

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

MARCELO VIZONÁ LIBERATO

Caracterização estrutural de endoglucanases da família GH5 e beta-glicosidases da família GH1: interação enzima-substrato.

São Carlos
2013

MARCELO VIZONÁ LIBERATO

Caracterização estrutural de endoglucanases da família GH5 e beta-glicosidases da família GH1: interação enzima-substrato.

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular
Orientador: Prof. Dr. Igor Polikarpov

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2013

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Vizoná Liberato, Marcelo

Caracterização estrutural de endoglucanases da família GH5 e beta-glicosidases da família GH1: interação enzima-substrato. / Marcelo Vizoná Liberato; orientador Igor Polikarpov - versão corrigida -- São Carlos, 2013.

141 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2013.

1. Celulose. 2. Bioetanol. 3. Endoglucanase. 4. Beta-glicosidase. I. Polikarpov, Igor, orient. II. Título.

A todas as pessoas que se esforçam em tornar o
mundo um pouco melhor.

AGRADECIMENTOS

Ao meu orientador Igor, por todo suporte, apoio e confiança.

Aos colegas de trabalho e amigos César e Vanessa, por terem compartilhado de forma direta para a produção dessa tese.

Ao professor Fernando Araripe Torres e toda sua equipe, por terem me recebido tão bem em Brasília.

Aos pesquisadores Sean McSweeney, Alexander Popov e toda a equipe do ESRF, por terem me recebido na França de forma tão cordial e hospitaleira e por todo o suporte oferecido.

A minha eterna companheira Cinthia, pelo companheirismo nas horas difíceis e empolgação nas horas boas, pelas broncas que se mostraram necessárias e por suportar que eu falasse sobre ases e mais ases durante quatro anos e ainda assim ter aceitado se tornar minha esposa.

Aos meus pais e irmão, por todo o amor e confiança. Eles, juntamente com minha esposa, são os principais responsáveis por quem eu me tornei.

Aos novos integrantes da minha família, Daniel, Clarice, Thais, Claudia e Dani, pelo carinho e suporte, principalmente na reta final de meu doutorado. Agradeço especialmente à Clarice pela ajuda com a correção da minha tese.

A todos os amigos e parentes que, mesmo sem entender nada do que eu estava falando, acreditaram no meu trabalho e esforço.

Aos funcionários do IFSC, que sempre foram prestativos e eficientes.

A Capes, CNPq e Ciência Sem Fronteiras pelo suporte financeiro ao Doutorado e Doutorado Sanduiche.

“Como é que pode! Aquelas pessoas passam a vida no meio do mato olhando e estudando os bichos e as plantas. Eu acho isso muito bonito.”

José Alfredo Vizoná (meu Avô)

RESUMO

LIBERATO, M.V. **Caracterização estrutural de endoglucanases da família GH5 e beta-glicosidases da família GH1: interação enzima-substrato.** 2013. 141 p. Tese (Doutorado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

A celulose é o biopolímero de maior abundância no mundo e tem potencial para se tornar fonte de energia renovável através de sua transformação em açúcares fermentáveis, que por sua vez serão transformados em etanol. A recalcitrância da celulose, principal dificuldade encontrada no processo, pode ser superada com o auxílio de enzimas (celulases). Ao menos três enzimas celulolíticas são necessárias para a degradação total da celulose, incluindo as celobiohidrolases, que hidrolisam as ligações glicosídicas das extremidades redutoras e não redutoras da cadeia, as endoglucanases, que clivam a cadeia de celulose amorfa aleatoriamente, e as beta-glicosidases, que produzem glicose através dos celoo-oligômeros. Mas para que esse processo se torne financeiramente viável é necessário conhecer o funcionamento, otimizar a atividade e aumentar a produção dessas celulases. Com o intuito de avançar na compreensão da função e estrutura dessas enzimas, o presente trabalho teve como objetivo o estudo estrutural de beta-glicosidases da família GH1 e endoglucanases da família GH5. Na primeira parte do trabalho, a expressão da endoglucanase II de *Trichoderma reesei* não foi alcançada, mesmo utilizando diferentes organismos e condições de expressão. Porém, na segunda etapa, foi obtida a expressão, purificação e os primeiros ensaios de cristalização de 11 beta-glicosidases bacterianas da família GH1 e 8 endoglucanases bacterianas da família GH5. Dentre elas, três beta-glicosidases e uma endoglucanase de *Bacillus licheniformis* foram cristalizadas e tiveram sua estrutura resolvida. As beta-glicosidases, apesar de possuírem o enovelamento similar, apresentaram variações no tamanho e posição das alças formadoras da fenda catalítica e divergem em relação a um dos aminoácidos importantes para a estabilização do substrato. Essas diferenças podem ajudar a explicar o mecanismo dessas enzimas para reconhecer substratos distintos. A endoglucanase da família GH5, possuindo dois módulos acessórios, foi cristalizada tanto na forma apo quanto complexada ao substrato celotetraose. O segundo módulo acessório possivelmente é um domínio de ligação à celulose (CBM) e seus resíduos aromáticos, que são responsáveis pela interação com o substrato, parecem complementar o sítio catalítico, sendo assim um novo mecanismo de auxílio enzimático de um CBM. O primeiro módulo acessório não possui um aparente sítio de interação com carboidratos e provavelmente funciona como um conector entre domínio catalítico e o CBM. O posicionamento do substrato no sítio de ligação é parecido com outras estruturas já determinadas, porém, suscita algumas dúvidas sobre a função dos resíduos catalíticos que é conservada na família. O carbono anomérico do substrato possui uma densidade eletrônica contínua com o glutamato da fita $\beta 4$ (que deveria ser o ácido/base) e está mais próximo dele que do glutamato da fita $\beta 7$ (que deveria ser o nucleófilo).

Palavras-chave: Celulose. Bioetanol. Endoglucanase. Beta-glicosidase.

ABSTRACT

LIBERATO, M.V. **Structural characterization of endoglucanases from family GH5 and beta-glucosidases from family GH1: enzyme-substrate interaction.** 2013. 141 p. Tese (Doutorado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

Cellulose is the most abundant biopolymer in the world and can become a renewable energy source through its transformation in fermentable sugars, which will be converted in bioethanol. The cellulose recalcitrance, main difficulty in the process, can be overcome with the aid of enzymes (cellulases). At least three cellulolytic enzymes are required for complete hydrolysis of cellulose, including cellobiohydrolases for hydrolyzing the glycosidic linkages from the reducing and non-reducing chain ends, endoglucanases for randomly cleaving cellulose chains in the amorphous regions, and beta-glucosidases for producing glucose from the solubilized cello-oligomers. But, to become a financially viable process it is necessary to know the mechanism, optimize the activity and improve the production of these cellulases. In order to advance the understanding of the structure and function of these enzymes, the present work intended to study the structure of beta-glucosidases from family GH1 and endoglucanases from family GH5. In the first part of the work, the expression of endoglucanase II from *Trichoderma reesei* was not achieved, even using different organisms and expression conditions. However, in the second part, the expression, purification and the crystallization first trials of eleven bacterial beta-glucosidases and eight bacterial endoglucanases were achieved. Among them, three beta-glucosidases and one endoglucanase from *Bacillus licheniformis* were crystallized and had their structures solved. Beta-glucosidases, although having a similar folding, showed variations in the length and position of the loops that form the catalytic cleft and diverge in relation to one of the amino acids that are important in substrate stabilization. These differences may help explain the mechanism of these enzymes to recognize distinct substrates. The endoglucanase, which has two accessory modules, was crystallized in the apo form and complexed with the substrate celotetraose. The second accessory module probably is a cellulose binding domain (CBM) and its aromatic residues, which are responsible for the substrate interaction, seem to complement the catalytic site. Therefore it can be a new mechanism of CBM assistance in the enzymatic activity. The first accessory module has no apparent interaction site with carbohydrates and probably works as a connector between the catalytic domain and CBM. The positioning of the substrate in the binding site is similar to other structures already solved but raises some questions about the role of the catalytic residues, that are conserved in the family. The anomeric carbon of the substrate has a continuous electron density with glutamate from sheet- β 4 (which should be the acid/base) and is closer to it than to glutamate from sheet- β 7 (which should be the nucleophile).

Keywords: Cellulose. Bioethanol. Endoglucanase. Beta-glucosidase.

LISTA DE FIGURAS

- Figura 1 - Da biomassa das plantas ao bioetanol. A celulose presente na parede celular das plantas é o biopolímero mais abundante no mundo e é uma potencial fonte de açúcar fermentável para a produção de bioetanol. A celulose, após um pré-tratamento, pode ser hidrolisada por celulases gerando unidades de glicose, gerando assim matéria prima para a fermentação alcoólica. Adaptado de¹² 30
- Figura 2 - Representação estrutural das camadas que formam a parede celular de um traqueídeo. Figura modificada de²⁹. 34
- Figura 3 - Modelo simplificado da hidrólise enzimática da celulose. As EGs primeiramente clivam a celulose nas regiões amorfas liberando extremidades redutoras e não redutoras de polissacarídeos para que as CBHs possam atacar e processivamente hidrolisar essas extremidades. As BGs hidrolisam as celobioses resultantes da atuação das EGs e CBHs, liberando glicose. Essa figura foi modificada de¹². 38
- Figura 4 - Mecanismos hidrolíticos de inversão e retenção das celulases. AH: resíduo catalítico ácido, B-: resíduo catalítico básico, Nuc: nucleófilo, and R: derivado de carboidrato. HOR*: nucleófilo exógeno, geralmente uma molécula de água. Adaptado⁷⁴. 43
- Figura 5 - Estrutura tridimensional de CBMs, representantes dos três grupos determinados pelo modo de ligação ao substrato. a) CBM tipo A de uma celobiohidrolase de *T. reesei*⁹²; b) CBM tipo B de endoglucanase e *C. fini*⁹³; c) CBM tipo C de xilanase 10A de *T. maritima*⁹⁴. Os resíduos aromáticos e alguns dos substratos estão representados em formato *stick*. 47
- Figura 6 - Vetor pGEM-T utilizado para multiplicação dos genes Eg1 e Eg2 em *E. coli*. 52
- Figura 7 - Vetor pPIC9K utilizado para expressão induzida por metanol das enzimas Eg1 e Eg2 em *P. pastoris*. 53
- Figura 8 - Vetor pPICZa utilizado para expressão autoinduzida das enzimas Eg1 e Eg2 em *P. pastoris*. 54
- Figura 9 - Vetor pET28a utilizado para expressão da enzima Eg2 em *E. coli*. 55
- Figura 10 - Eletroforese em géis de agarose da amplificação por PCR dos genes: a) EgI e b) Eg2. 67

- Figura 11 - Eletroforese em gel de agarose do teste de restrição de colônias transformadas com pPIC9K+Eg1. No interior do retângulo vermelho estão as bandas relativas aos fragmentos gerados pela clivagem com a enzima de restrição NcoI..... 68
- Figura 12 - Eletroforese em gel de agarose do teste de restrição de colônias transformadas com pPIC9K+Eg2. Os plasmídeos foram clivados com a enzima de restrição Sall. 68
- Figura 13 - Eletroforese em gel de agarose da linearização dos plasmídeos com a enzima de restrição SacI. MM. Padrão de peso molecular; 1. Eg1/pPIC9K clivado; 2. Eg1/pPIC9K não clivado; 3. Eg2/pPIC9K clivado; 4. Eg2/pPIC9K não clivado..... 69
- Figura 14 - Eletroforese em gel SDS do teste de expressão de Eg2 em *P. pastoris*. Foi aplicado nos geis a fração sobrenadante de cada cultura, com a exceção do gel “c”, colunas de 6 a 10, onde foi aplicada a fração precipitada. a) 1. Padrão de peso molecular; 2. 24 horas de indução, não transformado; 3. 48 horas de indução, não transformado; 4. 72 horas de indução, não transformado; 5. 96 horas de indução, não transformado; 6. 24 horas de indução, colônia 1; 7. 48 horas de indução, colônia 1; 8. 72 horas de indução, colônia 1; 9. 96 horas de indução, colônia 1. b) 1. Padrão de peso molecular; 2. 24 horas de indução, colônia 3; 3. 48 horas de indução, colônia 3; 4. 72 horas de indução colônia 3; 5. 96 horas de indução, colônia 3; 6. 24 horas de indução, colônia 4; 7. 48 horas de indução, colônia 4; 8. 72 horas de indução, colônia 4; 9. 96 horas de indução, colônia 4. c) 1. Padrão de peso molecular; 2. 24 horas de indução, colônia 5; 3. 48 horas de indução, colônia 5; 4. 72 horas de indução colônia 5; 5. 96 horas de indução, colônia 5; 6. Lisado antes da indução, colônia 5; 7. Lisado, 24 horas de indução, colônia 5; 8. Lisado, 48 horas de indução, colônia 5; 9. Lisado, 72 horas de indução, colônia 5; 10. Lisado, 96 horas de indução, colônia 5. d) 1. Padrão de peso molecular; 2. 24 horas de indução, pPIC9K sem inserto; 3. 96 horas de indução, pPIC9K sem inserto; 4. 96 horas de indução colônia 6; 5. 96 horas de indução, colônia 8; 6. 96 horas de indução, colônia 9. e) 1. Padrão de peso molecular; 2. 96 horas de indução colônia 10; 3. 96 horas de indução colônia 11; 4. 96 horas de indução colônia 12; 5. 96 horas de indução, colônia 13; 6. 96 horas de indução, colônia 14. 71
- Figura 15 - Eletroforese em gel SDS do teste de expressão de Eg1 e Eg2 em colônias de *P. pastoris* crescidas em 4 mg/mL de geneticina. Foi aplicado nos geis a fração sobrenadante de cada cultura após indução por 72 horas. Foram testadas 15 colônias para expressão de Eg1(barra vermelha) e 20 colônias para expressão de Eg2 (barra azul). Nas colunas indicadas pela letra C (controle) foram aplicadas amostras das culturas antes da indução..... 72
- Figura 16 - Western Blotting do teste de expressão de Eg2 em colônias de *P. pastoris*. Os números representam as colônias testadas. A única banda distinguível encontra-se dentro do retângulo vermelho..... 73

- Figura 17 - Eletroforese em gel SDS do teste de expressão da colônia 8 de Eg2 em de *P. pastoris* crescidas e induzidas em diferentes condições de acordo com a tabela 3. Foi aplicada nos geis a fração sobrenadante de cada cultura. Os números abaixo de cada gel representam as condições descritas na tabela 3. As amostras foram analisadas após 24 horas (barra vermelha), 48 horas (barra azul), 72 horas (barra laranja), 96 horas (barra verde) e 120 horas (barra roxa) de indução.....74
- Figura 18 - Eletroforese em gel SDS da purificação de Eg2 por cromatografia de afinidade. 1. Padrão de peso molecular; 2. Sobrenadante filtrado; 3. Sobrenadante filtrado e concentrado; 4. Fração não aderida à resina; 5. Primeira lavagem; 6. Segunda lavagem; 7. Terceira lavagem; 8. Eluição com 200 mM de imidazol; 9. Eluição com 500 mM de imidazol; 10. Resina após eluição.75
- Figura 19 - Eletroforese em gel de agarose da clivagem dos plasmídeos pPIC9K/Eg1 e pPIC9K/Eg2 pelas enzimas de restrição NotI e BstBI.76
- Figura 20 - Eletroforese em gel de agarose do teste de restrição de colônias transformadas com pPZα/Eg1 e pPZα/Eg2. Foram testadas quatro colônias de cada construção. As letras C e N representam as colunas dos plasmídeos clivados e não clivados, respectivamente.77
- Figura 21 - Eletroforese em gel de agarose da linearização dos plasmídeos pPZα/Eg1 e pPZα/Eg2 com a enzima de restrição SacI. As letras C e N representam as colunas dos plasmídeos clivados e não clivados, respectivamente.....77
- Figura 22 - Eletroforese em gel de agarose da amplificação por PCR do gene Eg2 para expressão em *E. coli* (a esquerda), e do teste de restrição com as enzimas NdeI e XhoI do plasmídeo pET28a/Eg2 extraído de duas colônias (a direita).78
- Figura 23 - Eletroforese em gel SDS do teste de expressão de Eg2 em diferentes cepas de *E. coli*. Foram testadas duas colônias de cada cepa: BL21; BL21 Codon Plus (CP); BL21 Ril (ril); e BL21 Origami2 (Ori). As letras P e S representam as frações precipitadas e solúveis respectivamente.....79
- Figura 24 - Eletroforese em gel SDS do teste de expressão de Eg2 a baixa temperatura em diferentes cepas de *E. coli*. Foram testadas três colônias da cepa BL21 Origami2 (Ori), duas colônias BL21 comum, duas colônias BL21 Codon Plus (CP) e duas colônias BL21 RIL (ril). As letras P e S representam as frações precipitadas e solúveis respectivamente.....80
- Figura 25 - Eletroforese em gel SDS da purificação de Eg2 por cromatografia de afinidade. 1. Padrão de peso molecular; 2. Fração insolúvel; 3. Fração solúvel; 4. Fração não aderida à resina; 5. Primeira lavagem; 6. Última lavagem; 7. Eluição com 200 mM de imidazol; 8. Eluição com 500 mM de imidazol; 9. Resina após eluição81

Figura 26 - Eletroforese em gel SDS do teste de expressão das enzimas H6a (75 kDa), H6b (63 kDa), H6c (53 kDa) e TEV. Foram testadas duas colônias de cada enzima, com a exceção da TEV (33 kDa). Foi aplicada nos géis a fração solúvel de cada lise.....	85
Figura 27 - Eletroforese em gel SDS do teste de expressão das enzimas: A) A9 (69 kDa), B5 (57 kDa), B9 (69 kDa), C6 (71 kDa), C9 (70 kDa), D5 (68 kDa), E5 (69 kDa), F5 (68 kDa) e F6 (68 kDa); B) F7 (60 kDa), F9 (72 kDa), F12 (76 kDa), G5 (70 kDa), G9 (71 kDa) e H5 (71 kDa). Foi aplicada nos géis a fração solúvel de cada lise.	86
Figura 28 - Eletroforese em gel SDS do teste de expressão das enzimas B5 (57 kDa), E5 (69 kDa), E12 (56 kDa), F1 (62 kDa), F4 (53 kDa), F6 (68 kDa), F7 (60 kDa), F9 (72 kDa), F12 (76 kDa), G9 (71 kDa) e H5 (71 kDa). Foram aplicadas nos géis as frações: A) Solúvel; B) Precipitado.	86
Figura 29 - Eletroforese em gel SDS do teste de expressão das enzimas F1 (62 kDa), F4 (53 kDa) e F9 (72 kDa). As letras S e P representam, respectivamente, as frações solúveis e precipitadas de cada cultura.	87
Figura 30 - Eletroforese em gel SDS da purificação por afinidade e clivagem da cauda de histidinas e tiorredoxina das enzima H6a, H6b, H6c, A9, B9, D5, E5 e H5. Os números indicados representam: 1) Fração precipitada; 2) Fração Solúvel; 3) Fração não aderida à resina; 4) Primeira lavagem; 5) Última lavagem; 6) Eluição com 200 mM de imidazol; 7) Clivagem com TEV; 8) Segunda passagem pela coluna; 9) Resina após eluição.	89
Figura 31 - Eletroforese em gel SDS da purificação por afinidade e clivagem da cauda de histidinas e tiorredoxina das enzima B5, F6, F12, F7, F4, F5, G5, C6 e C9. Os números indicados representam: 1) Fração precipitada; 2) Fração Solúvel; 3) Fração não aderida à resina; 4) Primeira lavagem; 5) Última lavagem; 6) Eluição com 200 mM de imidazol; 7) Clivagem com TEV; 8) Segunda passagem pela coluna; 9) Resina após eluição.	90
Figura 32 - Eletroforese em gel SDS da purificação por exclusão das enzima B5 e F12. As setas vermelhas indicam as frações que foram utilizadas para concentração e ensaio de cristalização. As frações foram aplicadas no gel de acordo com o tempo de passagem pela coluna.	91
Figura 33 - Teste enzimático da enzima H6a em diferentes temperaturas.	92
Figura 34 - Teste enzimático da enzima H6a em diferentes pHs.....	92
Figura 35 - Atividade relativa da enzima H6a frente a diferentes substratos.	93

- Figura 36 - Fotos dos diferentes formatos de cristais da enzima H6a. Note no item D a presença de duas formas cristalinas na mesma gota. As condições de cristalização de cada item foram: A) 25 % de PEG 3350, 0,2 M de cloreto de sódio e 0,1 M de Tris, pH 8,5; B) 25 % de PEG 3350, 0,2 M de sulfato de amônio e 0,1 M de Hepes, pH 7,5; C) 25 % de PEG 3350 e 0,1 M de Tris, pH 8,5; D) 22,5 % de PEG 4000, 14 % de isopropanol e 0,1 M de citrato de sódio, pH 6,0; E) 25 % de PEG 4000, 18 % de isopropanol e 0,1 M de citrato de sódio, pH 6,0.....96
- Figura 37 - Modelo cristalográfico da enzima H6. O domínio catalítico está representado pela cor de laranja, o módulo 1 é magenta e o módulo 2 é verde. 100
- Figura 38 - Sobreposição do modelo cristalográfico de H6 (domínio catalítico na cor de laranja e módulo 2 na cor verde) com a sialidase de *Micromonospora viridifaciens* (cor azul, PDB: 4HU0). Em amarelo está representada uma celotetraose ligada à sialidase, ocupando do subsítio -2 ao +2. 101
- Figura 39 - Modelo cristalográfico do domínio catalítico da enzima H6, constituído de 8 fitas β e oito α -hélices. Os resíduos catalíticos E159 e E268 estão evidenciados no centro da figura. 102
- Figura 40 - Celotetraose ligada a enzima H6. A) Modelo da celotetraose e dos resíduos catalíticos ajustados na densidade eletrônica (1σ) obtida pelos dados experimentais. Três das quatro glicoses ocupam os subsítios -1, -2 e -3. Observe a continuidade da densidade eletrônica entre a celotetraose e o resíduo E159. B) Estabilização do substrato no sítio ativo pelos resíduos indicados. As linhas tracejadas representam ligações de hidrogênio. 103
- Figura 41 - Modelo cristalográfico dos módulos 1 (em magenta) e 2 (em verde) da enzima H6. Os módulos foram superpostos e são constituídos de 7 fitas β e uma pequena α -hélice..... 105
- Figura 42 - Fotos dos cristais de H6b e H6c. A) H6b, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5; B) H6b, condição: 4 % de PEG 8000 e 0,1 M de Hepes, pH 7,0; C) H6c, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5. 106
- Figura 43 - Padrões de difração das proteínas: A) H6b, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5; B) H6c, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5..... 107
- Figura 44 - Modelos cristalográficos das enzimas D5 (em verde), F5 (em ciano) e G5 (em magenta). Abaixo está representada a sobreposição das três estruturas evidenciando a face em que se encontra o sítio ativo. As linhas pontilhadas (com as cores relacionadas com cada estrutura) representam a distância em angstroms entre as alças. No centro está representado o inibidor gluco-hidroximolactam (PDB:2J78)¹⁴¹ ligado no sítio ativo. 111

- Figura 45 - Resíduos responsáveis pela interação com substrato. Estão representadas as estruturas D5 (verde), F5 (ciano) e G5 (magenta). No centro estão modelados os inibidores de beta-glicosidase gluco-hidroximolactam (PDB: 2J78)¹⁴¹ e p-nitrofenil glucopiranosídeo (PDB: 3WBA)¹⁴⁴, cujas estruturas foram superpostas..... 112
- Figura 46 - Representação da interação do carboidrato (amarelo) com os resíduos catalíticos E159 e E268 da enzima H6 (laranja). À esquerda o carboidrato está representado com o oxigênio glicosídico e à direita este oxigênio está ausente. 117
- Figura 47 – Representação em superfície dos modelos de D5 (verde) e G5 (magenta). No centro de cada enzima está modelado o inibidor de beta-glicosidase gluco-hidroximolactam através da sobreposição com a estrutura 2J78¹⁴¹ 120

LISTA DE TABELAS

Tabela 1 - Algumas das funções desempenhadas por membros da família GH5.....	41
Tabela 2 - Relação das enzimas estudadas. Cada uma está designada por um código para facilitar seu reconhecimento. A tabela expõe a família pertencente, gene (código GenBank), organismo e o tamanho de cada enzima.	50
Tabela 3 - Condições selecionadas para teste de expressão da enzima Eg2.	74
Tabela 4 - Determinação dos valores de T _m (em °C) de cada enzima testada nos diferentes tampões e pHs. Os maiores valores encontrados estão destacados em vermelho. Os valores designados com um X representam que a determinação do T _m não foi possível.	94
Tabela 5 - Estatísticas dos dados cristalográficos da enzima H6a coletados em quatro grupos espaciais diferentes na forma apo (conjuntos de A a D) e complexada com o substrato celopentaose (conjunto E).	97
Tabela 6 - Condições de cristalização das enzimas D5, F5 e G5. Estão destacadas em negrito as condições em que foram obtidos os cristais com conjunto de dados coletados. As demais condições foram encontradas inicialmente pelos ensaios em robô de cristalização.....	107
Tabela 7 - Estatísticas dos dados cristalográficos das enzimas D5, F5 e G5.	109
Tabela 8 - Alinhamento dos resíduos responsáveis pela interação com substrato das enzimas D5, F5, G5 e de outras beta-glicosidases publicadas (designadas pelo código PDB). As posições foram numeradas de acordo com a ordem sequencial crescente dos aminoácidos.....	112

LISTA DE ABREVIATURAS E SIGLAS

CAZy	Carbohydrate-Active enZYmes
CBH	Celobioidrolase
CMC	Carboximetilcelulose
EC	Enzyme Commission
EG	Endoglucanase
EMBL	European Molecular Biology Laboratory
ESRF	European Synchrotron Radiation Facility
GH	Hidrolases de glicosídeos
GT	Transferases de glicosídeos
IPTG	Isopropil- β -D-1-tiogalactosideo
LB	Luria Broth
LIC	Clonagem Independente de Ligação
LNLS	Laboratório Nacional de Luz Síncrotron
PCR	Polimerase Chain Reaction
PDB	Protein Data Bank
PEG	Polietilenoglicol
PEG MME	Polietilenoglicol Monometil Éter
PMSF	Phenylmethanesulfonyl Fluoride
RMSD	Root-Mean-Square Deviation
TEV	Tobacco Etch Virus

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Visão global do processo de produção do etanol de segunda geração	30
1.2	Estrutura da biomassa lignocelulósica	31
1.2.1	Parede celular	32
1.2.1.2	Parede celular secundária.....	33
1.2.1.3	Componentes da parede celular	34
1.2.1.3.1	Celulose	34
1.2.1.3.2	Hemicelulose	35
1.2.1.3.3	Pectina.....	36
1.2.1.3.4	Lignina	36
1.2.1.3.5	Proteína	36
1.3	Enzimas.....	37
1.3.1	Celulases	37
1.3.2	CAZy	40
1.3.3	GH5.....	40
1.3.4	GH1.....	43

1.3.5	CBM.....	44
2	OBJETIVOS.....	49
3	MATERIAIS E MÉTODOS	51
3.1	Biblioteca de cDNA de <i>T. harzianum</i>	51
3.2	Clonagens dos genes Eg1 e Eg2 de <i>T. harzianum</i> para expressão em <i>P. pastoris</i>	51
3.2.1	pPIC9K.....	52
3.2.2	pPZa.	53
3.3	Clonagem do gene Eg2 de <i>T. harzianum</i> para expressão em <i>E. coli</i>	54
3.4	Clonagem dos genes das famílias GH1 e GH5 para expressão em <i>E. coli</i>	55
3.5	Preparação de células de <i>E. coli</i> competentes.	56
3.6	Transformação das bactérias.	56
3.7	Preparação de células de <i>P. pastoris</i> competentes.	56
3.8	Transformação de <i>P. pastoris</i>	57
3.9	Teste de expressão em <i>E. coli</i>	58
3.10	Teste de expressão em <i>P. pastoris</i>	58
3.11	<i>Western Blotting</i>	59
3.12	Expressão em <i>E. coli</i>	59
3.13	Purificação por afinidade e clivagem da cauda de histidinas.	60

3.14	Produção e purificação da protease TEV.....	61
3.15	Purificação por exclusão de tamanho.	61
3.16	Teste de termoestabilidade.....	62
3.17	Determinação da atividade.....	63
3.18	Cristalização.....	64
3.19	Difração de raios-X.....	64
3.20	Tratamento dos dados de difração de raios-X.	65
4	ENDOGLUCANASE I E ENDOGLUCANASE II DE <i>T. reesei</i>	67
4.1	RESULTADOS	67
4.1.1	Clonagem de Eg1 e Eg2 para expressão em <i>P. pastoris</i> pelo vetor pPIC9K.	65
4.1.2	Expressão de Eg1 e Eg2 em <i>P. pastoris</i> pelo vetor pPIC9K.	67
4.1.3	Clonagem de Eg1 e Eg2 para expressão em <i>P. pastoris</i> pelo vetor pPZa.....	75
4.1.4	Clonagem de Eg2 para expressão em <i>E. coli</i>	76
4.1.5	Expressão de Eg2 em <i>E. coli</i>	77
4.2	DISCUSSÃO	79
4.2.1	Expressão em <i>P. pastoris</i>	79
4.2.2	Expressão em <i>E. coli</i>	831
5	ENDOGLUCANASES DA FAMÍLIA GH5 E BETA-GLICOSIDASES DA FAMÍLIA GH1.....	85

5.1	RESULTADOS.....	83
5.1.1	Expressão.....	83
5.1.2	Purificação.....	85
5.1.3	Atividade enzimática da enzima H6 e suas construções.....	89
5.1.4	Teste de termoestabilidade.....	93
5.1.5	Cristalização e modelo estrutural da enzima H6.....	95
5.1.6	Cristalização das truncagens da enzima H6.....	105
5.1.7	Cristalização e modelo estrutural das enzimas D5, F5 e G5.....	107
5.2	DISCUSSÃO.....	113
5.2.1	Atividade da H6.....	114
5.2.2	Estrutura da H6.....	115
5.2.3	Beta-glicosidases D5, F5 e G5.....	118
6	CONCLUSÕES.....	121
	REFERÊNCIAS.....	123
	ANEXO A.....	139
	ANEXO B.....	141

1 INTRODUÇÃO

A necessidade humana de energia vem se tornando cada vez maior de acordo com o aumento da população e com o maior gasto energético *per capita*. Essa necessidade vem acompanhada por um maior interesse científico e industrial em fontes alternativas de energia como os biocombustíveis¹. Dentre eles, o mais utilizado no mundo hoje em dia é o bioetanol². Seu processo de produção utiliza apenas fontes de energia renováveis, tornando-o ambientalmente favorável, e há uma grande variedade de fontes naturais que podem fornecer açúcares para sua produção, classificadas de acordo com sua composição: sacarose (cana-de-açúcar, beterraba e sorgo), amido (trigo, milho e cevada) e biomassa lignocelulósica (madeira, palha e gramíneas)³.

O etanol de primeira geração, produzido através de sacarose e amido, já possui tecnologia e procedimentos industriais bem desenvolvidos. O amido é primeiramente sacarificado, geralmente por enzimas amilolíticas, transformando-se em açúcares simples que, subsequentemente, são fermentados por leveduras para produzir o etanol. Entretanto, para que o etanol possa substituir os combustíveis fósseis como fonte de energia, especialmente no setor de transportes, um grande volume precisa ser produzido. Por consequência, o uso de amido e sacarose como substratos para a produção de etanol parece ser insuficiente, pois esses materiais são também utilizados como alimento, igualmente necessários às necessidades humanas⁴.

A biomassa de plantas, que inclui bagaço de cana, palha de arroz e de milho, é considerada como lixo e é frequentemente eliminada por fogo. Hoje ela está se tornando uma fonte atrativa para a produção de etanol (etanol de segunda geração) graças a sua abundância, disponibilidade e baixo custo, além de não ser consumida como alimento. A utilização desse material, ao invés de sua queima, pode fazer com que a poluição do ar seja reduzida e que a produção de etanol aumente, sem a necessidade de maior área de plantio.

Os principais componentes estruturais da biomassa são celulose, hemicelulose e lignina, encontradas principalmente nas paredes celulares das plantas (figura 1). Mas a heterogeneidade e complexidade desses componentes fazem da produção de etanol por material lignocelulósico um processo complicado, pois as leveduras não têm a capacidade de hidrolisar as ligações β -1,4-glicosídicas que formam a celulose⁵. Embora os tratamentos químicos como a hidrólise ácida possam quebrar a lignocelulose, seus subprodutos podem

agir como inibidores, levando a uma baixa produtividade. Portanto, a hidrólise da celulose por enzimas (celulases) é um meio alternativo viável, enquanto os outros componentes podem ser primeiramente removidos por outras enzimas (hemicelulases) ou por pré-tratamentos físico-químicos⁶⁻⁷.

Inicialmente avaliado pelo Laboratório Nacional de Energia Renovável (NREL, EUA), o custo das celulases representa 20 % da produção de etanol, fazendo com que a venda tenha altos preços. Deste modo, a redução do custo das enzimas é primordial para que a técnica seja aplicável. Várias estratégias têm sido propostas para a redução do custo, como maximização da secreção de enzimas pelas leveduras através de mutagênese⁸, transgenia de plantas para expressar celulases heterólogas⁹, utilização de enzimas termofílicas¹⁰ e aumento da eficiência das enzimas por engenharia de proteínas¹¹.

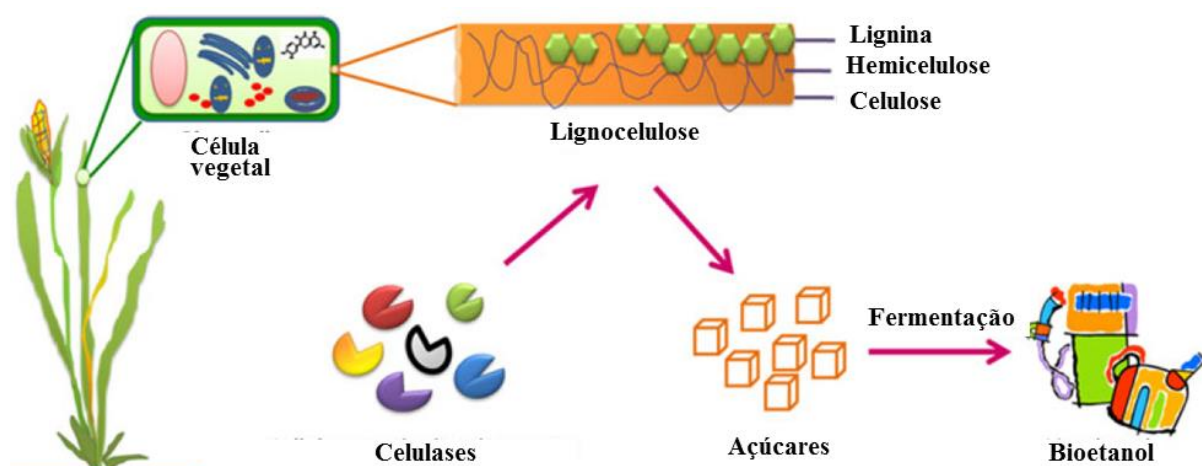


Figura 1 - Da biomassa das plantas ao bioetanol. A celulose presente na parede celular das plantas é o biopolímero mais abundante no mundo e é uma potencial fonte de açúcar fermentável para a produção de bioetanol. A celulose, após um pré-tratamento, pode ser hidrolisada por celulases gerando unidades de glicose, gerando assim matéria prima para a fermentação alcoólica. Adaptado de¹².

