

UNIVERSIDADE DE RIBEIRÃO PRETO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

THIAGO MATEUS ROSA DOS SANTOS

MODELAGEM COMPUTACIONAL E DOCKING MOLECULAR DE  
ENZIMAS SECRETADAS POR *Xanthomonas albilineans*

RIBEIRÃO PRETO  
2024

THIAGO MATEUS ROSA DOS SANTOS

MODELAGEM COMPUTACIONAL E DOCKING MOLECULAR DE  
ENZIMAS SECRETADAS POR *Xanthomonas albilineans*

Tese apresentada à Universidade de Ribeirão  
Preto - UNAERP, como requisito para a obtenção  
do título de Doutor em Biotecnologia.

Orientador: Prof. Dr. Mozart Marins

RIBEIRÃO PRETO  
2024

Ficha catalográfica preparada pelo Centro de Processamento  
Técnico da Biblioteca Central da UNAERP

- Universidade de Ribeirão Preto -

ROSA-SANTOS, Thiago Mateus, 1985-

S237m            Modelagem computacional e docking molecular de enzimas  
secretadas por *Xanthomonas albilineans* / Thiago Mateus Rosa dos  
Santos. - Ribeirão Preto, 2024.  
177 f. : il. color.

Orientador: Prof.º Dr.º Mozart Marins.

Tese (Doutorado) - Universidade de Ribeirão Preto, UNAERP,  
Biotecnologia, 2024.

1. Bioinformática. 2. CAZymes. 3. Celulose. 4. Hidrólise. 5.  
Modelagem molecular. II. Título.

CDD 660.6

**THIAGO MATEUS ROSA DOS SANTOS**

**MODELAGEM COMPUTACIONAL E DOCKING MOLECULAR DE ENZIMAS  
SECRETADAS POR *Xanthomonas albilineans***

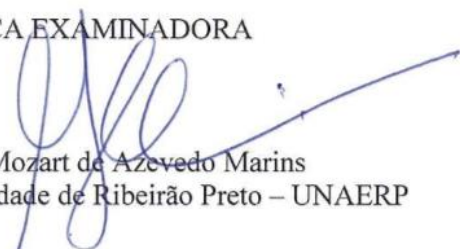
Tese de Doutorado apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade de Ribeirão Preto, para obtenção do título de Doutor em Biotecnologia.

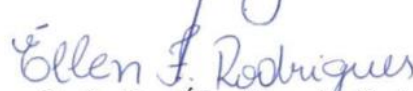
Área de Concentração: Biotecnologia Aplicada à Agroindústria e ao Meio Ambiente

Data da defesa: 04 de junho de 2024

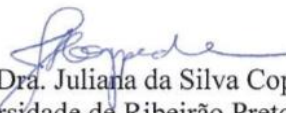
Resultado: Aprovado


**BANCA EXAMINADORA**

  
Prof. Dr. Mozart de Azevedo Marins  
Presidente/Universidade de Ribeirão Preto – UNAERP

  
Prof. Dra. Ellen Francine Rodrigues  
PPGTA - Universidade de Ribeirão Preto – UNAERP

  
Prof. Dr. Guilherme Sgobbi Zagui  
PPGTA - Universidade de Ribeirão Preto – UNAERP

  
Prof. Dra. Juliana da Silva Coppede  
PPGB - Universidade de Ribeirão Preto – UNAERP

  
Prof. Dr. Pablo Rodrigo Sanches  
PPGB - Universidade de Ribeirão Preto – UNAERP

**RIBEIRÃO PRETO  
2024**

## **Dedico**

*Ao meu pai, Deoclides (in memoriam), por ter sido o meu suporte e exemplo, e à minha mãe, Aparecida, por ser o meu espelho de força nos momentos difíceis.*

## **AGRADECIMENTOS**

À minha família - pai, mãe, irmã, cunhado e sobrinha - simplesmente por serem a minha família.

Ao meu orientador, Prof. Dr. Mozart Marins, pela confiança e paciência ao assumir a minha orientação.

À Profa. Dra. Sonia Marli Zingaretti, pela oportunidade oferecida.

Aos amigos, Dr. Tiago Paz, Érica Carvalho, Dra. Marielle Cascaes e Dra. Érica Gelfuso, pelos puxões de orelha e risadas ao longo dos anos. Um agradecimento especial ao Dr. Tiago Paz, cuja colaboração foi fundamental na conclusão desta tese.

Às amigas de pós-graduação, Me. Luane da Guia Vieira e Dra. Mariana Landenberger dos Santos, pelo apoio, amizade e convivência.

Aos amigos de infância, Me. Rafael Silveira e Eder Oliveira, pelo companheirismo ao longo dos anos.

À Mariana Cavalari, pelo apoio constante e acompanhamento durante esses anos.

Ao Programa de Pós-Graduação em Biotecnologia da Universidade de Ribeirão Preto, pela oportunidade de realizar o curso de doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudo.

A todos que colaboraram, direta ou indiretamente, para a realização deste projeto.

## RESUMO

A crescente demanda energética, juntamente com o impacto ambiental provocado pela queima de combustíveis fósseis (petróleo, gás natural e carvão mineral), impõe a necessidade do desenvolvimento de fontes alternativas e renováveis de energia. O etanol celulósico, impulsionado pelo avanço tecnológico, tem recebido crescente relevância como integrante na solução para essa questão energética. Esse tipo de etanol é produzido a partir dos carboidratos complexos (celulose e hemicelulose) extraídos da biomassa lignocelulósica (madeira, gramíneas, resíduos agrícolas, entre outros), uma matéria-prima de baixo custo e abundante no planeta. O processo de obtenção desses carboidratos é dependente das enzimas ativas em carboidratos (CAZymes), que são responsáveis pela biotransformação e degradação dos componentes lignocelulósicos em partes menores. Atualmente, as CAZymes utilizadas durante o processo de produção do etanol celulósico possuem um alto custo e baixa eficiência, sendo necessário a exploração de novas enzimas. Embora diversas bactérias, fungos, leveduras e animais decompositores possuam CAZymes naturais, o número de organismos analisados é limitado. A bactéria *Xanthomonas albilineans*, um fitopatógeno cujo principal hospedeiro é a cana-de-açúcar, uma gramínea reconhecida por sua capacidade de produção de biomassa e energia, desenvolveu ao longo de seu processo evolutivo mecanismos otimizados para invadir e sobreviver no xilema da planta. Um dos principais mecanismos de infecção utilizados por esta bactéria envolve a secreção de enzimas adaptadas para degradar a parede celular vegetal e utilizar os produtos da quebra como fonte de carbono para sua sobrevivência. Essa capacidade em degradar a parede celular vegetal, transforma *Xanthomonas albilineans* em um alvo de estudo que pode resultar na identificação de novas CAZymes com potencial para a produção de biocombustíveis. Em razão dos métodos experimentais para determinar a estrutura de proteínas demandarem um alto recurso financeiro, este trabalho teve como objetivo utilizar ferramentas computacionais para identificar CAZymes secretadas por *Xanthomonas albilineans* que estão envolvidas na degradação da parede celular vegetal, através da hidrólise de celulose. Foram utilizados métodos de bioinformática estrutural para realizar uma triagem do proteoma da bactéria e a predição estrutural de cinco enzimas (XALc\_0484, XALc\_0865, XALc\_0874, XALc\_2967 e XALc\_2969) secretadas pelo patógeno. Estas enzimas foram selecionadas para análise subsequente de interação com os oligossacarídeos de celulose, por intermédio de docking molecular. Entre os principais métodos utilizados para a predição estrutural das enzimas, os métodos que empregam algoritmos de Inteligência Artificial apresentaram os melhores modelos, com destaque para as estruturas obtidas no banco de dados do AlphaFold. A análise estrutural revelou que quatro dessas enzimas possuem domínios estruturais conservados que as classificam entre as famílias de hidrolases glicosídicas, GH5 (XALc\_2967 e XALc\_2969), GH6 (XALc\_0484) e GH12 (XALc\_0865). XALc\_0874 apresentou semelhanças estruturais com alginato liases da família PL6. Com base nos resultados das simulações de docking foi possível inferir que as enzimas XALc\_0484, XALc\_0865, XALc\_2967 e XALc\_2969 provavelmente atuam na degradação da celulose, com preferência por ligantes maiores como os tetrassacarídeos. Além disso, XALc\_0484 obteve os melhores resultados entre todas as enzimas analisadas, justificando sua posição de destaque para a sobrevivência de *X. albilineans* e um potencial alvo para uso industrial. As ferramentas computacionais utilizadas provaram ser essenciais para os estudos de biologia estrutural.

Palavras-chave: bioinformática; CAZymes; celulasas; celulose; hidrólise; modelagem molecular.

## ABSTRACT

The increasing energy demand, along with the environmental impact caused by the burning of fossil fuels (oil, natural gas, and coal), imposes the need for the development of alternative and renewable energy sources. Cellulosic ethanol, driven by technological advancement, has been gaining increasing relevance as part of the solution to this energy issue. This type of ethanol is produced from complex carbohydrates (cellulose and hemicellulose) extracted from lignocellulosic biomass (wood, grasses, agricultural residues, among others), a low-cost and abundant raw material on the planet. The process of obtaining these carbohydrates depends on carbohydrate-active enzymes (CAZymes), which are responsible for the biotransformation and degradation of lignocellulosic components into smaller parts. Currently, the CAZymes used during the cellulosic ethanol production process have high costs and low efficiency, necessitating the exploration of new enzymes. Although various bacteria, fungi, yeasts, and decomposer animals possess natural CAZymes, the number of organisms analyzed is limited. The bacterium *Xanthomonas albilineans*, a phytopathogen whose main host is sugarcane, a grass recognized for its biomass and energy production capacity, has developed optimized mechanisms over its evolutionary process to invade and survive in the plant xylem. One of the main infection mechanisms used by this bacterium involves the secretion of enzymes adapted to degrade the plant cell wall and use the breakdown products as a carbon source for its survival. This ability to degrade the plant cell wall makes *Xanthomonas albilineans* a study target that could result in the identification of new CAZymes with potential for biofuel production. Due to the high financial resources required for experimental methods to determine protein structures, this work aimed to use computational tools to identify CAZymes secreted by *Xanthomonas albilineans* involved in the degradation of the plant cell wall through cellulose hydrolysis. Structural bioinformatics methods were used to screen the bacterium's proteome and structurally predict five enzymes (XALc\_0484, XALc\_0865, XALc\_0874, XALc\_2967, and XALc\_2969) secreted by the pathogen. These enzymes were selected for subsequent analysis of interaction with cellulose oligosaccharides through molecular docking. Among the main methods used for the structural prediction of enzymes, those employing Artificial Intelligence algorithms produced the best models, with particular emphasis on the structures obtained from the AlphaFold database. Structural analysis revealed that four of these enzymes have conserved structural domains that classify them among glycoside hydrolase families, GH5 (XALc\_2967 and XALc\_2969), GH6 (XALc\_0484), and GH12 (XALc\_0865). XALc\_0874 showed structural similarities with alginate lyases of the PL6 family. Based on the results of the docking simulations, it was inferred that the enzymes XALc\_0484, XALc\_0865, XALc\_2967, and XALc\_2969 likely act in cellulose degradation, with a preference for larger ligands such as tetrasaccharides. Furthermore, XALc\_0484 achieved the best results among all analyzed enzymes, justifying its prominent role in the survival of *X. albilineans* and its potential as an industrial target. The computational tools used proved to be essential for structural biology studies.

Keywords: bioinformatics; CAZymes; cellulases; cellulose; hydrolysis; molecular modeling.

## LISTA DE ILUSTRAÇÕES

- Figura 1** - Estrutura principal da parede celular vegetal.....22
- Figura 2** - (a) Composição química da celulose formada por moléculas de D-glicose. Essas moléculas estão unidas por ligações glicosídicas  $\beta$  (1  $\rightarrow$  4) que ocorrem entre os átomos 1 e 4 da D-glicose. (b) Representação da rede estrutural da celulose, formada pela união das moléculas de D-glicose, sendo o componente principal das microfibrilas.....23
- Figura 3** - Ação coordenada das quatro principais celulases envolvidas no processo de hidrólise da celulose. A endo- $\beta$ -1,4-glucanase (EC 3.2.1.4), representada em rosa, atua clivando aleatoriamente a região amorfa interna da celulose. A  $\beta$ -1,4-celobiosidase/celobiohidrolase que degrada a celulose atuando nas extremidades não redutoras (EC 3.2.1.91), é representada em verde, e redutora (EC 3.2.1.176) da celulose, em amarelo. A  $\beta$ -glicosidase (EC 3.2.1.21), retratada em azul, converte a celobiose em glicose.....24
- Figura 4** - Representação esquemática de uma hidrolase glicosídica multidomínio secretada por *Xanthomonas albilineans* GPE PC73. A enzima é composta por um domínio catalítico e um módulo de ligação ao carboidrato 2 (CBM2), unidos pelo domínio linker.....29
- Figura 5** - Metodologia aplicada durante as etapas de triagem do proteoma (azul), predição dos modelos (laranja), avaliação dos modelos (verde), predição dos sítios de ligação (cinza) e simulações de docking molecular (rosa). .....34
- Figura 6** - Diagrama de Venn ilustrando a quantidade de proteínas identificadas na análise do proteoma de *Xanthomonas albilineans* GPE PC73. As proteínas são categorizadas com base na presença de peptídeo sinal (roxo), localização extracelular (azul) e a possibilidade de serem CAZymes (laranja). .....39
- Figura 7** - Composição multidomínio de cada enzima selecionada (XALc\_0484, XALc\_0865, XALc\_0874, XALc\_2967 e XALc\_2969). Os comprimentos são expressos em número de resíduos de aminoácidos que compõem o peptídeo sinal (PS) (losango rosa), domínio catalítico (retângulo azul), domínio linker (retângulo laranja) e domínio CBM2 (elipse amarela). N e C sinalizam as regiões N- e C-terminais, respectivamente. ....43
- Figura 8** - Representação estrutural de alguns modelos preditos erroneamente para a enzima XALc\_0484. O domínio catalítico está representado em ciano-claro, o linker em rosa-escuro e o CBM2 em amarelo-claro. (a) O servidor SWISS-MODEL não conseguiu modelar o domínio linker. (b) O servidor I-Tasser uniu os três domínios em um modelo de domínio único. (c) O software MODELLER não conseguiu enovelar corretamente o domínio CBM2, mantendo-o estendido.....44
- Figura 9** - Sequência dos resíduos que compõem a enzima XALc\_0484, destacado em laranja a região do domínio linker, entre os resíduos 462 - 611, composta majoritariamente por glicinas. ....50

**Figura 10** - Representação estrutural da enzima XALc\_0484 predita pelos algoritmos (a) RoseTTAFold, (b) AlphaFold e (c) ColabFold. O domínio catalítico está representado em ciano-claro, o linker em rosa-escuro e o CBM2 em amarelo-claro.....51

**Figura 11** - Representação estrutural e gráfica de qualidade local dos modelos preditos pelos algoritmos (a) AlphaFold, (b) ColabFold e (c) RoseTTAFold para a enzima XALc\_0484. O círculo vermelho nos gráficos identifica a região linker e o colchete vermelho na estrutura de RoseTTAFold indica o domínio linker do modelo.....52

**Figura 12** - Visualização estrutural em modo “cartoon” das cinco enzimas obtidas do banco de dados do AlphaFold, incluindo o peptídeo sinal. (a) XALc\_0484, (b) XALc\_0865, (c) XALc\_0874, (d) XALc\_2967 e (e) XALc\_02969. A coloração está representada conforme a pontuação de confiança pLDDT, sendo: azul = alta precisão (> 90), ciano = adequado (70 - 90), amarelo = cautela (50 - 70) e laranja = baixa confiança (< 50).....56

**Figura 13** - Sobreposição estrutural entre os domínios catalíticos dos modelos obtidos do banco de dados do AlphaFold, com as estruturas experimentais homólogas identificadas pelo FoldSeek (magenta). (a) XALc\_0484 (ciano-claro) sobreposta com CbsA (PDB: 5xyh) de *Xanthomonas oryzae* pv. *oryzae*. (b) XALc\_0865 (rosa-claro) sobreposta com Cel12A (PDB: 2bw8) de *Rhodothermus marinus*. (c) XALc\_2967 (amarelo-claro) e (d) XALc\_2969 (verde-claro), ambas sobrepostas com XccCel5A (PDB:4tuf\_C) de *Xanthomonas campestris* pv. *campestris*.....58

**Figura 14** - Visualização estrutural em modo “cartoon” da enzima XALc\_0874, obtida do banco de dados do AlphaFold, sem o peptídeo sinal. (a) Domínio catalítico em azul, linker em rosa-escuro, e domínio CBM2 em amarelo-claro. (b) Domínio CBM2 em vista frontal. (c) Domínio catalítico em vista lateral, (d) posterior e (e) frontal.....59

**Figura 15** - Sobreposição estrutural do domínio catalítico da enzima XALc\_0874 (azul), obtida do banco de dados do AlphaFold, sem o peptídeo sinal, com a enzima alginato liase (PDB: 7o7a) (magenta) de *Pseudopedobacter saltans*.....60

**Figura 16** - Representação estrutural do domínio catalítico de XALc\_0484. As  $\alpha$ -hélices são representadas em ciano-claro, as fitas- $\beta$  em rosa-escuro e as alças em branco. (a) Visão frontal da estrutura de XALc\_0484, destacando o centro do domínio  $\beta$ -barril central formado por sete fitas- $\beta$  (rosa-escuro). As regiões N- e C-terminais estão ocultas por duas  $\alpha$ -hélices. (b) Visão lateral da estrutura de XALc\_0484, onde é possível visualizar as sete fitas- $\beta$  (rosa-escuro) paralelas, rotacionando para formar o  $\beta$ -barril central. N e C sinalizam as regiões N- e C-terminais, respectivamente. ....61

**Figura 17** - Representação estrutural do domínio catalítico de XALc\_0865. As  $\alpha$ -hélices são representadas em azul-claro, as fitas- $\beta$  em rosa-claro e as alças em branco. (a) Visão frontal da estrutura de XALc\_0865, evidenciando as duas folhas- $\beta$ , denominadas por A e B. As fitas- $\beta$  individuais que compõem cada folha estão numeradas (A1 - A5 e B1 - B10), conforme a posição em cada folha. N e C sinalizam as regiões N- e C-terminais, respectivamente, ressaltando que o N-terminal se inicia na fita- $\beta$  B1. (b) Visão lateral da estrutura de XALc\_0865, onde é possível visualizar as folhas- $\beta$  A e B sobrepostas, compondo o domínio sanduíche- $\beta$  e curvadas

para formar o sítio de ligação. C sinaliza a região C-terminal, visto que o N-terminal estava oculto por uma folha- $\beta$  e, portanto, não pôde ser representado..... 62

**Figura 18** - Representação estrutural do domínio catalítico de XALc\_2967. As  $\alpha$ -hélices são representadas em amarelo-claro, as fitas- $\beta$  em laranja e as alças em branco. **(a)** Visão frontal da estrutura de XALc\_2967, evidenciando o centro do domínio barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássico formado por oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela ao fundo. N e C sinalizam as regiões N- e C-terminais, respectivamente. **(b)** Visão posterior da estrutura de XALc\_2967, onde é possível visualizar oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela que auxilia na proteção contra o solvente. N e C sinalizam as regiões N- e C-terminais, respectivamente.... 63

**Figura 19** - Representação estrutural do domínio catalítico de XALc\_2969. As  $\alpha$ -hélices são representadas em verde-claro, as fitas- $\beta$  em laranja e as alças em branco. **(a)** Visão frontal da estrutura de XALc\_2969, evidenciando o centro do domínio barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássico formado por oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela ao fundo. N e C sinalizam as regiões N- e C-terminais, respectivamente. **(b)** Visão lateral da estrutura de XALc\_2969, onde é possível visualizar oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela na parte inferior da estrutura, que auxilia na proteção contra o solvente. N e C sinalizam as regiões N- e C-terminais, respectivamente. .... 64

**Figura 20** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_0484, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb. 66

**Figura 21** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_0865, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb. 67

**Figura 22** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_2967, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb. 67

**Figura 23** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_2969, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb. 68

**Figura 24** - Gráficos de Ramachandran de cada CAZyme. **(a)** XALc\_0484; **(b)** XALc\_0865. **(c)** XALc\_2967; **(d)** XALc\_2969..... 70

**Figura 25** - Estrutura química dos ligantes selecionados para o docking..... 72

**Figura 26** - Estrutura de XALc\_0484 complexada com os cinco ligantes selecionados; **(a)**  $\beta$ -D-glucose; **(b)**  $\beta$ -celobiose; **(c)**  $\beta$ -celotriose; **(d)**  $\beta$ -celotetraose; **(e)** Metil-4..... 75

**Figura 27** - Visão geral dos principais subsítios negativos e positivos de XALc\_0484. Em roxo os principais resíduos que participam do processo catalítico. A celohexaose (laranja) complexada com XALc\_0484 foi obtida a partir de sobreposição estrutural com a celobiohidrolase GH6 (Cel6B) de *Thermobifida fusca* (PDB: 4b4f). .... 76

**Figura 28** - Estrutura de XALc\_0484 complexada com  $\beta$ -D-glicose posicionada no subsítio - 4. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). .... 78

- Figura 29** - Estrutura de XALc\_0484 complexada com  $\beta$ -celobiose posicionada nos subsítios -1 e +1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 79
- Figura 30** - Estrutura de XALc\_0484 complexada com  $\beta$ -celotriose posicionada nos subsítios -1 a +2. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 80
- Figura 31** - Estrutura de XALc\_0484 complexada com Metil-4 posicionado nos subsítios -2 a +2. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 81
- Figura 32** - Estrutura de XALc\_0484 complexada com  $\beta$ -celotetraose posicionada nos subsítios -2 a +2. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 81
- Figura 33** - Estrutura de XALc\_0865 complexada com os quatro ligantes selecionados; (a)  $\beta$ -D-glicose; (b)  $\beta$ -celobiose; (c)  $\beta$ -celotriose; (d)  $\beta$ -celotetraose. .... 84
- Figura 34** - Visão geral dos principais subsítios negativos e positivos de XALc\_0865. Em roxo os principais resíduos que participam do processo catalítico. As duas celotetraoses (laranja) complexadas com XALc\_0865 foram obtidas a partir de sobreposição estrutural com as endoglucanases GH12 (Cel12A) de *Humicola grisea* (PDB: 1uu5 e 1uu6). ..... 85
- Figura 35** - Estrutura de XALc\_0865 complexada com  $\beta$ -celotetraose posicionada nos subsítios -2 a +2. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 86
- Figura 36** - (a) Estrutura de XALc\_0865 complexada com  $\beta$ -celobiose posicionada nos subsítios -2 e -1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å); (b) Visão superficial dos subsítios ocupados por  $\beta$ -celobiose (laranja) e  $\beta$ -D-glicose (ciano). ..... 88
- Figura 37** - Estrutura de XALc\_0865 complexada com  $\beta$ -D-glicose posicionada no subsítio -2. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 89
- Figura 38** - Estrutura de XALc\_0865 complexada com  $\beta$ -celotriose posicionada nos subsítios -1 a +2. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em

angstroms (Å); **(b)** Visão superficial da conformação e subsítios ocupados por  $\beta$ -celotriose (laranja) e  $\beta$ -celotetraose (ciano). ..... 90

**Figura 39** - Estrutura de XALc\_2967 complexada com os cinco ligantes selecionados: **(a)**  $\beta$ -D-glicose; **(b)**  $\beta$ -celobiose; **(c)**  $\beta$ -celotriose; **(d)**  $\beta$ -celotetraose. .... 92

**Figura 40** - Visão geral dos principais subsítios negativos e positivos de XALc\_2967. Em roxo os principais resíduos que participam do processo catalítico. A celotetraose (laranja) complexada com XALc\_2967 foi obtida a partir de sobreposição estrutural com o domínio catalítico da endoglucanase GH5\_1 (Cel5A) de *Acidothermus cellulolyticus* (PDB: 1ece). As linhas tracejadas em amarelo representam as ligações de hidrogênio entre alguns resíduos que estabilizam o sítio ativo. A linha tracejada em verde-claro indica a distância entre os dois glutamatos catalíticos. Todas as distâncias são mostradas em angstroms (Å). ..... 93

**Figura 41** - Estrutura de XALc\_2967 complexada com  $\beta$ -celobiose posicionada nos subsítios -1 e +1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 95

**Figura 42** - Estrutura de XALc\_2967 complexada com  $\beta$ -D-glicose posicionada no subsítio -1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 96

**Figura 43** - Estrutura de XALc\_2967 complexada com  $\beta$ -celotriose posicionada nos subsítios -2 a +1. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 96

**Figura 44** - Estrutura de XALc\_2967 complexada com  $\beta$ -celotetraose posicionada nos subsítios -3 a +1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). ..... 97

**Figura 45** - **(a)** Alinhamento entre as sequências do domínio catalítico de XALc\_2969 e XALc\_2967. Os resíduos idênticos estão em negrito. Os triângulos vermelhos sinalizam os dois glutamatos catalíticos e os triângulos azuis apontam os seis resíduos estabilizadores. **(b)** Sobreposição estrutural entre XALc\_2969 (verde-claro) e XALc\_2967 (amarelo-claro). ..... 99

**Figura 46** - Sobreposição estrutural entre os oito (Glu182, Glu305, Arg84, His136, Asn181, His261, Tyr263 e Trp337) principais resíduos que participam do processo catalítico de XALc\_2969 (verde-claro), XALc\_2967 (amarelo-claro) e XccCel5A (PDB: 4tuf) (magenta). ..... 100

**Figura 47** - Visão geral dos principais subsítios negativos e positivos de XALc\_2969. Em roxo os principais resíduos que participam do processo catalítico. A celotetraose (laranja) complexada com XALc\_2969 foi obtida a partir de sobreposição estrutural com o domínio catalítico da endoglucanase GH5\_1 (Cel5A) de *Acidothermus cellulolyticus* (PDB: 1ece). . 100

- Figura 48** - Estrutura de XALc\_2969 complexada com os cinco ligantes selecionados; (a)  $\beta$ -D-glicose; (b)  $\beta$ -celobiose; (c)  $\beta$ -celotriose; (d)  $\beta$ -celotetraose. .... 102
- Figura 49** - Estrutura de XALc\_2969 complexada com  $\beta$ -D-glicose posicionada no subsítio -1. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). .... 103
- Figura 50** - Estrutura de XALc\_2969 complexada com  $\beta$ -celobiose posicionada nos subsítios -1 e +1. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). .... 104
- Figura 51** - Estrutura de XALc\_2969 complexada com  $\beta$ -celotriose posicionada nos subsítios -1 a +2. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). .... 104
- Figura 52** - Estrutura de XALc\_2969 complexada com  $\beta$ -celotetraose posicionada nos subsítios -2 a +1. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å). .... 105
- Figura 53** - Conformação e posição adotada por  $\beta$ -celotetraose no sítio ativo de XALc\_2969. (a) Visão ampliada do complexo, evidenciando que o final não-redutor de  $\beta$ -celotetraose está posicionado fora do sítio ativo de XALc\_2969; (b) Visão lateral direita do sítio ativo de XALc\_2969 complexado com  $\beta$ -celotetraose; (c) Visão lateral esquerda do sítio ativo de XALc\_2969 complexado com  $\beta$ -celotetraose. .... 106
- Figura 54** - (a) Sobreposição estrutural entre o domínio catalítico de XALc\_2969 (verde-claro), XALc\_2967 (amarelo-claro), XccCel5A (PDB: 4tuf) (cinza) e EGPh (PDB: 3qhn) (bege); (b) Ampliação da sobreposição estrutural evidenciando o sítio ativo sobreposto das quatro enzimas. O círculo vermelho delimita a região reduzida do sítio ativo de XALc\_2969. .... 107
- Figura 55** - (a) Domínio catalítico de quatro endoglucanases GH5\_1: XALc\_2969 (verde-claro), XALc\_2967 (amarelo-claro), XccCel5A (PDB: 4tuf) (cinza) e EGPh (PDB: 3qhn) (bege); (b) Ampliação do domínio catalítico evidenciando o sítio ativo das quatro enzimas. O círculo vermelho delimita a região de interesse para comparação visual entre os sítios ativos. .... 109
- Figura 56** - Ampliação e visão frontal da região não-redutora (subsítio -3) do sítio ativo das quatro endoglucanases GH5\_1: (a) XALc\_2969 (verde-claro); (b) XALc\_2967 (amarelo-claro); (c) XccCel5A (PDB: 4tuf) (cinza); (d) EGPh (PDB: 3qhn) (algodão/bege). O círculo vermelho delimita a região de interesse para comparação visual entre os sítios ativos. .... 110
- Figura A1** - Visualização gráfica da estimativa de qualidade local determinada pelo QMEANDisCo para as quatro CAZymes de *Xanthomonas albilineans*, previstas pelos algoritmos de inteligência artificial, AlphaFold, ColabFold e RoseTTAFold. (a)

XALc\_0865\_AlphaFold, (b) XALc\_0865\_ColabFold, (c) XALc\_0865\_RoseTTAFold, (d) XALc\_0874\_AlphaFold, (e) XALc\_0874\_ColabFold, (f) XALc\_0874\_RoseTTAFold, (g) XALc\_2967\_AlphaFold, (h) XALc\_2967\_ColabFold, (i) XALc\_2967\_RoseTTAFold, (j) XALc\_2969\_AlphaFold, (k) XALc\_2969\_ColabFold e (l) XALc\_2969\_RoseTTAFold. Os gráficos para a enzima XALc\_0484 estão representados na Figura 9..... 171

**Figura A2** - Visualização gráfica da estimativa de qualidade local determinada pelo QMEANDisCo para o domínio catalítico das cinco CAZymes de *Xanthomonas albilineans*, previstas pelos algoritmos de inteligência artificial, AlphaFold, ColabFold e RoseTTAFold. (a) XALc\_0484\_AlphaFold, (b) XALc\_0484\_ColabFold, (c) XALc\_0484\_RoseTTAFold, (d) XALc\_0865\_AlphaFold, (e) XALc\_0865\_ColabFold, (f) XALc\_0865\_RoseTTAFold, (g) XALc\_0874\_AlphaFold, (h) XALc\_0874\_ColabFold, (i) XALc\_0874\_RoseTTAFold, (j) XALc\_2967\_AlphaFold, (k) XALc\_2967\_ColabFold, (l) XALc\_2967\_RoseTTAFold, (m) XALc\_2969\_AlphaFold, (n) XALc\_2969\_ColabFold e (o) XALc\_2969\_RoseTTAFold.. 172

**Figura A3** - Previsão de topologia para as enzimas (a) XALc\_0484, (b) XALc\_0865, (c) XALc\_2967 e (d) XALc\_2969, geradas pelo software PDBsum. As fitas- $\beta$  são representadas pelas setas e as  $\alpha$ -hélices pelos bastonetes. N e C sinalizam as regiões N- e C-terminais, respectivamente. .... 173

**Figura A4** - Previsão de estrutura secundária para a enzima XALc\_0484, gerada pelo software PDBsum..... 174

**Figura A5** - Previsão de estrutura secundária para a enzima XALc\_0865, gerada pelo software PDBsum..... 175

**Figura A6** - Previsão de estrutura secundária para a enzima XALc\_02967, gerada pelo software PDBsum..... 176

**Figura A7** - Previsão de estrutura secundária para a enzima XALc\_02969, gerada pelo software PDBsum..... 177

## LISTA DE TABELAS

<b>Tabela 1</b> - Enzimas ativas em carboidratos secretadas por <i>Xanthomonas albilineans</i> GPE PC73 que possuem localização extracelular.....	41
<b>Tabela 2</b> - Pontuação de confiança global pLDDT (AlphaFold e ColabFold) e IDDT (RoseTTAFold), determinada por cada algoritmo de predição estrutural. Em negrito os maiores escores para cada CAZyme. ....	45
<b>Tabela 3</b> - Análise de qualidade global, realizada pelo software QMEANDisCo, entre as estruturas obtidas pelos três métodos de predição. Em negrito as maiores pontuações para cada CAZyme. ....	46
<b>Tabela 4</b> - Análise de qualidade local, realizada pelo software MolProbity, entre as estruturas obtidas pelos três métodos de predição. Em negrito os melhores valores para cada CAZyme. (continua).....	48
<b>Tabela 5</b> - Análise de qualidade global realizada pelo software QMEANDisCo entre os domínios catalíticos das estruturas obtidas pelos três métodos de predição. Em negrito as maiores pontuações para cada CAZyme. ....	53
<b>Tabela 6</b> - Análise de qualidade local realizada pelo software MolProbity entre os domínios catalíticos das estruturas obtidas pelos três métodos de predição. Em negrito os melhores valores para cada CAZyme. (continua) .....	54
<b>Tabela 7</b> - Posição dos melhores sítios de ligação preditos pelo software PrankWeb para cada CAZyme. ....	65
<b>Tabela 8</b> - Valores expressos sobre a qualidade dos ângulos diédricos phi $\phi$ e psi $\psi$ dos resíduos que compõem cada CAZyme. ....	69
<b>Tabela 9</b> - Principais ligantes selecionados para cada família enzimática .....	71
<b>Tabela 10</b> - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc_0484 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.....	74
<b>Tabela 11</b> - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc_0865 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.....	83
<b>Tabela 12</b> - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc_2967 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.....	94
<b>Tabela 13</b> - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc_2969 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.....	101

**Tabela 14** - Pontuações de afinidade de ligação para cada ligante comum entre as quatro enzimas. As melhores pontuações estão em negrito..... 111

**Tabela A1** - Lista das 486 proteínas identificadas no proteoma de *Xanthomonas albilineans* que possuem peptídeo sinal, conforme previsão realizada pelo software SignalP. As proteínas são categorizadas conforme a probabilidade de possuírem um dos cinco tipos de peptídeo sinal: o peptídeo “padrão” - Sec/SPI (SP); o peptídeo sinal proveniente de lipoproteínas - Sec/SPII (LIPO); peptídeo sinal de transportadores TAT - Tat/SPI (TAT); peptídeo sinal de lipoproteínas transportadas por TAT - Tat/SPII (TATLIPO); pilina e peptídeo sinal semelhante a pilina - Sec/SPIII (PILIN). ..... 127

**Tabela A2** - Lista das 109 proteínas do proteoma de *Xanthomonas albilineans*, identificados pelo software CELLO como tendo localização extracelular. A tabela apresenta a probabilidade de localização em cinco regiões principais: extracelular, membrana interna ou externa, periplasmática e citoplasmática. .... 154

**Tabela A3** - Possíveis CAZymes identificadas no servidor dbCAN3 utilizando os softwares HMMER e DIAMOND para realizar busca nos bancos de dados CAZy e dbCAN\_sub. .... 160

# SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>20</b>
1.1 BIOENERGIA E CELULOSE .....	20
1.2 HIDRÓLISE ENZIMÁTICA DA CELULOSE .....	23
1.3 A BACTÉRIA <i>Xanthomonas albilineans</i> .....	26
1.4 BIOINFORMÁTICA ESTRUTURAL .....	30
<b>2 OBJETIVOS</b> .....	<b>33</b>
2.1 OBJETIVO GERAL .....	33
2.2 OBJETIVOS ESPECÍFICOS .....	33
<b>3 MATERIAL E MÉTODOS</b> .....	<b>34</b>
3.1 OBTENÇÃO DAS SEQUÊNCIAS PRIMÁRIAS DE <i>Xanthomonas albilineans</i> .....	35
3.2 ANÁLISE <i>in silico</i> DO PROTEOMA E IDENTIFICAÇÃO DAS CAZymes .....	35
3.3 MODELAGEM MOLECULAR .....	35
3.3.1 Triagem das Estruturas Molde .....	35
3.3.2 Modelagem Molecular das Enzimas .....	36
3.3.3 Visualização dos Modelos .....	37
3.4 SIMULAÇÕES DE DOCKING MOLECULAR .....	37
3.4.1 Preparo dos Ligantes .....	37
3.4.2 Preparo das Enzimas .....	37
3.4.3 Docking Molecular .....	38
<b>4 RESULTADOS</b> .....	<b>39</b>
4.1 TRIAGEM <i>in silico</i> DO PROTEOMA DE <i>Xanthomonas albilineans</i> .....	39
4.2 MODELAGEM MOLECULAR DAS CAZymes .....	43
4.3 ANÁLISE COMPARATIVA DAS ESTRUTURAS TRIDIMENSIONAIS .....	56
4.4 PREDIÇÃO DOS SÍTIOS DE LIGAÇÃO .....	64
4.5 ESTUDOS DE DOCKING MOLECULAR DAS CAZymes DE <i>Xanthomonas albilineans</i> .....	71
4.5.1 Triagem e Seleção dos Ligantes .....	71
4.5.2 Simulações de Docking Molecular das CAZymes .....	73
4.5.3 Docking Molecular da Enzima Xalc_0484 .....	73
4.5.4 Docking Molecular da Enzima Xalc_0865 .....	83
4.5.5 Docking Molecular da Enzima Xalc_2967 .....	91
4.5.6 Docking Molecular da Enzima Xalc_2969 .....	97

4.5.7 Análise Comparativa dos Resultados de Docking .....	111
<b>5 CONCLUSÃO.....</b>	<b>112</b>
<b>REFERÊNCIAS.....</b>	<b>114</b>
<b>APÊNDICE A - Triagem <i>in silico</i> do proteoma de <i>Xanthomonas albilineans</i>. .....</b>	<b>127</b>
<b>APÊNDICE B - Análises estruturais e de qualidade das CAZymes de <i>Xanthomonas albilineans</i> selecionadas.....</b>	<b>171</b>

# 1 INTRODUÇÃO

## 1.1 BIOENERGIA E CELULOSE

A queima de combustíveis fósseis prevalece como a principal fonte de energia global, sendo também um importante contribuinte para a poluição ambiental (CHETTRI; VERMA; VERMA, 2020). No ano de 2022, aproximadamente 82% da energia consumida mundialmente derivou-se da queima de petróleo (31,57%), carvão (26,73%) e gás (23,49%) (RITCHIE; PABLO, 2023). Somente essas três matrizes energéticas representam o equivalente a três quartos das emissões de gases responsáveis pelo efeito estufa (RITCHIE; PABLO, 2023; RITCHIE; ROSADO; ROSER, 2023). Além dos impactos ambientais, os combustíveis fósseis centralizam o poder energético. As principais reservas e tecnologias para extração, refino e produção, pertencem a um número restrito de países, criando uma vulnerabilidade global em caso de interrupção do fornecimento de combustível por essas nações (CHETTRI; VERMA; VERMA, 2020; RITCHIE; PABLO, 2023; RITCHIE; ROSADO; ROSER, 2023). Apesar da eficiência energética dos combustíveis fósseis, a dependência global representa um risco potencial, além dos problemas ambientais inerentes ao seu uso contínuo.

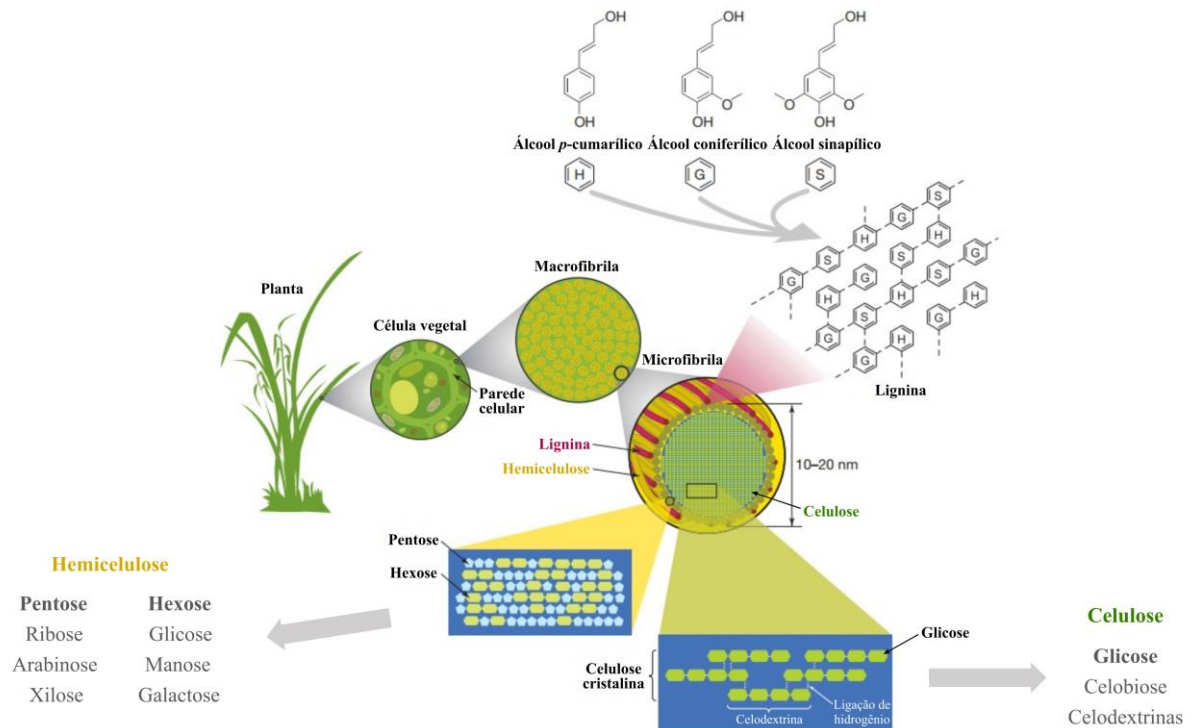
A transição para fontes de energia mais limpas, renováveis e descentralizadas é uma realidade. Ao longo de décadas, pesquisas com matrizes energéticas alternativas, como a eólica, solar e celulose vegetal, têm apresentado resultados promissores. No entanto, até o momento, essas fontes ainda não atingiram a eficiência necessária para substituir os combustíveis fósseis, indicando a necessidade de mais pesquisas na área. Com respeito a celulose vegetal, além de ser uma matéria-prima crucial para diversos produtos industriais como papel, fibras, produtos consumíveis e farmacêuticos, tem se mostrado como fonte de energia particularmente promissora, especialmente para a produção de biocombustível (CHUNDAWAT et al., 2011; MAITAN-ALFENAS; VISSER; GUIMARÃES, 2015; THERASME et al., 2021; ZABED et al., 2016). Essa possibilidade reside não apenas em sua versatilidade como matéria-prima, mas também em suas características favoráveis para a produção sustentável de energia (CHUNDAWAT et al., 2011; HAPPS et al., 2024; MAITAN-ALFENAS; VISSER; GUIMARÃES, 2015).

Diferentemente do etanol “comum”, produzido principalmente a partir dos açúcares obtidos do milho e da cana-de-açúcar, o etanol celulósico utiliza a lignocelulose proveniente de diversas fontes, como gramíneas, madeira lenhosa, resíduos industriais, agrícolas e qualquer derivado vegetal como insumo (CHUNDAWAT et al., 2011; HAPPS et al., 2024; ZABED et

al., 2016). Este tipo de etanol possui o potencial de substituir plantas alimentícias como matéria-prima, evitando a competição entre a produção de alimentos e energia. Além disso, é fundamental destacar o seu impacto ambiental positivo: o dióxido de carbono absorvido pelas plantas durante a fotossíntese reduziria significativamente a pegada de carbono oriunda da produção e utilização do biocombustível (RUBIN, 2008; THERASME et al., 2021; ZABED et al., 2016). Em toda a cadeia produtiva, o etanol celulósico reduziria em aproximadamente 60% as emissões de gases do efeito estufa, contra os aproximados 41% de redução de emissão previstos para o etanol “comum” (THERASME et al., 2021; ZABED et al., 2016). Além disso, a lignocelulose é um insumo de baixo custo, diversificado, não comestível por humanos e uma das fontes de carbono mais abundantes no planeta (FIGUEIREDO et al., 2019; LAMPUGNANI et al., 2018; MCNAMARA; MORGAN; ZIMMER, 2015).

A lignocelulose consiste em uma complexa matriz de polissacarídeos que compõem a estrutura resistente e flexível da parede celular vegetal. Ela recebe essa denominação por ser composta, principalmente, por lignina, celulose e hemicelulose, cujas concentrações podem variar entre as espécies vegetais e a idade da planta (LAMPUGNANI et al., 2018; MCNAMARA; MORGAN; ZIMMER, 2015; RUBIN, 2008). Outros constituintes podem ser encontrados em proporções menores, tais como: pectina, lipídios, minerais e proteínas estruturais (ALVES et al., 2019; LAMPUGNANI et al., 2018). Celulose, hemicelulose e lignina entrelaçam-se para formar as microfibrilas que, por sua vez, organizam-se em macrofibrilas (**Figura 1**). Além de proporcionar estabilidade estrutural à parede celular, as macrofibrilas desempenham um papel fundamental na forma e desenvolvimento das plantas (CHETTRI; VERMA; VERMA, 2020; FIGUEIREDO et al., 2019; LAMPUGNANI et al., 2018; MCNAMARA; MORGAN; ZIMMER, 2015).

**Figura 1** - Estrutura principal da parede celular vegetal.

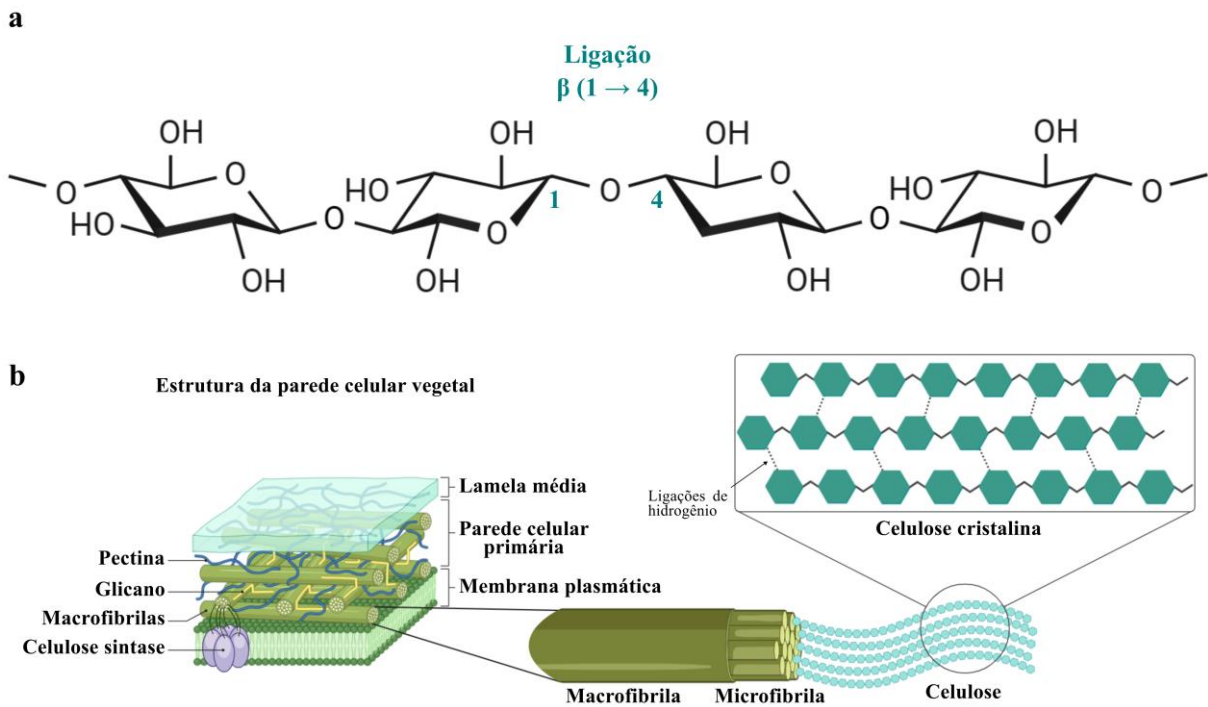


**Fonte:** Adaptada de Rubin, 2008.

Entre os três constituintes principais da lignocelulose presentes na maioria das plantas, a celulose se destaca em maior proporção, representando aproximadamente 51% da parede celular e em algumas espécies como o algodão, a celulose pode ser responsável por cerca de 90% da parede celular (FIGUEIREDO et al., 2019; LAMPUGNANI et al., 2018; ZABED et al., 2016).

A celulose é uma molécula simultaneamente simples e complexa. Quimicamente simples, por ser um polissacarídeo linear formado exclusivamente por resíduos de D-glicose unidos por ligações glicosídicas  $\beta$  (1  $\rightarrow$  4), difíceis de serem quebradas por enzimas (**Figura 2a**). Complexa devido à estrutura resultante da ligação  $\beta$  entre os átomos 1 e 4. Esta ligação rotaciona os monômeros de D-glicose em  $180^\circ$ , formando uma rede estrutural sólida, orientada em fibras cristalinas e de alta resistência mecânica (**Figura 2b**) (GARDNER; BLACKWELL, 1974; HON, 1994; LAMPUGNANI et al., 2018; MCNAMARA; MORGAN; ZIMMER, 2015).

**Figura 2 - (a)** Composição química da celulose formada por moléculas de D-glicose. Essas moléculas estão unidas por ligações glicosídicas  $\beta$  (1  $\rightarrow$  4) que ocorrem entre os átomos 1 e 4 da D-glicose. **(b)** Representação da rede estrutural da celulose, formada pela união das moléculas de D-glicose, sendo o componente principal das microfibrilas.



**Fonte:** Elaborada pelo autor com o auxílio de Biorender.com.

Estruturalmente, a função da celulose como principal componente na formação das microfibrilas (**Figura 2b**) é garantir a resistência mecânica essencial na manutenção da integridade da parede celular da planta (HON, 1994; LAMPUGNANI et al., 2018). Portanto, o processo para a obtenção da glicose a partir da celulose e, a subsequente produção de etanol celulósico, envolvem diversas etapas (CHUNDAWAT et al., 2011; MCNAMARA; MORGAN; ZIMMER, 2015; RUBIN, 2008; ZABED et al., 2016).

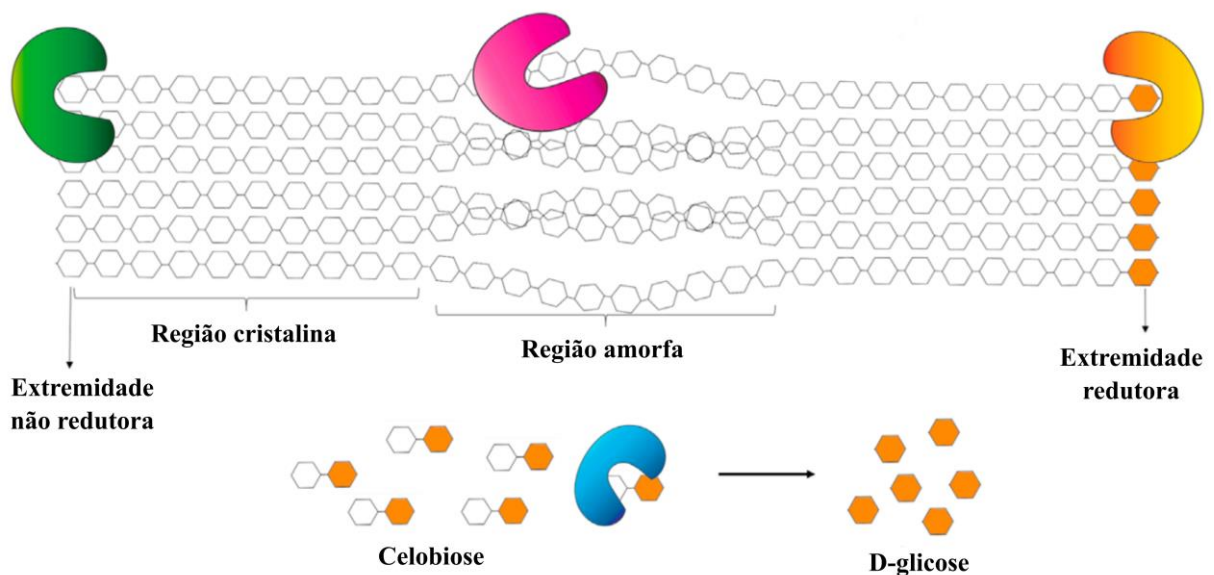
## 1.2 HIDRÓLISE ENZIMÁTICA DA CELULOSE

As etapas iniciais para a obtenção da glicose incluem a hidrólise da celulose, que pode ser realizada por intermédio de aditivos químicos e/ou enzimas biológicas (MAITAN-ALFENAS; VISSER; GUIMARÃES, 2015; RUBIN, 2008; ZABED et al., 2016). A hidrólise enzimática é preferível à abordagem química, visto que demonstra benefícios significativos, como o menor consumo de energia e o impacto ambiental reduzido. Entretanto, o alto custo das enzimas e a eficiência limitada na hidrólise da celulose tornam, até o momento, a produção em larga escala e a comercialização do etanol celulósico inviáveis (CHUNDAWAT et al., 2011; KUMAR; MURTHY, 2013; TAKEDA et al., 2022). Todavia, a viabilidade da produção pode

ser obtida por intermédio da compreensão do processo enzimático, identificação das condições ideais de hidrólise e busca contínua por novas enzimas.

A hidrólise enzimática da celulose envolve a ação sinérgica de quatro enzimas principais, denominadas celulasas. Cada enzima apresenta uma atividade catalítica específica, conforme classificado pela Comissão de Enzimas (EC, do inglês - *Enzyme Commission*) (CORNISH-BOWDEN, 2014), e trabalham em conjunto para converter a celulose em glicose. As endo- $\beta$ -1,4-glucanases (EC 3.2.1.4) clivam aleatoriamente as ligações glicosídicas  $\beta$  (1  $\rightarrow$  4) internas presentes nas regiões amorfas da celulose. Este processo resulta na liberação de oligossacarídeos, que possuem extremidades redutoras e não redutoras. As  $\beta$ -1,4-celobiosidases/celobiohidrolases degradam a celulose ao clivar os oligossacarídeos em celobiose, um dissacarídeo. Essas enzimas podem atuar tanto nas extremidades redutoras (EC 3.2.1.176) quanto nas não redutoras (EC 3.2.1.91) da porção cristalina da celulose. Por fim, as  $\beta$ -glicosidases (EC 3.2.1.21) hidrolisam a celobiose, convertendo-a em glicose (**Figura 3**) (CHUNDAWAT et al., 2011; KUMAR; MURTHY, 2013; MAITAN-ALFENAS; VISSER; GUIMARÃES, 2015; ZABED et al., 2016). Essas celulasas, responsáveis por degradar a celulose, são classificadas como enzimas ativas em carboidratos (CAZymes, do inglês *Carbohydrate-Active enZymes*) (CANTAREL et al., 2009).

**Figura 3** - Ação coordenada das quatro principais celulasas envolvidas no processo de hidrólise da celulose. A endo- $\beta$ -1,4-glucanase (EC 3.2.1.4), representada em rosa, atua clivando aleatoriamente a região amorfa interna da celulose. A  $\beta$ -1,4-celobiosidase/celobiohidrolase que degrada a celulose atuando nas extremidades não redutoras (EC 3.2.1.91), é representada em verde, e redutora (EC 3.2.1.176) da celulose, em amarelo. A  $\beta$ -glicosidase (EC 3.2.1.21), retratada em azul, converte a celobiose em glicose.



Fonte: Adaptada de Polizeli et al., 2016.

Além das celulases, as CAZymes englobam todas as enzimas que desempenham papéis essenciais na biotransformação dos carboidratos e glicoconjugados, abrangendo processos como degradação, modificação, biossíntese, metabolismo e transporte (CANTAREL et al., 2009; LOMBARD et al., 2014). A diversidade dos carboidratos naturais é refletida no amplo espectro de CAZymes registradas e classificadas em seis classes principais: Hidrolases glicosídicas (GH); Glicosiltransferases (GT); Liases de polissacarídeos (PL); Esterases de carboidratos (CE); Atividades auxiliares (AA), e Módulos de ligação a carboidratos (CBM). Essas classes representam as amplas funcionalidades dessas enzimas, sendo essenciais em processos como a hidrólise da celulose e outras biotransformações (CANTAREL et al., 2009; LOMBARD et al., 2014).

A classificação das CAZymes baseia-se principalmente em sua atividade catalítica, similaridades estruturais, sequências de aminoácidos e na presença de resíduos catalíticos conservados. Até o momento, todas as CAZymes identificadas estão catalogadas e classificadas no banco de dados CAZy (<https://www.cazy.org>). Além das classes enzimáticas, a categorização também é realizada por família, subfamília e clã (DRULA et al., 2022). As GHs destacam-se com o maior número de representantes, distribuídos em 187 famílias. Sendo as principais enzimas utilizadas para a degradação da lignocelulose, desempenham um papel fundamental ao catalisarem a hidrólise das ligações glicosídicas em carboidratos complexos (HENRISSAT, 1991).

As GTs estão classificadas em 117 famílias e atuam na transferência de grupos glicosil entre duas ou mais moléculas (doadoras e receptoras). Durante a transferência, essas enzimas catalisam a formação da ligação glicosídica para gerar um glicosídeo (CAMPBELL et al., 1997). As PLs, distribuídas em 43 famílias, clivam polissacarídeos contendo ácido urônico, como o ramnogalacturano, um componente da pectina encontrada na parede celular vegetal (LOMBARD et al., 2010). Esterases de carboidratos (CE), estão distribuídas em 20 famílias e clivam ésteres de cadeia pequena, favorecendo o relaxamento da parede vegetal ao interferir nas ligações do ácido poligalacturônico, o principal componente da pectina (BIELY, 2012; LOMBARD et al., 2010). As AAs atuam em conjunto com outras CAZymes promovendo reações de oxidação durante a degradação da biomassa. Nessa classe estão inclusas as enzimas que contribuem para a hidrólise da lignina, e até o momento, estão classificadas em 17 famílias (LEVASSEUR et al., 2013). Os CBMs não apresentam atividade catalítica, sendo classificados como módulos associados, distribuídos em 100 famílias (BORASTON et al., 2004). A sua principal função é ligar-se ao carboidrato e auxiliar o processo catalítico de outras enzimas, principalmente as GHs e PLs. Após o reconhecimento e ligação do CBM ao carboidrato, o

domínio catalítico da enzima é direcionado ao substrato, facilitando o acesso, interação e, consequente, hidrólise (BERNARDES et al., 2019; COURTADE et al., 2018). O auxílio do CBM aumenta a afinidade, velocidade e eficiência catalítica, bem como, a estabilidade térmica do domínio catalítico ao qual está associado (BORASTON et al., 2004).

As CAZymes são naturalmente sintetizadas por uma variedade de organismos procaríotos e eucariotos, sendo os microrganismos que incluem fungos, leveduras, arqueias e bactérias, os principais produtores (ASPEBORG et al., 2012; DAVIES; HENRISSAT, 1995). Mesmo com os avanços tecnológicos, estima-se que apenas cerca de 1% desses microrganismos foram isolados, cultivados e extensivamente avaliados. Isto significa que aproximadamente 99% dos organismos com potencial para serem fontes comerciais de CAZymes podem estar sendo negligenciados (ROSNOW et al., 2016). Essa sub-representação reflete nos principais coquetéis enzimáticos comerciais, predominantemente compostos por GHs provenientes de fungos aeróbios como *Trichoderma* e *Aspergillus* spp. (CHETTRI; VERMA; VERMA, 2020). Diante desse cenário, a exploração de novas CAZymes torna-se essencial para o desenvolvimento de diversas áreas industriais e biotecnológicas, como a produção de biocombustíveis (CHETTRI; VERMA; VERMA, 2020).

### 1.3 A BACTÉRIA *Xanthomonas albilineans*

Entre as bactérias produtoras de CAZymes, estão as representantes do gênero *Xanthomonas*. Esse gênero é composto por 27 espécies de bactérias fitopatogênicas, responsáveis por infectar diversas espécies vegetais, incluindo culturas de interesse comercial (SADDLER; BRADBURY, 2005). As *Xanthomonas* possuem um alto grau de especificidade tanto pelo hospedeiro, quanto pelo local de infecção e desenvolvimento (LEYNS et al., 1984). Entre as *Xanthomonas*, a *Xanthomonas albilineans* destaca-se como uma das espécies mais distintas (HAUBEN et al., 1997; YOUNG et al., 2008). Ela infecta exclusivamente plantas da família *Poaceae*, como milho, sorgo e diversas gramíneas, sendo a cana-de-açúcar a sua principal hospedeira. A *X. albilineans* é o agente responsável pela “Escaldadura das folhas”, uma das doenças mais relevantes dessa cultura vital para produção de biocombustíveis (BIRCH, 2001; PIERETTI et al., 2012).

A cana-de-açúcar, reconhecida por sua robustez, alto teor de sacarose armazenado nos colmos e eficiência na produção de biomassa (WANG et al., 2014), é uma planta comercialmente importante em regiões tropicais e subtropicais, onde o clima quente e úmido favorecem o seu cultivo (TEJERA et al., 2007). Além de ser uma cultura de interesse comercial,

devido à produção de açúcar e etanol “comum”, a cana-de-açúcar é uma importante fonte de biomassa lignocelulósica (MOORE; MARETZKI, 1996; VETTORE, 2003). Contudo, essa cultura é impactada pela infecção por *Xanthomonas albilineans*, que coloniza principalmente o xilema da planta. Os principais sintomas da escaldadura das folhas estão presentes nas folhas e envolvem o surgimento de uma faixa branca, estreita (clorose) e bem definida. Posteriormente a clorose aumenta, as folhas murcham, necrosam e ocorre a morte gradual do hospedeiro. Em muitos casos a multiplicação bacteriana e o estabelecimento da infecção não apresentam sintomas visíveis, representando uma ameaça significativa para a plantação (BIRCH, 2001; CHATTERJEE; ALMEIDA; LINDOW, 2008; MENSI et al., 2014; ROTT et al., 2000). Todo esse processo de invasão e colonização do xilema por *X. albilineans* é dependente de CAZymes especializadas em degradar a lignocelulose, abundantemente produzida pela cana-de-açúcar. Algumas variedades de cana-de-açúcar, focadas em bioenergia, foram projetadas visando produzirem mais biomassa lignocelulósica, do que sacarose livre (FANELLI et al., 2020; MEENA et al., 2020; SILVA et al., 2021).

O xilema vegetal é um ambiente desafiador para os patógenos devido à escassez de nutrientes, sendo composto predominantemente por tecido morto e lignificado. No entanto, mesmo nesse cenário, a seiva do xilema contém pequenas frações de produtos resultantes da quebra da parede celular vegetal. Ácidos orgânicos, aminoácidos, açúcares e íons inorgânicos, presentes em pequenas quantidades, são essenciais para as bactérias, permitindo que patógenos sistêmicos, como a *X. albilineans*, prosperem nesse ambiente (FATIMA; SENTHIL-KUMAR, 2015; PIERETTI et al., 2012). A sobrevivência nesse ambiente exige um alto nível de especialização por parte dos patógenos (JACKSON et al., 2011; LINDEBERG, 2012).

