the ends of  $\alpha$  helices. **Science**, v. 240, n. 4859, p. 1648-1652, 1988.

ROJAS-TAPIAS, D. et al. Effect of inoculation with plant growth-promoting bacteria (PGPB) on amelioration of saline stress in maize (*Zea mays*). **Applied Soil Ecology**, v. 61, p. 264-272, 2012.

ROSENBLUETH, M.; MARTÍNEZ-ROMERO, E. Bacterial endophytes and their interactions with hosts. **Molecular plant-microbe interactions**, v. 19, n. 8, p. 827-837, 2006.

RYAN, R. P.; GERMAINE, K.; FRANKS, A.; RYAN, D. J & DOWLING, D. N. Bacterial endophytes: recent developments and applications. **FEMS Microbiology Letters**, v. 278, n. 1, p. 1–9, 2008.

SAEED, Q., XIUKANG, W., HAIDER, F. U., KUČERIK, J., MUMTAZ, M. Z., HOLATKO, J., & MUSTAFA, A. Rhizosphere bacteria in plant growth promotion, biocontrol, and bioremediation of contaminated sites: A comprehensive review of effects and mechanisms. **International Journal of Molecular Sciences**, v. 22, n. 19, p. 10529, 2021.

ŠALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **Journal of molecular biology**, v. 234, n. 3, p. 779-815, 1993.

SANTOS DIAS, A.; SANTOS, C. C. Bactérias promotoras de crescimento de plantas: conceitos e potencial de uso. Nova Xavantina-MT: Pantanal, 2022. E-book: 100p.

SANTOS FILHO, O. A.; ALENCASTRO, R. B. de. Modelagem de proteínas por homologia. **Química Nova**, v. 26, p. 253-259, 2003.

SANTOS, D. R. et al. Isolamento e seleção de bactérias antagonistas a fitopatógenos e detecção de genes associados à produção de compostos bioativos. 2014.

SANTOS, L. M. L. **Análise proteômica da bactéria *Gluconacetobacter diazotrophicus*: estabelecimento de um banco proteômico, análise diferencial e implicações na anotação do genoma**. 2007. 184f. Dissertação (Mestrado em Ciências Biológicas – Biofísica) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.

SAPRE, S.; GONTIA-MISHRA, I.; TIWARI, S. *Klebsiella* sp. confers enhanced tolerance to salinity and plant growth promotion in oat seedlings (*Avena sativa*). **Microbiological Research**, v. 206, p. 25-32, 2018.

SARAF, M.; PANDYA, U.; THAKKAR, A. Role of allelochemicals in plant growth promoting rhizobacteria for biocontrol of phytopathogens. **Microbiological research**, v. 169, n. 1, p. 18-29, 2014.

SCAVINO, Ana Fernández; PEDRAZA, Raúl O. The role of siderophores in plant growth-promoting bacteria. In: **Bacteria in agrobiolgy: crop productivity**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 265-285.

SCHULZ, B.; BOYLE, C. The endophytic continuum. *Mycoll Res* 109 (6): 661–686. 2005.

SCHWEDE, T. et al. SWISS-MODEL: an automated protein homology-modeling server. **Nucleic acids research**, v. 31, n. 13, p. 3381-3385, 2003.

SCHWEDE, T. Protein modeling: what happened to the “protein structure gap”? **Structure**, v. 21, n. 9, p. 1531-1540, 2013.

SENGUPTA, S., & DEY, S. Microbial exo-polysaccharides (EPS): role in agriculture and environment. **Agric Food**, v. 1, n. 12, p. 4-8, 2019.

SERRATO, R. V. et al. Structural studies of an exopolysaccharide produced by *Gluconacetobacter diazotrophicus* Pal5. **Carbohydrate polymers**, v. 98, n. 1, p. 1153-1159, 2013.

SESSITSCH, A.; REITER, B.; BERG, G. Endophytic bacterial communities of field-grown potato plants and their plant-growth-promoting and antagonistic abilities. **Canadian journal of microbiology**, v. 50, n. 4, p. 239-249, 2004.