1.1 Visão global do processo de produção do etanol de segunda geração

A conversão de lignocelulose em etanol consiste em quatro grandes unidades operacionais: pré-tratamento, hidrólise, fermentação e destilação¹³.

O pré-tratamento tem a finalidade de aumentar o tamanho dos poros da matéria prima e reduzir a cristalinidade da celulose. No pré-tratamento ácido, a camada de hemicelulose é hidrolisada, enquanto que no tratamento alcalino, parte da lignina é removida e a

hemicelulose é hidrolisada por hemicelulases. Portanto, o pré-tratamento é necessário para expor as fibras de celulose ao ataque enzimático ou, ao menos, tornar a celulose mais acessível. Esse processo deve ser feito com uma mínima formação de compostos que podem inibir a fermentação pelos micro-organismos¹⁴.

A segunda etapa, a hidrólise enzimática, promove a despolimerização das fibras de celulose e outros polissacarídeos¹⁵. Os produtos finais das reações com as enzimas são açúcares monoméricos, que serão fermentados pelas bactérias, leveduras ou fungos filamentosos na terceira etapa do processo. Contrariamente à produção de etanol por sacarose ou amido, a produção por lignocelulose se dá por uma fermentação de vários carboidratos diferentes com a presença de compostos inibitórios (ácidos orgânicos de baixo peso molecular, derivados de furano e compostos fenólicos) liberados e formados durante o pré-tratamento e/ou hidrólise da matéria prima¹⁶.

Após a purificação final (por destilação ou outra técnica de separação), o etanol está pronto para ser utilizado como combustível. Uma parte da lignina, a principal parte sólida que sobra da biomassa, pode ser queimada para fornecer calor e eletricidade para o processo, enquanto o restante é conservado por ser um coproduto valioso¹⁶.

1.2 Estrutura da biomassa lignocelulósica

Os materiais lignocelulósicos são os biopolímeros mais abundantes na Terra. Considera-se que esses compostos abranjam 50 % da biomassa mundial, chegando a 200 milhões de toneladas produzidas por ano¹⁷. Em geral, os materiais lignocelulósicos utilizados para a produção de etanol podem ser divididos em seis grupos: resíduos de plantações (bagaço de cana, palha e casca de milho, trigo, arroz, etc.), madeiras duras (álamo), madeiras menos duras (pinheiros e abetos), descartes celulósicos (papel, jornais, material reciclável), biomassa herbácea (gramíneas) e dejetos sólidos municipais¹⁷.

Numerosos estudos para o desenvolvimento da produção de etanol por biomassa lignocelulósica têm sido feitos pelo mundo. Entretanto, o principal fator limitante é o alto grau de complexidade inerente ao processamento da matéria prima¹⁸. Isto está relacionado com a natureza e composição da biomassa lignocelulósica. Como pode ser notado, as fontes de biomassa, que são atualmente utilizadas para a produção de etanol, têm origem vegetal, e praticamente todo o material lignocelulósico está localizado nas paredes celulares.

1.2.1 Parede celular

Uma das características que distingue a célula vegetal da célula animal é a presença de parede celular que a envolve. Essa parede celular tem várias funções relacionadas ao crescimento, desenvolvimento, sustentação e defesa das plantas¹⁹. A parede celular primária é formada nos estágios iniciais da diferenciação celular. Quando é necessário um maior suporte estrutural, uma segunda parede celular é produzida na parte interior da parede celular primária. Essas camadas de parede celulares são divergentes em estrutura, organização e composição, portanto, possuem funções diferentes. Geralmente, as paredes celulares são formadas por uma organização de vários tipos de polissacarídeos, fenilpropanoides e proteínas estruturais²⁰⁻²¹. Esses polímeros, principalmente os polissacarídeos, representam a maior parte da biomassa das plantas que é utilizada pela humanidade como fonte de madeira, papel, alimento e combustível.

1.2.1.1 Parede celular primária

A parede celular primária das plantas é formada na lamela média das células em crescimento, logo depois da divisão celular. Na maioria dos tipos celulares, ela possui uma arquitetura simples e sua espessura varia de 0,1 a 1,0 μm enquanto que, em células do colênquima ou epidérmicas, ela pode ser mais espessa e com várias camadas²². A parede primária tem muitas funções importantes, como suporte estrutural e mecânico, proteção contra patógenos e desidratação, manutenção e determinação do formato da célula e regulação do turgor celular. Além disso, ela proporciona plasticidade, permitindo a expansão e divisão celular durante o crescimento.

As paredes celulares primárias são classificadas em três tipos de acordo com sua composição. O tipo 1, encontrado em dicotiledôneas e monocotiledôneas não comelinoides, consiste em quantidades aproximadamente iguais de celulose e xiloglucano²³. A estrutura formada por esses dois tipos de carboidratos está embebida em uma matriz rica em pectinas em que três tipos diferentes (homogalacturonano, ramnogalacturonano I e

rhamnogalacturonano II) interligam covalentemente as microfibrilas. O tipo 2, encontrado em espécies de gramíneas como o arroz²⁴, é caracterizado por microfibrilas com a mesma estrutura do tipo 1, porém, utiliza o glucuroarabinosilano como principal conector das microfibrilas através de ligações de hidrogênio. As paredes de tipo 2 também possuem baixas quantidades de pectina e de xiloglucano não decorados com arabinose e fucose, possuindo assim estrutura similar às paredes das dicotiledôneas. A recentemente descrita parede primária tipo 3, encontrada tipicamente em pteridófitas, possui mananos como principal glicano conector, chegando a 40 % (em mols) dos glicanos responsáveis por essa conexão nas amostras analisadas²⁵. Da mesma forma que o tipo 2, o tipo 3 também possui baixa quantidade de pectinas.

A parede primária é o principal componente estrutural dos alimentos derivados de plantas e é especialmente abundante nas bebidas derivadas de plantas²⁶. Por isso, a estrutura e organização da parede celular primária tem gerado um grande interesse por parte dos cientistas, nutricionistas e indústria de alimentos.

1.2.1.2 Parede celular secundária

A parede celular secundária das plantas é formada após a expansão e alongação celular entre a membrana plasmática e a parede primária, envolvendo células especializadas como os elementos de vasos. Com a formação da parede secundária, a parede celular se torna mais espessa e rígida e sua contribuição para a biomassa das plantas é muito maior do que da parede primária. A parede secundária de angiospermas arbóreas e de coníferas é a principal matéria bruta para a produção de papel, madeira, combustíveis e fibras têxteis²⁷⁻²⁸.

Geralmente, a parede secundária está dividida em três camadas designadas S1, S2 e S3²⁸, onde as microfibrilas celulósicas estão ordenadas de forma paralela, mas com diferentes ângulos (figura 2). A camada S1 é mais fina (0,1-0,35 μm) e suas microfibrilas formam um ângulo de 60-80° em relação à S2. A camada S2 é a mais espessa (1-10 μm) e dominante na parede secundária. Ela é a principal responsável pelo fornecimento do suporte mecânico das células. A camada 3 é relativamente fina (0,5-1,1 μm) e suas microfibrilas formam um ângulo de 60-90° em relação à S2²⁹.

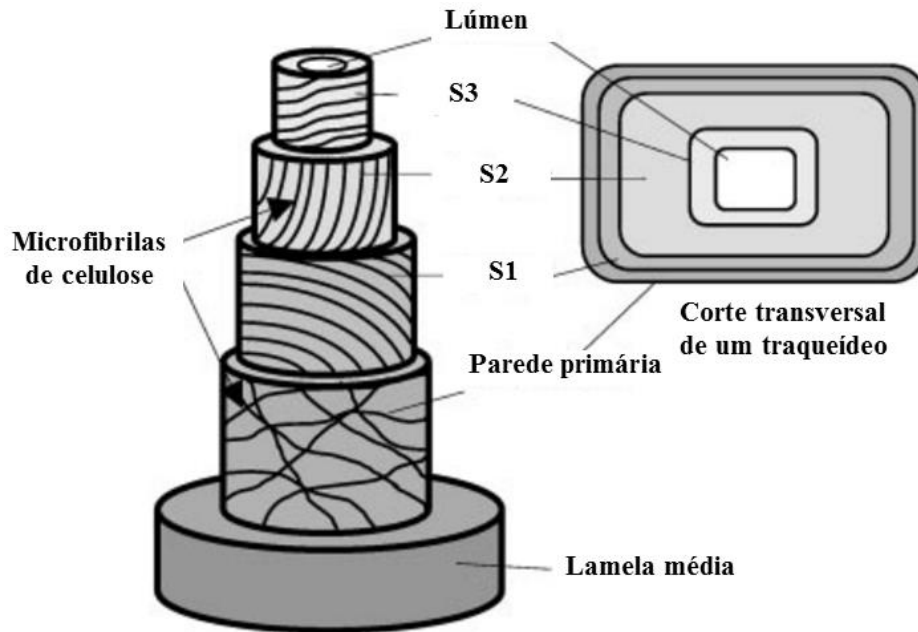


Figura 2 - Representação estrutural das camadas que formam a parede celular de um traqueídeo. Figura modificada de²⁹.

1.2.1.3 Componentes da parede celular

A parede celular primária de plantas arbóreas é constituída tipicamente de 20-30 % de celulose, 30-70 % de hemicelulose, 5-35% de pectinas, até 5% de glicoproteínas e uma quantidade reduzida de outros componentes como ésteres fenólicos³⁰⁻³¹.

A parede secundária do mesmo grupo de vegetais geralmente contem 35-50 % de celulose, 20-50 % de hemicelulose, até 20 % de lignina e uma pequena quantidade de pectinas e proteínas³⁰⁻³¹. Entretanto, os componentes e suas proporções podem variar entre as diferentes espécies, tecidos e até mesmo tipos celulares na mesma planta³⁰.

1.2.1.3.1 Celulose

A celulose é o polissacarídeo mais abundante da parede celular das plantas. Este carboidrato é sintetizado na membrana plasmática e suas microfibrilas são depositadas diretamente na matriz celular³². A celulose e suas microfibrilas são estáveis, insolúveis e resistentes à digestão enzimática. A molécula de celulose é uma cadeia de glicanos

conectados por ligações β -1,4-glicosídicas contendo de 500 a 14000 unidades de glicose. Cada cadeia é estabilizada por ligações de hidrogênio intramoleculares e cerca de 40 cadeias são conectadas através de ligações de hidrogênio intermoleculares para formar as microfibrilas. Essas microfibrilas são redes cristalinas ou semicristalinas que possuem 4-10 nm de espessura, embora possam chegar a 30 nm em algumas algas³³.

1.2.1.3.2 Hemicelulose

Os polissacarídeos hemicelulósicos são encontrados tanto na parede celular primária quanto na secundária, em que eles geralmente estão fortemente ligados à superfície celulósica através de ligações de hidrogênio. O galactomanano, glicomanano, glucuronoarabinosilano, arabinoglucuronosilano, glucuronosilano e xiloglucano são os exemplos mais comuns e abundantes de hemiceluloses das paredes celulares de plantas³⁰. Nas paredes primárias de dicotiledôneas, o xiloglucano é a principal hemicelulose, enquanto glucuronoarabinosilano é mais abundante em gramíneas. Já na parede secundária das dicotiledôneas e das gramíneas a principal hemicelulose é o glucuronosilano³⁴.

A hemicelulose é caracterizada por ligações β -1,4-glicosídicas entre os monossacarídeos da cadeia principal, podendo ser linear ou ramificada. A cadeia principal pode ser formada pelos monossacarídeos glicose, xilose, manose e galactose, enquanto nas ramificações são encontrados arabinose, ácido glucurônico e galactose. Sua hidrofobicidade está relacionada com suas ramificações, chegando a ser razoavelmente hidrofílica quando nelas são encontrados grupos ácidos³⁵.

O xiloglucano é a hemicelulose mais abundante e, conseqüentemente, a mais bem estudada. Ele é sintetizado no Complexo de Golgi e carregado para a membrana celular por vesículas³⁶. É caracterizado por uma cadeia principal de glicoses, conectadas por ligações β -1,4-glicosídicas que são frequentemente substituídas por xiloses e conectadas por ligações β -1,6. Os resíduos de xilose podem conter ramificações que são formadas, geralmente, por galactoses e fucoses³⁷.

1.2.1.3.3 Pectina

A pectina é um polissacarídeo altamente hidrofílico e é encontrado em abundância na lamela média e na parede celular primária do tipo I. As pectinas se associam a hemiceluloses para formar a matriz hidratada em que as microfibrilas de celulose estão embebidas³⁸. As pectinas possuem muitas funções, como controle da porosidade da parede, regulação do pH e reconhecimento de compostos, tanto para a defesa celular quanto para composições simbióticas³⁸. A cadeia principal das pectinas é formada por unidades de ácido D-galacturônico conectados por ligações α -1,4. As pectinas estão divididas em três grupos de acordo com modificações na cadeia principal e suas ramificações: homogalacturonano, rhamnogalacturonano-I e “galacturonanos substituídos”³⁹.

1.2.1.3.4 Lignina

A lignina é uma molécula hidrofóbica e complexa encontrada na lamela média e na parede celular secundária. Ela é encontrada preenchendo os espaços entre os demais componentes da parede celular e sua principal função é dar rigidez e impermeabilidade às células, promovendo uma proteção contra ataques físicos e biológicos⁴⁰. A lignina é composta por três precursores polifenólicos (álcoois cumarílico, coniferílico e sinapílico) ligados covalentemente e não-covalentemente⁴¹.

A proporção de ligninas está diretamente relacionada com a utilização da planta pelo homem. As plantas com maior concentração de lignina são muito utilizadas em construções enquanto as plantas com menor concentração são usadas na produção de papel⁴².

1.2.1.3.5 Proteína

Além dos carboidratos, a parede celular das plantas é composta por uma grande variedade de proteínas. Possivelmente, centenas de proteínas diferentes são encontradas na parede celular⁴³, podendo ser funcionais ou estruturais⁴⁴.

As proteínas estruturais são encontradas em todas as partes da parede celular, mas é especialmente abundante na parede primária. Elas são classificadas em grandes grupos de acordo com a composição de aminoácidos: glicoproteínas ricas em hidroxiprolinas; proteínas ricas em glicinas; proteínas ricas em prolinas; entre outros⁴⁵.

As proteínas funcionais, como as enzimas oxidativas e enzimas hidrolíticas, estão envolvidas com o desenvolvimento celular, proteção contra patógenos, expansão celular e maturação da parede celular⁴⁶.

1.3 Enzimas

Hoje em dia, as enzimas provocam um considerável interesse industrial por causa de sua excelente especificidade para substratos e por sua habilidade em atuar em condições suaves, oferecendo, assim, um processo ambientalmente favorável para a obtenção de produtos-alvo⁴⁷. O valor estimado para as enzimas no mercado mundial é de cerca de 2,3 bilhões de dólares por ano, sendo que as enzimas relacionadas aos alimentos constituem a maior fatia do mercado⁴⁸. Em relação à aplicação, as enzimas estão distribuídas em indústrias de alimentos (45 %), detergentes (34 %), têxteis (11 %) couro (3 %) e papel (1,2 %)⁴⁸. Atualmente, as celulases compõem o terceiro maior volume de enzimas utilizadas no mundo⁴⁹ e são empregadas no processamento de algodão, reciclagem de papel, produção de detergentes, extração de sucos e aditivos alimentares de animais⁵⁰. Mesmo com essa grande utilização, a demanda por celulases provavelmente vai aumentar com a implantação da produção do bioetanol de segunda geração⁴⁹.

1.3.1 Celulases

Celulase é um termo coletivo que se refere a enzimas capazes de hidrolisar celulose⁵¹. A celulose é um homopolímero de glicoses, e sua organização faz com que ela seja uma estrutura altamente ordenada e coesa. Para conseguir a hidrólise completa da celulose, três categorias de celulases são necessárias. A primeira, as endoglucanases (EG, endo-1,4-β-D-glucanase, EC 3.2.1.4), ataca preferencialmente as regiões amorfas da celulose e cliva

randomicamente as ligações internas das cadeias, gerando extremidades reductoras e não reductoras para o ataque das enzimas pertencentes à segunda categoria, as celobiohidrolases (CBH, 1,4- β -D-glucano-celobiohidrolase, EC 3.2.1.91). As CBHs hidrolisam os carboidratos pelas extremidades de uma maneira processiva, liberando celobiose como principal produto. Por último, as beta-glicosidasas (BG, β -D-glucosideglucanohidrolase, EC 3.2.1.21) hidrolisam a celobiose em glicoses e também podem liberar glicoses das extremidades não reductoras de polissacarídeos solúveis (figura 3)⁵².

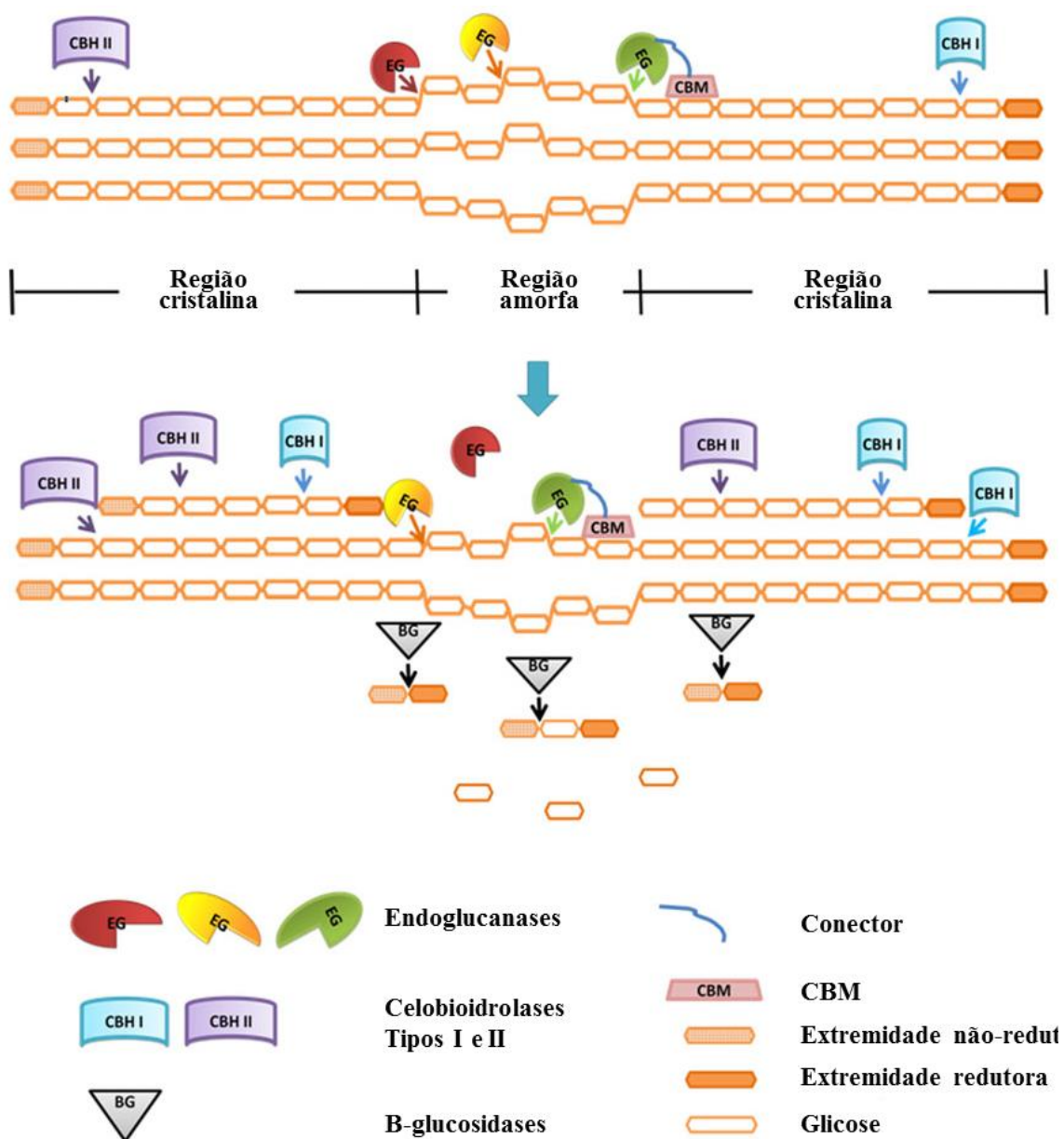


Figura 3 - Modelo simplificado da hidrólise enzimática da celulose. As EGs primeiramente clivam a celulose nas regiões amorfas liberando extremidades reductoras e não reductoras de polissacarídeos para que as CBHs possam atacar e processivamente hidrolisar essas extremidades. As BGs hidrolisam as celobioses resultantes da atuação das EGs e CBHs, liberando glicose. Essa figura foi modificada de¹².

Em oposição aos substratos solúveis que podem se difundir até o sítio ativo das enzimas, a celulose é insolúvel. Desta forma, as celulasas têm que se difundir, ancorar e posicionar o segmento do polímero de celulose no seu sítio ativo⁵³. A maioria das celulasas são proteínas modulares formadas por domínios catalíticos discretos com um ou mais módulos de ligação à celulose (CBMs), que estão ligados por conectores flexíveis⁵⁴. O CBM funciona como uma sonda celulósica, cuja principal função é manter a enzima ligada à celulose e aumentar a concentração efetiva de enzimas na superfície da celulose⁵⁵.

As celulasas são encontradas em plantas, insetos, bactérias, fungos e arqueas⁵⁶. Na natureza, as bactérias e fungos desempenham um papel vital como decompositores que aceleram o decaimento da biomassa de plantas⁵⁷. As bactérias aeróbias produzem as enzimas celulolíticas em uma forma “livre”, em que os diversos componentes trabalham sinergisticamente durante a hidrólise da celulose. Em contraste, muitas bactérias anaeróbicas produzem um complexo enzimático conhecido como celulosomo⁵⁸.

O celulosomo é um complexo multienzimático discreto formado por várias subunidades. No celulosomo de *Clostridium thermocellum*, a proteína primária denominada “scaffoldin” contém nove módulos de coesina, que respondem pela ligação das subunidades catalíticas ao complexo. Também faz parte desse complexo um CBM da família 3 responsável pela adesão à celulose⁵⁹. Alguns artigos têm mostrado que os celulosomos são mais eficientes na hidrólise da celulose do que as celulasas não complexadas, contudo a produção do celulosomo sob condições anaeróbicas é demorada e cara para a célula⁵⁹⁻⁶⁰.

Como as bactérias aeróbicas, os fungos aeróbios também produzem EGs, CBHs e BGs que trabalham sinergisticamente na hidrólise da celulose. A característica significativa dos fungos celulolíticos que atrai nosso interesse é o fato de que eles produzem grande quantidade de enzimas extracelulares. Por exemplo, várias linhagens do fungo filamentosso *Trichoderma* são conhecidas como superprodutoras de EGs e CBHs, enquanto espécies de *Aspergillus* são grandes produtoras de BGs e pectinases⁶¹.

Atualmente, as celulasas comerciais ainda são derivadas de fungos, embora as celulasas bacterianas estejam se mostrando competitivas em relação à eficiência na hidrólise da celulose. O *Trichoderma reesei* é o mais conhecido produtor de celulasas em escala industrial e ele tem sido estudado desde sua descoberta durante a Segunda Guerra Mundial. O *T. reesei* é considerado um poderoso fungo para a degradação de celulose cristalina, pois pode secretar grandes quantidades de celulase (acima de 100 g/L) com alta atividade específica⁴⁹⁻⁶². Outras espécies, como *Humicola insolens*, são utilizadas como produtores de enzimas

comerciais com diferentes propósitos. Por exemplo, as celulases de *H. insolens* são geralmente utilizadas em na produção de têxteis e detergentes, presumivelmente pela presença de celulases específicas que agem em condições levemente alcalinas e sob elevadas temperaturas⁶³. Para a sacarificação de celulose, alguns fungos pertencentes aos gêneros *Penicillium*, *Acremonium* e *Chrysosporium* estão se mostrando promissores para competir com *T. reesei* em alguns aspectos, como nível de produção enzimática e desempenho hidrolítico. Maiores informações sobre outros organismos com potencial para produção de celulases pode ser encontrado na revisão de Gusakov, 2011⁶².

1.3.2 CAZy

CAzy (Carbohydrate-Active enZYmes) é um banco de dados que contém uma classificação e informações associadas a enzimas envolvidas na síntese, metabolismo e transporte de carboidratos (CAZimas)⁶⁴. Nesse banco de dados estão incluídas as hidrolases de glicosídeos (GH), glicosiltransferases (GT), liases de polissacarídeos, esterases de carboidratos e proteínas de ligação a carboidratos.

A classificação de CAZimas baseada em sequências foi iniciada em 1991 como complemento ao sistema numérico amplamente utilizado Enzyme Commission (EC), que é fundamentado apenas nas atividades enzimáticas. Dado o predomínio da evolução convergente nas enzimas que clivam ligações glicosídicas, assim como a promiscuidade catalítica de várias delas, a classificação baseada em sequência provou ser um meio robusto para unificar informações sobre estrutura, especificidade e mecanismos das mesmas⁶⁵. Inicialmente motivada por uma necessidade de dispor as celulases em grupos estruturais distintos⁶⁶, a primeira classificação de GHs continha 35 famílias⁶⁵. Acompanhando o crescente interesse em glicobiologia, esse número vem aumentando continuamente e hoje conta com 132 famílias.

1.3.3 GH5

Atualmente, um das famílias mais numerosas é a GH5, historicamente conhecida como “família de celulasas A” por ter sido a primeira família de celulasas descrita⁶⁶. A GH5 é um exemplo de família com ampla variedade de funções. Apesar da função mais encontrada ser a de endo- β -1,4-glucanase, ela contém perto de 20 atividades determinadas experimentalmente denotadas com um número EC e algumas de suas enzimas já possuem estrutura cristalográfica resolvida (tabela 1). A abundância de enzimas GH5 em diferentes nichos ecológicos tem sido evidenciada pela sua frequente identificação em metagenomas de diversas comunidades microbiais⁶⁷⁻⁶⁸⁻⁶⁹ e genomas de organismos específicos⁷⁰.

Assim como outras famílias, as GH5s são frequentemente encontradas como parte de proteínas multimodulares contendo outros módulos catalíticos, CBMs ou módulos com funções ainda não identificadas⁷⁰.

Tabela 1 - Algumas das funções desempenhadas por membros da família GH5.

Função	E.C.	Subfamília	Estrutura representativa
endo- β -1,4-glucanase	3.2.1.4	GH5_1	2ZUN
β -mannosidase	3.2.1.25	GH5_7	IRH9
endo- β -1,4-mannanase	3.2.1.78	GH5_8	2WHL
endoglycoceramidase	3.2.1.123	GH5_28	2OSX
licheninase	3.2.1.73	GH5_36	IVJZ
glucan β -1,3-glicosidase	3.2.1.58	GH5_9	3N9K
Ainda não determinada	3.2.1.*	GH5_34	2Y8K

A família GH5 pertence ao clã GH-A que agrupa 19 famílias de hidrolases de glicosídeos, formando o maior grupo de famílias descritas no CAZy evolutivamente relacionadas. O clã é um grupo de famílias que se origina de um ancestral comum, mas muito distante. Embora a similaridade sequencial seja pequena, os membros possuem um enovelamento proteico e um mecanismo catalítico conservado.

Famílias como a GH5 foram originalmente definidas com um número muito pequeno de sequências. Com o aumento dos dados sequenciais, a relação entre algumas famílias

originais mudou o suficiente para que seus membros fossem reexaminados. Recentemente, uma análise detalhada de estruturas tridimensionais levou à reclassificação de várias sequências da família GH5 para a família GH30⁷¹. Em 2012, os membros da família GH5 foram separados em 51 subfamílias de acordo com a similaridade sequencial⁷².

Em relação ao mecanismo de ação, as famílias de GH estão separadas em duas classes: retenção ou inversão⁷³, dependendo da mudança da configuração do oxigênio glicosídico durante a reação (figura 4).

No mecanismo de inversão o resíduo catalítico ácido doa um próton para o carbono anomérico enquanto o resíduo catalítico básico remove um próton de uma molécula de água, aumentando sua nucleofilicidade e facilitando seu ataque ao centro anomérico. No mecanismo de retenção o resíduo catalítico ácido/base funciona em duas etapas, primeiro como ácido (glicosilação) e posteriormente como base (deglicosilação). Durante a primeira etapa o resíduo ácido/base doa um próton para o oxigênio glicosídico enquanto o resíduo nucleófilo forma uma ligação covalente com o carbono anomérico (estado intermediário). Na segunda etapa, o resíduo ácido/base desprotonado age como uma base para retirar um próton de uma molécula de água que, em seguida, ataca o intermediário enzima-glicosil. Com os dois passos de inversão, a estereoquímica do centro anomérico é mantida.

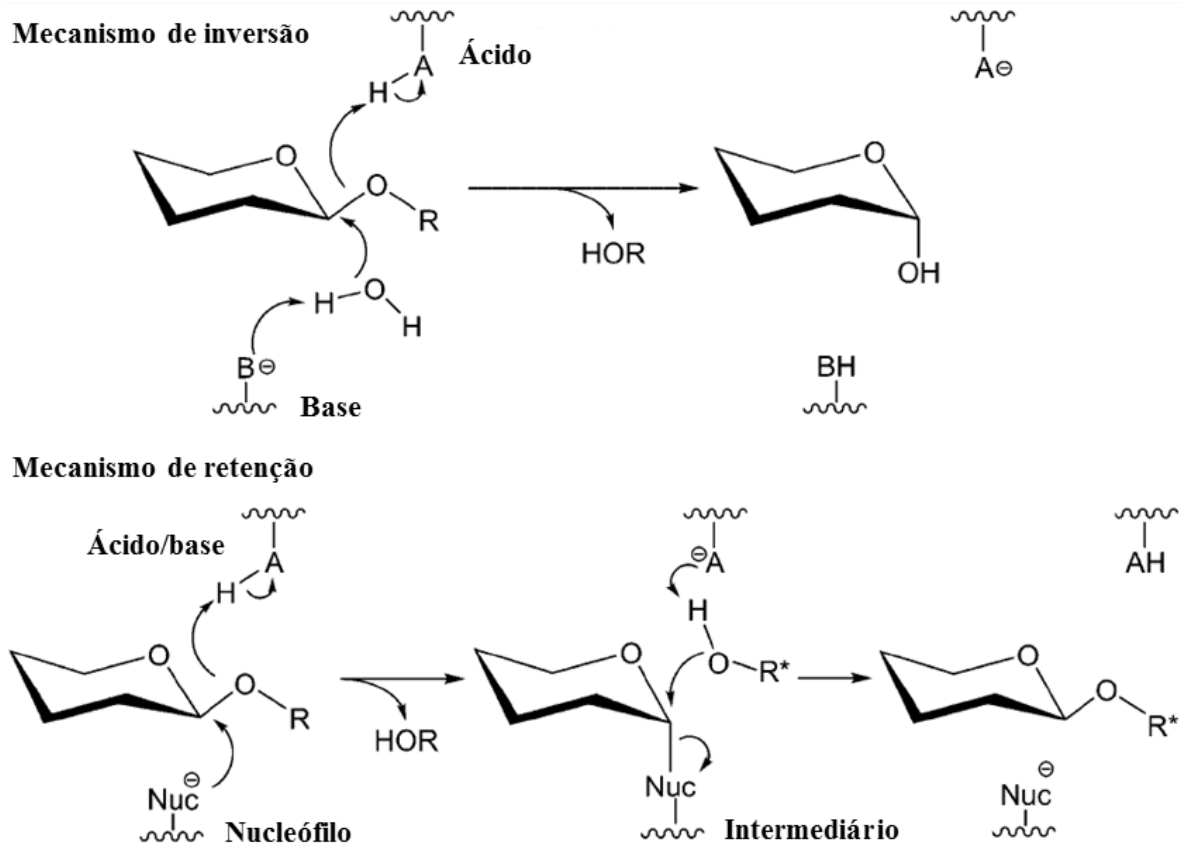


Figura 4 - Mecanismos hidrolíticos de inversão e retenção das celulases. AH: resíduo catalítico ácido, B⁻: resíduo catalítico básico, Nuc: nucleófilo, and R: derivado de carboidrato. HOR*: nucleófilo exógeno, geralmente uma molécula de água. Adaptado⁷⁴.

As GH5, por atuarem com o mecanismo de retenção, têm o potencial de serem hidrolases e transglicosilases. No sítio ativo, um par de glutamatos conservados, atuando como ácido/base e nucleófilo, são responsáveis pela reação. Dependendo do nucleófilo exógeno atuante na segunda etapa da reação, a enzima pode atuar como hidrolase (se for uma água) ou transglicosidase (se for outro carboidrato)⁷⁵.

Até a data de impressão dessa tese já foram reportadas 40 estruturas cristalográficas de enzimas da família GH5, de acordo com o CAZy. A primeira estrutura resolvida foi de uma endoglucanase de *Clostridium thermocellum* em 1995⁷⁶. Como membros do Clã GH-A, elas possuem um enovelamento clássico conhecido como barril TIM (β/α)₈ com os dois resíduos catalíticos separados sequencialmente por aproximadamente 200 resíduos e localizados nas extremidades carboxiterminais das fitas- β 4 (ácido/base) e 7 (nucleófilo). Além dos resíduos catalíticos, o sítio ativo é formado por alguns resíduos também conservados como W313, H198, Y200 e R46⁷⁶. A fenda catalítica é formada geralmente por alças e outros elementos de estrutura secundária que estão conectados às extremidades carboxiterminais das fitas- β que

formam o barril. Esses elementos que formam a fenda são variáveis dentre os membros da família GH5.

1.3.4 GH1

De forma geral, beta-glicosidases são enzimas que hidrolisam ligações glicosídicas liberando resíduos glicosil de terminais não redutores de glicosídeos e oligossacarídeos. Essas enzimas são encontradas em todos os reinos dos seres vivos, Archea, Eubactérias e Eucariotos, sendo empregadas em variadas funções como conversão de biomassa, quebra de glicolipídeos e glicosídeos, catabolismo de parede celular, defesa, entre outras⁷⁷.