A distinção de *X. albilineans* em relação às demais espécies do gênero *Xanthomonas*, provavelmente, está relacionada à redução significativa de seu genoma durante o processo evolutivo. Esta redução favoreceu o desenvolvimento de mecanismos de infecção especializados e adaptados para a sobrevivência no sistema vascular da planta (AGARWAL et al., 2023; PIERETTI et al., 2009, 2012). Em média, as bactérias do gênero *Xanthomonas* possuem um único cromossomo circular, variando entre 4,8 Mb e 5,3 Mb de tamanho, enquanto o genoma de *X. albilineans* foi reduzido para 3,8 Mb (PIERETTI et al., 2009; RYAN et al., 2011). Essa redução genômica afetou clusters gênicos essenciais para codificar importantes mecanismos de infecção, tais como o de resposta de hipersensibilidade e patogenicidade-sistema de secreção do tipo 3 (Hrp-T3SS) e tipo 6 (Hrp-T6SS), bem como clusters gênicos para produção de goma xantana e outros. Esses clusters comuns e essenciais para a maioria das

bactérias do gênero *Xanthomonas* estão ausentes no genoma de *X. albilineans* (AGARWAL et al., 2023; GUPTA et al., 2024; LI et al., 2022; PIERETTI et al., 2009, 2015).

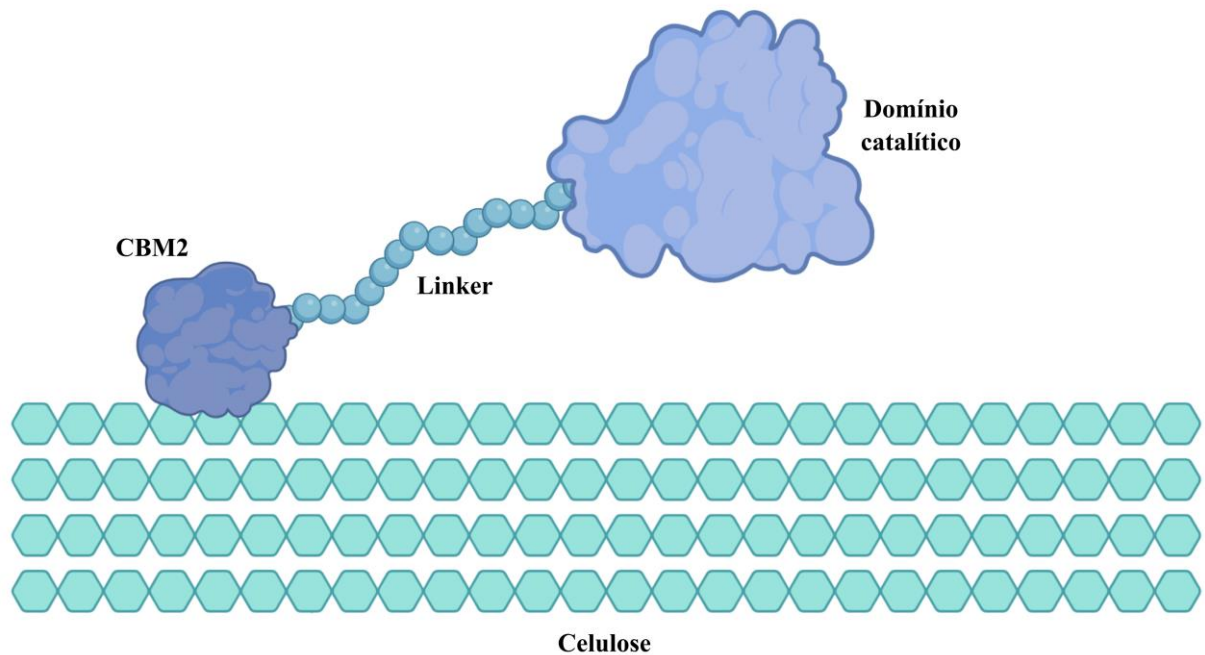
Apesar da ausência desses mecanismos de infecção essenciais para bactérias fitopatogênicas, *X. albilineans* se destaca como o único organismo capaz de sintetizar a albicidina, um peptídeo não ribossomal com propriedades fitotóxicas e antibióticas (BIRCH, 2001). A albicidina é responsável pela indução da clorose nas folhas e, simultaneamente, atua como agente antibiótico, assegurando a exclusividade do ambiente de desenvolvimento ao eliminar outros microrganismos (BIRCH, 2001; BIRCH; PATIL, 1987; FINKING; MARAHIEL, 2004; ZHANG; XU; BIRCH, 1999). Além disso, *X. albilineans* possui um “sistema de secreção do tipo 3 da Ilha de Patogenicidade de Salmonela-1” (T3SS SPI-1), observado em patógenos transmitidos por insetos. Estudos recentes indicam que além dos meios de transmissão descritos na literatura, a princípio, *X. albilineans* possui a capacidade de ser transmitida por insetos (BINI et al., 2023). No genoma de *X. albilineans*, também é encontrado um sistema de secreção do tipo II (T2SS), responsável por secretar as CAZymes que degradam a parede celular do hospedeiro (PIERETTI et al., 2012). É fundamental ressaltar que, entre os diversos mecanismos de infecção utilizados por *X. albilineans*, as adaptações que permitiram a este patógeno singular sobreviver em ambientes adversos, como o xilema, podem também ter influenciado na síntese de CAZymes especializadas na degradação da parede celular do hospedeiro (MENSI et al., 2014; PIERETTI et al., 2012, 2015).

Além de fornecer estrutura para a planta, a parede celular vegetal é um importante sistema de defesa físico-químico utilizado para limitar o ataque dos patógenos (FIGUEIREDO et al., 2019). Por esse motivo, as bactérias vasculares secretam um conjunto de enzimas com a finalidade de decompô-la e auxiliar no avanço do patógeno pelos tecidos da planta (JACOBS et al., 2012; JOHNSON et al., 2006; TAYI et al., 2016). Os produtos resultantes da quebra da parede celular podem ser utilizados como recurso alimentar, tanto pela liberação da glicose, quanto pela exposição dos nutrientes presentes no citosol e organelas da célula alvo (GLAZEBROOK, 2005; JACOBS et al., 2012; LALUK; MENGISTE, 2010).

Entre as enzimas secretadas por *X. albilineans* estirpe GPE PC73, quatro hidrolases glicosídicas (GHs) (XALc\_0484, XALc\_0865, XALc\_2967 e XALc\_2969) possuem um domínio catalítico na região N-terminal e um módulo de ligação ao carboidrato da família 2 (CBM2) na região C-terminal. Os domínios catalíticos (GHs) e os CBM2s das quatro enzimas estão unidos por linkers ricos em glicina, prolina, serina e/ou treonina (**Figura 4**) (PIERETTI et al., 2009, 2012, 2015). As principais funções do linker são conectar o domínio catalítico da enzima ao CBM e auxiliar no direcionamento do domínio catalítico ao substrato (BORASTON

et al., 2004). Geralmente, as regiões que formam os domínios linkers são desordenadas para manter a flexibilidade estrutural e, assim como os módulos CBM, os linkers são capazes de intensificar a eficiência da enzima (HU; OHUE, 2024; TOMPA, 2002). O complexo enzimático formado pela união dos domínios catalítico, linker e CBM, demonstrou ser mais eficiente para o processo de hidrólise, em comparação ao domínio catalítico isolado (NAKAMURA et al., 2016, 2020; PAYNE et al., 2013). Entre os membros do gênero *Xanthomonas*, o gene que codifica a enzima XALc\_0484, está presente apenas nas espécies colonizadoras do xilema (PIERETTI et al., 2012, 2015).

**Figura 4** - Representação esquemática de uma hidrolase glicosídica multidomínio secretada por *Xanthomonas albilineans* GPE PC73. A enzima é composta por um domínio catalítico e um módulo de ligação ao carboidrato 2 (CBM2), unidos pelo domínio linker.



**Fonte:** Elaborada pelo autor com o auxílio de Biorender.com.

A *Xanthomonas albilineans* sempre foi considerada uma bactéria de sobrevivência limitada ao xilema da planta. No entanto, Mensi et al., (2014), por intermédio de microscopia eletrônica, observaram que em até 15 dias após a infecção o patógeno não se restringia ao xilema, sendo encontrado em diversos tecidos do hospedeiro, incluindo floema, esclerênquima, parênquima e outros tecidos não vasculares. Essa capacidade é única entre as bactérias fitopatogênicas, pois patógenos vasculares geralmente vivem exclusivamente no xilema, enquanto patógenos apoplásticos residem apenas em áreas intercelulares do mesófilo (KUNKEL; CHEN, 2006). Até o momento, *X. albilineans* é a única bactéria fitopatogênica

relatada por infectar e sobreviver em tecidos distintos do hospedeiro, mesmo sem a presença de um sistema Hrp-T3SS funcional em seu genoma (MENSI et al., 2014). Especula-se que essa capacidade esteja relacionada ao conjunto característico de enzimas secretadas para degradar as paredes celulares primárias e secundárias do hospedeiro. Provavelmente essas enzimas apresentam atividade bifuncional, sendo adaptadas para serem utilizadas tanto na obtenção de alimento, quanto na locomoção pelos tecidos da planta, não se restringindo apenas ao xilema (MENSI et al., 2014; PIERETTI et al., 2012, 2015).

Devido à capacidade de *X. albilineans* em degradar a complexa parede celular vegetal, o estudo das enzimas secretadas por essa bactéria pode resultar na identificação de novas CAZymes com potencial para a produção de biocombustíveis. A função biológica de uma enzima está intrinsecamente relacionada à sua estrutura tridimensional, cujo conhecimento é fundamental para a sua compreensão (KELZ et al., 2022). Os métodos experimentais para determinar as estruturas de proteínas, como a cristalografia de raios X e a microscopia crioeletrônica (CryoEM, do inglês *CRYO-Electron microscopy*), demandam recursos significativos em termos financeiros e tempo de pesquisa. A aplicação desses métodos para cada enzima identificada em análises de genomas ou transcriptomas seria impraticável, tornando-se um obstáculo para a descoberta de novas CAZymes. (KELZ et al., 2022). Portanto, para circundar essas limitações, reduzindo o tempo e o custo de pesquisa, estão sendo empregados métodos de biologia computacional, complementados por técnicas de engenharia de proteínas, em projetos destinados à identificação de CAZymes para aplicações industriais (CHETTRI; VERMA; VERMA, 2020; DE CARVALHO et al., 2019).

#### 1.4 BIOINFORMÁTICA ESTRUTURAL

Dentre os principais métodos de biologia computacional, a modelagem de proteínas destaca-se como ferramenta essencial para realizar a predição estrutural de enzimas e auxiliar na triagem virtual com base em características catalíticas e ligantes de maior afinidade (CHETTRI; VERMA; VERMA, 2020; DE CARVALHO et al., 2019; KELZ et al., 2022; SAMAEI-DARYAN; GOLIAEI; EBRAHIM-HABIBI, 2017). Entre as técnicas de modelagem de proteínas, a modelagem comparativa tem sido amplamente utilizada por demandar menor tempo computacional e, geralmente, apresenta resultados mais promissores (HADDAD; ADAM; HEGER, 2020; HAMEDUH et al., 2020).

Em situações em que a modelagem comparativa se torna inviável, utilizam-se métodos como o *threading*, baseado na predição de modelos a partir de fragmentos de várias estruturas

conhecidas e o *ab initio*, que efetua a modelagem a partir do zero, fundamentando-se apenas na sequência de aminoácidos e cálculos físicos. Ambos os métodos exigem o uso de softwares que se baseiam no conhecimento estrutural de proteínas, previsão de ângulos de torção, campos de força e inserção de átomos por intermédio de modelos matemáticos. Entretanto, é importante ressaltar que esses métodos, além de demandarem maior capacidade computacional, podem apresentar resultados com menor precisão (SCHELER, 2019).

Com o advento dos algoritmos de inteligência artificial e a sua adoção pelos principais métodos de modelagem de proteínas, foi criado o algoritmo de aprendizado profundo (*deep learning*) denominado AlphaFold2 (AF2) (JUMPER et al., 2021). Atualmente, o AF2 é considerado a ferramenta de modelagem que superou os demais métodos computacionais anteriores, predizendo modelos tridimensionais com a precisão próxima de estruturas experimentais e, em alguns casos, auxiliando na resolução dessas estruturas (AKDEL et al., 2022; GUPTA et al., 2021; JUMPER et al., 2021; MARX, 2022; PEREIRA et al., 2021; VARADI et al., 2022). Diversos trabalhos foram desenvolvidos com o intuito de avaliar a confiabilidade das predições obtidas por AF2 e os resultados foram positivos. Além de estruturas com alta qualidade, também é possível obter dados sobre a flexibilidade dos resíduos da proteína e a sua dinâmica (AKDEL et al., 2022; BORDIN et al., 2023; BOUATTA; SORGER; ALQURAISHI, 2021; GUO et al., 2022; MARX, 2022; PEREIRA et al., 2021; SCARDINO; DI FILIPPO; CAVASOTTO, 2023; URBAN; POMPON, 2022; VARADI et al., 2022; VERKUIL et al., 2022). Posteriormente, surgiram diversos algoritmos baseados em AF2, como, por exemplo, o RoseTTAFold (BAEK; BAKER, 2022). Em colaboração com o EMBL's - European Bioinformatics Institute, foi criado o primeiro banco de dados de estruturas proteicas preditas exclusivamente pelo AF2 denominado AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk>) (VARADI et al., 2022).

O AlphaFold2 pode ser considerado um divisor de águas para a ciência, mas a infraestrutura computacional necessária para efetuar as simulações é alta e muitos grupos de pesquisa não possuem esse suporte (JUMPER et al., 2021). Com base nessa limitação, alguns projetos foram desenvolvidos para tornar o AF2 mais acessível. Dentre eles, o ColabFold (MIRDITA et al., 2022) é atualmente o projeto que obtém os melhores resultados. Esta plataforma utiliza o AlphaFold2 como base, mas otimiza algumas etapas do processo, tornando-se mais rápido e menos exigente em termos de recursos computacionais, mantendo a acurácia próxima do AlphaFold2 original. ColabFold se tornou uma ferramenta tão acessível e com baixo custo computacional que pode ser utilizada por intermédio da plataforma Google Colaboratory (<https://github.com/sokrypton/ColabFold>), estando disponível para todos os

grupos de pesquisa (MIRDITA et al., 2022). Portanto, como pode ser observado, a biologia computacional, bem como as ferramentas de modelagem de proteínas, promove o avanço em diversas áreas da pesquisa biológica.

Neste contexto, a utilização dessas ferramentas computacionais de última geração pode ser determinante para a descoberta de novas CAZymes. Por ser uma bactéria sistêmica, especializada em degradar a parede celular, a obtenção de informações sobre as CAZymes secretadas por *Xanthomonas albilineans* pode favorecer o descobrimento de produtos com potencial biotecnológico voltados para a agroindústria e bioenergia. Utilizando uma abordagem *in silico*, pretende-se selecionar e gerar modelos computacionais das CAZymes secretadas por *X. albilineans* GPE PC73 e obter dados da interação entre essas enzimas e a celulose.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

O objetivo deste trabalho foi identificar CAZymes com atividade celulósica no proteoma de *Xanthomonas albilineans* GPE PC73, utilizando ferramentas de bioinformática estrutural.

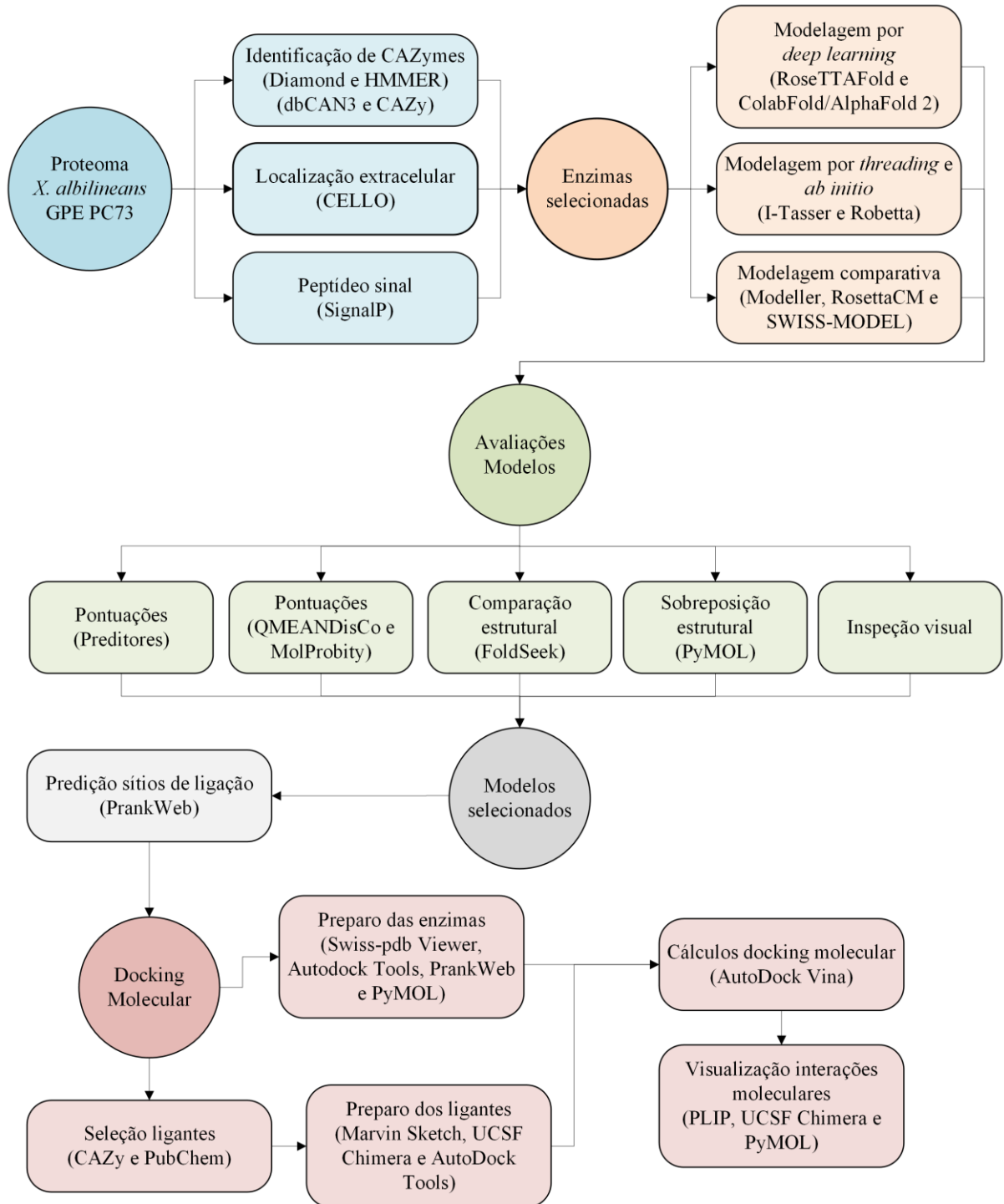
### 2.2 OBJETIVOS ESPECÍFICOS

- Realizar análises *in silico* do proteoma de *X. albilineans* GPE PC73 para identificar e selecionar as CAZymes extracelulares possivelmente responsáveis pela degradação de celulose;
- Realizar triagens em bancos de dados de estruturas de proteínas visando a identificação de estruturas tridimensionais adequadas para serem utilizadas como modelo em métodos de modelagem comparativa;
- Aplicar métodos de modelagem de proteínas, incluindo modelagem comparativa, *threading*, *ab initio* e *deep learning*, para gerar modelos tridimensionais das CAZymes de *X. albilineans* GPE PC73 selecionadas;
- Avaliar os modelos tridimensionais obtidos, utilizando análises de parâmetros visuais, comparativos, interações interatômicas, estereoquímicos e de contatos, para selecionar os modelos tridimensionais mais adequados;
- Identificar os possíveis sítios de ligação dos modelos estruturais selecionados;
- Selecionar os principais substratos alvos das CAZymes modeladas;
- Realizar simulações de docking molecular entre as CAZymes modeladas e os substratos selecionados;
- Avaliar as interações moleculares entre as CAZymes e os substratos.

### 3 MATERIAL E MÉTODOS

A **Figura 5** representa o fluxograma resumido da metodologia desenvolvida ao longo do projeto.

**Figura 5** - Metodologia aplicada durante as etapas de triagem do proteoma (azul), predição dos modelos (laranja), avaliação dos modelos (verde), predição dos sítios de ligação (cinza) e simulações de docking molecular (rosa).



Fonte: Elaborada pelo autor.

### 3.1 OBTENÇÃO DAS SEQUÊNCIAS PRIMÁRIAS DE *Xanthomonas albilineans*

As sequências do proteoma de *Xanthomonas albilineans* utilizadas nesse estudo foram obtidas a partir do sequenciamento da estirpe GPE PC73, depositado no GenBank pelo identificador NC\_013722.1 (PIERETTI et al., 2009). O proteoma está disponível nos servidores do Institut National de la Recherche Agronomique (INRAE - França) (<http://iant.toulouse.inra.fr/X.albilineans>) (PIERETTI et al., 2009). Esta estirpe foi isolada no ano de 2003 em Petit-Bourg, uma comunidade localizada nas ilhas de Guadalupe, região do Caribe sob jurisdição francesa (CHAMPOISEAU et al., 2006). O seu genoma e proteoma foram completamente sequenciados em 2009, tornando-se padrão de referência (MIRANDA et al., 2023; PIERETTI et al., 2009, 2015; ZHANG et al., 2020).

### 3.2 ANÁLISE *in silico* DO PROTEOMA E IDENTIFICAÇÃO DAS CAZymes

Todas as sequências de aminoácidos que compõem o proteoma de *X. albilineans* GPE PC73 foram analisadas por ferramentas computacionais, tais como os softwares: SignalP 6.0, para predição do peptídeo sinal (TEUFEL et al., 2022); CELLO 2.5, para predição da localização subcelular (YU et al., 2006); HMMER e DIAMOND, para realizar buscas por possíveis CAZymes nos bancos de dados Carbohydrate Active Enzymes data base (CAZy) (<http://www.cazy.org/>) e dbCAN3 (<https://bcb.unl.edu/dbCAN2/>) (BUCHFINK; REUTER; DROST, 2021; DRULA et al., 2022; FINN; CLEMENTS; EDDY, 2011; ZHENG et al., 2023). Os alinhamentos entre as sequências e análises de anotação funcional foram realizados utilizando-se os softwares e os servidores BlastP, BlastPDB (ALTSCHUL et al., 1990), Mega11 (TAMURA; STECHER; KUMAR, 2021), SMART (LETUNIC; KHEDKAR; BORK, 2021), InterPro (PAYSAN-LAFOSSE et al., 2023) e PDBsum (LASKOWSKI et al., 1997).

### 3.3 MODELAGEM MOLECULAR

#### 3.3.1 Triagem das Estruturas Molde

A partir das sequências de aminoácidos das possíveis CAZymes identificadas e selecionadas, foi realizado uma busca por estruturas tridimensionais (com sequências de aminoácidos similares), que poderiam servir como moldes experimentais, para o método de modelagem comparativa. Os bancos de dados utilizados para a busca foram, o Protein Data

Bank (PDB) (<https://www.rcsb.org/>) (BURLEY et al., 2023) e o SWISS-MODEL (<https://swissmodel.expasy.org/>) (WATERHOUSE et al., 2018).

### 3.3.2 Modelagem Molecular das Enzimas

Para a obtenção das melhores predições estruturais (modelos tridimensionais) das CAZymes de *X. albilineans* GPE PC73 selecionadas, foram utilizados diversos métodos de modelagem computacional, bem como critérios de avaliação para determinar a qualidade dos modelos. As simulações de modelagem comparativa foram realizadas utilizando-se os softwares MODELLER 9.25 (ALI, 1995), RosettaCM 3.12 (LEAVER-FAY et al., 2011; SONG et al., 2013) e o servidor SWISS-MODEL (WATERHOUSE et al., 2018).

Devido ao alto custo computacional, os métodos de *threading* e *ab initio* foram realizados nos servidores I-Tasser (YANG; ZHANG, 2015) e Robetta (RAMAN et al., 2009). Devido à disponibilização dos métodos que empregam *deep learning*, como o RoseTTAFold (BAEK et al., 2021) e AlphaFold2 (JUMPER et al., 2021), as CAZymes também foram submetidas a esse tipo de modelagem. As predições com o algoritmo RoseTTAFold foram realizadas por intermédio do servidor Robetta. O AlphaFold2 foi executado na plataforma ColabFold (MIRDITA et al., 2022). As configurações definidas no ColabFold foram para o uso do campo de força Amber e o banco de dados PDB70 como base de busca. Em todas as predições, as sequências de aminoácidos de cada enzima foram submetidas sem os resíduos que compõem o peptídeo sinal.

Além das estruturas resultantes dos sete métodos de predição utilizados, modelos obtidos diretamente do banco de dados do AlphaFold (AlphaFold Protein Structure Database - <https://alphafold.ebi.ac.uk/>) (VARADI et al., 2022), também foram analisados nas etapas de avaliação e seleção estrutural.

A qualidade de todas as estruturas preditas foram avaliadas através de cinco parâmetros: as pontuações internas de cada preditor (MODELLER, SWISS-MODEL, I-Tasser, Robetta, RoseTTAFold e AlphaFold2), as pontuações dos softwares de avaliação estrutural QMEANDisCo e MolProbity 4.4 (CHEN et al., 2010; STUDER et al., 2020), comparação estrutural utilizando o software FoldSeek (VAN KEMPEN et al., 2023) para identificar estruturas experimentais homólogas, sobreposição estrutural utilizando o software PyMOL 2.5 (DELANO, 2002) e inspeção visual.

### 3.3.3 Visualização dos Modelos

As visualizações e edições manuais dos modelos tridimensionais foram realizadas utilizando-se os softwares UCSF Chimera 1.17 (PETTERSEN et al., 2004) e PyMOL 2.5 (DELANO, 2002).

## 3.4 SIMULAÇÕES DE DOCKING MOLECULAR

As simulações de docking foram realizadas entre as estruturas tridimensionais selecionadas e os substratos de interesse.

### 3.4.1 Preparo dos Ligantes

Os ligantes foram selecionados a partir do banco de dados CAZy. Para cada família de enzimas, foram selecionados apenas os oligossacarídeos de celulose em conformação  $\beta$ , com quatro ou mais recorrências entre as estruturas proteicas depositadas no CAZy.

As estruturas químicas dos ligantes selecionados foram obtidas em 2D no banco de dados PubChem (<https://pubchem.ncbi.nlm.nih.gov>) (KIM et al., 2023). A conversão para modelos 3D, otimizações energéticas, bem como a adição (ou subtração) dos hidrogênios apropriados para pH 7,0, foram realizadas no software Marvin Sketch 23.4 (CSIZMADIA, 1999). O programa UCSF Chimera foi utilizado para adicionar as cargas das moléculas pelo método de Gasteiger (WANG et al., 2006) e o AutoDock Tools 1.5.7 (MORRIS et al., 2009; SANNER, 1999) para definir as rotações das ligações químicas.

### 3.4.2 Preparo das Enzimas

Os domínios catalíticos das possíveis CAZymes selecionadas nas etapas anteriores foram utilizados para os estudos de docking molecular. As reconstruções de possíveis cadeias laterais incompletas foram realizadas pelo software Swiss-pdb Viewer 4.1.0 (GUEX; PEITSCH, 1997). Os estados de protonação dos aminoácidos foram normalizados no AutoDock Tools, bem como a adição de cargas pelo método de Gasteiger.

### 3.4.3 Docking Molecular

Os cálculos de docking molecular foram realizados com o software AutoDock Vina 1.2.5 (TROTT; OLSON, 2009). Foram utilizados scripts adaptados para aumentar a aleatoriedade e maximizar os resultados do algoritmo genético do software. Em todas as simulações, o receptor (proteína) foi mantido rígido, enquanto o ligante foi considerado flexível, permitindo assim um aumento no número de conformações e melhores resultados. As coordenadas dos eixos X, Y e Z, assim como o tamanho da área de encaixe, foram definidos com base nas predições dos sítios de ligação realizadas pelo servidor PrankWeb (<https://prankweb.cz/>) (JAKUBEC et al., 2022). As interações moleculares foram analisadas com o auxílio do Protein-Ligand Interaction Profiler (PLIP) (SALENTIN et al., 2015), e as representações gráficas dos complexos ligante-proteína foram geradas utilizando o UCSF Chimera e o PyMOL. As validações das simulações de docking molecular foram realizadas seguindo o protocolo de redocking.

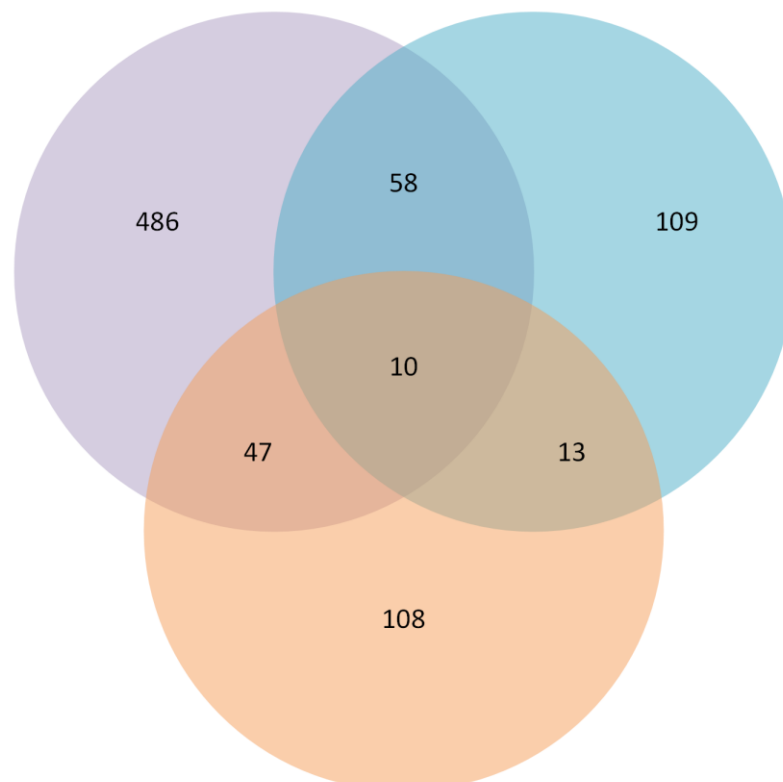
## 4 RESULTADOS E DISCUSSÃO

Os resultados descritos nas próximas seções foram obtidos a partir da análise *in silico* do proteoma de *Xanthomonas albilineans*, especificamente da estirpe GPE PC73 (PIERETTI et al., 2009).

### 4.1 TRIAGEM *in silico* DO PROTEOMA DE *Xanthomonas albilineans*

O proteoma de *Xanthomonas albilineans* GPE PC73 consiste em 3.208 proteínas sequenciadas. Dentre estas, 364 proteínas são previstas para serem secretadas pela bactéria (PIERETTI et al., 2009, 2012). Para identificar especificamente as enzimas envolvidas no processo de degradação da celulose vegetal, foram aplicados vários filtros de triagem. A primeira triagem foi realizada com o software de predição SignalP 6.0 para identificar as proteínas que possuem peptídeo sinal. No total, foram preditas 486 proteínas com peptídeo sinal (**Figura 6 e APÊNDICE A, Tabela A1**), ou seja, provavelmente essas proteínas serão secretadas para ambientes internos e externos da bactéria.

**Figura 6** - Diagrama de Venn ilustrando a quantidade de proteínas identificadas na análise do proteoma de *Xanthomonas albilineans* GPE PC73. As proteínas são categorizadas com base na presença de peptídeo sinal (roxo), localização extracelular (azul) e a possibilidade de serem CAZymes (laranja).



**Fonte:** Elaborada pelo autor.

Em seguida, foi necessário distinguir as proteínas destinadas aos ambientes intercelulares dos extracelulares. Para identificar apenas as proteínas que são secretadas para o ambiente externo da bactéria foi realizada uma segunda triagem com o software de predição de localização subcelular (extracelular, intracelular ou periplasmática) CELLO. Dentre as 3.208 proteínas que compõem o proteoma de *X. albilineans* GPE PC73, apenas 58 foram identificadas como possíveis enzimas com peptídeo sinal e localização extracelular (**Figura 6, APÊNDICE A, Tabela A1 e Tabela A2**). Determinar a localização subcelular de uma enzima é importante devido à função biológica que ela desempenha. Geralmente, as CAZymes secretadas para o espaço extracelular bacteriano têm como principal função degradar polissacarídeos complexos, convertendo-os em oligossacarídeos. Estes oligossacarídeos são transportados para o interior da bactéria por intermédio de transportadores de membrana específicos, tais como TonB, SusC, SusD, entre outros. Após internalizados, os oligossacarídeos são degradados por CAZymes intracelulares ou periplasmáticas e convertidos em monossacarídeos para serem utilizados em diversas funções metabólicas (PIERETTI et al., 2015; STROBEL; CALDWELL; DAWSON, 1995).

A terceira análise do proteoma, focada exclusivamente na identificação de enzimas ativas em carboidratos (CAZymes), foi realizada utilizando os softwares HMMER e DIAMOND, disponíveis no servidor dbCAN3. Esta análise resultou na identificação de 108 possíveis CAZymes (**Figura 6. APÊNDICE A, Tabela A3**). Dentre estas, 101 já estão catalogadas no banco de dados CAZy como enzimas pertencentes à *Xanthomonas albilineans* GPE PC73.

Das 108 CAZymes identificadas, 60 são classificadas como hidrolases glicosídicas (GHs), sete como esterases de carboidratos (CEs), uma como liase de polissacarídeo (PL), 31 como glicosiltransferases (GTs), cinco como módulos de ligação a carboidrato (CBMs) isolados e quatro como enzimas de atividades auxiliares (AAs).

O cruzamento de dados entre todas as análises do proteoma, peptídeo sinal, localização subcelular (extracelular) e dbCAN3, resultou no total de 10 CAZymes. Provavelmente, essas enzimas são secretadas pela bactéria para degradar a parede celular da planta. Entre elas, nove estão catalogadas no CAZy, com exceção de XALc\_1160, descrita como lipoproteína putativa e módulo de ligação ao carboidrato da família 50 (CBM50) isolado (**Tabela 1**).

**Tabela 1** - Enzimas ativas em carboidratos secretadas por *Xanthomonas albilineans* GPE PC73 que possuem localização extracelular.

Classe e Família	Identificador	Descrição funcional	Tamanho (aa)
GH6 + CBM2	XALc_0484	1,4-beta-celobiosidase	707
GH28	XALc_0811	Precursor de poligalacturonase	474
GH12 + CBM2	XALc_0865	Precursor de endoglucanase	500
CBM2	XALc_0874	Proteína hipotética	667
CBM50	XALc_1160	Lipoproteína putativa	267
GH23	XALc_1263	Transglicosilase lítica de mureína	315
CE8	XALc_1915	Pectina esterase	422
GH18	XALc_2183	Hidrolase glicosídica	434
GH5 + CBM2	XALc_2967	Endoglucanase extracelular	483
GH5 + CBM2	XALc_2969	Precursor de endoglucanase	549

**Fonte:** Elaborada pelo autor.

**Nota:** aa = aminoácidos

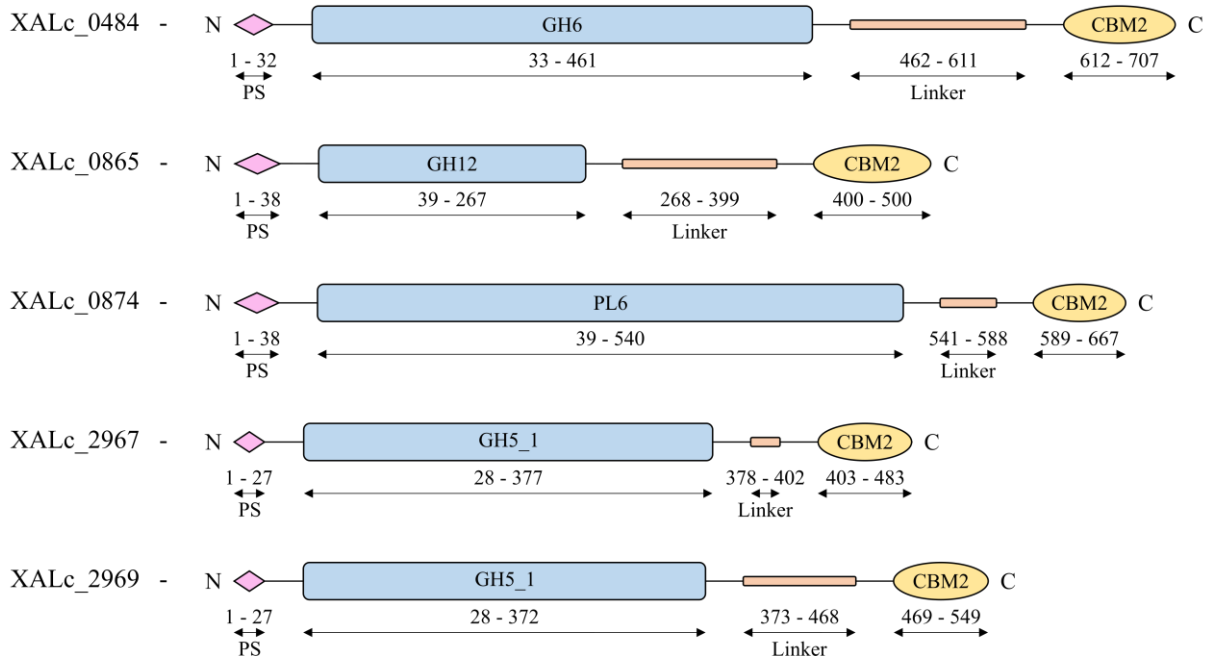
Conforme mencionado anteriormente, a atividade catalítica é um dos principais critérios para classificar as CAZymes em classes enzimáticas e famílias. Embora o número EC não seja o único critério para essa classificação, ele pode ser utilizado como parâmetro inicial de seleção. Segundo o CAZy, dentre as diversas atividades distribuídas entre as nove famílias enzimáticas identificadas (**Tabela 1**), somente as famílias GH6, GH12 e GH5 possuem CAZymes com ECs especificamente relacionados à hidrólise da celulose. A família GH6 inclui a endo- $\beta$ -1,4-glucanase (EC 3.2.1.4) e a  $\beta$ -1,4-celobiosidase/celobiohidrolase (EC 3.2.1.91). A família GH12 contém apenas a endo- $\beta$ -1,4-glucanase (EC 3.2.1.4). Já a família GH5 engloba a endo- $\beta$ -1,4-glucanase (EC 3.2.1.4), a  $\beta$ -glicosidase (EC 3.2.1.21), a exo- $\beta$ -1,4-glucanase/celodextrinase (EC 3.2.1.74) e a  $\beta$ -1,4-celobiosidase/celobiohidrolase (EC 3.2.1.91). Visto que os CBMs não possuem atividade catalítica, eles não recebem uma classificação numérica por um EC.

Contudo, é possível verificar que os módulos de ligação ao carboidrato (CBM) estão presentes na maioria das CAZymes identificadas na triagem do proteoma de *X. albilineans* GPE PC73 (**Tabela 1**). Em específico, os CBM2, que além de reconhecer e direcionar o domínio catalítico, desordenam a estrutura empacotada da celulose cristalina, tornando-a mais acessível para a hidrólise (BERNARDES et al., 2019; BORASTON et al., 2004). Visto a importância do CBM2 para o processo de hidrólise, juntamente com a classificação EC específica para

degradação de celulose, as cinco enzimas que pertencem as famílias CBM2, GH5, GH6 e GH12 foram selecionadas para as etapas posteriores de predição estrutural e estudos de docking molecular. Além disso, entre as 101 enzimas catalogadas no CAZy como pertencentes à *X. albilineans* GPE PC73, essas cinco selecionadas são as únicas representantes de suas famílias. É importante salientar que, normalmente, os aminoácidos que compõem os CBMs estão inseridos na sequência de uma proteína multidomínio, conectados ao domínio catalítico. São raras as exceções em que supostos CBMs são codificados de modo independente (GILBERT; KNOX; BORASTON, 2013). Portanto, é improvável que as anotações funcionais das enzimas XALc\_0874 (CBM2) e XALc\_1160 (CBM50) possam estar completas, visto que elas foram descritas apenas como módulos CBMs (**Tabela 1**).

Além dos domínios CBM2, as enzimas selecionadas (XALc\_0484, XALc\_0865, XALc\_0874, XALc\_2967 e XALc\_2969) possuem em sua sequência proteica domínios linker (**Figura 7**). Os domínios linker são regiões sequenciais que possuem comprimento variado, formadas por repetições de dois ou três resíduos, além de serem estruturalmente desordenados e possuírem alta flexibilidade (HU; OHUE, 2024; NAKAMURA et al., 2016, 2020; PAYNE et al., 2013; TOMPA, 2002). Por esse motivo, os domínios linker precisam ser considerados durante estudos de biologia estrutural, visto que podem interferir efetivamente na qualidade da estrutura tridimensional obtida. Os linkers das enzimas selecionadas de *X. albilineans* GPE PC73 são compostos por repetições de dois ou mais resíduos de prolina, treonina, serina e/ou glicina e estão entre os de maior comprimento no gênero *Xanthomonas* (**Figura 7**) (PIERETTI et al., 2012).

**Figura 7** - Composição multidomínio de cada enzima selecionada (XALc\_0484, XALc\_0865, XALc\_0874, XALc\_2967 e XALc\_2969). Os comprimentos são expressos em número de resíduos de aminoácidos que compõem o peptídeo sinal (PS) (losango rosa), domínio catalítico (retângulo azul), domínio linker (retângulo laranja) e domínio CBM2 (elipse amarela). N e C sinalizam as regiões N- e C-terminais, respectivamente.



**Fonte:** Elaborada pelo autor.

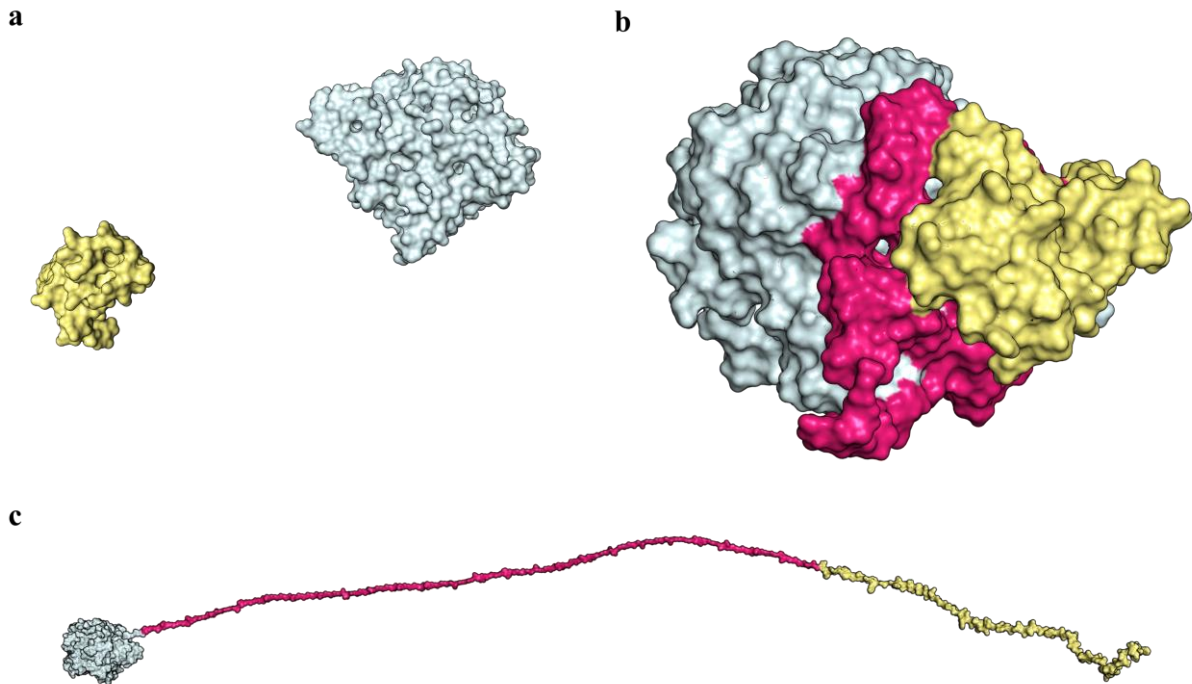
## 4.2 MODELAGEM MOLECULAR DAS CAZymes

Dentre os métodos de modelagem empregados para as cinco enzimas (XALc\_0484, XALc\_0865, XALc\_0874, XALc\_2967 e XALc\_2969), as melhores estruturas tridimensionais foram obtidas utilizando a última versão dos algoritmos RoseTTAFold e AlphaFold2, sendo executados pelas plataformas Robetta e ColabFold, respectivamente. Entre os algoritmos que utilizam *deep learning* para predição de modelos estruturais, o AlphaFold2 é considerado padrão ouro (BORDIN et al., 2023; JUMPER et al., 2021; JUMPER; HASSABIS, 2022).

Os modelos gerados pelos outros algoritmos, incluindo os utilizados pelos softwares, MODELLER, RosettaCM, I-Tasser, SWISS-MODEL e Robetta, apresentaram conformações inadequadas para as regiões dos domínios linker e dos domínios CBM2 (**Figura 8**). A razão para isso está nas proteínas multidomínios que ainda representam um desafio para os algoritmos de predição. Esse desafio é acentuado, especialmente, pela flexibilidade dos linkers que podem adotar diversas conformações. Quando um linker é predito incorretamente, isso pode interferir na conformação dos domínios conectados por ele, resultando em uma estrutura final da proteína que pode ser inadequada para análises futuras (KURKCUOGLU et al., 2012; ZHOU et al., 2019). Os algoritmos utilizados pelos softwares supracitados identificaram a região linker das

enzimas como uma estrutura mais rígida, de menor grau de liberdade, resultando em modelos de domínio único ou estendidos (**Figura 8**).

**Figura 8** - Representação estrutural de alguns modelos preditos erroneamente para a enzima XALc\_0484. O domínio catalítico está representado em ciano-claro, o linker em rosa-escuro e o CBM2 em amarelo-claro. (a) O servidor SWISS-MODEL não conseguiu modelar o domínio linker. (b) O servidor I-Tasser uniu os três domínios em um modelo de domínio único. (c) O software MODELLER não conseguiu enovelar corretamente o domínio CBM2, mantendo-o estendido.



**Fonte:** Elaborada pelo autor.

As melhores predições de RoseTTAFold e ColabFold, juntamente com as estruturas obtidas do banco de dados do AlphaFold, foram submetidas a diferentes métodos de avaliação. Esses métodos visam determinar a qualidade estrutural e selecionar as mais adequadas para os estudos de docking molecular. A primeira avaliação foi realizada com base na pontuação de confiança emitida por cada algoritmo no final da predição das estruturas. Tanto o ColabFold quanto o AlphaFold utilizam como parâmetro de qualidade as estimativas de confiança local (por resíduo) que geram um escore global (média dos valores de confiança local) de qualidade do modelo predito. Esta pontuação de confiança, denominada pLDDT, baseia-se na variação estrutural e distância interatômica em relação à plausibilidade estereoquímica. A escala de pontuação pLDDT varia de 0 - 100 pontos, sendo: > 90 (alta precisão), 70 - 90 (adequado), 50 - 70 (cautela) e < 50 baixa confiança, ou pertencem a regiões com alta desordem estrutural e provável flexibilidade demasiada (JUMPER et al., 2021; JUMPER; HASSABIS, 2022; TUNYASUVUNAKOOL et al., 2021).

RoseTTAFold utiliza a pontuação de confiança IDDT, o qual segue os mesmos princípios básicos de pLDDT, sendo equivalentes (JUMPER et al., 2021; MARIANI et al., 2013). Portanto, os escores obtidos por ambos os algoritmos, pLDDT e IDDT, podem ser comparados com o objetivo de avaliar preliminarmente a qualidade das estruturas preditas por cada algoritmo.

Um fator relevante nesse contexto é a presença do peptídeo sinal na proteína. A sua conformação interfere tanto na predição quanto na avaliação estrutural. Por este motivo, os resíduos que compõem os peptídeos sinais (**Figura 7**) foram removidos das sequências de aminoácidos antes de serem submetidos à predição pelo RoseTTAFold e ColabFold. Contudo, as estruturas disponíveis no banco de dados do AlphaFold possuem o peptídeo sinal e, sendo assim, foi necessário efetuar o cálculo de pLDDT manualmente, ou seja, desconsiderando os resíduos que compõem o peptídeo sinal (**Tabela 2**).

**Tabela 2** - Pontuação de confiança global pLDDT (AlphaFold e ColabFold) e IDDT (RoseTTAFold), determinada por cada algoritmo de predição estrutural. Em negrito os maiores escores para cada CAZyme.

CAZyme	AlphaFold	ColabFold	RoseTTAFold
XALc_0484	80,60	<b>82,21</b>	69,21
XALc_0865	73,28	<b>76,68</b>	55,25
XALc_0874	86,50	<b>86,82</b>	65,48
XALc_2967	92,30	<b>92,62</b>	80,71
XALc_2969	85,43	<b>86,13</b>	70,75

**Fonte:** Elaborada pelo autor.

A comparação entre as pontuações revela que ColabFold foi o melhor dos preditores utilizados para todas as estruturas proteicas modeladas (**Tabela 2**). A diferença mínima de escore encontrada entre as estruturas obtidas no banco do AlphaFold e as preditas pelo ColabFold é justificada por compartilharem do mesmo algoritmo base (AlphaFold2) (MIRDITA et al., 2022; VARADI et al., 2022). Com relação a RoseTTAFold, foi encontrada uma diferença significativa nos valores de pontuação, quando comparado aos demais métodos preditivos (**Tabela 2**). Mesmo sendo metodologias de pontuação equivalentes, o cálculo adotado por cada algoritmo pode ter influenciado no valor final. Levando isso em consideração, foi realizada uma nova avaliação de qualidade dos modelos utilizando-se dois softwares

independentes aos algoritmos preditores. Os softwares utilizados foram o QMEANDisCo e o MolProbity, disponíveis no servidor Swiss-Model (<https://swissmodel.expasy.org/assess>). É importante destacar que as estruturas avaliadas nesses dois últimos softwares estavam sem o peptídeo sinal.

A pontuação de qualidade global do QMEANDisCo é representada na escala entre 0,0 (ruim) e 1,0 (bom). Considerando a margem de erro de  $\pm 0,05$  adotada pelo software, com exceção para as enzimas XALc\_0484 e XALc\_0865, não houve diferença de escore entre os três métodos de predição (**Tabela 3**). Conforme o esperado, AlphaFold e ColabFold obtiveram escores semelhantes. Em relação a RoseTTAFold, os resultados foram melhores em comparação aos determinados pela pontuação interna de seu preditor (**Tabela 2**). Três (XALc\_0484, XALc\_0865 e XALc\_2969) das cinco enzimas preditas utilizando o RoseTTAFold tiveram os melhores valores de qualidade entre os preditores (**Tabela 3**).

**Tabela 3** - Análise de qualidade global, realizada pelo software QMEANDisCo, entre as estruturas obtidas pelos três métodos de predição. Em negrito as maiores pontuações para cada CAZyme.

CAZyme	AlphaFold	ColabFold	RoseTTAFold
XALc_0484	0,69	0,70	<b>0,76</b>
XALc_0865	0,53	0,54	<b>0,64</b>
XALc_0874	<b>0,58</b>	0,57	0,55
XALc_2967	0,76	<b>0,77</b>	0,76
XALc_2969	0,71	0,71	<b>0,72</b>

**Fonte:** Elaborada pelo autor.

O QMEANDisCo é um dos melhores softwares utilizados para avaliar a qualidade de modelos estruturais preditos computacionalmente (ROBIN et al., 2021). A qualidade dos modelos é determinada a partir de pontuações globais e locais (por resíduo de aminoácido). A pontuação global da estrutura é calculada por meio de redes neurais que combinam os métodos estatísticos QMEAN e DisCo. Ao prever estimativas de qualidade global, que são obtidas pela média das pontuações locais, o método QMEAN se mostra adequado para a avaliação de modelos individuais (BENKERT; TOSATTO; SCHOMBURG, 2008). Por sua vez, o método DisCo realiza previsões apenas da qualidade local, por intermédio da estimativa de distância entre os resíduos e as compara com informações de estruturas experimentais homólogas à

avaliada (STUDER et al., 2020). Os resultados desses métodos são combinados pelas redes neurais para obtenção do IDDT local, sendo, a média deste, definida como a pontuação global (STUDER et al., 2020; WATERHOUSE et al., 2018).

O MolProbity é um avaliador de estruturas tridimensionais amplamente reconhecido e utilizado, principalmente, para avaliar a qualidade de estruturas experimentais. Assim como o QMEANDisCo, o MolProbity determina a qualidade dos modelos avaliados por meio de pontuações ao nível local e global (WILLIAMS et al., 2018). A qualidade estereoquímica das estruturas é avaliada através da análise de contato dos átomos, possibilitando a identificação de diversos problemas estéricos, valores extremos dos ângulos e ligações incoerentes. A ferramenta utiliza um conjunto de algoritmos baseados em física e conhecimento empírico para efetuar as avaliações (WILLIAMS et al., 2018). Apesar de o MolProbity ser capaz de realizar diversos tipos de análises, a versão disponibilizada pelo Swiss-Model inclui apenas algumas análises mais relevantes (CHEN et al., 2010; WILLIAMS et al., 2018). Estas incluem: MolProbity Score, uma pontuação global que reflete a qualidade da estrutura obtida pela combinação de todas as pontuações geométricas; Clash Score, para avaliar e classificar estatisticamente o número de sobreposições estéricas de 0,4 Å (ou mais) entre átomos que não possuem ligação direta (CHEN et al., 2010; WILLIAMS et al., 2018); Bad Bonds e Bad Angles, que se referem a avaliações de geometria para verificar o comprimento e ângulo de ligações covalentes em relação ao conhecimento pré-determinado para estruturas de proteínas (ENGH; HUBER, 1991, 2006); e C-Beta Deviations, que se refere à validação da geometria estrutural, analisando a posição ideal do átomo C $\beta$  dos resíduos que compõe a cadeia principal (backbone) da proteína. Qualquer desvio identificado na estrutura pode ser considerado um erro de geometria (CHEN et al., 2010; WILLIAMS et al., 2018).

Os resultados das avaliações realizadas pelo MolProbity identificaram problemas estruturais em todos os modelos. No entanto, e contrariando os resultados de QMEANDisCo (**Tabela 3**), as estruturas preditas por RoseTTAFold apresentaram divergências estereoquímicas que, provavelmente, inviabilizariam a sua utilização (**Tabela 4**). O principal parâmetro negativo identificado nessas estruturas está relacionado com as altas taxas de clashes, indicando inúmeras sobreposições atômicas irregulares. O clashe score classifica estatisticamente o número de sobreposições estéricas de no mínimo 0,4 Å entre átomos que não possuem ligação direta (CHEN et al., 2010; WILLIAMS et al., 2018). Dada a alta incidência de clashes nas estruturas previstas pelo RoseTTAFold, é necessário um exame mais detalhado, pois a qualidade da estrutura é diretamente proporcional ao menor número de clashes (WILLIAMS et al., 2018).

**Tabela 4** - Análise de qualidade local, realizada pelo software MolProbity, entre as estruturas obtidas pelos três métodos de predição. Em negrito os melhores valores para cada CAZyme. (continua)

<b>XALc_0484</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	1,72	<b>1,71</b>	2,99
Clash Score	0	0,87	<b>0,77</b>	160,35
C-Beta Deviations	0	1	8	<b>0</b>
Bad Bonds	0	<b>0 / 4.833</b>	<b>0 / 4833</b>	8 / 4.833
Bad Angles	0	117 / 6.582	128 / 6582	<b>12 / 6.582</b>
<b>XALc_0865</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	2,21	<b>1,87</b>	2,99
Clash Score	0	1,48	<b>1,32</b>	135,17
C-Beta Deviations	0	3	4	<b>0</b>
Bad Bonds	0	<b>0 / 3.277</b>	<b>0 / 3.277</b>	5 / 3.277
Bad Angles	0	84 / 4.448	97 / 4.448	<b>7 / 4.448</b>
<b>XALc_0874</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	<b>1,46</b>	1,56	3,14
Clash Score	0	<b>2,07</b>	2,3	156,96
C-Beta Deviations	0	3	3	<b>0</b>
Bad Bonds	0	<b>0 / 4.574</b>	<b>0 / 4.574</b>	1 / 4.573
Bad Angles	0	36 / 6.232	47 / 6.232	<b>15 / 6.230</b>

(conclusão)

XALc_2967				
Parâmetro	Ideal	AlphaFold	ColabFold	RoseTTAFold
MolProbity Score	< melhor	<b>1,42</b>	1,43	2,95
Clash Score	0	<b>1,76</b>	1,91	165,05
C-Beta Deviations	0	<b>0</b>	1	<b>0</b>
Bad Bonds	0	<b>0 / 3.618</b>	<b>0 / 3.618</b>	3 / 3.618
Bad Angles	0	24 / 4.942	16 / 4.942	<b>9 / 4.942</b>

XALc_2969				
Parâmetro	Ideal	AlphaFold	ColabFold	RoseTTAFold
MolProbity Score	< melhor	1,74	<b>1,44</b>	2,78
Clash Score	0	1,79	<b>1,37</b>	166,96
C-Beta Deviations	0	6	<b>0</b>	<b>0</b>
Bad Bonds	0	<b>0 / 3.900</b>	6 / 3.900	9 / 3.900
Bad Angles	0	80 / 5.306	62 / 5.306	<b>9 / 5.306</b>

**Fonte:** Elaborada pelo autor.

A maior parte dos erros estruturais identificados pelo MolProbity (valores em desacordo com os ideais, **Tabela 4**), novamente estão associados aos domínios linkers. Mesmo possuindo conformação visualmente adequada, provavelmente, à baixa variabilidade dos resíduos que compõem essas regiões interferem na avaliação do modelo final. A organização desses resíduos raramente resultaria em uma estrutura secundária de fita- $\beta$  ou  $\alpha$ -hélice, o que torna difícil para os algoritmos baseados em *deep learning* preverem a conformação adequada, uma vez que foram treinados com base em estruturas experimentais obtidas no PDB. Outro fator relevante que interfere na predição correta é o tamanho considerável do linker dessas enzimas (BAEK et al., 2021; JUMPER et al., 2021; MIRDITA et al., 2022). Por exemplo, a sequência de aminoácidos da enzima XALc\_0484 é composta por um total de 707 resíduos, dos quais, somente o domínio linker é formado por 150 (posições 462 - 611) desses resíduos (**Figura 7** e **Figura 9**), sendo o de maior comprimento entre as cinco CAZymes analisadas. Domínios linker

com esse tamanho não são frequentemente encontrados no PDB. Isso reduz o volume de dados utilizado durante o treinamento desses algoritmos preditivos, resultando na baixa qualidade conformacional desses domínios (GUO et al., 2022; SCARDINO; DI FILIPPO; CAVASOTTO, 2023; TUNYASUVUNAKOOL et al., 2021; URBAN; POMPON, 2022). É importante ressaltar que uma baixa pontuação em uma estrutura predita computacionalmente não, necessariamente, indica uma conformação incorreta, principalmente em domínios linkers longos como o de XALc\_0484. Em alguns casos, uma baixa pontuação de qualidade pode indicar uma região desordenada e flexível, exigindo uma avaliação mais cuidadosa do modelo (TOMPA, 2002; TUNYASUVUNAKOOL et al., 2021).

**Figura 9** - Sequência dos resíduos que compõem a enzima XALc\_0484, destacado em laranja a região do domínio linker, entre os resíduos 462 - 611, composta majoritariamente por glicinas.

XALc\_0484

```

1 MKTSKSTGNRLKYKATASLLSVCLLSLSAHAQSRVDNPFVVGATSYVNP 50
51 YTVDIDSSIAKVSDPALIAKMQTAKTYPTFIWLDSIGAIYGGNRNAGLLG 100
101 LRDHLNSALAQKTAGQTMLVGLVIYDMPGRDCHALSSNGELPLTQDALAR 150
151 YKTEYIDVIAGILADPQYKDLRIVNVVEPDSLPLNLVTNLSTPACGQASST 200
201 GIYEAAIKYAINKLHAIPTNYNYLDIAHSGWLGWPNMMDKTIPLYTRIFQ 250
251 DTTAGLASIDGFATNTANTTPTTEPNLPNSDMNVGGQPLRSSTFYQWNPI 300
301 FDENHYVQTLYNNFVSAGWSSQIGFLIDTSRNGWGGPNRPTSASGNDVNT 350
351 YVNSGRVDQRLARGNWCNPSGAGMGAPPTTAPLPHVDAFVHIKEPGVSDG 400
401 SSRYIPNDQKGADGMCDERYTTPDGVLGTGALPNSPISGAWFHNQFVMLV 450
451 QNAYPAIPGSTSGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGG 500
501 STGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGST 550
551 GGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGGGGSTGG 600
601 GGSTGGGGSTANPFDVQYTVNSDWGAGYCETISVTNKRGSKAQWNITVP 650
651 ILGKVSSSNINLWSGMWTNPNTDASGAVQSLQVSGMDWNRVYVDPQQTTTLG 700
701 FCATRSN 707

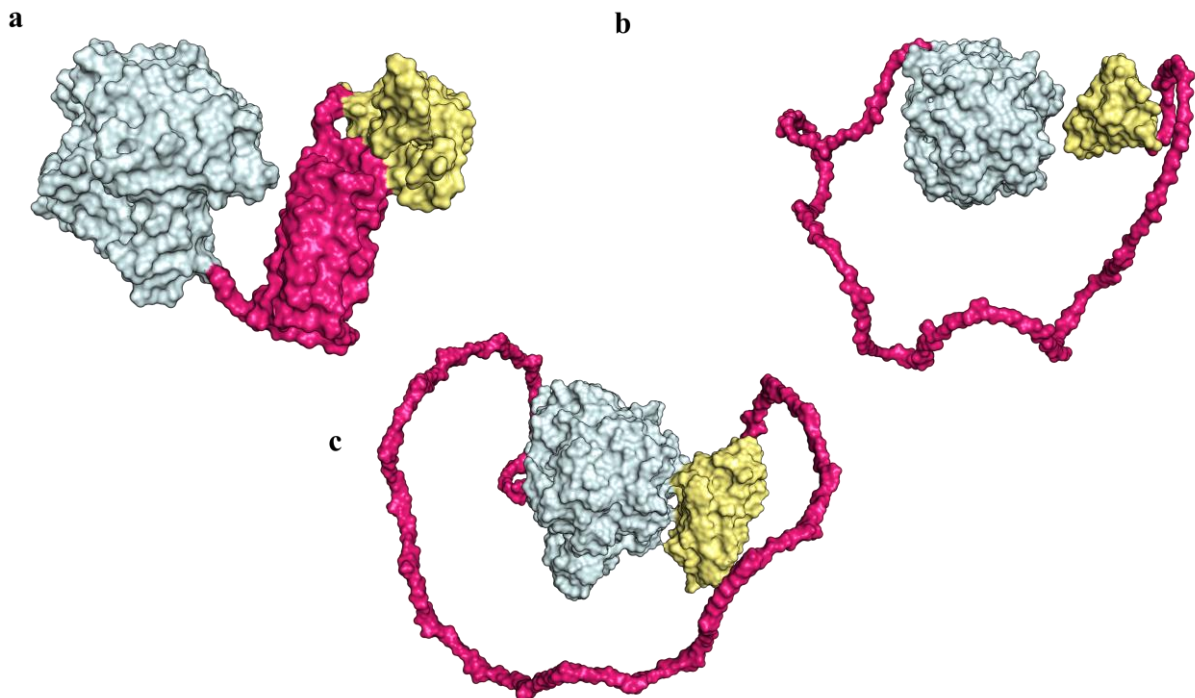
```

**Fonte:** Elaborada pelo autor.

Os erros de predição conformacional dos domínios linker podem ser visualmente verificados, por exemplo, na enzima XALc\_0484 (**Figura 10**). Os linkers precisam ser flexíveis, visto que, enquanto a enzima se mantém fixada ao substrato, por intermédio do CBM, o linker se estende para aumentar a área de atuação do domínio catalítico (COURTADE et al., 2018). Em XALc\_0484, a predição da região linker pelo RoseTTAFold resultou em uma estrutura mais ordenada e menos flexível. Essa região, modelada erroneamente, colaborou para aumentar tanto os valores do escore global, quanto os índices de clashes (**Figura 10a** e **Figura 11c**). Por outro lado, a estrutura obtida no banco de dados do AlphaFold possui o domínio linker

em conformação correta e aparentemente flexível, além de manter as regiões catalítica e CBM2, separadas (**Figura 10b**). O modelo predito pelo ColabFold apresenta um problema recorrente entre as demais enzimas preditas por ele. Pela avaliação visual, o domínio linker aparenta estar estruturalmente correto, entretanto, as regiões catalítica e CBM2, estão conectadas ao invés de separadas e independentes (**Figura 10c**).

