

Laboratório Nacional de Biorrenováveis (LNBR)  
Centro Nacional de Pesquisa em Energia e Materiais (CNPEM)

## Projeto para Programa Institucional de Bolsas de Iniciação Científica (PIBIC)

### Investigação dos aspectos dinâmicos de arabinofuranosidases por meio de simulações computacionais

Orientadora: Dra. Mariana Abrahão Bueno de Moraes

Coorientadora: Dra. Gabriela de Lima Menezes

Campinas, abril de 2026

## Resumo

As enzimas pertencentes à classe das hidrolases glicosídicas possuem grande importância para bioprocessos voltados à produção de biorrenováveis. Dentre elas, uma família conservada em bactérias especializadas na degradação da biomassa vegetal destaca-se, a GH43. Em nosso grupo de pesquisa, temos estudados diversos membros dessa família, entre elas a enzima 2533, do ponto de vista mecanístico, entretanto não foi possível obter sua estrutura tridimensional experimental. Este projeto tem por objetivo analisar o comportamento dessa proteína a partir de modelos gerados por algoritmos preditivos (como o AlphaFold2) que serão submetidos a simulações de dinâmica molecular. Com os resultados gerados, pretendemos investigar a interação da proteína com seu possível substrato e também compreender os motivos para a dificuldade de sua cristalização. Especialmente, avaliaremos o papel do domínio acessório nesse processo, tanto para a estabilidade da enzima, quanto para a sua interação com diferentes carboidratos. Posteriormente, esperamos verificar a possível conservação ou divergência do mecanismo de reconhecimento de carboidratos de outra arabinofuranosidase da mesma família, através da realização de simulações com outra enzima cuja estrutura foi determinada experimentalmente.

## Introdução

Enzimas ativas em carboidratos são de grande interesse para a área de biorrenováveis, na medida em que são essas moléculas que podem ser convertidas em biocombustíveis e outras moléculas. Nesse sentido, a desconstrução de polissacarídeos complexos é necessária para o aproveitamento dos açúcares para, por exemplo, a produção de etanol de segunda geração. Esse, contudo, não é um problema de solução facilmente escalável, de modo que o estudo e caracterização de novas enzimas que atuem de maneira sinérgica faz-se necessário.

A enzima 2533, encontrada na bactéria *Xanthomonas citri*, é uma hidrolase glicosídica (GH) da família 43, pertencente à subfamília 29<sup>1</sup>. Acredita-se que ela seja ativa em arabinóxilano, clivando a ligação glicosídica entre monômeros distintos e linearizando a cadeia de xiloses. É, portanto, um potencial agente para a despolimerização de polissacarídeos da hemicelulose, parte da parede celular vegetal, o que a torna interessante para estudo. A enzima possui um domínio catalítico e um domínio acessório, cuja função permanece desconhecida. Tentativas anteriores, realizadas pelo grupo, para cristalizar essa proteína e resolver sua estrutura tridimensional não produziram cristais bem formados. Assim, as abordagens computacionais possuem grande potencial para auxiliar a entender seu comportamento em solução bem como determinantes estruturais para o reconhecimento do substrato e a possível função do domínio acessório. Este projeto visa empregar técnicas computacionais de modelagem, *docking* e dinâmica molecular para o estudo dos aspectos estruturais e dinâmicos associados ao reconhecimento do substrato e atividade da enzima selecionada.

## Objetivos

Este projeto tem por objetivo estudar a estrutura e o comportamento da enzima 2533, incluindo interações interdomínio e domínio-substrato. Para tanto, serão realizados protocolos de dinâmica molecular utilizando o modelo de AlphaFold2 da proteína, disponível no UniProt, e o *software* AMBER. Dessa forma, pretende-se compreender a dificuldade de cristalização, o *docking* do carboidrato no domínio catalítico e, especialmente, a influência (ou falta dela) do domínio acessório. O processo será repetido com a estrutura cristalográfica recém-resolvida de outra GH da mesma família, buscando semelhanças de mecanismo. Procura-se obter, desse modo, informações inéditas acerca de duas enzimas com potencial para a área de biorrenováveis, além de esclarecimento sobre sua família como um todo.

## Materiais e Métodos

A simulação será feita com o modelo do AlphaFold2<sup>2</sup> e o *software* AMBER<sup>3</sup>. Será usado o campo de força Glycam-06j<sup>4</sup> com água de três pontos como solvente explícito. Anteriormente, o mesmo grupo de pesquisa utilizou o campo CHARMM36<sup>5</sup> no *software* GROMACS<sup>6</sup>, obtendo bons resultados para a proteína em água, inclusive preservando a coordenação do íon de cálcio no domínio catalítico sem aplicar restrições. No entanto, as simulações com um modelo de substrato (duas xiloses com uma arabinose) exigiram fortes restrições de posição e distância nos resíduos catalíticos e nos átomos relevantes para a ligação glicosídica.