Apesar da breve definição, existe um grande número de moléculas na natureza que possuem diferentes resíduos glicosil em terminais não redutores, fazendo com que as beta-glicosidases sejam encontradas com diferentes números EC (Enzyme Commission). São exemplos as glucocerebrosidases (3.2.1.45), glucano 1,4- β -glicosidases (3.2.1.58), glucano 1,3- β -glicosidases (3.2.1.74), β -D galactosidases (3.2.1.23), β -D mannosidases (3.2.1.25), etc⁷⁷. A variabilidade das beta-glicosidases também é observada na classificação das hidrolases de glicosídeos, possuindo membros nas famílias GH1, GH3, GH5, GH9 e GH30, sendo que a maioria pertence ao Clã GH-A⁶⁵⁻⁷⁰.

Dentre as famílias, a GH1 é a que possui o maior número de membros. Suas enzimas são encontradas em todas as classes de organismos e exercem variadas funções relacionadas ao metabolismo de glicolipídeos em mamíferos, digestão em insetos, metabolismo da parede celular em plantas, patogenicidade de microorganismos, etc⁷⁷.

As GH1, como as GH5, pertencem ao Clã GH-A e, portanto, possuem o mesmo enovelamento barril TIM (β/α)₈, com os resíduos catalíticos situados nas extremidades carboxiterminais das fitas- β 4 (ácido/base) e 7 (nucleófilo). O mecanismo de retenção é conservado na família e também similar à família GH5 (figura 4). O tamanho das enzimas GH1 varia consideravelmente, dependendo da presença de domínios auxiliares e de outros domínios GH1 na mesma cadeia polipeptídica. Considerando somente o domínio catalítico, são encontradas enzimas constituídas de 440 a 550 resíduos, de acordo com o comprimento das alças e a presença de outros pequenos elementos de estrutura secundária conectados às extremidades carboxiterminais das fitas- β do barril TIM⁷⁸.

Embora os resíduos responsáveis pelo mecanismo hidrolítico estejam bem caracterizados, a maneira como as beta-glicosidases reconhecem e interagem com seus substratos, o que na maioria das vezes determina a diversidade de funções, continua não sendo tão clara⁷⁷.

As GH1 podem ser pouco específicas, onde apenas uma enzima hidrolisa β -D-glicosídeos, β -D-fucosídeos, β -D-mannosídeos, β -D-galactosídeos e α -L-arabinosídeos, ou clivar apenas um tipo de substrato. Embora exista grande quantidade de informação estrutural, resta ainda ser determinado como as enzimas GH1 podem ser preferencialmente beta-glicosidases ou beta-manosidases, ou ainda apresentar atividade para diferentes substratos, ainda que os resíduos utilizados sejam conservados⁷⁷⁻⁷⁹⁻⁸⁰⁻⁸¹.

1.3.5 CBM

Os carboidratos são moléculas envolvidas em muitos processos nos organismos vivos como metabolismo, suporte estrutural, acúmulo de energia, etc. Para lidar com essa multiplicidade de funções, muitas enzimas atuantes em carboidratos adquiriram um módulo não catalítico que interage especificamente com mono, oligo e polissacarídeos. Em geral, esses domínios de ligação a carboidratos (CBM) são módulos auxiliares com enovelamento autônomo com poder de reconhecimento de diferentes arranjos de carboidratos⁸².

O papel principal dos CBMs é reconhecer e se ligar especificamente a carboidratos. As consequências biológicas desse evento resultam em diferentes funções como o aumento da hidrólise de substratos insolúveis, redução da distância do domínio catalítico ao substrato e rompimento da estrutura de polissacarídeos⁸³.

Os CBMs podem ser encontrados em todos os grupos de seres vivos, em proteínas que reconhecem polissacarídeos como celulose, quitina, β -glicanos, amido, glicogênio, xilano arabinofurano, lactose, etc⁸³. Ocasionalmente, os CBMs também podem ser encontrados isolados, como uma única proteína, como a proteína CBP21 de *Serratia marcesens*, que tem afinidade por quitina⁸⁴, e as proteínas E7 e E8 de *Thermobifida fusca*⁸⁵.

Nas CAZimas como as hidrolases de glicosídeos ou transferases de glicosídeos os CBMs podem estar localizados nas porções amino e/ou carboxiterminal⁸⁶, ou entre dois módulos catalíticos⁸⁷, como uma única unidade ou arranjadas em tandem. Esses CBMs podem ser idênticos ou homólogos como nas α -amilases de lactobacilos⁸⁸, mas em alguns casos os

CBMs em tandem podem incluir módulos com diferentes especificidades por substratos. Por exemplo, a α -sialidase de *Clostridium perfringens* possui dois CBMs, um que se liga a galactose e outro a ácido siálico⁸⁹. Outro exemplo é a xilanase 5 de *Paenibacillus sp*, flanqueada por dois CBMs que ligam xilano e celulose separadamente⁹⁰.

Os CBMs estão classificados CAZy de acordo com a similaridade de suas sequências de aminoácidos. Atualmente existem 55 famílias, sendo que a mais numerosa é a família 48, cujos membros têm afinidade por glicogênio⁶⁴. Das 55 famílias, 40 já possuem membros com a estrutura tridimensional determinada e, embora apresentem diversas funções e pertençam a grupos taxonômicos distintos, é observado um enovelamento bem comum. O elemento conformacional mais importante é a folha- β . As arquiteturas formadas por essas folhas- β foram classificadas em sete famílias, sendo o Sanduiche- β o mais recorrente⁹¹. A topologia dos sítios de ligação dos CBMs, que está relacionada com a classe de substratos que eles reconhecem, também tem sido usada como critério para classificá-los em três tipos diferentes⁸² (figura 5). O tipo A possui uma superfície plana hidrofóbica composta de resíduos aromáticos que se ligam às superfícies planas dos polissacarídeos cristalinos, como celulose ou quitina. Em contraste, o sítio de ligação do tipo B, que se liga a celulose amorfa ou xilano, exibe uma fenda em que resíduos aromáticos interagem com cadeias polissacarídicas livres. Por último, o tipo C liga-se apenas a monossacarídeos, dissacarídeos ou trissacarídeos por restrições estéricas causadas pelo sítio de ligação.

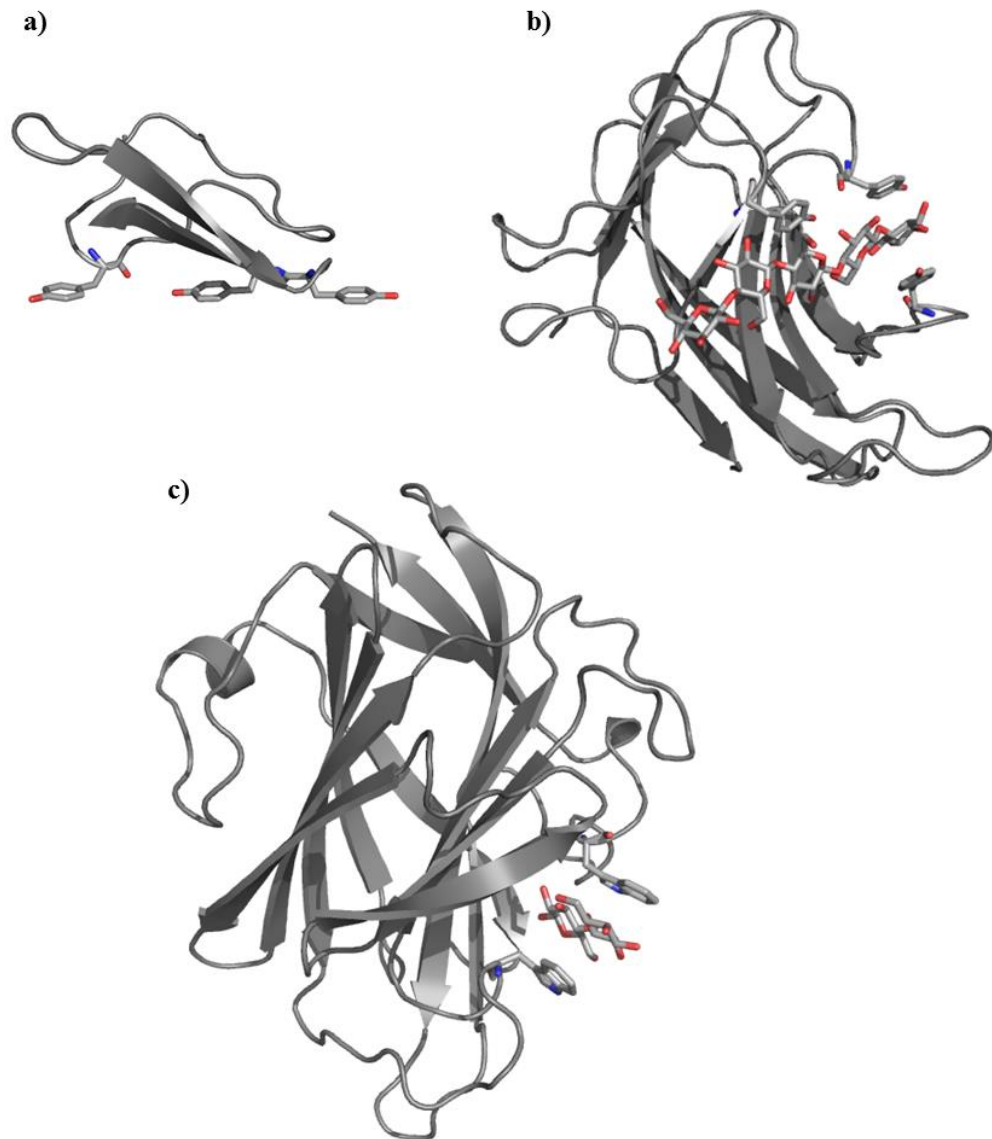


Figura 5 - Estrutura tridimensional de CBMs, representantes dos três grupos determinados pelo modo de ligação ao substrato. a) CBM tipo A de uma celobioidrolase de *T. reesei*⁹²; b) CBM tipo B de endoglucanase e *C. fini*⁹³; c) CBM tipo C de xilanase 10A de *T. maritima*⁹⁴. Os resíduos aromáticos e alguns dos substratos estão representados em formato *stick*.

Os resíduos envolvidos na ligação dos substratos têm sido identificados por NMR, mutagênese sítio dirigida e cristalografia⁹⁵⁻⁹⁶⁻⁹⁷⁻⁹⁸⁻⁹⁹. Em geral, é conhecida a importância dos resíduos aromáticos, especialmente os triptofanos. Eles formam interações de empilhamento (CH- π) com os anéis de açúcares resultando em uma forte ligação de Van der Waals, estabilizando assim a estrutura. Ademais, as cadeias laterais de outros aminoácidos polares podem formar ligações de hidrogênio com o carboidrato auxiliando na estabilização. Muitos estudos sugerem que a orientação dos resíduos aromáticos é responsável pela diferente especificidade das diferentes famílias de CBMs⁸³.

2 OBJETIVOS

Devido à grande demanda mundial por fontes de energia ambientalmente favoráveis, a produção de bioetanol através de biomassa lignocelulósica tornou-se uma excelente alternativa. Mas para que esse método se torne financeiramente viável, é de suma importância que as enzimas utilizadas neste processo sejam eficientes, resistentes e produzidas em grande quantidade.

Embora o mecanismo catalítico seja bem conhecido para as enzimas do Clã GH-A, os meios em que a especificidade para os substratos são estabelecidos têm se mostrado divergentes mesmo entre membros muito próximos. Este fato, juntamente com a alta diversidade de beta-glicosidases e endoglucanases encontradas, limitam as conclusões que podem ser feitas através de sequências genômicas sobre a especificidade e função dos novos membros. Todavia, as informações fornecidas por estudos estruturais podem dar um ponto de partida para essa compreensão. Conforme mais estruturas de enzimas com diferentes especificidades são determinadas, a predição da função de novas enzimas e o poder de melhoramento delas, através de engenharia de proteínas, vai se tornando mais plausível e aplicável.

Tendo por base essa necessidade, o presente trabalho teve como objetivo geral o estudo estrutural e funcional de celulasas de duas famílias muito representativas, GH1 e GH5.

Os objetivos específicos iniciais foram: clonagem e expressão das endoglucanases I (Eg1, família GH7) e II (Eg2, família GH5) de *T. reesei*; caracterização enzimática da Eg2 e comparação com a já caracterizada Eg1; cristalização e resolução da estrutura tridimensional de Eg2.

Posteriormente novos objetivos específicos foram definidos: expressão e purificação de endoglucanases bacterianas da família GH5 e beta-glicosidases bacterianas da família GH1 (tabela 2); cristalização e resolução da estrutura tridimensional das celulasas expressas; caracterização funcional das enzimas que tiverem sua estrutura resolvida; estudo estrutural e funcional dos módulos que compõem as celulasas com a estrutura resolvida.

Tabela 2 - Relação das enzimas estudadas. Cada uma está designada por um código para facilitar seu reconhecimento. A tabela expõe a família pertencente, gene (código GenBank), organismo e o tamanho de cada enzima.

Código	CAZY	Organismo	GenBank	Nº de aminoácidos	Peso molecular (Da)
A9	GH1	<i>L. gasseri</i>	ABJ59607.1	465	68939
B9	GH1	<i>L. gasseri</i>	ABJ59900.1	475	69297,4
C6	GH1	<i>B. licheniformis</i>	AAU39684.1	489	70571,1
C9	GH1	<i>L. gasseri</i>	ABJ59750.1	484	69947,2
D5	GH1	<i>B. licheniformis</i>	AAU43027.1	469	67715,7
E5	GH1	<i>B. licheniformis</i>	AAU41434.1	471	69223,1
F5	GH1	<i>B. licheniformis</i>	AAU43012.1	472	68143,9
F9	GH1	<i>L. gasseri</i>	ABJ61042.1	492	71594,7
G5	GH1	<i>B. licheniformis</i>	AAU39345.1	478	69748,6
G9	GH1	<i>L. gasseri</i>	ABJ59596.1	496	71232,4
H5	GH1	<i>B. licheniformis</i>	AAU42981.1	478	70802,3
F1	GH5	<i>H. turkmenica</i>	ADB63457.1	443	62466,1
F4	GH5	<i>P. chrysosporium</i>	AAU12275.2	369	53178,6
B5	GH5	<i>B. licheniformis</i>	AAU40778.1	364	56588,3
F6	GH5	<i>B. licheniformis</i>	AAU23613.2	485	68013,9
H6	GH5	<i>B. licheniformis</i>	AAU40777.1	533	74987,8
F7	GH5	<i>B. adolescentis</i>	BAF39985.1	402	59975,9
E12	GH5	<i>X. fastidiosa</i>	AAF85505.1	356	53596,9
F12	GH5	<i>X. fastidiosa</i>	AAF83620.1	565	75944

3 MATERIAIS E MÉTODOS

3.1 Biblioteca de cDNA de *T. harzianum*.

A extração de RNA foi feita a partir de células de *T. harzianum* cultivadas na presença de celulose pela ex-aluna de mestrado Renata Alves. As células congeladas foram maceradas em nitrogênio líquido, homogeneizadas com reagente TriZol (Invitrogen) e o RNA foi extraído de acordo com as instruções do fabricante. Após quantificação por espectrofotometria e análise por eletroforese em gel de agarose, foi realizada a síntese de cDNA por meio de transcrição reversa a partir do RNA total utilizando o kit “First Strand cDNA Synthesis” (Fermentas).

Todo o processo foi feito com material estéril.

3.2 Clonagens dos genes Eg1 e Eg2 de *T. harzianum* para expressão em *P. pastoris*.

Os genes relativos às enzimas Eg1 e Eg2 foram amplificados por PCR, a partir da biblioteca de cDNA descrita no item 3.1, através dos oligonucleotídeos: 5' GAA CCT AGG ATA CAT CAT CAT CAT CAT CTG GTT CCG CGT GGA TCC CCA CAA CCG GGT ACC AGC ACC 3' e 5' GAT CAT CCT AGG TTA CGT AGT AGA CCC AAT GTC TCC CCA GCG 3' para Eg1 e 5' GAA CCT AGG CAT CAT CAT CAT CAT CAT CTG GTT CCG CGT GGA TCC CCA GGC ACT TGC GTT ACC TCG AAG 3' e 5' GAG CGT CCT AGG TTA CTT TCT TGC GAG ACA CGA GCT GAC 3' para Eg2. O oligonucleotídeo complementar à região aminoterminal dos genes possui também uma região codificante para cauda de histidinas e para o sítio de clivagem por trombina, além de sítio para a enzima de restrição AvrII. No oligonucleotídeo complementar à região carboxiterminal foi adicionado um códon de terminação e um sítio para a enzima de restrição AvrII.

Os oligonucleotídeos foram desenhados a partir das sequências gênicas depositadas sob o código GenBank M15665.1 para Eg1 e DQ178347.1 para Eg2. O domínio catalítico da Eg1 foi determinado por alinhamento da sequência de aminoácidos com a sequência da estrutura cristalográfica de endoglucanase I de *Trichoderma reesei* já determinada¹⁰⁰. O

domínio catalítico da Eg2 foi determinado de acordo com artigos publicados¹⁰¹⁻¹⁰². Portanto, a enzima Eg1 foi definida entre os resíduos 24 e 293, e a Eg2 entre os resíduos 102 a 418.

A reação de PCR foi feita de acordo com o protocolo descrito pelo manual da enzima Phusion High-Fidelity DNA Polimerase (New England Biolabs).

Depois de amplificados, os genes foram ligados ao vetor pGEM-T (figura 6) e foi feita a transformação de células de *E. coli* (cepa DH5α) para multiplicação do vetor.

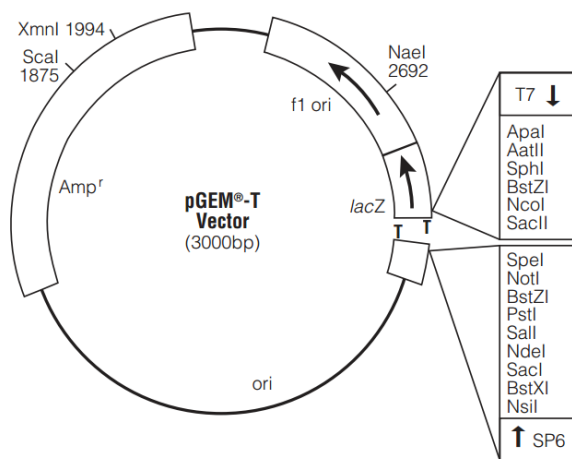


Figura 6 - Vetor pGEM-T utilizado para multiplicação dos genes Eg1 e Eg2 em *E. coli*.

3.2.1 pPIC9K.

Os genes foram extraídos do vetor pGEM-T por clivagem com a enzima AvrII e ligados ao vetor pPIC9K (figura 7), também clivado pela enzima AvrII, e o vetor final inserido em células de *E. coli* (cepa DH5α) para sua propagação. As colônias geradas tiveram o vetor extraído e a presença do gene e sua orientação foram confirmadas por teste de restrição com as enzimas NcoI e SalI para os genes Eg1 e Eg2, respectivamente.

O vetor correto foi então linearizado pela enzima de restrição SacI e utilizado na transformação de células de *P. pastoris*.

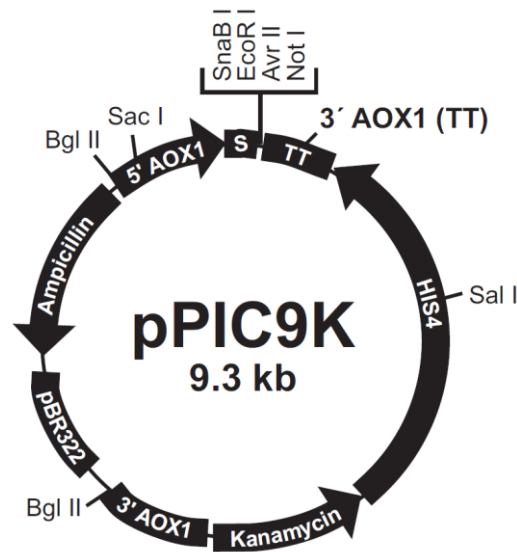


Figura 7 - Vetor pPIC9K utilizado para expressão induzida por metanol das enzimas Eg1 e Eg2 em *P. pastoris*.

3.2.2 pPZa.

Os genes de Eg1 e Eg2 também foram inseridos em vetor pPZa. Este vetor é uma modificação do pPICZa (figura 8), feita pelo grupo do professor Fernando Araripe Torres do Laboratório de Biologia Molecular da Universidade de Brasília, em que o promotor AOX foi substituído pelo promotor constitutivo PGK1. Neste processo, os genes foram separados do vetor pPIC9K pela clivagem com as enzimas NotI e BstBI e ligados no vetor pPZa, que também foi clivado pelas mesmas enzimas.

O vetor final foi inserido em células de *E. coli* (cepa DH5a) para sua propagação. As colônias geradas tiveram o plasmídeo extraído e a presença do gene foi confirmada por teste de restrição com as enzimas NotI e BstBI.

Da mesma forma descrita no item anterior, o plasmídeo final foi linearizado pela enzima de restrição SacI e utilizado na transformação de células de *P. pastoris*.

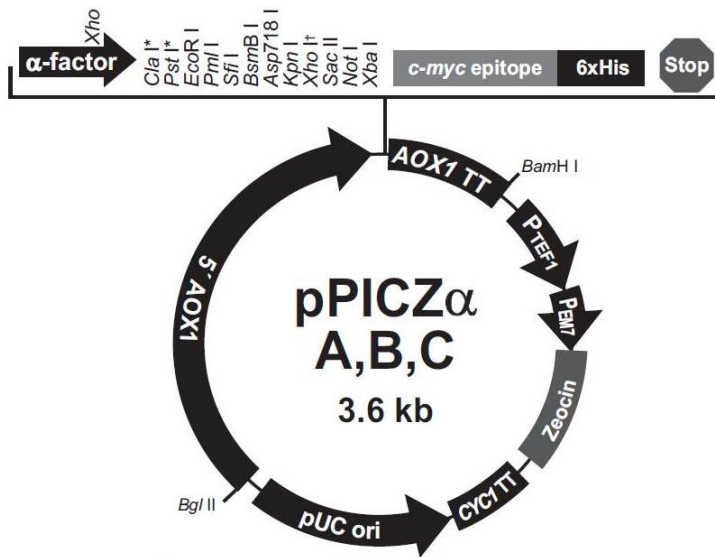


Figura 8 - Vetor pPICZ α utilizado para expressão autoinduzida das enzimas Eg1 e Eg2 em *P. pastoris*.

3.3 Clonagem do gene Eg2 de *T. harzianum* para expressão em *E. coli*.

Os genes das enzimas Eg1 e Eg2 foram amplificados da mesma forma descrita no item 2.2 através dos oligonucleotídeos: 5' CAT ATG GGC ACT TGC GTT ACC TCG AAG 3' e 5' CTC GAG TTA CTT TCT TGC GAG ACA CGA GCT 3'. Ao oligonucleotídeo complementar à região aminoterminal, foi adicionado um sítio para a enzima de restrição NdeI. Ao oligonucleotídeo complementar à região carboxiterminal, foi adicionado um códon de terminação e um sítio para a enzima de restrição XhoI. A reação de PCR foi feita de acordo com o protocolo descrito pelo manual da enzima Phusion High-Fidelity DNA Polimerase (New England Biolabs).

Depois de amplificados, os genes foram ligados ao vetor pET28a (figura 9), também clivado pelas enzimas NdeI e XhoI, e o vetor final inserido em células de *E. coli* (cepa DH5 α) para sua propagação. As colônias geradas tiveram o vetor extraído e a presença do gene foi confirmada por teste de restrição com as enzimas NdeI e XhoI.

O vetor final foi utilizado na transformação de células de *E. coli* (cepa B121 Rosetta 2) para expressão da enzima.

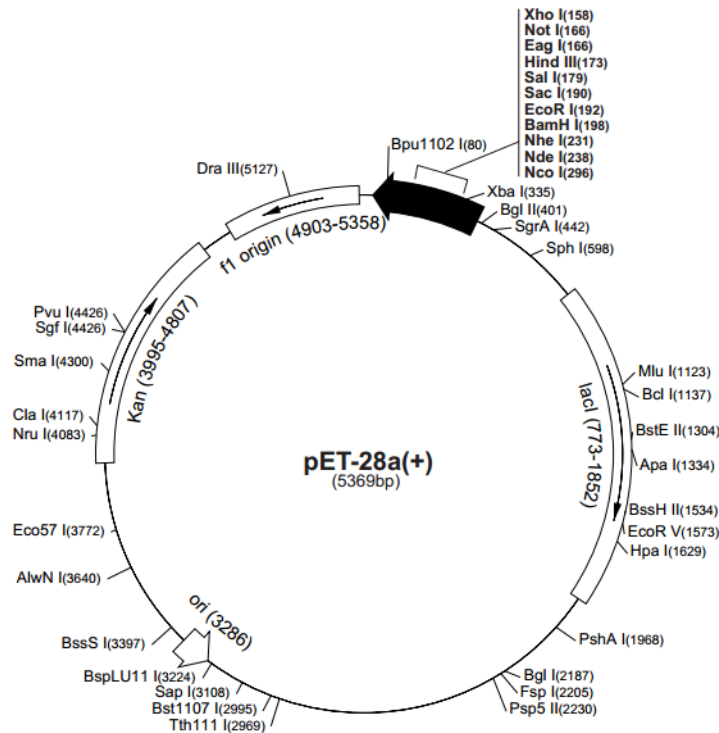


Figura 9 - Vetor pET28a utilizado para expressão da enzima Eg2 em *E. coli*.

3.4 Clonagem dos genes das famílias GH1 e GH5 para expressão em *E. coli*.

Os genes que codificam as proteínas das famílias GH1 e GH5 foram clonados pelo pós-doutorando César Moisés Camilo através do método de Clonagem Independente de Ligação (LIC)¹⁰³. Graças ao método empregado, não é necessária a utilização de enzimas de restrição nem de ligação, o que dá nome à técnica. Para uma descrição detalhada verifique o artigo de Aslanidis¹⁰⁴. Os genes foram inseridos em vetor pET-Trx, modificado e gentilmente cedido pelo Dr. Arie Geerlof (EMBL, Hamburgo - Alemanha). Esse vetor, além de estar preparado para a clonagem pelo método LIC, também acrescenta à proteína alvo um domínio tiorredoxina e uma cauda de histidinas, além de dar resistência à canamicina para a célula bacteriana. O domínio tiorredoxina, pelo seu poder redutor, pode ajudar a evitar agregações causadas por ligações dissulfeto entre as enzimas. A cauda de histidina permite a purificação por afinidade.

Os clones das duas construções truncadas da enzima H6 também foram gerados pelo pós-doutorando César Moisés Camilo através do método LIC. Os limites de cada construção foram determinados de acordo com a estrutura cristalográfica da construção completa. A

primeira construção, H6c1, foi definida entre os resíduos 1 a 432. A segunda construção, H6c2, foi definida entre os resíduos 1 a 335.

3.5 Preparação de células de *E. coli* competentes.

Para que as bactérias estejam aptas a receber o vetor, um tratamento das células se faz necessário. Um pré-inóculo de 5 mL é feito com meio Luria Broth (LB) sem a adição de antibióticos e incubado por 16 horas a 37 °C com rotação de 150 rpm. A partir dele, um inóculo é feito com 500 mL de meio LB e incubado a 37 °C com rotação de 150 rpm até atingir a densidade ótica a 600 nm (DO_{600}) de 0,3. A cultura então é centrifugada a 1600 g a 4 °C por 10 minutos. Em seguida, as células precipitadas são ressuspensas cuidadosamente em 100 mL de tampão A (60 mM de cloreto de cálcio, 10 mM de PIPES pH 7,0 e 15 % de glicerol) a 4 °C. Este passo é repetido mais uma vez e, por fim, após nova centrifugação, as células são ressuspensas em 12 mL de tampão A a 4 °C, aliquotadas em tubos de micro centrífuga e estocadas em freezer a -80 °C.

3.6 Transformação das bactérias.

Para a inserção dos vetores nas células bacterianas foi utilizada a transformação química por temperatura. Neste método, as células competentes são descongeladas a 4 °C e incubadas por 30 minutos com 10 ng do vetor a ser inserido. Após a incubação, é feito um choque térmico no qual as células são passadas para 42 °C em banho térmico por 1 minuto e novamente voltam ao gelo por 2 minutos. Em seguida, é adicionado às células 1 mL de meio LB que são incubados a 37 °C por uma hora em agitação de 150 rpm. Após a incubação, as células são espalhadas em placas de LB/ágar contendo o antibiótico que o vetor inserido confere resistência. Por fim, as placas são mantidas a 37 °C por 16 horas.

3.7 Preparação de células de *P. pastoris* competentes.

Foi feito pré-inóculo das colônias de *P. pastoris* em 10 mL de meio YPD (1 % de extrato de levedura, 2 % de peptona e 2 % de glicose, relação peso/volume) a 30 °C por 16 horas. Em seguida, inoculou-se 500 mL de meio YPD, deixando-se crescer a 30 °C até a DO₆₀₀ de 1,2 a 1,5. A cultura foi então centrifugada a 1500 g por 5 minutos a 4 °C e ressuspensa em 500 mL de água estéril a 4 °C. O passo anterior foi repetido com ressuspensão em 250 mL de água. Mais uma vez foi feita a centrifugação e as células foram ressuspensas em 20 mL de solução de sorbitol 1M a 4 °C. Por último as células são novamente centrifugadas e ressuspensas em 1 mL de solução de sorbitol 1M a 4 °C. As células eram, então, separadas em alíquotas de 40 µL e utilizadas no mesmo dia para transformação.

Todos os passos foram feitos com material estéril e dentro de um fluxo laminar. Após o crescimento, todo o processo foi feito a baixa temperatura mantendo-se todas as soluções em gelo.

3.8 Transformação de *P. pastoris*.

As linhagens de *P. pastoris* utilizadas foram a GS-115 (Invitrogen) para transformação com o plasmídeo pPIC9K e a X33 (cedida pelo professor Fernando Araripe Torres) para transformação com o plasmídeo pPZa.

As alíquotas de células recém-preparadas para receber os vetores eram misturadas com 10 µg do plasmídeo linearizado e transferidas para uma cubeta de eletroporação de 0,2 cm gelada. Após incubação por 5 minutos, foi dado pulso elétrico com eletroporador (BioRad) ajustado a 25 µF, 200 ohm e 2000 V. Imediatamente as células foram diluídas em 1 mL de solução de sorbitol 1M a 4 °C e incubou-se a 30 °C por 2 horas sob agitação.

Nos casos em que foi utilizado o plasmídeo pPIC9K, logo após a incubação, as células eram espalhadas em placas contendo meio MD (1,34 % de *Yeast Nitrogen Base* sem aminoácidos, 4×10^{-5} % biotina e 2% de glicose, relação peso/volume) sólido e diferentes concentrações do antibiótico Geneticina (de 0 a 4 mg/mL) e incubadas em estufa a 30 °C até o aparecimento das colônias.

Com a utilização do plasmídeo pPZa, as células eram espalhadas em placas contendo meio YPDS (1 % de extrato de levedura, 2 % de peptona, 2 % de glicose e 0,2 % de sorbitol, relação peso/volume) contendo diferentes concentrações do antibiótico Zeocina (de 25 a 250 µg/mL).

3.9 Teste de expressão em *E. coli*.

A capacidade de expressão das proteínas pelas colônias geradas com a transformação de *E. coli* foi testada inoculando-se 5 mL de meio LB, acrescido dos antibióticos adequados, e incubando-as por 12 horas a 37 °C sob agitação de 150 rpm. Após o crescimento a expressão foi induzida com a adição de 1 mM de isopropil-β-D-1-tiogalactosídeo (IPTG) ao meio e incubado por 6 horas a 37 °C sob agitação de 150 rpm. Em seguida as células foram centrifugadas a 5000 g a 4 °C por 20 minutos. As células precipitadas foram então ressuspensas em 200 µL de tampão de ressuspensão (100 mM de cloreto de sódio, 50 mM de Tris-HCl pH 8,0, 10 % glicerol, 1 mM de inibidor de protease Phenylmethanesulfonyl Fluoride (PMSF) e 250 µg/mL de lisozima) e incubadas em gelo por 30 minutos. A lise das células foi feita por choque térmico, congelando-as com nitrogênio líquido e descongelando-as a 4 °C. Após a lise, as amostras eram centrifugadas a 15000 g a 4 °C por 10 minutos, os sobrenadantes eram coletados e aplicados em gel dodecil-sulfato de sódio (SDS) para análise das proteínas expressadas. As colônias que apresentaram expressão solúvel eram então congeladas a -80 °C como estoque para expressão em maior escala. Em relação às enzimas que não se apresentavam solúveis no teste de expressão, novas colônias eram testadas com variações no protocolo.

3.10 Teste de expressão em *P. pastoris*.

As colônias que cresceram em placas com as maiores concentrações de antibióticos e algumas com menor concentração foram selecionadas para teste de expressão.

Com cada colônia era feito um pré-inóculo de 5 mL em meio BMGY (1,34 % de *Yeast Nitrogen Base* sem aminoácidos, 4×10^{-5} % de biotina, 2% de peptona, 1 % de extrato de

levedura, 1 % de glicerol e 100 mM de fosfato de potássio pH 6,0) e incubado a 30 °C sob agitação de 250 rpm até atingir uma DO₆₀₀ de 2-4. Após o crescimento, as células eram centrifugadas a 2000 g a temperatura ambiente e ressuspensas em 10 mL de meio indutor BMMY (1,34 % de *Yeast Nitrogen Base* sem aminoácidos, 4×10^{-5} % de biotina, 2% de peptona, 1 % de extrato de levedura, 0,5 % de metanol e 100 mM de fosfato de potássio pH 6,0). A nova cultura voltava a ser incubada com as mesmas condições por 4 dias. A cada dia era adicionado 50 µL de metanol 100 % (concentração final de 0,5 %) para a indução ser mantida. Durante os quatro dias foram retiradas alíquotas, que foram analisadas por gel SDS e pela atividade.

Visando uma boa aeração, os cultivos foram feitos em frascos com dez vezes o volume do meio de cultura. Outros testes de expressão foram feitos variando-se as condições de cultivo, como temperatura, pH, meio de cultura, etc.