**Figura 10** - Representação estrutural da enzima XALc\_0484 predita pelos algoritmos (a) RoseTTAFold, (b) AlphaFold e (c) ColabFold. O domínio catalítico está representado em ciano-claro, o linker em rosa-escuro e o CBM2 em amarelo-claro.

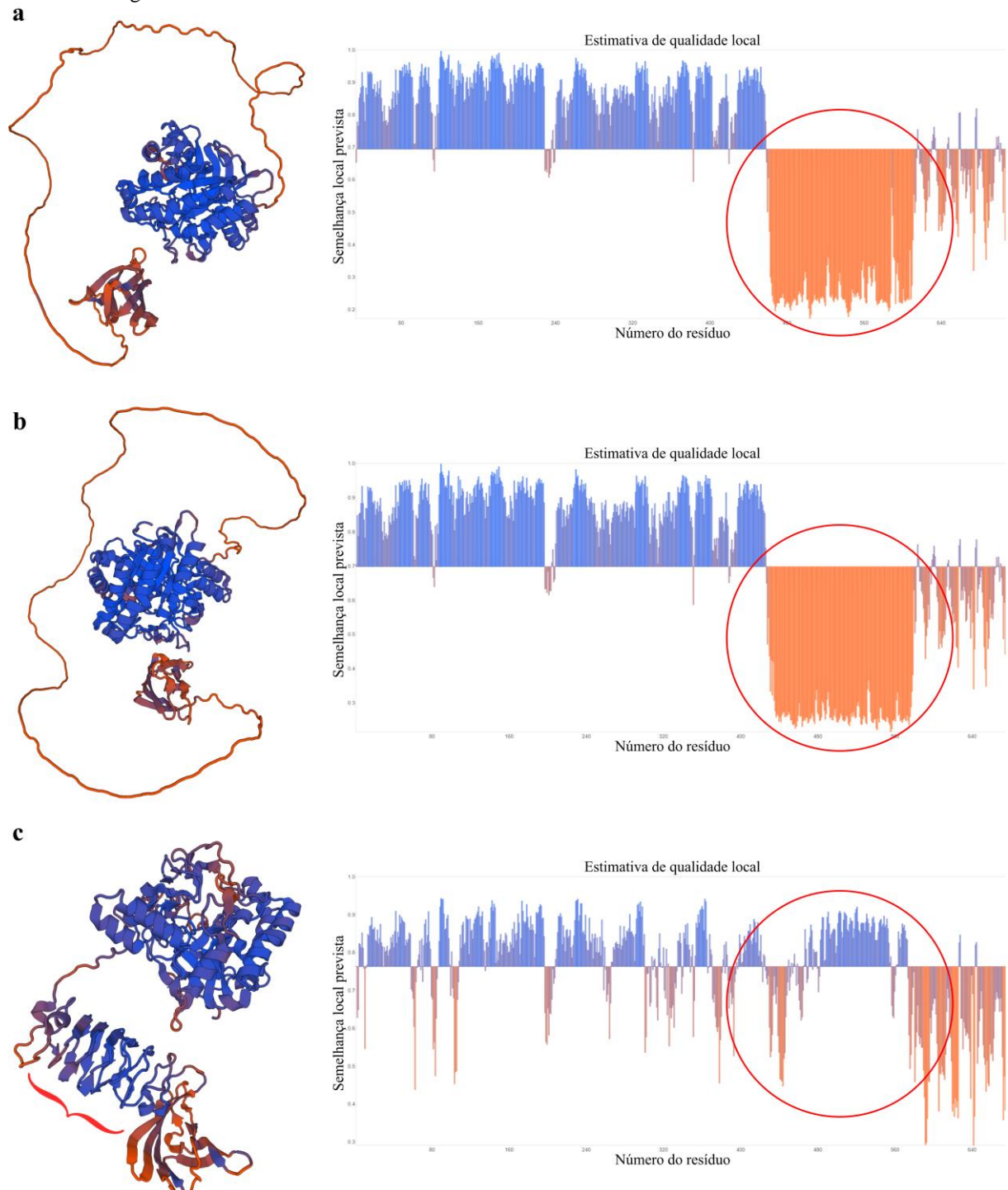


**Fonte:** Elaborada pelo autor.

Além da pontuação de qualidade global, o QMEANDisCo também disponibiliza graficamente o nível de qualidade local para cada resíduo. Visualmente, o resultado de confiabilidade local do QMEANDisCo é representado por uma variação de cores, do laranja para o azul, indicando do pior para o melhor, respectivamente (**APÊNDICE B, Figura A1**). Pelo gráfico da **Figura 11c** é possível observar que o domínio linker da enzima XALc\_0484, modelado pelo RoseTTAFold, apresenta um bom escore de qualidade local e, conseqüentemente, contribuiu para o aumento da pontuação global (**Tabela 3**). Porém, conforme relatado anteriormente, a conformação da estrutura não está correta. Esta região deveria ser formada por uma longa alça, ao invés das fitas- $\beta$  e  $\alpha$ -hélices preditas pelo RoseTTAFold (**Figura 11c, colchete vermelho**). Os gráficos gerados pelo QMEANDisCo, a partir dos modelos preditos pelo ColabFold e pelo AlphaFold, refletem praticamente o mesmo

resultado, consistente com a similaridade do escore global (**Tabela 4**). A região do domínio linker de XALc\_0484, modelada por ambos os softwares, apresentou uma pontuação de baixa qualidade. Mesmo assim, pela análise visual, é possível verificar que a conformação em alça do domínio linker foi predita corretamente pelos algoritmos (**Figura 11a e b**).

**Figura 11** - Representação estrutural e gráfica de qualidade local dos modelos preditos pelos algoritmos (a) AlphaFold, (b) ColabFold e (c) RoseTTAFold para a enzima XALc\_0484. O círculo vermelho nos gráficos identifica a região linker e o colchete vermelho na estrutura de RoseTTAFold indica o domínio linker do modelo.



**Fonte:** Elaborada pelo autor.

O mesmo “padrão” de modelagem do domínio linker de XALc\_0484 foi repetido pelo RoseTTAFold nas demais enzimas. Apesar disso, é possível que a região de interesse – nesse caso, o domínio catalítico – esteja adequadamente modelada. Portanto, os domínios catalíticos de cada enzima (**Figura 7**) foram isolados manualmente, utilizando-se o software PyMOL, e submetidos a uma nova análise de qualidade estrutural com os softwares QMEANDisCo e MolProbity (**Tabela 5** e **Tabela 6**).

**Tabela 5** - Análise de qualidade global realizada pelo software QMEANDisCo entre os domínios catalíticos das estruturas obtidas pelos três métodos de predição. Em negrito as maiores pontuações para cada CAZyme.

CAZyme	AlphaFold	ColabFold	RoseTTAFold
XALc_0484 (GH6)	<b>0,88</b>	<b>0,88</b>	0,81
XALc_0865 (GH12)	0,65	<b>0,66</b>	<b>0,66</b>
XALc_0874 (PL6)	<b>0,61</b>	<b>0,61</b>	<b>0,61</b>
XALc_2967 (GH5)	<b>0,85</b>	<b>0,85</b>	0,82
XALc_2969 (GH5)	<b>0,87</b>	0,86	<b>0,82</b>

**Fonte:** Elaborada pelo autor.

Conforme esperado, os escores globais e locais dos domínios catalíticos, avaliados pelo QMEANDisCo (**Tabela 5** e **APÊNDICE B, Figura A2**), foram superiores aos escores das estruturas integrais, que incluem todos os domínios (**Tabela 3**). Isso decorre das características mencionadas da região linker, resultando em uma avaliação de baixa qualidade pelos softwares. Esses valores negativos, quando somados aos cálculos para a obtenção da pontuação global, podem influenciar no julgamento e seleção das estruturas, exigindo uma análise mais detalhada, inclusive visual. Contudo, segundo os resultados obtidos pelo software MolProbity, os modelos preditos pelo RoseTTAFold continuam apresentando altas incidências de clashes (**Tabela 6**). Por esse motivo, esses modelos foram excluídos das análises subsequentes.

**Tabela 6** - Análise de qualidade local realizada pelo software MolProbity entre os domínios catalíticos das estruturas obtidas pelos três métodos de predição. Em negrito os melhores valores para cada CAZyme. (continua)

<b>XALc_0484_GH6</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	1,00	<b>0,92</b>	2,88
Clash Score	0	0,78	<b>0,47</b>	175,86
C-Beta Deviations	0	1	1	<b>0</b>
Bad Bonds	0	<b>0 / 3.345</b>	<b>0 / 3.345</b>	8 / 3.345
Bad Angles	0	12 / 4.584	14 / 4.584	<b>11 / 4.584</b>

<b>XALc_0865_GH12</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	0,95	<b>0,88</b>	2,96
Clash Score	0	0,88	<b>0,29</b>	144,13
C-Beta Deviations	0	<b>0</b>	<b>0</b>	<b>0</b>
Bad Bonds	0	<b>0 / 1.840</b>	<b>0 / 1.840</b>	3 / 1.840
Bad Angles	0	7 / 2.517	13 / 2.517	<b>6 / 2.517</b>

<b>XALc_0874_CBM2</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	<b>1,27</b>	1,31	3,160
Clash Score	0	<b>1,84</b>	1,98	161,90
C-Beta Deviations	0	<b>0</b>	1	<b>0</b>
Bad Bonds	0	<b>0 / 3.701</b>	<b>0 / 3.701</b>	<b>0 / 3.700</b>
Bad Angles	0	19 / 5.051	22 / 5.051	<b>14 / 5.049</b>

(conclusão)

<b>XALc_2967_GH5</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	<b>1,25</b>	1,27	2,95
Clash Score	0	<b>1,86</b>	<b>1,86</b>	171,39
C-Beta Deviations	0	<b>0</b>	<b>0</b>	<b>0</b>
Bad Bonds	0	<b>0 / 2.838</b>	<b>0 / 2.838</b>	2 / 2.838
Bad Angles	0	12 / 3.878	11 / 3.878	<b>8 / 3.878</b>

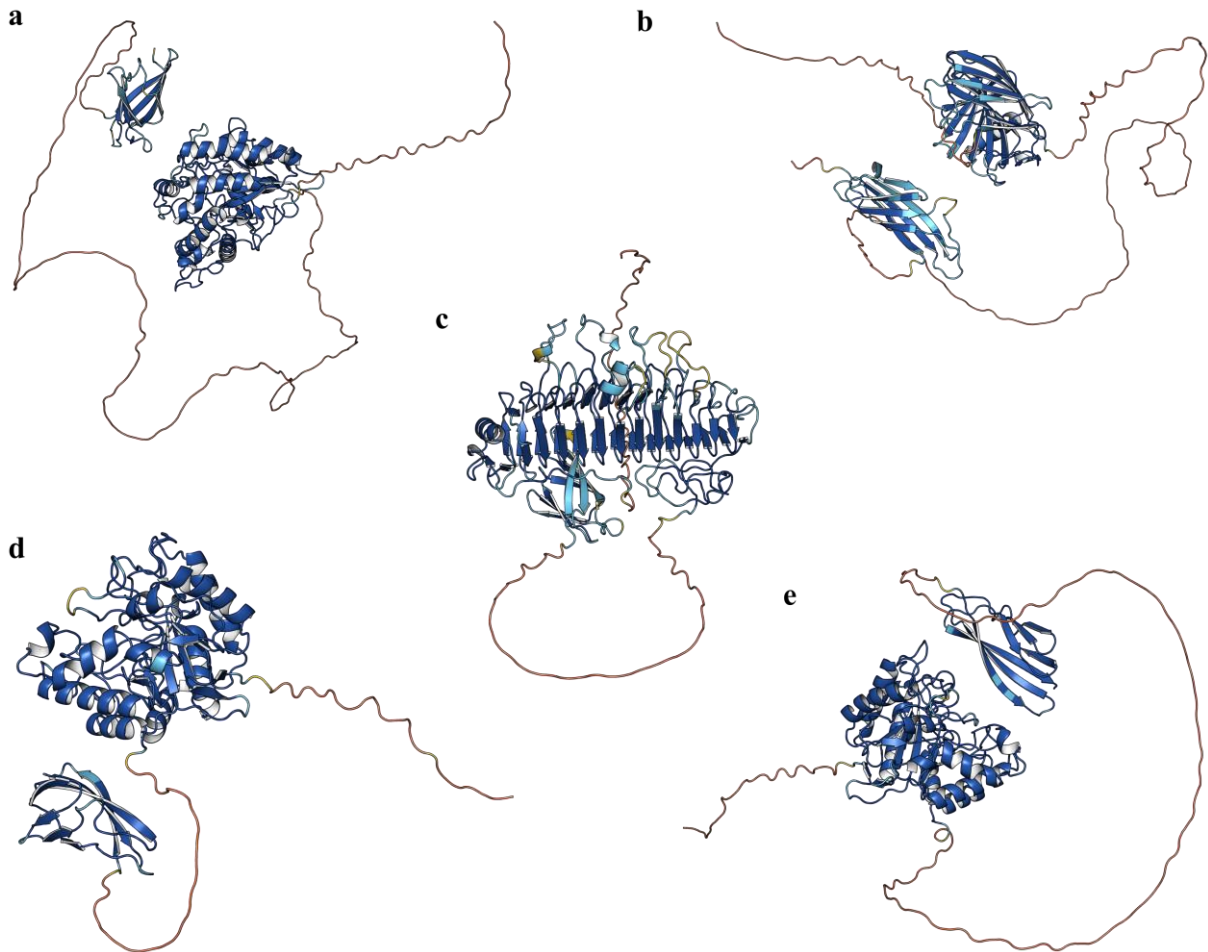
<b>XALc_2969_GH5</b>				
<b>Parâmetro</b>	<b>Ideal</b>	<b>AlphaFold</b>	<b>ColabFold</b>	<b>RoseTTAFold</b>
MolProbity Score	< melhor	<b>1,21</b>	1,39	2,84
Clash Score	0	1,91	<b>1,72</b>	183,45
C-Beta Deviations	0	<b>0</b>	<b>0</b>	<b>0</b>
Bad Bonds	0	<b>0 / 2.799</b>	6 / 2.799	8 / 2.799
Bad Angles	0	11 / 3.824	25 / 3.824	<b>7 / 3.824</b>

**Fonte:** Elaborada pelo autor.

Novamente, os modelos preditos pelo ColabFold e pelo AlphaFold apresentaram pontuações muito semelhantes quando avaliados pelos softwares QMEANDisCo e MolProbity (**Tabela 5** e **Tabela 6**). No entanto, exceto para a enzima XALc\_2967, o problema de união dos domínios catalítico e CBM2, identificado anteriormente na inspeção visual de XALc\_0484 (**Figura 10c**), persiste nos demais modelos preditos pelo ColabFold (**dados não mostrados**). Isso continua sendo um obstáculo para selecionar essas estruturas para as próximas análises, portanto, os modelos preditos pelo ColabFold também foram descartados.

Por fim, entre todos os métodos utilizados para obter estruturas tridimensionais viáveis para cada CAZyme, os modelos obtidos diretamente do banco de dados do AlphaFold (**Figura 12**) foram selecionados para serem utilizados nas próximas análises.

**Figura 12** - Visualização estrutural em modo “cartoon” das cinco enzimas obtidas do banco de dados do AlphaFold, incluindo o peptídeo sinal. (a) XALc\_0484, (b) XALc\_0865, (c) XALc\_0874, (d) XALc\_2967 e (e) XALc\_02969. A coloração está representada conforme a pontuação de confiança pLDDT, sendo: azul = alta precisão (> 90), ciano = adequado (70 - 90), amarelo = cautela (50 - 70) e laranja = baixa confiança (< 50).



**Fonte:** Elaborada pelo autor.

As divergências estruturais identificadas nos modelos preditos pelos três algoritmos, conforme observados nos resultados das análises de qualidade local e global, bem como na inspeção visual, destacam a importância de analisar-se a sequência de resíduos, o processo de enovelamento proteico e a conformação final da proteína, visto que, a sua função biológica é determinada por sua estrutura tridimensional (BOUATTA; SORGER; ALQURAISHI, 2021; HOU et al., 2022; KELZ et al., 2022; RICHARDSON, 1981).

#### 4.3 ANÁLISE COMPARATIVA DAS ESTRUTURAS TRIDIMENSIONAIS

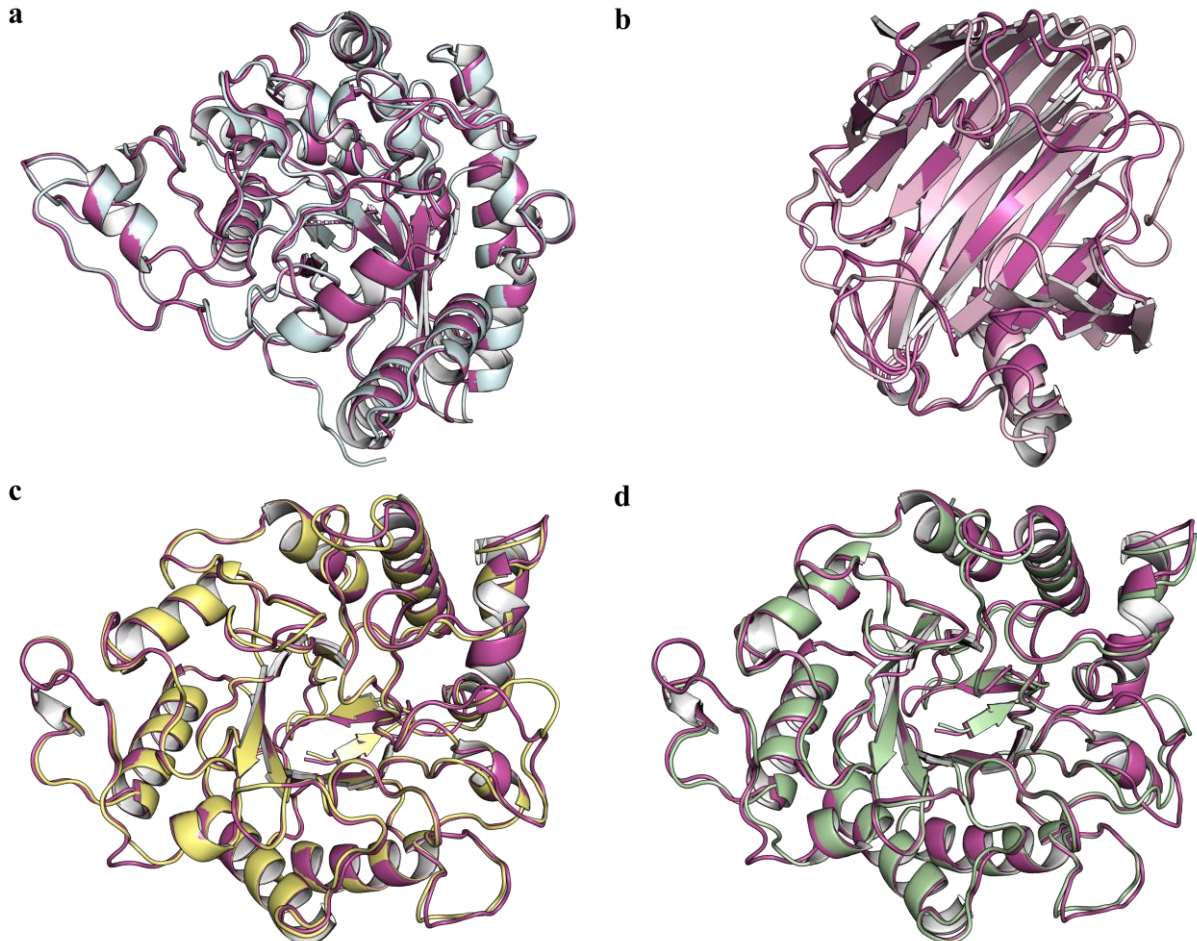
Após a análise visual, os cinco modelos selecionados foram submetidos a uma comparação com estruturas experimentais de alta qualidade. A sobreposição proteica, mais sensível que o alinhamento de sequências tradicional (MAHLICH et al., 2018), permite

identificar proteínas homólogas e uma melhor inferência de características estruturais, funcionais e evolutivas para as proteínas de interesse (ILLERGÅRD; ARDELL; ELOFSSON, 2009). A alta qualidade das estruturas experimentais disponíveis nos bancos de dados possibilita avaliar a precisão das estruturas preditas computacionalmente, como as geradas pelo AlphaFold2. Além disso, fornece informações detalhadas sobre a proteína de interesse. Devido a sua importância, o alinhamento por sobreposição estrutural foi realizado para confirmar as famílias e classes enzimáticas das CAZymes selecionadas (**Tabela 1**).

Utilizando o software FoldSeek (VAN KEMPEN et al., 2023), foi selecionado no PDB as estruturas experimentais que apresentam maior semelhança estrutural com os domínios catalíticos das cinco enzimas de *X. albilineans* GPE PC73. O FoldSeek é um software que revolucionou os métodos de busca por proteínas homólogas (VAN KEMPEN et al., 2023). Esse software realiza uma busca estrutural de proteínas, comparando a estrutura de interesse com as proteínas depositadas nos bancos de dados. Para isso, o software analisa e compara as interações terciárias entre os aminoácidos das estruturas. Entre os indicadores de resultados do FoldSeek, o TM-score avalia a similaridade estrutural entre duas proteínas. Valores de TM-score próximos a 1 indicam alta similaridade entre as proteínas (VAN KEMPEN et al., 2023; ZHANG; SKOLNICK, 2004).

Com base nos resultados obtidos pelo FoldSeek, a sobreposição foi realizada por meio do software PyMOL. A enzima XALc\_0484 foi sobreposta com a enzima CbsA (TM-score = 0,986 e probabilidade = 1,00), uma celobiohidrolase da família GH6 de *Xanthomonas oryzae* pv. *oryzae* (PDB: 5xyh, resolução; 1,86 Å) (**Figura 13a**) (TAYI et al., 2018). A sobreposição de XALc\_0865 foi realizada com Cel12A (TM-score = 0,842 e probabilidade = 0,99), uma endoglucanase GH12 de *Rhodothermus marinus* (PDB: 2bw8, resolução; 1,54 Å) (**Figura 13b**) (CRENNELL et al., 2006). Devido à semelhança estrutural entre si, tanto XALc\_2967, quanto XALc\_2969, foram sobrepostas com a enzima XccCel5A (TM-score = 0,980 e probabilidade = 1,0 e, TM-score = 0,984 e probabilidade = 1,0, respectivamente), uma endoglucanase da família GH5\_1 de *Xanthomonas campestris* pv. *campestris* (PDB:4tuf\_C, resolução; 2,70 Å) (**Figura 13c e d**) (PUHL et al., 2019). A partir dos resultados obtidos pelo FoldSeek e pelas sobreposições estruturais, é possível inferir que o domínio catalítico de quatro das cinco enzimas pertence às famílias e classes em que foram descritas: XALc\_0484 (GH6), XALc\_0865 (GH12), XALc\_2967 e XALc\_2969 (GH5) (**Tabela 1**).

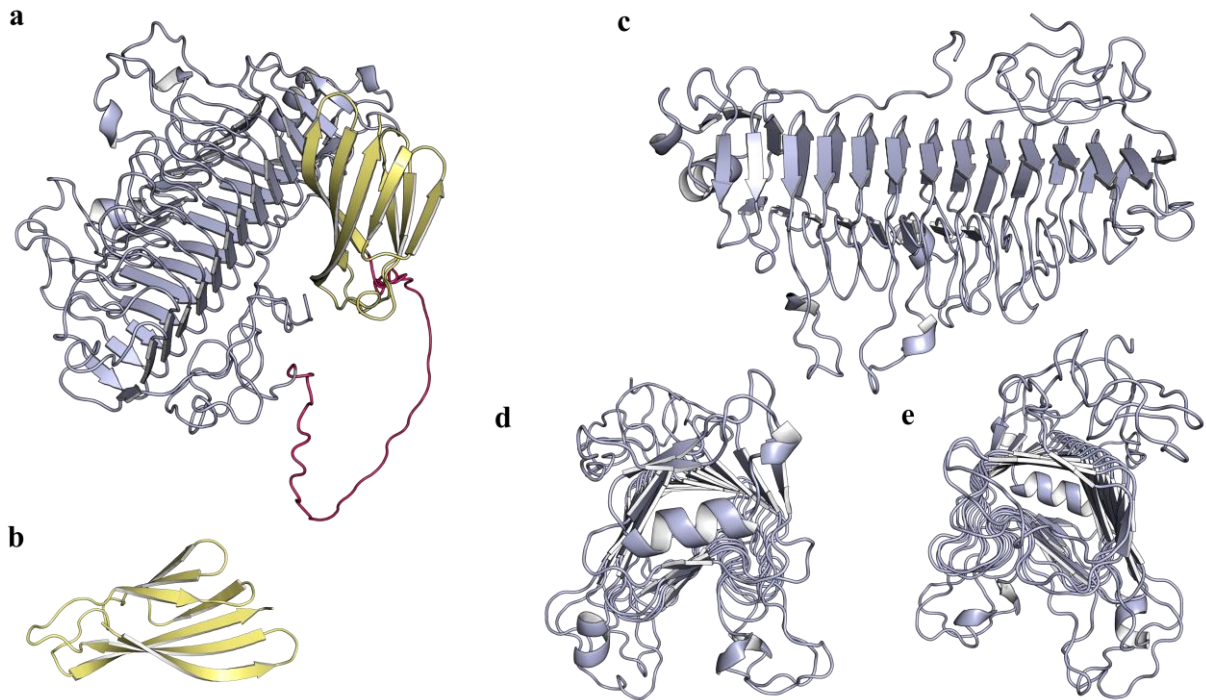
**Figura 13** - Sobreposição estrutural entre os domínios catalíticos dos modelos obtidos do banco de dados do AlphaFold, com as estruturas experimentais homólogas identificadas pelo FoldSeek (magenta). **(a)** XALc\_0484 (ciano-claro) sobreposta com CbsA (PDB: 5xyh) de *Xanthomonas oryzae* pv. *oryzae*. **(b)** XALc\_0865 (rosa-claro) sobreposta com Cel12A (PDB: 2bw8) de *Rhodothermus marinus*. **(c)** XALc\_2967 (amarelo-claro) e **(d)** XALc\_2969 (verde-claro), ambas sobrepostas com XccCel5A (PDB:4tuf\_C) de *Xanthomonas campestris* pv. *campestris*.



**Fonte:** Elaborada pelo autor.

A análise visual da enzima XALc\_0874 revelou que esta não é composta somente por um domínio da família CBM2, conforme descrito pelas anotações funcionais disponíveis nos bancos de dados (**Tabela 1** e **Figura 14a** e **b**). Semelhante às outras quatro enzimas, XALc\_0874 possui um domínio CBM2 com conformação sanduíche- $\beta$  clássico, um domínio linker e um domínio catalítico (**Figura 14a** e **b**). O domínio catalítico de XALc\_0874 é extenso (502 resíduos) para ser considerado um CBM2 e a estrutura é formada, majoritariamente, por fitas- $\beta$  paralelas em conformação tubular (**Figura 14a** e **c**). Estruturalmente, esse domínio se aproxima ao domínio catalítico das enzimas que pertencem à classe enzimática polissacarídeo liase da família 6 (PL6).

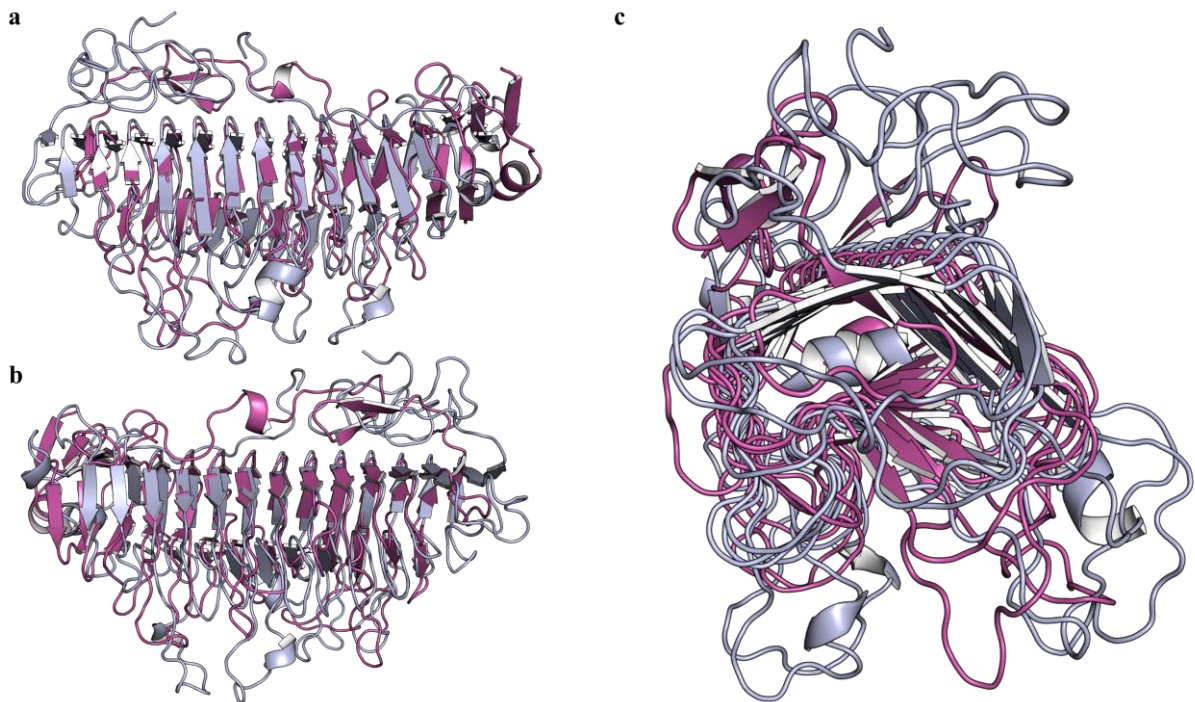
**Figura 14** - Visualização estrutural em modo “cartoon” da enzima XALc\_0874, obtida do banco de dados do AlphaFold, sem o peptídeo sinal. (a) Domínio catalítico em azul, linker em rosa-escuro, e domínio CBM2 em amarelo-claro. (b) Domínio CBM2 em vista frontal. (c) Domínio catalítico em vista lateral, (d) posterior e (e) frontal.



**Fonte:** Elaborada pelo autor.

Ainda a partir dos resultados obtidos pelo FoldSeek, foi realizado uma sobreposição estrutural entre os domínios catalíticos de XALc\_0874 e da enzima alginato liase (TM-score = 0,691 e probabilidade = 0,82) de *Pseudopedobacter saltans* (PDB: 7o7a, resolução; 1,99 Å) (**Figura 15**). Essa enzima alginato liase é comprovadamente uma representante da família PL6 (VIOLOT et al., 2021) e a similaridade de características estruturais entre ambas sugere que XALc\_0874 também pertence à família PL6, sendo uma enzima multidomínio (catalítico, linker e CBM2) (**Figura 14a** e **Figura 15**). A confirmação desta hipótese seria possível mediante a realização de testes de atividade enzimática e análises subsequentes. Essas observações realçam a importância da modelagem de proteínas na identificação de características estruturais e funções biológicas, além de auxiliarem em revisões nas anotações funcionais dos bancos de dados.

**Figura 15** - Sobreposição estrutural do domínio catalítico da enzima XALc\_0874 (azul), obtida do banco de dados do AlphaFold, sem o peptídeo sinal, com a enzima alginato liase (PDB: 7o7a) (magenta) de *Pseudopedobacter saltans*.



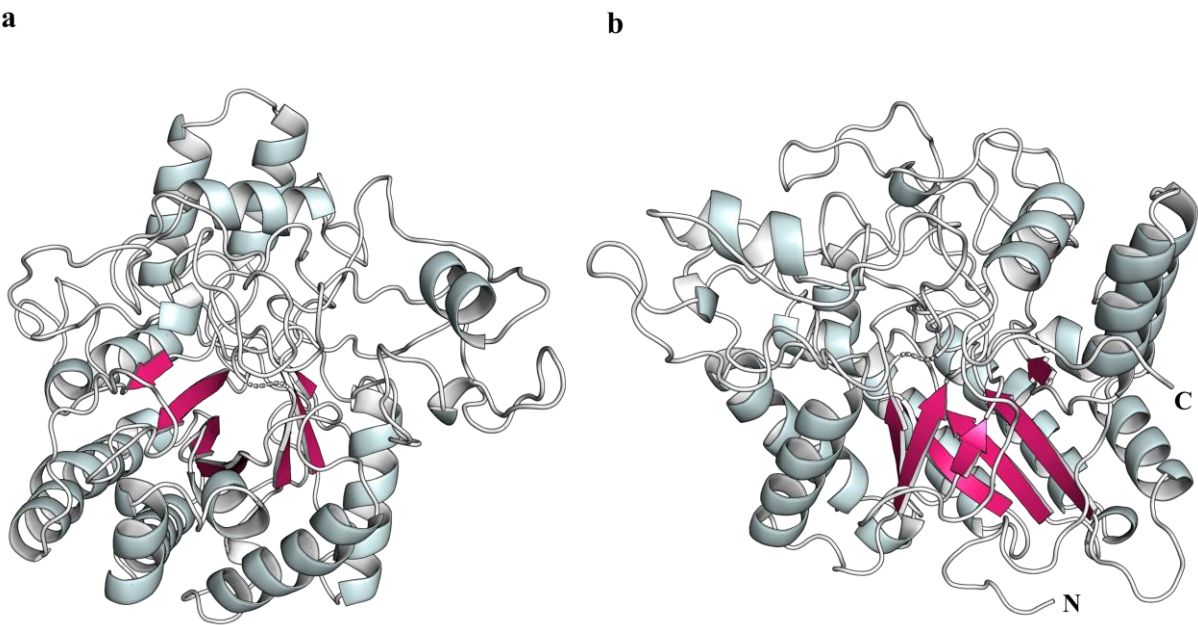
**Fonte:** Elaborada pelo autor.

Com o intuito de identificar novas CAZymes com capacidade para degradar celulose, a enzima XALc\_0874, que provavelmente é representante da família PL6, foi excluída das análises subsequentes. Isso se deve ao fato de que as enzimas PL6 possuem atividade de alginato liase (EC 4.2.2.-), sendo reconhecidas por atuarem principalmente na despolimerização do alginato, não por interagirem diretamente com a cadeia de celulose (LOMBARD et al., 2010; MATHIEU et al., 2016). Visto que o alginato possui propriedades físicas de interesse industrial, sendo encontrado principalmente na parede celular de algas marrons e em algumas bactérias Gram-negativas (MATHIEU et al., 2016), a presença da enzima XALc\_0874 no proteoma de *X. albilineans* GPE PC73 a caracteriza como um alvo interessante para estudos posteriores.

Além da sobreposição, as quatro enzimas restantes (XALc\_0484, XALc\_0865, XALc\_2967 e XALc\_2969) compartilham características estruturais altamente conservadas em GHs, que podem auxiliar na identificação da família enzimática, a qual a enzima pertence (DAVIES; HENRISSAT, 1995; HENRISSAT et al., 1995). Uma dessas características é a presença de domínios específicos como o domínio barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássico, composto principalmente por oito folhas- $\beta$  dispostas em formato de túnel e comumente encontrado em GHs (COPLEY; BORK, 2000; JUERS; HUBER; MATTHEWS, 1999; WIERENGA, 2001).

As enzimas da família GH6, como a XALc\_0484, se destacam por possuírem um domínio barril  $\alpha/\beta$  modificado, denominado de  $\beta$ -barril central, uma estrutura cilíndrica formada por sete fitas- $\beta$  principais (SANDGREN et al., 2013; TAYI et al., 2018; VARROT; SCHÜLEIN; DAVIES, 1999). Segundo os dados de predição topológica e estruturas secundárias obtidos pelo PDBsum (**APÊNDICE B, Figura A3 - Figura A7**), o domínio catalítico de XALc\_0484 é composto por um total de 19  $\alpha$ -hélices e 11 fitas- $\beta$ . As fitas- $\beta$  2-6, 9 e 11 são responsáveis por formar o  $\beta$ -barril central (**Figura 16**). Esse resultado reforça a indicação de que XALc\_0484 pertence à família GH6.

**Figura 16** - Representação estrutural do domínio catalítico de XALc\_0484. As  $\alpha$ -hélices são representadas em ciano-claro, as fitas- $\beta$  em rosa-escuro e as alças em branco. **(a)** Visão frontal da estrutura de XALc\_0484, destacando o centro do domínio  $\beta$ -barril central formado por sete fitas- $\beta$  (rosa-escuro). As regiões N- e C-terminais estão ocultas por duas  $\alpha$ -hélices. **(b)** Visão lateral da estrutura de XALc\_0484, onde é possível visualizar as sete fitas- $\beta$  (rosa-escuro) paralelas, rotacionando para formar o  $\beta$ -barril central. N e C sinalizam as regiões N- e C-terminais, respectivamente.



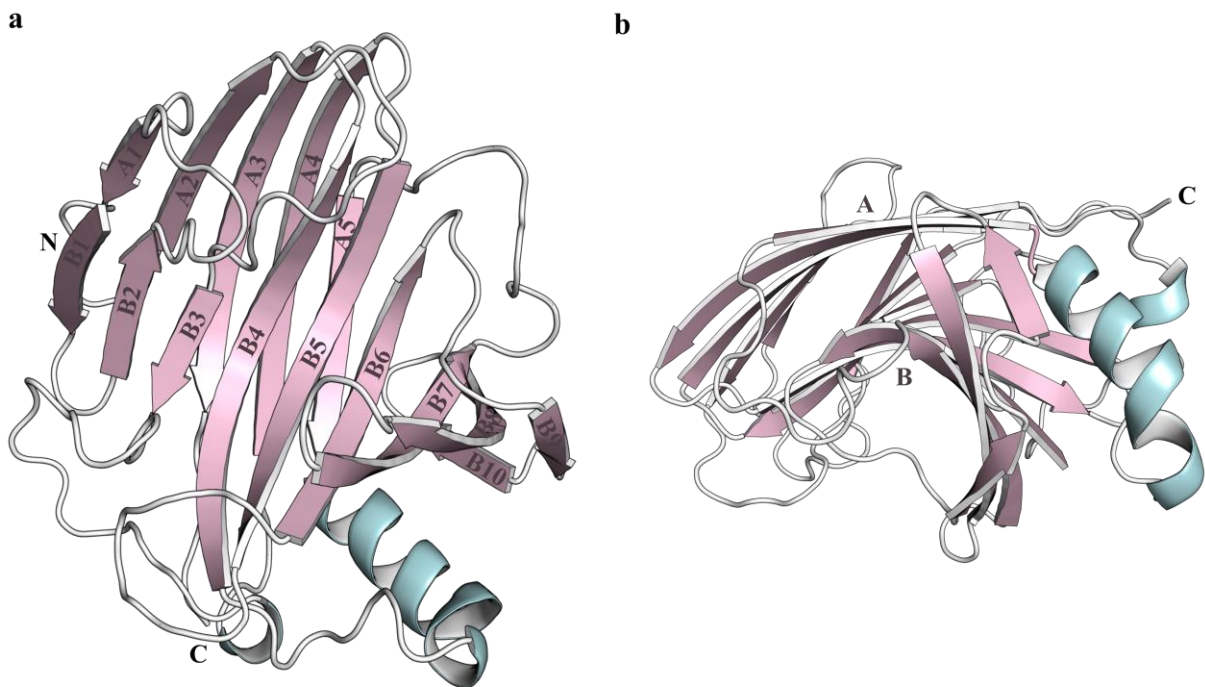
**Fonte:** Elaborada pelo autor.

Dentre as enzimas da família GH12, o domínio catalítico é frequentemente encontrado em uma conformação conhecida como sanduíche- $\beta$ . Esta conformação é formada por duas folhas- $\beta$ , A e B, compostas por fitas- $\beta$  antiparalelas e sobrepostas. As duas folhas- $\beta$ , principalmente a interna, se curvam para formar um extenso sítio de ligação (CRENNELL et al., 2006; SANDGREN et al., 2003; SULZENBACHER et al., 1999). O modelo predito para o domínio catalítico de XALc\_0865 apresentou a estrutura sanduíche- $\beta$  típica das enzimas GH12.

Esta estrutura é formada pelas folhas- $\beta$  A e B, compostas por cinco e dez fitas- $\beta$ , respectivamente. Além disso, o modelo também apresenta duas pequenas  $\alpha$ -hélices (**Figura 17**).

Com exceção das fitas- $\beta$  B6, B7, B9 e B10, as demais fitas- $\beta$  são antiparalelas, conforme o esperado para as enzimas desta família (**Figura 17**) (CRENNELL et al., 2006; CRENNELL; HREGGVIDSSON; NORDBERG KARLSSON, 2002; PRATES et al., 2013; SANDGREN et al., 2003; SULZENBACHER et al., 1999).

**Figura 17** - Representação estrutural do domínio catalítico de XALc\_0865. As  $\alpha$ -hélices são representadas em azul-claro, as fitas- $\beta$  em rosa-claro e as alças em branco. **(a)** Visão frontal da estrutura de XALc\_0865, evidenciando as duas folhas- $\beta$ , denominadas por A e B. As fitas- $\beta$  individuais que compõem cada folha estão numeradas (A1 - A5 e B1 - B10), conforme a posição em cada folha. N e C sinalizam as regiões N- e C-terminais, respectivamente, ressaltando que o N-terminal se inicia na fita- $\beta$  B1. **(b)** Visão lateral da estrutura de XALc\_0865, onde é possível visualizar as folhas- $\beta$  A e B sobrepostas, compondo o domínio sanduíche- $\beta$  e curvadas para formar o sítio de ligação. C sinaliza a região C-terminal, visto que o N-terminal estava oculto por uma folha- $\beta$  e, portanto, não pôde ser representado.



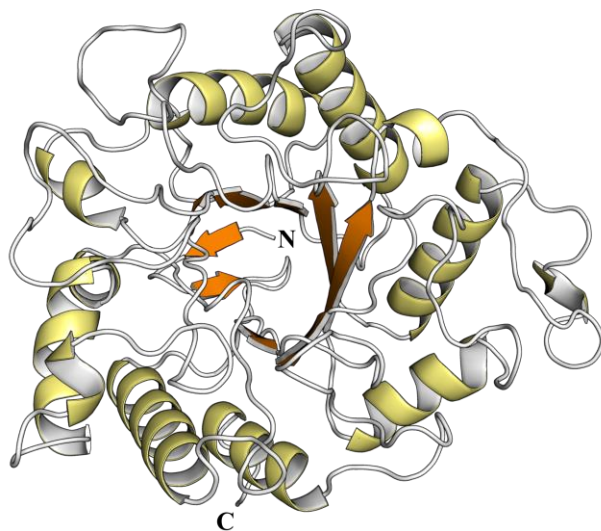
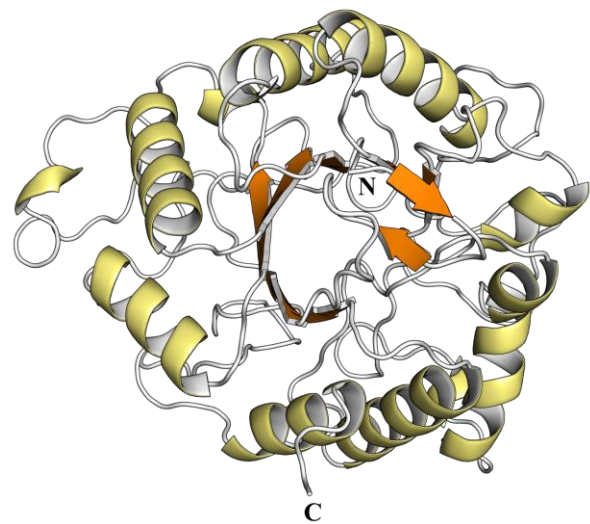
**Fonte:** Elaborada pelo autor.

A família enzimática GH5 é a maior entre as GHs. Os seus representantes apresentam, em sua maioria, a conformação barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássica. Esta conformação é composta principalmente por oito  $\alpha$ -hélices e fitas- $\beta$ . As oito fitas- $\beta$  internas são dispostas paralelamente, formando a estrutura central do barril, sendo recobertas pelas oito  $\alpha$ -hélices na parte externa (HENRISSAT et al., 1995; JENKINS et al., 1995; WIERENGA, 2001). Os domínios catalíticos de XALc\_2967 e XALc\_2969 são exemplos desta conformação. Eles são compostos por 12  $\alpha$ -

hélices e 12 fitas- $\beta$ , e 12  $\alpha$ -hélices e 11 fitas- $\beta$ , respectivamente. Devido à grande semelhança estrutural entre si, a conformação barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássica pode ser observada na estrutura predita de ambas as enzimas. Isso evidencia que elas são representantes da família GH5 (**Figura 18 e Figura 19**).

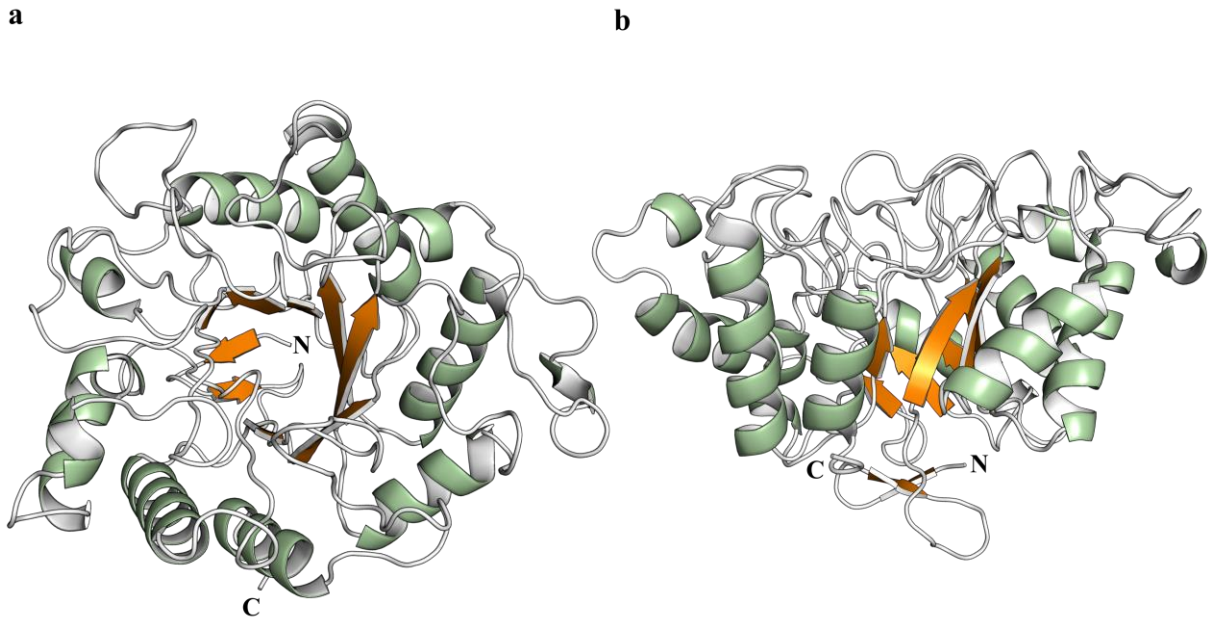
Além das oito fitas- $\beta$ , é possível visualizar uma nona fita- $\beta$  antiparalela na parte inferior do barril. Esta fita auxilia na vedação da estrutura para proteger do solvente (**Figura 18 e Figura 19**) (HENRISSAT et al., 1995; JENKINS et al., 1995; WIERENGA, 2001).

**Figura 18** - Representação estrutural do domínio catalítico de XALc\_2967. As  $\alpha$ -hélices são representadas em amarelo-claro, as fitas- $\beta$  em laranja e as alças em branco. **(a)** Visão frontal da estrutura de XALc\_2967, evidenciando o centro do domínio barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássico formado por oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela ao fundo. N e C sinalizam as regiões N- e C-terminais, respectivamente. **(b)** Visão posterior da estrutura de XALc\_2967, onde é possível visualizar oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela que auxilia na proteção contra o solvente. N e C sinalizam as regiões N- e C-terminais, respectivamente.

**a****b**

**Fonte:** Elaborada pelo autor.

**Figura 19** - Representação estrutural do domínio catalítico de XALc 2969. As  $\alpha$ -hélices são representadas em verde-claro, as fitas- $\beta$  em laranja e as alças em branco. **(a)** Visão frontal da estrutura de XALc\_2969, evidenciando o centro do domínio barril-TIM ( $\alpha/\beta$ )<sub>8</sub> clássico formado por oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela ao fundo. N e C sinalizam as regiões N- e C-terminais, respectivamente. **(b)** Visão lateral da estrutura de XALc\_2969, onde é possível visualizar oito fitas- $\beta$  (laranja) paralelas e a nona fita antiparalela na parte inferior da estrutura, que auxilia na proteção contra o solvente. N e C sinalizam as regiões N- e C-terminais, respectivamente.



**Fonte:** Elaborada pelo autor.

A validação estrutural, proveniente da análise comparativa e visual, sustenta a utilização desses modelos nas simulações de docking molecular, visando compreender as interações específicas com os substratos de interesse.

#### 4.4 PREDIÇÃO DOS SÍTIOS DE LIGAÇÃO

A análise dos sítios de ligação é essencial para compreender as interações específicas entre as enzimas e os substratos, visto que essas interações são fundamentais para diversos processos biológicos. A identificação precisa desses sítios a partir da estrutura enzimática auxilia na determinação de mecanismos moleculares, funções enzimáticas, design de medicamentos, entre outros (ALASDAIR; JACKSON; JACKSON, 2006; BALLANTE, 2021). Embora a identificação experimental do sítio de ligação seja mais confiável, é um processo dispendioso e extenso (ALASDAIR; JACKSON; JACKSON, 2006; KONC; JANEŽIČ, 2014; XIE; XIE; BOURNE, 2011). Nesse contexto, métodos computacionais, foram desenvolvidos para otimizar a previsão de sítios de ligação, mantendo a alta precisão dos resultados

(ALASDAIR; JACKSON; JACKSON, 2006; JAKUBEC et al., 2022; KRIVÁK; HOKSZA, 2018).

O servidor PrankWeb foi utilizado para prever os possíveis sítios de ligação nos domínios catalíticos das quatro CAZymes de *X. albilineans* GPE PC73. A análise realizada identificou prováveis sítios de ligação nos domínios catalíticos de todas as CAZymes. Foi observado que algumas enzimas apresentam diversas cavidades em sua superfície, das quais o algoritmo de predição as identifica como possíveis sítios de ligação. Por exemplo, três sítios de ligação foram identificados no domínio catalítico de XALc\_0484, enquanto apenas um foi identificado em XALc\_2967. Todos os sítios de ligação previstos para cada enzima foram classificados conforme a sua pontuação integral e probabilidade de ocorrer na posição predeterminada. As predições mais precisas para cada CAZyme estão descritas na **Tabela 7**.

**Tabela 7** - Posição dos melhores sítios de ligação preditos pelo software PrankWeb para cada CAZyme.

CAZymes	Probabilidade	Coordenadas	Resíduos ID
XALc_0484 GH6	0,939	x = 5,5072 y = -3,0694 z = 3,1095	48, 50, 54, 81, 82, 84, 95, 98, 104, 125, 130, 131, 133, 134, 135, 136, 138, 179, 180, 188, 227, 228, 230, 231, 266, 267, 268, 364, 366, 393, 394, 397, 399, 400, 409, 411, 412, 414, 416, 424, 425, 438 e 439
XALc_0865 GH12	0,798	x = 6,7748 y = -0,5348 z = 0,7330	25, 26, 27, 29, 31, 32, 34, 35, 36, 38, 50, 52, 54, 82, 84, 87, 95, 96, 98, 130, 136, 138, 151, 153, 159, 160, 161, 162, 163, 186, 244, 248 e 250
XALc_2967 GH5	0,594	x = 5,1788 y = 9,6531 z = 1,9232	51, 54, 55, 61, 136, 141, 142, 143, 182, 224, 236, 237, 238, 264, 268, 269, 271, 305, 341, 346, 347, 348 e 350
XALc_2969 GH5	0,572	x = 8,2441 y = 0,3713 z = -0,5504	51, 54, 55, 136, 137, 143, 182, 236, 237, 263, 267, 303, 337, 343, 344 e 345

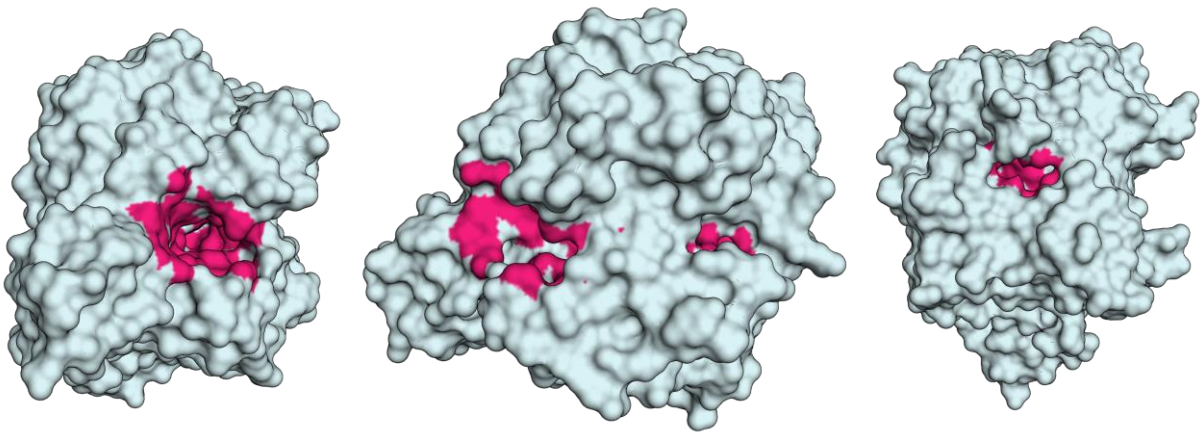
**Fonte:** Elaborada pelo autor.

As pontuações de probabilidade mais baixas para XALc\_2967 e XALc\_2969 estão provavelmente relacionadas com a quantidade de informações disponíveis sobre estruturas experimentais semelhantes. Dados evolutivos e estruturais estão entre as informações que o PrankWeb utiliza para construir o modelo preditivo (JAKUBEC et al., 2022; KRIVÁK;

HOKSZA, 2018). As enzimas XALc\_2967 e XALc\_2969 pertencem à família GH5, uma das maiores entre as GHs (ASPEBORG et al., 2012; DRULA et al., 2022), refletindo no alto número de estruturas depositadas nos bancos de dados experimentais. Isso resulta em uma grande variedade de formatos de sítios de ligação, podendo levar a pontuações de probabilidade mais baixas, conforme as determinadas pelo software (**Tabela 7**).

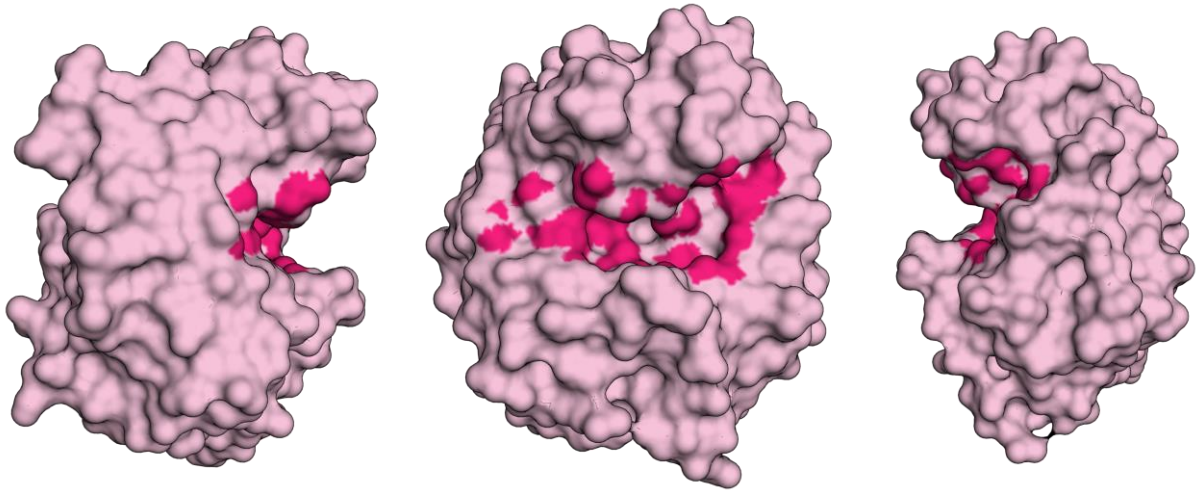
Dentre os dados fornecidos pelo PrankWeb, a identificação dos resíduos que constituem o sítio de ligação, juntamente com as coordenadas dos eixos X, Y e Z (**Tabela 7**), foram utilizadas como ponto inicial para configurar a área de encaixe dos experimentos de docking. Com base na classificação e inspeção visual, o tamanho da área de encaixe para algumas enzimas foi determinado utilizando-se os dados de predição de dois ou mais sítios de ligação. Por exemplo, dos três sítios preditos para XALc\_0484, os resíduos que compõem os sítios 1 e 2 foram selecionados para configurar a área de encaixe utilizada na simulação de docking. Os sítios preditos para as quatro enzimas podem ser visualizados nas **Figura 20 - Figura 23**.

**Figura 20** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_0484, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb.



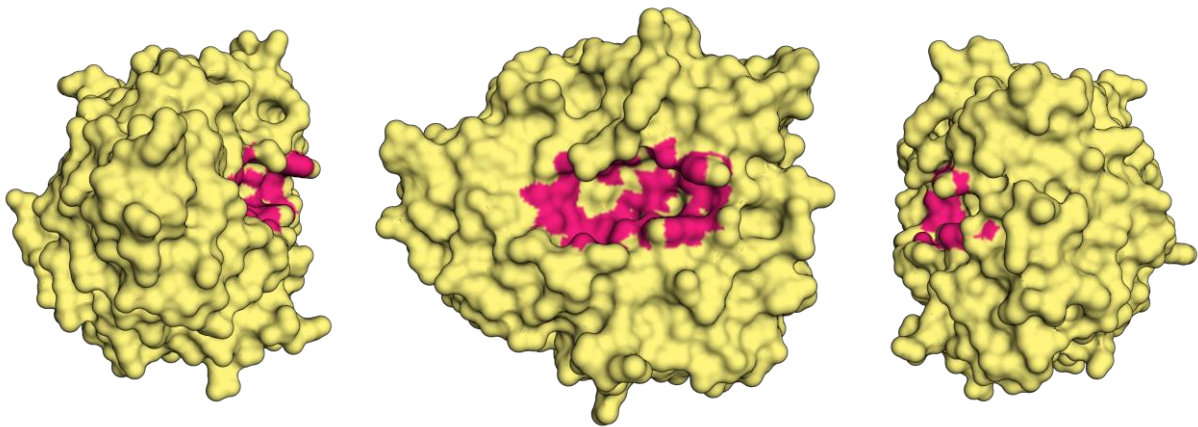
**Fonte:** Elaborada pelo autor.

**Figura 21** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_0865, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb.



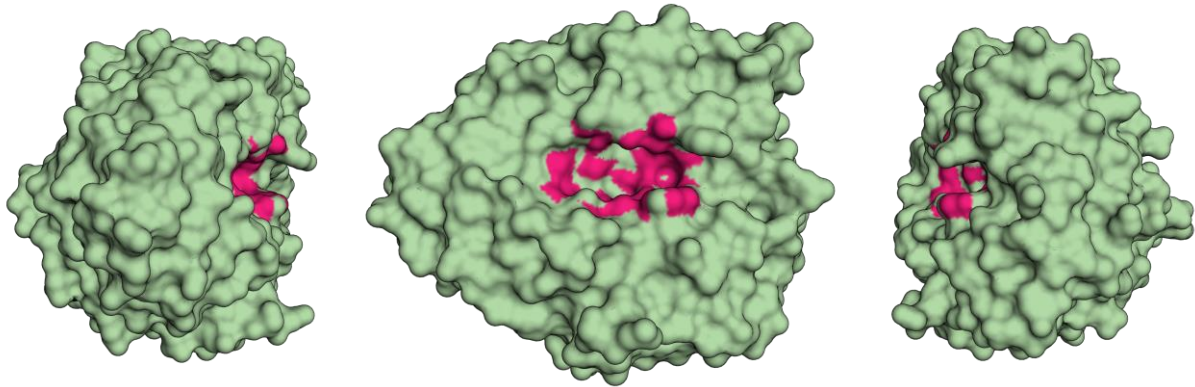
**Fonte:** Elaborada pelo autor.

**Figura 22** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_2967, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb.



**Fonte:** Elaborada pelo autor.

**Figura 23** - Visualização superficial (lateral esquerda, frontal e lateral direita) da enzima XALc\_2969, destacando em rosa-escuro o sítio de ligação predito pelo servidor PrankWeb.



**Fonte:** Elaborada pelo autor.

A confirmação da posição dos sítios de ligação preditos foi obtida por meio de sobreposição estrutural e inspeção visual. A sobreposição estrutural foi realizada utilizando duas ou três estruturas experimentais, as quais foram obtidas no PDB. Essas estruturas estão na forma *holo*, ou seja, enzimas que estão complexadas com um ligante. Além disso, essas enzimas obtidas do PDB são homólogas a cada CAZyme analisada. Após realizar a sobreposição e avaliar as cavidades que correspondem aos sítios de ligação das duas enzimas, tanto no modelo predito quanto na estrutura experimental (**dados não mostrados**), foi possível confirmar a posição do sítio de ligação para cada CAZyme. A ausência de clashes e outras discrepâncias estruturais na região dos sítios foram examinadas utilizando-se os dados obtidos pelo MolProbity, juntamente com os gráficos de Ramachandran.