Isso motiva a mudança para o campo de força Glycam; diferentemente do CHARMM, ele possui suporte para carboidratos, mas não está disponível no GROMACS, por isso o AMBER se torna necessário. Serão rodadas simulações da enzima 2533 com o fragmento de arabinoxilano tanto com quanto sem seu domínio acessório a fim de analisar diferenças, além de comparar com os seguintes resultados prévios: os 500 ns de simulação a 303 K já realizados revelam *loops* flexíveis no domínio catalítico, alguns com flutuação maior até que a do *linker*, o que pode explicar a dificuldade de formação de cristais. Apesar disso, os dois domínios se comportaram como rígidos e, após um período de cerca de 100 ns grande mobilidade, estabilizaram; a variação das distâncias interdomínio e até mesmo a flutuação dos resíduos do *linker* diminuíram depois.

O projeto utilizará a mesma divisão de domínios, resíduos catalíticos e coordenação com o cálcio que a da simulação anterior. Essa, por sua vez, foi feita com base em análise com o *software* VMD<sup>7</sup>, utilizando as estruturas secundárias para a definição do  $\beta$ -propeller e do domínio acessório, e alinhamento com outras GH43\_29, em especial a de PDB 8HCJ<sup>8</sup>, para definição dos resíduos relevantes. Nesse caso, além da sobreposição visual com análise de orientação e distância, foi observado o alinhamento MUSCLE<sup>9</sup>, já que os resíduos catalíticos são bem conservados nessa família, e a observação no UniProt<sup>10</sup> de que a ASP159 seria o regulador de

pKa. Os resíduos de conservação com o cálcio também se mostraram bem conservados; a 8HCJ também conta com esse íon.

O protocolo será repetido para a GH inédita, com a diferença de que seu arquivo .pdb conterà a estrutura obtida experimentalmente, não de um modelo computacional. Para ambos, serão usados os computadores do Laboratório Nacional de Biorrenováveis (LNBR) e o supercomputador Sdumont do Laboratório Nacional de Computação Científica (LNCC), valendo-se dos programas AMBER para rodar as dinâmicas, PyMOL<sup>11</sup> e VMD para análise visual, Aliview<sup>12</sup> para alinhamento da estrutura primária, XMGrace e JupyterLab (Python) para análise dos dados, além dos dados já obtidos com GROMACS.

### **Análise dos resultados**

Para todas as trajetórias geradas, será feita a análise visual no VMD e a análise quantitativa dos seguintes dados: distâncias entre os átomos da ligação glicosídica, entre o centro de massa de cada domínio e entre os C $\alpha$  das extremidades do *linker*; raio de giro (Rg), desvio (RMSD), flutuação (RMSF) e o número de estruturas secundárias pelo tempo. Nas simulações com mais de um domínio, serão calculados Rg e RMSD de cada um deles, além do RMSF do *linker* e do número e tipo de contatos entre os domínios. Caso seja observada, mais uma vez, uma estabilização após o início da etapa de produção, o RMSF será calculado antes e depois dessa configuração ser atingida, além da média global. Dessa forma, serão analisadas quantitativamente as interações interdomínio, domínio-substrato e proteína-substrato, além da estabilidade, compactação, rigidez e mobilidade da proteína e seus domínios individualmente, ao longo do tempo, das duas enzimas abordadas neste projeto e compará-las com os resultados prévios obtidos com o GROMACS.

### **Plano de Trabalho**

O cronograma de execução do projeto foi dividido trimestralmente e segue na Tabela 1:

**Tabela 1: Cronograma de execução do projeto.**

	1° trimestre	2° trimestre	3° trimestre	4° trimestre
Revisão bibliográfica	X	X	X	X
Análise estrutural e docagem molecular	X	X		
Simulações de dinâmica molecular (MD) clássicas	X	X	X	X
Escrita do relatório parcial		X		
Simulações de MD com outra enzima da família e comparação entre possíveis mecanismos			X	X
Escrita do relatório final			X	X

## Referências

1. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**, D233–D238 (2009).
2. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).
3. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688 (2005).
4. Kirschner, K. N. *et al.* GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622–655 (2008).
5. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
6. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J Comput Chem* **26**, 1701–1718 (2005).
7. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).
8. Vishwakarma, P. *et al.* Deciphering the structural and biochemical aspects of xylosidase from *Pseudopedobacter saltans*. *International Journal of Biological Macromolecules* **291**, 139042 (2025).
9. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
10. Apweiler, R. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research* **32**, 115D – 119 (2004).
11. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
12. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).