3.11 *Western Blotting.*

Após a eletroforese em gel SDS, as proteínas foram transferidas para membrana de nitrocelulose *Hybond ECL* (Amersham Biosciences) usando um tampão preparado com 500 mM de Tris-HCl, 1 M de glicina, 0,1 % de SDS, 10 % de metanol e pH ajustado a 8,3. A membrana foi bloqueada durante 16 horas a 4 °C com solução de leite desnatado 5 % em PBS-Tween 0,1 % e incubada com anticorpo primário anti-His (Sigma) diluído em PBS (1:1000) por 2 horas. Em seguida, a membrana foi três vezes lavada com PBS-Tween 0,1 % por 10 minutos. Após as lavagens, a membrana foi incubada com anticorpo secundário anti-camundongo conjugado com Peroxidase (Sigma) na diluição 1:5000 por 1 hora a temperatura ambiente. Foram repetidas as lavagens com PBS-Tween e a membrana foi revelada com o reagente *AP Conjugate Substrate* (Bio-rad) segundo instruções do fabricante.

3.12 Expressão em *E. coli*.

A expressão das proteínas para os estudos era feita da mesma maneira descrita no item 3.9, porém utilizando maior volume de meio de cultura e diferente método de lise.

Considerando que cada proteína expressa heterologamente possui um rendimento diferente, os volumes de meio de cultura variavam entre 1 L e 5 L, dependendo da quantidade de expressão de cada proteína.

A lise das células era feita por sonicação. Aplicavam-se 5 ciclos de 10 segundos de pulsos intercalados com 50 segundos de descanso em gelo. O sobrenadante, que continha a proteína, era então coletado e purificado. Alíquotas da fração sobrenadante e do precipitado eram separadas para posterior análise em gel SDS.

3.13 Purificação por afinidade e clivagem da cauda de histidinas.

As enzimas expressas continham, devido ao vetor utilizado, uma cauda de seis histidinas localizada na sua porção amino terminal, permitindo assim sua purificação por cromatografia de afinidade. A resina utilizada foi a Talon Superflow (Clontech) contendo íons de níquel imobilizados.

O sobrenadante coletado depois da lise das células era tratado com DNase, para reduzir a viscosidade, e aplicado na coluna de purificação contendo de 2 a 5 mL de resina. Uma vez que a cauda de histidinas tem afinidade pelos íons níquel, a enzima se fixava na resina enquanto todas as outras passavam diretamente. A resina era então lavada com dez vezes o volume de coluna com tampão de lavagem (100 mM de cloreto de sódio, 50 mM de Tris-HCl pH 8,0 e 5 mM de imidazol) e eluída com 5 vezes o volume de coluna com tampão de eluição (100 mM de cloreto de sódio, 50 mM de Tris-HCl pH 7,0 e 200 mM de imidazol).

Uma vez pura, a enzima não precisava mais da cauda de histidinas, nem do domínio tiorredoxina. Para a retirada dessas regiões, a fração eluída era acrescida de 1 mg da protease Tobacco Etch Virus (TEV) e incubada por 12 horas a 4 °C. Como o imidazol reduz a atividade da TEV, foi necessária a diluição da enzima por dez vezes em tampão de eluição sem imidazol, antes da aplicação da protease.

Após a clivagem, a proteína era novamente aplicada na coluna de afinidade para que, desta vez, a cauda ficasse retida e ela passasse diretamente. A enzima era então dialisada com tampão mínimo (50 mM de cloreto de sódio e 20 mM de Tris-HCl pH 7,0) e concentrada para posteriores estudos.

Todos os tampões utilizados eram mantidos a 4 °C, apesar da purificação ter sido feita a temperatura ambiente. Em todos os passos foram retiradas alíquotas para análise em gel SDS.

3.14 Produção e purificação da protease TEV.

A produção da protease TEV, utilizada na clivagem da cauda de histidinas e da tiorredoxina, foi feita utilizando-se o mesmo protocolo descrito nos itens 2.11 e 2.12. O plasmídeo de expressão da protease TEV foi gentilmente cedido pela ex-aluna de doutorado Ana Cristina Puhl.

3.15 Purificação por exclusão de tamanho.

Em alguns casos, a purificação por afinidade não foi suficiente para obtenção de proteína pura, apresentando alguns poucos contaminantes com diferentes pesos moleculares. Nesses casos, foi utilizada a cromatografia por exclusão de tamanho, com a coluna Superdex 200 HR 10/30 (GE Healthcare).

Com o auxílio de um aparelho de *High-performance liquid chromatography* (HPLC), a resina era equilibrada com tampão mínimo, e volumes de 250 µL de enzima concentrada eram aplicados. O ajuste de parâmetros, como fluxo e pressão máxima, foi determinado de acordo com o manual da coluna. Os picos observados no cromatograma eram coletados em alíquotas de 500 µL e analisados por gel SDS. As alíquotas que apresentavam a enzima pura eram concentradas para posteriores estudos.

Todas as purificações por exclusão de tamanho foram feitas à temperatura ambiente.

3.16 Teste de termoestabilidade.

A termoestabilidade, além de uma medida física da capacidade da enzima de se manter enovelada corretamente em determinada temperatura, passou a ser uma medida valiosa para determinação de condições com maior probabilidade de cristalização de proteínas.

A solução em que a proteína se encontra, antes do experimento de cristalização, pode ser crucial para a obtenção de cristais. Baseados nesse fato, dois artigos publicados em 2006¹⁰⁵ e 2011¹⁰⁶ apresentaram um método rápido de análise da termoestabilidade das proteínas produzidas em diferentes condições. Mostraram também que as proteínas mais termoestáveis têm maior probabilidade de gerar cristais.

Os testes de termoestabilidade foram feitos com a ajuda do técnico Philippe Mas no Laboratório Europeu de Biologia Molecular (EMBL) de Grenoble, França.

Em uma placa de PCR de 96 poços, as enzimas foram diluídas a 20 μ M em solução contendo 5x SYPRO Orange (Invitrogen). As amostras foram então induzidas à desnaturação térmica em um PCR quantitativo em tempo real por gradiente de temperatura de 25 a 75 °C em passos de 1 °C e 1 minuto. O desenovelamento da proteína foi monitorado pelo aumento da intensidade da fluorescência da sonda SYPRO Orange, que era medida a cada minuto utilizando excitação a 492 nm e emissão a 516 nm.

A intensidade relativa de emissão de fluorescência era plotada como uma função da temperatura e o T_m (temperatura na qual metade das enzimas encontra-se estável) de cada amostra era calculado como a média entre a linha de base e o ponto com maior intensidade de fluorescência.

A molécula de SYPRO Orange tem afinidade por regiões hidrofóbicas das proteínas e, quando interagem, aumentam a intensidade da fluorescência emitida a 516 nm. Desta forma, conforme as proteínas vão perdendo seu enovelamento, as regiões hidrofóbicas que estão interiorizadas se expõem e mais moléculas de Sypro vão se ligando e aumentando a intensidade da emissão de fluorescência.

3.17 Determinação da atividade.

Na determinação da atividade das enzimas foi utilizado como substrato padrão a carboximetilcelulose (CMC). Outros substratos também foram utilizados para teste de especificidade, como beta-manano, avicel, arabinano, galactomanano, lichenano, beta-glicano e xiloglucano.

O teste era feito incubando-se 50 μ L de reação contendo 1 % (relação massa/volume) de substrato, 50 mM de agente tamponante (de acordo com o pH definido), 100 mM de cloreto de sódio e a enzima, cuja concentração era determinada empiricamente para que se pudesse ter uma boa medida no final da reação¹⁰⁷. Considerou-se boa medida aquela compreendida entre o branco e o valor máximo representado na porção linear da curva padrão feita com diferentes concentrações de glicose.

A incubação era feita tubos de PCR, e o termociclador era utilizado para que a temperatura de reação escolhida fosse mantida constante.

O tempo de reação também era determinado empiricamente levando-se em conta uma boa medida no final da reação e a manutenção da velocidade inicial da reação. Tempos muito longos fazem com que o substrato se extinga, a velocidade de reação diminua e a determinação da atividade seja alterada.

A quantidade de açúcares redutores liberados na reação foi determinada pelo método colorimétrico do ácido 3-5-dinitrossalicílico (DNS)¹⁰⁸ e a absorbância medida em espectrofotômetro a 540 nm. Como a solução de DNS é extremamente básica, ela mesma era utilizada para parar a reação adicionando-se 100 μ L à mesma.

A unidade de atividade enzimática (U) foi definida como: a quantidade de enzima capaz de liberar 1 μ mol de açúcares redutores (a partir do substrato) por minuto. A atividade específica foi determinada como o número de unidades U por miligrama de enzima.

O mesmo método foi utilizado variando-se a temperatura (de 30 a 90 °C) e o pH (de 3,0 a 7,0) das reações, determinando-se assim as condições ótimas para a clivagem dos substratos pelas enzimas.

3.18 Cristalização.

A cristalização foi feita concentrando-se as enzimas puras e dialisadas em tampão mínimo a 10 mg/mL. A concentração foi posteriormente modificada de acordo com os resultados obtidos nos primeiros testes. Como primeiros testes de cristalização, foi utilizada a técnica de difusão de vapor em gota sentada, com o auxílio do robô de cristalização Honeybee. Foram usadas placas de cristalização de 96 poços em que a gota era formada por 1 μ L da solução de enzima com 1 μ L da solução cristalizante. O poço era preenchido com 80 μ L da solução cristalizante. Após vedar a placa, ela era mantida a temperatura constante de 18 °C. Nesses testes iniciais, as soluções de cristalização provinham de kits de cristalização como “Crystal Screen” e “Index” da Hampton, e “Classics Suite” e “MbClass Suite” da Qiagen.

Uma vez encontradas condições promissoras, elas eram refinadas manualmente utilizando-se placas de 24 ou 48 poços com a técnica de difusão de vapor em gota suspensa. As gotas eram formadas por 2 μ L da solução de enzima com 2 μ L da solução cristalizante e os poços eram preenchidos com 250 μ L das soluções cristalizantes.

Durante o Doutorado Sanduiche feito no *European Synchrotron Radiation Facility* (ESRF) em Grenoble, França, os experimentos iniciais de cristalização eram feitos pelo robô Cartesian PixSys 4200 (Genomic Solutions, Reino Unido), onde as gotas eram formadas por 0,1 μ L da solução de enzima com 0,1 μ L da solução cristalizante. Os kits de cristalização utilizados foram os mesmo dos experimentos feitos no Brasil.

3.19 Difração de raios-X.

Os cristais obtidos nos experimentos de cristalização eram transferidos para uma solução crioprotetora, feita com a mesma solução cristalizante acrescida de 15 a 20 % de etilenoglicol ou glicerol, e congelados em nitrogênio líquido. No Brasil, os cristais eram testados em experimentos de difração de raios-x na linha MX-2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP), equipada com detector CCD modelo MARMOSAIC 225. Na França, foram testados na linha BM23-1 do ESRF, equipada com detector PILATUS

6M-F. Os cristais que apresentavam bom poder de difração tinham seu conjunto de dados coletados. Durante os experimentos, os cristais eram mantidos a temperatura de $-173\text{ }^{\circ}\text{C}$.

3.20 Tratamento dos dados de difração de raios-X.

Os dados coletados foram indexados e integrados pelo programa XDS¹⁰⁹ e as intensidades das reflexões foram escalonadas pelo programa Aimless¹¹⁰⁻¹¹¹. O número de moléculas na unidade assimétrica foi estimado pelo programa Mathews_coef¹¹²⁻¹¹³. As fases iniciais foram determinadas com o programa Phaser¹¹⁴ por substituição molecular. Os modelos iniciais utilizados foram encontrados no *Protein Data Bank* (PDB) através de procura por sequências com alta identidade. Os modelos gerados foram aperfeiçoados através de ciclos de refinamento com o programa PHENIX¹¹⁵, intercalados com análises e ajustes manuais pelo programa COOT¹¹⁶. As figuras das estruturas resolvidas foram geradas com o programa PYMOL.

ENDOGLUCANASE I E ENDOGLUCANASE II DE *T. reesei*.

4.1 RESULTADOS

4.1.1 Clonagem de Eg1 e Eg2 para expressão em *P. pastoris* pelo vetor pPIC9K.

Os genes foram amplificados por PCR a partir de uma biblioteca de cDNA e foram confirmados por gel de agarose a 1 %. As bandas referentes aos genes que codificam as enzimas Eg1 (1243 pb) e Eg2 (987 pb) apresentaram-se únicas e com tamanho esperado (figura 10). Os fragmentos foram subclonados no plasmídeo pGEM-T e utilizados na transformação de células de *E. coli* (DH5 α) competentes para propagação.

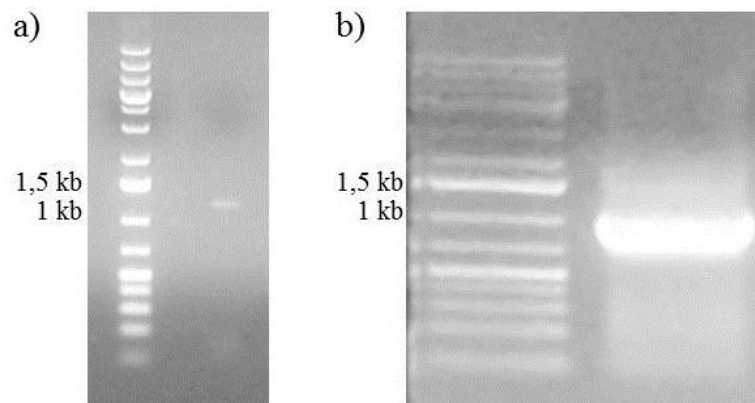


Figura 10 - Eletroforese em géis de agarose da amplificação por PCR dos genes: a) EgI e b) Eg2.

Os plasmídeos pGEM-T contendo os fragmentos que codificam as endoglucanases foram clivados com a enzima de restrição AvrII, assim como o vetor pPIC9K. Em seguida, os fragmentos liberados foram ligados em pPIC9k e utilizados na transformação de células de *E. coli* (DH5 α). Como os genes estão flanqueados pelo mesmo sítio de restrição da enzima AvrII, após a ligação foi feito um teste de restrição das diferentes colônias geradas com as enzimas NcoI para endoglucanase I e SalI para endoglucanase II para determinar qual colônia apresentava o plasmídeo com o gene inserido na orientação correta. No caso da Eg1, a

clivagem do vetor com gene na posição correta geraria fragmentos com tamanhos 7890 pb e 2610 pb, na posição invertida seriam 7124 pb e 3376 pb. Com a Eg2 os fragmentos do gene na orientação correta teriam 7611 pb e 2633 pb, e os fragmentos com o gene invertido teriam 8054 pb e 2190 pb (figura 11 e 12). Desta forma, as colônias escolhidas foram as colônias 3 tanto da endoglucanase I como endoglucanase II.

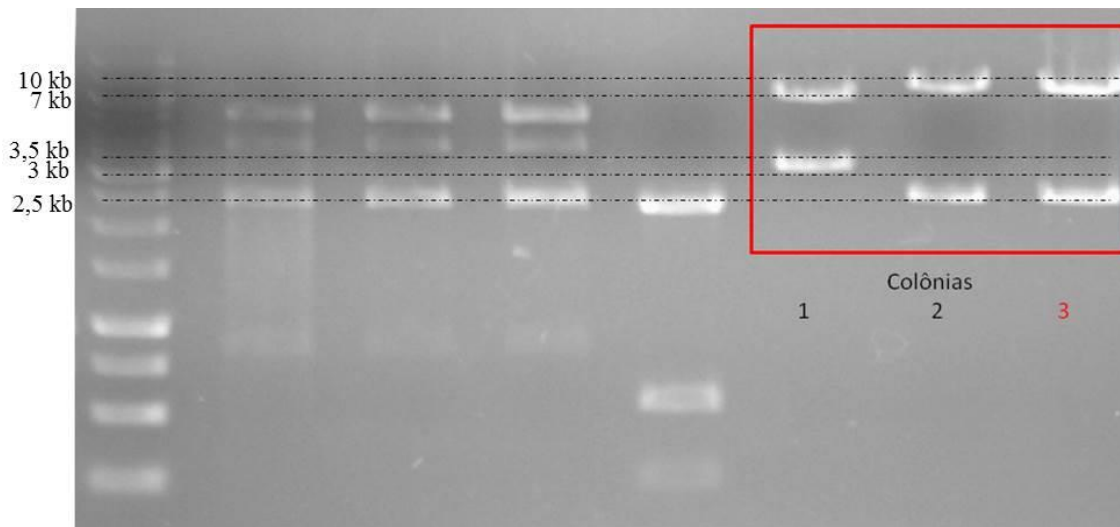


Figura 11 - Eletroforese em gel de agarose do teste de restrição de colônias transformadas com pPIC9K+Eg1. No interior do retângulo vermelho estão as bandas relativas aos fragmentos gerados pela clivagem com a enzima de restrição NcoI.

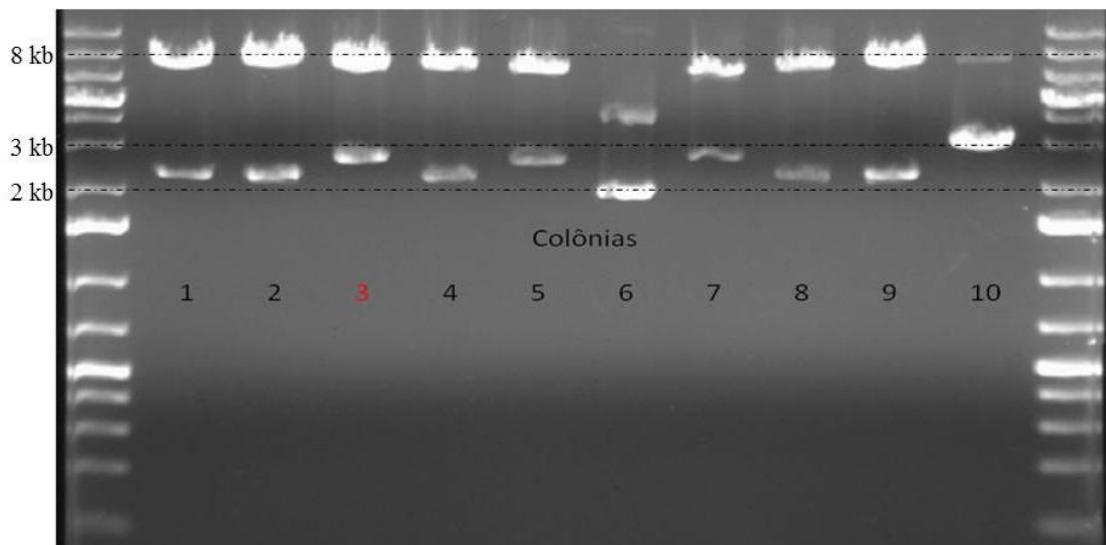


Figura 12 - Eletroforese em gel de agarose do teste de restrição de colônias transformadas com pPIC9K+Eg2. Os plasmídeos foram clivados com a enzima de restrição SalI.

As colônias selecionadas foram cultivadas e grande quantidade de plasmídeo foi extraído e purificado com o kit “PureYield Plasmid Midiprep System” (Promega).

Para a transformação de *P. pastoris*, os plasmídeos recombinantes foram inicialmente linearizados com a enzima de restrição SacI (figura 13), de acordo com as recomendações do fabricante (Fermentas).

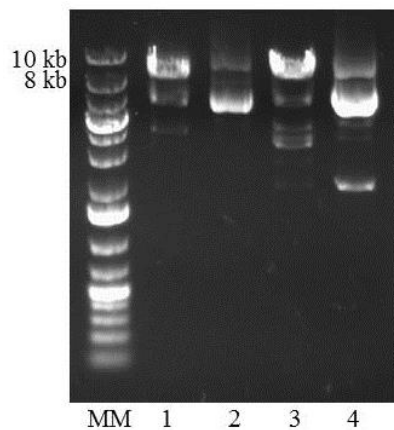


Figura 13 - Eletroforese em gel de agarose da linearização dos plasmídeos com a enzima de restrição SacI. MM. Padrão de peso molecular; 1. Eg1/pPIC9K clivado; 2. Eg1/pPIC9K não clivado; 3. Eg2/pPIC9K clivado; 4. Eg2/pPIC9K não clivado.

Após a eletroporação, as células foram espalhadas em placas com meio MD. Como a linhagem GS-115 não possui o gene *HIS4*, essencial para a produção do aminoácido histidina, somente as células que obtiveram a inserção do plasmídeo pPIC9K crescem em meio mínimo deficiente em histidina. Para uma segunda etapa de seleção de transformantes, as colônias que cresceram em meio mínimo foram plaqueadas em meio YPD contendo concentrações crescentes do antibiótico geneticina. Foram selecionadas para testes de expressão, as colônias que cresceram na concentração máxima de geneticina de 4 mg/mL, como indicado pelo fabricante (Invitrogen), na qual se espera que as células possuam até 12 cópias do gene inserido no genoma da levedura. Como controle, células de *P. pastoris* GS-115 não transformadas também foram plaqueadas nas mesmas condições descritas acima e, como esperado, nenhuma colônia cresceu.

4.1.2 Expressão de Eg1 e Eg2 em *P. pastoris* pelo vetor pPIC9K.

Antes de começar com os testes de expressão as colônias transformadas foram replicadas em novas placas e numeradas para facilitar a identificação. Também foram feitos estoques em glicerol de cada colônia testada e congelados a -80 °C.

Para identificação das bandas relativas às enzimas por SDS-PAGE, foi feito o cálculo de peso molecular de Eg1 e Eg2 através do programa online Protparam¹¹⁷, resultando em 41,06 kDa e 35,27 kDa respectivamente. Porém, as enzimas possuem sítios de glicosilação, confirmados através dos programas NetNGlyc¹¹⁸ e NetOGlyc¹¹⁹, ficando impossível prever o aumento do peso molecular gerado pela adição de carboidratos. De acordo com o artigo de Qin¹⁰², a construção de endoglucanase II de *T. reesei* possui 48 kDa quando expressa pelo próprio fungo e 57 kDa quando expressa heterologicamente em *S. cerevisiae*.

Como o plasmídeo pPIC9K confere uma região N-terminal sinalizadora para exportação (fator- α), a enzima expressa deve ser encontrada no meio extracelular. Desta forma, as análises por SDS-PAGE foram feitas a partir do sobrenadante, após centrifugação da cultura.

O primeiro teste de expressão foi realizado com as colônias 1 a 14 de Eg2. As colônias foram induzidas em meio BMMY a 30 °C sob agitação de 250 rpm por 4 dias. Alíquotas foram coletadas e, após centrifugação, a fração do sobrenadante foi aplicada em gel SDS-PAGE. Como controle, foram testadas também colônias de *P. pastoris* não transformadas e colônias transformadas com o plasmídeo pPIC9k sem inserto.

Os géis SDS-PAGE do primeiro teste deixam claro que nenhuma banda mais intensa foi notada acima dos pesos moleculares mínimos esperados (figura 14). Mesmo bandas mais fracas não pareceram promissoras quando comparadas com o padrão de expressão da *P. pastoris* sem transformação (figura 14a) e da *P. pastoris* transformada com o plasmídeo sem inserto (figura 14d). As colônias 1, 3, 4 e 5 foram observadas após 24, 48, 72 e 96 horas de indução e não se nota praticamente nenhuma diferença entre elas (figuras 14a, 14b e 14c). O mesmo acontece com as demais colônias que foram analisadas apenas após 96 horas de indução (figuras 14d e 14e).

Para conferir se a proteína estava sendo expressa de forma intracelular, o precipitado das amostras da colônia 5 também foi analisado em gel SDS-PAGE. Na figura 14c (coluna 7) nota-se a presença de uma banda um pouco mais intensa do que as outras após a indução.

Porém essa banda está localizada numa posição próxima a uma banda de menor intensidade na colônia não transformada (figura 14a).

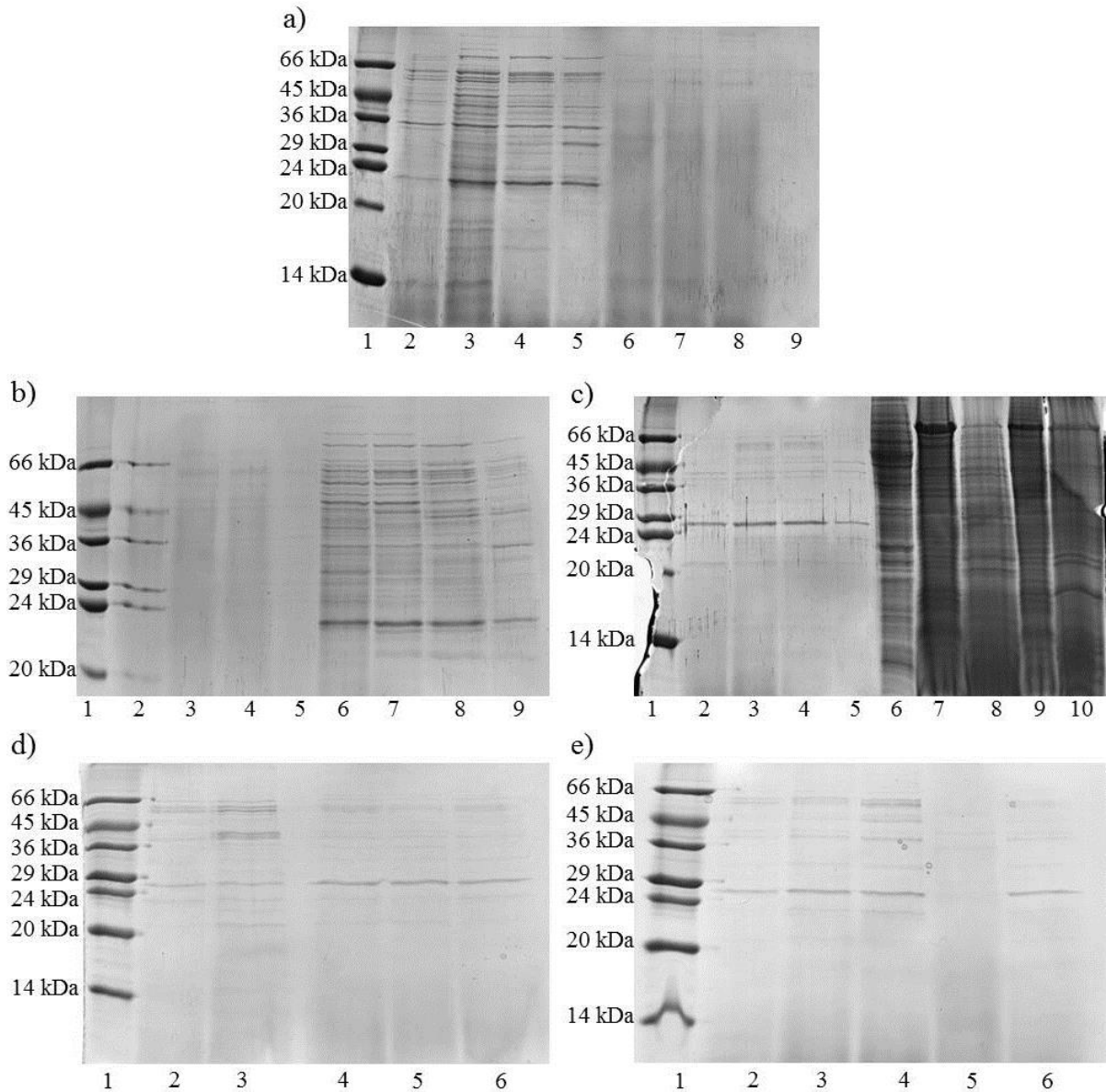


Figura 14 - Eletroforese em gel SDS do teste de expressão de Eg2 em *P. pastoris*. Foi aplicado nos gels a fração sobrenadante de cada cultura, com a exceção do gel "c", colunas de 6 a 10, onde foi aplicada a fração precipitada. a) 1. Padrão de peso molecular; 2. 24 horas de indução, não transformado; 3. 48 horas de indução, não transformado; 4. 72 horas de indução, não transformado; 5. 96 horas de indução, não transformado; 6. 24 horas de indução, colônia 1; 7. 48 horas de indução, colônia 1; 8. 72 horas de indução, colônia 1; 9. 96 horas de indução, colônia 1. b) 1. Padrão de peso molecular; 2. 24 horas de indução, colônia 3; 3. 48 horas de indução, colônia 3; 4. 72 horas de indução colônia 3; 5. 96 horas de indução, colônia 3; 6. 24 horas de indução, colônia 4; 7. 48 horas de indução, colônia 4; 8. 72 horas de indução, colônia 4; 9. 96 horas de indução, colônia 4. c) 1. Padrão de peso molecular; 2. 24 horas de indução, colônia 5; 3. 48 horas de indução, colônia 5; 4. 72 horas de indução colônia 5; 5. 96 horas de indução, colônia 5; 6. Lisado antes da indução, colônia 5; 7. Lisado, 24 horas de indução, colônia 5; 8. Lisado, 48 horas de indução, colônia 5; 9. Lisado, 72 horas de indução, colônia 5; 10. Lisado, 96 horas de indução, colônia 5. d) 1. Padrão de peso molecular; 2. 24 horas de indução, pPIC9K sem inserto; 3. 96 horas de indução, pPIC9K sem inserto; 4. 96 horas de indução colônia 6; 5. 96 horas de indução, colônia 8; 6. 96 horas de indução, colônia 9. e) 1. Padrão de peso molecular; 2. 96 horas de indução colônia 10; 3. 96 horas de indução colônia 11; 4. 96 horas de indução colônia 12; 5. 96 horas de indução, colônia 13; 6. 96 horas de indução, colônia 14.

O segundo teste de expressão foi realizado com 20 colônias para expressão de Eg2 e 15 colônias para expressão de Eg1. As colônias foram induzidas em meio BMMY a 28 °C sob agitação de 250 rpm por 4 dias. Logo após foram coletadas alíquotas que foram centrifugadas e a fração do sobrenadante foi aplicada em gel SDS-PAGE. A colônia 3 de Eg1 não foi analisada por ter apresentado sinais de contaminação durante a indução.

Neste segundo teste notou-se uma diferença no padrão de bandas das colônias 5 e 12 de Eg1 e das colônias 8 e 12 de Eg2 em relação às demais, que se apresentaram como o controle (figura 15). Mesmo assim, diferentemente do esperado para uma expressão induzida, nenhuma banda que pudesse indicar claramente a presença da enzima destacou-se das outras.

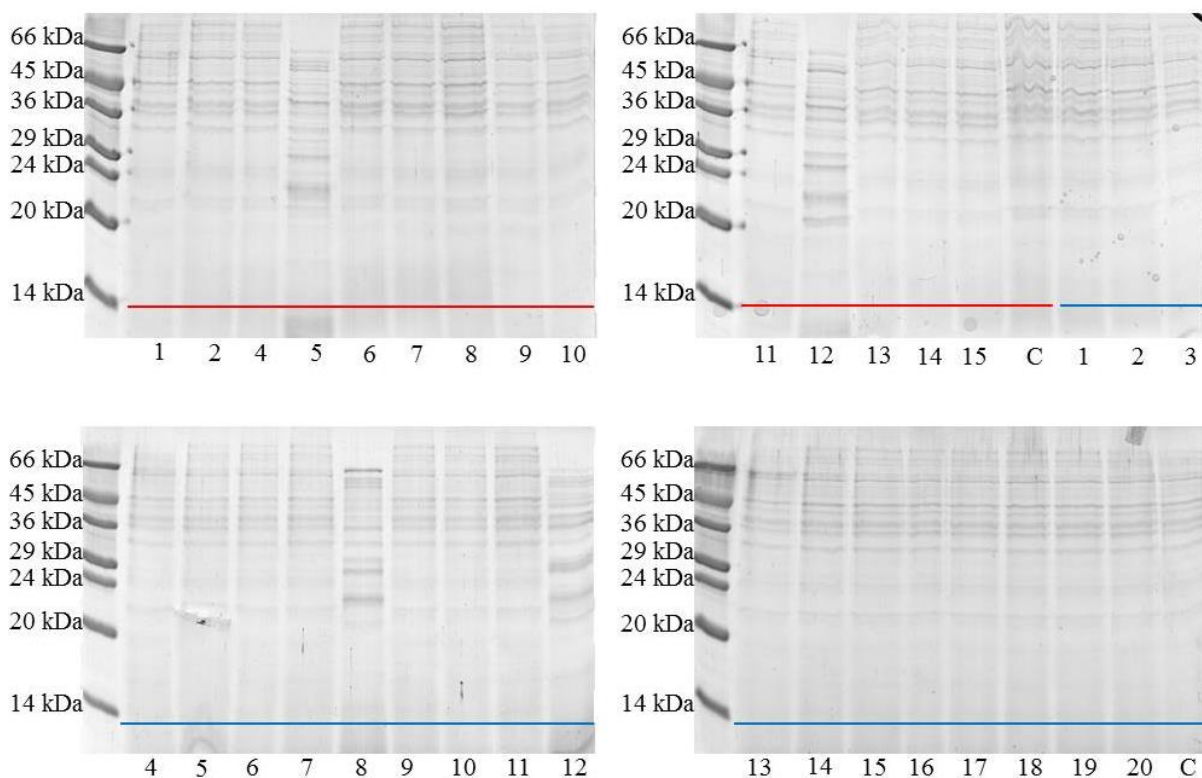


Figura 15 - Eletroforese em gel SDS do teste de expressão de Eg1 e Eg2 em colônias de *P. pastoris* crescidas em 4 mg/mL de geneticina. Foi aplicado nos gels a fração sobrenadante de cada cultura após indução por 72 horas. Foram testadas 15 colônias para expressão de Eg1 (barra vermelha) e 20 colônias para expressão de Eg2 (barra azul). Nas colunas indicadas pela letra C (controle) foram aplicadas amostras das culturas antes da indução.

Para tentar identificar a enzima entre as bandas apresentadas no gel foi feito um experimento de *Western Blotting* com anticorpo primário anti-his. O *Western Blotting* foi feito com as mesmas amostras do terceiro gel da figura 15 (abaixo e à esquerda), onde foram

aplicados os sobrenadantes das colônias 4 a 12 de Eg2. No resultado é possível distinguir apenas uma banda fraca relativa à colônia 8. A banda se apresenta entre 50 e 70 kDa, de acordo com o possível peso molecular (figura 16). Aparentemente ela pode ser a mesma banda da colônia 8 representada na figura 15, logo abaixo de 66 kDa.

Como a colônia 12 de Eg2 gerou o mesmo padrão de bandas da colônia 8 no gel SDS, esperava-se que o *Western Blotting* também revelasse uma banda de mesmo tamanho. Mas como a banda-alvo apresenta-se bem mais fraca, possivelmente não foi suficiente para ser identificada no experimento de *Western Blotting*.

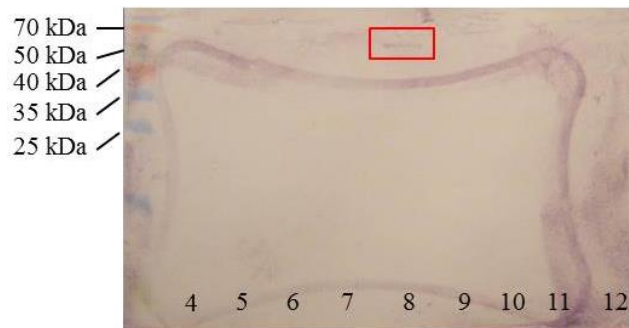


Figura 16 - Western Blotting do teste de expressão de Eg2 em colônias de *P. pastoris*. Os números representam as colônias testadas. A única banda distinguível encontra-se dentro do retângulo vermelho.