Os ângulos phi ( $\varphi$ ) e psi ( $\psi$ ) representam os únicos ângulos rotacionáveis de uma proteína e contribuem para manter a variação conformacional e a flexibilidade da enzima. No entanto, as possibilidades de rotações e combinações desses ângulos são naturalmente limitadas para prevenir conflitos estruturais. O gráfico de Ramachandran avalia a qualidade do modelo, representando visualmente as combinações estereoquímicas entre os ângulos  $\varphi$  e  $\psi$  (RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963; RAMACHANDRAN; SASISEKHARAN, 1968; RICHARDSON, 1981). Como resultado, os resíduos analisados são distribuídos graficamente em regiões favoráveis (verde), regiões permitidas (verde-claro) e regiões desfavoráveis ou não permitidas (branco) (**Figura 24**). Os modelos estruturais considerados de alta qualidade apresentam, no mínimo, 90% dos seus resíduos em regiões

favoráveis. Os domínios catalíticos das quatro CAZymes excedem este mínimo necessário, apresentando uma quantidade de resíduos em regiões favoráveis que os classifica como de alta qualidade e aptos para os estudos de docking molecular (**Tabela 8 e Figura 24**).

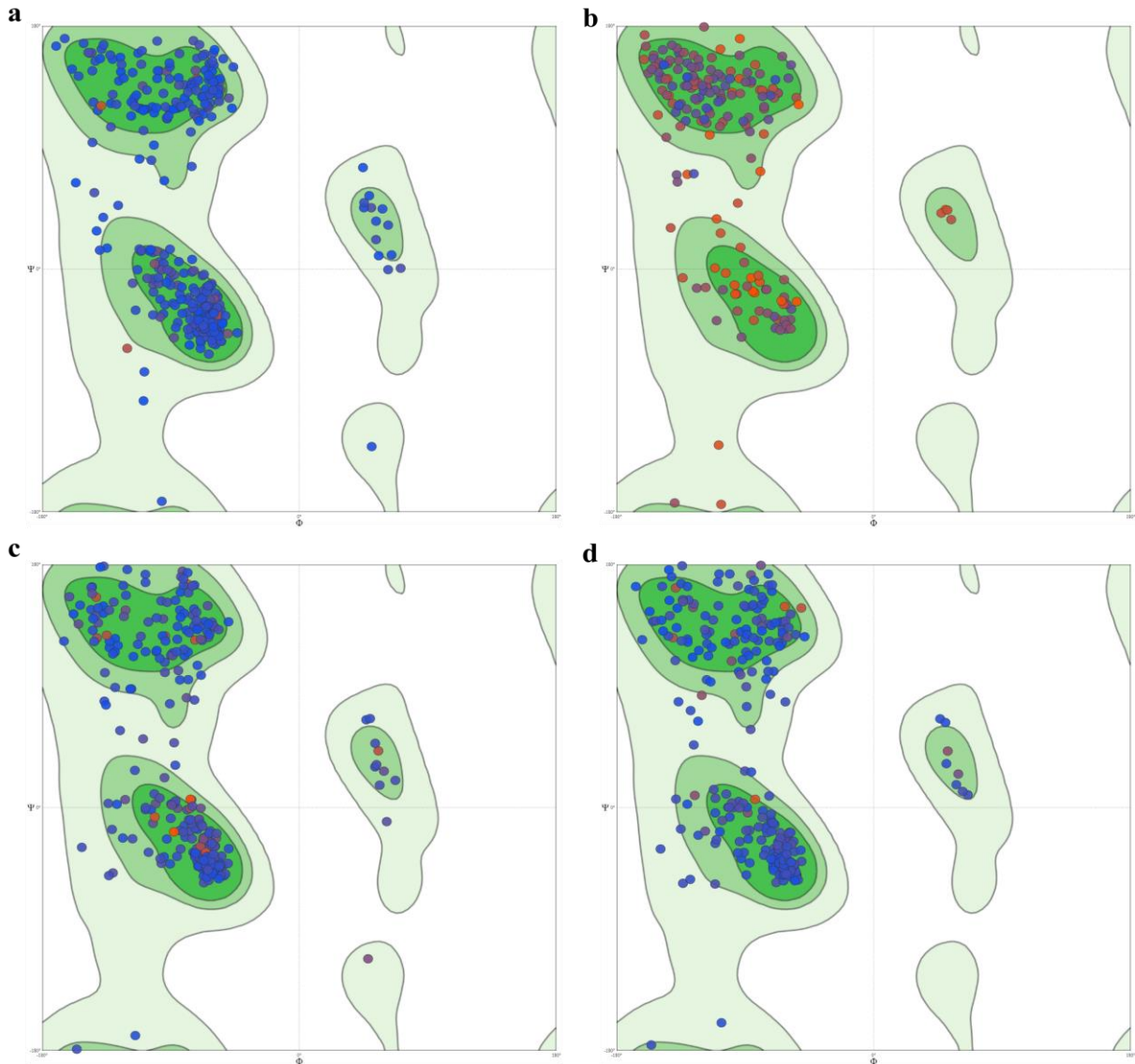
**Tabela 8** - Valores expressos sobre a qualidade dos ângulos diédricos phi  $\phi$  e psi  $\psi$  dos resíduos que compõem cada CAZyme.

CAZymes	Resíduos favoráveis	Resíduos permitidos	Resíduos desfavoráveis
XALc_0484	96,25%	0,00%	0,28%
XALc_0865	96,92%	0,00%	0,00%
XALc_2967	95,69%	0,00%	0,00%
XALc_2969	96,21%	0,29%	0,35%

**Fonte:** Elaborada pelo autor.

**Nota:** Ideal para resíduos favoráveis > 98%, resíduos permitidos < 0.2% e resíduos desfavoráveis < 1%.

**Figura 24** - Gráficos de Ramachandran de cada CAZyme. (a) XALc\_0484; (b) XALc\_0865. (c) XALc\_2967; (d) XALc\_2969.



**Fonte:** Elaborada pelo autor.

Ao final, com base na confirmação dos sítios de ligação por meio de análises estruturais, incluindo a sobreposição com estruturas homólogas e a avaliação de conformações permitidas, todos os domínios catalíticos foram considerados aptos para as simulações de docking molecular.

## 4.5 ESTUDOS DE DOCKING MOLECULAR DAS CAZymes DE *Xanthomonas albilineans*

### 4.5.1 Triagem e Seleção dos Ligantes

Para selecionar os principais ligantes de celulose para as simulações de docking, foi realizado uma triagem no banco de dados do CAZy. Dessa análise, foram selecionados cinco ligantes que correspondem a cada família enzimática (GH5, GH6 e GH12). Esses ligantes, que representam as principais unidades estruturais da celulose, apresentam uma conformação espacial da glicose em padrão  $\beta$  e estão entre os mais comumente utilizados em estudos de cristalização envolvendo estruturas experimentais de CAZymes (**Tabela 9**).

**Tabela 9** - Principais ligantes selecionados para cada família enzimática.

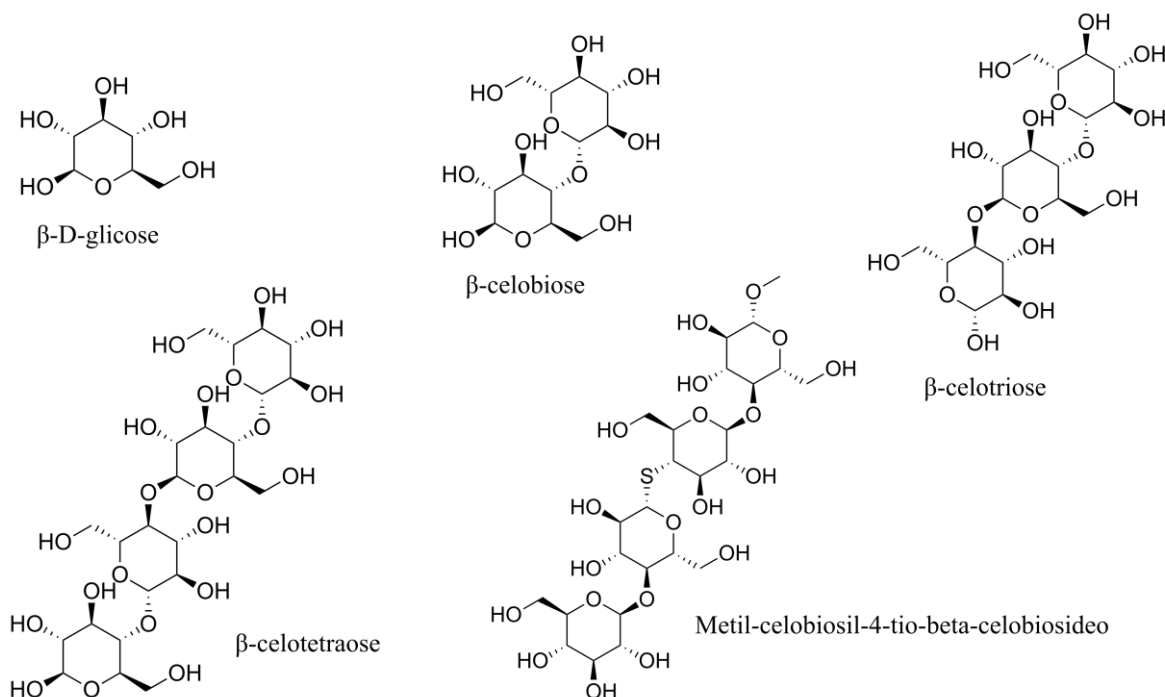
<b>Família</b>	<b>Ligante</b>	<b>PubChem ID</b>
GH6	$\beta$ -D-glicose	64689
	$\beta$ -celobiose	10712
	$\beta$ -celotriose	440950
	$\beta$ -celotetraose	5288428
	Metil-celobiosil-4-tio-beta-celobiosideo	5288410
GH12	$\beta$ -D-glicose	64689
	$\beta$ -celobiose	10712
	$\beta$ -celotetraose	5288428
GH5	$\beta$ -D-glicose	64689
	$\beta$ -celobiose	10712
	$\beta$ -celotriose	440950
	$\beta$ -celotetraose	5288428

**Fonte:** Elaborada pelo autor.

É importante salientar que, apesar das diferenças entre as famílias enzimáticas, cada qual caracterizada por enzimas com especificidade de substrato, há uma recorrência dos

principais ligantes entre elas, visto que são celulasas especializadas na degradação desses polissacarídeos (**Tabela 9 e Figura 25**).

**Figura 25** - Estrutura química dos ligantes selecionados para o docking.



**Fonte:** Elaborada pelo autor.

Entre os ligantes selecionados, a  $\beta$ -D-glicose é a unidade fundamental na composição da celulose vegetal, visto que, conforme relatado anteriormente, a celulose é um polissacarídeo formado principalmente por monômeros de  $\beta$ -D-glicose ligadas entre si por meio de ligações glicosídicas (LAMPUGNANI et al., 2018; MCNAMARA; MORGAN; ZIMMER, 2015). A união de duas dessas unidades resulta na  $\beta$ -celobiose, um dissacarídeo essencial para a estrutura da celulose e o principal produto da hidrólise enzimática por celulasas de diversos microrganismos. Os oligossacarídeos  $\beta$ -celotriose e  $\beta$ -celotetraose, constituídos por três e quatro unidades de glicose, respectivamente, também contribuem para a manutenção estrutural da celulose e são os principais substratos de muitas enzimas hidrolíticas (ANNAMALAI; RAJESWARI; SIVAKUMAR, 2016; DAVIES; HENRISSAT, 1995; GARDNER; BLACKWELL, 1974; LAMPUGNANI et al., 2018; MCNAMARA; MORGAN; ZIMMER, 2015).

O ligante metil-celobiosil-4-tio- $\beta$ -celobiosideo (Metil-4) é uma versão não-hidrolisável da  $\beta$ -celotetraose, modificada pela substituição de um átomo de oxigênio por um átomo de enxofre na ligação glicosídica central. Essa modificação, desestabiliza os estados de transição

da glicose, permitindo que esse estado intermediário da reação enzimática seja capturado e analisado. Este ligante é utilizado principalmente em estudos de interação enzima-substrato, pois permite uma análise mais detalhada da dinâmica da reação (REVERBEL-LEROY et al., 1998; VARROT et al., 2002). Em resumo, esses ligantes são oligossacarídeos de celulose, que diferem apenas em tamanho (**Figura 25**).

#### 4.5.2 Simulações de Docking Molecular das CAZymes

Os estudos de docking molecular foram conduzidos com a finalidade de explorar as interações entre os domínios catalíticos das quatro CAZymes de *Xanthomonas albilineans* GPE PC73 e os ligantes selecionados. Em todas as simulações, os ligantes foram docados nos sítios de ligação dos domínios catalíticos preditos das enzimas. No decorrer dessas simulações, o software de docking (AutoDock Vina) gerou diversas poses (posicionamentos) para cada ligante, refletindo as possíveis interações com os resíduos dos sítios de ligação predefinidos. Os escores de afinidade de ligação (em kcal/mol) resultantes dessas interações foram calculados pela função de pontuação do AutoDock Vina e as poses mais favoráveis foram classificadas pelo menor escore de afinidade (**Tabela 10 - Tabela 13**) (DOS SANTOS; FERREIRA; ANDRICOPULO, 2018).

#### 4.5.3 Docking Molecular da Enzima Xalc\_0484

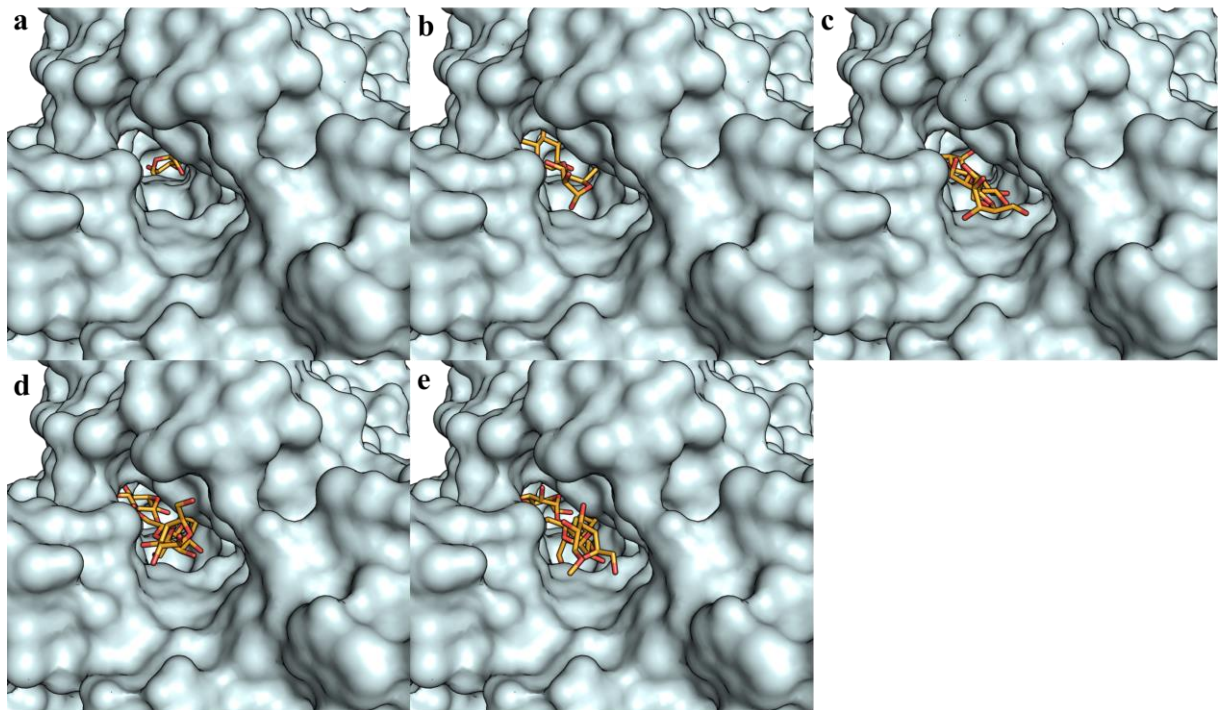
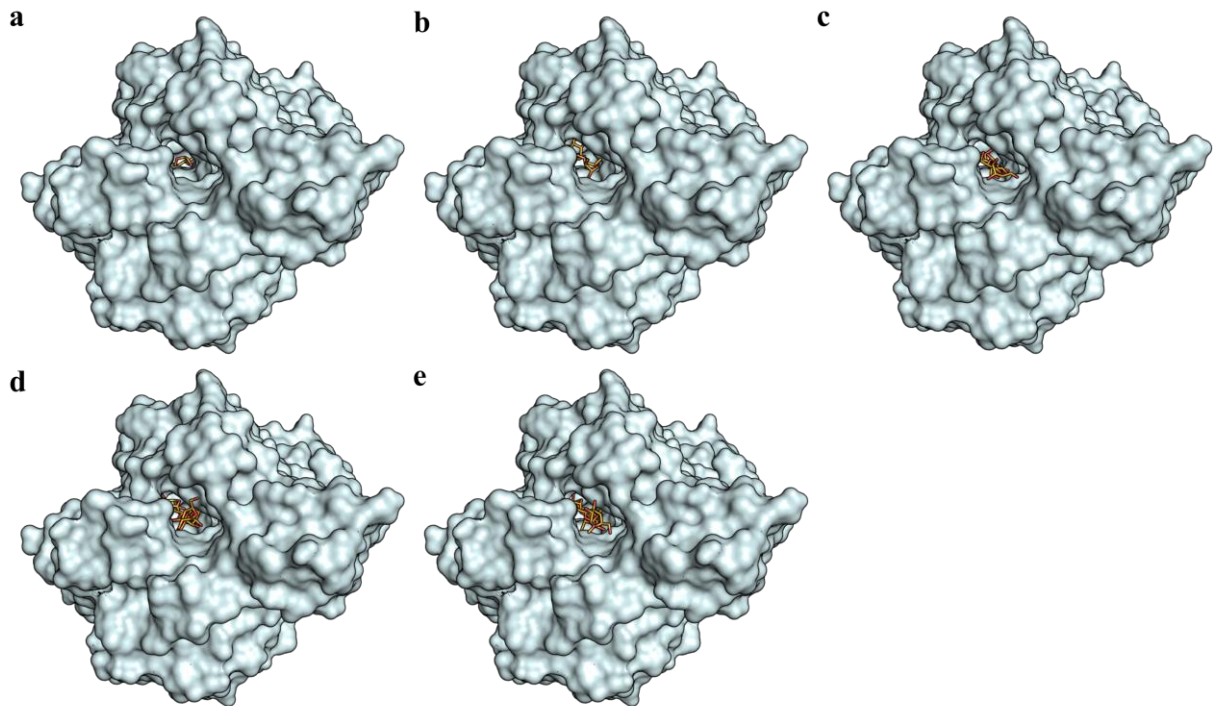
A enzima XALc\_0484 interagiu com os cinco ligantes conforme posicionados no sítio de ligação previsto (**Figura 26**). A análise dessas interações revelou que Tyr125, Asp131, His133, Asn138, His228, Asn268, Lys393 e Asp399 foram os resíduos que interagiram com maior frequência com os ligantes, desempenhando um papel importante na hidrólise da celulose (**Tabela 10**). Conforme mencionado anteriormente, XALc\_0484 é uma hidrolase glicosídica (GH) da família 6 (GH6) (**Figura 13, Figura 16 e Tabela 1**). É importante ressaltar que o seu sítio ativo é envolto por um túnel, fato que a caracteriza, estruturalmente, como uma celobiohidrolase (CBH) da mesma família. As CBHs são reconhecidas por sua alta capacidade hidrolítica, atuando de forma processiva e unidirecional na hidrólise de longas cadeias de celulose em moléculas de celobiose (ANNAMALAI; RAJESWARI; SIVAKUMAR, 2016; DAVIES; HENRISSAT, 1995; SANDGREN et al., 2013; TAKEDA et al., 2022; TAYI et al., 2018; WU et al., 2013).

**Tabela 10** - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc\_0484 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.

<b>Ligante</b>	<b>Afinidade</b>	<b>Ligações de Hidrogênio</b>	<b>Pontes salinas</b>
$\beta$ -D-glicose	-5,84	Asn48 (2,94), Trp82 (3,77), Asn95 (3,79), His104 (3,30) e Glu394 (3,85).	-
$\beta$ -celobiose	-8,57	Tyr125 (3,06), Asp131 (3,10), Asn138 (3,23), Asn138 (3,14) e Asp180 (3,68).	-
$\beta$ -celotriose	-10,19	Tyr125 (2,83), Asp131 (3,56), His133 (3,16), Thr187 (4,04), Asn188 (3,12), His228 (2,99), Asn268 (3,28), Arg363 (3,89) e Asp399 (3,56).	Lys393 (5,46)
$\beta$ -celotetraose	-9,83	Tyr125 (3,79), Tyr125 (3,79), Arg130 (3,86), Asp131 (4,09), His133 (3,13), Asn138 (3,57), Asn138 (3,17), Asn184 (4,10), Asn188 (2,90), His228 (3,30), Asn268 (3,14), Arg331 (4,10), Lys393 (4,01), Asp399 (3,00) e Ser438 (3,13).	Lys393 (5,11)
Metil-4	-9,90	Tyr125 (3,50), Arg130 (3,15), His133 (3,30), Asn138 (3,37), Asn138 (3,68), His228 (3,01), Asn268 (3,04), Arg331 (4,07), Gly364 (3,00), Trp366 (3,31), Lys393 (3,42), Asp399 (3,11), Asp399 (3,11) e Asp399 (3,29).	Lys393 (4,77)

**Fonte:** Elaborada pelo autor.

**Figura 26** - Estrutura de XALc\_0484 complexada com os cinco ligantes selecionados; (a)  $\beta$ -D-glucose; (b)  $\beta$ -celobiose; (c)  $\beta$ -celotriose; (d)  $\beta$ -celotetraose; (e) Metil-4.

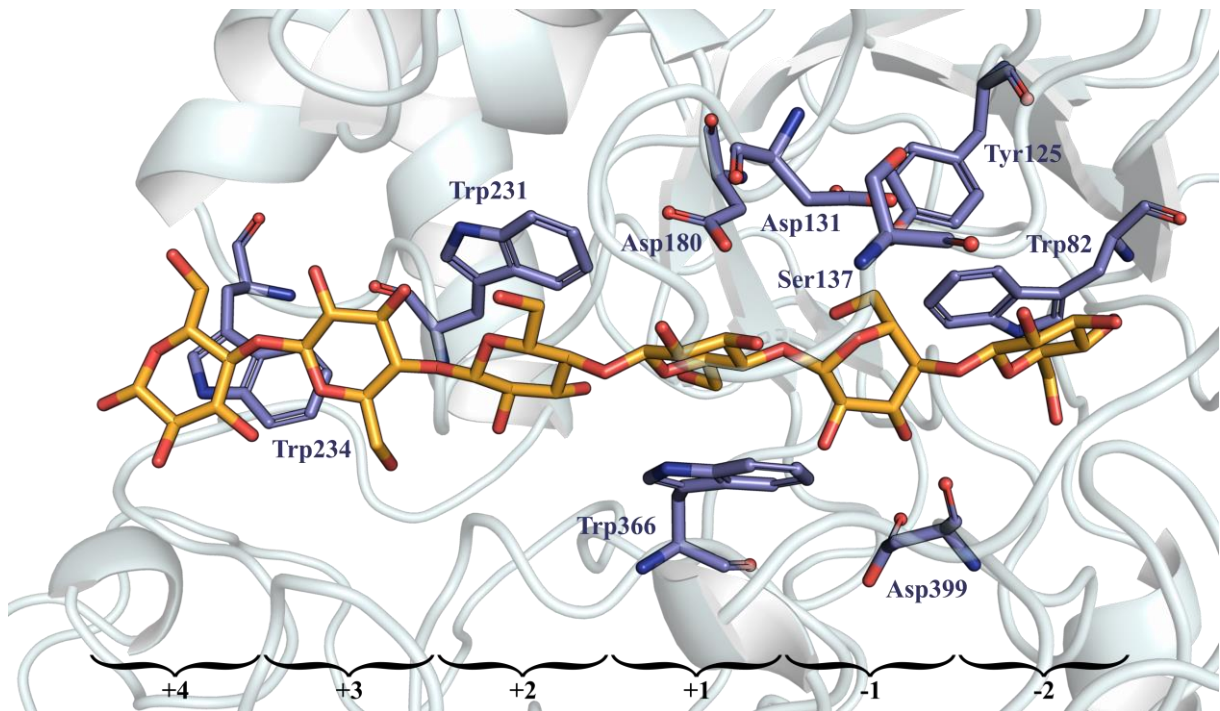


**Fonte:** Elaborada pelo autor.

Os sítios ativos das hidrolases glicosídicas (GHs) são divididos em partes menores denominadas subsítios. Esses subsítios são numerados com base na posição em que as moléculas de glicose ocupam ao longo de todo sítio ativo. Em CBHs GH6, como a XALc\_0484, essa enumeração segue uma sequência (-4, -3, -2, -1, +1, +2, +3, +4) (DAVIES; WILSON;

HENRISSAT, 1997), indo da extremidade não redutora (negativa) para a redutora (positiva) (**Figura 27**). A ligação glicosídica a ser clivada está posicionada entre os subsítios -1 e +1. Vale destacar que em CBHs GH6 os subsítios -2 e -1 são comumente referenciados como subsítios do produto, pois abrigam a porção celobiose removida da extremidade não redutora da cadeia de celulose (**Figura 27**) (SANDGREN et al., 2013; TAKEDA et al., 2022; WU et al., 2013).

**Figura 27** - Visão geral dos principais subsítios negativos e positivos de XALc\_0484. Em roxo os principais resíduos que participam do processo catalítico. A celohexaose (laranja) complexada com XALc\_0484 foi obtida a partir de sobreposição estrutural com a celobiohidrolase GH6 (Cel6B) de *Thermobifida fusca* (PDB: 4b4f).



Fonte: Elaborada pelo autor.

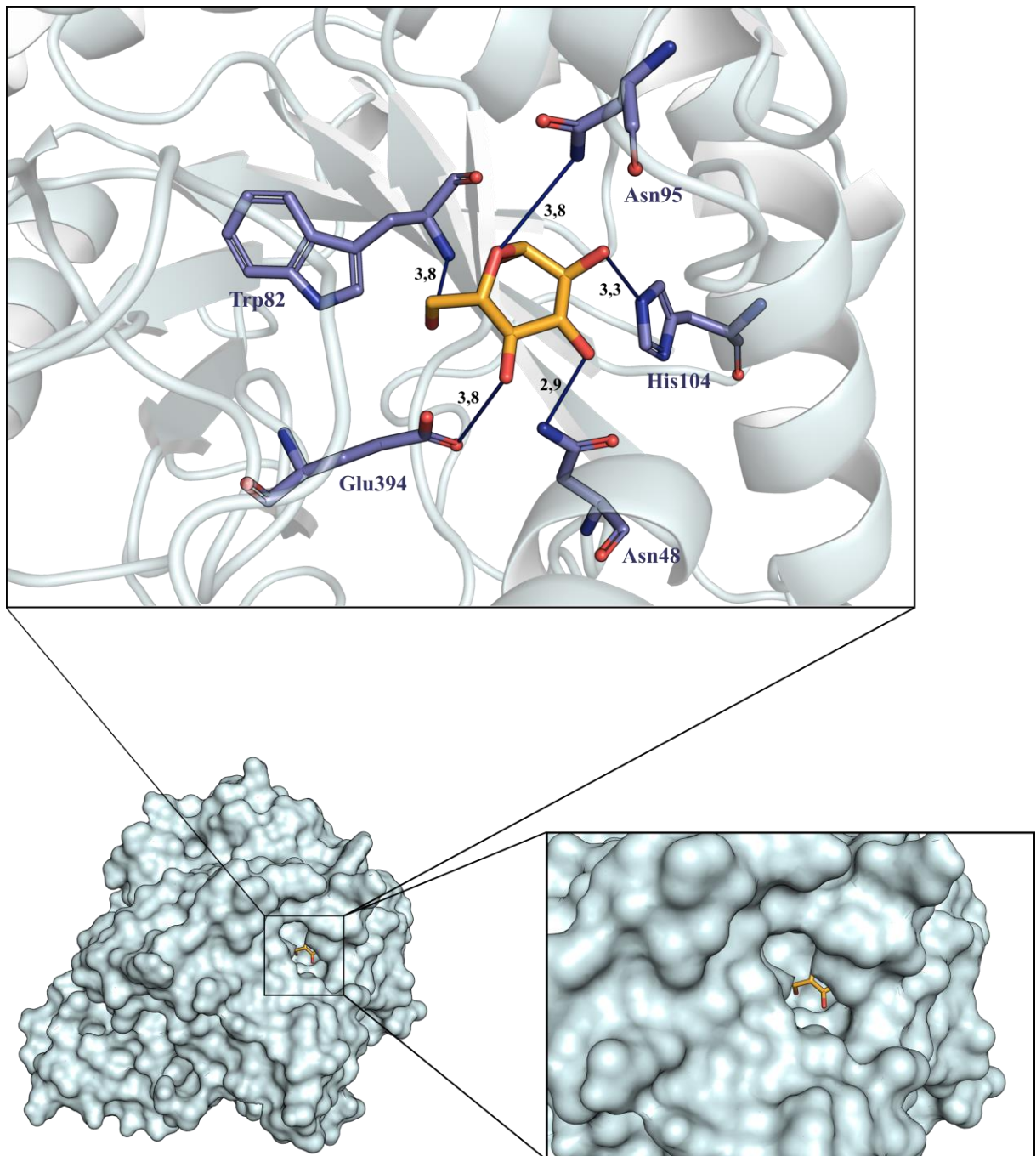
As celobiohidrolases da família GH6 realizam a hidrólise por meio do mecanismo ácido-base de inversão, o qual consiste na inversão da configuração anomérica do açúcar durante a catálise. Nesse processo catalítico, a enzima utiliza dois resíduos de aspartatos (ácidos carboxílicos) altamente conservados e posicionados em lados opostos do sítio catalítico para romper a ligação glicosídica (**Figura 27**). O primeiro aspartato atua como ácido geral, doando um próton ao açúcar, enquanto o segundo age como base geral, aceitando um próton do açúcar. Essa interação promove a instabilidade do açúcar, levando à formação de um estado de transição semelhante ao íon oxocarbênio e consequentemente a quebra da ligação glicosídica (DAVIES; HENRISSAT, 1995; MCCARTER; STEPHEN WITHERS, 1994).

Em XALc\_0484, o Asp180 atua como o ácido catalítico, enquanto o Asp131 interage com a Ser137 e moléculas de água para formar uma rede hídrica e desempenhar o papel de base

catalítica (SANDGREN et al., 2013; TAKEDA et al., 2022; TAYI et al., 2018; WU et al., 2013). Outros resíduos conservados, incluindo Tyr125, Asp399 e cinco triptofanos (Trp82, Trp231, Trp234, Trp297 e Trp366) são essenciais para o processo catalítico (**Figura 27**). O Tyr125 é responsável por distorcer o substrato no sítio ativo, mantendo sua conformação adequada para a hidrólise. O Asp399 participa do processo estabilizando o substrato, e os cinco triptofanos (Trp82, Trp231, Trp234, Trp297 e Trp366) distribuídos entre os subsítios, guiam e ancoram o substrato ao longo do sítio ativo (SANDGREN et al., 2013; TAKEDA et al., 2022; TAYI et al., 2018; WU et al., 2013).

Em se tratando dos resultados de docking molecular, cinco interações foram observadas para o ligante  $\beta$ -D-glicose. No entanto, esse conjunto de interações não são típicos para o processo de hidrólise (**Tabela 10**). Esse ligante foi posicionado no subsítio -4, próximo à saída do túnel que envolve o sítio catalítico, indicando que estaria pronto para ser expelido pela enzima (**Figura 28**).

**Figura 28** - Estrutura de XALc\_0484 complexada com  $\beta$ -D-glicose posicionada no subsítio -4. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).

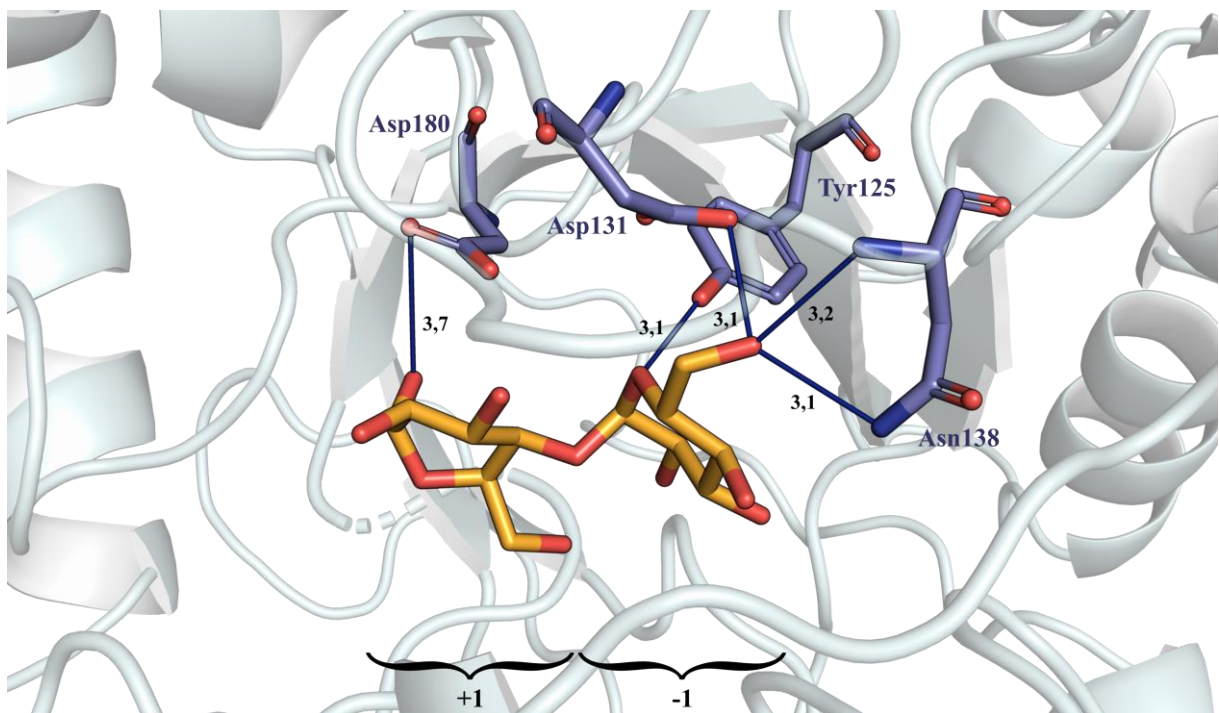


**Fonte:** Elaborada pelo autor.

Para o ligante  $\beta$ -celobiose, foi observado um escore de afinidade de -8,57 kcal/mol, sendo este o único ligante a interagir com o Asp180 (**Tabela 10**). Esse ligante foi posicionado nos subsítios de clivagem -1 e +1, ao invés dos subsítios de produto, localizados em -2 e -1 (**Figura 29**). Isto se deve ao fato de que o AutoDock Vina seleciona a pose de melhor afinidade entre diversas possibilidades, sem considerar características naturais como o tamanho do

substrato. Enzimas da família GH6, como XALc\_0484, apresentam maior afinidade por substratos maiores, geralmente com quatro ou mais monômeros de glicose. Normalmente, substratos menores como  $\beta$ -D-glicose e  $\beta$ -celobiose atuam como inibidores enzimáticos (ANNAMALAI; RAJESWARI; SIVAKUMAR, 2016; TAKEDA et al., 2022; WU et al., 2013) e, portanto, a clivagem por XALc\_0484 seria pouco provável, contradizendo os resultados de docking para  $\beta$ -celobiose (**Tabela 10**).

**Figura 29** - Estrutura de XALc\_0484 complexada com  $\beta$ -celobiose posicionada nos subsítios -1 e +1. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



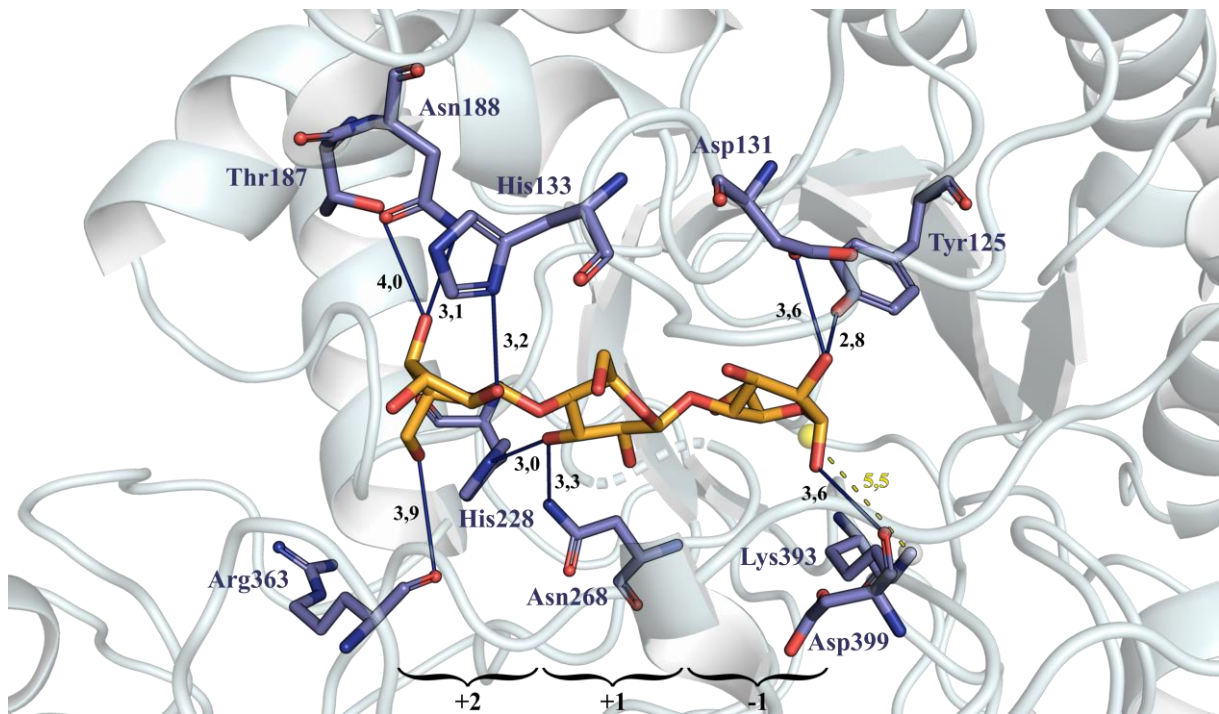
Fonte: Elaborada pelo autor.

Em relação a  $\beta$ -celotriose, esta mostrou a maior afinidade de ligação (-10,19 kcal/mol) entre os ligantes simulados para a enzima XALc\_0484. Esta interação proteína-ligante ocorreu por intermédio de nove ligações de hidrogênio e uma ponte salina, envolvendo um total de 10 resíduos do sítio de ligação, incluindo três resíduos catalíticos conservados (Tyr125, Asp131 e Asp399) (**Tabela 10**).

Ainda assim, é pouco provável que este polímero seja eficientemente hidrolisado por celobiohidrolases GH6 bacterianas, como a XALc\_0484. As celobioses são, naturalmente, o principal produto das CBHs GH6. A posição ocupada pela  $\beta$ -celotriose nos subsítios -1 a +2 sugere que o produto desta hidrólise seria um monômero de glicose (**Figura 30**), diferindo dos

resultados de atividade enzimática observados em CBHs GH6 experimentais (KOIVULA et al., 2002; SANDGREN et al., 2013; TAKEDA et al., 2022; TAYI et al., 2018).

**Figura 30** - Estrutura de XALc\_0484 complexada com  $\beta$ -celotriose posicionada nos subsítios -1 a +2. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).

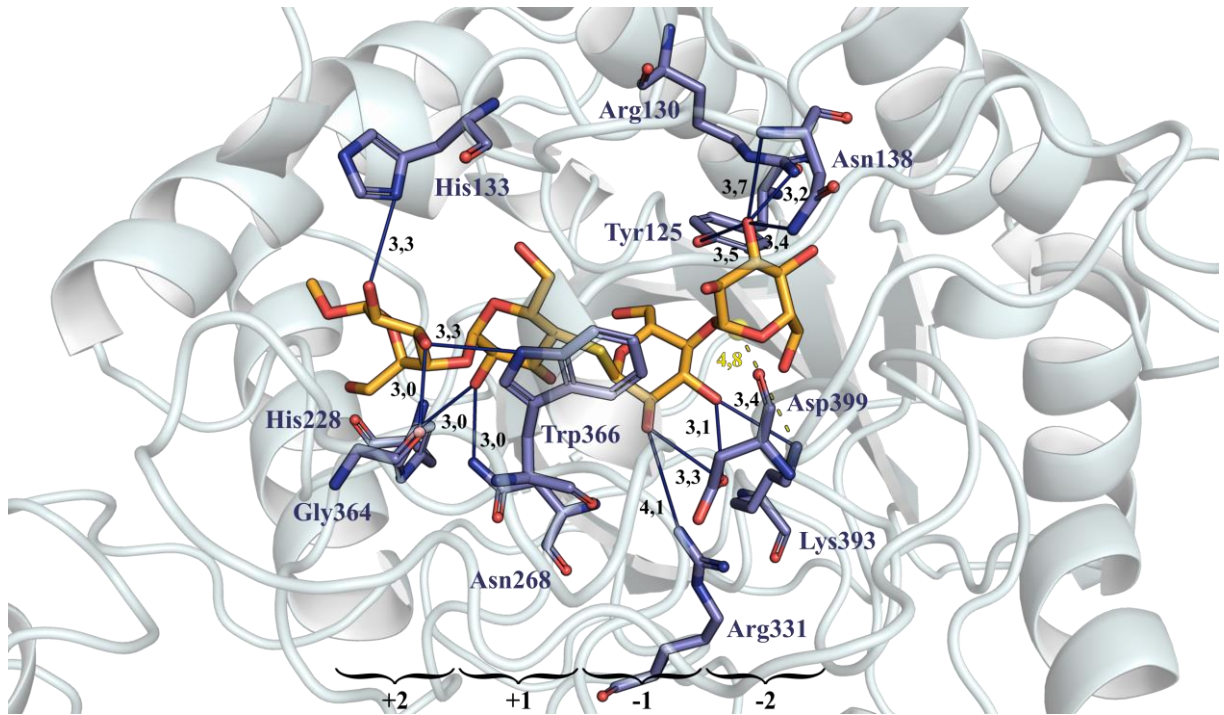


Fonte: Elaborada pelo autor.

Além disso, entre as celobiohidrolases GH6 estruturalmente mais próximas de XALc\_0484, encontradas no banco de dados PDB, a ligação com celotrioses não é comum e, quando ocorre, a sua hidrólise é incompleta ou muito lenta (KOIVULA et al., 2002; TAKEDA et al., 2022; TAYI et al., 2018). A posição ocupada por  $\beta$ -celotriose no sítio ativo (**Figura 30**), em conjunto com dados experimentais de enzimas próximas, reforça a hipótese de que, apesar de apresentar alto escore de afinidade, XALc\_0484 não seria eficiente na degradação deste polímero.

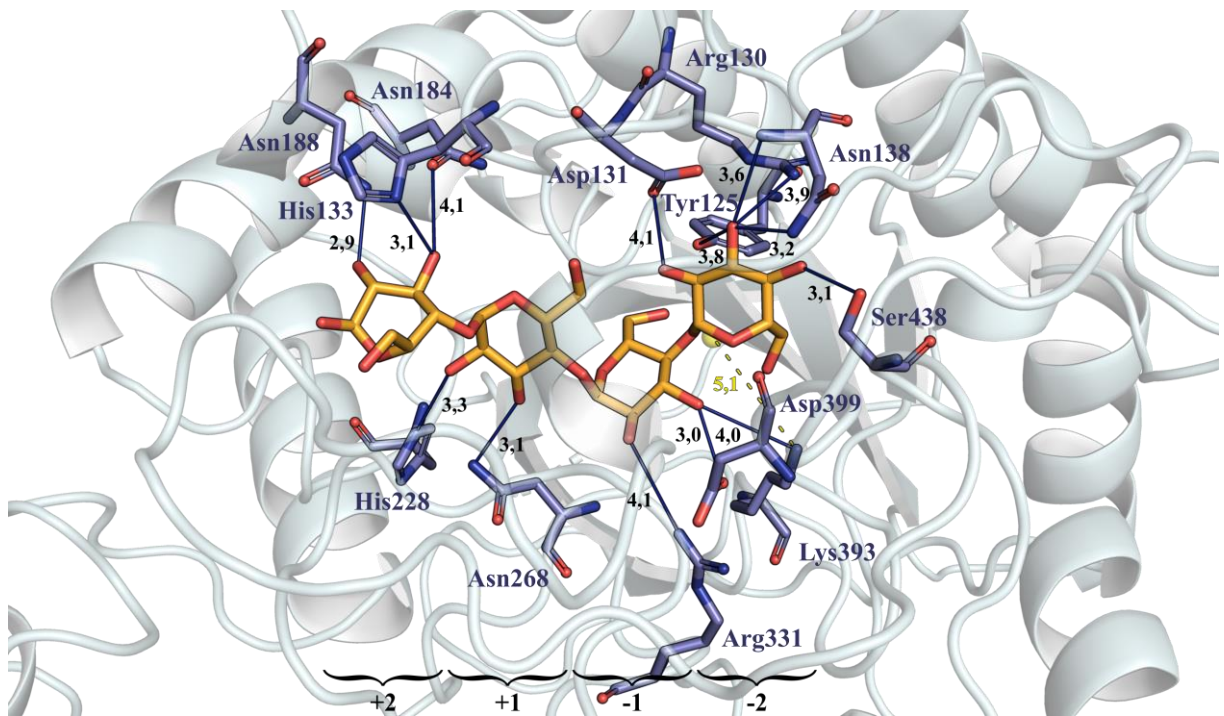
Em termos de afinidade de ligação, Metil-4 e  $\beta$ -celotetraose ocupam o segundo lugar, com pontuações de -9,90 e -9,83 kcal/mol, respectivamente (**Tabela 10**). Ambos os ligantes, compostos por quatro monômeros de glicose, foram posicionados nos subsítios -2 a +2, mantendo uma conformação semelhante. Sendo assim, provavelmente, os dois ligantes seriam clivados em celobioses (**Figura 31** e **Figura 32**).

**Figura 31** - Estrutura de XALc\_0484 complexada com Metil-4 posicionado nos subsítios -2 a +2. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

**Figura 32** - Estrutura de XALc\_0484 complexada com  $\beta$ -celotetraose posicionada nos subsítios -2 a +2. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

No entanto, é importante destacar que Metil-4 não é um composto natural. Este ligante foi projetado para ser não-hidrolisável e é frequentemente utilizado em estudos de interações enzimáticas (REVERBEL-LEROY et al., 1998; VARROT et al., 2002). Essa condição resultou em uma alteração na quantidade de resíduos (11 resíduos) que interagiram por intermédio de 14 ligações de hidrogênio, sendo três destas estabelecidas apenas com Asp399, o que manteve a estabilidade do substrato (**Tabela 10**). Outra interação relevante ocorreu com Tyr125, indicando que o substrato foi reconhecido e distorcido para a conformação ideal. Entretanto, por ser um ligante não-hidrolisável que contém um átomo de enxofre substituindo um átomo de oxigênio na ligação glicosídica central (**Figura 25**), não houve interação com os principais resíduos catalíticos Asp131 e Asp180, impossibilitando a provável clivagem da ligação (**Tabela 10 e Figura 31**).

Em comparação,  $\beta$ -celotetraose, sendo o segundo tetrassacarídeo, interagiu com dois resíduos a mais (13) do que Metil-4 e, conseqüentemente, obteve um aumento na quantidade de ligações de hidrogênio (15), incluindo a interação com Asp131. Isso indica que o sistema possivelmente estaria apto para a hidrólise (**Tabela 10 e Figura 32**). Semelhante a  $\beta$ -celotriose, ambos os ligantes tetrassacarídeos formaram longas pontes salinas com o resíduo Lys393 (**Tabela 10**). Vale ressaltar que, mesmo sendo um resíduo importante para a hidrólise, Ser137 dificilmente interage com o substrato, pois participa principalmente da rede hídrica com Asp131, e, portanto, não é mencionado nos resultados de docking (**Tabela 10**).

Os escores de afinidade e as análises visuais dos subsítios sugerem que, apesar de não possuir a pontuação mais elevada, XALc\_0484 apresenta afinidade significativa por  $\beta$ -celotetraose e Metil-4, conforme relatado para outras CBHs GH6 (NAKAMURA et al., 2020; SANDGREN et al., 2013; TAKEDA et al., 2022; TAYI et al., 2018; VARROT et al., 2002; WU et al., 2013). Todavia, seria relevante realizar testes de atividade enzimática com XALc\_0484 e os ligantes  $\beta$ -celobiose e  $\beta$ -celotriose, dado o posicionamento teoricamente desfavorável nos subsítios -1 a +2 e o fato de  $\beta$ -celotriose ser o ligante com maior escore de afinidade, mesmo esperando-se baixa eficiência desta atividade (KOIVULA et al., 2002; TAKEDA et al., 2022; TAYI et al., 2018). Esses testes de atividade enzimática poderiam fornecer uma validação experimental dos resultados de docking e oferecer uma compreensão mais completa da interação entre XALc\_0484 e seus potenciais substratos.

#### 4.5.4 Docking Molecular da Enzima Xalc\_0865

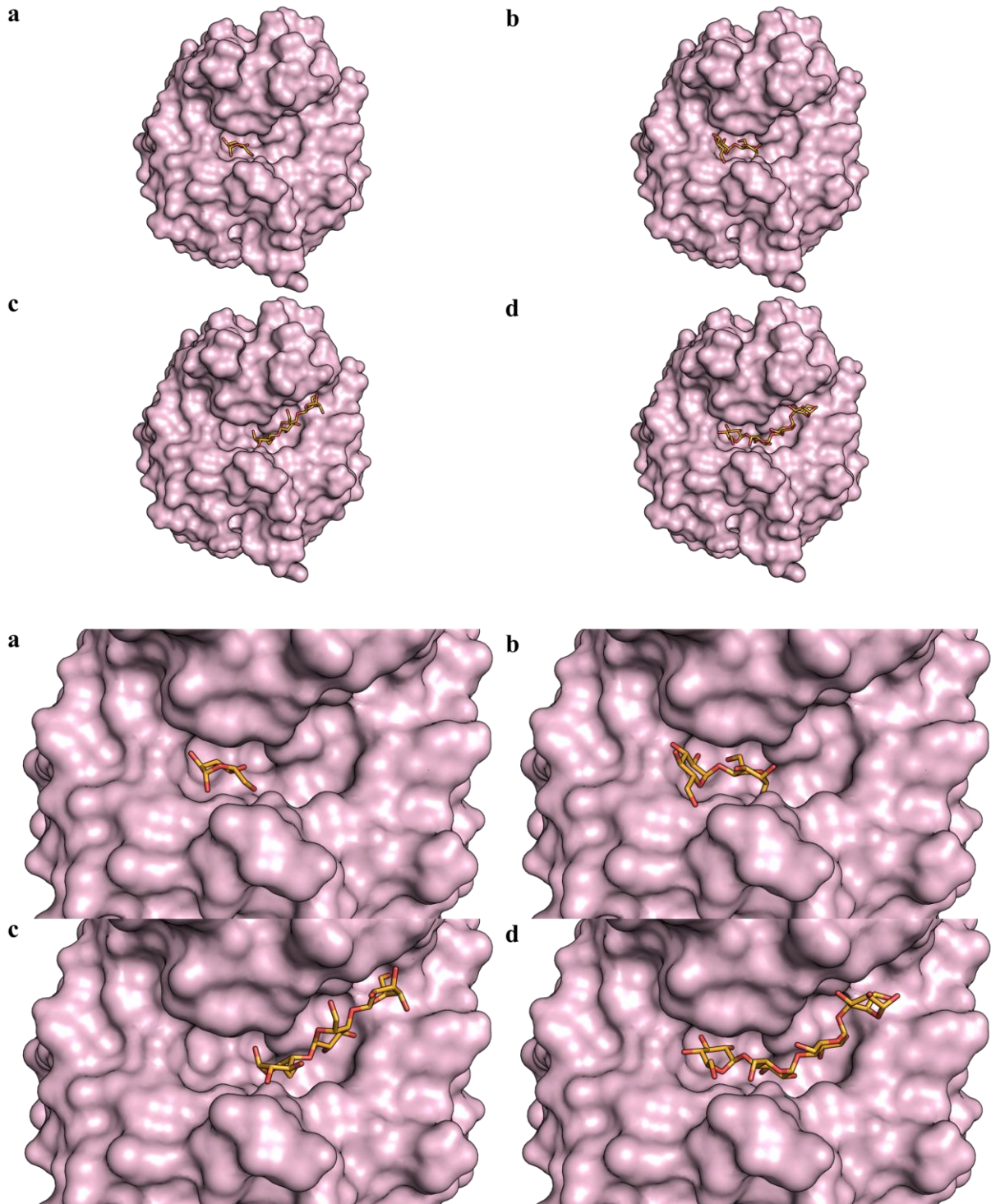
A triagem por ligantes da família GH12 resultou apenas em  $\beta$ -D-glicose,  $\beta$ -celobiose e  $\beta$ -celotetraose (**Tabela 9**). Contudo, a  $\beta$ -celotriose também foi adicionada ao conjunto de ligantes para as simulações de docking. Os resultados das simulações de docking molecular com a enzima XALc\_0865 demonstraram que o escore de afinidade aumentou conforme o incremento no número de monômeros de glicose nos ligantes (**Tabela 11 e Figura 33**).

**Tabela 11** - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc\_0865 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.

Ligante	Afinidade	Ligações de Hidrogênio
$\beta$ -D-glicose	-4,83	Asn52 (3,35), His95 (3,40), Asp136 (2,88), Asp136 (2,98), Asn186 (3,75), Tyr191 (2,95), Tyr191 (2,95) e Glu248 (3,68).
$\beta$ -celobiose	-6,44	Asn52 (3,27), His95 (3,12), Tyr96 (2,80), Tyr96 (2,80), Glu151 (2,81), Asn186 (4,07) e Tyr191 (3,61).
$\beta$ -celotriose	-7,13	Asn82 (4,04), Leu84 (3,3), His130 (2,87), Gly131 (3,96), Asn159 (2,87), Tyr161 (3,94), Asn186 (3,20) e Glu248 (3,87).
$\beta$ -celotetraose	-7,78	Asn52 (2,88), His95 (3,25), Tyr161 (3,87), Asn186 (3,08), Asn186 (3,77), Tyr191 (3,77), Glu248 (3,39) e Glu 248 (4,00).

**Fonte:** Elaborada pelo autor.

**Figura 33** - Estrutura de XALc\_0865 complexada com os quatro ligantes selecionados; (a)  $\beta$ -D-glicose; (b)  $\beta$ -celobiose; (c)  $\beta$ -celotriose; (d)  $\beta$ -celotetraose.



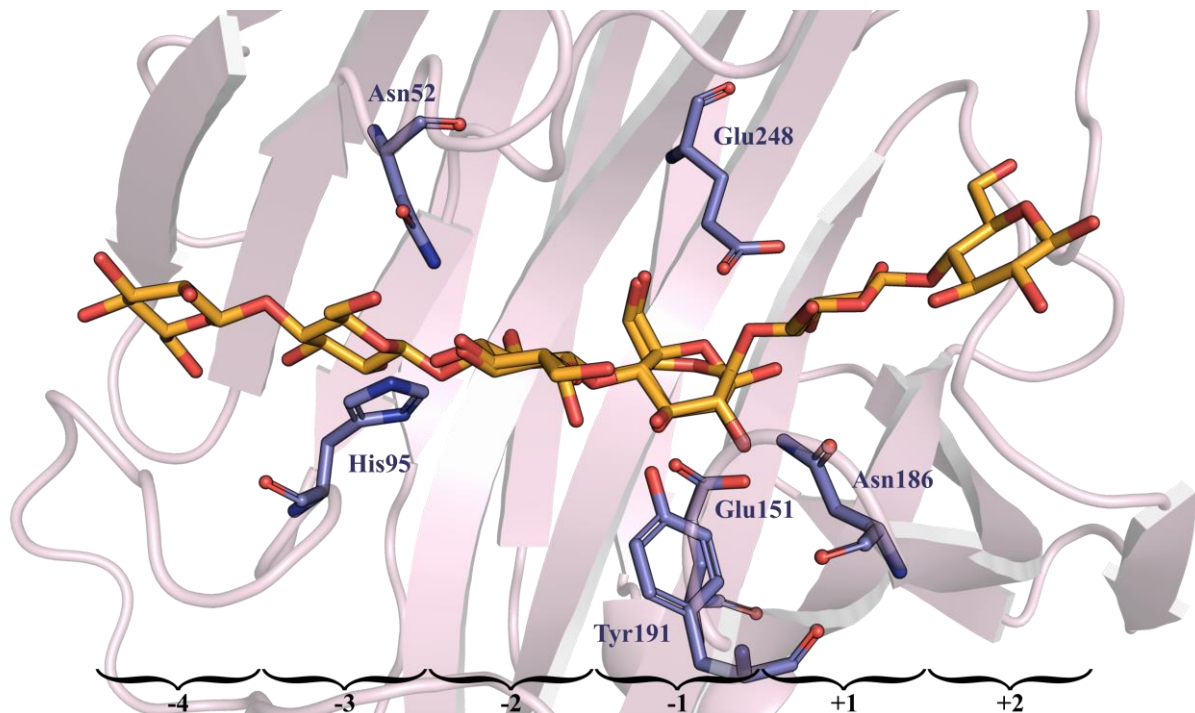
**Fonte:** Elaborada pelo autor.

Ao contrário das CBHs GH6, as endoglucanases GH12, como a enzima XALc\_0865, atuam nas regiões redutoras das cadeias de celulose e utilizam o mecanismo de retenção de duplo deslocamento para catalisar a hidrólise da ligação glicosídica. Esse mecanismo se baseia na utilização de um nucleófilo catalítico para formar um intermediário covalente, glicosil-

enzima. Em seguida, o intermediário é hidrolisado com a assistência de um resíduo ácido-base. Após a hidrólise, a configuração do carbono anomérico é retida, preservando assim a conformação do carboidrato hidrolisado (DAVIES; HENRISSAT, 1995).

Os principais resíduos catalíticos dessas enzimas são representados por dois glutamatos altamente conservados, localizados em lados opostos do subsítio -1. Em XALc\_0865, esses resíduos são Glu151, que desempenha o papel de nucleófilo catalítico, e Glu248, que atua como ácido-base durante a hidrólise (**Figura 34**). As interações entre o glutamato ácido-base (Glu248) e a celulose ocorrem por intermédio de longas ligações de hidrogênio (CRENNELL et al., 2006; SANDGREN et al., 2003), conforme evidenciado nos resultados de docking para os ligantes  $\beta$ -D-glicose,  $\beta$ -celotriose e  $\beta$ -celotetraose (**Tabela 11**). Além disso, outros resíduos essenciais como Asn52, His95, Asn186 e Tyr191 atuam como estabilizadores tanto do substrato no sítio ativo, quanto na manutenção da conformação estrutural da enzima (**Figura 34**). His95, além de ser um estabilizador, também auxilia durante o processo de hidrólise atuando como ácido-base, doando e/ou recebendo prótons (CRENNELL et al., 2006; SANDGREN et al., 2003, 2004; SULZENBACHER et al., 1999).

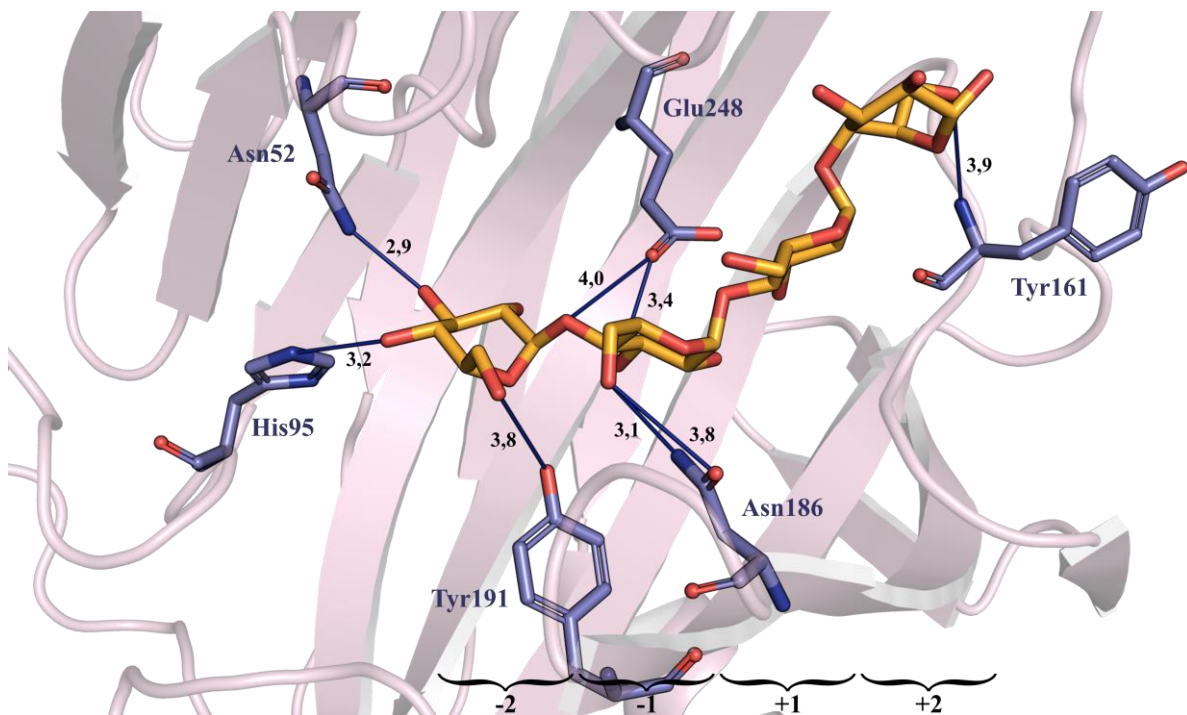
**Figura 34** - Visão geral dos principais subsítios negativos e positivos de XALc\_0865. Em roxo os principais resíduos que participam do processo catalítico. As duas celotetraoses (laranja) complexadas com XALc\_0865 foram obtidas a partir de sobreposição estrutural com as endoglucanases GH12 (Cel12A) de *Hemicola grisea* (PDB: 1uu5 e 1uu6).



Fonte: Elaborada pelo autor.