SILVA, H. S. A. et al. Microrganismos endofíticos: potencial de uso como agentes de biocontrole da ferrugem do cafeeiro. 2006.

SILVA, L. X.; BASTOS, L. L.; SANTOS, L. H. Modelagem computacional de proteínas. 2021.

SILVA, R. et al. *Gluconacetobacter diazotrophicus* changes the molecular mechanisms of root development in *Oryza sativa* L. growing under water stress. **International Journal of Molecular Sciences**, v. 21, n. 1, p. 333, 2020.

SIMÕES, S. N. **Uma abordagem de integração de dados de redes PPI e expressão gênica para priorizar genes relacionados a doenças complexas**. 2015. Tese de Doutorado. Universidade de São Paulo.

SINGH, M. et al. Endophytic bacteria: a new source of bioactive compounds. **3 Biotech**, v. 7, p. 1-14, 2017.

SINGH, R. P.; JHA, P. N. The PGPR *Stenotrophomonas maltophilia* SBP-9 augments resistance against biotic and abiotic stress in wheat plants. **Frontiers in microbiology**, v. 8, p. 1945, 2017.

SOARES, A. N. G. **Interatômica da cavidade oral**. 2014. Tese de Doutorado.

SOLANO, B. R.; MAICAS, J. B.; MAÑERO, FJ Gutiérrez. Physiological and molecular mechanisms of plant growth promoting rhizobacteria (PGPR). **Plant-bacteria interactions: Strategies and techniques to promote plant growth**. Weinheim: Wiley, p. 41-52, 2008.

SOMERS, E.; VANDERLEYDEN, J.; SRINIVASAN, M. Rhizosphere bacterial signalling: a love parade beneath our feet. **Critical reviews in microbiology**, v. 30, n. 4, p. 205-240, 2004.

SPRENT, J. I.; DE FARIA, S. M. Mechanisms of infection of plants by nitrogen fixing organisms. **Plant and Soil**, v. 110, p. 157-165, 1988.

STEPHAN, M. P.; OLIVEIRA, M.; TEIXEIRA, K. R. S.; MARTINEZ-DRETS, G.; DOBEREINER, J. Physiology and nitrogen fixation of *Acetobacter diazotrophicus*. **FEMS Microbiology Letters**, Rio de Janeiro, v. 77, n. 1, p. 67-72, 1991.

STEVENSON, G. et al. Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. **Journal of bacteriology**, v. 178, n. 16, p. 4885-4893, 1996.

STUDER, G. et al. Modeling of protein tertiary and quaternary structures based on evolutionary information. **Computational Methods in Protein Evolution**, p. 301-316, 2019.

SUAREZ, C. et al. Plant growth-promoting effects of *Hartmannibacter diazotrophicus* on summer barley (*Hordeum vulgare* L.) under salt stress. **Applied Soil Ecology**, v. 95, p. 23-30, 2015.

SUKWEENADHI, J. et al. *Paenibacillus yonginensis* DCY84T induces changes in Arabidopsis thaliana gene expression against aluminum, drought, and salt stress. **Microbiological research**, v. 172, p. 7-15, 2015.

SULTANA, S. et al. Isolation and identification of salt-tolerant plant-growth-promoting rhizobacteria and their application for rice cultivation under salt stress. **Canadian journal of microbiology**, v. 66, n. 2, p. 144-160, 2020.

SZKLARCZYK, D. et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. **Nucleic acids research**, v. 43, n. D1, p. D447-D452, 2015.

SZKLARCZYK, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. **Nucleic acids research**, v. 47, n. D1, p. D607-D613, 2019.

SZKLARCZYK, D. et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. **Nucleic acids research**, p. gkw937, 2016.

THOMASHOW, L. S. et al. Production of the antibiotic phenazine-1-carboxylic acid by fluorescent *Pseudomonas* species in the rhizosphere of wheat. **Applied and Environmental Microbiology**, v. 56, n. 4, p. 908-912, 1990.

TING, D. et al. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. **PLoS computational biology**, v. 6, n. 4, p. e1000763, 2010.