Com intuito de melhorar a expressão, novos testes foram feitos com a colônia 8 de Eg2 variando-se a temperatura e o pH do meio de cultura de acordo com a tabela 3. As colônias foram induzidas em meio BMMY sob agitação de 250 rpm por 5 dias. Foram coletadas alíquotas a cada 24 horas após indução, que foram centrifugadas e a fração do sobrenadante foi aplicada em gel SDS. Porém, apesar da variação das condições, não foi verificada melhora alguma (figura 17). Além disso, as bandas que possivelmente representariam a Eg2 (figuras 15 e 16) não se apresentaram mais intensas nos géis. É possível observar uma banda bem fraca, em posição logo abaixo do marcador de 66 kDa, logo após 24 horas de indução nas sete condições testadas, mas sua identificação se torna mais difícil conforme se aumenta o tempo de indução.

Tabela 3 - Condições selecionadas para teste de expressão da enzima Eg2.

Condições	Temperatura	pH
1	28 °C	7,5
2	25 °C	7,5
3	23 °C	7,5
4	28 °C	6,0
5	25 °C	6,0
6	23 °C	6,0
7	28 °C	4,5

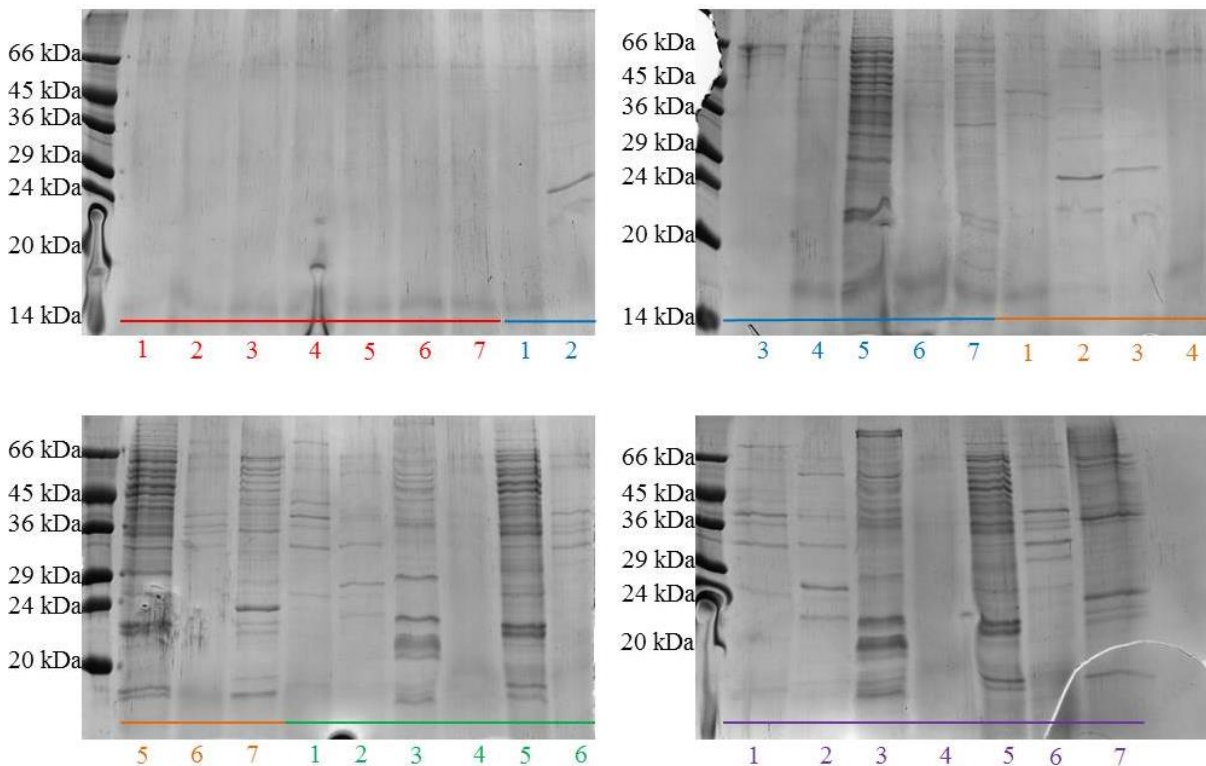


Figura 17 - Eletroforese em gel SDS do teste de expressão da colônia 8 de Eg2 em de *P. pastoris* crescidas e induzidas em diferentes condições de acordo com a tabela 3. Foi aplicada nos geis a fração sobrenadante de cada cultura. Os números abaixo de cada gel representam as condições descritas na tabela 3. As amostras foram analisadas após 24 horas (barra vermelha), 48 horas (barra azul), 72 horas (barra laranja), 96 horas (barra verde) e 120 horas (barra roxa) de indução.

Em uma última tentativa de determinar a presença da enzima clonada, foi feita uma expressão da colônia 8 de Eg2 em maior escala com as mesmas condições descritas no segundo teste de expressão. Após a indução de 4 dias a cultura foi centrifugada e o sobrenadante foi concentrado e aplicado em uma coluna de afinidade com níquel imobilizado.

Após três lavagens com 3x o volume de coluna, foi feita a eluição com 200 e 500 mM de imidazol.

Todas as etapas da purificação foram analisadas em gel SDS (figura 18). Infelizmente, não foi observada nenhuma banda acima de 45 kDa nas frações eluídas. Apenas uma banda foi observada contendo pouco menos de 24 kDa, que possivelmente representa alguma proteína da própria *P. pastoris* que possui alguma afinidade pela coluna de níquel. Comparando-se a fração sobrenadante com a fração não aderida na coluna chegamos à conclusão de que todas as bandas mais fortes passaram diretamente pela resina.

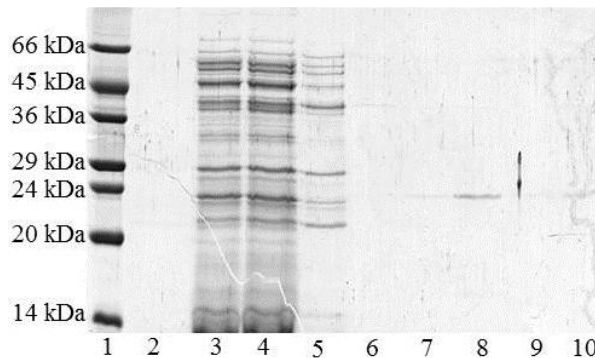


Figura 18 - Eletroforese em gel SDS da purificação de Eg2 por cromatografia de afinidade. 1. Padrão de peso molecular; 2. Sobrenadante filtrado; 3. Sobrenadante filtrado e concentrado; 4. Fração não aderida à resina; 5. Primeira lavagem; 6. Segunda lavagem; 7. Terceira lavagem; 8. Eluição com 200 mM de imidazol; 9. Eluição com 500 mM de imidazol; 10. Resina após eluição.

Durante todos os testes de expressão foram feitas medidas de atividade das amostras coletadas. As amostras eram misturadas com solução contendo CMC (concentração final de 1 %) e acetato de sódio (concentração final de 50 mM) pH 5,0 e mantidas a temperatura constante de 50 °C. Logo após, a quantidade de açúcares redutores liberados era revelada por reação com DNS e comparada com reações controle. Todas as amostras testadas não apresentaram atividade.

4.1.3 Clonagem de Eg1 e Eg2 para expressão em *P. pastoris* pelo vetor pPZa.

A clonagem em vetor pPZa foi feita no Laboratório de Biologia Molecular da Universidade de Brasília sob a coorientação do Professor Dr. Fernando Araripe Torres.

O primeiro passo foi a retirada dos genes do plasmídeo original pPIC9K e a inserção dos mesmos no plasmídeo pPZ α , onde o promotor AOX foi substituído pelo promotor constitutivo PGK1.

Para tal, pPIC9K/Eg1, pPIC9K/Eg2 e pPZ α foram clivados com as enzimas de restrição NotI e BstBI (figura 19). O sítio de clivagem da NotI está no poli-linker dos dois plasmídeos e a BstBI cliva em duas posições do pPIC9K. A posição que libera o gene está logo antes do fator- α , mesma posição encontrada no pPZ α , e a outra se encontra no lado oposto da região de inserção do gene. De acordo com as enzimas utilizadas, os insertos liberados deveriam ter os tamanhos 1463 kb, para Eg1, e 1266 kb, para Eg2.

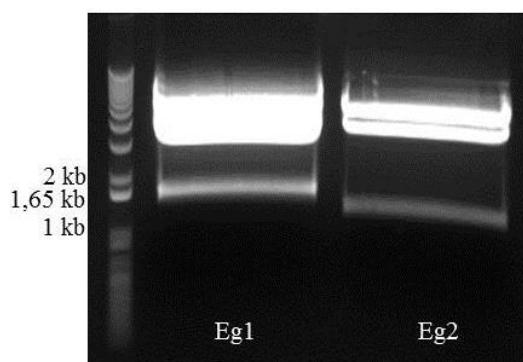


Figura 19 - Eletroforese em gel de agarose da clivagem dos plasmídeos pPIC9K/Eg1 e pPIC9K/Eg2 pelas enzimas de restrição NotI e BstBI.

As bandas relativas aos genes foram purificadas do gel e foi feita ligação no plasmídeo pPZ α com a enzima ligase (New England Biolabs) por 2 horas a temperatura ambiente. O plasmídeo final foi utilizado para a transformação de células de *E. coli* (DH5 α) competente.

O antibiótico utilizado para a seleção dos transformantes, tanto em *E. coli* como em *P. pastoris*, é a zeocina. Como ela é sensível a elevadas concentrações de sal, as placas em que as células transformadas foram espalhadas continham meio LB com baixa concentração de cloreto de sódio (< 90 mM).

Para conferir a inserção dos genes foi feito teste de restrição com as enzimas NotI e BstBI em quatro colônias de cada construção. Em todos os casos o inserto estava presente com o número de pares de bases esperado (figura 20).

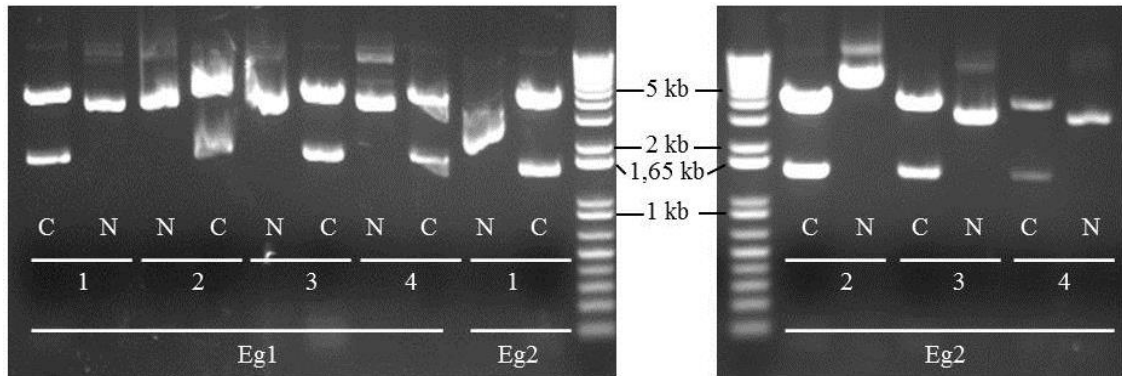


Figura 20 - Eletroforese em gel de agarose do teste de restrição de colônias transformadas com pPZ α /Eg1 e pPZ α /Eg2. Foram testadas quatro colônias de cada construção. As letras C e N representam as colunas dos plasmídeos clivados e não clivados, respectivamente.

Foi selecionada uma colônia de cada construção e grande quantidade de plasmídeo foi produzido e purificado com o kit midi-prep (Promega). Para a transformação em *P. pastoris*, os vetores foram linearizados com a enzima de restrição SacI (figura 21). Apesar da Eg1 não ter sido linearizada totalmente, a porção clivada é majoritária. As bandas digeridas foram purificadas do gel e utilizadas na transformação de *P. pastoris* competente. Após as transformações, as placas foram incubadas por até cinco dias, uma vez que a zeocina perde sua atividade com o tempo.

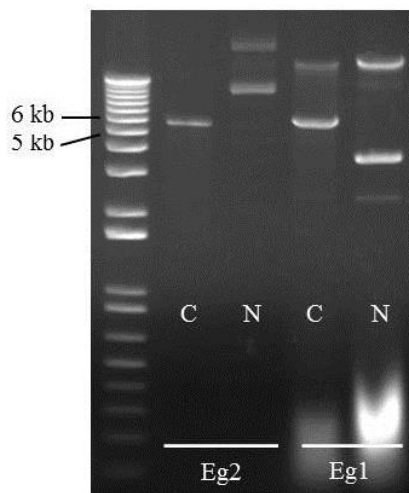


Figura 21 - Eletroforese em gel de agarose da linearização dos plasmídeos pPZ α /Eg1 e pPZ α /Eg2 com a enzima de restrição SacI. As letras C e N representam as colunas dos plasmídeos clivados e não clivados, respectivamente.

Várias tentativas de transformação foram feitas variando-se o protocolo e utilizando-se diferentes concentrações de antibiótico. Em todas as condições testadas não houve sucesso.

Mesmo após cinco dias, nenhuma colônia foi observada. De forma inesperada, nenhuma colônia cresceu, mesmo sem utilizar antibiótico.

4.1.4 Clonagem de Eg2 para expressão em *E. coli*.

Como as tentativas de expressão da enzima Eg2 em *P. pastoris* levaram uma quantidade considerável de tempo e não tiveram sucesso, procedemos à tentativa de expressão em *E. coli*, como já foi descrito na literatura¹⁰¹. Deste modo, realizamos a clonagem do gene de Eg2 no plasmídeo pET28a. Apesar das bactérias não terem a capacidade de fazer glicosilação de enzimas, o estudo estrutural da endoglucanase não-glicosilada também seria de grande importância.

Após a amplificação do gene por PCR (figura 22), este foi subclonado no vetor pGEM-T e utilizado em transformação de células de *E. coli* (DH5 α) para sua propagação. O gene foi então retirado do pGEM-T por clivagem com as enzimas de restrição NdeI e XhoI e inserido no vetor de expressão pET28a, que também havia sido clivado pelas mesmas enzimas. Novamente, células de *E. coli* (DH5 α) foram transformadas, agora com o novo vetor, e duas das colônias resultantes foram selecionadas para teste de restrição do plasmídeo (figura 22). O vetor pET28a/Eg2 foi então utilizado na transformação das cepas de *E. coli*.

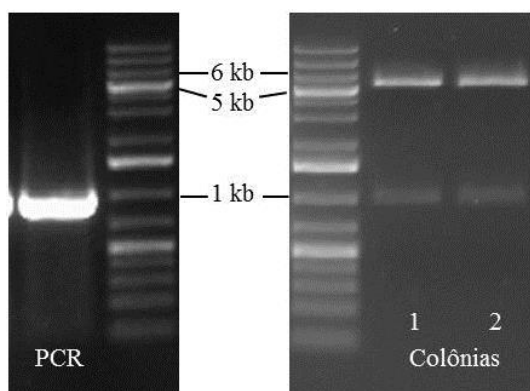


Figura 22 - Eletroforese em gel de agarose da amplificação por PCR do gene Eg2 para expressão em *E. coli* (a esquerda), e do teste de restrição com as enzimas NdeI e XhoI do plasmídeo pET28a/Eg2 extraído de duas colônias (a direita).

4.1.5 Expressão de Eg2 em *E. coli*.

A expressão da enzima Eg2 foi primeiramente testada em condições descritas no item materiais e métodos. No teste de expressão, diferentes cepas de *E. coli* foram utilizadas: BL21; BL21 Codon Plus; BL21 Ril; e BL21 Origami2.

Após a indução, as células foram separadas do meio de cultura e lisadas em tampão de ressuspensão (100 mM de cloreto de sódio, 50 mM de Tris-HCl pH 8,0, 10 % glicerol, 1 mM de inibidor de protease PMSF e 250 µg/mL de lisozima) através de sonicação. Novamente eram centrifugadas e as frações solúveis e precipitadas eram aplicadas e analisadas em gel SDS.

O resultado obtido mostra que a Eg2 foi expressa em todas as cepas (com a exceção da colônia 2 da cepa BL21 pLysS), porém insolúvel (figura 23). Fica clara no gel a presença de duas fortes bandas nas frações precipitadas entre os marcadores de 36 e 29 kDa. Como o peso molecular calculado para a Eg2 é de 35,27 kDa, é provável que ela seja uma das duas bandas.

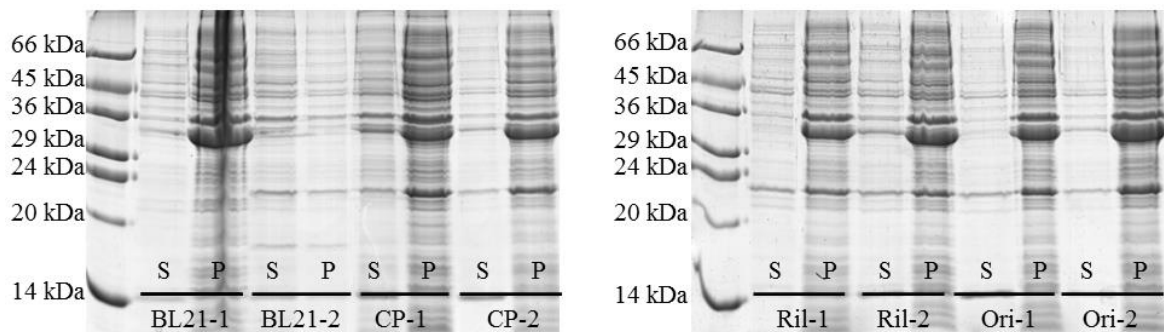


Figura 23 - Eletroforese em gel SDS do teste de expressão de Eg2 em diferentes cepas de *E. coli*. Foram testadas duas colônias de cada cepa: BL21; BL21 Codon Plus (CP); BL21 Ril (ril); e BL21 Origami2 (Ori). As letras P e S representam as frações precipitadas e solúveis respectivamente.

De acordo com o artigo que descreve a expressão de Eg2 em *E. coli*, a enzima somente apareceu na fração solúvel quando foi feito o crescimento das bactérias e a indução da expressão a baixas temperaturas¹⁰¹. Portanto, novo teste de expressão foi feito em *shaker* com temperatura controlada a 18 °C. Mas, diferentemente da descrição no artigo, a Eg2 continuou na fração insolúvel (figura 24).

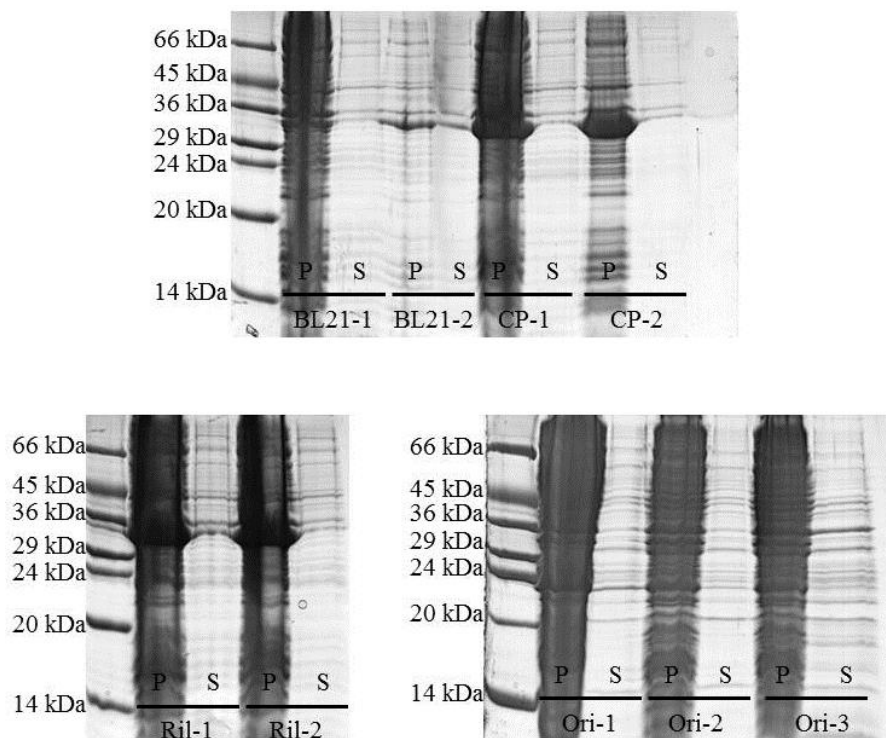


Figura 24 - Eletroforese em gel SDS do teste de expressão de Eg2 a baixa temperatura em diferentes cepas de *E. coli*. Foram testadas três colônias da cepa BL21 Origami2 (Ori), duas colônias BL21 comum, duas colônias BL21 Codon Plus (CP) e duas colônias BL21 RIL (ril). As letras P e S representam as frações precipitadas e solúveis respectivamente.

Vários outros testes foram feitos, variando-se as condições de crescimento, condições de indução, método de lise e tampão de ressuspensão, mas em todos eles a enzima sempre se encontrava na fração insolúvel.

Em uma última tentativa de conseguir ao menos pequena quantidade da enzima, foi feita uma expressão em maior escala da colônia 1 da cepa BL21 Ril e o sobrenadante foi purificado por cromatografia de afinidade. Para evitar precipitação, a purificação foi feita a baixa temperatura (4 °C). A análise das etapas da purificação mostra que praticamente toda a enzima encontra-se na fração insolúvel (figura 25). Nas frações de eluição aparecem pequenas bandas de diferentes tamanhos, sendo que as duas mais fortes estão bem distantes do peso molecular esperado. Além disso, a quantidade total de proteínas presentes nas eluições foi estimada em torno de 100 µg.

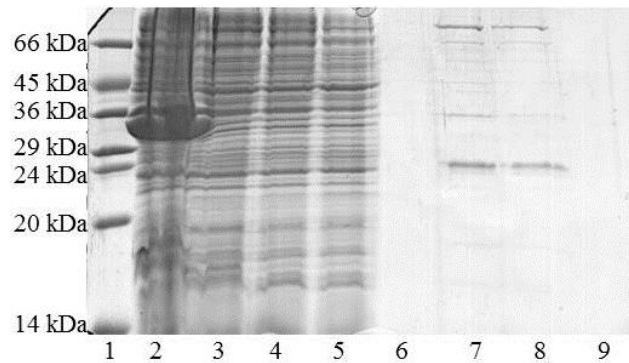


Figura 25 - Eletroforese em gel SDS da purificação de Eg2 por cromatografia de afinidade. 1. Padrão de peso molecular; 2. Fração insolúvel; 3. Fração solúvel; 4. Fração não aderida à resina; 5. Primeira lavagem; 6. Última lavagem; 7. Eluição com 200 mM de imidazol; 8. Eluição com 500 mM de imidazol; 9. Resina após eluição

4.2 DISCUSSÃO

4.2.1 Expressão em *P. pastoris*.

A expressão de proteínas heterólogas em leveduras vem sendo utilizada com sucesso há mais de vinte anos. Geralmente, as leveduras são selecionadas como hospedeiras por combinar uma fácil manipulação genética com rápido crescimento em meios de cultura de baixo custo e por sua capacidade de promover modificações pós-traducionais¹²⁰. Uma das mais empregadas, tanto no meio acadêmico quanto industrial, é a *P. pastoris*, cujo sucesso decorre principalmente da facilidade de regulação do promotor do gene que codifica a primeira enzima da via de utilização de metanol, a álcool oxidase I (AOX1)¹²¹. Em células alimentadas com metanol (indutor do promotor), a AOX1 chega a responder por 30 % do total de proteínas solúveis¹²².

Até então, nenhum aluno do grupo tinha experiência com expressão heteróloga em leveduras, mas, graças as suas características, a *P. pastoris* foi escolhida como organismo de expressão das enzimas Eg1 e Eg2, que são proteínas originalmente de fungos e possuem modificações pós-traducionais, no caso, a glicosilação.

As fases de clonagem e transformação das células competentes foram feitas com relativa facilidade. As análises em géis de agarose confirmaram, sem deixar dúvidas, todas as etapas de clonagem. As transformações aparentemente foram feitas da forma correta, pois

várias colônias de *P. pastoris* GS-115 cresceram em meios seletivos, até em concentração máxima de antibiótico indicada.

Em todos os testes de expressão feitos não foi observada, em gel SDS, nenhuma banda de expressão que pudesse indicar a presença das enzimas-alvo. A única banca mais intensa foi observada na fração lisada da colônia 5 (figura 14c, item 7), mas aparentemente essa banda também aparece com menor intensidade nas colônias controle. Ademais, as bandas um pouco mais intensas do que a média são observadas tanto no padrão de bandas gerado pelas colônias positivas como no padrão das colônias controle.

Um fator que dificultou a análise foi a imprecisão do peso molecular das enzimas. O cálculo de peso molecular a partir das sequências de aminoácidos definiu a Eg1 com 41,06 kDa e a Eg2 com 35,27 kDa, porém a glicosilação feita por *P. pastoris* pode aumentar de 2 a mais de 60 kDa o peso molecular da proteína¹²³. Ou seja, qualquer banda acima do peso molecular calculado poderia representar a enzima desejada.

De acordo com protocolo da Invitrogen a temperatura de crescimento e indução é um fator crítico para a expressão heteróloga. Indica-se a manutenção da temperatura a 30 °C, que não deve ser ultrapassada. Em *shakers* que não possuem bom controle de temperatura é aconselhado fixar o termostato em 28 °C, evitando assim que possíveis variações não interfiram com a expressão. Além disso, vários artigos citam uma melhora considerável na expressão utilizando diferentes temperaturas¹²⁴⁻¹²⁵. Por isso várias temperaturas de crescimento e indução foram testadas. Mas, como pôde ser observado, a variação da temperatura não surtiu efeito na expressão das enzimas.

Após todas as tentativas descritas para expressão utilizando o vetor pPIC9K, foi feita a inserção dos genes em plasmídeo pPZa e a transformação de células de *P. pastoris* X-33. Todos os experimentos com o vetor pPZa foram feitos no Laboratório de Biologia Molecular da Universidade de Brasília sob a supervisão do Professor Dr. Fernando Araripe Torres. A mudança tanto do vetor quanto da cepa da levedura se deu principalmente por causa da experiência que o grupo possuía, além de gerar condições diferentes para a expressão.

A retirada dos genes do vetor pPIC9K e a inserção dos mesmos no vetor pPZa também foram feitas com facilidade e, em pouco tempo, já tínhamos os vetores linearizados e prontos para a transformação das células de *P. pastoris*.

Todos os experimentos de transformação fracassaram. As primeiras tentativas de transformação foram feitas com células competentes previamente preparadas e não foi observada nenhuma colônia após cinco dias de incubação. A partir de então, várias tentativas foram feitas preparando-se as células competentes no mesmo dia da transformação, sem surtir

efeito. Em um último experimento, após serem transformadas, as células foram plaqueadas em meios de cultura com diferentes concentrações de antibiótico, mas, mesmo na ausência de antibiótico, nenhuma colônia foi observada. Esse resultado também foi observado fazendo-se a transformação com vetores de outros alunos, cujos resultados eram positivos.

De acordo com o próprio Professor Dr. Fernando Araripe Torres, provavelmente os estoques de células de *P. pastoris* estavam contaminados. Corroboram essa hipótese a turbidez anormal (mais intensa) dos meios de cultura após o inóculo e o tempo de crescimento duas ou três vezes maior do que o normal para atingir a DO₆₀₀ esperada.

Várias razões podem explicar os resultados negativos obtidos na expressão das Eg1 e Eg2 em *P. pastoris* e não necessariamente apenas uma delas responde pelo todo. Os genes de Eg1 e Eg2 foram sequenciados para confirmar se não havia alguma mutação, mas os vetores não. Uma única mutação em algumas regiões específicas do vetor é suficiente para que a transformação ou a expressão não ocorram corretamente. O mecanismo de seleção das colônias pode ter falhado por possível perda de atividade do antibiótico ou por problemas com os ingredientes dos meios de cultura. Contaminações anteriores ou posteriores nas diferentes etapas podem ter impedido a expressão.

Mesmo com os vários possíveis motivos responsáveis pelos resultados negativos, nenhum deles pareceu ser mais contundente. Por ter-se passado muito tempo e por não se possuir uma resposta clara e direta sobre o que fazer para que a expressão em *P. pastoris* desse certo, decidiu-se tentar a expressão das enzimas em *E. coli* paralelamente com estudos de outras endoglucanases.

4.2.2 Expressão em *E. coli*.

A expressão das endoglucanases de *T. reesei* em *E. coli* passaram a ser uma opção assim que os experimentos de expressão em *P. pastoris* se mostraram ineficazes, ao menos nas condições que eles foram feitos. Apesar das bactérias não terem a capacidade de fazer modificações pós-traducionais, o estudo funcional e estrutural da Eg2 de *T. reesei* não glicosilada seria de grande importância, uma vez que ainda não havia estrutura cristalográfica resolvida dessa proteína e poucos estudos haviam sido feitos com ela.

A construção do gene de Eg2 inserida no vetor de expressão pET28a foi a mesma utilizada para expressão em *P. pastoris*. Essa construção foi baseada em um artigo que descreve sua expressão em *E. coli*¹⁰¹.

Diferentemente dos ensaios com a *P. pastoris*, os testes de expressão feitos mostram claramente que a proteína foi expressa em todas as linhagens de células testadas, mas em nenhuma ela foi obtida na fração solúvel. Várias condições foram testadas variando-se temperatura, tempo de crescimento, tempo de indução, meio de cultura, método de lise e tampão de ressuspensão, mas, mesmo seguindo rigorosamente as condições descritas no artigo que relata sua expressão, não foi possível obtê-la na fração solúvel.

Ao mesmo tempo em que esses resultados foram obtidos, a estrutura do domínio catalítico da endoglucanase II de *T. reesei* foi publicada¹²⁶. Baseando-se neste artigo e considerando-se que o protocolo utilizado para a expressão da proteína foi o mesmo testado neste trabalho, chega-se a conclusão de que, provavelmente, a construção do gene foi crítica para sua expressão na forma solúvel. Na construção descrita pelo artigo, o domínio catalítico foi acrescido, na sua região amino-terminal, de sete resíduos de aminoácidos. Porém, ao invés de utilizar os aminoácidos da sequência nativa (STPPTSS), foi utilizada a sequência: “MGGSGSG”. A alteração na solubilidade de proteínas pela adição ou remoção de poucos aminoácidos já foi reportada em vários outros trabalhos¹²⁷⁻¹²⁸.

Apesar da provável resposta para o problema da solubilidade da Eg2 ter sido encontrada, sua estrutura e suas características enzimáticas foram publicadas¹²⁶⁻¹²⁹. Desta forma, decidiu-se por encerrar seu estudo e dedicar-se a outras endoglucanases que se tornavam promissoras.

5 ENDOGLUCANASES DA FAMÍLIA GH5 E BETA-GLICOSIDASES DA FAMÍLIA GH1

5.1 RESULTADOS

5.1.1 Expressão.

A primeira rodada de teste de expressão foi feita com as três construções da enzima H6 e a protease TEV. O protocolo descrito se apresentou satisfatório já que todas foram expressas na forma solúvel. Após a lise das bactérias, foi feita centrifugação e a fração solúvel foi analisada em gel SDS. As bandas apresentadas não deixam dúvida sobre presença das enzimas na fração solúvel (figura 26).

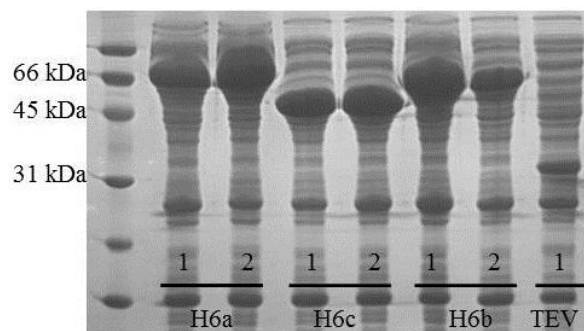


Figura 26 - Eletroforese em gel SDS do teste de expressão das enzimas H6a (75 kDa), H6b (63 kDa), H6c (53 kDa) e TEV. Foram testadas duas colônias de cada enzima, com a exceção da TEV (33 kDa). Foi aplicada nos geis a fração solúvel de cada lise.

Da mesma forma, uma segunda rodada de testes de expressão foi feita com as enzimas A9, B5, B9, C6, C9, D5, E5, F5 e F6. Mais uma vez o protocolo usado foi satisfatório já que seis das nove enzimas foram expressas na forma solúvel. Apesar das bandas serem menos intensas, com exceção da F5, a expressão é evidente. As enzimas B5, E5 e F6 não apresentaram banda de expressão distinta (figura 27a).

A terceira rodada de testes, feita com as enzimas F7, F9, F12, G5, G9 e H5, apresentou apenas as enzimas G5 e H5 na forma solúvel. As demais não possuem uma evidente banda no gel (figura 27b).

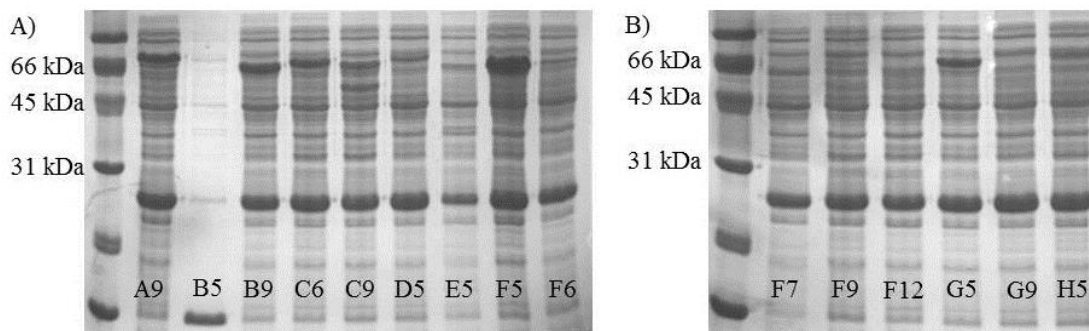


Figura 27 - Eletroforese em gel SDS do teste de expressão das enzimas: A) A9 (69 kDa), B5 (57 kDa), B9 (69 kDa), C6 (71 kDa), C9 (70 kDa), D5 (68 kDa), E5 (69 kDa), F5 (68 kDa) e F6 (68 kDa); B) F7 (60 kDa), F9 (72 kDa), F12 (76 kDa), G5 (70 kDa), G9 (71 kDa) e H5 (71 kDa). Foi aplicada nos geis a fração solúvel de cada lise.