$\beta$ -celotetraose, com um escore de -7,78 kcal/mol, foi o ligante que demonstrou maior afinidade por XALc\_0865, ocupando os subsítios -2 a +2, em uma posição favorável para a hidrólise (**Figura 35**). Este ligante estabeleceu oito ligações de hidrogênio, mediadas por seis resíduos do sítio ativo, incluindo Asn52, His95, Asn186, Tyr191 e o ácido-base Glu248 (**Tabela 11**). Contudo, é importante salientar que apenas o resíduo Glu248 interagiu diretamente com  $\beta$ -celotetraose, enquanto Glu151, que atua como nucleófilo, não estabeleceu ligações de hidrogênio durante as simulações de docking (**Tabela 11 e Figura 35**).

**Figura 35** - Estrutura de XALc\_0865 complexada com  $\beta$ -celotetraose posicionada nos subsítios -2 a +2. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).

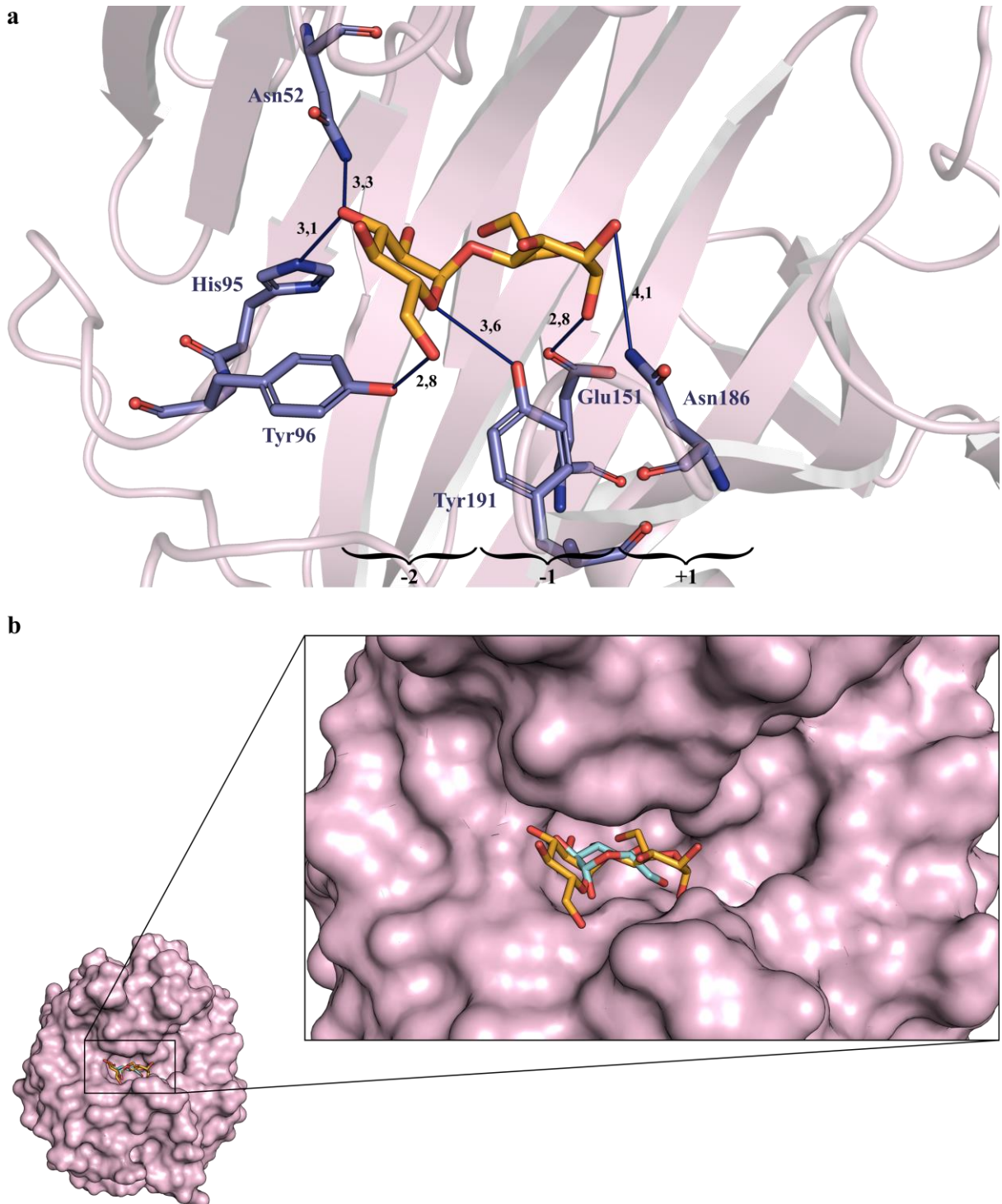


Fonte: Elaborada pelo autor.

Diferentemente de  $\beta$ -celotetraose e os demais ligantes,  $\beta$ -celobiose, que obteve o terceiro melhor escore de afinidade (-6,44 kcal/mol), foi o único ligante que estabeleceu ligações de hidrogênio com o nucleófilo catalítico Glu151, além de outros resíduos estabilizadores, como Asn52, His95 e Asn186 (**Tabela 11 e Figura 36**). No entanto, as enzimas GH12 requerem substratos com cadeias glicosídicas mais longas, uma vez que o reconhecimento do substrato é dependente da presença de glicose no subsítio -2 e a hidrólise ocorre na ligação entre os subsítios -1 e +1 (CRENNELL et al., 2006; CRENNELL; HREGGVIDSSON; NORDBERG KARLSSON, 2002; SULZENBACHER et al., 1999). Ligantes menores, como  $\beta$ -celobiose e  $\beta$ -D-glicose, ocuparam apenas os subsítios negativos -2

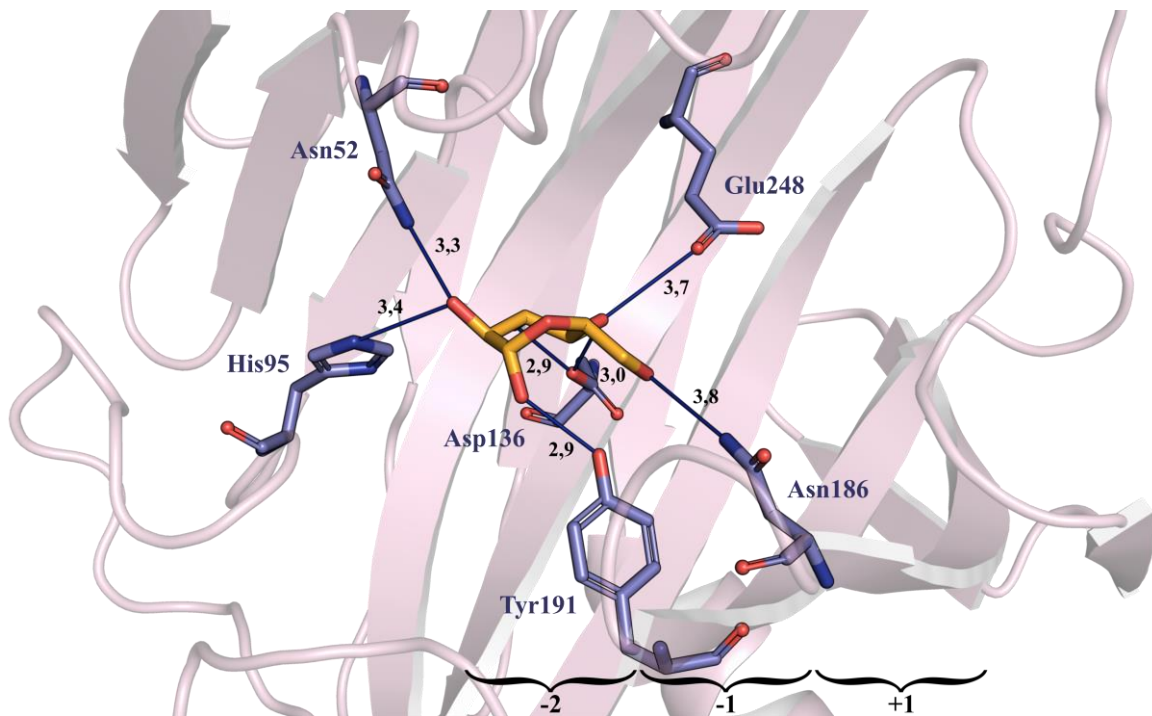
a -1, e -2, respectivamente (**Figura 36a, b e Figura 37**). Como não atingem o tamanho necessário para interagir com o subsítio -2, essencial para o reconhecimento, e ultrapassar as posições dos subsítios de clivagem -1 e +1, o mecanismo catalítico é ativado, mas o substrato pequeno não é clivado (CRENNELL et al., 2006). Além disso, assim como em XALc\_0484 (CBH GH6), na maioria dos estudos experimentais com GH12, a celobiose ocupa a posição dos subsítios do produto (+1 e +2 em enzimas redutoras ou -2 e -1 em não redutoras), por serem frequentemente os principais produtos da hidrólise (CRENNELL et al., 2006; SANDGREN et al., 2004; SULZENBACHER et al., 1999). Durante as simulações de docking molecular com  $\beta$ -celobiose, a pose mais favorável foi alcançada quando o ligante se posicionou nos subsítios -2 e -1 de XALc\_0865, destinados ao substrato de enzimas redutoras (**Figura 36**). Este resultado sugere que, embora  $\beta$ -celobiose interaja com o nucleófilo Glu151 e ocupe a posição de substrato no sítio de ligação, seu tamanho seria um obstáculo para a ocorrência de hidrólise e clivagem.

**Figura 36** - (a) Estrutura de XALc\_0865 complexada com  $\beta$ -celobiose posicionada nos subsítios -2 e -1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å); (b) Visão superficial dos subsítios ocupados por  $\beta$ -celobiose (laranja) e  $\beta$ -D-glicose (ciano).



**Fonte:** Elaborada pelo autor.

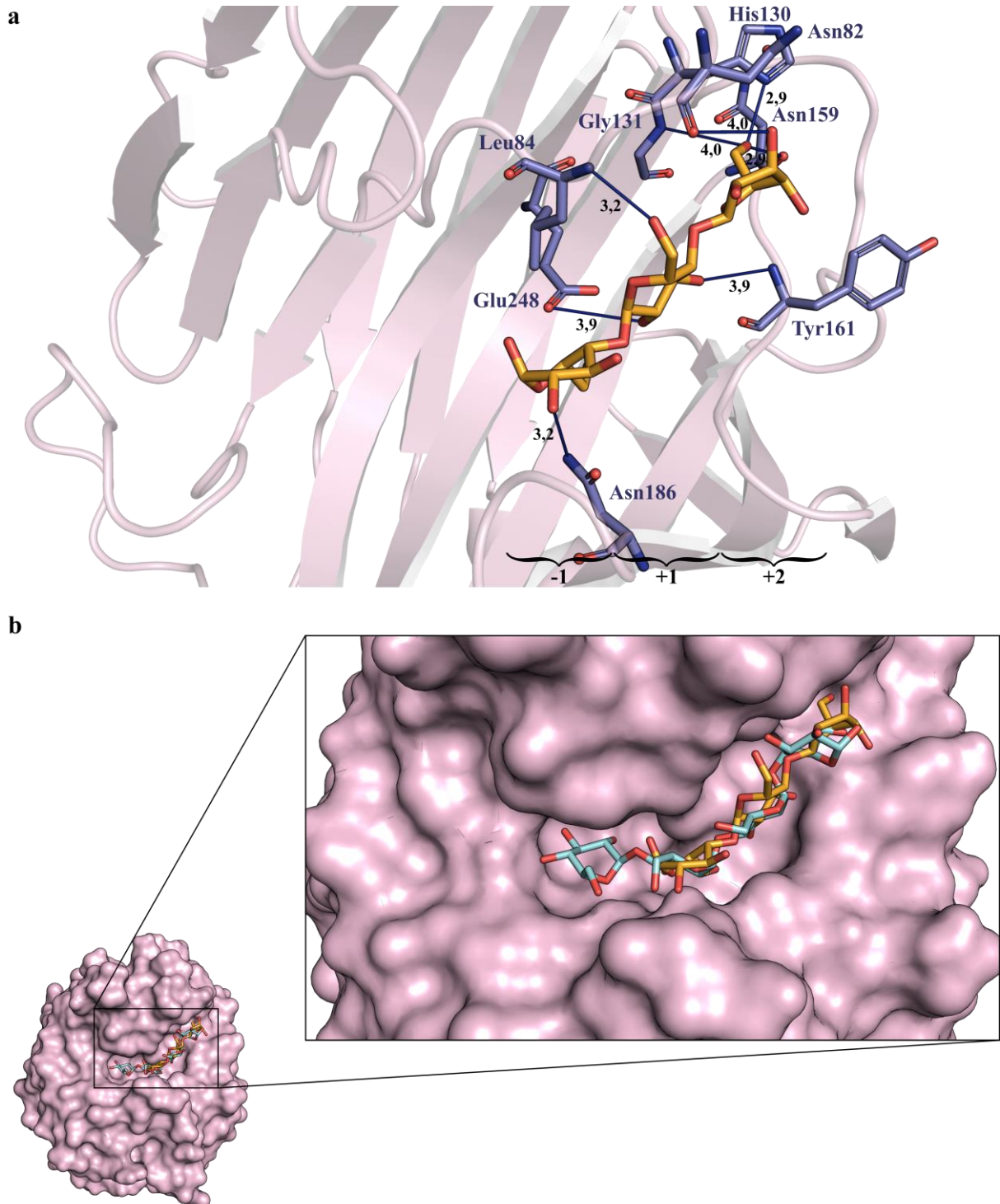
**Figura 37** - Estrutura de XALc\_0865 complexada com  $\beta$ -D-glicose posicionada no subsítio -2. As principais interações por ligações de hidrogênio (linha contínua azul-escuro), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

Os ligantes  $\beta$ -D-glicose e  $\beta$ -celotriose apresentaram escores de afinidade de -4,83 e -7,13 kcal/mol, respectivamente.  $\beta$ -D-glicose formou ligação de hidrogênio com o resíduo Glu248 e estabilizadores, entretanto, conforme mencionado anteriormente, este ligante não estaria em posição adequada para ser clivado (nem poderia, visto se tratar apenas de um monômero de glicose que não possui ligações glicosídicas) ou ser expelido pela enzima como produto de hidrólise (**Tabela 11** e **Figura 37**). Por outro lado,  $\beta$ -celotriose formou uma ligação de hidrogênio apenas com um resíduo catalítico (Glu248) e um estabilizador (Asn186) conservado, sendo posicionada nos subsítios -1 a +2 e não seria reconhecida para clivagem (**Tabela 11** e **Figura 38a**). Apesar de estar posicionada nos mesmos subsítios e manter uma conformação similar à  $\beta$ -celotetraose, o ligante  $\beta$ -celotriose interagiu com resíduos incomuns em comparação com os demais ligantes (**Figura 35** e **Figura 38**) e outras endoglucanases GH12 (CRENNELL et al., 2006; CRENNELL; HREGGVIDSSON; NORDBERG KARLSSON, 2002; SANDGREN et al., 2004; SULZENBACHER et al., 1999).

**Figura 38** - Estrutura de XALc\_0865 complexada com  $\beta$ -celotriose posicionada nos subsítios -1 a +2. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å); **(b)** Visão superficial da conformação e subsítios ocupados por  $\beta$ -celotriose (laranja) e  $\beta$ -celotetraose (ciano).



**Fonte:** Elaborada pelo autor.

Esses resultados destacam a complexidade das interações enzima-ligante e a importância de considerar não apenas a afinidade do ligante, mas também a posição e quais resíduos específicos estão envolvidos nessas interações. Em um mecanismo de retenção de

duplo deslocamento, como o observado em XALc\_0865, o resíduo Glu151 atua como um nucleófilo, formando um intermediário covalente com o substrato (CRENNELL et al., 2006; CRENNELL; HREGGVIDSSON; NORDBERG KARLSSON, 2002; DAVIES; HENRISSAT, 1995; SANDGREN et al., 2003, 2004; SULZENBACHER et al., 1999). Este processo dinâmico envolve uma mudança conformacional na enzima, que pode não ser totalmente capturada em uma simulação de docking estática. Embora as simulações de docking sejam excelentes para prever a orientação e posição de um ligante em um sítio de ligação, elas apresentam limitações ao capturar a dinâmica completa de uma reação enzimática, incluindo a ativação do nucleófilo e a formação de intermediários covalentes (BENDER et al., 2021; DOS SANTOS; FERREIRA; ANDRICOPULO, 2018; KIRCHMAIR et al., 2012; TAO et al., 2020; THOMAS; PARRILL; BAKER, 2022; TORRES et al., 2019). Portanto, é possível que Glu151 não apareça interagindo com os ligantes nas simulações de docking, visto que sua ativação e subsequente ataque nucleofílico ao substrato são eventos dinâmicos que ocorrem durante a catálise. Essa hipótese poderia ser analisada por meio de técnicas de dinâmica molecular ou experimentos bioquímicos *in vitro*.

#### 4.5.5 Docking Molecular da Enzima Xalc\_2967

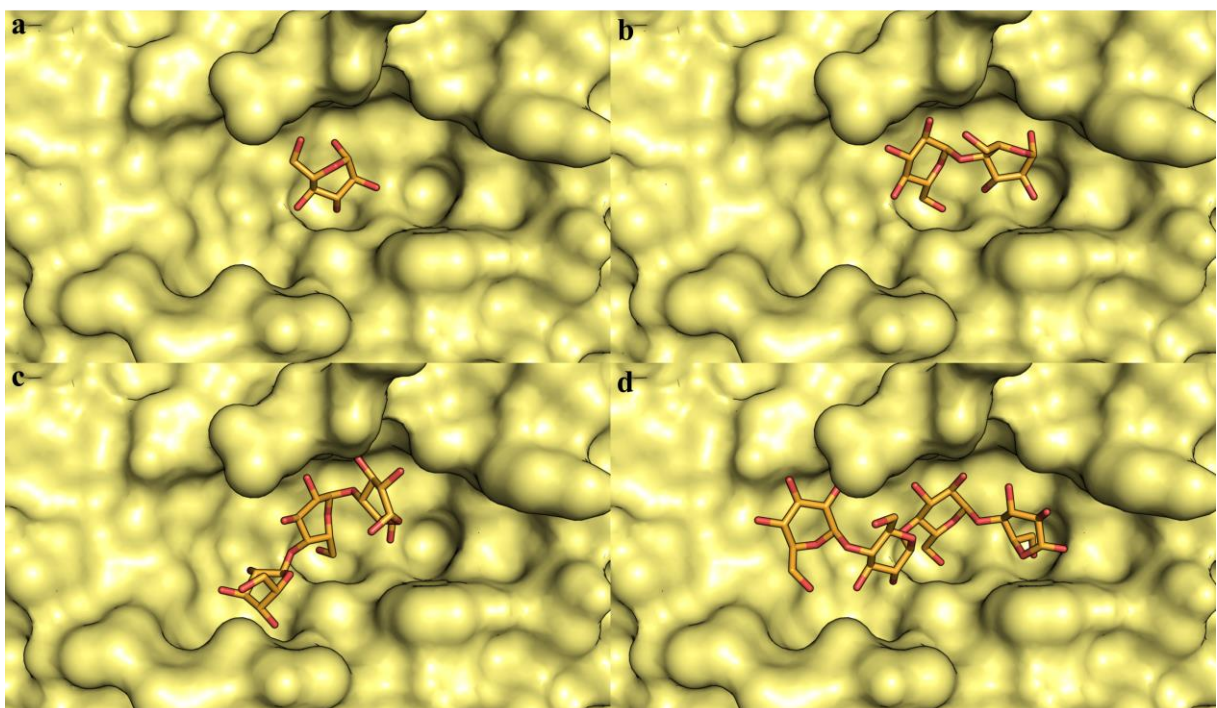
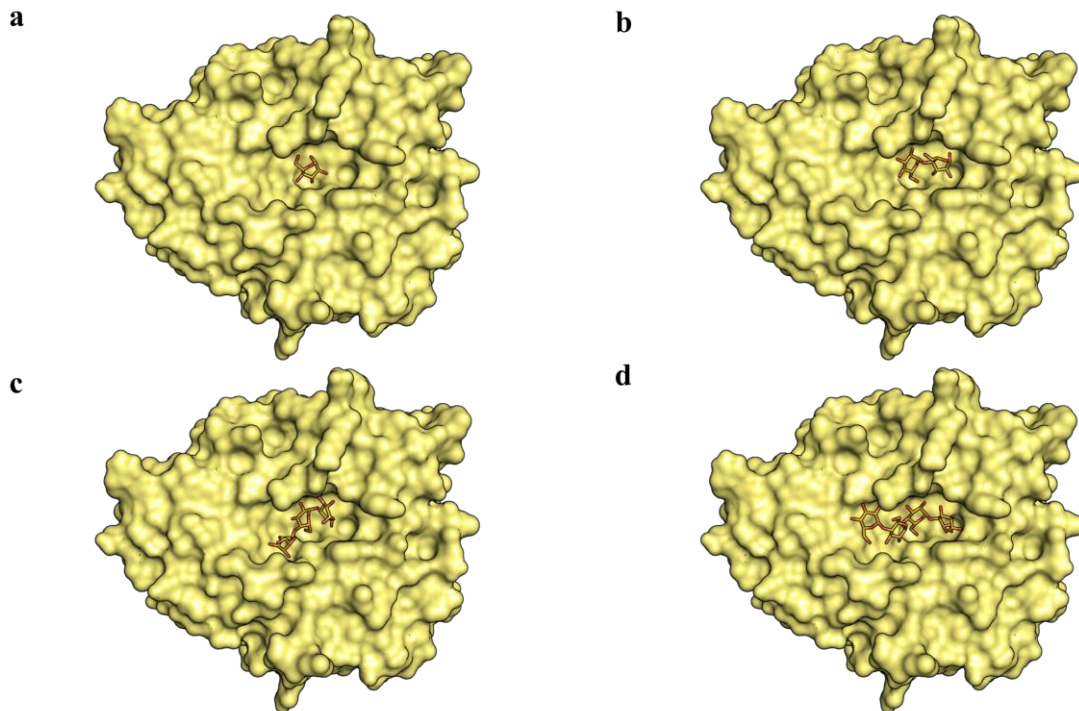
Com base nas análises de sobreposição estrutural previamente realizadas (**Figura 13**), sugere-se que XALc\_2967 corresponda a uma endoglucanase pertencente à classe enzimática das hidrolases glicosídicas, mais especificamente à família 5 e subfamília 1 (GH5\_1). No entanto, a confirmação dessa classificação exigiria testes de atividade enzimática. A família GH5 é uma das maiores famílias de GHs, abrangendo uma grande diversidade com 32 tipos de atividades enzimáticas associadas a um amplo espectro de carboidratos (ASPEBORG et al., 2012; DRULA et al., 2022; HENRISSAT et al., 1989).

O processo de subdivisão em subfamílias, como GH5\_1, permite a organização e agrupamento de enzimas com maior correlação sequencial, especificidade enzimática e substrato comum (ASPEBORG et al., 2012). A subfamília GH5\_1, é uma das 57 subfamílias na família GH5 e predominantemente composta por enzimas com atividade de endo- $\beta$ -1,4-glucanase (EC 3.2.1.4). Outras duas atividades enzimáticas classificadas para GH5\_1, mas com menos representantes, incluem liquenase/endo- $\beta$ -1,3-glucanase (EC 3.2.1.73) e exo- $\beta$ -1,4-glucanase/celodextrinase (EC 3.2.1.74) (ASPEBORG et al., 2012; DRULA et al., 2022).

No banco de dados CAZy as estruturas experimentais de algumas endoglucanases da subfamília GH5\_1 estão complexadas apenas com  $\beta$ -celobiose e/ou  $\beta$ -celotetraose. Contudo, as

simulações de docking molecular com XALc\_2967 foram realizadas incluindo outros ligantes frequentemente utilizados em complexos experimentais de enzimas da família GH5 (**Tabela 9**). Após as simulações de docking, todos os quatro ligantes selecionados foram posicionados no sítio ativo de XALc\_2967 (**Figura 39**).

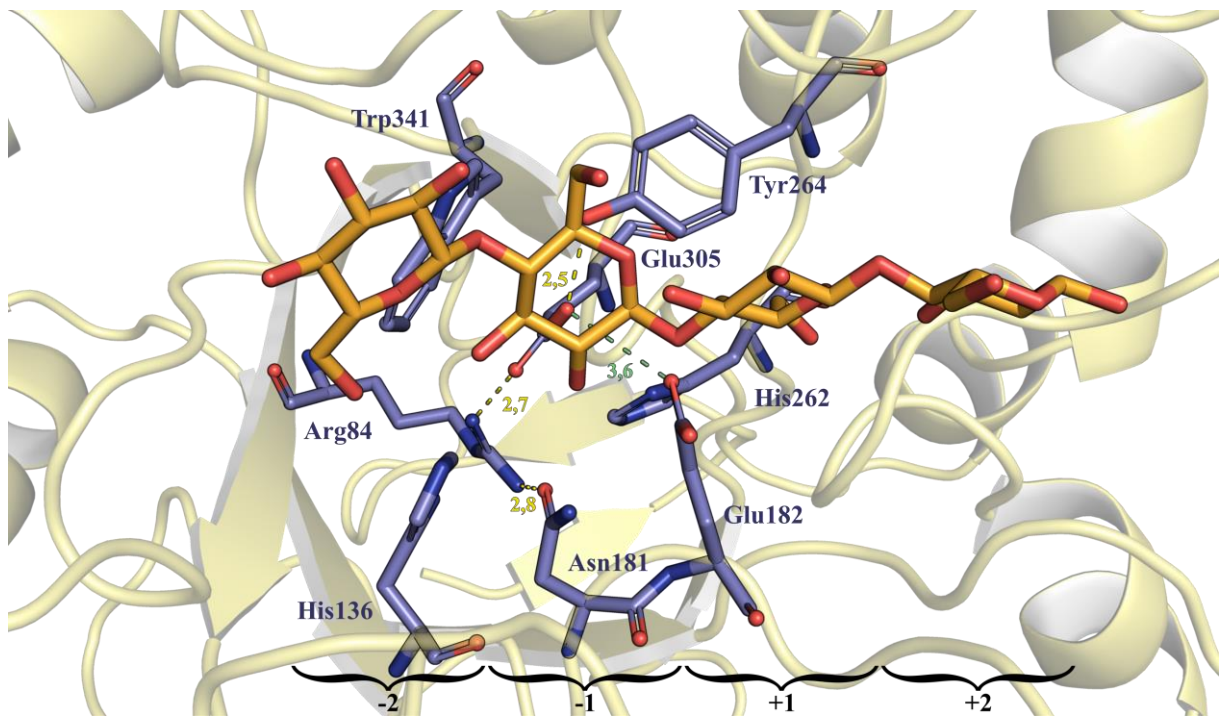
**Figura 39** - Estrutura de XALc\_2967 complexada com os cinco ligantes selecionados: (a)  $\beta$ -D-glicose; (b)  $\beta$ -celobiose; (c)  $\beta$ -celotriose; (d)  $\beta$ -celotetraose.



**Fonte:** Elaborada pelo autor.

Assim como XALc\_0865, XALc\_2967 também é classificada como uma endoglucanase (EC 3.2.1.4) e utiliza o mesmo mecanismo de retenção de duplo deslocamento para catalisar a hidrólise da ligação glicosídica. Nas GH5\_1, os principais resíduos catalíticos também consistem em dois glutamatos conservados, que atuam como nucleófilo e ácido-base (KIM; ISHIKAWA, 2011; PUHL et al., 2019; SAKON et al., 1996). Em XALc\_2967, Glu305, posicionado no subsítio -1, atua como nucleófilo, enquanto Glu182, localizado entre os subsítios -1 e +1, atua como ácido-base. Ambos os resíduos mantêm uma distância de 3,6 Å entre si (**Figura 40**), uma característica comum em muitas GHs que utilizam o mecanismo de retenção (DAVIES; HENRISSAT, 1995).

**Figura 40** - Visão geral dos principais subsítios negativos e positivos de XALc\_2967. Em roxo os principais resíduos que participam do processo catalítico. A celotetraose (laranja) complexada com XALc\_2967 foi obtida a partir de sobreposição estrutural com o domínio catalítico da endoglucanase GH5\_1 (Cel5A) de *Acidothermus cellulolyticus* (PDB: 1ece). As linhas tracejadas em amarelo representam as ligações de hidrogênio entre alguns resíduos que estabilizam o sítio ativo. A linha tracejada em verde-claro indica a distância entre os dois glutamatos catalíticos. Todas as distâncias são mostradas em angstroms (Å).



Fonte: Elaborada pelo autor.

Além desses dois glutamatos principais, outros seis resíduos, que compreendem Arg84, His136, Asn181, His262, Tyr264 e Trp341, desempenham papéis importantes no processo de hidrólise e estão distribuídos nos subsítios -2 a +1. Esses resíduos não apenas contribuem para a estabilização do substrato, mas também estabelecem ligações de hidrogênio entre si,

mantendo a estabilidade da estrutura do sítio ativo (**Figura 40**) (PUHL et al., 2019; SABINI et al., 2000; SANTOS et al., 2012).

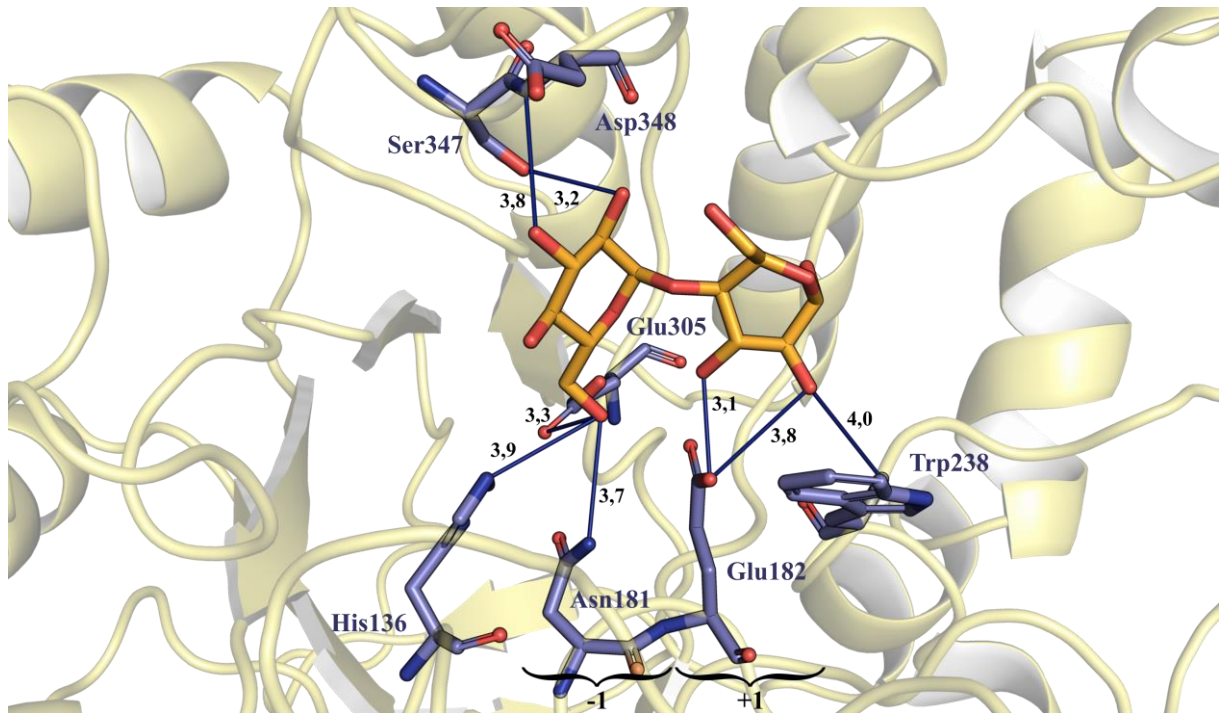
O padrão de afinidade observado entre XALc\_2967 e os ligantes assemelha-se ao observado em XALc\_0865, apresentando um aumento no escore de ligação correspondente ao aumento no tamanho do ligante. O ligante com a melhor afinidade foi a  $\beta$ -celotetraose, com pontuação de -8,74 kcal/mol, enquanto a  $\beta$ -D-glicose apresentou a menor afinidade e pontuação de -5,26 kcal/mol (**Tabela 12**). Todos os ligantes estabeleceram interações com o resíduo catalítico ácido-base Glu182. Ainda assim, replicando os resultados obtidos para XALc\_0865, a  $\beta$ -celobiose foi o único ligante que interagiu com o nucleófilo catalítico Glu305 e, neste caso, estabelecendo também uma interação com o ácido-base Glu182 (**Tabela 12 e Figura 41**). Adicionalmente, todos os ligantes interagiram com pelo menos dois resíduos (His136 e Asn181) importantes para a estabilização do substrato posicionado no subsítio de clivagem -1 (**Tabela 12 e Figura 41 - Figura 44**).

**Tabela 12** - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc\_2967 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.

Ligante	Afinidade	Ligações de Hidrogênio	Pontes salinas
$\beta$ -D-glicose	-5,26	His136 (3,01), Asn181 (3,64), Ser347 (3,88), Ser347 (2,92) e Asp348 (3,88).	-
$\beta$ -celobiose	-7,27	His136 (3,87), Asn181 (3,66), Glu182 (3,05), Glu182 (3,80), Trp238 (4,01), Glu305 (3,34), Ser347 (3,18) e Asp348 (3,84).	-
$\beta$ -celotriose	-8,33	Glu54 (3,52), His136 (3,62), Ile143 (3,33), Glu182 (2,90), Gly268 (2,77) e Ser347 (2,87).	His136 (5,13)
$\beta$ -celotetraose	-8,74	His136 (3,12), Asn141 (3,78), Asn181 (3,63), Glu182 (3,07), Trp238 (3,05), Trp238 (3,94), Gly268 (3,69), Asn346 (3,16), Ser347 (3,02), Asp348 (3,17) e Asp348 (3,86)	-

**Fonte:** Elaborada pelo autor.

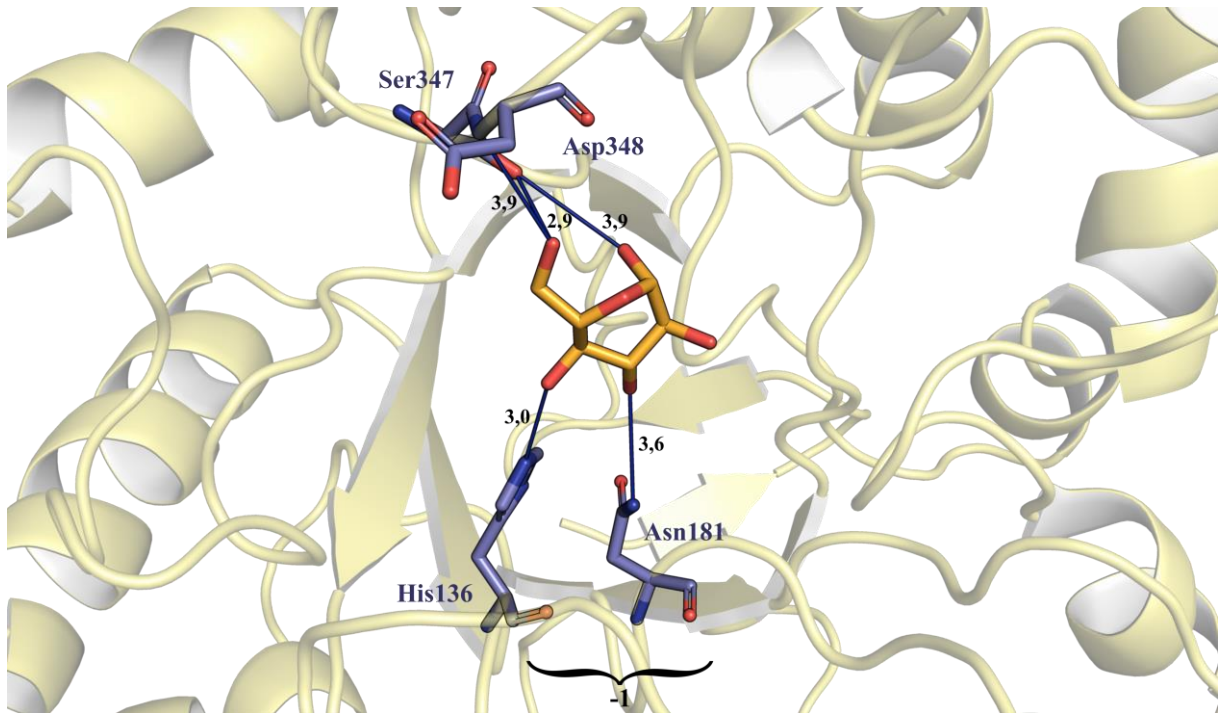
**Figura 41** - Estrutura de XALc\_2967 complexada com  $\beta$ -celobiose posicionada nos subsítios -1 e +1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

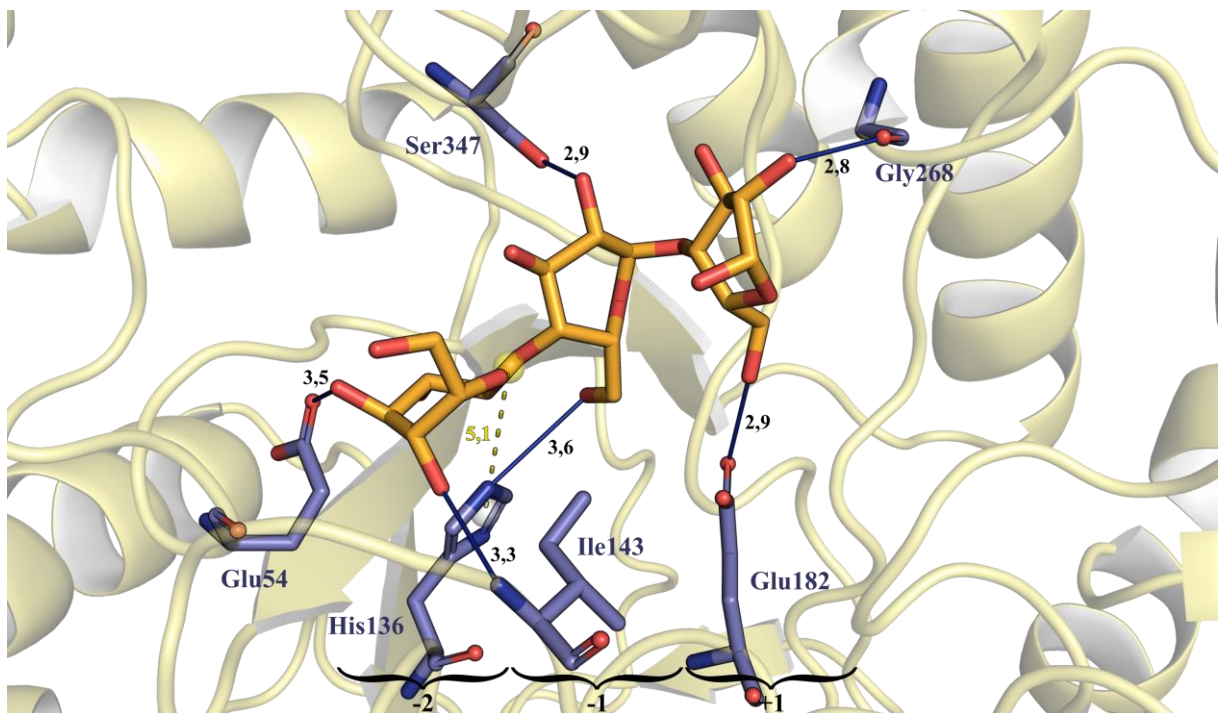
É relevante destacar que XALc\_2967 apresentou proximidade estrutural com endoglucanases GH5\_1 (PDB: 4tuf, 3axx e 1ece) que liberam celobiose como principal produto de hidrólise (KIM; ISHIKAWA, 2011; PUHL et al., 2019; SAKON et al., 1996). Entretanto, em XALc\_2967, os ligantes foram posicionados desde os subsítios negativos até o subsítio positivo +1 (**Figura 41** - **Figura 44**). Inclusive, a  $\beta$ -celotetraose, que teoricamente seria o substrato preferencial para endoglucanases GH5\_1 hidrolisarem em celobiose, ocupou os subsítios -3 a +1 (**Figura 44**). Este resultado de interações proteína-ligante difere do observado em outras GH5\_1 complexadas com o tetrassacarídeo (BAKER et al., 2005; KIM; ISHIKAWA, 2011; PUHL et al., 2019; SAKON et al., 1996). No entanto, essas observações sugerem a possibilidade de que XALc\_2967 possa liberar um monômero de glicose como principal produto da reação de hidrólise, um cenário contrário aos resultados de sobreposição estrutural (**Figura 13**). Contudo, considerando a diversidade de enzimas na família GH5 e a natureza versátil das endoglucanases, seria plausível encontrar uma enzima GH5\_1 com preferência por hidrolisar oligossacarídeos em glicose.

**Figura 42** - Estrutura de XALc\_2967 complexada com  $\beta$ -D-glicose posicionada no subsítio -1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



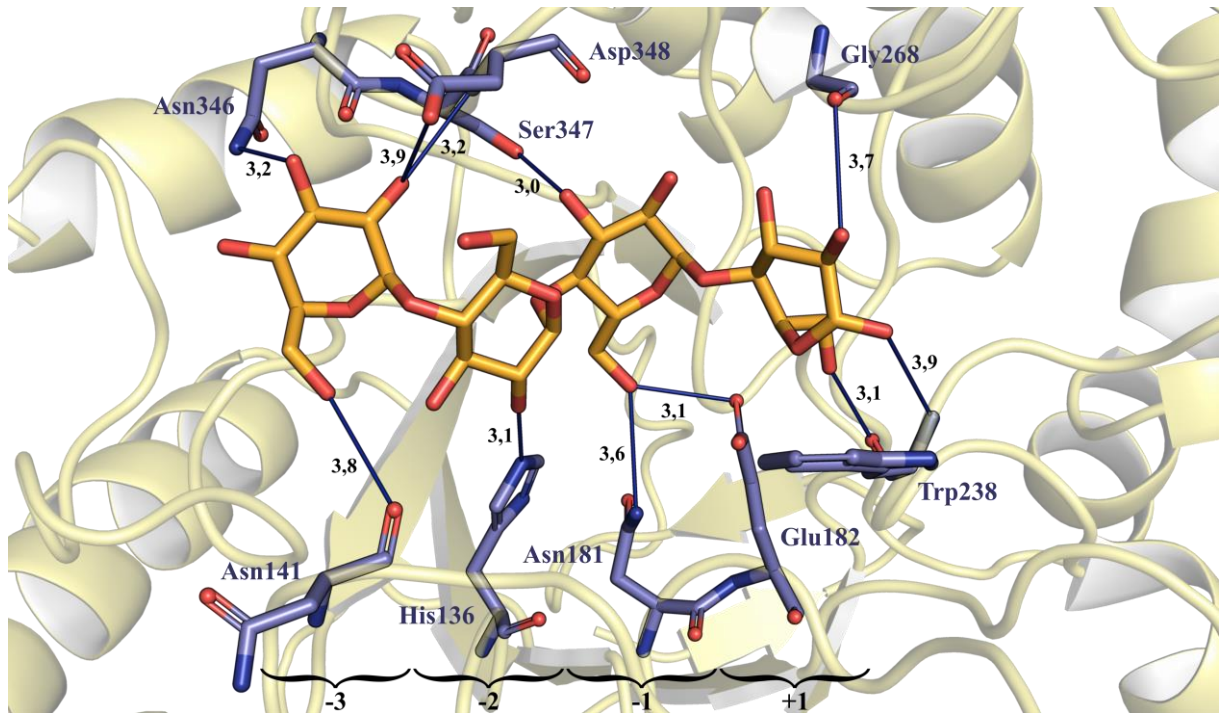
Fonte: Elaborada pelo autor.

**Figura 43** - Estrutura de XALc\_2967 complexada com  $\beta$ -celotriose posicionada nos subsítios -2 a +1. As principais interações por pontes salinas (linha tracejada amarela) e ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

**Figura 44** - Estrutura de XALc\_2967 complexada com  $\beta$ -celotetraose posicionada nos subsítios -3 a +1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

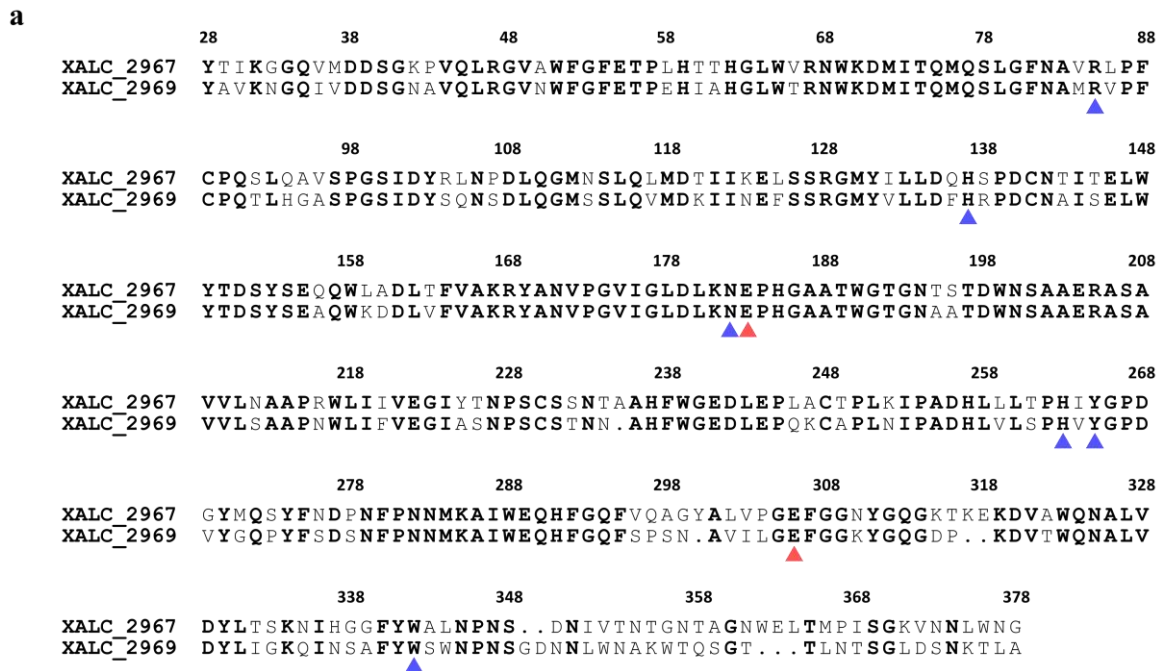
Durante as simulações de docking, algumas poses indicaram a ocupação nos subsítios +1 e +2, sugerindo uma possível clivagem em celobiose (**dados não mostrados**). No entanto, vale ressaltar que, em 400 simulações por ligante, as poses de maior afinidade para os cinco ligantes selecionados ocuparam os subsítios negativos até o positivo +1. Este padrão de posicionamento sugere que as conformações preferenciais estariam em concordância com o mecanismo catalítico, inferindo que XALc\_2967 possa apresentar uma atividade enzimática diferente daquela anotada nos bancos de dados. Dada a descoberta de novas CAZymes, incluindo novas GH5 (BAKER et al., 2005; KIM; ISHIKAWA, 2011; MARTINS et al., 2022; PUHL et al., 2019; SAKON et al., 1996), seria indispensável conduzir pesquisas adicionais, como a atividade enzimática *in vitro*, para confirmar essa hipótese e obter uma compreensão mais completa do mecanismo de ação de XALc\_2967.

#### 4.5.6 Docking Molecular da Enzima Xalc\_2969

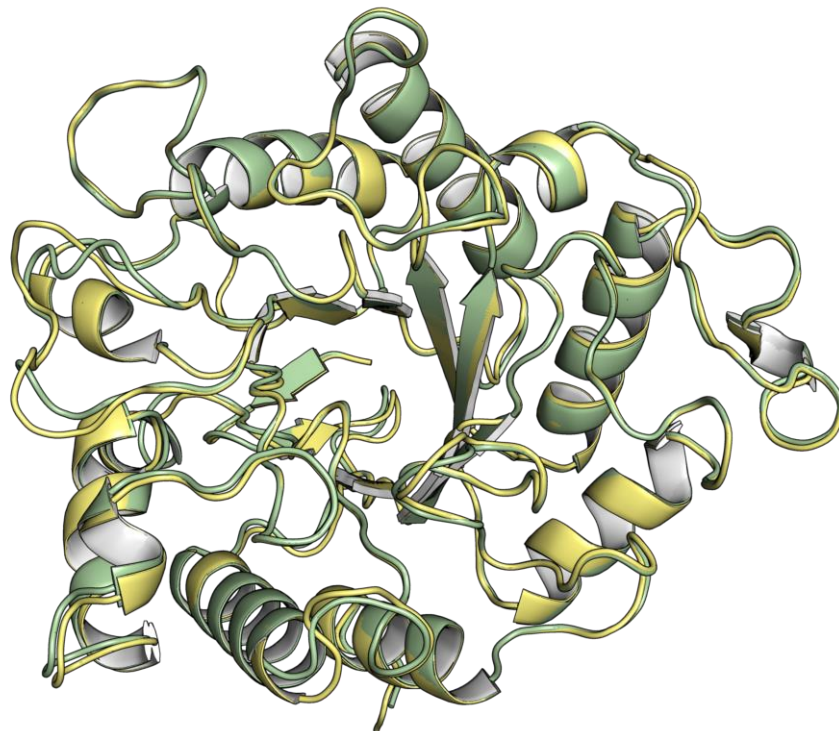
XALc\_2969 e XALc\_2967 são notoriamente semelhantes em termos de sequência e estrutura, compartilhando os mesmos glutamatos catalíticos posicionados nos subsítios -1 (nucleófilo) e entre -1 e +1 (ácido-base) (**Figura 45 - Figura 47**). Em relação aos aminoácidos

catalíticos, a principal distinção entre as duas enzimas, reside na posição desses resíduos na sequência de aminoácidos. Em XALc\_2967 o nucleófilo é representado por Glu305, enquanto em XALc\_2969 é representado por Glu303. Entretanto, o glutamato ácido-base permanece como Glu182 em ambas as enzimas. Além disso, alguns dos seis resíduos importantes para o processo de hidrólise e estabilização estrutural, sendo eles, Arg84, His136, Asn181, His261, Tyr263 e Trp337, também apresentam variações em sua posição na sequência (**Figura 45 e Figura 46**). É importante destacar que os três últimos resíduos, que diferem de XALc\_2967 (His262, Tyr264 e Trp341), ocupam a mesma posição sequencial observada na endoglucanase GH5\_1 de *Xanthomonas campestris* pv. *campestris*, XccCel5A (PDB: 4tuf) (PUHL et al., 2019) (**Figura 46**).

**Figura 45 - (a)** Alinhamento entre as seqüências do domínio catalítico de XALc\_2969 e XALc\_2967. Os resíduos idênticos estão em negrito. Os triângulos vermelhos sinalizam os dois glutamatos catalíticos e os triângulos azuis apontam os seis resíduos estabilizadores. **(b)** Sobreposição estrutural entre XALc\_2969 (verde-claro) e XALc\_2967 (amarelo-claro).

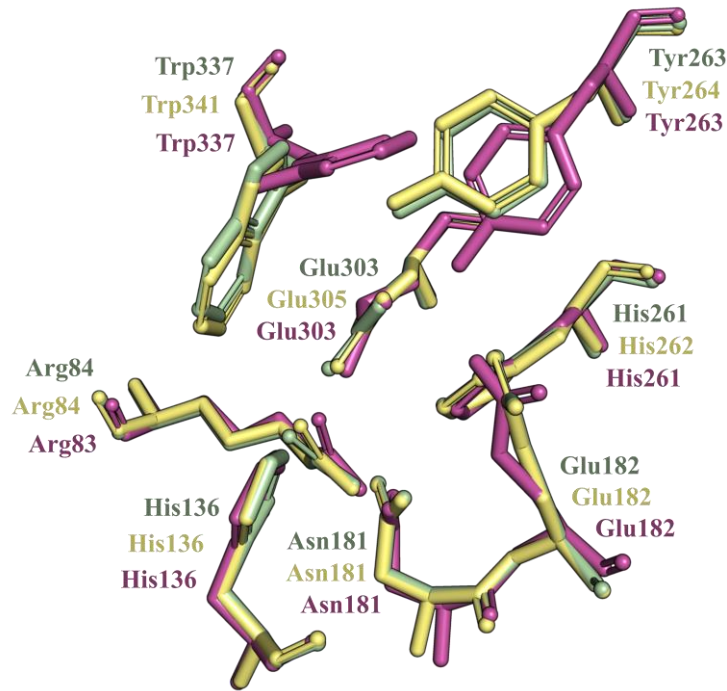


**b**



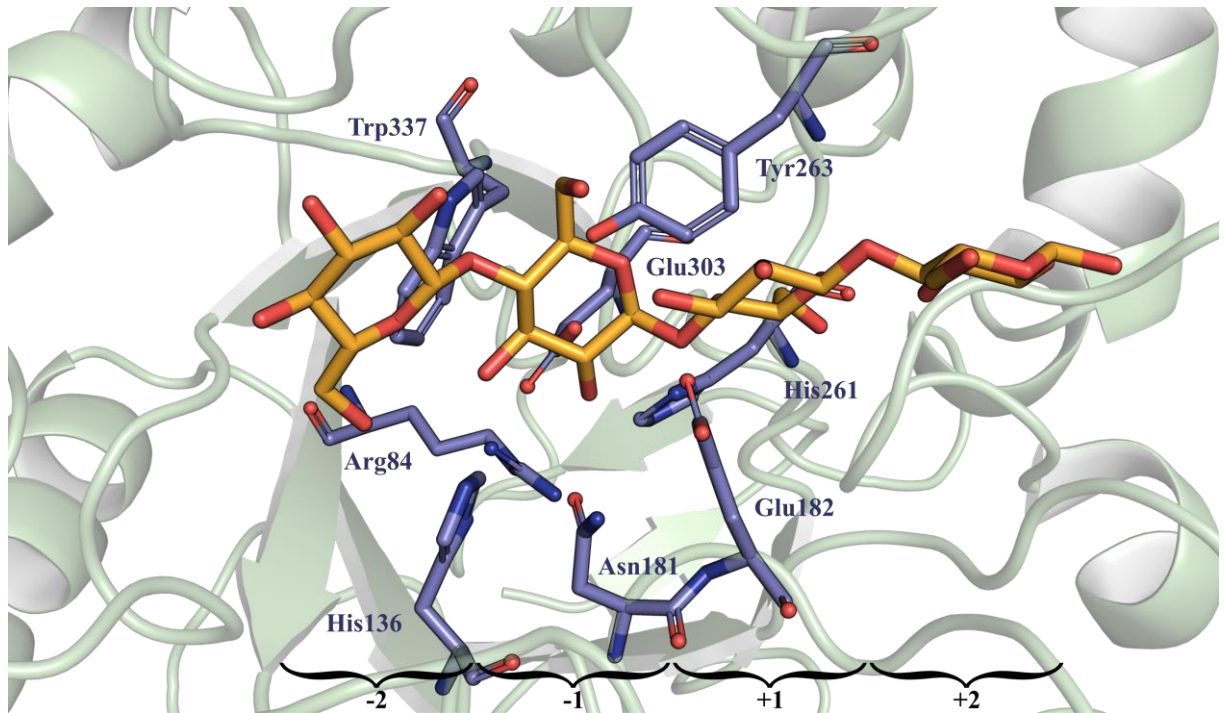
Fonte: Elaborada pelo autor.

**Figura 46** - Sobreposição estrutural entre os oito (Glu182, Glu305, Arg84, His136, Asn181, His261, Tyr263 e Trp337) principais resíduos que participam do processo catalítico de XALc\_2969 (verde-claro), XALc\_2967 (amarelo-claro) e XccCel5A (PDB: 4tuf) (magenta).



Fonte: Elaborada pelo autor.

**Figura 47** - Visão geral dos principais subsítios negativos e positivos de XALc\_2969. Em roxo os principais resíduos que participam do processo catalítico. A celotetraose (laranja) complexada com XALc\_2969 foi obtida a partir de sobreposição estrutural com o domínio catalítico da endoglucanase GH5\_1 (Cel5A) de *Acidothermus cellulolyticus* (PDB: 1ece).



Fonte: Elaborada pelo autor.

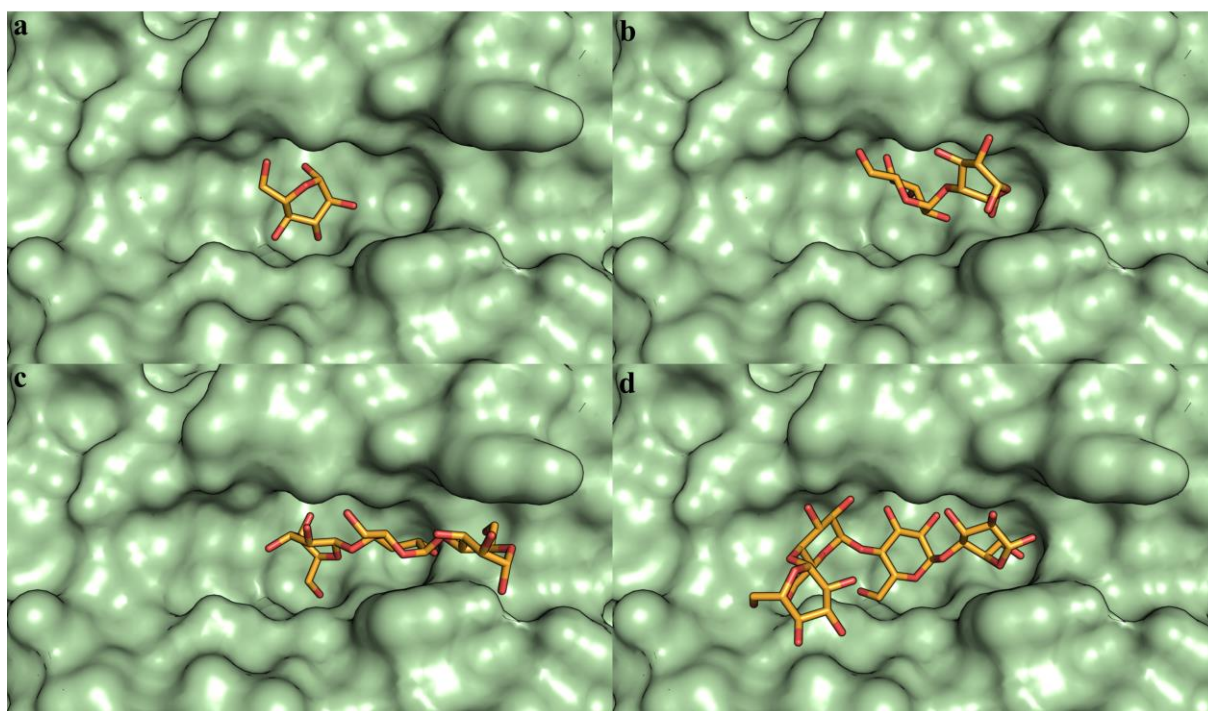
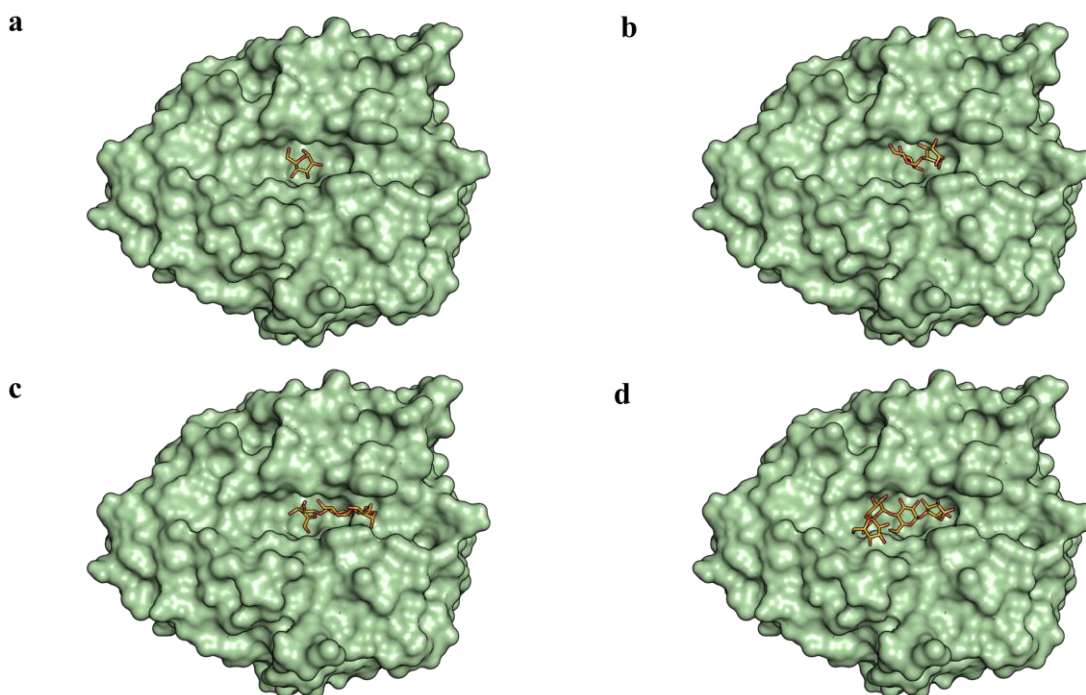
Apesar das diferenças na posição sequencial, é importante observar que todos os oito resíduos envolvidos no processo catalítico mantêm a mesma posição e conformação estrutural nas três enzimas (**Figura 46**). Portanto, os dados obtidos para XALc\_2967 podem ser extrapolados para XALc\_2969. Seguindo o padrão observado nas endoglucanases anteriores (XALc\_0865 e XALc\_2967), o aumento no escore de afinidade esteve diretamente relacionado ao aumento no tamanho do ligante (**Tabela 13**). Contudo, embora os ligantes posicionados no sítio de ligação de XALc\_2969 (**Figura 48**) estabeleçam mais interações com os resíduos em comparação com as interações de XALc\_2967, os escores de afinidade para XALc\_2969 foram ligeiramente inferiores (**Tabela 12 e Tabela 13**).

**Tabela 13** - Afinidade de ligação e interações proteína-ligante geradas a partir das simulações de docking molecular entre a enzima XALc\_2969 e os ligantes selecionados. As distâncias das ligações, representadas em angstroms (Å), estão descritas entre parênteses.

<b>Ligante</b>	<b>Afinidade</b>	<b>Ligações de Hidrogênio</b>
$\beta$ -D-glicose	-5,53	His136 (3,15), Asn181 (3,59), Glu303 (3,80), Trp337 (3,96), Asp345 (3,98) e Asp345 (3,50).
$\beta$ -celobiose	-6,87	Glu182 (2,84), Glu182 (3,55), Tyr263 (2,87), Val267 (3,03), Glu303 (3,10), Trp337 (2,81), Ser343 (2,71), Gly344 (3,58) e Asp345 (3,60).
$\beta$ -celotriose	-7,49	His136 (3,73), Asn181 (3,29), His235 (3,17), Trp237 (2,82), Trp237 (3,26), Glu303 (2,97) e Asp345 (3,70).
$\beta$ -celotetraose	-8,31	His136 (3,28), Arg137 (2,93), Arg137 (3,19), Asn141 (2,91), Trp237 (3,08), Trp237 (3,74), Trp337 (3,19), Gly344 (3,37), Asp345 (3,26), Asp345 (2,71) e Asp345 (3,96).