TRIVEDI, P.; LEACH, J. E.; TRINGE, S. G.; SA, T., & SINGH, B. K. Plant–microbiome interactions: from community assembly to plant health. **Nature reviews microbiology**, v. 18, n. 11, p. 607-621, 2020.

UNIPROT: the universal protein knowledgebase in 2023. **Nucleic Acids Research**, v. 51, n. D1, p. D523-D531, 2023.

UPADHYAY, A.; SRIVASTAVA, S. Phenazine-1-carboxylic acid is a more important contributor to biocontrol *Fusarium oxysporum* than pyrrolnitrin in *Pseudomonas fluorescens* strain Psd. **Microbiological research**, v. 166, n. 4, p. 323-335, 2011.

VELMOUROUGANE, K.; PRASANNA, R.; SAXENA, A. K. Agriculturally important microbial biofilms: present status and future prospects. **Journal of basic microbiology**, v. 57, n. 7, p. 548-573, 2017.

VERLI, H. Bioinformática: da biologia à flexibilidade molecular. 2014.

WALPORT, L. J. et al. The characterization of protein interactions—what, how and how much?. **Chemical Society Reviews**, v. 50, n. 22, p. 12292-12307, 2021.

WANDERLEY, K. M. et al. Regulação da biossíntese de exopolissacarídeos de *Gluconacetobacter diazotrophicus* por fosforilação de uma Tirosina quinase. 2019.

WANDERLEY, K. M. et al. Tyrosine Kinase Self-Phosphorylation Controls Exopolysaccharide Biosynthesis in *Gluconacetobacter diazotrophicus* Strain Pal5. **Life**, v. 11, n. 11, p. 1231, 2021.

WANG, Q. et al. Rhizosphere bacteria containing 1-aminocyclopropane-1-carboxylate deaminase increase growth and photosynthesis of pea plants under salt stress by limiting Na<sup>+</sup> accumulation. **Functional Plant Biology**, v. 43, n. 2, p. 161-172, 2016.

WOLYNES, P. G. Evolution, energy landscapes and the paradoxes of protein folding. **Biochimie**, v. 119, p. 218-230, 2015.

XIE, H. et al. Implications of endophytic microbiota in *Camellia sinensis*: a review on current understanding and future insights. **Bioengineered**, v. 11, n. 1, p. 1001-1015, 2020.

XU, D.; ZHANG, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. **Biophysical Journal**, v. 101, n. 10, p. 2525–2534, 2011.

XU, Z. et al. Contribution of *bacillomycin D* in *Bacillus amyloliquefaciens* SQR9 to antifungal activity and biofilm formation. **Applied and environmental microbiology**, v. 79, n. 3, p. 808-815, 2013.

YANG Z. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**, 9:40, 2008.

YANG, J.; ZHANG, Y. I-TASSER server: New development for protein structure and function predictions. **Nucleic Acids Research**, v. 43, n. W1, p. W174–W181, 2015.

YAO, L. et al. Growth promotion and protection against salt stress by *Pseudomonas putida* Rs-198 on cotton. **European Journal of Soil Biology**, v. 46, n. 1, p. 49-54, 2010.

YOSHIDA, Y. et al. Identification of a genetic locus essential for serotype b-specific antigen synthesis in *Actinobacillus actinomycetemcomitans*. **Infection and immunity**, v. 66, n. 1, p. 107-114, 1998.

ZHANG, H. W.; SONG, Y. C.; TAN, R. X. Biology and chemistry of endophytes. **Natural product reports**, v. 23, n. 5, p. 753-771, 2006.

ZHANG, Y. I-TASSER: Fully automated protein structure prediction in CASP8. **Proteins: Structure, Function, and Bioinformatics**, v. 77, n. S9, p. 100-113, 2009.

ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. **Proteins: Structure, Function, and Bioinformatics**, v. 57, n. 4, p. 702-710, 2004.

ZHENG, W. et al. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. **Cell reports methods**, v. 1, n. 3, 2021.

ZHOU, X. et al. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. **Nature Protocols**, v. 17, n. 10, p. 2326-2353, 2022.