Na quarta rodada, foi utilizado o mesmo protocolo das rodadas anteriores para as enzimas E12, F1 e F4. Ao mesmo tempo, foi feito teste de expressão com indução a 20 °C com as enzimas B5, E5, F6, F7, F9, F12 e G9. A partir deste momento, as frações precipitadas também foram analisadas para comprovar a expressão das enzimas.

Com a redução da temperatura, as enzimas B5, E5, F6, F7, F12 e G9 tornaram-se solúveis (figura 28a) enquanto as enzimas testadas pela primeira vez apareceram na forma precipitada (figura 28b). A enzima F9 não apareceu em nenhum dos géis, indicando que ela não foi expressa.

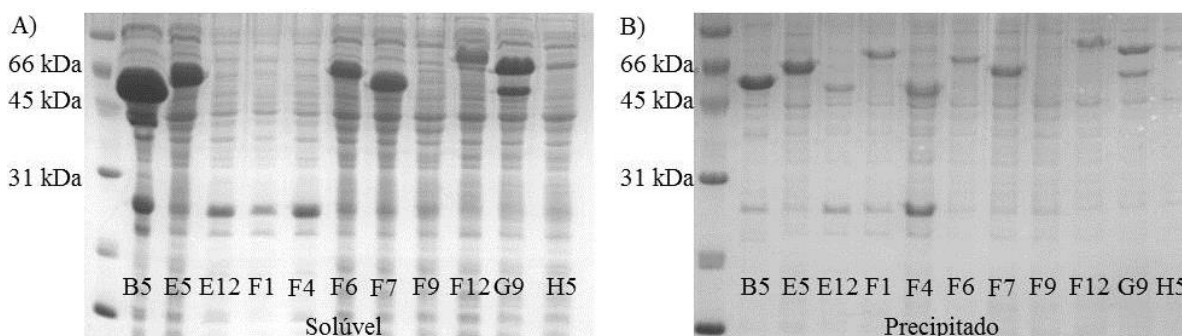


Figura 28 - Eletroforese em gel SDS do teste de expressão das enzimas B5 (57 kDa), E5 (69 kDa), E12 (56 kDa), F1 (62 kDa), F4 (53 kDa), F6 (68 kDa), F7 (60 kDa), F9 (72 kDa), F12 (76 kDa), G9 (71 kDa) e H5 (71 kDa). Foram aplicadas nos geis as frações: A) Solúvel; B) Precipitado.

A quinta rodada de teste de expressão foi feita com crescimento das bactérias e indução da expressão a 20 °C. Este teste foi feito com as enzimas F1, F4 e com F9 provinda de nova transformação. A enzima F4 se apresentou solúvel, a F1 mostrou-se no precipitado e a enzima F9 não apresentou banda de expressão (figura 29).

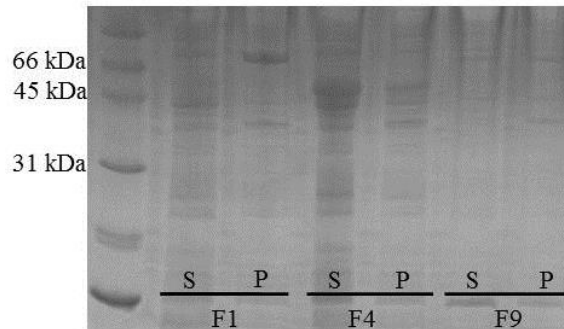


Figura 29 - Eletroforese em gel SDS do teste de expressão das enzimas F1 (62 kDa), F4 (53 kDa) e F9 (72 kDa). As letras S e P representam, respectivamente, as frações solúveis e precipitadas de cada cultura.

O protocolo de expressão em maior escala foi feito de acordo com o teste de expressão em que cada enzima foi encontrada na fração solúvel. Desta forma, todas elas, com a exceção de F1 e F9, foram expressas em litros de meio de cultura. O rendimento de cada uma estava de acordo com o tamanho das bandas de expressão apresentadas nos testes. Enzimas como H6 e suas construções, G5 e F5 rendiam por volta de 60 mg de proteína por litro de meio de cultura. As demais tinham rendimento variável entre 5 e 15 mg/L.

5.1.2 Purificação.

O método de purificação por afinidade se mostrou muito eficiente com as enzimas estudadas. Na maioria dos casos, ela foi suficiente para obtenção das enzimas com alto grau de pureza.

Todas as etapas da purificação foram analisadas em gel SDS juntamente com as etapas de clivagem da cauda de histidinas e tioredoxina (figuras 30 e 31). Observando-se a coluna 5 de cada gel, percebe-se que a composição e o volume do tampão de lavagem utilizado foram suficientes para a eliminação dos contaminantes. Em alguns casos, a expressão da enzima-

alvo foi tão grande que a resina foi saturada e, mesmo na última lavagem, ainda é possível observar uma pequena quantidade da enzima sendo eliminada da coluna.

A coluna 6 de cada gel, mostradas nas figuras 30 e 31, representa a fração em que as proteínas foram eluídas da resina de purificação; a coluna 8 mostra a fração em que as enzimas estão clivadas e separadas de suas caudas. Quando era verificado que a clivagem da cauda não tinha sido total, nova quantidade da protease TEV era aplicada, deixava-se incubar por mais vinte quatro horas e a passagem pela resina era feita novamente.

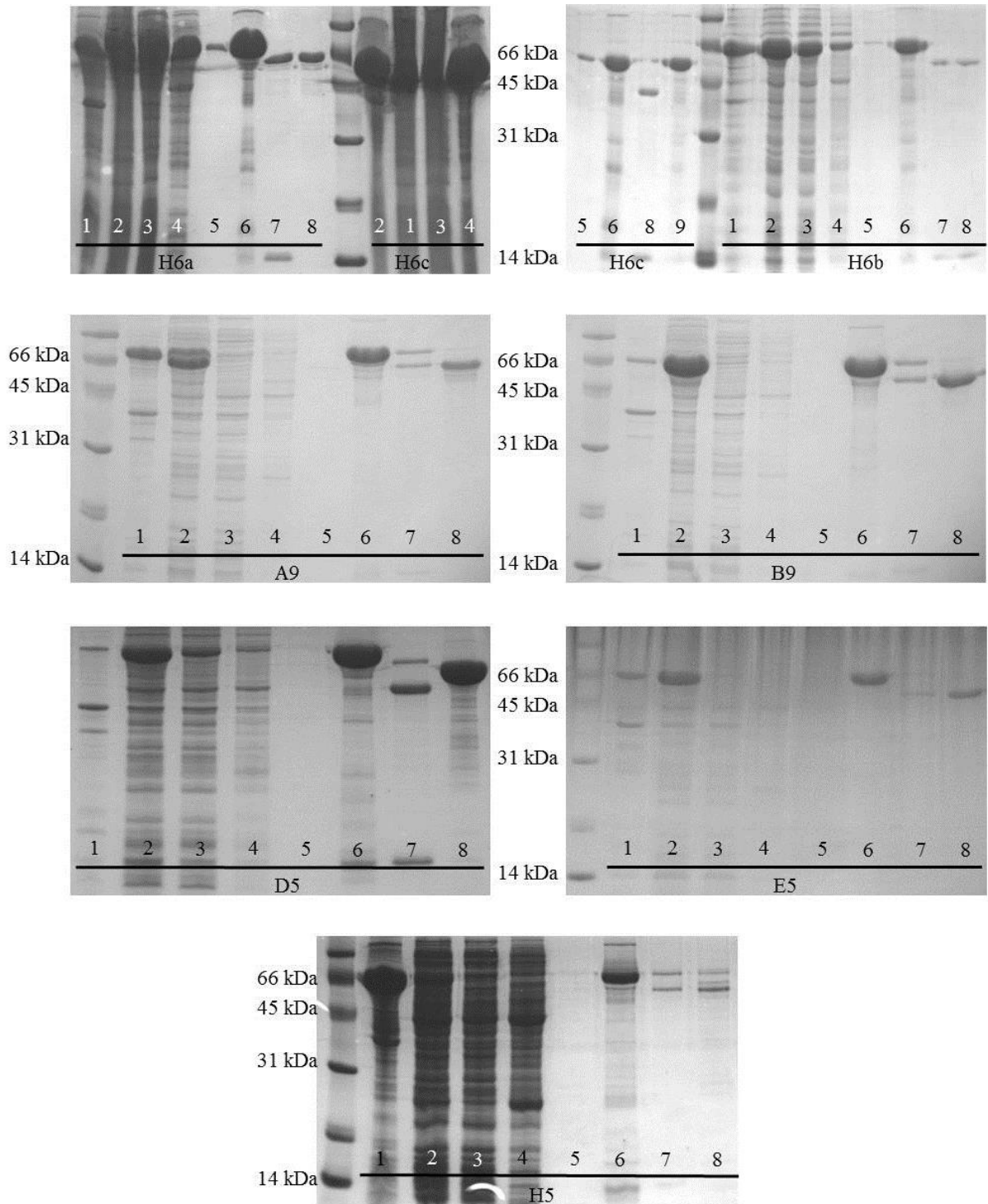


Figura 30 - Eletroforese em gel SDS da purificação por afinidade e clivagem da cauda de histidinas e tiorredoxina das enzima H6a, H6b, H6c, A9, B9, D5, E5 e H5. Os números indicados representam: 1) Fração precipitada; 2) Fração Solúvel; 3) Fração não aderida à resina; 4) Primeira lavagem; 5) Última lavagem; 6) Eluição com 200 mM de imidazol; 7) Clivagem com TEV; 8) Segunda passagem pela coluna; 9) Resina após eluição.

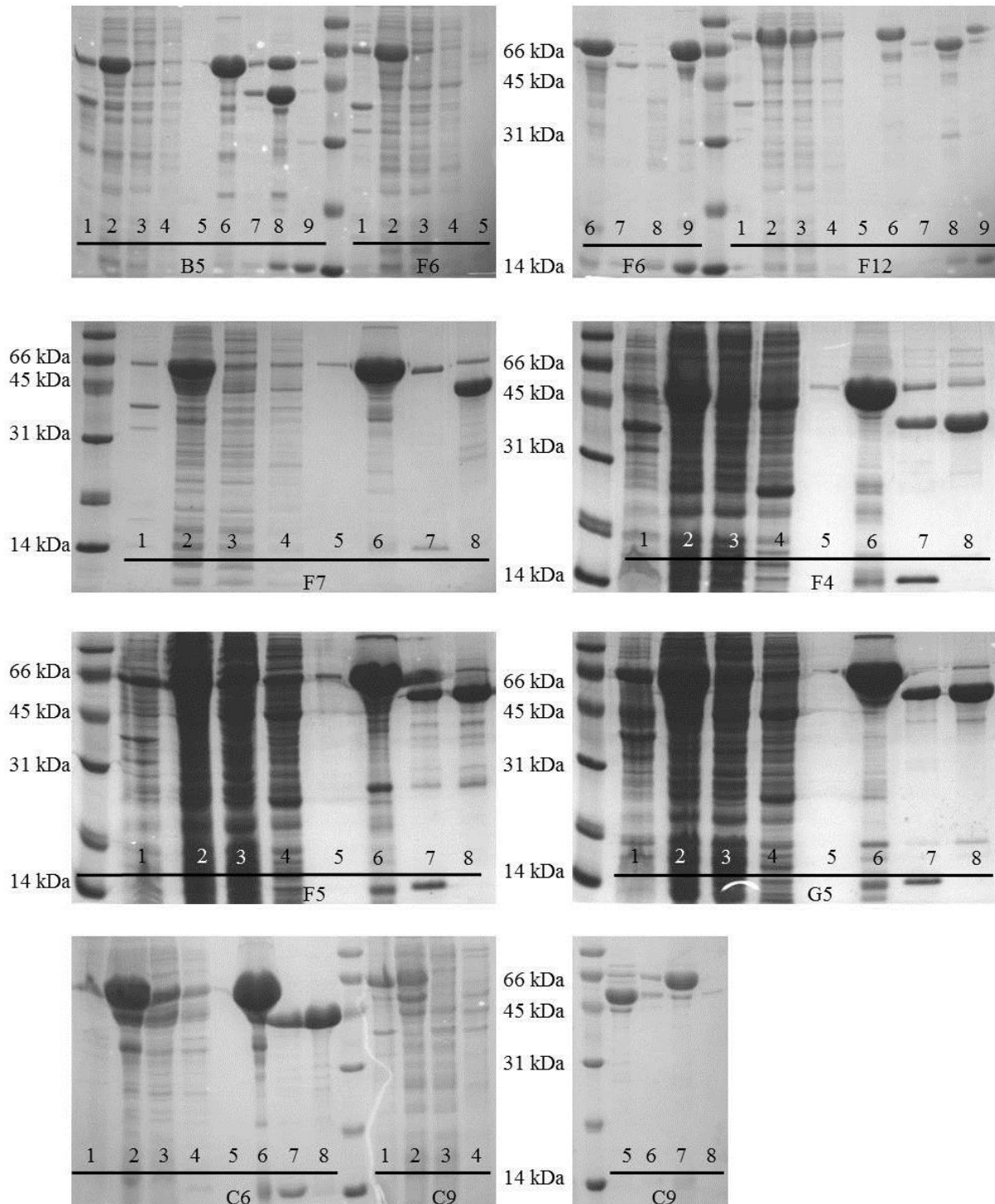


Figura 31 - Eletroforese em gel SDS da purificação por afinidade e clivagem da cauda de histidinas e tiorredoxina das enzimas B5, F6, F12, F7, F4, F5, G5, C6 e C9. Os números indicados representam: 1) Fração precipitada; 2) Fração Solúvel; 3) Fração não aderida à resina; 4) Primeira lavagem; 5) Última lavagem; 6) Eluição com 200 mM de imidazol; 7) Clivagem com TEV; 8) Segunda passagem pela coluna; 9) Resina após eluição.

As enzimas B5 e F12, após a purificação por afinidade, apresentaram uma quantidade significativa de contaminantes e foram submetidas à purificação por exclusão de tamanho. Frações de 0,5 mL foram coletadas e os picos do cromatograma foram analisados em gel

SDS. Na figura 32, as setas vermelhas indicam as frações mais puras e que foram utilizadas para os experimentos subsequentes.

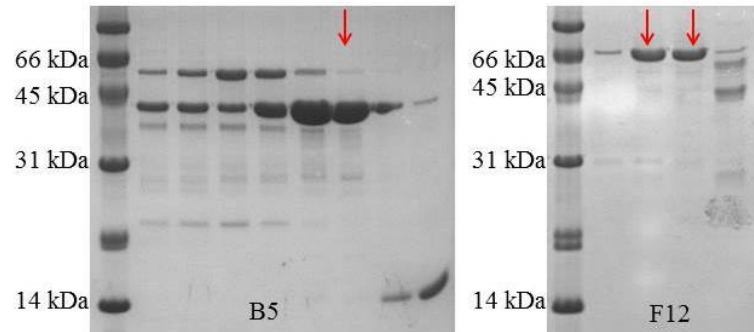


Figura 32 - Eletroforese em gel SDS da purificação por exclusão das enzima B5 e F12. As setas vermelhas indicam as frações que foram utilizadas para concentração e ensaio de cristalização. As frações foram aplicadas no gel de acordo com o tempo de passagem pela coluna.

5.1.3 Atividade enzimática da enzima H6 e suas construções.

As enzimas H6a, H6b e H6c foram submetidas a diferentes testes enzimáticos com o intuito de estabelecer suas condições ótimas de atividade e seu grau de especificidade. Por ser descrita como endoglucanase, a carboximetilcelulose foi selecionada como substrato para os primeiros testes de atividade.

A figura 33 exibe a atividade específica da enzima H6a em diferentes temperaturas. A temperatura ótima de atividade foi determinada em 55 °C, onde a enzima possui atividade de 11,4 U/mg. Apesar de a temperatura ótima de atividade ser 55 °C, a enzima apresenta uma grande estabilidade térmica, pois ainda possui 50 % da atividade a 80 °C e 16 % a 90 °C.

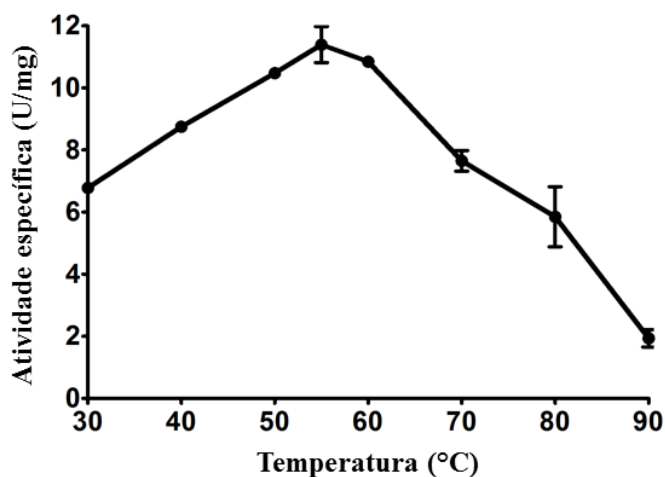


Figura 33 - Teste enzimático da enzima H6a em diferentes temperaturas.

Fixando-se a temperatura em 55 °C, foi calculada a atividade da H6a em diferentes pHs (figura 34). O pH ótimo para a atividade da enzima foi determinado em 4,0. De acordo com o gráfico, a atividade cai bruscamente com a redução do pH a partir do ótimo. Em pH 3,5 a atividade é de 19 % da máxima, perdendo totalmente a atividade em pH 3,0. Com valores maiores de pH, a atividade cai lentamente chegando a 50 % da atividade em pH 5,5 e 19 % em pH 7,0.

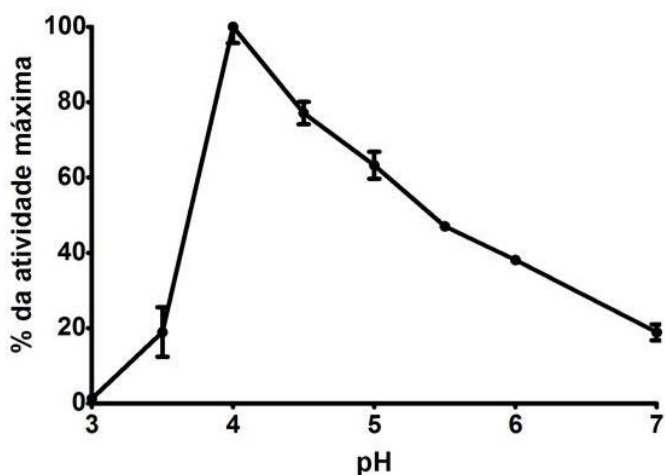


Figura 34 - Teste enzimático da enzima H6a em diferentes pHs.

Uma vez definidas as condições ótimas de atividade, foi feito teste de especificidade, onde diferentes substratos foram submetidos à clivagem pela enzima H6a nas condições ótimas. A figura 35 apresenta o gráfico das atividades obtidas com os diferentes substratos. A

enzima H6a apresentou maior atividade para beta-glicano, chegando a 34,2 U/mg. Mesmo com a preferência por beta-glicano, a enzima possui boa atividade para outros substratos, como lichenano, xiloglucano e CMC. Os demais substratos testados não foram clivados pela enzima.

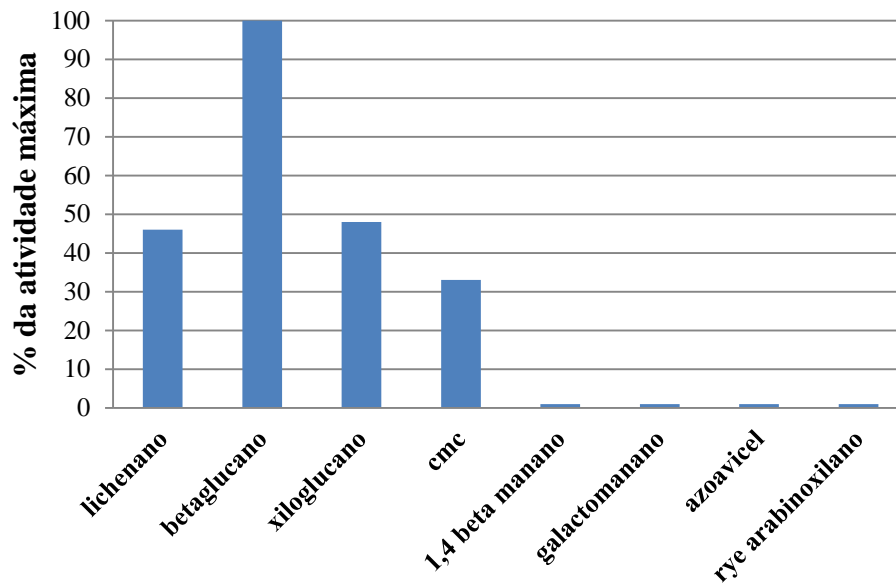


Figura 35 - Atividade relativa da enzima H6a frente a diferentes substratos.

Todos os testes enzimáticos feitos com a enzima H6a também foram feitos com as outras construções, H6b e H6c. Porém, ambas não apresentaram atividade alguma. Todos os testes foram feitos com as enzimas recém-produzidas e purificadas.

5.1.4 Teste de termoestabilidade.

Todas as enzimas purificadas tiveram sua termoestabilidade medida em diferentes tampões com intuito de definir qual era a condição mais favorável para a cristalização. Após a diluição até concentração adequada (em torno de 1,5 mg/mL), foram feitos experimentos utilizando diferentes tampões e pHs. Os tampões utilizados, bem como os Tm calculados, estão descritos na tabela 4.

O tampão de eluição utilizado na purificação das enzimas era o Tris pH 7,5. Nas condições em que o Tm excedia pelo menos 5 °C o Tm do tampão de eluição, a enzima era dialisada para o novo tampão e utilizada nos experimentos de cristalização. Nos demais casos, o tampão Tris pH 7,5 era mantido.

Por exemplo, a enzima A9 possui Tm de 45 °C em tampão Tris pH 7,5 e Tm de 60 °C em tampão de ácido cítrico pH 5,0. Pelo aumento do Tm e, portanto, da estabilidade, o ácido cítrico foi escolhido como tampão. Já no caso das enzimas H6a e H6b, como o tampão que gerou maior Tm era o próprio Tris pH 7,5 ou a diferença era menor do que 5 °C, o tampão de eluição foi mantido.

Tabela 4 - Determinação dos valores de Tm (em °C) de cada enzima testada nos diferentes tampões e pHs. Os maiores valores encontrados estão destacados em vermelho. Os valores designados com um X representam que a determinação do Tm não foi possível.

Tampão	pH	A9	B5	B9	D5	E5	F4	F5	F6	F7	F12	G5	H5	H6a	H6b	H6c
Ácido cítrico	5,0	60	57	62	43	67	53	61	49	49	43	48	55	71	60	51
Citrato de sódio	5,5	59	48	63	49	66	X	46	54	51	46	X	58	X	48	X
MÊS	5,8	49	55	62	46	62	54	62	51	49	44	X	50	70	61	52
Fosf. de potássio	6,0	55	52	62	52	66	53	63	56	52	45	X	59	70	60	53
Citrato de sódio	6,0	55	53	61	52	63	55	63	58	51	47	X	60	70	62	55
MES	6,0	56	57	62	49	69	53	61	52	51	43	47	52	70	61	52
MES	6,2	56	57	62	51	68	43	48	53	52	43	43	53	66	57	41
MES	6,5	55	57	61	53	54	53	60	54	49	43	49	54	62	52	54
Fosf. de sódio	6,6	50	55	59	54	63	54	62	55	50	45	52	60	70	61	53
Fosf. de sódio	7,0	48	57	58	55	69	46	58	56	46	45	39	61	69	59	45
Fosf. de potássio	7,0	49	53	59	55	63	55	63	56	48	45	50	60	70	61	54
Hepes	7,0	48	50	58	54	66	51	63	55	49	44	X	56	71	59	51
Acet. de amônio	7,3	53	66	61	53	68	36	51	54	51	44	X	56	61	56	40
Fosf. de sódio	7,5	44	54	55	54	63	54	63	55	47	44	X	60	69	61	53
Tris	7,5	45	57	57	54	63	54	63	54	48	43	52	57	69	62	50
Imidazol	8,0	43	48	55	54	60	54	62	53	47	40	50	53	66	58	54
Hepes	8,0	42	55	56	54	63	54	63	53	45	44	53	56	69	62	52
Tris	8,0	41	54	56	54	62	53	61	53	46	43	51	57	70	60	52
Tricina	8,0	43	52	57	55	63	55	63	54	44	44	49	56	70	61	55
Bicina	8,0	42	53	56	55	62	55	63	54	43	43	52	56	70	61	55
Bicina	8,5	37	53	53	54	64	50	62	53	33	44	X	55	67	61	49
Tris	8,5	38	47	54	54	57	30	29	53	44	43	46	56	45	48	X
Bicina	9,0	33	59	49	52	65	52	63	53	X	43	50	54	66	62	37
Água		49	54	59	54	62	54	63	54	51	43	50	55	71	61	48

A enzima que apresentou maior T_m foi a H6a (71 °C). Curiosamente, suas construções apresentaram T_m consideravelmente menor. Retirando-se o modulo 2 na posição carboxiterminal (H6b), o T_m foi reduzido em aproximadamente 10 °C. Retirando-se os dois módulos acessórios (H6c), a redução aproximada foi de 20 °C. Apesar da função dos módulos ainda não ser comprovada, o experimento mostra a importância dos mesmos no aumento da estabilidade da enzima.

As enzimas C6 e C9, por terem sido as últimas a serem expressas e não haver mais tempo, não foram submetidas aos experimentos de termoestabilidade.

5.1.5 Cristalização e modelo estrutural da enzima H6.

Logo no primeiro ensaio de cristalização feito no IFSC com o robô *Honey Bee*, foram obtidos cristais da enzima H6a. Em três dias, cristais pontiagudos em forma de lança (figuras 36B e 36D) já estavam totalmente formados em quatro condições diferentes: 1. 20 % de PEG 4000, 20 % de propanol e 0,1 M de citrato de sódio, pH 5,6; 2. 0,2 M de sulfato de amônio, 30 % de PEG MME 5000 e 0,1 M de MES, pH 6,5; 3. 0,2 M de acetato de amônio, 30 % de PEG 4000 e 0,1 M de citrato de sódio, pH 5,6; 4. 0,2 M de sulfato de amônio, 30 % de PEG 8000 e 0,1 M de cacodilato de sódio, pH 6,5. Porém, em testes de difração eles não alcançaram boa resolução e apresentaram reflexões alongadas.

A partir das condições iniciais encontradas foram feitos refinamentos, tanto no IFSC quanto no ESRF, onde vários cristais foram obtidos e vários conjuntos de dados foram coletados. Diferentes formatos de cristais foram obtidos levando de 3 dias a 2 meses para crescer (figura 36). A diferença no tempo de crescimento se dava mesmo entre cristais com mesmo formato e mesmo grupo espacial determinado.

O tipo cristalino da figura 36E foi o único que não teve seus dados coletados por ter baixo poder de difração.

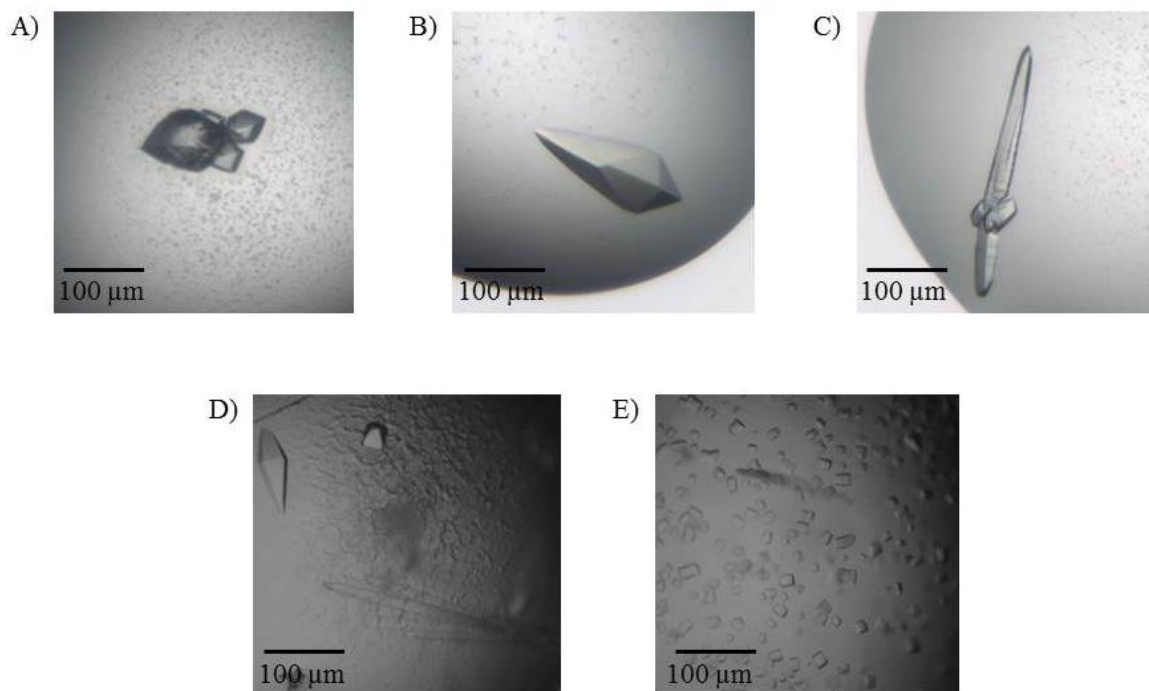


Figura 36 - Fotos dos diferentes formatos de cristais da enzima H6a. Note no item D a presença de duas formas cristalinas na mesma gota. As condições de cristalização de cada item foram: A) 25 % de PEG 3350, 0,2 M de cloreto de sódio e 0,1 M de Tris, pH 8,5; B) 25 % de PEG 3350, 0,2 M de sulfato de amônio e 0,1 M de HEPES, pH 7,5; C) 25 % de PEG 3350 e 0,1 M de Tris, pH 8,5; D) 22,5 % de PEG 4000, 14 % de isopropanol e 0,1 M de citrato de sódio, pH 6,0; E) 25 % de PEG 4000, 18 % de isopropanol e 0,1 M de citrato de sódio, pH 6,0.

Todos os conjuntos de dados coletados foram processados e os que obtiveram melhores estatísticas e resolução foram refinados. As estatísticas de coleta de dados e refinamento estão resumidas na tabela 5. Os conjuntos de dados foram nomeados de A a D de acordo com os itens da figura 36, onde estão representadas as fotos de seus respectivos cristais. O conjunto de dados D foi coletado do cristal mais fino e alongado localizado na parte inferior da figura 36D. O conjunto de dados E foi coletado de um cristal similar ao do conjunto D, que foi submetido ao experimento de *soaking* com o substrato celopentaose por 24 horas.

Os modelos gerados a partir dos conjuntos de dados foram validados pelos programas Procheck¹³⁰ e Molprobit¹³¹, e apresentaram boas estatísticas. Algumas delas estão resumidas na tabela 5.

O conjunto A, por possuir resolução atômica, levou mais tempo para o refinamento. Como sua estrutura foi resolvida há pouco tempo, a estrutura reportada aqui é o melhor modelo construído até então. Apesar de ainda não estar pronto, já possui boas estatísticas e será usado apenas para discussões em que seus dados não deixem dúvidas.

Tabela 5 - Estatísticas dos dados cristalográficos da enzima H6a coletados em quatro grupos espaciais diferentes na forma apo (conjuntos de A a D) e complexada com o substrato celopentaose (conjunto E).

Conjunto de dados	A	B	C	D	E
Estatísticas de coleta					
Fonte	Sincrotron ESRF	Sincrotron ESRF	Sincrotron LNLS	Sincrotron ESRF	Sincrotron ESRF
Comprimento de onda (Å)	0,97	0,97	1,46	0,97	0,97
Resolução (Å)	44,30-1,10 (1,11-1,10)	44,40-2,30 (2,34-2,30)	46,12-1,75 (1,76-1,75)	43,90-1,70 (1,72-1,70)	57,10-1,75 (1,76-1,75)
Grupo espacial	P2 ₁ 2 ₁ 2 ₁	I4 ₁ 22	P2 ₁	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cela unitária					
a (Å)	90,44	92,27	62,10	90,70	91,00
b (Å)	91,31	92,27	154,00	90,70	91,00
c (Å)	122,77	326,85	64,60	120,20	124,30
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 117, 90	90, 90, 90	90, 90, 90
Nº de reflexões únicas	385868	31946	107468	53165	53268
Multiplicidade	4,4 (3,3)	6,4 (6,3)	3,6 (3,4)	2,6 (2,7)	5,5 (5,5)
CC1/2	0,99 (0,65)	0,96 (0,65)	0,99 (0,62)	0,99 (0,66)	0,99 (0,65)
Completeza (%)	98,2 (94,6)	99,8 (99,8)	99,0 (99,0)	95,9 (95,9)	99,6 (99,6)
<I>/<σ(I)>	14,0 (1,9)	4,8 (1,7)	10,9 (1,8)	8,5 (2,1)	10,3 (1,9)
R _{pim} ^a	0,03 (0,39)	0,10 (0,41)	0,06 (0,46)	0,05 (0,37)	0,05 (0,38)
Estatísticas de refinamento					
Rwork ^b (%)	15,4	18,8	15,3	16,4	17,9
Rfree ^b (%)	16,6	23,5	18,8	19,4	20,8
B-fator médio (Å ²)					
Proteína	15,1	39,3	15,2	22,0	21,7
Solvente	29,3	47,2	27,0	34,8	36,3
Ligante	X	X	X	X	29,8
Rmsd comprimento de ligação (Å)	0,005	0,007	0,014	0,007	0,005
Rmsd ângulos de ligação (°)	1,12	1,06	1,43	1,09	0,99
Ramachandran outliers (%)	0,1	0	0,2	0,2	0

^a $R_{pim} = \sum_{hkl} \sqrt{(1/(n-1)) \sum_{j=1}^n (|I_{hkl,j} - \langle I_{hkl} \rangle|) / \sum_{hkl} \sum_j (I_{hkl,j})}$, onde I_{hkl} é o valor da intensidade para uma única reflexão dentro de um conjunto de reflexões simetricamente relacionadas, e $\langle I_{hkl} \rangle$ é a média das intensidades para um conjunto j de reflexões simetricamente relacionadas. ^b R_{work} ou $R_{free} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$, onde F_{obs} e F_{calc} correspondem aos fatores de estrutura observados e calculados, respectivamente. Para o cálculo do R_{free} , foram utilizadas somente 5 % das reflexões omitidas do refinamento. Os valores em parênteses se referem à camada mais externa das faixas de resolução.