**Fonte:** Elaborada pelo autor.

**Figura 48** - Estrutura de XALc\_2969 complexada com os cinco ligantes selecionados; (a)  $\beta$ -D-glicose; (b)  $\beta$ -celobiose; (c)  $\beta$ -celotriose; (d)  $\beta$ -celotetraose.

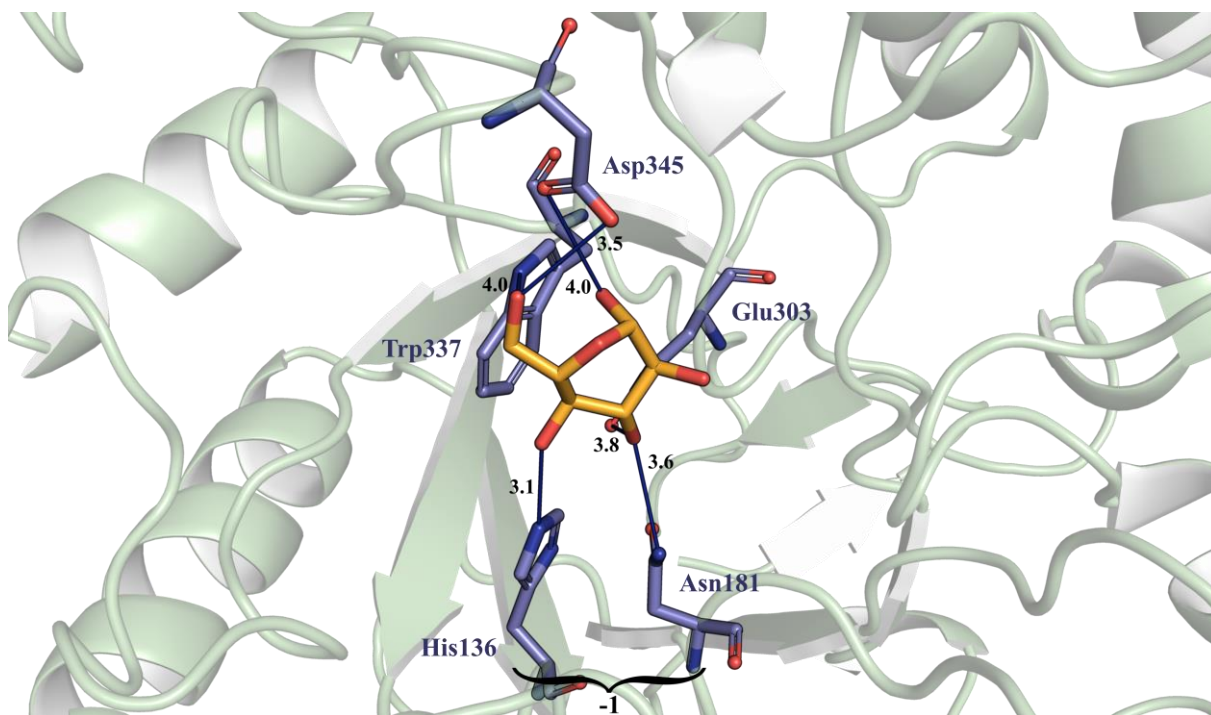


Fonte: Elaborada pelo autor.

Com exceção da  $\beta$ -celotetraose que não interagiu com nenhum dos glutamatos catalíticos, os demais ligantes estabeleceram interações com o nucleófilo Glu303, diferentemente do observado para XALc\_2967 e apenas  $\beta$ -celobiose interagiu com o glutamato ácido-base Glu182 (Tabela 12 e Tabela 13). Assim como em XALc\_2967, a maioria dos

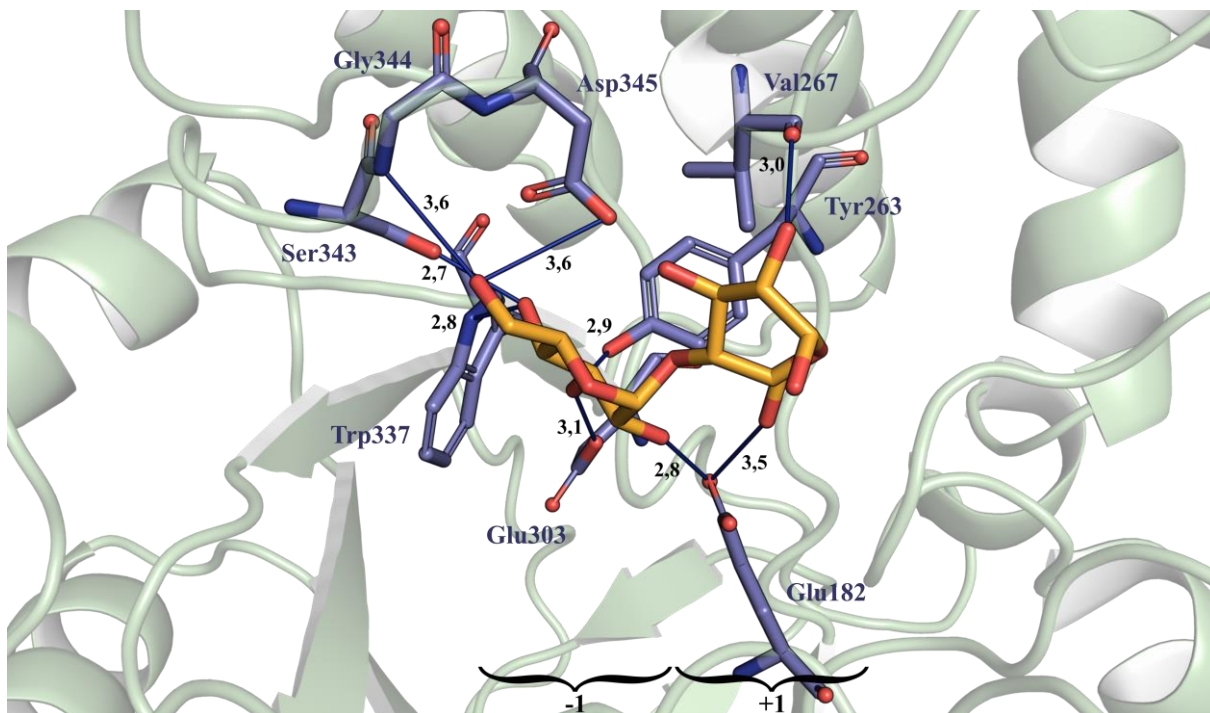
ligantes foram posicionados nos subsítios negativos até o positivo +1 (**Figura 49 - Figura 52**).  $\beta$ -celotriose foi o único ligante posicionado nos subsítios -1 a +2 (**Figura 51**), sugerindo que poderia ser hidrolisado em celobiose. No entanto, a sua interação ocorreu apenas com um glutamato catalítico, o Glu303 (**Tabela 13**).

**Figura 49** - Estrutura de XALc\_2969 complexada com  $\beta$ -D-glicose posicionada no subsítio -1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



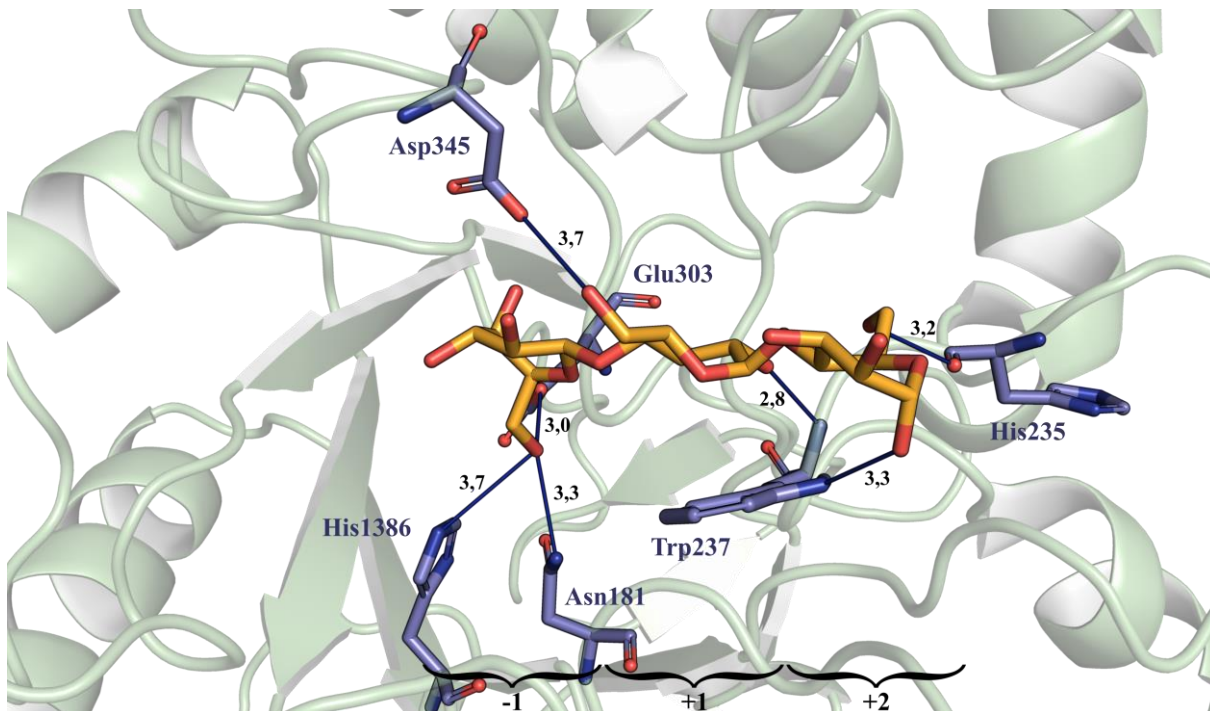
Fonte: Elaborada pelo autor.

**Figura 50** - Estrutura de XALc\_2969 complexada com  $\beta$ -celobiose posicionada nos subsítios -1 e +1. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.

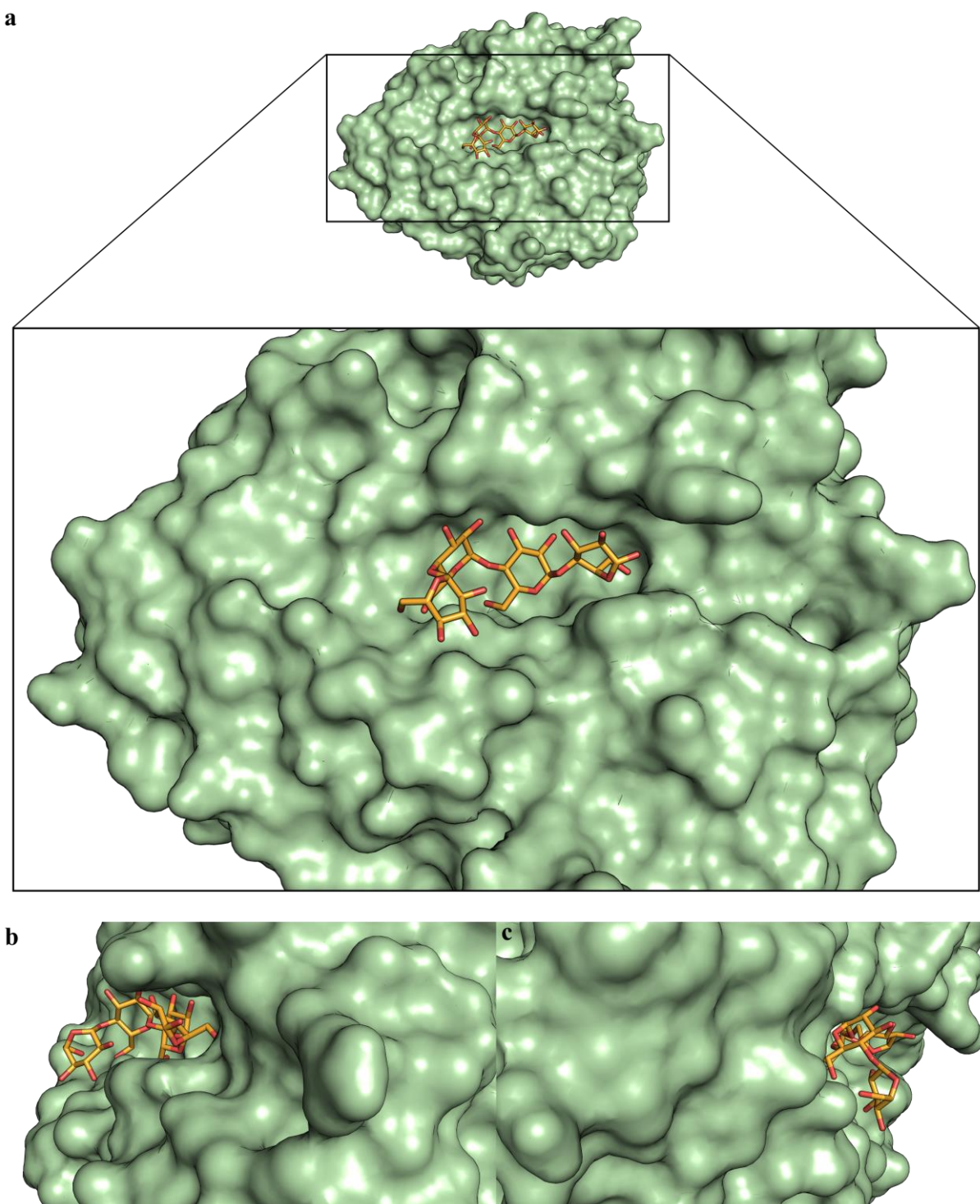
**Figura 51** - Estrutura de XALc\_2969 complexada com  $\beta$ -celotriose posicionada nos subsítios -1 a +2. As principais interações por ligações de hidrogênio (linha contínua azul-escura), entre, os resíduos de aminoácidos (roxo) e o ligante (laranja), são mostradas com as distâncias em angstroms (Å).



Fonte: Elaborada pelo autor.



**Figura 53** - Conformação e posição adotada por  $\beta$ -celotetraose no sítio ativo de XALc\_2969. **(a)** Visão ampliada do complexo, evidenciando que o final não-redutor de  $\beta$ -celotetraose está posicionado fora do sítio ativo de XALc\_2969; **(b)** Visão lateral direita do sítio ativo de XALc\_2969 complexado com  $\beta$ -celotetraose; **(c)** Visão lateral esquerda do sítio ativo de XALc\_2969 complexado com  $\beta$ -celotetraose.



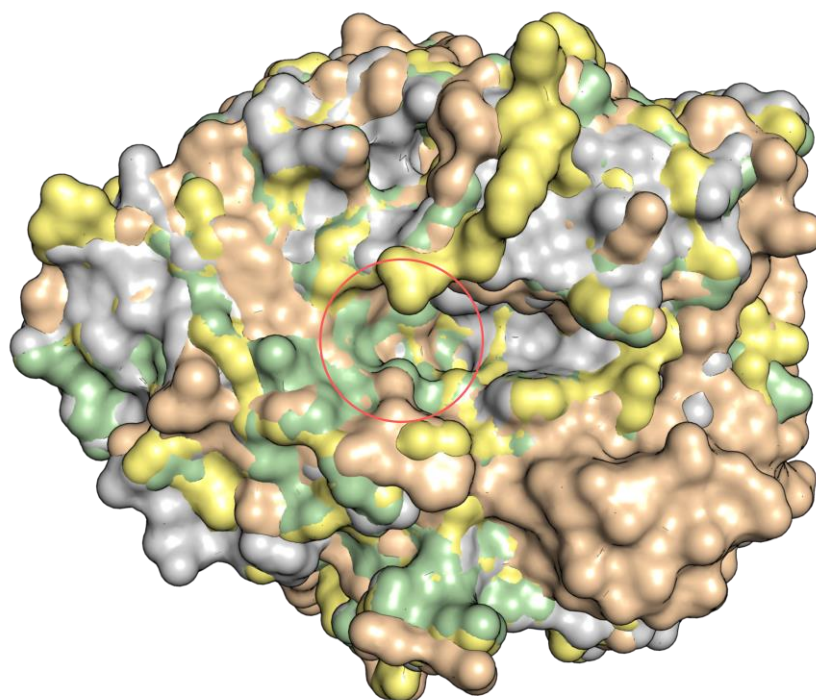
**Fonte:** Elaborada pelo autor.

Ao sobrepor estruturalmente os domínios catalíticos de XALc\_2969, XALc\_2967, XccCel5A de *X. campestris* pv. *campestris* (PDB: 4tuf) (PUHL et al., 2019) e EGPh de *Pyrococcus horikoshii* (PDB: 3qhn) (KIM; ISHIKAWA, 2011) é possível observar que o sítio

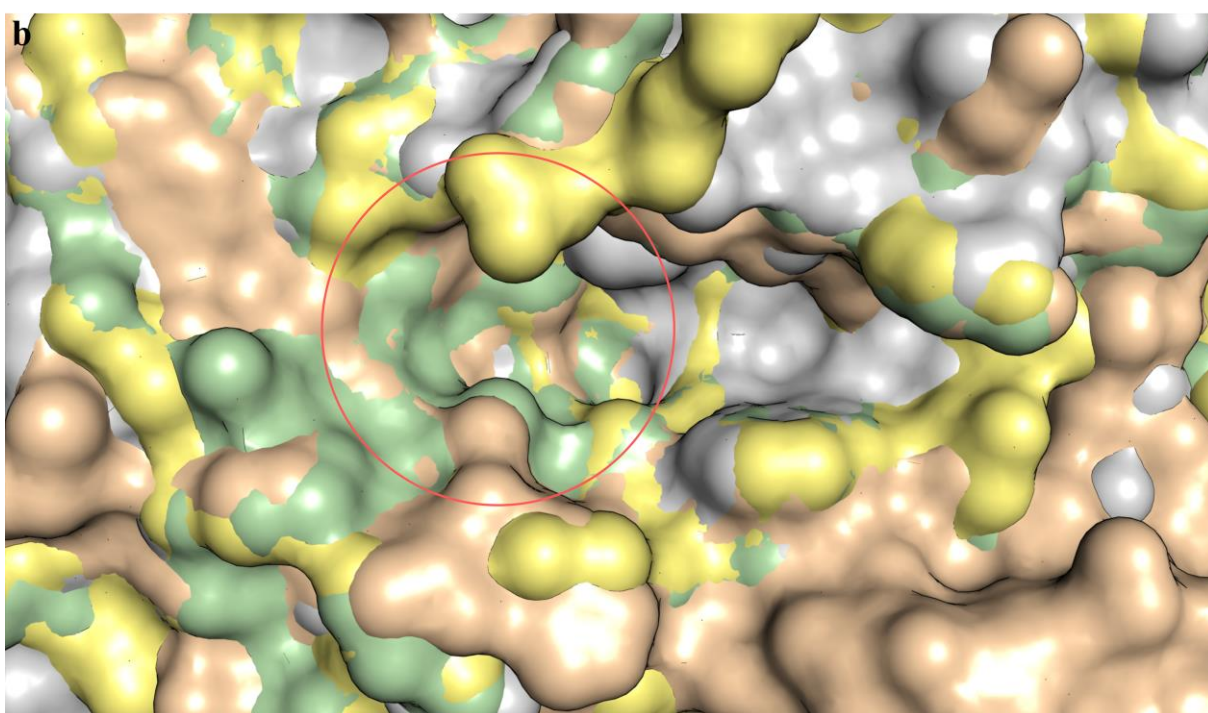
ativo de XALc\_2969 é menor na região do suposto subsítio -3, se destacando em relação as demais GH5\_1 (Figura 54).

**Figura 54** - (a) Sobreposição estrutural entre o domínio catalítico de XALc\_2969 (verde-claro), XALc\_2967 (amarelo-claro), XccCel5A (PDB: 4tuf) (cinza) e EGPh (PDB: 3qhn) (bege); (b) Ampliação da sobreposição estrutural evidenciando o sítio ativo sobreposto das quatro enzimas. O círculo vermelho delimita a região reduzida do sítio ativo de XALc\_2969.

**a**



**b**

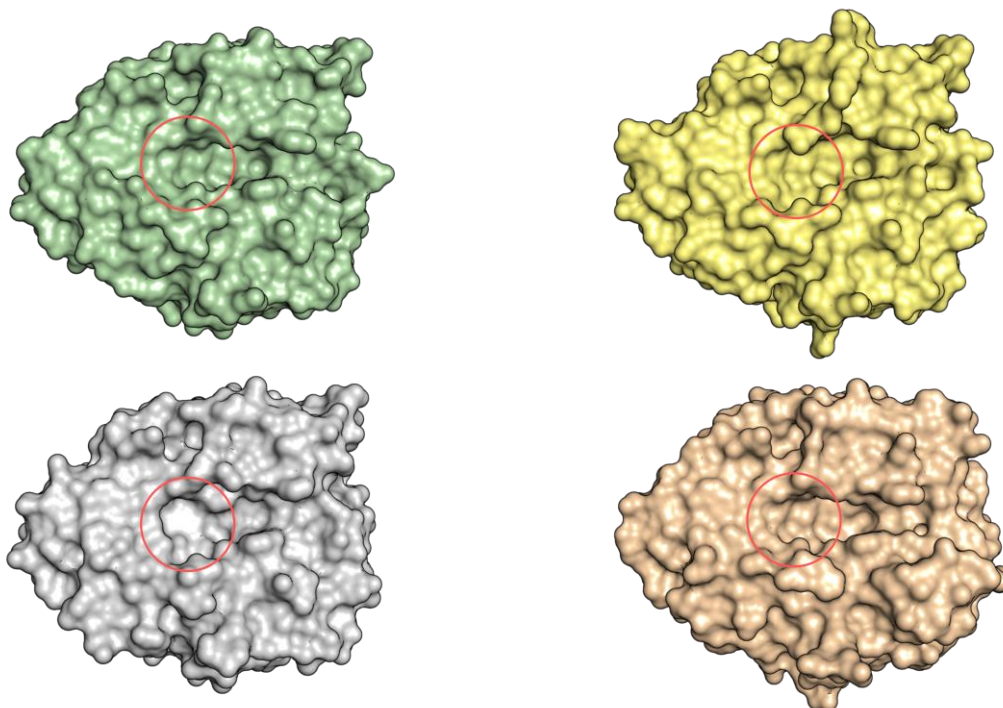


**Fonte:** Elaborada pelo autor.

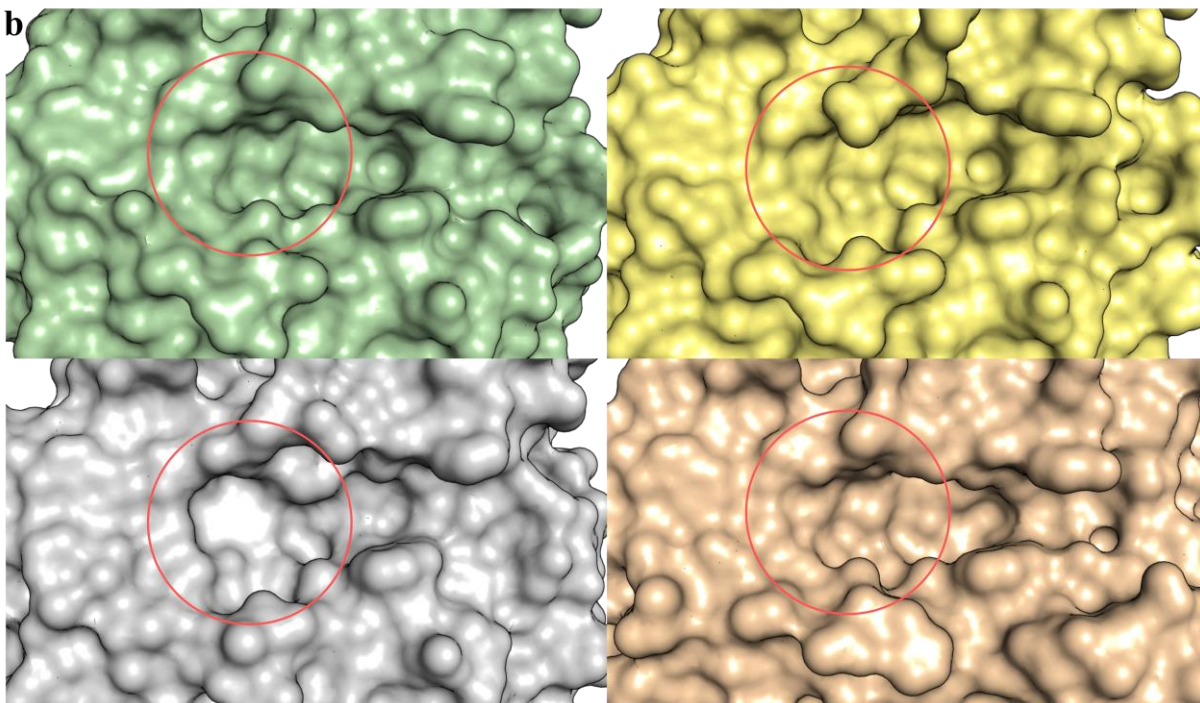
Visto que os subsítios -2 a +2 são a região principal para a ligação do substrato na maioria das GH5\_1, a presença de um subsítio -3 e subsequente não é consensual. Por exemplo, enzimas cristalinas de mutantes GH5\_1 (EGPh) de *P. horikoshii* (PDB: 3qhm, 3qhn e 3qho) exibem subsítios -3 e -4, favorecendo a ligação com substratos mais longos (KIM; ISHIKAWA, 2011). Por meio de simulações, foi demonstrado que a enzima XccCel5A de *X. campestris* pv. *campestris* (PDB: 4tuf) poderia interagir com substratos maiores em possíveis subsítios -3 e -4, mantendo a conformação ativa para a hidrólise (PUHL et al., 2019). Em ambos os casos, a extremidade do sítio ativo que acomodaria o subsítio -3 é desobstruída. Entretanto, em XALc\_2969, essa região da fenda catalítica é levemente mais inclinada, mantendo uma conformação mais côncava, impossibilitando o posicionamento eficiente de um ligante maior, o que poderia reduzir a possibilidade da existência de um subsítio -3 (**Figura 55 e Figura 56**). Algumas simulações de docking posicionaram  $\beta$ -celotetraose em subsítios -3 e -4, porém, além de continuarem não interagindo com os glutamatos catalíticos, as interações com os resíduos (**dados não mostrados**) presentes nesses subsítios não estão conforme a literatura (KIM; ISHIKAWA, 2011; PUHL et al., 2019; SAKON et al., 1996).

**Figura 55** - (a) Domínio catalítico de quatro endoglucanases GH5\_1: XALc\_2969 (verde-claro), XALc\_2967 (amarelo-claro), XccCel5A (PDB: 4tuf) (cinza) e EGPh (PDB: 3qhn) (bege); (b) Ampliação do domínio catalítico evidenciando o sítio ativo das quatro enzimas. O círculo vermelho delimita a região de interesse para comparação visual entre os sítios ativos.

**a**

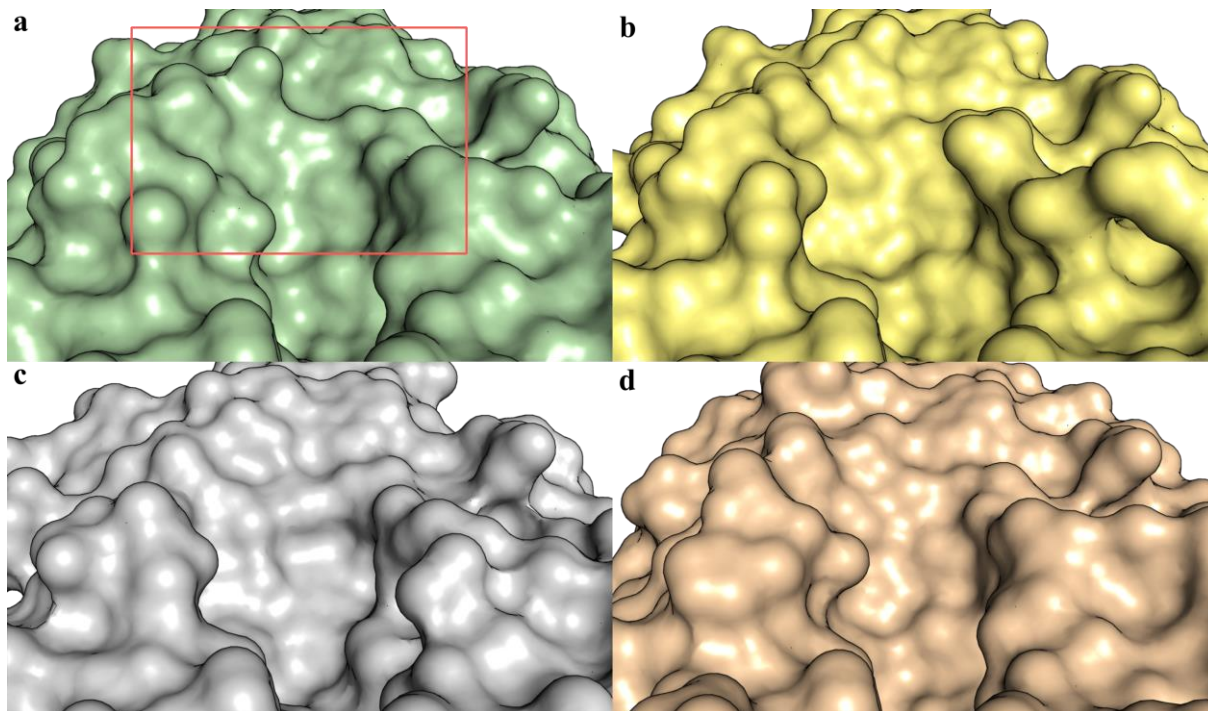


**b**



**Fonte:** Elaborada pelo autor.

**Figura 56** - Ampliação e visão frontal da região não-redutora (subsítio -3) do sítio ativo das quatro endoglucanases GH5\_1: **(a)** XALc\_2969 (verde-claro); **(b)** XALc\_2967 (amarelo-claro); **(c)** XccCel5A (PDB: 4tuf) (cinza); **(d)** EGPh (PDB: 3qhn) (algodão/bege). O círculo vermelho delimita a região de interesse para comparação visual entre os sítios ativos.



**Fonte:** Elaborada pelo autor.

Considerando que a conformação do sítio ativo de XALc\_2969 apresenta pequenas características que a diferenciam de outras GH5\_1 mais próximas, e que cadeias longas de polissacarídeos são geralmente projetadas para o solvente na posição do subsítio -3 (PUHL et al., 2019), conforme evidenciado pelos resultados de  $\beta$ -celotetraose (**Figura 52** e **Figura 53**), é possível inferir que XALc\_2969 não possui subsítio -3. Assim, os subsítios -2 a +2 seriam os únicos responsáveis pela ligação e hidrólise do substrato.

É importante enfatizar que a interpretação dos resultados de docking deve ser feita com cautela, visto que representam apenas um instantâneo das possíveis interações entre a enzima e o substrato. Com base nos resultados de XALc\_2969, não é possível inferir uma preferência específica por substrato e produto de hidrólise, dada a alternância na forma como os ligantes foram posicionados nos subsítios (**Figura 49** - **Figura 52**). Além disso, as simulações de docking não consideram a dinâmica entre a enzima e o substrato, que pode desempenhar um papel crucial na catálise, como evidenciado pela discrepância entre as interações com os glutamatos catalíticos e o posicionamento nos subsítios (**Tabela 13** e **Figura 49** - **Figura 52**).

#### 4.5.7 Análise Comparativa dos Resultados de Docking

Os resultados das simulações de docking sugerem que, teoricamente, as quatro enzimas analisadas têm potencial para atuar sobre as cadeias de celulose. Observou-se que ligantes maiores, que interagem com um número maior de subsítios e resíduos, apresentaram pontuações de afinidade superiores (**Tabela 10 - Tabela 13**). Isso era esperado, visto que geralmente as hidrolases glicosídicas (GHs) possuem mais afinidade com tetrassacarídeos ou substratos maiores (TAKEDA et al., 2022). Uma comparação entre as enzimas simuladas revelou que XALc\_0484 obteve as melhores pontuações de afinidade para os quatro ligantes comuns para as demais enzimas (**Tabela 14**).

**Tabela 14** - Pontuações de afinidade de ligação para cada ligante comum entre as quatro enzimas. As melhores pontuações estão em negrito.

	$\beta$ -D-glicose	$\beta$ -celobiose	$\beta$ -celotriose	$\beta$ -celotetraose
<b>XALc_0484</b>	<b>-5,84</b>	<b>-8,57</b>	<b>-10,19</b>	<b>-9,83</b>
XALc_0865	-4,83	-6,44	-7,13	-7,78
XALc_2967	-5,26	-7,27	-8,33	-8,74
XALc_2969	-5,53	-6,87	-7,49	-8,31

**Fonte:** Elaborada pelo autor.

Adicionalmente, além de XALc\_0484 demonstrar maior afinidade pelos carboidratos simulados, Pieretti et al. (2012) a consideram como essencial para a sobrevivência de *Xanthomonas albilineans* GPE PC73. Tal importância pode ser atribuída a características estruturais distintas e/ou otimizadas. Uma análise detalhada do domínio catalítico de XALc\_0484 e a sua comparação com estruturas cristalinas similares podem contribuir para verificar a existência de diferenças estruturais significativas em relação a outras celobiohidrolases da família GH6, conferindo a XALc\_0484 um caráter distintivo.

## 5 CONCLUSÃO

Dentre as 3.208 proteínas que compõem o proteoma de *Xanthomonas albilineans* GPE PC73, dez enzimas foram identificadas como CAZymes extracelulares dedicadas a degradar a parede celular do hospedeiro.

Em se tratando dos métodos de modelagem avaliados, os baseados em Inteligência Artificial (IA) apresentaram os melhores modelos para as enzimas integrais (domínio catalítico, linker e CBM2). Esses resultados refletem o impacto significativo que os algoritmos de predição baseados em IA ocasionam em várias áreas biológicas, principalmente na bioinformática estrutural, nos últimos anos.

As avaliações estruturais dos modelos gerados pelos três principais preditores, AlphaFold, ColabFold e RoseTTAFold, evidenciaram que apenas as pontuações de confiança (pLDDT e IDDT) não são suficientes como único critério para a seleção de modelos para uso subsequente. Essas pontuações refletem a confiança preditiva, mas não necessariamente a qualidade do modelo. Portanto, é essencial a utilização de outras ferramentas dedicadas à análise de divergências estereoquímicas e de contato, como QMEANDisCo, MolProbity e os gráficos de Ramachandran. Além disso, a conferência visual e a comparação estrutural com modelos experimentais de alta qualidade são fundamentais para identificar possíveis erros locais, ao invés de confiar exclusivamente nas pontuações globais fornecidas pelos preditores e/ou softwares de validação.

A análise comparativa por sobreposição estrutural entre os modelos preditos e estruturas experimentais permitiu adquirir confiabilidade nos modelos depositados no banco de estruturas do AlphaFold.

As simulações de docking forneceram informações importantes sobre a interação das enzimas com seus potenciais substratos, sendo essenciais no início do entendimento dos mecanismos enzimáticos. Entretanto, algumas interações com aminoácidos importantes nos processos catalíticos não foram observadas. De modo geral, esses resultados destacam a complexidade dessas interações e a necessidade de considerar outras abordagens, como estudos de dinâmica molecular, simulações QM/MM, análises bioquímicas e validação experimental, para complementar os resultados obtidos.

Com base nas pontuações de afinidade proteína-ligante obtidas para as quatro enzimas, é possível afirmar que XALc\_0484 apresenta potencial para futuros estudos e tem mais relevância para *Xanthomonas albilineans* em comparação com XALc\_0865, XALc\_2967 e XALc\_2969. Além dos resultados de docking, características estruturais de XALc\_0484, como

o fato de ser uma celobiohidrolase (CBH) – uma classe de enzimas altamente hidrolíticas com sítio ativo envolto por um túnel que dificultam o desacoplamento de uma cadeia de celulose antes de ser completamente hidrolisada – e possuir um domínio linker significativamente maior em relação a várias enzimas secretadas por outros representantes do gênero *Xanthomonas*, conferem a XALc\_0484 maior liberdade e área de busca para identificar cadeias de celulose, enquanto se mantêm fixada no substrato pelo CBM2.

Contudo, as ferramentas computacionais são indispensáveis para a contribuição da biologia estrutural.

## REFERÊNCIAS

- ≤ALI, A. Comparative protein modeling by satisfaction of spatial restraints. **Molecular Medicine Today**, v. 1, n. 6, p. 270–277, set. 1995.
- AGARWAL, V. et al. Pangenome insights into the diversification and disease specificity of worldwide *Xanthomonas* outbreaks. **Frontiers in Microbiology**, v. 14, n. July, p. 1–18, 5 jul. 2023.
- AKDEL, M. et al. A structural biology community assessment of AlphaFold2 applications. **Nature Structural & Molecular Biology**, v. 29, n. 11, p. 1056–1067, 7 nov. 2022.
- ALASDAIR, T. R.; JACKSON, L. M.; JACKSON, R. M. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. **Current Protein & Peptide Science**, v. 7, n. 5, p. 395–406, 1 out. 2006.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.
- ALVES, L. C. et al. Diel oscillations in cell wall components and soluble sugars as a response to short-day in sugarcane (*Saccharum* sp.). **BMC Plant Biology**, v. 19, n. 1, p. 215, 23 dez. 2019.
- ANNAMALAI, N.; RAJESWARI, M. V.; SIVAKUMAR, N. Cellobiohydrolases: role, mechanism, and recent developments. In: GUPTA, V. K. (Ed.). **Microbial Enzymes in Bioconversions of Biomass**. 1. ed. [s.l.] Springer, Cham, 2016. p. 29–35.
- ASPEBORG, H. et al. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). **BMC Evolutionary Biology**, v. 12, n. 1, p. 186, 20 dez. 2012.
- BAEK, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. **Science**, v. 373, n. 6557, p. 871–876, 20 ago. 2021.
- BAEK, M.; BAKER, D. Deep learning and protein structure modeling. **Nature Methods**, v. 19, n. 1, p. 13–14, 11 jan. 2022.
- BAKER, J. O. et al. Catalytically enhanced endocellulase Cel5A from *Acidothermus cellulolyticus*. **Applied Biochemistry and Biotechnology**, v. 121, n. 1–3, p. 0129–0148, 2005.
- BALLANTE, F. (ED.). **Protein-ligand interactions and drug design**. 1. ed. New York, NY: Springer US, 2021. v. 2266
- BENDER, B. J. et al. A practical guide to large-scale docking. **Nature Protocols**, v. 16, n. 10, p. 4799–4832, 24 out. 2021.
- BENKERT, P.; TOSATTO, S. C. E.; SCHOMBURG, D. QMEAN: a comprehensive scoring function for model quality assessment. **Proteins: Structure, Function, and Bioinformatics**, v. 71, n. 1, p. 261–277, 11 abr. 2008.
- BERNARDES, A. et al. Carbohydrate binding modules enhance cellulose enzymatic

hydrolysis by increasing access of cellulases to the substrate. **Carbohydrate Polymers**, v. 211, p. 57–68, 1 maio 2019.

BIELY, P. Microbial carbohydrate esterases deacetylating plant polysaccharides. **Biotechnology Advances**, v. 30, n. 6, p. 1575–1588, 1 nov. 2012.

BINI, A. P. et al. Transmission of *Xanthomonas albilineans* by the spittlebug, *Mahanarva fimbriolata* (Hemiptera: Cercopidae), in Brazil: first report of an insect vector for the causal agent of sugarcane leaf scald. **Journal of Insect Science**, v. 23, n. 6, p. 28–29, 1 nov. 2023.

BIRCH, R. G. *Xanthomonas albilineans* and the antipathogenesis approach to disease control. **Molecular Plant Pathology**, v. 2, n. 1, p. 1–11, 7 jan. 2001.

BIRCH, R. G.; PATIL, S. S. Correlation between albicidin production and chlorosis induction by *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. **Physiological and Molecular Plant Pathology**, v. 30, n. 2, p. 199–206, mar. 1987.

BORASTON, A. B. et al. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. **Biochemical Journal**, v. 382, n. 3, p. 769–781, 15 set. 2004.

BORDIN, N. et al. Novel machine learning approaches revolutionize protein knowledge. **Trends in Biochemical Sciences**, v. 48, n. 4, p. 345–359, abr. 2023.

BOUATTA, N.; SORGER, P.; ALQURAISHI, M. Protein structure prediction by AlphaFold2: are attention and symmetries all you need? **Acta Crystallographica Section D: Structural Biology**, v. 77, n. 8, p. 982–991, 1 ago. 2021.

BUCHFINK, B.; REUTER, K.; DROST, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. **Nature Methods**, v. 18, n. 4, p. 366–368, 7 abr. 2021.

BURLEY, S. K. et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. **Nucleic Acids Research**, v. 51, n. D1, p. D488–D508, 6 jan. 2023.

CAMPBELL, J. A. et al. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. **Biochemical Journal**, v. 326, n. 3, p. 929–939, 15 set. 1997.

CANTAREL, B. I. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. **Nucleic Acids Research**, v. 37, n. SUPPL. 1, p. D233–D238, 1 jan. 2009.

CHAMPOISEAU, P. et al. High variation in pathogenicity of genetically closely related strains of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen, in Guadeloupe. **Phytopathology**, v. 96, n. 10, p. 1081–1091, out. 2006.

CHATTERJEE, S.; ALMEIDA, R. P. P.; LINDOW, S. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. **Annual Review of Phytopathology**, v. 46, n. 1, p. 243–271, set. 2008.

CHEN, V. B. et al. MolProbity : all-atom structure validation for macromolecular

crystallography. **Acta Crystallographica Section D Biological Crystallography**, v. 66, n. 1, p. 12–21, 1 jan. 2010.

CHETTRI, D.; VERMA, A. K.; VERMA, A. K. Innovations in CAZyme gene diversity and its modification for biorefinery applications. **Biotechnology Reports**, v. 28, p. e00525, 1 dez. 2020.

CHUNDAWAT, S. P. S. et al. Deconstruction of lignocellulosic biomass to fuels and chemicals. **Annual Review of Chemical and Biomolecular Engineering**, v. 2, n. 1, p. 121–145, 15 jul. 2011.

COPLEY, R. R.; BORK, P. Homology among ( $\beta\alpha$ )<sub>8</sub> barrels: implications for the evolution of metabolic pathways. **Journal of Molecular Biology**, v. 303, n. 4, p. 627–641, nov. 2000.

CORNISH-BOWDEN, A. Current IUBMB recommendations on enzyme nomenclature and kinetics. **Perspectives in Science**, v. 1, n. 1–6, p. 74–87, maio 2014.

COURTADE, G. et al. The carbohydrate-binding module and linker of a modular lytic polysaccharide monooxygenase promote localized cellulose oxidation. **Journal of Biological Chemistry**, v. 293, n. 34, p. 13006–13015, 24 ago. 2018.

CRENNELL, S. J. et al. Dimerisation and an increase in active site aromatic groups as adaptations to high temperatures: X-ray solution scattering and substrate-bound crystal structures of *Rhodothermus marinus* endoglucanase Cel12A. **Journal of Molecular Biology**, v. 356, n. 1, p. 57–71, 10 fev. 2006.

CRENNELL, S. J.; HREGGVIDSSON, G. O.; NORDBERG KARLSSON, E. The structure of *Rhodothermus marinus* Cel12A, a highly thermostable family 12 endoglucanase, at 1.8 Å resolution. **Journal of Molecular Biology**, v. 320, n. 4, p. 883–897, jul. 2002.

CSIZMADIA, P. **MarvinSketch and MarvinView: molecule applets for the World Wide Web**. Proceedings of The 3rd International Electronic Conference on Synthetic Organic Chemistry. **Anais...Basel, Switzerland: MDPI**, 1 nov. 1999. Disponível em: <<http://sciforum.net/conference/ecsoc-3/paper/1775>>

DAVIES, G.; HENRISSAT, B. Structures and mechanisms of glycosyl hydrolases. **Structure**, v. 3, n. 9, p. 853–859, set. 1995.

DAVIES, G. J.; WILSON, K. S.; HENRISSAT, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. **Biochemical Journal**, v. 321, n. 2, p. 557–559, 15 jan. 1997.

DE CARVALHO, L. M. et al. Bioinformatics applied to biotechnology: a review towards bioenergy research. **Biomass and Bioenergy**, v. 123, p. 195–224, 1 abr. 2019.

DELANO, W. L. The PyMOL molecular graphics system. v. 40, n. 1, p. 82–92, 2002.

DOS SANTOS, R. N.; FERREIRA, L. G.; ANDRICOPULO, A. D. Practices in molecular docking and structure-based virtual screening. In: [s.l.: s.n.]. v. 1762p. 31–50.

DRULA, E. et al. The carbohydrate-active enzyme database: functions and literature. **Nucleic Acids Research**, v. 50, n. D1, p. D571–D577, 7 jan. 2022.

- ENGH, R. A.; HUBER, R. Accurate bond and angle parameters for X-ray protein structure refinement. **Acta Crystallographica Section A Foundations of Crystallography**, v. 47, n. 4, p. 392–400, 1 jul. 1991.
- ENGH, R. A.; HUBER, R. Structure quality and target parameters. In: **International Tables for Crystallography**. Chester, England: International Union of Crystallography, 2006. p. 382–392.
- FANELLI, A. et al. Biomass composition of two new energy cane cultivars compared with their ancestral *Saccharum spontaneum* during internode development. **Biomass and Bioenergy**, v. 141, p. 105696, 1 out. 2020.
- FATIMA, U.; SENTHIL-KUMAR, M. Plant and pathogen nutrient acquisition strategies. **Frontiers in Plant Science**, v. 6, n. September, p. 750, 17 set. 2015.
- FIGUEIREDO, R. et al. Suberin and hemicellulose in sugarcane cell wall architecture and crop digestibility: a biotechnological perspective. **Food and Energy Security**, v. 8, n. 3, p. e00163, 13 jul. 2019.
- FINKING, R.; MARAHIEL, M. A. Biosynthesis of nonribosomal peptides. **Annual Review of Microbiology**, v. 58, n. 1, p. 453–488, 1 out. 2004.
- FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research**, v. 39, n. suppl, p. W29–W37, 1 jul. 2011.
- GARDNER, K.; BLACKWELL, J. The hydrogen bonding in native cellulose. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 343, n. 1, p. 232–237, 20 mar. 1974.
- GILBERT, H. J.; KNOX, J. P.; BORASTON, A. B. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. **Current Opinion in Structural Biology**, v. 23, n. 5, p. 669–677, out. 2013.
- GLAZEBROOK, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. **Annual review of phytopathology**, v. 43, p. 205–27, jan. 2005.
- GUEX, N.; PEITSCH, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. **ELECTROPHORESIS**, v. 18, n. 15, p. 2714–2723, 14 jan. 1997.
- GUO, H.-B. et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. **Scientific Reports**, v. 12, n. 1, p. 10696, 23 dez. 2022.
- GUPTA, G. et al. Secretory molecules from secretion systems fine-tune the host-beneficial bacteria (PGPRs) interaction. **Frontiers in Microbiology**, v. 15, p. 1355750, 26 fev. 2024.
- GUPTA, M. et al. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. **bioRxiv**, p. 2021.05.10.443524, 12 maio 2021.
- HADDAD, Y.; ADAM, V.; HEGER, Z. Ten quick tips for homology modeling of high-resolution protein 3D structures. **PLOS Computational Biology**, v. 16, n. 4, p. e1007449, 2 abr. 2020.

HAMEDUH, T. et al. Homology modeling in the time of collective and artificial intelligence. **Computational and Structural Biotechnology Journal**, v. 18, p. 3494–3506, 2020.

HAPPS, R. M. et al. Economic and sustainability impacts of yield and composition variation in bioenergy crops: switchgrass (*Panicum virgatum* L.). **ACS Sustainable Chemistry & Engineering**, v. 12, n. 5, p. 1897–1910, 5 fev. 2024.

HAUBEN, L. et al. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. **International Journal of Systematic Bacteriology**, v. 47, n. 2, p. 328–335, 1 abr. 1997.

HENRISSAT, B. et al. Cellulase families revealed by hydrophobic cluster analysis. **Gene**, v. 81, n. 1, p. 83–95, set. 1989.

HENRISSAT, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemical Journal**, v. 280, n. 2, p. 309–316, 1 dez. 1991.

HENRISSAT, B. et al. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. **Proceedings of the National Academy of Sciences**, v. 92, n. 15, p. 7090–7094, 18 jul. 1995.

HON, D. N. S. Cellulose: a random walk along its historical path. **Cellulose**, v. 1, n. 1, p. 1–25, mar. 1994.

HOU, Q. et al. Ten quick tips for sequence-based prediction of protein properties using machine learning. **PLOS Computational Biology**, v. 18, n. 12, p. e1010669, 1 dez. 2022.

HU, W.; OHUE, M. SpatialPPI: three-dimensional space protein-protein interaction prediction with AlphaFold Multimer. **Computational and Structural Biotechnology Journal**, v. 23, p. 1214–1225, dez. 2024.

ILLERGÅRD, K.; ARDELL, D. H.; ELOFSSON, A. Structure is three to ten times more conserved than sequence — A study of structural response in protein cores. **Proteins: Structure, Function, and Bioinformatics**, v. 77, n. 3, p. 499–508, 15 nov. 2009.

JACKSON, R. W. et al. Bacterial pathogen evolution: breaking news. **Trends in Genetics**, v. 27, n. 1, p. 32–40, 1 jan. 2011.

JACOBS, J. M. et al. The in planta transcriptome of *Ralstonia solanacearum*: conserved physiological and virulence strategies during bacterial wilt of tomato. **mBio**, v. 3, n. 4, 31 ago. 2012.

JAKUBEC, D. et al. PrankWeb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures. **Nucleic Acids Research**, v. 50, n. W1, p. W593–W597, 5 jul. 2022.

JENKINS, J. et al.  $\beta$ -Glucosidase,  $\beta$ -galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold  $\beta/\alpha$  architecture and with two conserved glutamates near the carboxy-terminal ends of  $\beta$ -strands four and seven. **FEBS Letters**, v. 362, n. 3, p. 281–285, 10 abr. 1995.

JOHNSON, T. L. et al. Type II secretion: from structure to function. **FEMS Microbiology**

**Letters**, v. 255, n. 2, p. 175–186, 1 fev. 2006.

JUERS, D. H.; HUBER, R. E.; MATTHEWS, B. W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between  $\beta$ -galactosidase and other glycohydrolases. **Protein Science**, v. 8, n. 1, p. 122–136, 31 jan. 1999.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 26 ago. 2021.

JUMPER, J.; HASSABIS, D. Protein structure predictions to atomic accuracy with AlphaFold. **Nature Methods**, v. 19, n. 1, p. 11–12, 11 jan. 2022.

KELZ, J. I. et al. Active learning module for protein structure analysis using novel enzymes. **The Biophysicist**, v. 3, n. 1, p. 49–63, 1 jul. 2022.

KIM, H. W.; ISHIKAWA, K. Functional analysis of hyperthermophilic endocellulase from *Pyrococcus horikoshii* by crystallographic snapshots. **Biochemical Journal**, v. 437, n. 2, p. 223–230, 15 jul. 2011.

KIM, S. et al. PubChem 2023 update. **Nucleic Acids Research**, v. 51, n. D1, p. D1373–D1380, 6 jan. 2023.

KIRCHMAIR, J. et al. Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics, and mechanisms. **Journal of Chemical Information and Modeling**, v. 52, n. 3, p. 617–648, 26 mar. 2012.

KOIVULA, A. et al. The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. **Journal of the American Chemical Society**, v. 124, n. 34, p. 10015–10024, 1 ago. 2002.

KONC, J.; JANEŽIČ, D. Binding site comparison for function prediction and pharmaceutical discovery. **Current Opinion in Structural Biology**, v. 25, p. 34–39, abr. 2014.

KRIVÁK, R.; HOKSZA, D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. **Journal of Cheminformatics**, v. 10, n. 1, p. 39, 14 dez. 2018.

KUMAR, D.; MURTHY, G. S. Stochastic molecular model of enzymatic hydrolysis of cellulose for ethanol production. **Biotechnology for Biofuels**, v. 6, n. 1, p. 1–20, 2 maio 2013.

KUNKEL, B. N.; CHEN, Z. Virulence strategies of plant pathogenic bacteria. In: DWORKIN, M. et al. (Eds.). **The Prokaryotes**. New York, NY: Springer New York, 2006. p. 421–440.

KURKCUOGLU, Z. et al. Coupling between catalytic loop motions and enzyme global dynamics. **PLOS Computational Biology**, v. 8, n. 9, p. e1002705, 27 set. 2012.

LALUK, K.; MENGISTE, T. Necrotroph attacks on plants: wanton destruction or covert extortion? **The Arabidopsis Book**, v. 8, p. e0136, jan. 2010.

LAMPUGNANI, E. R. et al. Building a plant cell wall at a glance. **Journal of Cell Science**, v. 131, n. 2, 15 jan. 2018.

- LASKOWSKI, R. A. et al. PDBsum: a web-based database of summaries and analyses of all PDB structures. **Trends in Biochemical Sciences**, v. 22, n. 12, p. 488–490, 1 dez. 1997.
- LEAVER-FAY, A. et al. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. In: JOHNSON, M. L.; BRAND, L. (Eds.). **Methods in Enzymology**. [s.l.: s.n.]. v. 487p. 545–574.
- LETUNIC, I.; KHEDKAR, S.; BORK, P. SMART: recent updates, new developments and status in 2020. **Nucleic Acids Research**, v. 49, n. D1, p. D458–D460, 8 jan. 2021.
- LEVASSEUR, A. et al. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. **Biotechnology for Biofuels**, v. 6, n. 1, p. 41, 2013.
- LEYNS, F. et al. The host range of the genus *Xanthomonas*. **The Botanical Review**, v. 50, n. 3, p. 308–356, jul. 1984.
- LI, M. L. et al. Comparative genome analysis unravels pathogenicity of *Xanthomonas albilineans* causing sugarcane leaf scald disease. **BMC Genomics**, v. 23, n. 1, p. 1–15, 1 dez. 2022.
- LINDEBERG, M. Genome-enabled perspectives on the composition, evolution, and expression of virulence determinants in bacterial plant pathogens. **Annual Review of Phytopathology**, v. 50, n. 1, p. 111–132, 8 set. 2012.
- LOMBARD, V. et al. A hierarchical classification of polysaccharide lyases for glyco-genomics. **Biochemical Journal**, v. 432, n. 3, p. 437–444, 15 dez. 2010.
- LOMBARD, V. et al. The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Research**, v. 42, n. D1, p. D490–D495, jan. 2014.
- MAHLICH, Y. et al. HFSP: high speed homology-driven function annotation of proteins. **Bioinformatics**, v. 34, n. 13, p. i304–i312, 1 jul. 2018.
- MAITAN-ALFENAS, G. P.; VISSER, E. M.; GUIMARÃES, V. M. Enzymatic hydrolysis of lignocellulosic biomass: converting food waste in valuable products. **Current Opinion in Food Science**, v. 1, n. 1, p. 44–49, fev. 2015.
- MARIANI, V. et al. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. **Bioinformatics**, v. 29, n. 21, p. 2722–2728, 1 nov. 2013.
- MARTINS, M. P. et al. Glycoside hydrolase subfamily GH5\_57 features a highly redesigned catalytic interface to process complex hetero- $\beta$ -mannans. **Acta Crystallographica Section D: Structural Biology**, v. 78, n. 11, p. 1358–1372, 1 nov. 2022.
- MARX, V. Method of the year: protein structure prediction. **Nature Methods**, v. 19, n. 1, p. 5–10, 11 jan. 2022.
- MATHIEU, S. et al. Functional exploration of the polysaccharide lyase family PL6. **PLOS ONE**, v. 11, n. 7, p. e0159415, 20 jul. 2016.
- MCCARTER, J. D.; STEPHEN WITHERS, G. Mechanisms of enzymatic glycoside

hydrolysis. **Current Opinion in Structural Biology**, v. 4, n. 6, p. 885–892, 1 jan. 1994.

MCNAMARA, J. T.; MORGAN, J. L. W.; ZIMMER, J. A molecular description of cellulose biosynthesis. **Annual Review of Biochemistry**, v. 84, n. 1, p. 895–921, 2 jun. 2015.

MEENA, M. R. et al. Biomass potential of novel interspecific and intergeneric hybrids of *Saccharum* grown in sub-tropical climates. **Scientific Reports**, v. 10, n. 1, p. 21560, 9 dez. 2020.

MENSI, I. et al. Breaking dogmas: The plant vascular pathogen *Xanthomonas albilineans* is able to invade non-vascular tissues despite its reduced genome. **Open Biology**, v. 4, n. 2, p. 130116–130116, 12 fev. 2014.

MIRANDA, R. P. et al. Genome organization of four brazilian *Xanthomonas albilineans* strains does not correlate with aggressiveness. **Microbiology Spectrum**, v. 11, n. 3, 15 jun. 2023.

MIRDITA, M. et al. ColabFold: making protein folding accessible to all. **Nature Methods**, v. 19, n. 6, p. 679–682, 30 jun. 2022.

MOORE, P. H.; MARETZKI, A. Sugarcane. In: ZAMSKI, E.; SCHAFFER, A. A. (Eds.). **Photoassimilate distribution in plants and crops**. 1. ed. [s.l.] Marcel Dekker, 1996. p. 643–670.

MORRIS, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. **Journal of Computational Chemistry**, v. 30, n. 16, p. 2785–2791, 27 dez. 2009.

NAKAMURA, A. et al. Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase *Trichoderma reesei* Cel6A and its domains on crystalline cellulose. **Journal of Biological Chemistry**, v. 291, n. 43, p. 22404–22413, out. 2016.

NAKAMURA, A. et al. Domain architecture divergence leads to functional divergence in binding and catalytic domains of bacterial and fungal cellobiohydrolases. **Journal of Biological Chemistry**, v. 295, n. 43, p. 14606–14617, 23 out. 2020.

PAYNE, C. M. et al. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. **Proceedings of the National Academy of Sciences**, v. 110, n. 36, p. 14646–14651, 3 set. 2013.

PAYSAN-LAFOSSE, T. et al. InterPro in 2022. **Nucleic Acids Research**, v. 51, n. D1, p. D418–D427, 6 jan. 2023.

PEREIRA, J. et al. High-accuracy protein structure prediction in CASP14. **Proteins**, v. 89, n. 12, p. 1687–1699, 14 dez. 2021.

PETTERSEN, E. F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 1 out. 2004.

PIERETTI, I. et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae.

**BMC Genomics**, v. 10, n. 1, p. 616, jan. 2009.

PIERETTI, I. et al. Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. **BMC Genomics**, v. 13, n. 1, p. 658, jan. 2012.

PIERETTI, I. et al. What makes *Xanthomonas albilineans* unique amongst xanthomonads? **Frontiers in Plant Science**, v. 6, n. APR, p. 289, 24 abr. 2015.

POLIZELI, M. L. T. M. et al. Enzyme system from *Aspergillus* in current industrial uses and future applications in the production of second-generation ethanol. In: **New and Future Developments in Microbial Biotechnology and Bioengineering**. [s.l.] Elsevier, 2016. p. 127–140.

PRATES, É. T. et al. X-ray structure and molecular dynamics simulations of endoglucanase 3 from *Trichoderma harzianum*: structural organization and substrate recognition by endoglucanases that lack cellulose binding module. **PLoS ONE**, v. 8, n. 3, p. e59069, 14 mar. 2013.

PUHL, A. C. et al. Crystallographic structure and molecular dynamics simulations of the major endoglucanase from *Xanthomonas campestris* pv. *campestris* shed light on its oligosaccharide products release pattern. **International Journal of Biological Macromolecules**, v. 136, p. 493–502, 1 set. 2019.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. **Journal of Molecular Biology**, v. 7, n. 1, p. 95–99, 1 jul. 1963.

RAMACHANDRAN, G. N.; SASISEKHARAN, V. Conformation of polypeptides and proteins. **Advances in Protein Chemistry**, v. 23, n. C, p. 283–437, 1 jan. 1968.

RAMAN, S. et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. **Proteins**, v. 77, n. S9, p. 89–99, 1 jan. 2009.

REVERBEL-LEROY, C. et al. Crystallization of the catalytic domain of *Clostridium cellulolyticum* CelF cellulase in the presence of a newly synthesized cellulase inhibitor. **Acta Crystallographica Section D: Biological Crystallography**, v. 54, n. 1, p. 114–118, 1 jan. 1998.

RICHARDSON, J. S. The Anatomy and taxonomy of protein structure. In: ANFINSEN, C. B.; EDSALL, J. T.; RICHARDS, F. M. (Eds.). **Advances in Protein Chemistry**. [s.l.] Academic Press, 1981. v. 34p. 167–339.

RITCHIE, H.; PABLO, R. **Energy mix**. Disponível em: <<https://ourworldindata.org/energy-mix>>. Acesso em: 9 set. 2023.

RITCHIE, H.; ROSADO, P.; ROSER, M. **Energy**. Disponível em: <<https://ourworldindata.org/energy>>. Acesso em: 9 set. 2023.

ROBIN, X. et al. Continuous Automated Model EvaluatiOn (CAMEO) - Perspectives on the future of fully automated evaluation of structure prediction methods. **Proteins**, v. 89, n. 12, p. 1977–1986, 19 dez. 2021.

- ROSNOW, J. J. et al. Profiling microbial lignocellulose degradation and utilization by emergent omics technologies. **Critical Reviews in Biotechnology**, v. 37, n. 5, p. 626–640, 4 jul. 2016.
- ROTT, P. et al. **A guide to sugarcane diseases**. Montpellier: CIRAD Publication Service, 2000.
- RUBIN, E. M. Genomics of cellulosic biofuels. **Nature**, v. 454, n. 7206, p. 841–845, ago. 2008.
- RYAN, R. P. et al. Pathogenomics of *Xanthomonas*: Understanding bacterium-plant interactions. **Nature reviews. Microbiology**, v. 9, n. 5, p. 344–55, maio 2011.
- SABINI, E. et al. The three-dimensional structure of a *Trichoderma reesei*  $\beta$ -mannanase from glycoside hydrolase family 5. **Acta Crystallographica Section D: Biological Crystallography**, v. 56, n. 1, p. 3–13, 1 jan. 2000.
- SADDLER, G. S.; BRADBURY, J. F. *Xanthomonadales* ord. nov. In: **Bergey's Manual of Systematic Bacteriology**. Boston, MA: Springer US, 2005. p. 63–122.
- SAKON, J. et al. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. **Biochemistry**, v. 35, n. 33, p. 10648–10660, 1 jan. 1996.
- SALENTIN, S. et al. PLIP: fully automated protein-ligand interaction profiler. **Nucleic Acids Research**, v. 43, n. W1, p. W443–W447, 1 jul. 2015.
- SAMAEI-DARYAN, S.; GOLIAEI, B.; EBRAHIM-HABIBI, A. Characterization of surface binding sites in glycoside hydrolases: a computational study. **Journal of Molecular Recognition**, v. 30, n. 9, p. e2624, set. 2017.
- SANDGREN, M. et al. Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. **Protein Science**, v. 12, n. 4, p. 848–860, 1 abr. 2003.
- SANDGREN, M. et al. Crystal complex structures reveal how substrate is bound in the -4 to the +2 binding sites of *Humicola grisea* Cel12A. **Journal of Molecular Biology**, v. 342, n. 5, p. 1505–1517, 1 out. 2004.
- SANDGREN, M. et al. The structure of a bacterial cellobiohydrolase: the catalytic core of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B. **Journal of Molecular Biology**, v. 425, n. 3, p. 622–635, fev. 2013.
- SANNER, M. F. Python: a programming language for software integration and development. **Journal of Molecular Graphics and Modelling**, v. 17, n. 1, p. 57–61, fev. 1999.
- SANTOS, C. R. DOS et al. Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- $\beta$ -d-mannanase. **Journal of Structural Biology**, v. 177, n. 2, p. 469–476, fev. 2012.
- SCARDINO, V.; DI FILIPPO, J. I.; CAVASOTTO, C. N. How good are AlphaFold models for docking-based virtual screening? **iScience**, v. 26, n. 1, p. 105920, 20 jan. 2023.