A primeira estrutura resolvida foi a partir do conjunto de dados C. Após o processamento, redução e escalonamento dos dados, as fases foram determinadas por substituição molecular a partir da estrutura de uma endoglucanase de *Clostridium*

cellulovorans (36 % de identidade), cujo código PDB é 3NDY (não publicada). O resultado final apresentou um Z score (função de translação) de 15,1 e LLG de 150. Apesar da baixa identidade, o domínio catalítico possui um enovelamento muito parecido, tornando viável a utilização da substituição molecular. O baixo valor de LLG se deve principalmente à ausência dos módulos acessórios no modelo inicial.

As demais estruturas tiveram suas fases determinadas também por substituição molecular, utilizando-se a estrutura H6a_C como modelo.

5.1.5.1 Modelo cristalográfico.

Embora a proteína H6 tenha cristalizado em diferentes grupos espaciais, os modelos gerados em cada conjunto de dados são praticamente idênticos. A única diferença mais notável está em uma pequena alça formada pelos resíduos G183, G184 e K185, chegando a um RMSD de 3 Å em relação aos carbonos alfa. Essa alça está exposta ao solvente e está localizada no domínio catalítico no lado oposto ao sítio de interação com o substrato.

Praticamente em todos os modelos, a alça compreendida entre os resíduos L399 e G403 está ausente por não possuir densidade eletrônica bem definida. Apenas na cadeia A do modelo H6_C essa região foi completamente modelada. Os primeiros resíduos de cada modelo também estão ausentes por estarem desordenados; todos começam a ser modelados a partir do resíduo I18, com a exceção do modelo H6_D, que começa com o resíduo I19.

Os modelos H6_B, H6_D e H6_E possuem apenas uma molécula na unidade assimétrica, enquanto os modelos H6_A e H6_C possuem duas. Nos dois últimos casos, as moléculas que compõem a unidade assimétrica não apresentam diferenças significativas.

De forma geral, a estrutura tridimensional da enzima H6 é formada por três domínios distintos (figuras 37, 39 e 41): um domínio catalítico, compreendido entre os resíduos 1 e 331, e dois outros domínios acessórios menores, sendo o primeiro (módulo 1) delimitado pelos resíduos 332 e 429, e o segundo (módulo 2) pelos resíduos 430 e 533. Os diferentes domínios estão muito próximos uns dos outros, estando conectados por apenas alguns aminoácidos.

O módulo 1, além de estar ligado covalentemente ao domínio catalítico, possui contato com o mesmo através de interação hidrofóbica e algumas ligações de hidrogênio entre suas fitas $\beta 1$ e $\beta 8$ (porção inicial) e as hélices $\alpha 7$ e $\alpha 8$ do domínio catalítico.

A relação entre os módulos 1 e 2 se dá basicamente por estarem ligados covalentemente e por uma pequena área formada pela fitas $\beta 2$ e $\beta 8$ (porção final) do módulo 1 juntamente com duas alças do módulo 2, localizados entre as fitas $\beta 3$ e $\beta 4$ e entre as fitas $\beta 7$ e $\beta 8$.

Apesar de o domínio catalítico estar separado sequencialmente do módulo 2 por 98 aminoácidos (relativos ao módulo 1), eles interagem entre si em uma região relativamente próxima ao sítio de interação com o substrato. Essa interação se dá principalmente entre a alça que conecta os elementos $\beta 6$ e $\alpha 6$ do domínio catalítico com a alça que conecta os elementos $\beta 4$ e $\beta 5$ do módulo 2.

Até então, esse encaixe entre domínios foi observado apenas uma vez em uma estrutura de sialidase de *Micromonospora viridifaciens*, depositada no PDB com o código 1EUT¹³². Apesar do encaixe entre os domínios ser parecido, as estruturas não apresentam nenhum módulo acessório em comum.

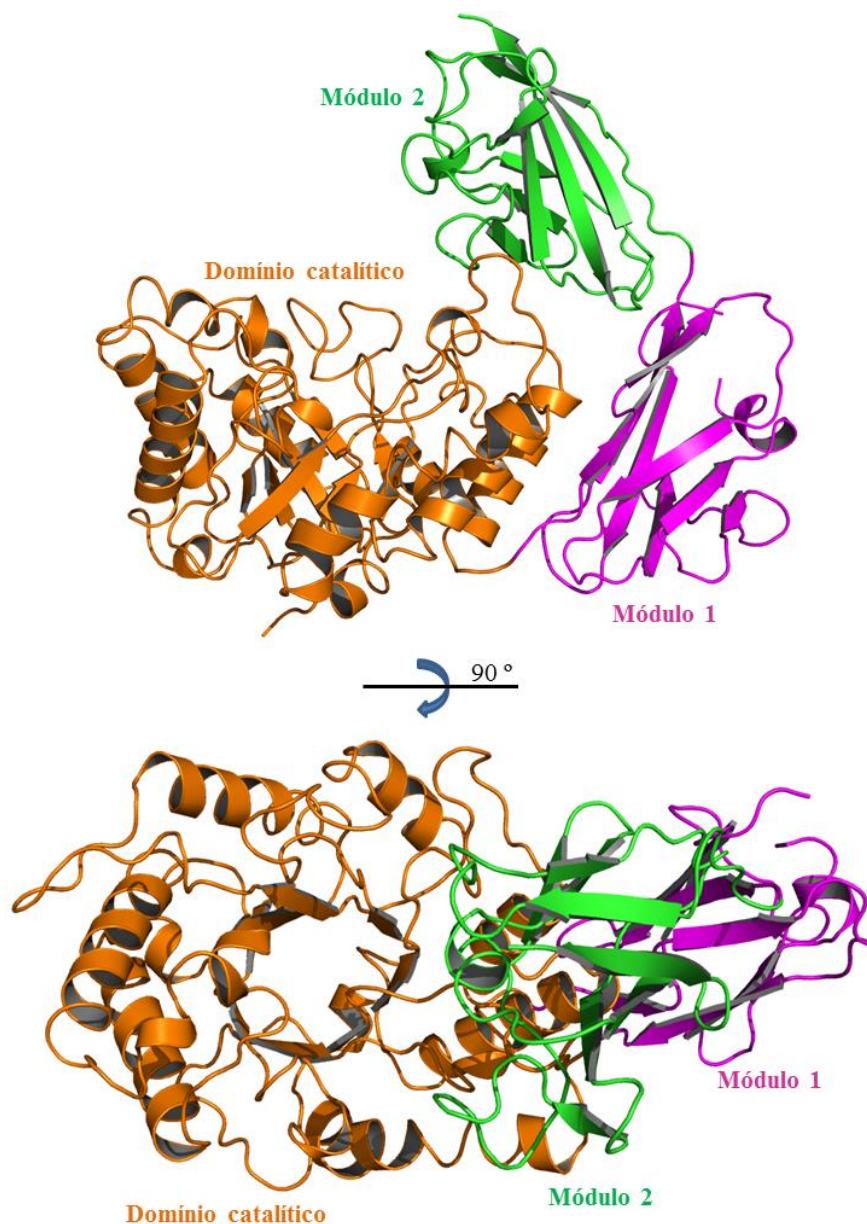


Figura 37 - Modelo cristalográfico da enzima H6. O domínio catalítico está representado pela cor de laranja, o módulo 1 está magenta e o módulo 2 está verde.

Além dessa interação pouco conhecida entre os domínios em celulasas, a interação entre o domínio catalítico e o módulo 2 possui uma característica peculiar. Três resíduos de triptofano (W225, W231 e W232), localizados na alça que conecta os elementos β_6 e α_6 do domínio catalítico, estão dispostos de forma alinhada com outros dois triptofanos (W479 e W481) localizados na alça que conecta os elementos β_4 e β_5 , do módulo 2 (figura 38).

O primeiro triptofano dessa sequência está próximo do sítio ativo e, a partir da sobreposição dessa estrutura com uma celulase da família GH5 (CelX) complexada com celotetraose (código PDB 4HU0, ainda não publicada), percebe-se que esse resíduo poderia

atuar na estabilização de substratos na posição +2 do sítio ativo. A glicose que está localizada no subsítio +2 é estabilizada por um triptofano da CelX, cuja cadeia lateral se posiciona paralelamente à molécula de glicose a 4,4 Å de distância. O resíduo W225 da H6 possui mesma orientação com distância equivalente (4,7 Å), porém, do lado oposto do carboidrato.

Pela proximidade e pelo alinhamento dos outros quatro triptofanos, é possível que eles sejam responsáveis pela interação com carboidratos, fazendo com que aumente a afinidade da enzima por substratos de cadeias longas¹³³. Desta forma, estariam presentes os subsítios +3, +4 e, talvez, até +5 e +6. A sobreposição entre H6 e CelX apresenta um RMSD de 1,635 Å e os resíduos catalítico estão na mesma posição.

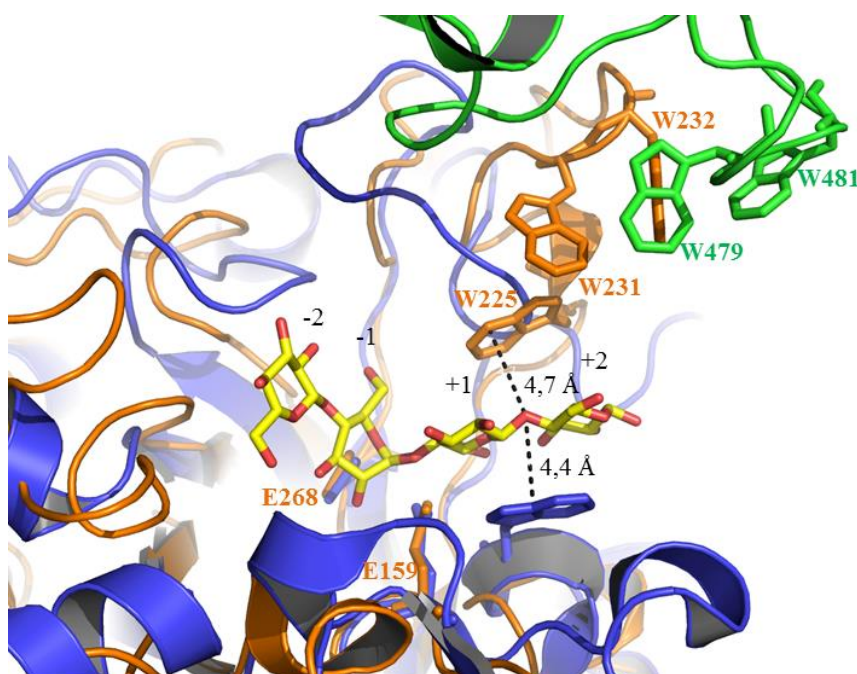


Figura 38 - Sobreposição do modelo cristalográfico de H6 (domínio catalítico na cor de laranja e módulo 2 na cor verde) com a sialidase de *Micromonospora viridifaciens* (cor azul, PDB: 4HU0). Em amarelo está representada uma cetotetraose ligada à sialidase, ocupando do subsítio -2 ao +2.

5.1.5.2 Domínio catalítico.

O domínio catalítico da enzima H6 é classificado como integrante da família GH5 e, como todos os membros dessa família, possui um enovelamento tipo barril (α/β)₈. Essa estrutura é formada por oito fitas- β internas alinhadas com oito α -hélices externas. As fitas- β

e as α -hélices estão conectadas sequencialmente e de forma intercalada através de alças e outros pequenos segmentos α -hélices como mostrado na figura 39.

Além do enovelamento, os resíduos catalíticos também são conservados em toda a família GH5. Dois glutamatos estão localizados na porção carboxiterminal das fitas- β , sendo que o resíduo E159 está posicionado na alça entre a fita β 4 e hélice α 4 e o resíduo E268 está na alça entre a fita β 7 e hélice α 7.

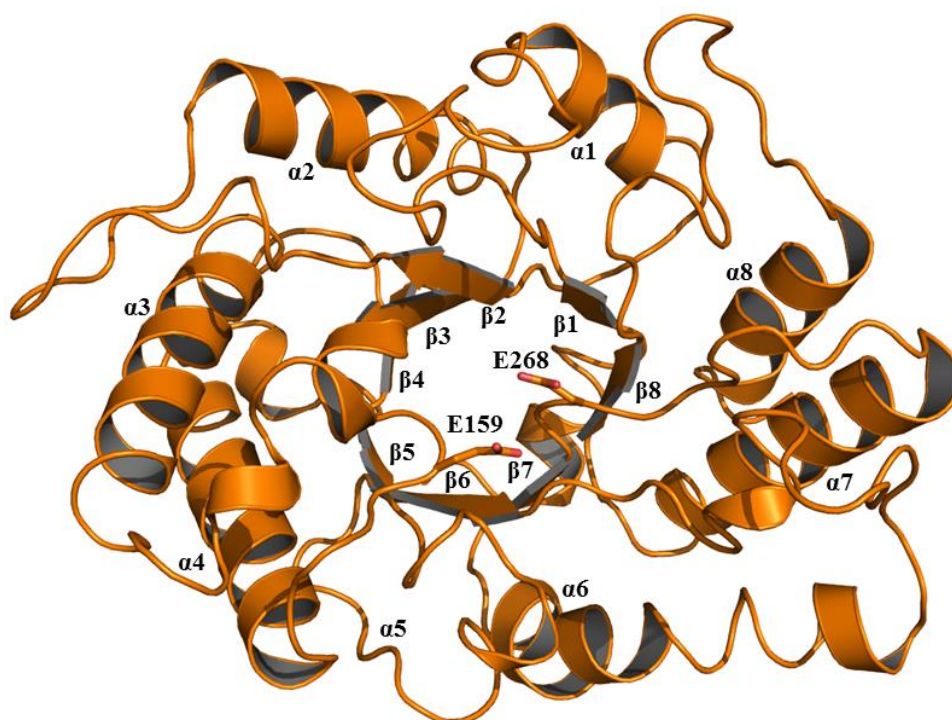


Figura 39 - Modelo cristalográfico do domínio catalítico da enzima H6, constituído de 8 fitas β e oito α -hélices. Os resíduos catalíticos E159 e E268 estão evidenciados no centro da figura.

5.1.5.3 Sítio catalítico.

Como já dito anteriormente, a estrutura cristalográfica H6a_E proveio de um *soaking* de um cristal de H6 com o substrato celopentaose. Porém, a densidade eletrônica experimental mostra claramente no sítio ativo uma celotetraose, cujo carbono anomérico da extremidade redutora está a apenas 2,9 Å do oxigênio OE2 (grupo carboxila) do resíduo catalítico E159 e a 3,2 Å do oxigênio OE1 do resíduo catalítico E268 (figura 40A). Além da densidade eletrônica bem definida, o carboidrato está delimitado em uma das pontas pelos resíduos catalíticos e

pela outra ponta pela cadeia lateral do resíduo R234 da molécula simétrica. Portanto, a celopentaose provavelmente foi clivada durante o período de 24 horas de *soaking*.

A celotetraose presente na estrutura ocupa os subsítios -3, -2 e -1 e está estabilizada por ligações de hidrogênio com os resíduos N36, H113, H114, N158, W301 e N303 e por uma interação CH- π ¹³³⁻¹³⁴⁻¹³⁵ com o resíduo W47 (figura 40B). Infelizmente, não há indício algum da presença da glicose resultante da clivagem da celopentaose. Possivelmente esta glicose ocupou o subsítio +1, como mostrado na estrutura 4HU0 da figura 38, antes de ser clivado.

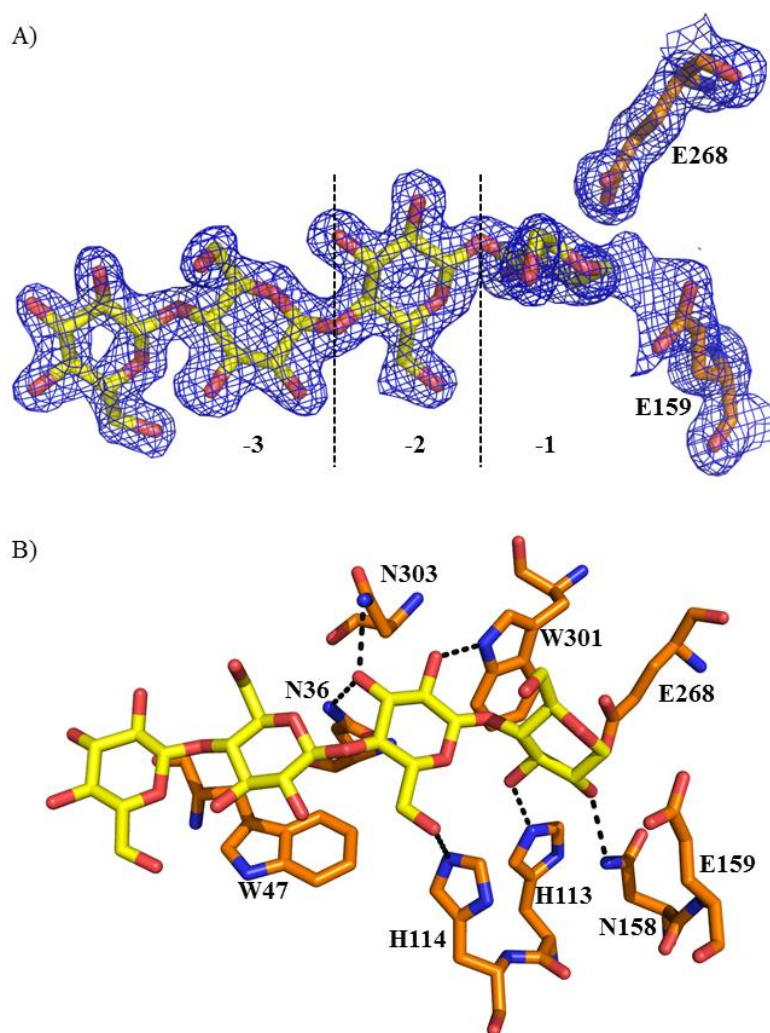


Figura 40 - Celotetraose ligada a enzima H6. A) Modelo da celotetraose e dos resíduos catalíticos ajustados na densidade eletrônica (1σ) obtida pelos dados experimentais. Três das quatro glicoses ocupam os subsítios -1, -2 e -3. Observe a continuidade da densidade eletrônica entre a celotetraose e o resíduo E159. B) Estabilização do substrato no sítio ativo pelos resíduos indicados. As linhas tracejadas representam ligações de hidrogênio.

5.1.5.4 Módulos acessórios.

A enzima H6 possui dois módulos acessórios ao domínio catalítico. O primeiro (módulo 1), situado entre o domínio catalítico e o módulo 2, é classificado como CBM X2¹³⁶ tanto pelo BLAST¹³⁷ quanto pelo PFAM¹³⁸ através da análise de sua sequência de aminoácidos. Já o módulo 2, situado na porção carboxiterminal, não possui classificação através da análise de sua sequência.

Os dois módulos possuem tamanhos equivalentes (93 e 103 aminoácidos) e suas estruturas terciárias são muito parecidas (figura 41). Ambos possuem um enovelamento sanduiche- β composto de duas folhas- β , com quatro fitas antiparalelas de um lado e três do outro e uma pequena α -hélice entre as fitas β 5 e β 6. Os módulos possuem um RMSD de 3,1 Å onde a diferença se dá basicamente pelas alças que conectam os elementos de estrutura secundária.

Curiosamente, os domínios possuem uma identidade sequencial de apenas 17 % em relação aos aminoácidos, apesar de possuírem o mesmo enovelamento. Outro fato curioso é a ausência de triptofanos no módulo 1, pois tais resíduos geralmente estão presentes em módulos responsáveis por interação com carboidratos¹³⁶. O módulo 2 possui 3 triptofanos, sendo que dois deles estão alinhados com outros triptofanos do domínio catalítico como descrito acima (no item 3.5).

O módulo 1 também se destaca por ser a região com maior mobilidade da molécula. No caso da estrutura H6a_E, ele possui um B-fator médio de 32,8 Å² comparado a 22,7 Å² do módulo 2 e 20,0 Å² do domínio catalítico. Nas outras estruturas de H6a, a relação entre os B-fatores médios dos domínios se mantém apesar dos valores absolutos serem diferentes.

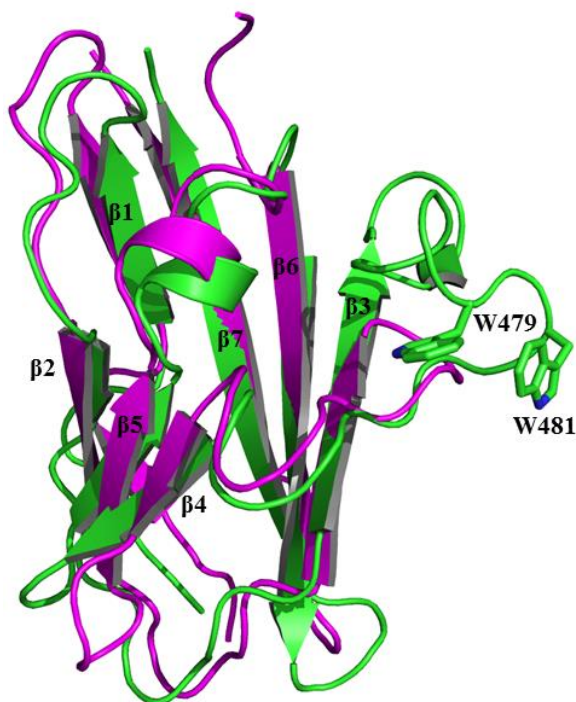


Figura 41 - Modelo cristalográfico dos módulos 1 (em magenta) e 2 (em verde) da enzima H6. Os módulos foram superpostos e são constituídos de 7 fitas β e uma pequena α -hélice.

5.1.6 Cristalização das truncagens da enzima H6

Além da enzima na sua forma completa (H6a), foram geradas duas outras construções da enzima H6. Na primeira, foi retirado o módulo 2 (H6b) e na segunda, retirou-se tanto o módulo 1 quanto o 2, deixando-se somente o domínio catalítico (H6c).

Nos testes iniciais de cristalização, tanto a enzima H6b quanto H6c foram cristalizadas em apenas alguns dias, em algumas condições que variavam em torno de: 4 % de PEG 8000 e 0,1 M de Tris pH 8,5; e 5 % de PEG 6000 e 0,1 M de Heps 7,0. Essas condições foram refinadas e os cristais testados (figura 42).

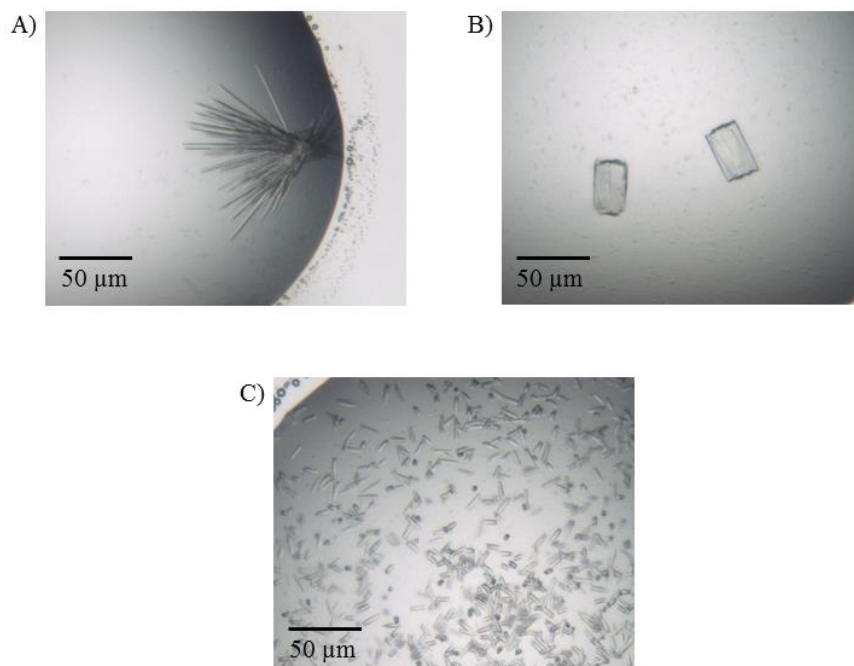


Figura 42 - Fotos dos cristais de H6b e H6c. A) H6b, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5; B) H6b, condição: 4 % de PEG 8000 e 0,1 M de Hepes, pH 7,0; C) H6c, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5.

No experimento de difração de raios-X ficou comprovado que os cristais eram realmente de proteínas. Apesar de serem observadas poucas reflexões e na região de baixa resolução, a distância entre elas indica uma cela unitária grande demais para ser um cristal de sal (figura 43). A resolução alcançada pela difração dos cristais não foi suficiente para que conjuntos de dados fossem coletados.

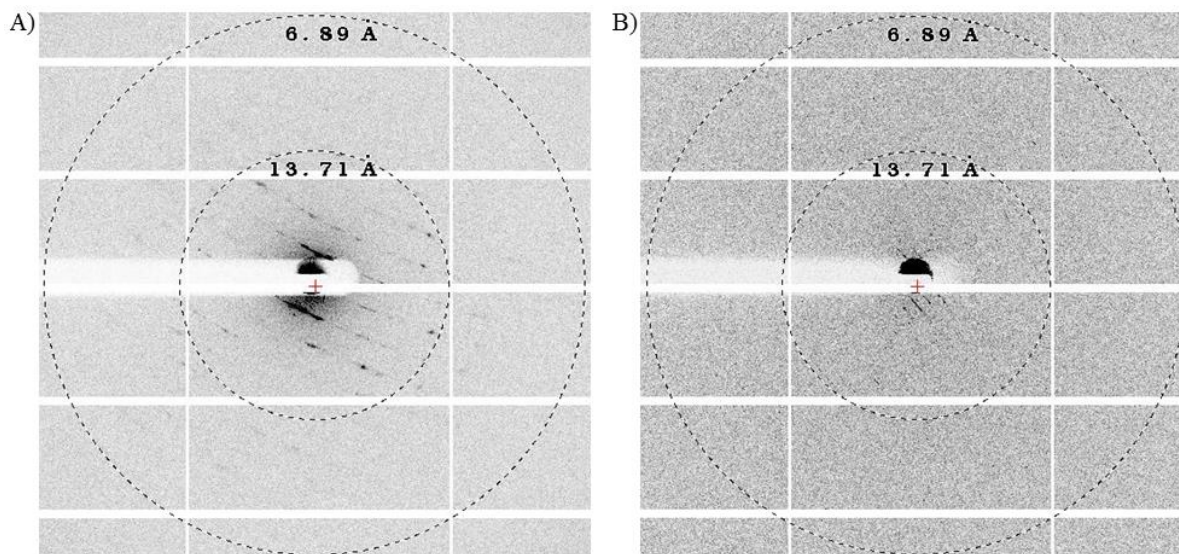


Figura 43 - Padrões de difração das proteínas: A) H6b, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5; B) H6c, condição: 5 % de PEG 6000 e 0,1 M de Tris, pH 8,5.

5.1.7 Cristalização e modelo estrutural das enzimas D5, F5 e G5

As três beta-glicosidasas de *B. licheniformis* foram cristalizadas já nos primeiros ensaios de cristalização. Nas primeiras condições encontradas para cada enzima (tabela 6), os cristais gerados não tinham bom poder de difração. Com o refinamento dessas condições, novos cristais foram gerados e foram feitas coletas de dados das três enzimas com resoluções de 2,3 Å, 2,4 Å e 1,9 Å para as enzimas D5, F5 e G5, respectivamente.

Tabela 6 - Condições de cristalização das enzimas D5, F5 e G5. Estão destacadas em negrito as condições em que foram obtidos os cristais com conjunto de dados coletados. As demais condições foram encontradas inicialmente pelos ensaios em robô de cristalização.

D5	0,2 M de sulfato de lítio, 25 % de PEG 4000 e 0,1 M de Bis-tris pH 5,5
	0,2 M de acetato de amônio, 25 % de PEG 4000 e 0,1 M de Tris pH 8,5
	0,1 M de sulfato de amônio, 15 % de PEG 4000 e 0,1 M de Tris pH 7,0
F5	20 % de PEG 3350 e 0,1 M de ácido málico pH 7,0
	20 % de PEG 3350 e 0,2 M de acetato de sódio pH 7,9
	0,3 M de cloreto de sódio, 25 % de PEG 3350 e 0,1 M de hepes pH 7,0
G5	25 % de PEG 4000 e 0,1 M de Tris pH 8,5
	20 % de PEG 4000 e 0,2 M de formato de sódio pH 7,2
	0,2 M de acetato de sódio, 15 % de PEG 4000 e 0,1 M de Tris pH 8,5

Após o escalonamento e redução dos dados, as fases foram encontradas por substituição molecular utilizando como modelo a beta-glicosidase de *E. coli*, cujo código PDB é 2XHY¹³⁹. Esse modelo foi selecionado por possuir maior grau de identidade com as enzimas, variando em torno de 50 %.

Os modelos finais apresentaram boas estatísticas que, juntamente com dados das coletas, estão resumidas na tabela 7.

Tabela 7 - Estatísticas dos dados cristalográficos das enzimas D5, F5 e G5.

Enzima	D5	F5	G5
Estatísticas de coleta			
Fonte	Sincrotron ESRF	Sincrotron ESRF	Sincrotron ESRF
Comprimento de onda (Å)	0,97	0,97	0,97
Resolução (Å)	49,1-2,3 (2,32-2,30)	55,9-2,4 (2,42-2,40)	58,2-1,9 (1,91-1,90)
Grupo espacial	P3 ₁ 21	P4 ₁ 2 ₁ 2	P2 ₁
Cela unitária			
a (Å)	98,27	99,44	100,00
b (Å)	98,27	99,44	71,36
c (Å)	285,73	223,97	169,30
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 101, 90
Nº de reflexões únicas	72098	46056	118102
Multiplicidade	5,5 (5,5)	16,1 (15,6)	2,7 (2,1)
CC1/2	0,99 (0,87)	0,99 (0,79)	0,99 (0,64)
Completeza (%)	100 (100)	99,9 (99,8)	98,0 (97,7)
$\langle I \rangle / \langle \sigma(I) \rangle$	10,6 (2,0)	25,9 (1,6)	7,6 (1,4)
R _{pim} ^a	0,06 (0,49)	0,02 (0,47)	0,06 (0,48)
Estatísticas de refinamento			
Rwork ^b (%)	16,9	20,0	15,6
Rfree ^b (%)	20,9	23,9	18,8
B-fator médio (Å ²)			
Proteína	33,5	56,3	36,6
Solvente	40,3	61,4	43,6
Rmsd comprimento de ligação (Å)	0,007	0,006	0,007
Rmsd ângulos de ligação (°)	1,05	0,91	1,03
Ramachandran outliers (%)	0,15	0,35	0

^a $R_{pim} = \sum_{hkl} \sqrt{1/(n-1)} \sum_{j=1}^n (|I_{hkl,j} - \langle I_{hkl} \rangle|) / \sum_{hkl} \sum_j (I_{hkl,j})$, onde I_{hkl} é o valor da intensidade para uma única reflexão dentro de um conjunto de reflexões simetricamente relacionadas, e $\langle I_{hkl} \rangle$ é a média das intensidades para um conjunto j de reflexões simetricamente relacionadas. ^b R_{work} ou $R_{free} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$, onde F_{obs} e F_{calc} correspondem aos fatores de estrutura observados e calculados, respectivamente. Para o cálculo do R_{free} , foram utilizadas somente 5 % das reflexões omitidas do refinamento. Os valores em parênteses se referem à camada mais externa das faixas de resolução.

Os cristais das enzimas D5, F5 e G5 possuem respectivamente 3, 2 e 4 moléculas na unidade assimétrica. Como todas pertencem à família GH1, apresentam o mesmo envelamento tipo barril $(\alpha/\beta)_8$, característico da família (figura 44)¹⁴⁰. Os elementos de estrutura secundária que formam a base estrutural do barril, e que caracterizam a família como

pertencente ao clã GH-A, são oito fitas- β internas alinhadas com oito α -hélices externas. Essa base estrutural e a numeração dos elementos de estrutura secundária são os mesmos descritos no item 3.5.2 para a enzima H6 da família GH5. A sobreposição das três estruturas revela que são muito parecidas, possuindo um RMSD de apenas 0,959 Å.

Diferentemente das enzimas da família GH5, as conexões entre os elementos de estrutura secundária que formam a base do barril são mais extensos, formadas por alças maiores e por algumas pequenas fitas- β e α -hélices. Também, são essas conexões que constituem as principais diferenças entre as enzimas D5, F5 e G5. As alças localizadas no lado oposto ao sítio ativo das enzimas são muito parecidas, mas a maioria das alças localizadas no mesmo lado é diferente entre si.

As alças 1, 4, 6 e 8 (que conectam os elementos β_1 , β_4 , β_6 e β_8 aos α_1 , α_4 , α_6 e α_8 , respectivamente) formam a entrada do sítio ativo. A enzima D5 possui uma abertura relativamente grande limitada de um lado pela alça 6 e pelo outro pelas alças 1 e 8, onde a distância entre as cadeias principais é de pouco mais de 14 Å (figura 44). Já a enzima G5 possui uma abertura muito menor, pois a cadeia principal da alça 6 está a apenas 5,5 Å da alça 8 e a 10,4 Å da alça 4.

A enzima F5 não possui densidade eletrônica definida em parte das alças 4 (resíduos 173-187), 6 (resíduos 307-324) e 8 (resíduos 442-446), portanto, não é possível definir a abertura do sítio ativo.