SCHELER, G. **Computational methods in protein evolution**. New York, NY: Springer New York, 2019. v. 1851

SILVA, P. P. DA et al. RB961552 - Highly productive sugarcane cultivar with fast inter-row closure. **Crop Breeding and Applied Biotechnology**, v. 21, n. 2, p. e32992126, 13 set. 2021.

SONG, Y. et al. High-resolution comparative modeling with RosettaCM. **Structure**, v. 21, n. 10, p. 1735–1742, 8 out. 2013.

STROBEL, H. J.; CALDWELL, F. C.; DAWSON, K. A. Carbohydrate transport by the anaerobic thermophile *Clostridium thermocellum* LQRI. **Applied and Environmental Microbiology**, v. 61, n. 11, p. 4012–4015, nov. 1995.

STUDER, G. et al. QMEANDisCo-distance constraints applied on model quality estimation. **Bioinformatics**, v. 36, n. 6, p. 1765–1771, 1 mar. 2020.

SULZENBACHER, G. et al. The crystal structure of a 2-fluorocellootriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 Å resolution. **Biochemistry**, v. 38, n. 15, p. 4826–4833, 1 abr. 1999.

TAKEDA, M. et al. Metagenomic mining and structure-function studies of a hyperthermostable cellobiohydrolase from hot spring sediment. **Communications Biology**, v. 5, n. 1, p. 247, 22 mar. 2022.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. **Molecular Biology and Evolution**, v. 38, n. 7, p. 3022–3027, 25 jun. 2021.

TAO, X. et al. Recent developments in molecular docking technology applied in food science: a review. **International Journal of Food Science & Technology**, v. 55, n. 1, p. 33–45, 6 jan. 2020.

TAYI, L. et al. Action of multiple cell wall-degrading enzymes is required for elicitation of innate immune responses during *Xanthomonas oryzae* pv. *oryzae* infection in rice. **Molecular Plant-Microbe Interactions**, v. 29, n. 8, p. 599–608, 19 ago. 2016.

TAYI, L. et al. A mutation in an exoglucanase of *Xanthomonas oryzae* pv. *oryzae*, which confers an endo mode of activity, affects bacterial virulence, but not the induction of immune responses, in rice. **Molecular Plant Pathology**, v. 19, n. 6, p. 1364–1376, 26 jun. 2018.

TEJERA, N. A. et al. Comparative analysis of physiological characteristics and yield components in sugarcane cultivars. **Field Crops Research**, v. 102, n. 1, p. 64–72, abr. 2007.

TEUFEL, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. **Nature Biotechnology**, v. 40, n. 7, p. 1023–1025, 3 jul. 2022.

THERASME, O. et al. Life cycle greenhouse gas emissions of ethanol produced via fermentation of sugars derived from shrub willow (*Salix* spp.) hot water extraction in the Northeast United States. **Biotechnology for Biofuels**, v. 14, n. 1, p. 52, 1 dez. 2021.

THOMAS, B. N.; PARRILL, A. L.; BAKER, D. L. Self-docking and cross-docking simulations of G protein-coupled receptor-ligand complexes: impact of ligand type and

- receptor activation state. **Journal of Molecular Graphics and Modelling**, v. 112, p. 108119, 1 maio 2022.
- TOMPA, P. Intrinsically unstructured proteins. **Trends in Biochemical Sciences**, v. 27, n. 10, p. 527–533, out. 2002.
- TORRES, P. H. M. et al. Key topics in molecular docking for drug design. **International Journal of Molecular Sciences**, v. 20, n. 18, p. 4574, 15 set. 2019.
- TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, p. NA-NA, 2009.
- TUNYASUVUNAKOOL, K. et al. Highly accurate protein structure prediction for the human proteome. **Nature**, v. 596, n. 7873, p. 590–596, 26 ago. 2021.
- URBAN, P.; POMPON, D. Confrontation of AlphaFold models with experimental structures enlightens conformational dynamics supporting CYP102A1 functions. **Scientific Reports**, v. 12, n. 1, p. 15982, 25 set. 2022.
- VAN KEMPEN, M. et al. Fast and accurate protein structure search with Foldseek. **Nature Biotechnology**, n. 3, 8 maio 2023.
- VARADI, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. **Nucleic Acids Research**, v. 50, n. D1, p. D439–D444, 7 jan. 2022.
- VARROT, A. et al. Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio- $\beta$ -cellobioside, at 1.9 Å. **Acta Crystallographica Section D: Biological Crystallography**, v. 58, n. 12, p. 2201–2204, 1 dez. 2002.
- VARROT, A.; SCHÜLEIN, M.; DAVIES, G. J. Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. **Biochemistry**, v. 38, n. 28, p. 8884–8891, 1 jul. 1999.
- VERKUIL, R. et al. Language models generalize beyond natural proteins. **bioRxiv**, p. 2022.12.21.521521, 2022.
- VETTORE, A. L. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, v. 13, n. 12, p. 2725–2735, 1 dez. 2003.
- VIOLOT, S. et al. Exploring molecular determinants of *polysaccharide lyase family 6–1 enzyme* activity. **Glycobiology**, v. 31, n. 11, p. 1557–1570, 18 dez. 2021.
- WANG, J. et al. Automatic atom type and bond type perception in molecular mechanical calculations. **Journal of Molecular Graphics and Modelling**, v. 25, n. 2, p. 247–260, out. 2006.
- WANG, L. et al. Economic and GHG emissions analyses for sugarcane ethanol in Brazil: Looking forward. **Renewable and Sustainable Energy Reviews**, v. 40, p. 571–582, dez. 2014.

WATERHOUSE, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, n. W1, p. W296–W303, 2 jul. 2018.

WIERENGA, R. . The TIM-barrel fold: a versatile framework for efficient enzymes. **FEBS Letters**, v. 492, n. 3, p. 193–198, 16 mar. 2001.

WILLIAMS, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. **Protein Science**, v. 27, n. 1, p. 293–315, 27 jan. 2018.

WU, M. et al. Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. **Journal of Biological Chemistry**, v. 288, n. 46, p. 33107–33117, nov. 2013.

XIE, L.; XIE, L.; BOURNE, P. E. Structure-based systems biology for analyzing off-target binding. **Current Opinion in Structural Biology**, v. 21, n. 2, p. 189–199, abr. 2011.

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, 1 jul. 2015.

YOUNG, J. M. et al. A multilocus sequence analysis of the genus *Xanthomonas*. **Systematic and Applied Microbiology**, v. 31, n. 5, p. 366–377, 1 out. 2008.

YU, C.-S. et al. Prediction of protein subcellular localization. **Proteins: Structure, Function, and Bioinformatics**, v. 64, n. 3, p. 643–651, 2 jun. 2006.

ZABED, H. et al. Fuel ethanol production from lignocellulosic biomass: an overview on feedstocks and technological approaches. **Renewable and Sustainable Energy Reviews**, v. 66, p. 751–774, dez. 2016.

ZHANG, H.-L. et al. Complete Genome Sequence Reveals Evolutionary and Comparative Genomic Features of *Xanthomonas albilineans* Causing Sugarcane Leaf Scald. **Microorganisms**, v. 8, n. 2, p. 182, 28 jan. 2020.

ZHANG, L.; XU, J.; BIRCH, R. G. Engineered detoxification confers resistance against a pathogenic bacterium. **Nature Biotechnology**, v. 17, n. 10, p. 1021–1024, out. 1999.

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, 8 dez. 2004.

ZHENG, J. et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. **Nucleic Acids Research**, v. 51, n. W1, p. W115–W121, 5 jul. 2023.

ZHOU, X. et al. Assembling multidomain protein structures through analogous global structural alignments. **Proceedings of the National Academy of Sciences**, v. 116, n. 32, p. 15930–15938, 6 ago. 2019.

## APÊNDICE A - Triagem *in silico* do proteoma de *Xanthomonas albilineans*.

**Tabela A1** - Lista das 486 proteínas identificadas no proteoma de *Xanthomonas albilineans* que possuem peptídeo sinal, conforme previsão realizada pelo software SignalP. As proteínas são categorizadas conforme a probabilidade de possuírem um dos cinco tipos de peptídeo sinal: o peptídeo “padrão” - Sec/SPI (SP); o peptídeo sinal proveniente de lipoproteínas - Sec/SPII (LIPO); peptídeo sinal de transportadores TAT - Tat/SPI (TAT); peptídeo sinal de lipoproteínas transportadas por TAT - Tat/SPII (TATLIPO); pilina e peptídeo sinal semelhante a pilina - Sec/SPIII (PILIN).

Proteína	Peptídeo sinal	NO_SP	SP	LIPO	TAT	TATLIPO	PILIN	Posição de clivagem (Pr: Probabilidade)
XALr_3249	SP	0,000337	0,998810	0,000235	0,000251	0,000199	0,000188	CS pos: 24-25. Pr: 0,9727
XALr_3257	LIPO	0,000000	0,000000	1,000020	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9969
XALr_3258	SP	0,000286	0,998968	0,000210	0,000213	0,000173	0,000164	CS pos: 26-27. Pr: 0,9739
XALr_3261	SP	0,171168	0,822180	0,003874	0,001107	0,000694	0,000951	CS pos: 37-38. Pr: 0,8960
XALr_3255.5	LIPO	0,000000	0,000003	0,999996	0,000000	0,000000	0,000000	CS pos: 22-23. Pr: 0,9966
XALq_3221	SP	0,000176	0,999267	0,000153	0,000161	0,000132	0,000131	CS pos: 21-22. Pr: 0,9764
XALq_3224	LIPO	0,000000	0,000000	1,000036	0,000000	0,000000	0,000000	CS pos: 15-16. Pr: 0,9968
XALq_3228	SP	0,342827	0,642358	0,003567	0,004714	0,001237	0,005276	CS pos: 26-27. Pr: 0,7811
XALq_3230	SP	0,000239	0,999095	0,000175	0,000198	0,000151	0,000145	CS pos: 24-25. Pr: 0,9785
XALp_3183	SP	0,000180	0,999236	0,000159	0,000176	0,000134	0,000143	CS pos: 15-16. Pr: 0,9745
XALp_3186	SP	0,000203	0,999182	0,000175	0,000174	0,000142	0,000138	CS pos: 24-25. Pr: 0,9789
XALp_3189	SP	0,330580	0,656003	0,004416	0,003320	0,001471	0,004202	CS pos: 26-27. Pr: 0,7923
XALp_3190	SP	0,000173	0,662306	0,337261	0,000108	0,000087	0,000084	CS pos: 24-25. Pr: 0,6458
XALc_0006	LIPO	0,000000	0,000000	1,000021	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9970

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0007	SP	0,000215	0,999105	0,000185	0,000186	0,000155	0,000148	CS pos: 29-30. Pr: 0,9680
XALc_0030	SP	0,000207	0,999145	0,000170	0,000187	0,000150	0,000141	CS pos: 24-25. Pr: 0,9770
XALc_0032	LIPO	0,000000	0,000000	0,999991	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9954
XALc_0036	SP	0,000176	0,999227	0,000158	0,000172	0,000137	0,000135	CS pos: 22-23. Pr: 0,9795
XALc_0038	SP	0,000185	0,665887	0,333544	0,000144	0,000118	0,000123	CS pos: 28-29. Pr: 0,6461
XALc_0043	LIPO	0,000006	0,000395	0,999588	0,000001	0,000000	0,000000	CS pos: 28-29. Pr: 0,9962
XALc_0045	SP	0,000230	0,999030	0,000196	0,000218	0,000175	0,000161	CS pos: 32-33. Pr: 0,9700
XALc_0048	LIPO	0,000000	0,000000	1,000020	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9969
XALc_0051	LIPO	0,000000	0,000000	1,000036	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9935
XALc_0052	LIPO	0,166672	0,000000	0,833307	0,000000	0,000000	0,000000	CS pos: 5-6. Pr: 0,8155
XALc_0053	LIPO	0,000000	0,000000	1,000040	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9966
XALc_0056	SP	0,166757	0,832684	0,000171	0,000145	0,000118	0,000121	CS pos: 51-52. Pr: 0,8255
XALc_0058	SP	0,000238	0,999006	0,000204	0,000219	0,000179	0,000173	CS pos: 29-30. Pr: 0,9749
XALc_0059	SP	0,001526	0,997522	0,000275	0,000269	0,000205	0,000221	CS pos: 9-10. Pr: 0,8202
XALc_0061	SP	0,000215	0,779227	0,220131	0,000164	0,000133	0,000128	CS pos: 27-28. Pr: 0,5041
XALc_0070	SP	0,000227	0,999103	0,000181	0,000197	0,000158	0,000156	CS pos: 22-23. Pr: 0,9762
XALc_0076	SP	0,000210	0,832463	0,166898	0,000161	0,000134	0,000152	CS pos: 22-23. Pr: 0,7930
XALc_0087	SP	0,000374	0,998635	0,000281	0,000282	0,000227	0,000198	CS pos: 25-26. Pr: 0,9747

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0090	SP	0,000200	0,999195	0,000175	0,000173	0,000136	0,000133	CS pos: 26-27. Pr: 0,9741
XALc_0094	SP	0,295169	0,697376	0,004502	0,001331	0,000763	0,000883	CS pos: 23-24. Pr: 0,7417
XALc_0096	SP	0,000173	0,999310	0,000145	0,000156	0,000126	0,000130	CS pos: 32-33. Pr: 0,9807
XALc_0097	SP	0,000282	0,999025	0,000205	0,000192	0,000154	0,000161	CS pos: 31-32. Pr: 0,4356
XALc_0102	SP	0,000349	0,998656	0,000276	0,000278	0,000240	0,000211	CS pos: 20-21. Pr: 0,9548
XALc_0114	SP	0,000202	0,999128	0,000162	0,000195	0,000154	0,000145	CS pos: 30-31. Pr: 0,9734
XALc_0118	LIPO	0,000000	0,000000	1,000043	0,000000	0,000000	0,000000	CS pos: 14-15. Pr: 0,9942
XALc_0121	SP	0,167074	0,831529	0,000440	0,000363	0,000290	0,000305	CS pos: 36-37. Pr: 0,7896
XALc_0122	SP	0,002933	0,996283	0,000223	0,000230	0,000172	0,000181	CS pos: 13-14. Pr: 0,9425
XALc_0134	SP	0,000280	0,998913	0,000243	0,000223	0,000173	0,000174	CS pos: 22-23. Pr: 0,9726
XALc_0138	SP	0,000229	0,999103	0,000170	0,000181	0,000149	0,000145	CS pos: 23-24. Pr: 0,9777
XALc_0168	SP	0,000254	0,999001	0,000199	0,000214	0,000167	0,000167	CS pos: 25-26. Pr: 0,9751
XALc_0189	SP	0,000204	0,999151	0,000163	0,000185	0,000146	0,000148	CS pos: 26-27. Pr: 0,9724
XALc_0215	LIPO	0,001712	0,202561	0,795447	0,000182	0,000049	0,000041	CS pos: 31-32. Pr: 0,8298
XALc_0217	LIPO	0,000000	0,000000	1,000013	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9955
XALc_0241	SP	0,000160	0,999321	0,000138	0,000151	0,000119	0,000119	CS pos: 22-23. Pr: 0,9821
XALc_0244	SP	0,000446	0,998604	0,000293	0,000256	0,000208	0,000205	CS pos: 12-13. Pr: 0,9134
XALc_0248	TAT	0,000000	0,000000	0,000000	1,000004	0,000000	0,000000	CS pos: 25-26. Pr: 0,9202

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0250	LIPO	0,000062	0,499787	0,500068	0,000038	0,000034	0,000033	CS pos: 18-19. Pr: 0,4958
XALc_0255	LIPO	0,000000	0,000000	0,999999	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9950
XALc_0256	LIPO	0,000000	0,000000	1,000003	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9944
XALc_0261	LIPO	0,000034	0,166573	0,833374	0,000017	0,000015	0,000015	CS pos: 19-20. Pr: 0,8270
XALc_0272	SP	0,000912	0,665817	0,000188	0,332804	0,000141	0,000156	CS pos: 42-43. Pr: 0,6415
XALc_0308	SP	0,000191	0,999192	0,000165	0,000175	0,000141	0,000142	CS pos: 18-19. Pr: 0,9770
XALc_0316	SP	0,000192	0,999211	0,000164	0,000170	0,000132	0,000132	CS pos: 22-23. Pr: 0,9754
XALc_0320	SP	0,000320	0,998753	0,000270	0,000275	0,000219	0,000186	CS pos: 25-26. Pr: 0,9739
XALc_0326	SP	0,167362	0,831750	0,000256	0,000244	0,000192	0,000178	CS pos: 10-11. Pr: 0,7965
XALc_0327	SP	0,000329	0,664386	0,334958	0,000128	0,000107	0,000104	CS pos: 27-28. Pr: 0,6473
XALc_0328	SP	0,000147	0,666162	0,333445	0,000094	0,000080	0,000078	CS pos: 27-28. Pr: 0,6531
XALc_0331	SP	0,000187	0,999190	0,000156	0,000182	0,000141	0,000142	CS pos: 27-28. Pr: 0,9734
XALc_0332	TAT	0,000000	0,000000	0,000000	0,999983	0,000000	0,000000	CS pos: 32-33. Pr: 0,8217
XALc_0344	LIPO	0,000000	0,000000	1,000027	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9950
XALc_0348	LIPO	0,000000	0,000127	0,999874	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9766
XALc_0368	SP	0,023227	0,954958	0,019900	0,000753	0,000665	0,000489	CS pos: 41-42. Pr: 0,9451
XALc_0376	SP	0,000189	0,999208	0,000159	0,000174	0,000134	0,000140	CS pos: 25-26. Pr: 0,9770
XALc_0391	SP	0,000289	0,998918	0,000223	0,000220	0,000179	0,000177	CS pos: 21-22. Pr: 0,9787

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0392	SP	0,159141	0,840357	0,000167	0,000122	0,000110	0,000111	CS pos: 15-16. Pr: 0,8139
XALc_0393	SP	0,000193	0,999222	0,000161	0,000169	0,000133	0,000135	CS pos: 20-21. Pr: 0,9778
XALc_0397	SP	0,002402	0,991467	0,001760	0,001607	0,001322	0,001435	CS pos: 28-29. Pr: 0,9678
XALc_0400	SP	0,242317	0,745828	0,007911	0,001360	0,000962	0,001624	CS pos: 29-30. Pr: 0,7939
XALc_0410	LIPO	0,000021	0,000688	0,999296	0,000001	0,000000	0,000000	CS pos: 26-27. Pr: 0,9957
XALc_0422	SP	0,110569	0,888822	0,000172	0,000167	0,000136	0,000136	CS pos: 50-51. Pr: 0,8735
XALc_0423	LIPO	0,000002	0,000249	0,999734	0,000001	0,000000	0,000000	CS pos: 27-28. Pr: 0,9973
XALc_0425	SP	0,000245	0,999069	0,000190	0,000197	0,000152	0,000147	CS pos: 26-27. Pr: 0,9743
XALc_0434	SP	0,399738	0,599295	0,000229	0,000305	0,000226	0,000210	CS pos: 34-35. Pr: 0,5882
XALc_0437	SP	0,000154	0,999353	0,000128	0,000144	0,000115	0,000113	CS pos: 21-22. Pr: 0,9838
XALc_0452	LIPO	0,000000	0,000000	1,000072	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9970
XALc_0469	SP	0,000227	0,999084	0,000182	0,000202	0,000157	0,000155	CS pos: 24-25. Pr: 0,6689
XALc_0481	LIPO	0,000000	0,000000	1,000035	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9968
XALc_0484	SP	0,000319	0,998834	0,000264	0,000232	0,000180	0,000187	CS pos: 32-33. Pr: 0,9753
XALc_0488	SP	0,000221	0,999147	0,000179	0,000183	0,000153	0,000141	CS pos: 25-26. Pr: 0,9734
XALc_0489	LIPO	0,000000	0,000000	1,000050	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9960
XALc_0541	SP	0,000228	0,999031	0,000197	0,000213	0,000171	0,000157	CS pos: 23-24. Pr: 0,9735
XALc_0547	SP	0,136471	0,862918	0,000180	0,000163	0,000131	0,000132	CS pos: 53-54. Pr: 0,8116

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0549	SP	0,000777	0,998057	0,000369	0,000327	0,000259	0,000237	CS pos: 18-19. Pr: 0,9331
XALc_0550	SP	0,317986	0,681133	0,000226	0,000257	0,000199	0,000183	CS pos: 30-31. Pr: 0,6331
XALc_0551	SP	0,000194	0,999170	0,000162	0,000188	0,000149	0,000150	CS pos: 29-30. Pr: 0,9751
XALc_0552	SP	0,000317	0,832161	0,000248	0,166927	0,000199	0,000181	CS pos: 20-21. Pr: 0,7788
XALc_0553	SP	0,000503	0,998351	0,000366	0,000333	0,000252	0,000216	CS pos: 27-28. Pr: 0,9718
XALc_0557	SP	0,096063	0,899097	0,001363	0,001273	0,001262	0,000943	CS pos: 22-23. Pr: 0,9558
XALc_0558	SP	0,000214	0,999117	0,000182	0,000197	0,000150	0,000154	CS pos: 20-21. Pr: 0,9753
XALc_0562	SP	0,000194	0,999148	0,000166	0,000175	0,000139	0,000148	CS pos: 24-25. Pr: 0,9795
XALc_0576	LIPO	0,000085	0,499723	0,500081	0,000062	0,000055	0,000056	CS pos: 15-16. Pr: 0,4934
XALc_0580	SP	0,337344	0,662001	0,000222	0,000157	0,000164	0,000135	CS pos: 55-56. Pr: 0,4490
XALc_0616	LIPO	0,000000	0,000000	1,000061	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9973
XALc_0636	SP	0,000227	0,999104	0,000194	0,000196	0,000160	0,000164	CS pos: 36-37. Pr: 0,9654
XALc_0643	SP	0,000516	0,998324	0,000340	0,000320	0,000264	0,000227	CS pos: 26-27. Pr: 0,9698
XALc_0644	SP	0,000593	0,997500	0,000588	0,000459	0,000457	0,000431	CS pos: 27-28. Pr: 0,9732
XALc_0646	SP	0,000227	0,832310	0,166987	0,000172	0,000151	0,000161	CS pos: 25-26. Pr: 0,8063
XALc_0647	SP	0,000202	0,999152	0,000168	0,000188	0,000143	0,000147	CS pos: 23-24. Pr: 0,9736
XALc_0658	TAT	0,000000	0,000000	0,000000	0,999991	0,000000	0,000000	CS pos: 35-36. Pr: 0,9188
XALc_0659	SP	0,001154	0,997272	0,000649	0,000385	0,000280	0,000266	CS pos: 28-29. Pr: 0,9610

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0666	PILIN	0,000005	0,000000	0,000000	0,000000	0,000000	0,999982	CS pos: 6-7. Pr: 0,7323
XALc_0676	SP	0,000192	0,999130	0,000171	0,000185	0,000147	0,000155	CS pos: 25-26. Pr: 0,9742
XALc_0679	LIPO	0,154593	0,006687	0,838705	0,000005	0,000003	0,000002	CS pos: 31-32. Pr: 0,8318
XALc_0692	SP	0,000438	0,998384	0,000326	0,000372	0,000268	0,000205	CS pos: 25-26. Pr: 0,9595
XALc_0694	LIPO	0,000002	0,000020	0,999984	0,000000	0,000000	0,000000	CS pos: 27-28. Pr: 0,9969
XALc_0700	SP	0,000341	0,998669	0,000285	0,000282	0,000234	0,000204	CS pos: 23-24. Pr: 0,9740
XALc_0709	LIPO	0,000000	0,000000	1,000018	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9970
XALc_0713	SP	0,000356	0,998495	0,000336	0,000305	0,000271	0,000240	CS pos: 20-21. Pr: 0,9726
XALc_0714	SP	0,404662	0,588415	0,001444	0,002342	0,001283	0,001837	CS pos: 4-5. Pr: 0,0145
XALc_0716	SP	0,000208	0,999029	0,000185	0,000215	0,000166	0,000171	CS pos: 14-15. Pr: 0,9574
XALc_0723	SP	0,000573	0,997551	0,000526	0,000463	0,000489	0,000398	CS pos: 25-26. Pr: 0,9720
XALc_0727	SP	0,000118	0,675225	0,324411	0,000085	0,000075	0,000075	CS pos: 26-27. Pr: 0,6581
XALc_0731	SP	0,000362	0,998570	0,000295	0,000325	0,000254	0,000215	CS pos: 27-28. Pr: 0,9702
XALc_0737	SP	0,000179	0,999235	0,000153	0,000173	0,000133	0,000141	CS pos: 20-21. Pr: 0,9797
XALc_0747	SP	0,001260	0,987569	0,007983	0,001467	0,000726	0,000997	CS pos: 21-22. Pr: 0,9737
XALc_0750	SP	0,023139	0,809381	0,166937	0,000224	0,000178	0,000162	CS pos: 43-44. Pr: 0,7771
XALc_0764	SP	0,333338	0,666317	0,000109	0,000086	0,000076	0,000073	CS pos: 17-18. Pr: 0,6467
XALc_0766	SP	0,000205	0,999160	0,000173	0,000193	0,000154	0,000150	CS pos: 28-29. Pr: 0,9759

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0771	SP	0,000150	0,666102	0,333468	0,000103	0,000094	0,000085	CS pos: 22-23. Pr: 0,6486
XALc_0781	LIPO	0,000025	0,166592	0,833373	0,000016	0,000014	0,000014	CS pos: 17-18. Pr: 0,8305
XALc_0810	LIPO	0,333325	0,000008	0,666688	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,6481
XALc_0811	SP	0,000758	0,996782	0,000791	0,000544	0,000623	0,000500	CS pos: 22-23. Pr: 0,9604
XALc_0812	LIPO	0,000000	0,000000	1,000047	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9976
XALc_0813	LIPO	0,000000	0,000000	1,000070	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9968
XALc_0814	SP	0,000200	0,999200	0,000164	0,000181	0,000133	0,000141	CS pos: 20-21. Pr: 0,9789
XALc_0818	SP	0,000242	0,999022	0,000201	0,000212	0,000163	0,000167	CS pos: 22-23. Pr: 0,9858
XALc_0819	SP	0,000210	0,999208	0,000165	0,000172	0,000133	0,000133	CS pos: 22-23. Pr: 0,9785
XALc_0822	LIPO	0,338375	0,018169	0,643452	0,000000	0,000000	0,000000	CS pos: 50-51. Pr: 0,5772
XALc_0823	SP	0,000277	0,998620	0,000269	0,000289	0,000232	0,000279	CS pos: 25-26. Pr: 0,9754
XALc_0843	SP	0,365008	0,631662	0,002734	0,000245	0,000178	0,000179	CS pos: 57-58. Pr: 0,6544
XALc_0849	SP	0,000783	0,781851	0,000377	0,216392	0,000397	0,000227	CS pos: 27-28. Pr: 0,7434
XALc_0851	LIPO	0,000000	0,000000	1,000024	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9969
XALc_0852	SP	0,000309	0,998936	0,000223	0,000216	0,000171	0,000159	CS pos: 21-22. Pr: 0,9784
XALc_0853	SP	0,000244	0,999097	0,000186	0,000198	0,000147	0,000147	CS pos: 23-24. Pr: 0,9688
XALc_0857	LIPO	0,000071	0,499756	0,500059	0,000046	0,000037	0,000037	CS pos: 24-25. Pr: 0,4753
XALc_0858	SP	0,000212	0,999090	0,000190	0,000212	0,000165	0,000167	CS pos: 31-32. Pr: 0,9701

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0865	SP	0,000277	0,892343	0,106691	0,000271	0,000219	0,000225	CS pos: 38-39. Pr: 0,8788
XALc_0866	SP	0,000172	0,999289	0,000152	0,000159	0,000125	0,000124	CS pos: 22-23. Pr: 0,9789
XALc_0874	SP	0,000197	0,999213	0,000163	0,000178	0,000136	0,000135	CS pos: 38-39. Pr: 0,8965
XALc_0885	SP	0,463359	0,534095	0,000653	0,000867	0,000561	0,000478	CS pos: 14-15. Pr: 0,3712
XALc_0887	LIPO	0,000000	0,000000	1,000018	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9969
XALc_0907	LIPO	0,004429	0,358790	0,636584	0,000075	0,000050	0,000050	CS pos: 28-29. Pr: 0,6569
XALc_0909	SP	0,000283	0,998911	0,000235	0,000218	0,000175	0,000173	CS pos: 22-23. Pr: 0,9794
XALc_0927	SP	0,000323	0,832229	0,166906	0,000226	0,000174	0,000164	CS pos: 26-27. Pr: 0,8148
XALc_0929	LIPO	0,086270	0,125356	0,787722	0,000248	0,000212	0,000213	CS pos: 25-26. Pr: 0,8202
XALc_0933	SP	0,000169	0,999272	0,000149	0,000164	0,000130	0,000128	CS pos: 23-24. Pr: 0,9749
XALc_0941	LIPO	0,000000	0,000000	1,000047	0,000000	0,000000	0,000000	CS pos: 23-24. Pr: 0,9975
XALc_0946	SP	0,000303	0,998747	0,000281	0,000274	0,000213	0,000203	CS pos: 35-36. Pr: 0,9663
XALc_0957	LIPO	0,000142	0,039940	0,959897	0,000002	0,000000	0,000000	CS pos: 35-36. Pr: 0,9926
XALc_0962	SP	0,000185	0,999213	0,000163	0,000173	0,000137	0,000137	CS pos: 29-30. Pr: 0,4357
XALc_0964	SP	0,000183	0,999212	0,000156	0,000175	0,000140	0,000141	CS pos: 24-25. Pr: 0,9839
XALc_0965	LIPO	0,000000	0,000000	1,000015	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9964
XALc_0967	SP	0,000438	0,831580	0,000360	0,167062	0,000323	0,000246	CS pos: 32-33. Pr: 0,8002
XALc_0975	SP	0,000266	0,998897	0,000238	0,000240	0,000190	0,000180	CS pos: 28-29. Pr: 0,9751

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_0982	SP	0,002810	0,996515	0,000191	0,000188	0,000148	0,000144	CS pos: 44-45. Pr: 0,9679
XALc_0991	SP	0,000391	0,998483	0,000310	0,000306	0,000268	0,000229	CS pos: 30-31. Pr: 0,9627
XALc_0993	SP	0,017525	0,978251	0,001840	0,000976	0,000730	0,000717	CS pos: 24-25. Pr: 0,9658
XALc_0994	LIPO	0,000000	0,000000	1,000042	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9980
XALc_1030	SP	0,000176	0,999227	0,000156	0,000161	0,000130	0,000130	CS pos: 19-20. Pr: 0,9790
XALc_1034	SP	0,000196	0,999167	0,000171	0,000185	0,000147	0,000149	CS pos: 24-25. Pr: 0,9770
XALc_1036	SP	0,000215	0,999079	0,000179	0,000205	0,000166	0,000175	CS pos: 23-24. Pr: 0,9755
XALc_1068	SP	0,000227	0,999108	0,000186	0,000185	0,000151	0,000148	CS pos: 20-21. Pr: 0,9789
XALc_1085	LIPO	0,000000	0,000000	1,000018	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9950
XALc_1095	LIPO	0,000000	0,000000	1,000046	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9967
XALc_1106	LIPO	0,000023	0,000071	0,999889	0,000001	0,000000	0,000000	CS pos: 25-26. Pr: 0,9945
XALc_1109	LIPO	0,000001	0,000004	0,999979	0,000000	0,000000	0,000000	CS pos: 26-27. Pr: 0,9961
XALc_1116	SP	0,000442	0,998315	0,000330	0,000387	0,000296	0,000238	CS pos: 24-25. Pr: 0,9304
XALc_1123	SP	0,000250	0,998914	0,000235	0,000234	0,000183	0,000186	CS pos: 26-27. Pr: 0,9658
XALc_1126	SP	0,000234	0,999052	0,000195	0,000194	0,000155	0,000164	CS pos: 31-32. Pr: 0,9753
XALc_1129	SP	0,000204	0,999166	0,000180	0,000167	0,000139	0,000134	CS pos: 22-23. Pr: 0,9767
XALc_1132	LIPO	0,000000	0,000000	1,000028	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9909
XALc_1136	SP	0,024184	0,974771	0,000291	0,000291	0,000245	0,000240	CS pos: 26-27. Pr: 0,4518

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_1141	LIPO	0,000000	0,000000	1,000014	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9942
XALc_1143	SP	0,166723	0,824574	0,008159	0,000222	0,000159	0,000151	CS pos: 54-55. Pr: 0,8105
XALc_1148	SP	0,000244	0,999001	0,000207	0,000210	0,000172	0,000167	CS pos: 26-27. Pr: 0,9466
XALc_1160	LIPO	0,000002	0,000045	0,999985	0,000000	0,000000	0,000000	CS pos: 25-26. Pr: 0,9946
XALc_1180	LIPO	0,000000	0,000000	1,000026	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9960
XALc_1185	LIPO	0,000000	0,000000	0,999997	0,000000	0,000000	0,000000	CS pos: 26-27. Pr: 0,9801
XALc_1192	SP	0,000501	0,998177	0,000367	0,000402	0,000305	0,000256	CS pos: 41-42. Pr: 0,9657
XALc_1193	SP	0,000231	0,999059	0,000190	0,000196	0,000155	0,000148	CS pos: 26-27. Pr: 0,9708
XALc_1194	SP	0,000335	0,998882	0,000254	0,000208	0,000168	0,000155	CS pos: 36-37. Pr: 0,9733
XALc_1198	SP	0,000360	0,998482	0,000327	0,000311	0,000265	0,000265	CS pos: 24-25. Pr: 0,6285
XALc_1199	TAT	0,000000	0,000000	0,000000	0,999990	0,000000	0,000000	CS pos: 25-26. Pr: 0,9035
XALc_1227	SP	0,000344	0,998592	0,000283	0,000307	0,000231	0,000242	CS pos: 43-44. Pr: 0,9515
XALc_1251	SP	0,190548	0,808626	0,000283	0,000196	0,000173	0,000154	CS pos: 52-53. Pr: 0,7923
XALc_1263	SP	0,000240	0,999067	0,000205	0,000187	0,000155	0,000155	CS pos: 22-23. Pr: 0,9748
XALc_1264	TAT	0,000000	0,000000	0,000000	0,666694	0,333316	0,000000	CS pos: 47-48. Pr: 0,4093
XALc_1266	SP	0,000257	0,999021	0,000204	0,000212	0,000168	0,000151	CS pos: 24-25. Pr: 0,9762
XALc_1278	SP	0,000335	0,998880	0,000225	0,000232	0,000177	0,000167	CS pos: 26-27. Pr: 0,9783
XALc_1281	PILIN	0,025004	0,031263	0,000002	0,000046	0,000150	0,943525	CS pos: 17-18. Pr: 0,6932

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_1282	PILIN	0,000003	0,000022	0,000000	0,000001	0,000000	0,999968	CS pos: 12-13. Pr: 0,7021
XALc_1283	PILIN	0,000402	0,001805	0,000000	0,000007	0,000022	0,997756	CS pos: 15-16. Pr: 0,6384
XALc_1285	SP	0,000244	0,999006	0,000202	0,000216	0,000167	0,000179	CS pos: 15-16. Pr: 0,9500
XALc_1286	PILIN	0,000073	0,000271	0,000000	0,000005	0,000007	0,999642	CS pos: 14-15. Pr: 0,7569
XALc_1291	SP	0,000193	0,999229	0,000163	0,000168	0,000132	0,000127	CS pos: 27-28. Pr: 0,9780
XALc_1301	SP	0,000167	0,999303	0,000144	0,000161	0,000128	0,000125	CS pos: 22-23. Pr: 0,9821
XALc_1305	SP	0,281791	0,714119	0,002180	0,000843	0,000508	0,000550	CS pos: 60-61. Pr: 0,6290
XALc_1306	SP	0,397304	0,601890	0,000210	0,000229	0,000183	0,000179	CS pos: 24-25. Pr: 0,4853
XALc_1314	LIPO	0,000000	0,000003	0,999998	0,000000	0,000000	0,000000	CS pos: 23-24. Pr: 0,9965
XALc_1315	LIPO	0,000000	0,000000	1,000030	0,000000	0,000000	0,000000	CS pos: 22-23. Pr: 0,9968
XALc_1318	SP	0,000162	0,999304	0,000132	0,000148	0,000119	0,000118	CS pos: 23-24. Pr: 0,9821
XALc_1319	LIPO	0,000115	0,333048	0,666734	0,000042	0,000039	0,000034	CS pos: 39-40. Pr: 0,5974
XALc_1328	SP	0,000217	0,999074	0,000188	0,000215	0,000169	0,000171	CS pos: 22-23. Pr: 0,9504
XALc_1329	SP	0,000128	0,666197	0,333436	0,000091	0,000084	0,000076	CS pos: 24-25. Pr: 0,6421
XALc_1358	SP	0,000168	0,999284	0,000148	0,000157	0,000127	0,000130	CS pos: 21-22. Pr: 0,9810
XALc_1363	SP	0,018232	0,640751	0,339913	0,000518	0,000271	0,000314	CS pos: 21-22. Pr: 0,6492
XALc_1386	SP	0,000224	0,999073	0,000200	0,000198	0,000170	0,000153	CS pos: 30-31. Pr: 0,9644
XALc_1400	SP	0,000164	0,999295	0,000143	0,000159	0,000129	0,000124	CS pos: 21-22. Pr: 0,9812

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_1421	SP	0,000214	0,999051	0,000198	0,000202	0,000166	0,000169	CS pos: 26-27. Pr: 0,9744
XALc_1422	LIPO	0,000000	0,000000	0,999990	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9686
XALc_1430	SP	0,180536	0,818739	0,000247	0,000179	0,000158	0,000145	CS pos: 58-59. Pr: 0,6596
XALc_1452	SP	0,005849	0,991784	0,001440	0,000412	0,000243	0,000307	CS pos: 21-22. Pr: 0,9719
XALc_1457	SP	0,000381	0,998367	0,000358	0,000345	0,000299	0,000270	CS pos: 30-31. Pr: 0,9744
XALc_1466	SP	0,000668	0,997807	0,000415	0,000419	0,000358	0,000333	CS pos: 43-44. Pr: 0,9424
XALc_1467	SP	0,000321	0,998950	0,000212	0,000215	0,000167	0,000151	CS pos: 30-31. Pr: 0,9776
XALc_1468	SP	0,000890	0,998198	0,000366	0,000229	0,000168	0,000166	CS pos: 41-42. Pr: 0,9707
XALc_1482	SP	0,332562	0,666310	0,000316	0,000320	0,000270	0,000232	CS pos: 6-7. Pr: 0,6032
XALc_1486	LIPO	0,000000	0,000000	1,000014	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9962
XALc_1491	SP	0,000229	0,999044	0,000204	0,000209	0,000171	0,000160	CS pos: 35-36. Pr: 0,9674
XALc_1541	SP	0,171098	0,810900	0,004227	0,005291	0,002224	0,006258	CS pos: 25-26. Pr: 0,9222
XALc_1555	LIPO	0,000167	0,002429	0,997379	0,000013	0,000001	0,000000	CS pos: 26-27. Pr: 0,9949
XALc_1556	LIPO	0,000267	0,498969	0,500196	0,000190	0,000188	0,000179	CS pos: 21-22. Pr: 0,4739
XALc_1557	LIPO	0,000000	0,000000	1,000003	0,000000	0,000000	0,000000	CS pos: 22-23. Pr: 0,9974
XALc_1568	SP	0,099962	0,888363	0,005701	0,002437	0,001336	0,002188	CS pos: 27-28. Pr: 0,9581
XALc_1569	SP	0,000323	0,998780	0,000263	0,000255	0,000198	0,000196	CS pos: 26-27. Pr: 0,9505
XALc_1587	LIPO	0,000000	0,000000	1,000027	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9962

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_1590	LIPO	0,000000	0,000000	1,000017	0,000000	0,000000	0,000000	CS pos: 22-23. Pr: 0,9969
XALc_1605	SP	0,000230	0,999075	0,000195	0,000191	0,000161	0,000162	CS pos: 22-23. Pr: 0,9782
XALc_1610	LIPO	0,000000	0,000004	1,000008	0,000000	0,000000	0,000000	CS pos: 22-23. Pr: 0,9971
XALc_1614	LIPO	0,000000	0,000000	1,000006	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9958
XALc_1615	LIPO	0,000000	0,000029	0,999971	0,000000	0,000000	0,000000	CS pos: 24-25. Pr: 0,9955
XALc_1651	SP	0,000612	0,998491	0,000277	0,000247	0,000197	0,000175	CS pos: 23-24. Pr: 0,9770
XALc_1656	SP	0,000218	0,999143	0,000169	0,000182	0,000143	0,000148	CS pos: 25-26. Pr: 0,9800
XALc_1657	SP	0,000254	0,999123	0,000181	0,000185	0,000143	0,000131	CS pos: 24-25. Pr: 0,9812
XALc_1661	SP	0,000188	0,999182	0,000176	0,000170	0,000138	0,000138	CS pos: 18-19. Pr: 0,9765
XALc_1662	LIPO	0,000065	0,166441	0,833399	0,000033	0,000029	0,000030	CS pos: 22-23. Pr: 0,8296
XALc_1663	LIPO	0,000000	0,000000	1,000044	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9952
XALc_1682	LIPO	0,000000	0,000000	1,000039	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9937
XALc_1691	LIPO	0,023277	0,101837	0,874861	0,000003	0,000003	0,000001	CS pos: 39-40. Pr: 0,9901
XALc_1705	TATLIPO	0,000000	0,000000	0,000000	0,333304	0,666681	0,000000	CS pos: 40-41. Pr: 0,5528
XALc_1727	SP	0,000326	0,998411	0,000595	0,000273	0,000197	0,000197	CS pos: 24-25. Pr: 0,9686
XALc_1744	SP	0,000306	0,998785	0,000280	0,000254	0,000191	0,000198	CS pos: 27-28. Pr: 0,8456
XALc_1762	LIPO	0,000000	0,000000	1,000035	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9943
XALc_1770	LIPO	0,000000	0,000000	1,000030	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9974

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_1775	LIPO	0,000000	0,000000	1,000029	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9978
XALc_1788	SP	0,000209	0,999154	0,000162	0,000180	0,000145	0,000144	CS pos: 12-13. Pr: 0,9550
XALc_1794	SP	0,000233	0,999098	0,000192	0,000195	0,000156	0,000153	CS pos: 39-40. Pr: 0,9728
XALc_1837	SP	0,045432	0,940744	0,006648	0,003306	0,001690	0,002169	CS pos: 37-38. Pr: 0,9603
XALc_1841	SP	0,000173	0,999247	0,000164	0,000163	0,000136	0,000131	CS pos: 26-27. Pr: 0,9792
XALc_1843	LIPO	0,000000	0,000000	1,000017	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9974
XALc_1866	SP	0,000220	0,999050	0,000188	0,000209	0,000160	0,000172	CS pos: 20-21. Pr: 0,9754
XALc_1871	PILIN	0,000000	0,000000	0,000000	0,000000	0,000000	1,000018	CS pos: 8-9. Pr: 0,7881
XALc_1884	LIPO	0,334987	0,229719	0,435123	0,000079	0,000045	0,000040	CS pos: 54-55. Pr: 0,2066
XALc_1885	SP	0,000188	0,999235	0,000160	0,000172	0,000129	0,000133	CS pos: 23-24. Pr: 0,9757
XALc_1897	SP	0,000228	0,999008	0,000204	0,000204	0,000169	0,000166	CS pos: 22-23. Pr: 0,5109
XALc_1907	SP	0,000179	0,999223	0,000155	0,000170	0,000139	0,000133	CS pos: 26-27. Pr: 0,9761
XALc_1910	SP	0,000229	0,998990	0,000194	0,000222	0,000174	0,000181	CS pos: 23-24. Pr: 0,9741
XALc_1914	SP	0,328670	0,670595	0,000192	0,000223	0,000174	0,000183	CS pos: 12-13. Pr: 0,5561
XALc_1915	SP	0,000199	0,999155	0,000179	0,000174	0,000141	0,000143	CS pos: 23-24. Pr: 0,9645
XALc_1923	SP	0,000221	0,999089	0,000186	0,000186	0,000156	0,000152	CS pos: 25-26. Pr: 0,9755
XALc_1934	LIPO	0,000000	0,000000	1,000028	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9971
XALc_1940	SP	0,000225	0,832433	0,166839	0,000203	0,000167	0,000155	CS pos: 21-22. Pr: 0,8212

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_1942	LIPO	0,000000	0,000000	1,000010	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9965
XALc_1946	SP	0,000423	0,998379	0,000317	0,000366	0,000284	0,000230	CS pos: 36-37. Pr: 0,9684
XALc_1947	LIPO	0,000278	0,196421	0,803239	0,000025	0,000016	0,000015	CS pos: 30-31. Pr: 0,8304
XALc_1948	TAT	0,000000	0,000000	0,000000	0,834889	0,165125	0,000000	CS pos: 56-57. Pr: 0,7684
XALc_1949	SP	0,000181	0,999255	0,000152	0,000166	0,000125	0,000134	CS pos: 18-19. Pr: 0,9760
XALc_1950	SP	0,000191	0,999232	0,000160	0,000169	0,000130	0,000134	CS pos: 20-21. Pr: 0,9754
XALc_1951	LIPO	0,000038	0,343721	0,656219	0,000017	0,000017	0,000016	CS pos: 15-16. Pr: 0,5123
XALc_1952	SP	0,000137	0,666105	0,333454	0,000102	0,000087	0,000086	CS pos: 19-20. Pr: 0,6488
XALc_1953	SP	0,000805	0,996860	0,000529	0,001098	0,000412	0,000300	CS pos: 43-44. Pr: 0,9666
XALc_1954	TAT	0,000000	0,000000	0,000000	0,999994	0,000000	0,000000	CS pos: 28-29. Pr: 0,9340
XALc_1956	LIPO	0,000004	0,166654	0,833347	0,000002	0,000003	0,000002	CS pos: 16-17. Pr: 0,8011
XALc_1957	SP	0,000162	0,999314	0,000140	0,000158	0,000124	0,000124	CS pos: 15-16. Pr: 0,9725
XALc_1959	SP	0,000205	0,665935	0,333489	0,000143	0,000123	0,000107	CS pos: 28-29. Pr: 0,6426
XALc_1984	SP	0,000176	0,999220	0,000156	0,000174	0,000137	0,000143	CS pos: 21-22. Pr: 0,9803
XALc_1989	SP	0,000184	0,832600	0,166827	0,000162	0,000130	0,000124	CS pos: 23-24. Pr: 0,8149
XALc_1990	LIPO	0,000000	0,000000	1,000013	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9954
XALc_1995	SP	0,000297	0,998803	0,000248	0,000243	0,000196	0,000191	CS pos: 26-27. Pr: 0,9661
XALc_2002	SP	0,000185	0,999207	0,000161	0,000173	0,000142	0,000138	CS pos: 22-23. Pr: 0,9775

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2006	SP	0,108383	0,723725	0,167007	0,000352	0,000275	0,000247	CS pos: 45-46. Pr: 0,5227
XALc_2009	LIPO	0,000000	0,000000	1,000026	0,000000	0,000000	0,000000	CS pos: 11-12. Pr: 0,9944
XALc_2014	SP	0,000323	0,998848	0,000237	0,000253	0,000195	0,000169	CS pos: 23-24. Pr: 0,9696
XALc_2021	SP	0,000200	0,999216	0,000160	0,000174	0,000137	0,000134	CS pos: 23-24. Pr: 0,9794
XALc_2022	SP	0,000155	0,999344	0,000128	0,000144	0,000114	0,000116	CS pos: 25-26. Pr: 0,7680
XALc_2023	SP	0,000195	0,999236	0,000154	0,000169	0,000132	0,000133	CS pos: 28-29. Pr: 0,9764
XALc_2024	SP	0,116729	0,594896	0,288074	0,000105	0,000097	0,000089	CS pos: 52-53. Pr: 0,5045
XALc_2025	SP	0,000304	0,998854	0,000265	0,000240	0,000191	0,000181	CS pos: 25-26. Pr: 0,9778
XALc_2034	SP	0,000543	0,997636	0,000511	0,000449	0,000470	0,000395	CS pos: 24-25. Pr: 0,9614
XALc_2045	SP	0,000180	0,999248	0,000154	0,000168	0,000130	0,000130	CS pos: 23-24. Pr: 0,9788
XALc_2047	SP	0,000199	0,999210	0,000166	0,000175	0,000135	0,000135	CS pos: 22-23. Pr: 0,9787
XALc_2055	SP	0,000207	0,999181	0,000174	0,000175	0,000140	0,000139	CS pos: 29-30. Pr: 0,9762
XALc_2063	SP	0,000178	0,999257	0,000155	0,000165	0,000128	0,000130	CS pos: 24-25. Pr: 0,9785
XALc_2069	SP	0,000163	0,999299	0,000143	0,000157	0,000123	0,000122	CS pos: 24-25. Pr: 0,9783
XALc_2080	LIPO	0,000000	0,000000	1,000028	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9922
XALc_2085	SP	0,000175	0,999260	0,000155	0,000162	0,000129	0,000130	CS pos: 20-21. Pr: 0,9780
XALc_2087	SP	0,000956	0,665845	0,332790	0,000149	0,000142	0,000123	CS pos: 32-33. Pr: 0,6404
XALc_2093	LIPO	0,000000	0,000000	1,000038	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9965

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2110	SP	0,000198	0,999205	0,000162	0,000173	0,000138	0,000141	CS pos: 40-41. Pr: 0,9732
XALc_2112	SP	0,000238	0,999057	0,000198	0,000206	0,000159	0,000160	CS pos: 31-32. Pr: 0,9643
XALc_2114	SP	0,026902	0,667421	0,305254	0,000149	0,000144	0,000119	CS pos: 42-43. Pr: 0,6409
XALc_2115	SP	0,316678	0,681711	0,000568	0,000415	0,000356	0,000271	CS pos: 23-24. Pr: 0,6034
XALc_2116	LIPO	0,000000	0,000000	1,000040	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9968
XALc_2118	SP	0,000261	0,998995	0,000222	0,000207	0,000165	0,000156	CS pos: 23-24. Pr: 0,9802
XALc_2125	SP	0,000311	0,998617	0,000297	0,000284	0,000263	0,000251	CS pos: 22-23. Pr: 0,9714
XALc_2129	SP	0,333181	0,665996	0,000385	0,000152	0,000141	0,000124	CS pos: 48-49. Pr: 0,6449
XALc_2130	SP	0,000191	0,999245	0,000155	0,000167	0,000130	0,000127	CS pos: 25-26. Pr: 0,9770
XALc_2131	SP	0,000185	0,832600	0,166816	0,000165	0,000129	0,000133	CS pos: 25-26. Pr: 0,7897
XALc_2138	SP	0,000210	0,999144	0,000174	0,000186	0,000150	0,000150	CS pos: 22-23. Pr: 0,9786
XALc_2148	LIPO	0,000000	0,000000	1,000042	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9974
XALc_2154	SP	0,000372	0,998314	0,000377	0,000351	0,000294	0,000298	CS pos: 34-35. Pr: 0,9627
XALc_2162	SP	0,000167	0,999295	0,000140	0,000156	0,000122	0,000124	CS pos: 25-26. Pr: 0,9436
XALc_2167	LIPO	0,000000	0,000000	1,000015	0,000000	0,000000	0,000000	CS pos: 14-15. Pr: 0,9962
XALc_2174	SP	0,000556	0,998318	0,000368	0,000298	0,000245	0,000222	CS pos: 20-21. Pr: 0,9708
XALc_2175	SP	0,000197	0,999165	0,000175	0,000183	0,000147	0,000145	CS pos: 19-20. Pr: 0,9684
XALc_2177	SP	0,000185	0,999193	0,000159	0,000171	0,000133	0,000138	CS pos: 26-27. Pr: 0,9780

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2180	SP	0,334237	0,664257	0,000507	0,000395	0,000337	0,000276	CS pos: 15-16. Pr: 0,5454
XALc_2183	SP	0,000169	0,830474	0,168982	0,000155	0,000124	0,000126	CS pos: 33-34. Pr: 0,5499
XALc_2187	SP	0,000286	0,832266	0,166903	0,000229	0,000187	0,000176	CS pos: 24-25. Pr: 0,8050
XALc_2194	SP	0,000167	0,999256	0,000145	0,000160	0,000124	0,000122	CS pos: 23-24. Pr: 0,9787
XALc_2195	SP	0,000432	0,998554	0,000296	0,000298	0,000229	0,000200	CS pos: 29-30. Pr: 0,9700
XALc_2200	SP	0,000181	0,666107	0,000125	0,333364	0,000113	0,000112	CS pos: 30-31. Pr: 0,6438
XALc_2214	SP	0,000389	0,998507	0,000325	0,000297	0,000247	0,000219	CS pos: 24-25. Pr: 0,9743
XALc_2216	SP	0,000289	0,832185	0,166880	0,000249	0,000209	0,000178	CS pos: 24-25. Pr: 0,8126
XALc_2217	SP	0,000365	0,992790	0,006041	0,000314	0,000259	0,000229	CS pos: 39-40. Pr: 0,9696
XALc_2218	SP	0,000211	0,999080	0,000199	0,000191	0,000158	0,000163	CS pos: 21-22. Pr: 0,9763
XALc_2219	LIPO	0,002439	0,071322	0,926243	0,000000	0,000000	0,000000	CS pos: 43-44. Pr: 0,9870
XALc_2220	SP	0,000304	0,998804	0,000245	0,000251	0,000204	0,000188	CS pos: 34-35. Pr: 0,9627
XALc_2221	SP	0,151364	0,826699	0,021261	0,000267	0,000223	0,000192	CS pos: 49-50. Pr: 0,8213
XALc_2222	TAT	0,000000	0,000000	0,000000	1,000004	0,000000	0,000000	CS pos: 34-35. Pr: 0,6471
XALc_2223	TAT	0,000000	0,000000	0,000000	1,000013	0,000000	0,000000	CS pos: 28-29. Pr: 0,9352
XALc_2224	SP	0,000177	0,999276	0,000153	0,000165	0,000127	0,000133	CS pos: 21-22. Pr: 0,9794
XALc_2226	SP	0,000807	0,997273	0,000665	0,000458	0,000378	0,000410	CS pos: 27-28. Pr: 0,9699
XALc_2227	SP	0,000181	0,999249	0,000150	0,000170	0,000130	0,000131	CS pos: 19-20. Pr: 0,9885

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2229	TAT	0,000000	0,000000	0,000000	0,999994	0,000000	0,000000	CS pos: 29-30. Pr: 0,9004
XALc_2233	SP	0,000413	0,998369	0,000354	0,000331	0,000291	0,000242	CS pos: 21-22. Pr: 0,9751
XALc_2245	SP	0,000215	0,999160	0,000173	0,000192	0,000155	0,000147	CS pos: 28-29. Pr: 0,9764
XALc_2248	LIPO	0,000000	0,000000	1,000060	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9966
XALc_2253	SP	0,000191	0,999200	0,000165	0,000173	0,000138	0,000143	CS pos: 24-25. Pr: 0,9786
XALc_2258	SP	0,000192	0,999188	0,000166	0,000168	0,000138	0,000139	CS pos: 19-20. Pr: 0,9742
XALc_2260	SP	0,000226	0,999069	0,000194	0,000204	0,000167	0,000160	CS pos: 24-25. Pr: 0,9783
XALc_2287	LIPO	0,000000	0,000000	1,000055	0,000000	0,000000	0,000000	CS pos: 15-16. Pr: 0,9975
XALc_2291	SP	0,000257	0,998998	0,000211	0,000196	0,000163	0,000160	CS pos: 22-23. Pr: 0,9708
XALc_2295	SP	0,000418	0,995868	0,003122	0,000255	0,000174	0,000174	CS pos: 42-43. Pr: 0,9672
XALc_2302	SP	0,000201	0,999206	0,000168	0,000177	0,000136	0,000134	CS pos: 30-31. Pr: 0,9741
XALc_2307	SP	0,000257	0,998990	0,000210	0,000217	0,000169	0,000171	CS pos: 36-37. Pr: 0,9679
XALc_2311	SP	0,000211	0,836472	0,000161	0,162938	0,000120	0,000109	CS pos: 23-24. Pr: 0,8198
XALc_2314	SP	0,000194	0,999201	0,000159	0,000169	0,000135	0,000135	CS pos: 25-26. Pr: 0,9782
XALc_2315	LIPO	0,000000	0,000000	1,000049	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9976
XALc_2319	SP	0,028556	0,921762	0,048996	0,000281	0,000215	0,000210	CS pos: 47-48. Pr: 0,9281
XALc_2320	SP	0,000485	0,997978	0,000466	0,000384	0,000388	0,000327	CS pos: 26-27. Pr: 0,9735
XALc_2321	TAT	0,000000	0,000000	0,000000	0,999998	0,000000	0,000000	CS pos: 25-26. Pr: 0,5845

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2338	SP	0,000703	0,996622	0,001312	0,000503	0,000365	0,000503	CS pos: 27-28. Pr: 0,9781
XALc_2339	SP	0,000191	0,999232	0,000162	0,000165	0,000133	0,000135	CS pos: 29-30. Pr: 0,9773
XALc_2343	SP	0,000208	0,999161	0,000181	0,000183	0,000151	0,000150	CS pos: 31-32. Pr: 0,9747
XALc_2349	SP	0,001277	0,996785	0,000418	0,000479	0,000721	0,000314	CS pos: 53-54. Pr: 0,8539
XALc_2350	LIPO	0,000000	0,000000	1,000008	0,000000	0,000000	0,000000	CS pos: 22-23. Pr: 0,9968
XALc_2361	LIPO	0,000000	0,000000	1,000064	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9976
XALc_2378	SP	0,085532	0,913516	0,000280	0,000266	0,000216	0,000196	CS pos: 38-39. Pr: 0,9367
XALc_2409	SP	0,000268	0,998932	0,000220	0,000242	0,000178	0,000171	CS pos: 31-32. Pr: 0,9605
XALc_2410	SP	0,000307	0,998843	0,000254	0,000236	0,000186	0,000187	CS pos: 14-15. Pr: 0,9251
XALc_2411	SP	0,499038	0,500476	0,000125	0,000153	0,000114	0,000103	CS pos: 23-24. Pr: 0,4923
XALc_2412	LIPO	0,000000	0,000000	1,000044	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9962
XALc_2417	SP	0,000108	0,666610	0,333052	0,000080	0,000069	0,000067	CS pos: 26-27. Pr: 0,6438
XALc_2422	LIPO	0,000000	0,000000	1,000043	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9968
XALc_2423	SP	0,000381	0,998580	0,000339	0,000266	0,000233	0,000210	CS pos: 22-23. Pr: 0,9739
XALc_2438	TAT	0,000000	0,000000	0,000000	0,999987	0,000000	0,000000	CS pos: 29-30. Pr: 0,9283
XALc_2439	TAT	0,000000	0,000000	0,000000	1,000009	0,000000	0,000000	CS pos: 25-26. Pr: 0,9345
XALc_2442	SP	0,000316	0,998754	0,000274	0,000261	0,000223	0,000195	CS pos: 23-24. Pr: 0,9779
XALc_2447	LIPO	0,000029	0,000647	0,999324	0,000001	0,000000	0,000000	CS pos: 28-29. Pr: 0,9968

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2448	SP	0,001359	0,997386	0,000469	0,000325	0,000256	0,000227	CS pos: 37-38. Pr: 0,9634
XALc_2462	SP	0,000293	0,665506	0,333616	0,000221	0,000197	0,000172	CS pos: 25-26. Pr: 0,6445
XALc_2468	LIPO	0,000000	0,000000	1,000027	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9971
XALc_2473	SP	0,000238	0,999090	0,000186	0,000192	0,000160	0,000152	CS pos: 30-31. Pr: 0,9783
XALc_2479	LIPO	0,000000	0,000000	1,000030	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9976
XALc_2485	SP	0,000251	0,998984	0,000203	0,000214	0,000171	0,000166	CS pos: 27-28. Pr: 0,9729
XALc_2487	SP	0,000217	0,999168	0,000171	0,000178	0,000141	0,000138	CS pos: 23-24. Pr: 0,9790
XALc_2491	SP	0,028328	0,780065	0,191083	0,000211	0,000153	0,000167	CS pos: 36-37. Pr: 0,7464
XALc_2504	SP	0,167849	0,830302	0,000683	0,000424	0,000391	0,000332	CS pos: 25-26. Pr: 0,7330
XALc_2515	SP	0,000193	0,999181	0,000160	0,000173	0,000145	0,000144	CS pos: 23-24. Pr: 0,9783
XALc_2516	SP	0,000257	0,999011	0,000203	0,000197	0,000160	0,000159	CS pos: 24-25. Pr: 0,9762
XALc_2518	SP	0,000211	0,999058	0,000190	0,000201	0,000162	0,000163	CS pos: 25-26. Pr: 0,9733
XALc_2524	LIPO	0,000001	0,001544	0,998456	0,000001	0,000001	0,000001	CS pos: 20-21. Pr: 0,9825
XALc_2526	LIPO	0,000000	0,000000	1,000022	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9951
XALc_2527	LIPO	0,000000	0,000000	1,000015	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9967
XALc_2531	LIPO	0,000101	0,499599	0,500061	0,000079	0,000073	0,000072	CS pos: 22-23. Pr: 0,4851
XALc_2542	PILIN	0,472654	0,002878	0,000006	0,000005	0,000004	0,524463	CS pos: 7-8. Pr: 0,4148
XALc_2545	LIPO	0,000030	0,166563	0,833354	0,000020	0,000018	0,000019	CS pos: 21-22. Pr: 0,7981

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2556	SP	0,449260	0,547662	0,000646	0,001021	0,000743	0,000701	CS pos: 14-15. Pr: 0,3775
XALc_2561	SP	0,000110	0,832836	0,166767	0,000102	0,000080	0,000079	CS pos: 20-21. Pr: 0,8183
XALc_2565	LIPO	0,000000	0,000000	1,000022	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9969
XALc_2568	LIPO	0,000000	0,000014	1,000007	0,000000	0,000000	0,000000	CS pos: 24-25. Pr: 0,9961
XALc_2572	LIPO	0,000078	0,499694	0,500057	0,000064	0,000054	0,000053	CS pos: 19-20. Pr: 0,4771
XALc_2580	SP	0,000620	0,998651	0,000231	0,000195	0,000151	0,000154	CS pos: 21-22. Pr: 0,9796
XALc_2582	SP	0,000195	0,832522	0,166835	0,000172	0,000144	0,000143	CS pos: 24-25. Pr: 0,8116
XALc_2585	SP	0,000198	0,999138	0,000169	0,000186	0,000145	0,000150	CS pos: 23-24. Pr: 0,9752
XALc_2597	SP	0,000208	0,999186	0,000164	0,000175	0,000142	0,000137	CS pos: 20-21. Pr: 0,9783
XALc_2641	LIPO	0,000000	0,000000	1,000019	0,000000	0,000000	0,000000	CS pos: 16-17. Pr: 0,9970
XALc_2643	SP	0,000293	0,998923	0,000229	0,000232	0,000180	0,000178	CS pos: 24-25. Pr: 0,9718
XALc_2654	LIPO	0,000000	0,000000	1,000059	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9962
XALc_2659	PILIN	0,010979	0,000121	0,000000	0,000001	0,000002	0,988881	CS pos: 12-13. Pr: 0,6055
XALc_2660	PILIN	0,015766	0,000000	0,000000	0,000000	0,000000	0,984211	CS pos: 6-7. Pr: 0,7013
XALc_2661	PILIN	0,000000	0,000001	0,000000	0,000000	0,000000	1,000009	CS pos: 8-9. Pr: 0,7675
XALc_2662	PILIN	0,022240	0,049765	0,000002	0,000026	0,000061	0,927894	CS pos: 18-19. Pr: 0,6392
XALc_2665	SP	0,000253	0,998948	0,000213	0,000238	0,000184	0,000182	CS pos: 28-29. Pr: 0,9655
XALc_2668	LIPO	0,000015	0,042092	0,957898	0,000009	0,000008	0,000009	CS pos: 17-18. Pr: 0,9514

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2679	SP	0,000407	0,832252	0,166826	0,000216	0,000183	0,000170	CS pos: 36-37. Pr: 0,7994
XALc_2686	LIPO	0,000000	0,000000	1,000074	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9970
XALc_2715	SP	0,000217	0,999142	0,000186	0,000191	0,000145	0,000146	CS pos: 25-26. Pr: 0,9759
XALc_2716	SP	0,000217	0,999091	0,000179	0,000198	0,000151	0,000156	CS pos: 15-16. Pr: 0,9578
XALc_2726	SP	0,000341	0,998632	0,000295	0,000284	0,000246	0,000210	CS pos: 22-23. Pr: 0,9758
XALc_2740	LIPO	0,000000	0,000000	1,000012	0,000000	0,000000	0,000000	CS pos: 19-20. Pr: 0,9964
XALc_2746	SP	0,000233	0,999070	0,000210	0,000194	0,000155	0,000152	CS pos: 30-31. Pr: 0,9659
XALc_2749	TAT	0,000000	0,000000	0,000000	0,999992	0,000000	0,000000	CS pos: 30-31. Pr: 0,9388
XALc_2750	SP	0,000210	0,999148	0,000183	0,000182	0,000147	0,000142	CS pos: 19-20. Pr: 0,9772
XALc_2752	SP	0,000272	0,998937	0,000223	0,000216	0,000181	0,000164	CS pos: 22-23. Pr: 0,9730
XALc_2761	SP	0,000184	0,832569	0,166864	0,000152	0,000124	0,000131	CS pos: 22-23. Pr: 0,8061
XALc_2762	SP	0,000158	0,832619	0,166810	0,000150	0,000128	0,000121	CS pos: 22-23. Pr: 0,6495
XALc_2774	LIPO	0,000132	0,499559	0,500112	0,000084	0,000081	0,000078	CS pos: 18-19. Pr: 0,4963
XALc_2777	SP	0,000182	0,999177	0,000163	0,000180	0,000143	0,000145	CS pos: 20-21. Pr: 0,9781
XALc_2778	SP	0,000328	0,998690	0,000282	0,000281	0,000233	0,000206	CS pos: 21-22. Pr: 0,9726
XALc_2784	LIPO	0,000000	0,000002	0,999995	0,000000	0,000000	0,000000	CS pos: 23-24. Pr: 0,9953
XALc_2788	SP	0,000192	0,999163	0,000174	0,000182	0,000141	0,000154	CS pos: 18-19. Pr: 0,9721
XALc_2798	SP	0,000189	0,999210	0,000163	0,000177	0,000147	0,000149	CS pos: 31-32. Pr: 0,9710

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2825	LIPO	0,000000	0,000000	1,000050	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9962
XALc_2834	SP	0,000407	0,998311	0,000329	0,000363	0,000287	0,000307	CS pos: 22-23. Pr: 0,9700
XALc_2844	SP	0,000154	0,666089	0,333464	0,000105	0,000097	0,000089	CS pos: 26-27. Pr: 0,6467
XALc_2859	SP	0,000450	0,998494	0,000307	0,000293	0,000250	0,000202	CS pos: 26-27. Pr: 0,9653
XALc_2866	SP	0,000230	0,999020	0,000218	0,000198	0,000167	0,000170	CS pos: 15-16. Pr: 0,9618
XALc_2867	SP	0,000235	0,999085	0,000205	0,000186	0,000149	0,000152	CS pos: 24-25. Pr: 0,9774
XALc_2869	LIPO	0,000000	0,000000	1,000026	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9947
XALc_2904	LIPO	0,000307	0,165249	0,833648	0,000250	0,000239	0,000302	CS pos: 24-25. Pr: 0,8234
XALc_2918	SP	0,002065	0,991554	0,003169	0,001208	0,000875	0,001114	CS pos: 23-24. Pr: 0,9763
XALc_2919	LIPO	0,000000	0,000000	1,000026	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9963
XALc_2928	LIPO	0,000000	0,000002	1,000006	0,000000	0,000000	0,000000	CS pos: 23-24. Pr: 0,9969
XALc_2933	SP	0,000119	0,666221	0,333432	0,000083	0,000075	0,000073	CS pos: 22-23. Pr: 0,6478
XALc_2946	SP	0,000897	0,995402	0,002311	0,000581	0,000400	0,000431	CS pos: 27-28. Pr: 0,9768
XALc_2949	SP	0,000240	0,998957	0,000263	0,000219	0,000167	0,000174	CS pos: 26-27. Pr: 0,9781
XALc_2962	SP	0,000441	0,998104	0,000453	0,000379	0,000319	0,000329	CS pos: 20-21. Pr: 0,9639
XALc_2967	SP	0,000191	0,999201	0,000163	0,000173	0,000138	0,000142	CS pos: 27-28. Pr: 0,9765
XALc_2969	SP	0,000196	0,999185	0,000167	0,000183	0,000148	0,000155	CS pos: 27-28. Pr: 0,9774
XALc_2972	SP	0,166727	0,666180	0,166782	0,000123	0,000094	0,000097	CS pos: 26-27. Pr: 0,6470

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_2973	SP	0,000392	0,998846	0,000242	0,000219	0,000170	0,000161	CS pos: 20-21. Pr: 0,9729
XALc_2978	SP	0,002308	0,996829	0,000251	0,000240	0,000186	0,000193	CS pos: 51-52. Pr: 0,9204
XALc_2980	LIPO	0,000108	0,499620	0,500079	0,000078	0,000068	0,000071	CS pos: 13-14. Pr: 0,4895
XALc_2983	TAT	0,000000	0,000000	0,000000	0,833751	0,166253	0,000000	CS pos: 35-36. Pr: 0,7701
XALc_2990	SP	0,000317	0,998952	0,000214	0,000201	0,000166	0,000152	CS pos: 19-20. Pr: 0,9771
XALc_2991	SP	0,000166	0,999268	0,000143	0,000155	0,000128	0,000128	CS pos: 23-24. Pr: 0,9359
XALc_2994	SP	0,000305	0,998947	0,000203	0,000219	0,000169	0,000156	CS pos: 22-23. Pr: 0,9769
XALc_2996	SP	0,000174	0,999250	0,000147	0,000157	0,000128	0,000123	CS pos: 22-23. Pr: 0,7645
XALc_2999	LIPO	0,000000	0,000000	1,000038	0,000000	0,000000	0,000000	CS pos: 17-18. Pr: 0,9975
XALc_3000	SP	0,000206	0,999146	0,000174	0,000184	0,000142	0,000138	CS pos: 21-22. Pr: 0,9773
XALc_3003	LIPO	0,000013	0,000666	0,999320	0,000003	0,000000	0,000000	CS pos: 24-25. Pr: 0,9968
XALc_3006	LIPO	0,060321	0,088809	0,850842	0,000004	0,000004	0,000002	CS pos: 41-42. Pr: 0,9774
XALc_3018	SP	0,000657	0,998393	0,000296	0,000267	0,000193	0,000191	CS pos: 26-27. Pr: 0,9738
XALc_3025	SP	0,000303	0,998690	0,000280	0,000275	0,000241	0,000217	CS pos: 19-20. Pr: 0,9718
XALc_3037	SP	0,000181	0,999257	0,000147	0,000158	0,000125	0,000121	CS pos: 17-18. Pr: 0,9731
XALc_3041	SP	0,000210	0,999129	0,000179	0,000194	0,000155	0,000150	CS pos: 34-35. Pr: 0,9685
XALc_3064	SP	0,000214	0,999145	0,000183	0,000194	0,000152	0,000148	CS pos: 30-31. Pr: 0,9702
XALc_3071	SP	0,000174	0,999251	0,000147	0,000170	0,000131	0,000141	CS pos: 20-21. Pr: 0,9806

<b>Proteína</b>	<b>Peptídeo sinal</b>	<b>NO_SP</b>	<b>SP</b>	<b>LIPO</b>	<b>TAT</b>	<b>TATLIPO</b>	<b>PILIN</b>	<b>Posição de clivagem (Pr: Probabilidade)</b>
XALc_3074	LIPO	0,000000	0,000000	1,000010	0,000000	0,000000	0,000000	CS pos: 15-16. Pr: 0,9976
XALc_3077	SP	0,022669	0,974351	0,000936	0,000747	0,000630	0,000665	CS pos: 20-21. Pr: 0,9546
XALc_3081	LIPO	0,000116	0,333087	0,666716	0,000032	0,000030	0,000024	CS pos: 15-16. Pr: 0,6616
XALc_3088	LIPO	0,000000	0,000000	1,000027	0,000000	0,000000	0,000000	CS pos: 20-21. Pr: 0,9942
XALc_3091	SP	0,000270	0,998947	0,000214	0,000230	0,000180	0,000171	CS pos: 35-36. Pr: 0,9665
XALc_3100	LIPO	0,000000	0,000000	1,000038	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9968
XALc_3104	SP	0,000141	0,832778	0,166786	0,000105	0,000092	0,000092	CS pos: 29-30. Pr: 0,8157
XALc_3105	SP	0,000313	0,998701	0,000318	0,000250	0,000211	0,000208	CS pos: 22-23. Pr: 0,9639
XALc_3109	LIPO	0,000028	0,000985	0,998967	0,000001	0,000000	0,000000	CS pos: 27-28. Pr: 0,9932
XALc_3122	SP	0,000123	0,832850	0,166771	0,000096	0,000083	0,000087	CS pos: 17-18. Pr: 0,8220
XALc_3133	SP	0,000196	0,999193	0,000165	0,000178	0,000138	0,000144	CS pos: 28-29. Pr: 0,9767
XALc_3135	SP	0,000137	0,832722	0,166789	0,000130	0,000112	0,000107	CS pos: 22-23. Pr: 0,8194
XALc_3141	LIPO	0,001965	0,062367	0,935672	0,000000	0,000001	0,000000	CS pos: 40-41. Pr: 0,9952
XALc_3142	LIPO	0,000000	0,000000	1,000017	0,000000	0,000000	0,000000	CS pos: 21-22. Pr: 0,9967
XALc_3143	SP	0,000135	0,666198	0,333447	0,000086	0,000083	0,000076	CS pos: 22-23. Pr: 0,6498
XALc_3147	SP	0,000443	0,998432	0,000307	0,000334	0,000261	0,000212	CS pos: 28-29. Pr: 0,9722
XALc_3155	LIPO	0,000000	0,000000	1,000031	0,000000	0,000000	0,000000	CS pos: 18-19. Pr: 0,9951
XALc_3157	LIPO	0,000039	0,166550	0,833373	0,000020	0,000017	0,000017	CS pos: 19-20. Pr: 0,8245

Proteína	Peptídeo sinal	NO_SP	SP	LIPO	TAT	TATLIPO	PILIN	Posição de clivagem (Pr: Probabilidade)
XALc_3161	TAT	0,000000	0,000000	0,000000	1,000008	0,000000	0,000000	CS pos: 31-32. Pr: 0,7783
XALc_3162	SP	0,000229	0,999120	0,000183	0,000196	0,000144	0,000144	CS pos: 28-29. Pr: 0,9751
XALc_3165	SP	0,000303	0,998844	0,000238	0,000240	0,000196	0,000185	CS pos: 39-40. Pr: 0,9723
XALc_3100,7	SP	0,000324	0,998927	0,000198	0,000211	0,000172	0,000176	CS pos: 46-47. Pr: 0,9460

Fonte: Elaborada pelo autor.

**Tabela A2** - Lista das 109 proteínas do proteoma de *Xanthomonas albilineans*, identificados pelo software CELLO como tendo localização extracelular. A tabela apresenta a probabilidade de localização em cinco regiões principais: extracelular, membrana interna ou externa, periplasmática e citoplasmática.

ID	Localização	Extracelular	Membrana externa	Periplasmática	Membrana interna	Citoplasmática
XALr_3254	Extracelular	1,836134	1,147810	1,517992	0,059261	0,438803
XALr_3258	Extracelular	1,940080	0,932245	1,610834	0,377371	0,139470
XALq_3225	Extracelular	1,707332	0,706616	1,685293	0,268538	0,632221
XALc_0023	Extracelular	2,542280	0,573220	1,238047	0,177857	0,468595
XALc_0056	Extracelular	2,515273	0,553882	1,469353	0,192843	0,268647
XALc_0090	Extracelular	2,562428	2,163358	0,252123	0,009884	0,012206
XALc_0138	Extracelular	2,385072	0,504855	1,819769	0,264785	0,025521
XALc_0168	Extracelular	1,750129	0,993666	1,735870	0,444713	0,075621
XALc_0189	Extracelular	2,020214	0,403211	1,810750	0,245890	0,519935
XALc_0211	Extracelular	1,857222	0,417831	0,694779	0,537122	1,493047

<b>ID</b>	<b>Localização</b>	<b>Extracelular</b>	<b>Membrana externa</b>	<b>Periplasmática</b>	<b>Membrana interna</b>	<b>Citoplasmática</b>
XALc_0225	Extracelular	3,671816	0,720714	0,353210	0,121843	0,132419
XALc_0228	Extracelular	2,048572	1,814819	0,597960	0,129703	0,408944
XALc_0230	Extracelular	2,149452	1,503423	1,178127	0,071674	0,097324
XALc_0259	Extracelular	3,788406	0,320475	0,633763	0,139056	0,118300
XALc_0391	Extracelular	2,844112	1,128300	0,395305	0,123558	0,508725
XALc_0456	Extracelular	1,990161	0,100844	1,608843	0,320088	0,980065
XALc_0484	Extracelular	2,162542	1,102000	0,552870	0,789628	0,392963
XALc_0551	Extracelular	2,558972	0,912485	1,220600	0,081390	0,226554
XALc_0636	Extracelular	2,393499	0,989715	0,943625	0,388065	0,285097
XALc_0666	Extracelular	2,201000	0,448038	1,170397	0,555254	0,625313
XALc_0676	Extracelular	2,671188	1,174265	0,921672	0,139099	0,093777
XALc_0727	Extracelular	3,000903	0,634840	1,050031	0,091548	0,222677
XALc_0802	Extracelular	3,960687	0,141305	0,485072	0,089852	0,323084
XALc_0811	Extracelular	1,664915	1,210236	1,477305	0,240353	0,407189
XALc_0865	Extracelular	1,647322	1,156747	0,656141	0,853453	0,686339
XALc_0874	Extracelular	3,804701	0,510729	0,212914	0,371483	0,100173
XALc_0883	Extracelular	2,422398	0,873470	1,392242	0,169695	0,142196
XALc_0927	Extracelular	3,443379	0,517196	0,722120	0,135920	0,181385
XALc_1042	Extracelular	1,998806	0,692185	1,445471	0,237882	0,625658

<b>ID</b>	<b>Localização</b>	<b>Extracelular</b>	<b>Membrana externa</b>	<b>Periplasmática</b>	<b>Membrana interna</b>	<b>Citoplasmática</b>
XALc_1123	Extracelular	1,955274	0,383744	1,454902	0,344771	0,861309
XALc_1132	Extracelular	2,324487	0,536677	1,284199	0,272672	0,581966
XALc_1160	Extracelular	1,841333	1,228707	1,429286	0,275466	0,225208
XALc_1234	Extracelular	1,691377	0,571456	0,827502	0,394356	1,515309
XALc_1245	Extracelular	2,271421	0,738366	1,291485	0,456178	0,242551
XALc_1263	Extracelular	1,410240	0,971870	1,328295	0,670019	0,619577
XALc_1281	Extracelular	1,583352	1,266632	0,843658	0,802642	0,503716
XALc_1282	Extracelular	2,257994	0,490239	1,298700	0,644978	0,308089
XALc_1283	Extracelular	2,174641	1,522732	0,626357	0,598382	0,077890
XALc_1284	Extracelular	2,663394	1,309744	0,666512	0,212046	0,148302
XALc_1285	Extracelular	2,930259	1,400598	0,109037	0,465915	0,094191
XALc_1301	Extracelular	2,215591	0,679427	1,454513	0,185041	0,465428
XALc_1305	Extracelular	2,956251	0,827913	0,309971	0,659957	0,245907
XALc_1391	Extracelular	2,310126	0,669257	1,756761	0,117821	0,146035
XALc_1397	Extracelular	1,715565	0,937523	1,258330	0,323116	0,765464
XALc_1400	Extracelular	3,019638	0,183659	1,570993	0,119145	0,106565
XALc_1416	Extracelular	4,574725	0,260001	0,109284	0,034635	0,021355
XALc_1417	Extracelular	4,492957	0,298244	0,114052	0,062440	0,032306
XALc_1418	Extracelular	4,248913	0,619214	0,118567	0,009779	0,003527

<b>ID</b>	<b>Localização</b>	<b>Extracelular</b>	<b>Membrana externa</b>	<b>Periplasmática</b>	<b>Membrana interna</b>	<b>Citoplasmática</b>
XALc_1419	Extracelular	3,975909	0,866237	0,092024	0,050556	0,015273
XALc_1420	Extracelular	1,883206	0,867720	1,544591	0,081757	0,622725
XALc_1423	Extracelular	2,870582	1,066421	0,961001	0,069667	0,032330
XALc_1425	Extracelular	4,278957	0,584826	0,108746	0,016701	0,010770
XALc_1426	Extracelular	3,700458	0,444123	0,767551	0,041787	0,046081
XALc_1428	Extracelular	1,621583	1,211368	0,917997	0,069363	1,179689
XALc_1431	Extracelular	1,400781	0,821098	0,858591	0,523414	1,396117
XALc_1466	Extracelular	1,610698	1,520812	1,269909	0,397301	0,201281
XALc_1477	Extracelular	2,844242	1,865633	0,206431	0,024788	0,058906
XALc_1488	Extracelular	1,731863	0,517540	1,135791	0,381985	1,232821
XALc_1539	Extracelular	1,956691	0,459542	1,678185	0,245476	0,660106
XALc_1569	Extracelular	1,744988	0,726917	1,696310	0,484895	0,346889
XALc_1661	Extracelular	2,593167	1,106973	0,841786	0,093619	0,364457
XALc_1842	Extracelular	2,823788	0,906049	0,876863	0,340249	0,053051
XALc_1843	Extracelular	1,969861	1,283236	1,058143	0,484440	0,204322
XALc_1855	Extracelular	1,738666	0,526762	1,507467	0,222899	1,004208
XALc_1884	Extracelular	2,969993	0,729912	0,329921	0,708964	0,261211
XALc_1909	Extracelular	3,633967	0,206178	0,616366	0,183472	0,360017
XALc_1915	Extracelular	3,940324	0,819694	0,210021	0,011604	0,018357

<b>ID</b>	<b>Localização</b>	<b>Extracelular</b>	<b>Membrana externa</b>	<b>Periplasmática</b>	<b>Membrana interna</b>	<b>Citoplasmática</b>
XALc_1916	Extracelular	4,547467	0,347593	0,077868	0,019303	0,007769
XALc_1937	Extracelular	1,995587	0,577882	1,496191	0,232760	0,697579
XALc_1956	Extracelular	2,471701	0,903862	1,068830	0,114292	0,441316
XALc_2002	Extracelular	2,548402	0,391727	1,562211	0,177517	0,320144
XALc_2021	Extracelular	4,211279	0,242939	0,258269	0,146582	0,140932
XALc_2023	Extracelular	2,657660	1,980686	0,220708	0,090029	0,050918
XALc_2113	Extracelular	2,594518	1,319469	0,296985	0,626174	0,162855
XALc_2114	Extracelular	2,254909	1,698459	0,305668	0,602535	0,138428
XALc_2131	Extracelular	2,086921	1,026625	1,726542	0,097263	0,062647
XALc_2162	Extracelular	4,877539	0,091322	0,025019	0,003522	0,002598
XALc_2183	Extracelular	2,806786	1,704061	0,396757	0,020761	0,071635
XALc_2194	Extracelular	1,784567	1,681651	1,196267	0,215247	0,122269
XALc_2203	Extracelular	1,709668	0,500599	1,237720	0,378105	1,173906
XALc_2258	Extracelular	2,278220	0,842315	1,532008	0,186200	0,161257
XALc_2278	Extracelular	3,272175	0,993022	0,161482	0,458614	0,114706
XALc_2281	Extracelular	3,152651	0,985188	0,157658	0,572414	0,132088
XALc_2282	Extracelular	3,386200	0,903155	0,241514	0,332433	0,136698
XALc_2285	Extracelular	1,524426	1,506492	1,214645	0,167732	0,586707
XALc_2288	Extracelular	2,804440	1,355992	0,116465	0,590495	0,132607

<b>ID</b>	<b>Localização</b>	<b>Extracelular</b>	<b>Membrana externa</b>	<b>Periplasmática</b>	<b>Membrana interna</b>	<b>Citoplasmática</b>
XALc_2290	Extracelular	2,769120	1,176669	0,283044	0,613657	0,157509
XALc_2295	Extracelular	2,572928	1,482814	0,901303	0,015720	0,027235
XALc_2314	Extracelular	2,316397	1,300854	0,992127	0,163560	0,227062
XALc_2339	Extracelular	2,983715	0,949337	0,630142	0,234275	0,202532
XALc_2402	Extracelular	3,085189	0,512638	0,512218	0,170847	0,719108
XALc_2404	Extracelular	2,065844	0,269762	1,632769	0,215779	0,815846
XALc_2409	Extracelular	2,023485	1,949674	0,295716	0,596092	0,135034
XALc_2410	Extracelular	2,611147	1,387173	0,249865	0,608172	0,143644
XALc_2524	Extracelular	2,494091	1,386418	0,450417	0,546174	0,122900
XALc_2531	Extracelular	3,687271	0,233855	0,904678	0,138707	0,035490
XALc_2665	Extracelular	2,436750	1,162542	1,263535	0,083188	0,053987
XALc_2666	Extracelular	2,593211	1,024074	0,357734	0,725074	0,299905
XALc_2794	Extracelular	2,529449	1,494678	0,394586	0,219489	0,361799
XALc_2798	Extracelular	1,673904	1,389496	1,325809	0,214821	0,395971
XALc_2918	Extracelular	2,456346	0,754504	1,643150	0,060991	0,085009
XALc_2930	Extracelular	2,806999	0,713686	0,807702	0,286649	0,384965
XALc_2967	Extracelular	3,834152	0,312919	0,820820	0,013760	0,018350
XALc_2969	Extracelular	2,622882	1,113851	0,510846	0,534446	0,217974
XALc_2978	Extracelular	2,629135	0,476416	1,421111	0,316225	0,157113

ID	Localização	Extracelular	Membrana externa	Periplasmática	Membrana interna	Citoplasmática
XALc_2987	Extracelular	2,213241	0,383864	0,713718	0,103993	1,585184
XALc_2999	Extracelular	1,814180	0,352543	1,450296	0,608922	0,774061
XALc_3071	Extracelular	2,191190	0,385293	1,813509	0,199388	0,410619
XALc_3088	Extracelular	2,180081	0,202889	1,458084	0,555863	0,603081

**Fonte:** Elaborada pelo autor.

**Tabela A3** - Possíveis CAZymes identificadas no servidor dbCAN3 utilizando os softwares HMMER e DIAMOND para realizar busca nos bancos de dados CAZy e dbCAN\_sub.

ID	EC	HMMER	dbCAN_sub	DIAMOND
XALc_0036	3.2.1.8	GH10(33-321)	GH10_e0	GH10
XALc_0039	-	-	-	GH43_1
XALc_0040	3.2.1.37 3.2.1.55	GH43_1(20-329)	GH43_e53	GH43_1
XALc_0058	3.2.1.37 3.2.1.55 3.2.1.21 3.2.1.6 3.2.1.73	GH3(67-317)	GH3_e0	GH3
XALc_0059	-	GH43_12(38-329)+CBM91(365-544)	GH43_e318+CBM91_e1	GH43_12
XALc_0060	-	-	-	GH43_12
XALc_0061	-	CE20(117-229)+CE20(426-548)	CE20_e33	CE20
XALc_0062	3.2.1.139 3.2.1.131 3.2.1.-	GH67(90-756)	GH67_e0	GH67
XALc_0083	2.4.1.321	GH94(73-794)	GH94_e4	GH94
XALc_0092	-	-	-	GH1
XALc_0107	-	GT2(7-168)	GT2	GT2

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_0140	-	GT2(13-134)	GT2	GT2
XALc_0307	-	-	-	GH13_23
XALc_0319	-	-	-	GH13
XALc_0364	-	-	-	GT2
XALc_0365	-	GT1(97-424)	GT1_e33	GT1
XALc_0408	-	-	-	GH3
XALc_0419	3.2.1.74	GH3(78-299)	GH3_e3	GH3
XALc_0441	-	-	-	GH36
XALc_0453	-	-	-	CE8
XALc_0462	-	-	-	GT2
XALc_0479	-	AA6(4-193)	AA6_e2	-
XALc_0484	3.2.1.91 3.2.1.4 3.2.1.8 -	GH6(60-406)	GH6_e0+CBM2_e84	CBM2+GH6
XALc_0485	2.4.1.255	GT41(7-562)	GT41_e81	GT41
XALc_0488	-	-	-	CBM63+GH 5
XALc_0508	-	-	-	GH1
XALc_0511	-	-	-	GH38
XALc_0567	-	-	-	GH103
XALc_0571	-	GT4(197-340)	GT4_e1620	GT4

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_0595	-	GH23(96-245)	GH23_e878+CBM50_e541	CBM50+GH 23
XALc_0603	-	GT9(72-298)	GT9_e167	GT9
XALc_0620	-	GT83(8-380)	GT83_e87	-
XALc_0658	-	-	-	CBM32
XALc_0697	2.4.1.15	GT20(3-454)	GT20_e1	GT20
XALc_0698	-	GH15(220-582)	GH15_e57	GH15
XALc_0712	-	-	-	GT4
XALc_0730	-	GH13_4(104-502)	GH13_e80	GH13_4
XALc_0743	-	GT4(198-330)	GT4_e505	GT4
XALc_0746	2.4.99.-	GT30(41-219)	GT30_e0	GT30
XALc_0772	-	-	-	GT1
XALc_0808	-	-	-	GH103
XALc_0811	3.2.1.15	GH28(95-422)	GH28_e73	GH28
XALc_0812	-	GH103(68-368)	GH103_e12	GH103
XALc_0836	-	CBM63(93-168)	CBM63_e7	CBM63
XALc_0841	-	CE9(7-378)	CE9_e59	CE9
XALc_0843	-	GH23(518-647)	GH23_e765	GH23
XALc_0863	-	CE1(21-272)	-	-

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_0865	-	GH12(106-255)	GH12_e0+CBM2_e84	CBM2+GH1 2
XALc_0866	-	PL4_1(25-522)	PL4_e0	PL4_1
XALc_0869	-	CE4(79-197)	CE4_e139	CE4
XALc_0874	3.2.1.4	CBM2(591-666)	CBM2_e21	CBM2
XALc_0918	-	-	-	GT4
XALc_0946	3.2.1.55	GH51(43-529)	GH51_e19	GH51
XALc_0959	-	-	AA2_e1	AA0
XALc_0975	-	GH23(205-336)	GH23_e543+CBM50_e851	CBM50+GH 23
XALc_1016	-	-	-	GT2
XALc_1020	-	-	-	GH13
XALc_1056	-	-	-	GT1
XALc_1057	-	-	-	GT1
XALc_1058	-	-	-	GT1
XALc_1071	-	-	-	GH28
XALc_1117	3.2.1.21 3.2.1.74	GH3(317-529)	GH3_e92	GH3
XALc_1144	-	GT1(12-419)	GT1_e33	GT1
XALc_1145	-	-	-	GT1
XALc_1151	-	-	-	GH13_31

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_1160	-	-	CBM50_e36	CBM50
XALc_1193	-	CE20(107-223)+CE20(399-530)	CE20_e33	CE20
XALc_1194	3.2.1.177	GH31(386-843)	GH31_e87	GH31
XALc_1209	3.2.1.21 3.2.1.74 3.2.1.- 3.2.1.6 3.2.1.155	GH3(149-369)	GH3_e227	GH3
XALc_1225	-	-	-	GH28
XALc_1263	-	GH23(201-303)	GH23_e893	GH23
XALc_1314	-	GH144(58-489)	GH144_e10	GH144
XALc_1318	-	-	CBM32_e120	CBM32+GH 0
XALc_1323	-	-	-	GH23
XALc_1324	-	-	-	GH23
XALc_1327	-	AA4(37-240)	-	-
XALc_1398	-	GT2(8-136)+GT2(582-750)	GT2+GT2	GT0+GT2
XALc_1420	-	GH73(232-365)	GH73_e204	GH73
XALc_1421	-	-	-	GH73
XALc_1445	-	-	-	GT4
XALc_1450	-	-	-	GT2
XALc_1474	-	-	-	GH28
XALc_1482	-	GH23(20-125)	GH23_e667	-
XALc_1523	-	-	-	GT1

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_1529	-	-	-	GT1
XALc_1550	-	-	-	GT1
XALc_1551	-	-	-	GT1
XALc_1624	-	-	-	GH28
XALc_1668	-	-	-	GT2
XALc_1750	-	GH108(8-97)	GH108_e2	-
XALc_1766	-	-	-	GH39
XALc_1767	-	-	-	GH39
XALc_1838	-	-	GH23_e876	GH23
XALc_1879	-	-	-	GH28
XALc_1915	-	CE8(77-392)	CE8_e2	CE8
XALc_1916	3.2.1.15	GH28(68-448)	GH28_e0	GH28
XALc_1923	4.2.2.n1	GH103(52-345)	GH103_e1	GH103
XALc_1934	3.2.1.4 3.2.1.- 3.2.1.151	GH9(121-569)	GH9_e26	GH9
XALc_1945	-	-	-	GH152
XALc_1947	3.2.1.22 2.4.1.-	GH27(263-501)	GH27_e14	GH27
XALc_1948	-	GH146(69-586)	GH146_e1	GH146
XALc_1950	-	-	-	GH43_29
XALc_1951	-	GH43_29(39-333)	GH43_e237	GH43_29

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_1952	3.2.1.187	GH121(13-1424)	GH121_e0	GH121
XALc_1957	-	-	-	GH43_29
XALc_2007	-	-	-	GT1
XALc_2038	2.4.1.182	GT19(42-402)	GT19_e17	GT19
XALc_2040	-	-	-	GH13_30
XALc_2148	-	CE1(39-242)	-	-
XALc_2183	3.2.1.14 1.14.99.53	GH18(37-422)	GH18_e285	GH18
XALc_2186	-	-	-	GT2
XALc_2188	-	-	-	GH1
XALc_2191	-	GT2(18-178)	GT2	GT2
XALc_2192	-	GT51(74-232)	GT51_e99	GT51
XALc_2199	-	AA3_2(2-524)	AA3_e57	AA3_2
XALc_2217	3.2.1.51	GH29(78-446)	GH29_e0	GH29
XALc_2218	3.2.1.14 3.2.1.-	GH18(54-338)	GH18_e3	GH18
XALc_2219	3.2.1.52	GH20(207-551)	GH20_e51	GH20
XALc_2220	3.2.1.25	GH2(38-759)	GH2_e75	GH2
XALc_2221	3.2.1.37 3.2.1.55 3.2.1.21 3.2.1.6 3.2.1.73	GH3(78-320)	GH3_e0	GH3
XALc_2222	3.2.1.-	GH125(68-470)	GH125_e1	GH125
XALc_2223	3.2.1.- 3.2.1.113 3.2.1.114 3.2.1.24	GH92(269-764)	GH92_e0	GH92

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_2224	3.2.1.23	GH35(39-354)	GH35_e43	GH35
XALc_2226	-	GH29(27-325)	GH29_e44+CBM32_e126	GH29
XALc_2227	-	GH33(40-330)	GH33_e86	GH33
XALc_2249	-	GT51(174-348)	GT51_e14	GT51
XALc_2250	-	GT2(74-192)	GT2+GT4_e3741	GT2+GT4
XALc_2260	-	CE16(33-297)	-	-
XALc_2293	-	CE4(65-206)	CE4_e233	-
XALc_2294	-	GT4(181-331)	GT4_e504	GT4
XALc_2321	1.1.3.-	AA7(110-292)	AA7_e0	-
XALc_2354	-	GT51(58-236)	GT51_e145	GT51
XALc_2374	-	-	-	GH28
XALc_2435	-	-	-	GT83
XALc_2438	-	-	-	CBM32
XALc_2480	-	-	-	GH13_26
XALc_2514	-	CE3(34-176)	-	-
XALc_2534	-	GH3(53-277)	GH3_e187	GH3
XALc_2544	-	-	-	GH28
XALc_2578	-	-	-	GH65
XALc_2589	-	-	GT2	GT2

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_2591	-	GT5(19-479)	GT5_e0	GT5
XALc_2592	3.2.1.141 3.2.1.1 -	GH13_10(136-435)	CBM48_e0+GH13_e120	CBM48+GH13_10
XALc_2593	-	GH77(306-638)	GH77_e33	GH77
XALc_2594	5.4.99.15	GH13_26(24-334)	GH13_e105	GH13_26
XALc_2596	3.2.1.68 2.4.1.25 3.2.1.-	CBM48(19-109)+GH13_11(189-537)	CBM48_e3+GH13_e48	CBM48+GH13_11
XALc_2631	-	-	-	GH28
XALc_2649	2.4.1.18	CBM48(122-209)+GH13_9(276-576)	CBM48_e2+GH13_e200	CBM48+GH13_9
XALc_2652	-	GT2(11-148)	GT2	GT2
XALc_2653	-	-	GT2	GT0
XALc_2675	-	-	-	GH5_11
XALc_2680	-	-	-	CBM50+GH25
XALc_2688	-	-	-	GT87
XALc_2689	-	-	-	GT2+GT4
XALc_2693	-	-	-	GT4
XALc_2697	-	-	-	GT2
XALc_2701	-	GT2(9-135)	GT2	GT2
XALc_2705	-	GT4(225-368)+GT4(661-801)	GT4_e1614+GT4_e1614	GT4

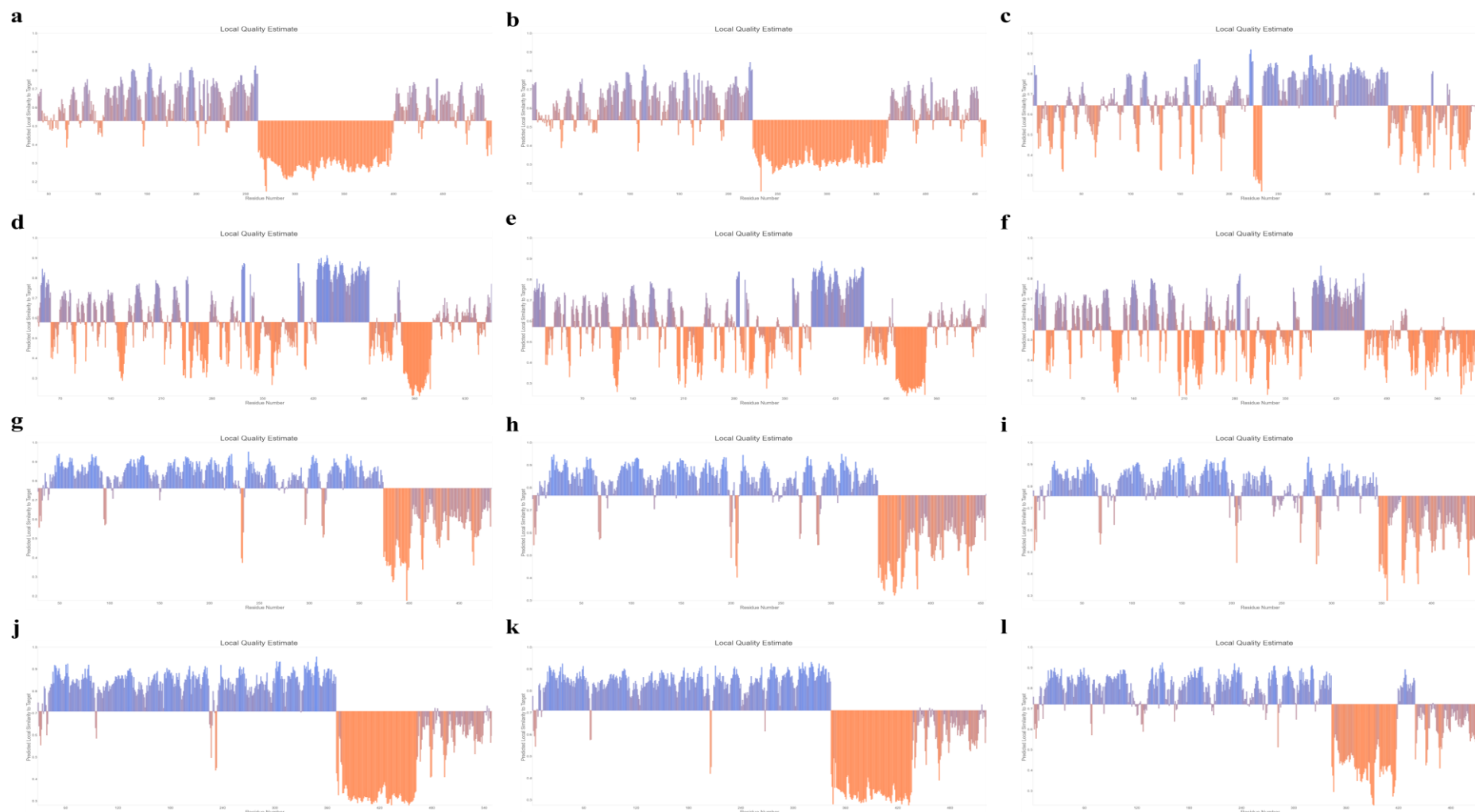
<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_2711	-	-	GT4_e3001	GT4
XALc_2749	-	GH109(40-197)	-	-
XALc_2784	3.2.1.8	GH98(73-391)+CBM35(679-783)	GH98_e0+CBM35_e1	CBM35+GH 98
XALc_2808	3.5.1.108	CE11(4-275)	CE11_e22	CE11
XALc_2813	-	-	-	GT28
XALc_2814	2.4.1.227	GT28(194-352)	GT28_e46	GT28
XALc_2843	-	-	-	GH13_11
XALc_2852	-	-	GT2	GT2
XALc_2854	-	-	GT4_e1723	GT4
XALc_2855	-	GT83(4-542)	GT83_e10	GT83
XALc_2859	-	-	CBM50_e1035	CBM50
XALc_2873	-	-	-	GH92
XALc_2875	-	-	-	GH125
XALc_2877	-	-	-	GH3
XALc_2880	-	-	-	GH18
XALc_2906	-	-	-	GT58
XALc_2956	-	AA6(3-186)	-	-
XALc_2958	-	-	GT2	GT2

<b>ID</b>	<b>EC</b>	<b>HMMER</b>	<b>dbCAN_sub</b>	<b>DIAMOND</b>
XALc_2967	3.2.1.4	GH5_1(37-352)+CBM2(407-481)	GH5_e1+CBM2_e21	CBM2+GH5_1
XALc_2969	3.2.1.4	GH5_1(37-348)+CBM2(473-548)	GH5_e1+CBM2_e21	CBM2+GH5_1
XALc_2972	3.2.1.28	GH37(63-538)	GH37_e6	GH37
XALc_2975	-	-	-	GH13_3
XALc_2980	-	GH42(38-398)	GH42_e30	GH42
XALc_2987	-	GH25(20-208)	GH25_e20	GH25
XALc_3121	-	-	-	GH103
XALc_3122	-	GH103(34-329)	GH103_e3	GH103
XALp_3190	-	GH23(43-173)	GH23_e79	GH23
XALq_3230	-	-	GH23_e95	GH23
XALr_3249	-	GH23(93-190)	GH23_e653	GH23

**Fonte:** Elaborada pelo autor.

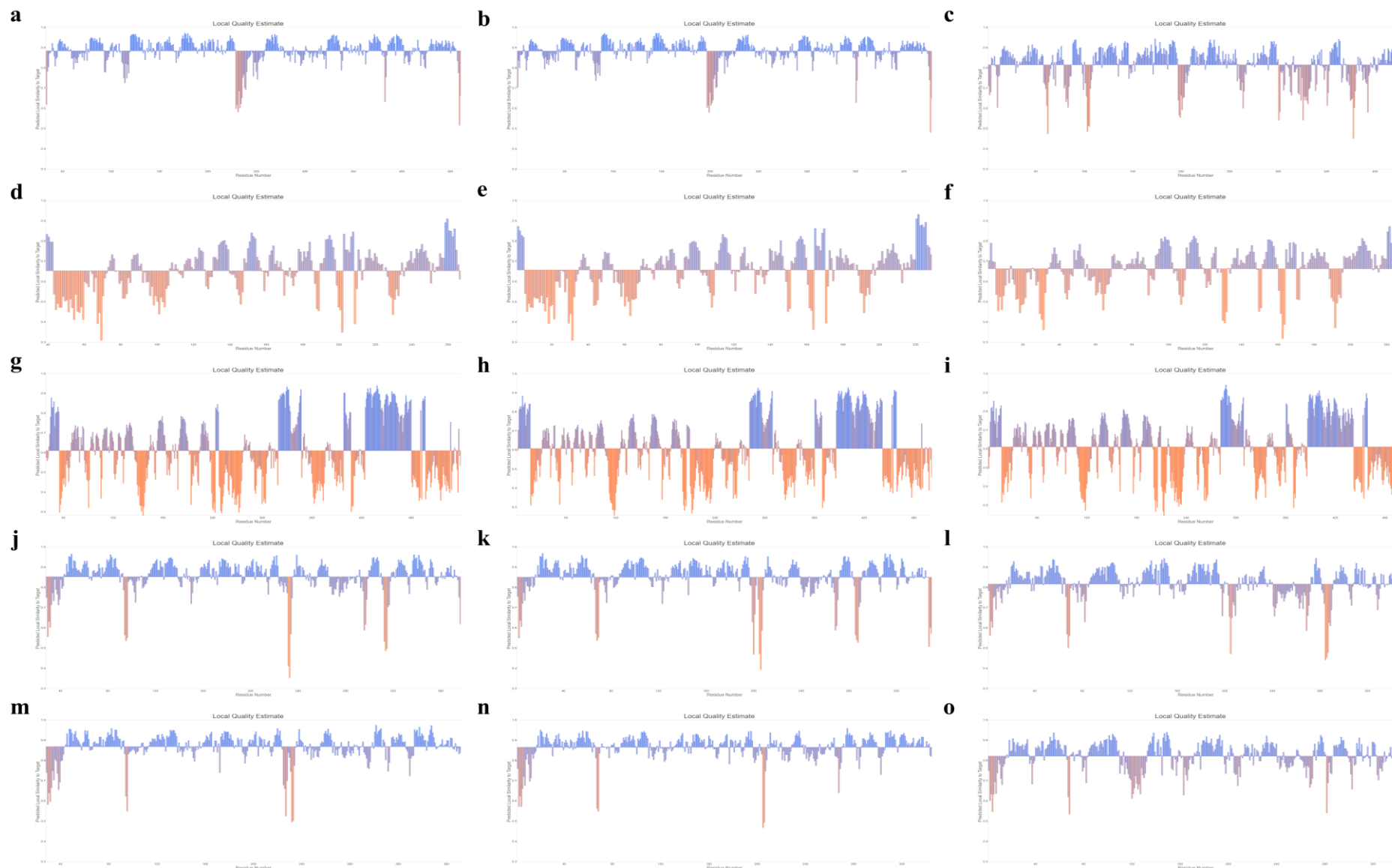
## APÊNDICE B - Análises estruturais e de qualidade das CAZymes de *Xanthomonas albilineans* selecionadas.

**Figura A1** - Visualização gráfica da estimativa de qualidade local determinada pelo QMEANDisCo para as quatro CAZymes de *Xanthomonas albilineans*, previstas pelos algoritmos de inteligência artificial, AlphaFold, ColabFold e RoseTTAFold. **(a)** XALc\_0865\_AlphaFold, **(b)** XALc\_0865\_ColabFold, **(c)** XALc\_0865\_RoseTTAFold, **(d)** XALc\_0874\_AlphaFold, **(e)** XALc\_0874\_ColabFold, **(f)** XALc\_0874\_RoseTTAFold, **(g)** XALc\_2967\_AlphaFold, **(h)** XALc\_2967\_ColabFold, **(i)** XALc\_2967\_RoseTTAFold, **(j)** XALc\_2969\_AlphaFold, **(k)** XALc\_2969\_ColabFold e **(l)** XALc\_2969\_RoseTTAFold. Os gráficos para a enzima XALc\_0484 estão representados na Figura 9.



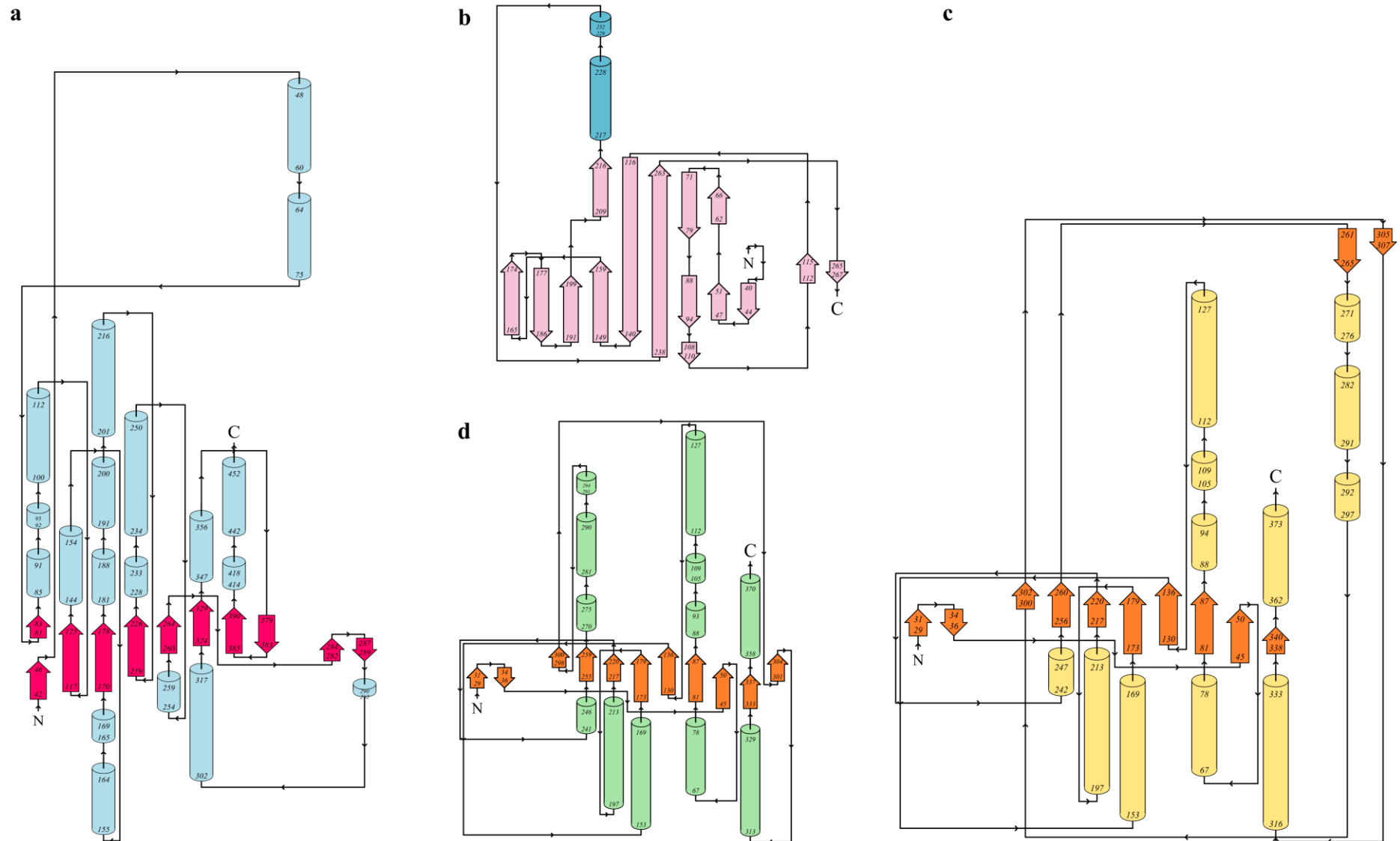
Fonte: Elaborada pelo autor.

**Figura A2** - Visualização gráfica da estimativa de qualidade local determinada pelo QMEANDisCo para o domínio catalítico das cinco CAZymes de *Xanthomonas albilineans*, previstas pelos algoritmos de inteligência artificial, AlphaFold, ColabFold e RoseTTAFold. **(a)** XALc\_0484\_AlphaFold, **(b)** XALc\_0484\_ColabFold, **(c)** XALc\_0484\_RoseTTAFold, **(d)** XALc\_0865\_AlphaFold, **(e)** XALc\_0865\_ColabFold, **(f)** XALc\_0865\_RoseTTAFold, **(g)** XALc\_0874\_AlphaFold, **(h)** XALc\_0874\_ColabFold, **(i)** XALc\_0874\_RoseTTAFold, **(j)** XALc\_2967\_AlphaFold, **(k)** XALc\_2967\_ColabFold, **(l)** XALc\_2967\_RoseTTAFold, **(m)** XALc\_2969\_AlphaFold, **(n)** XALc\_2969\_ColabFold e **(o)** XALc\_2969\_RoseTTAFold.



Fonte: Elaborada pelo autor.

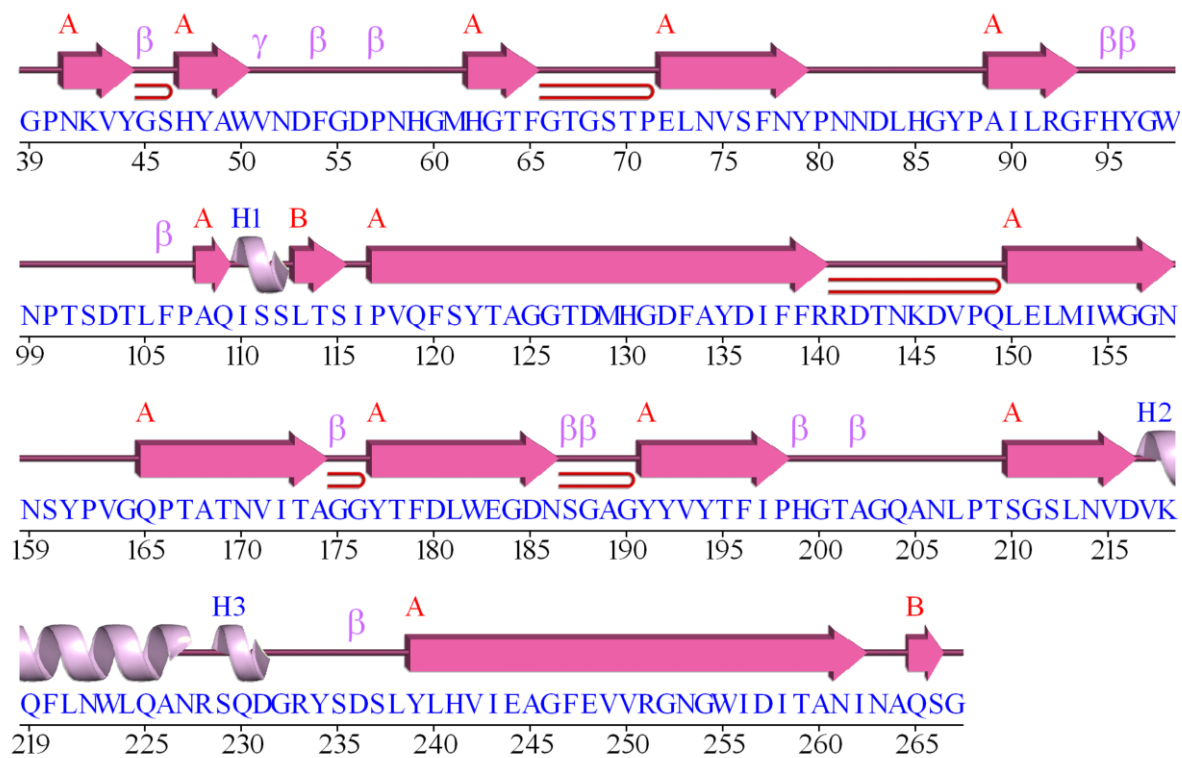
**Figura A3** - Previsão de topologia para as enzimas (a) XALc\_0484, (b) XALc\_0865, (c) XALc2967 e (d) XALc\_2969, geradas pelo software PDBsum. As fitas- $\beta$  são representadas pelas setas e as  $\alpha$ -hélices pelos bastonetetes. N e C sinalizam as regiões N- e C-terminais, respectivamente.



Fonte: Elaborada pelo autor.



**Figura A5** - Previsão de estrutura secundária para a enzima XALc\_0865, gerada pelo software PDBsum.

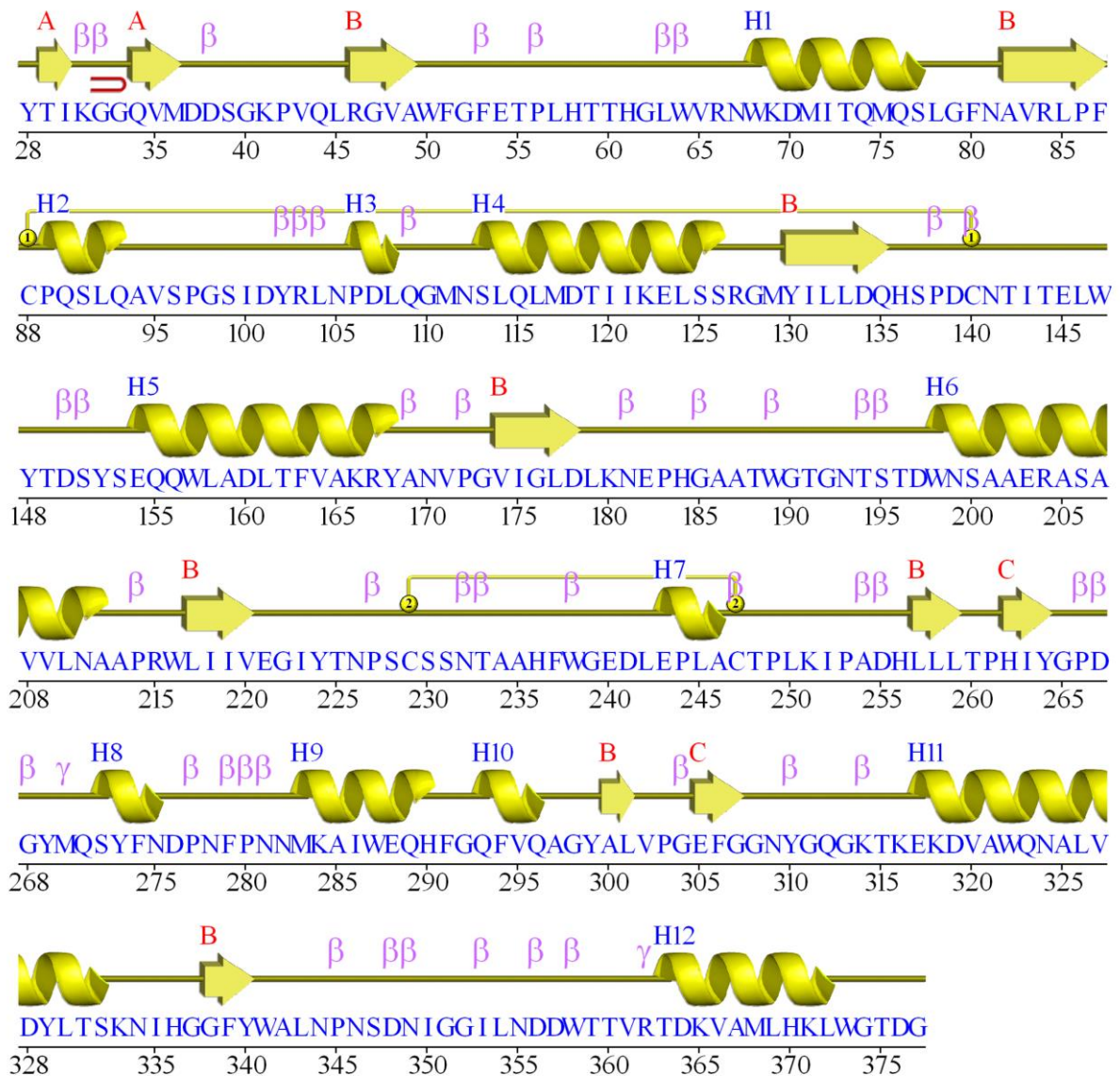


**Key:**




Sec. struc: Helices labelled H1, H2, ... and strands by their sheets A, B, ...  
 Helix Strand  
 Motifs:  $\beta$  beta turn  $\gamma$  gamma turn beta hairpin

**Fonte:** Elaborada pelo autor com o auxílio do PDBsum.

**Figura A6** - Previsão de estrutura secundária para a enzima XALc\_02967, gerada pelo software PDBsum.

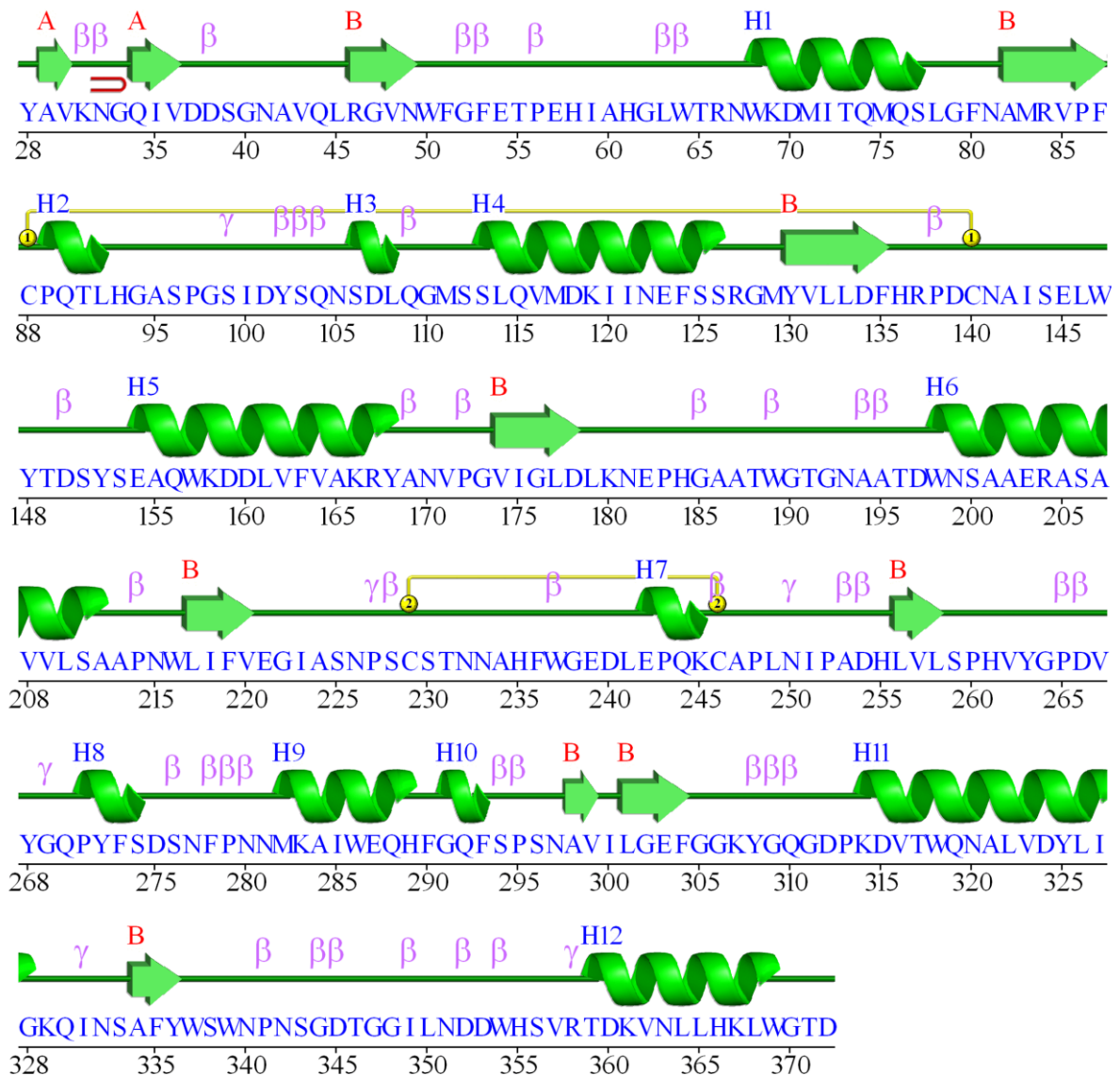


**Key:**

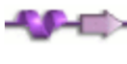


Sec. struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...  
 Motifs:  $\beta$  beta turn  $\gamma$  gamma turn  beta hairpin  
 Disulphides:  disulphide bond

**Fonte:** Elaborada pelo autor com o auxílio do PDBsum.

**Figura A7** - Previsão de estrutura secundária para a enzima XALc\_02969, gerada pelo software PDBsum.



**Key:**

Sec. struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...  
 Helix Strand  
 Motifs:  $\beta$  beta turn  $\gamma$  gamma turn  beta hairpin  
 Disulphides:  disulphide bond

**Fonte:** Elaborada pelo autor com o auxílio do PDBsum.