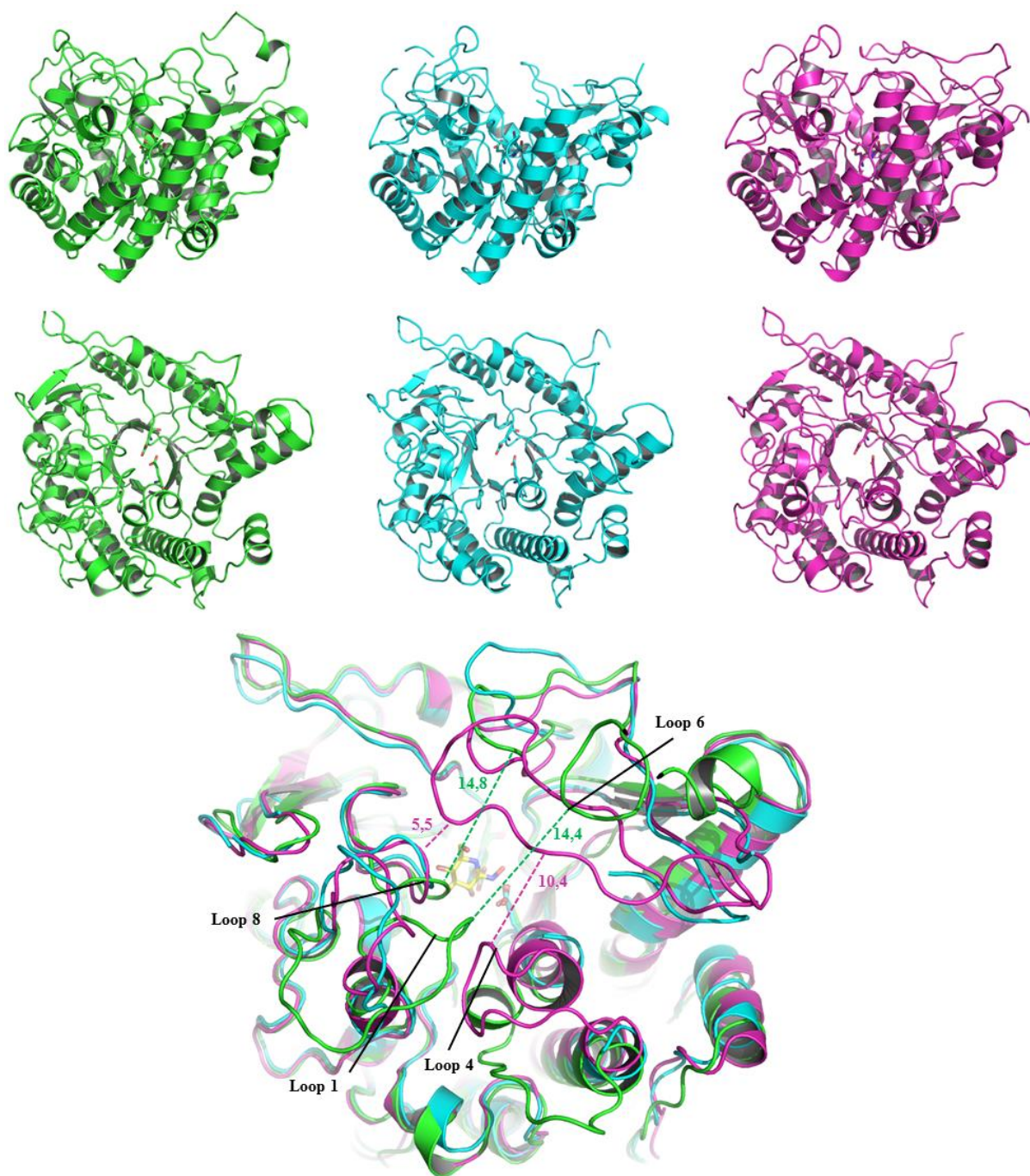


Figura 44 - Modelos cristalográficos das enzimas D5 (em verde), F5 (em ciano) e G5 (em magenta). Abaixo está representada a sobreposição das três estruturas evidenciando a face em que se encontra o sítio ativo. As linhas pontilhadas (com as cores relacionadas com cada estrutura) representam a distância em angstroms entre as alças. No centro está representado o inibidor gluco-hidroximolactam (PDB:2J78)¹⁴¹ ligado no sítio ativo.

Os resíduos que interagem com o substrato no sítio catalítico são muito conservados entre os membros da família GH1. De acordo com a estrutura de outras beta-glicosidases publicadas¹⁴¹⁻¹⁴⁴, nove resíduos são responsáveis pela estabilização do substrato, incluindo os resíduos catalíticos (tabela 8 e figura 45). O alinhamento estrutural e sequencial dessas três

estruturas publicadas com as beta-glicosidases aqui estudadas mostra que esses resíduos são os mesmos nas diferentes enzimas e possuem as mesmas disposições, com exceção de duas posições (8 e 9). Na posição 8 as três enzimas possuem uma serina enquanto outras enzimas (com exceção de 4B3K) possuem um glutamato. Na posição 9 as enzimas D5 e F5 possuem, respectivamente, uma alanina e uma treonina no lugar de um triptofano.

Tabela 8 - Alinhamento dos resíduos responsáveis pela interação com substrato das enzimas D5, F5, G5 e de outras beta-glicosidases publicadas (designadas pelo código PDB). As posições foram numeradas de acordo com a ordem sequencial crescente dos aminoácidos.

Estrutura	Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	Pos. 6	Pos. 7	Pos. 8	Pos. 9
D5	Q22	H129	N174	E175	Y307	E368	W415	S422	A423
F5	Q23	H121	N165	E166	Y298	E371	W418	S425	T426
G5	Q23	H124	N169	E170	Y301	E378	W425	S432	W433
4B3K	Q18	H119	N164	E165	Y299	E366	W416	S423	W416
2CET	Q20	H121	N165	E166	Y295	E351	W398	E405	W406
3F5J	Q29	H130	N175	E176	Y315	E386	W433	E440	W441
3WBA	Q31	H132	N177	E178	Y321	E394	W444	E451	W452

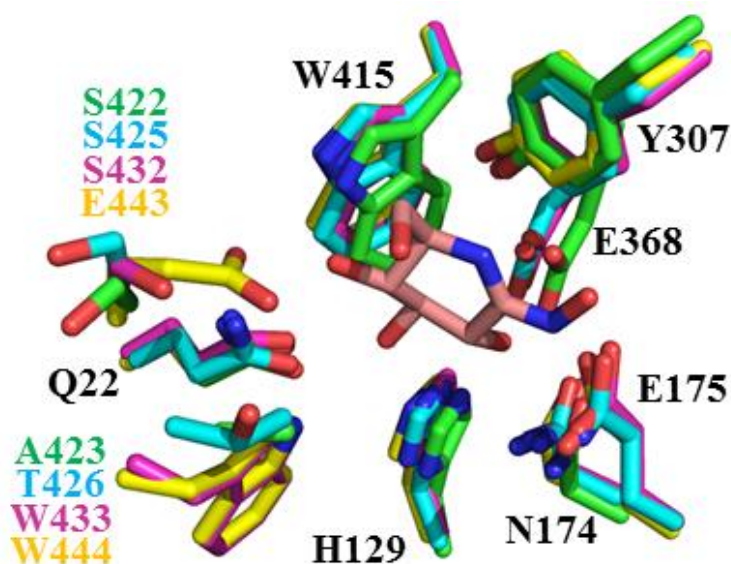


Figura 45 - Resíduos responsáveis pela interação com substrato. Estão representadas as estruturas D5 (verde), F5 (ciano), G5 (magenta) e 3WBA¹⁴⁴. No centro está modelado o inibidor de beta-glicosidase glucohidroximolactam (PDB: 2J78)¹⁴¹, cuja estrutura foi superposta. O número dos resíduos em preto está de acordo com a estrutura D5.

5.2 DISCUSSÃO

As hidrolases de glicosídeos formam uma grande família de enzimas responsáveis pela quebra das ligações formadas entre os vários tipos existentes de carboidratos e entre os carboidratos com outros tipos moleculares. A quantidade de enzimas descritas, que pertencem a essa grande família, já é muito grande e novos membros são constantemente descobertos. Por ser um grupo tão numeroso, as hidrolases de glicosídeos foram classificadas, até agora, em 132 famílias⁶⁵⁻¹⁴⁵⁻¹⁴⁶. Ainda assim, algumas famílias foram também classificadas em subfamílias, como é o caso da família GH5⁷².

Embora uma enorme quantidade de GHs tenha sido identificada, suas características funcionais e estruturais estão longe de ser totalmente compreendidas. Desta forma, esse estudo se propôs a avançar nos estudos estruturais do maior número possível de enzimas das famílias GH1 e GH5. Para tal fim, as enzimas aqui estudadas, juntamente com enzimas de outras famílias, foram clonadas e testadas, inicialmente, pela capacidade de expressão por método de alto rendimento¹⁰³⁻¹⁴⁷ feito pelo pós-doutorando César Moisés Camilo. Essa rápida clonagem e avaliação primária da expressão foi decisiva para que o todo o trabalho descrito nessa tese tenha sido feito em tempo adequado.

As famílias GH1 e GH5 são dois dos maiores grupos dentre as GHs, contendo, respectivamente, 4929 e 4016 sequências depositadas de acordo com o CAZY⁶⁴. Além do grande número de membros, as enzimas possuem muitas funções diferentes e são encontradas tanto sozinhas (apenas domínio catalítico) como conectadas sequencialmente com um número variado de módulos acessórios. Toda essa variabilidade se contrasta com o enovelamento do tipo barril (β/α)₈ que é conservado em todos os membros dessas duas famílias. De alguma forma essa mesma estrutura básica tem a capacidade de se adequar a diversas funções e é nesse ponto que os resultados obtidos nesse trabalho trazem algumas novidades.

A princípio, as enzimas GH1 e GH5, cujos clones estavam disponíveis, eram conhecidas apenas pelas suas sequências. Não havia nenhuma informação sobre atividade catalítica, número de domínios, limite dos possíveis domínios, preferência por substratos, etc. Como o foco do trabalho era o estudo estrutural, todas elas foram expressas e purificadas conjuntamente e logo foram destinadas aos experimentos de cristalização.

Por ter sido a primeira a ser cristalizada e ter sua estrutura determinada, a caracterização básica da atividade da enzima H6 foi feita, bem como a produção das novas construções, com o intuito de compreender melhor a função dos diferentes módulos

acessórios. Com o início do Doutorado Sanduiche feito no ESRF (Grenoble, França) foi dado prosseguimento nas tentativas de cristalização das demais enzimas e da proteína H6 ligada a substratos. Infelizmente, não houve tempo hábil para a caracterização enzimática das demais enzimas purificadas, que será feito posteriormente.

De todas as enzimas clonadas, apenas E12, F1 e F9 não foram obtidas na forma solúvel. Algumas tentativas de variação do protocolo de expressão foram feitas, sem que fosse observada sua presença na fração solúvel. É possível que um maior esforço na tentativa de expressá-las, com uma maior variação nas condições de expressão ou utilizando-se diferentes cepas de *E. coli*, possa torná-las solúveis. Mas, como o tempo até o fim do doutorado era curto, optei por seguir com o estudo das outras enzimas.

5.2.1 Atividade da H6.

Na caracterização enzimática, a H6 mostrou-se uma enzima versátil. Primeiramente, ela possui boa termoestabilidade. Apesar de ter uma atividade ótima a 55 °C, é capaz de chegar a 80 °C com 50 % da atividade em CMC. Além disso, nos testes de termoestabilidade, ela se mostrou igualmente estável em praticamente todos os tampões e pHs testados, com o Tm variando entre 66 e 71 °C. Outra característica de sua versatilidade é a capacidade de clivar diferentes tipos de carboidratos: lineares com ligação β -1,4-glicosídica (CMC), lineares com mistura de ligações β -1,4 e β -1,3-glicosídicas (lichenano) e ramificados (xiloglucano). Por ter a capacidade de clivar celulose e hemicelulose, duas das moléculas mais abundantes nas paredes celulares, a H6a torna-se um excelente candidato para aplicação na produção de etanol de segunda geração.

Muitas celulasas possuem módulos acessórios para auxiliá-las na clivagem dos substratos. Para compreender melhor a função dos módulos acessórios e também dos domínios catalíticos, muitos trabalhos expressaram e caracterizaram os domínios acessórios separadamente. Na maioria dos casos, a retirada do CBM faz com que a enzima tenha uma atividade reduzida¹⁴⁸⁻¹⁴⁹, principalmente com substratos insolúveis¹⁵⁰. Em alguns casos, a retirada do CBM não altera a atividade do domínio catalítico¹⁵¹ e, em outros, a retirada do CBM inativa completamente a atividade enzimática¹⁵²⁻¹⁵³. A enzima H6 parece fazer parte da terceira categoria, pois, mesmo variando-se as condições e os substratos, as truncagens da

enzima H6 não apresentaram atividade, comprovando que ambos os módulos são absolutamente importantes para a atividade enzimática.

Juntamente com a eliminação da atividade da H6, a retirada dos módulos acessórios reduz de forma considerável o T_m da enzima. A retirada do módulo 2 reduz o T_m em 9 °C e a retirada dos dois módulos acessórios reduz em 15 °C. Duas publicações retratam a redução da temperatura ótima da atividade de celulasas das famílias GH9¹⁵⁴ e GH10¹⁵⁵ pela retirada de CBM. Porém, nenhum artigo relacionando termoestabilidade (independente da atividade) com os CBMs foi encontrado.

5.2.2 Estrutura da H6.

A estrutura geral da enzima H6 revelou a presença de três domínios, sendo um domínio catalítico da família GH5 na porção aminoterminal e outros dois módulos acessórios conectados subsequentemente. A presença de três domínios não era esperada, pois somente dois haviam sido previstos pelo BLAST através de sua sequência de aminoácidos.

O domínio catalítico possui o mesmo enovelamento característico da família GH5, porém, os domínios acessórios não possuem semelhança significativa com nenhuma estrutura depositada no PDB, de acordo com a busca estrutural feita com o programa DALI¹⁵⁶.

A presença de pequenos módulos ligados aos domínios catalíticos de celulasas já foi descrita em alguns artigos. Em boa parte, esses módulos possuem um enovelamento tipo sanduiche- β , como descrito para a maioria dos CBMs¹⁵⁷, porém, com função não comprovada¹⁵⁸. Na estrutura de uma beta-galactosidase ainda não publicada (PDB 1FN9) são reconhecíveis três pequenos módulos acessórios conectados ao domínio catalítico, possuindo o mesmo enovelamento básico. Em uma endogliceramidase da família GH5, o domínio acessório possui o enovelamento sanduiche- β e, como descrito para os módulos acessórios da H6, sua sequência não foi identificada pelo BLAST¹⁵⁹. Em todos os casos, os domínios acessórios estão localizados longe do sítio ativo dos domínios catalíticos e, portanto, não parecem influenciar a interação enzima-substrato.

A primeira grande novidade que a estrutura da H6 traz é exatamente o posicionamento dos módulos acessórios. O módulo 1 localiza-se lateralmente ao domínio catalítico, da mesma forma como em outras estruturas¹⁵⁷⁻¹⁵⁸⁻¹⁵⁹. Já o módulo 2 está localizado muito próximo ao sítio catalítico e o alinhamento dos triptofanos dos dois domínios indicam fortemente a

formação de maior número de subsítios, acarretando em uma maior afinidade a carboidratos longos. Infelizmente, na estrutura resolvida com ligante, o carboidrato está situado nos subsítios negativos (anteriores à posição de clivagem) e os possíveis novos subsítios se localizam nas posições positivas (posteriores à posição de clivagem). Vários experimentos já estão sendo planejados pelo nosso grupo para comprovar se essa função é real. Uma vez confirmada, será a primeira vez que se observa a complementação do sítio ativo por um módulo acessório.

A enzima H6a foi cristalizada em quatro grupos espaciais diferentes, gerando cristais com formatos diferentes. Em um dos tipos cristalinos ($P4_32_12$), o soaking com a celopentaose foi efetivo e a estrutura da enzima ligada ao substrato foi resolvida. Todas as cinco estruturas (quatro na forma apo e uma com substrato), apesar dos grupos espaciais diferentes, não apresentaram nenhuma variação de estrutura proteica considerável. Esse fato evidencia, mais uma vez, a grande estabilidade dessa proteína e se torna bom indício de que o enovelamento não tem grande influencia do empacotamento cristalino.

Inesperadamente, a presença do substrato no sítio catalítico também não alterou a estrutura da enzima, nem mesmo as cadeias laterais dos resíduos que interagem com o ligante. Apesar de curioso esse evento já havia sido observado com uma endoglucanase de *Thermotoga maritima* pertencente à família GH5¹⁶⁰. No trabalho publicado, a endoglucanase com um dos resíduos catalíticos mutados teve sua estrutura cristalográfica resolvida na forma apo e ligada a vários carboidratos, inclusive a celotetraose. Da mesma forma, as enzimas apo e holo não apresentaram diferença significativa.

A questão mais importante suscitada pela estrutura da H6a com substrato trata da posição em que o substrato se encontra e como ele foi modificado. Assim como foi descrito nos resultados, a celopentaose provavelmente foi clivada tornando-se uma celotetraose que permaneceu no sítio ativo e uma glicose que se desligou da enzima.

De acordo com o mecanismo de clivagem proposto (figura 4), na primeira etapa o resíduo nucleófilo ataca o carbono anomérico do glicano que está localizado no subsítio -1, formando um intermediário enzima-substrato, enquanto o outro resíduo catalítico protona o oxigênio glicosídico. Consequentemente, o carboidrato que está localizado nos subsítios positivos é liberado. Na segunda etapa, o intermediário é clivado através de uma hidrólise e o carboidrato restante, que estava localizado nos subsítios negativos, é liberado¹⁶⁰.

Por essa descrição, seria esperado que a celotetraose encontrada fosse um dos produtos da reação e estivesse completa, ou que ela fizesse parte do estado intermediário e que, além de

estar ligada covalentemente com a enzima, não possuísse o oxigênio glicosídico. Porém, a posição em que o glicano (subsítio -1) se encontra não deixa clara a situação.

O carbono anomérico está a uma distância de 2,9 Å do oxigênio do resíduo catalítico E268. Como essa distância é incompatível com uma ligação covalente (que possui ~ 1,5 Å) é pouco provável que o modelo gerado seja a representação do estado intermediário. Porém, a densidade eletrônica gerada pelos dados experimentais mostra claramente uma continuidade entre o carbono anomérico do carboidrato e o oxigênio do resíduo E268, indicando um fluxo de elétrons nessa posição. Se a celotetraose for o produto da reação, deveria existir um oxigênio ligado ao carbono anomérico, conhecido com o oxigênio glicosídico. Esse estaria a 1,5 Å do oxigênio do resíduo catalítico E268, havendo um impedimento estérico (figura 46). Apesar da qualidade e precisão dos dados, os resultados descritos tornam complexa a interpretação da situação.

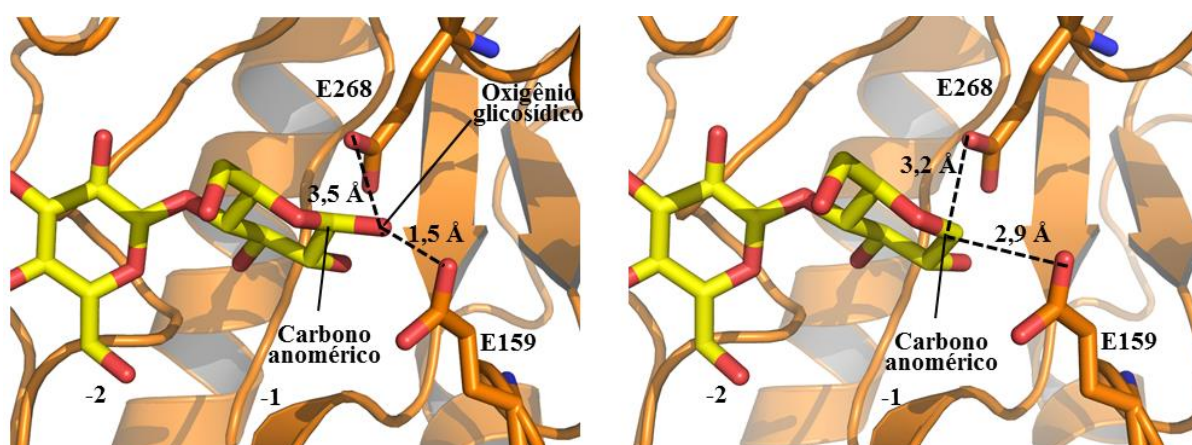


Figura 46 - Representação da interação do carboidrato (amarelo) com os resíduos catalíticos E159 e E268 da enzima H6 (laranja). À esquerda o carboidrato está representado com o oxigênio glicosídico e à direita este oxigênio está ausente.

O posicionamento do carboidrato no sítio ativo traz mais uma curiosidade. Como descrito na literatura¹⁶¹, o primeiro resíduo catalítico (localizado na alça entre a fita $\beta 4$ e hélice $\alpha 4$) atua como ácido/base e o segundo (localizado na alça entre a fita $\beta 7$ e hélice $\alpha 7$) é o nucleófilo. Entretanto, na estrutura da enzima H6, o resíduo E159 está mais próximo do carboidrato do que o resíduo E268. Além disso, temos uma densidade eletrônica contínua entre o resíduo E159 e o carboidrato. Esses dois fatos indicam que, no caso da enzima H6, o nucleófilo pode ser o resíduo E159, enquanto o E268 agiria como ácido/base.

Em uma revisão, Vuong¹⁶² sugere que, com a identificação de uma diversidade de mecanismos catalíticos encontrados dentro do grupo das hidrolases de glicosídeos, a visão clássica de Koshland não está sempre correta. Nessa revisão, também são sugeridas algumas técnicas para se determinar qual é o resíduo nucleófilo, como o *Docking*. Desta forma, experimentos já estão sendo feitos para confirmação da função de cada resíduo catalítico.

5.2.3 Beta-glicosidases D5, F5 e G5

De todas as enzimas que foram expressas, quatro foram cristalizadas, sendo uma endoglucanase e três beta-glicosidases. Por acaso, as quatro enzimas cristalizadas pertencem ao mesmo organismo, a bactéria *Bacillus licheniformis*. Trata-se de uma bactéria pertencente ao grupo *Subtilis*¹⁶³, juntamente com *Bacillus subtilis* e *Bacillus pumilus* e é geralmente encontrada no solo e em penas de aves, estando relacionada à contaminação de produtos alimentares¹⁶⁴.

Por pertencerem ao mesmo organismo, não faz sentido imaginar que as enzimas D5, F5 e G5 tenham a mesma especificidade. Apesar de serem extremamente parecidas, dois pontos de divergência importantes foram detectados através de suas estruturas.

O primeiro ponto é a presença de diferentes aminoácidos em duas das posições que compõem o sítio catalítico. Essa diferença é significativa, pois os aminoácidos possuem polaridade e tamanho diferentes. Na posição 9, a enzima D5 possui uma alanina, que é caracterizada por uma cadeia lateral pequena e apolar, a enzima F5 possui uma treonina, cuja cadeia lateral é pouco maior do que a alanina e é polar e a enzima G5, como as outras beta-glicosidases usadas para comparação, possui um triptofano que tem uma cadeia lateral grande e é apolar, apesar de poder formar ligação de hidrogênio através de um nitrogênio. Na posição 8 as três enzimas possuem o mesmo aminoácido que, apesar de ser polar, possui cadeia lateral curta, em comparação com as demais enzimas que possuem um aminoácido de cadeia lateral polar e longa. A variação nessas posições pode fazer com que os diferentes carboidratos tenham diferentes afinidades pelo sítio ativo. Desta forma, as três enzimas não seriam redundantes para a bactéria e teriam funções distintas. No momento, experimentos para a caracterização da atividade das beta-glicosidases estão sendo preparados. Com isso, esperamos comprovar a diferença na função das mesmas.

O segundo ponto divergente está relacionado com as alças 1, 4, 6 e 8 (figuras 44 e 47), que formam a entrada do sítio catalítico¹⁶⁵. A enzima D5 possui uma fenda bem aberta, permitindo assim a passagem de carboidratos grandes. Já a enzima G5 é bem fechada, possuindo apenas uma pequena abertura entre a alça 6 e a hélice 4. As alças da enzima F5 não puderam ser modeladas por não possuírem densidade eletrônica definida, portanto, não há como inferir o tamanho da abertura do sítio ativo. Há apenas uma pequena evidência de que ela tenha um tamanho intermediário. A porção da alça 6, que pôde ser modelada, é muito parecida com a alça 6 da enzima G5 (figura 44). Porém a alça da F5 possui 4 resíduos a menos do que a alça da G5. As outras alças, 1, 4 e 8, também são divergentes, porém, menos significativas do que a alça 6, por serem menores e possuírem o mesmo número de aminoácidos. Como a alça 6 é a principal determinante da diferença no tamanho da fenda entre D5 e G5, pode ser que a enzima F5 tenha uma fenda de tamanho intermediário.

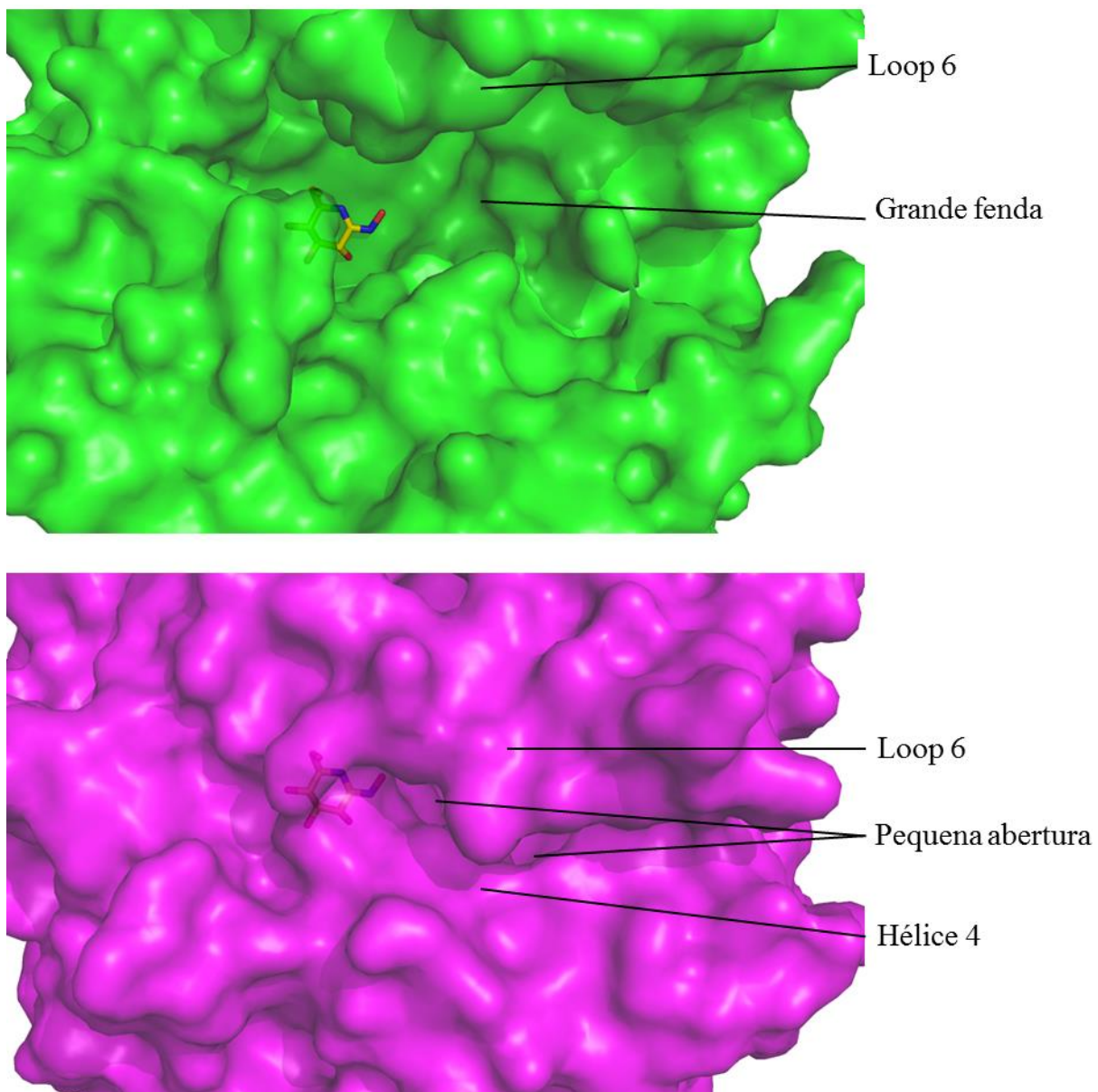


Figura 47 – Representação em superfície dos modelos de D5 (verde) e G5 (magenta). No centro de cada enzima está modelado o inibidor de beta-glicosidase gluco-hidroximolactam através da sobreposição com a estrutura 2J78¹⁴¹.

6 CONCLUSÕES

O período de quatro anos destinados a cumprir os objetivos propostos, que a princípio parecia mais que suficiente, mostrou-se extremamente curto. Talvez seja o caso da maioria dos alunos de doutorado.

Os dois primeiros anos dedicados à expressão da endoglucanase II de *T. reesei* passaram rapidamente e, infelizmente, o objetivo não foi alcançado. Várias tentativas foram feitas, variando-se os organismos de expressão, condições de expressão e métodos de transformação, mas a enzima não foi obtida na forma solúvel. Durante esse período, tive a oportunidade de trabalhar com o professor Fernando Araripe Torres e sua equipe (UnB) e, apesar de não chegar ao resultado esperado, aprendi muito sobre o mecanismo de expressão em *P. pastoris*.

Os dois anos finais foram voltados para o estudo das endoglucanases e beta-glicosidases bacterianas que estavam disponíveis. Na primeira parte do trabalho, feita no Brasil, várias das enzimas foram expressas e purificadas e a endoglucanase H6 teve sua atividade caracterizada, foi cristalizada e sua estrutura tridimensional foi resolvida. E nos últimos meses sete meses, trabalhando no ESRF da França, essa mesma enzima foi novamente cristalizada e sua estrutura foi resolvida na forma apo a 1,1 Å de resolução e complexada a um substrato a 1,75 Å. Essas estruturas permitiram a descoberta de uma nova forma de interação entre CBMs e domínios catalíticos e trouxeram mais informações sobre o funcionamento da atividade catalítica. Ainda no ESRF, foram cristalizadas três beta-glicosidases (D5, F5 e G5) pertencentes ao mesmo organismo da endoglucanase H6 (*Bacillus licheniformis*) e suas estruturas foram resolvidas. Apesar de não ter havido tempo para caracterizar as atividades das beta-glicosidases, a análise dos modelos cristalográficos indicam que algumas características estruturais divergentes podem fazer com que elas apresentem especificidades diferentes para os substratos. Essa descoberta, unida aos ensaios enzimáticos que serão feitos, pode indicar alvos para mutações visando o melhoramento dessas enzimas ou de outras pertencentes à mesma família.

Durante o doutorado, ainda tive a oportunidade de colaborar com os professores Dr. Flávio Henrique Silva (UFSCar), Dr. Munir Skaf (UNICAMP) e suas respectivas equipes, em que cristalicei e resolvi a estrutura da endoglucanase III de *Trichoderma harzianum*. Esse trabalho em colaboração gerou a publicação de dois artigos publicados, em ACTA Crystallographica F (anexo A) e em PLOS ONE (anexo B).

Por fim, os resultados obtidos foram satisfatórios e, como prova de sua importância, motivaram não somente novas perguntas e novas pesquisas a serem feitas, mas também a vontade do autor dessa tese em continuar a trabalhar nessa área.

REFERÊNCIAS

- 1 AGARWAL, A. Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines. **Progress in Energy and Combustion Science**, v. 33, n. 3, p. 233-271, 2007.
- 2 DOGARIS, I.; MAMMA, D.; KEKOS, D. Biotechnological production of ethanol from renewable resources by *Neurospora crassa*: an alternative to conventional yeast fermentations? **Applied Microbiology and Biotechnology**, v. 97, n. 4, p. 1457-1473, 2013.
- 3 BALAT, M.; BALAT, H.; OZ, C. Progress in bioethanol processing. **Progress in Energy and Combustion Science**, v. 34, n. 5, p. 551-573, 2008.
- 4 KNOCKE, C.; VOGT, J. Biofuels - challenges & chances: how biofuel development can benefit from advanced process technology. **Engineering in Life Sciences**, v. 9, n. 2, p. 96-99, 2009.
- 5 HIMMEL, M. E. et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. **Science**, v. 315, n. 5813, p. 804-807, 2007.
- 6 BEUKES, N.; PLETSCHKE, B. I. Effect of lime pre-treatment on the synergistic hydrolysis of sugarcane bagasse by hemicellulases. **Bioresour Technol**, v. 101, n. 12, p. 4472-4478, 2010.
- 7 BEUKES, N.; PLETSCHKE, B. I. Effect of alkaline pre-treatment on enzyme synergy for efficient hemicellulose hydrolysis in sugarcane bagasse. **Bioresource Technology**, v. 102, n. 8, p. 5207-5213, 2011.
- 8 LI, X. H. et al. Enhanced cellulase production of the *Trichoderma viride* mutated by microwave and ultraviolet. **Microbiological Research**, v. 165, n. 3, p. 190-198, 2010.
- 9 YU, L. X. et al. Expression of thermostable microbial cellulases in the chloroplasts of nicotine-free tobacco. **Journal of Biotechnology**, v. 131, n. 3, p. 362-369, 2007.
- 10 TAYLOR, L. E. et al. Heterologous expression of glycosyl hydrolases in planta: a new departure for biofuels. **Trends in Biotechnology**, v. 26, n. 8, p. 413-424, 2008.

11 FENEL, F. et al. A de novo designed N-terminal disulphide bridge stabilizes the *Trichoderma reesei* endo-1,4-beta-xylanase II. **Journal of Biotechnology**, v. 108, n. 2, p. 137-143, 2004.

12 PHITSUWAN, P. et al. Present and potential applications of cellulases in agriculture, biotechnology, and bioenergy. **Folia Microbiologica**, v. 58, n. 2, p. 163-176, 2013.

13 TAHERZADEH, M.; KARIMI, K. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review. **International Journal of Molecular Sciences**, v. 9, n. 9, p. 1621-1651, 2008.

14 OLOFSSON, K.; BERTILSSON, M.; LIDEN, G. A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks. **Biotechnology for Biofuels**, v. 1, 2008. doi: 10.1186/1754-6834-1-7

15 MARGEOT, A. et al. New improvements for lignocellulosic ethanol. **Current Opinion in Biotechnology**, v. 20, n. 3, p. 372-380, 2009.

16 HAHN-HAGERDAL, B. et al. Bio-ethanol - the fuel of tomorrow from the residues of today. **Trends in Biotechnology**, v. 24, n. 12, p. 549-556, 2006.

17 DEMAIN, A. Biosolutions to the energy problem. **Journal of Industrial Microbiology & Biotechnology**, v. 36, n. 3, p. 319-332, 2009.

18 SANCHEZ, O.; CARDONA, C. Trends in biotechnological production of fuel ethanol from different feedstocks. **Bioresource Technology**, v. 99, n. 13, p. 5270-5295, 2008.

19 CARPITA, N.; GIBEAUT, D. Structural models of primary-cell walls in flowering plants: consistency of molecular-structure with the physical-properties of the walls during growth. **Plant Journal**, v. 3, n. 1, p. 1-30, 1993.

20 POPPER, Z.; FRY, S. Xyloglucan-pectin linkages are formed intra-protoplasmically, contribute to wall-assembly, and remain stable in the cell wall. **Planta**, v. 227, n. 4, p. 781-794, 2008.

21 ROPPOLO, D. et al. A novel protein family mediates Casparian strip formation in the endodermis. **Nature**, v. 473, n. 7347, p. 380-383, 2011.

22 LEROUX, O. Collenchyma: a versatile mechanical tissue with dynamic cell walls. **Annals of Botany**, v. 110, n. 6, p. 1083-1098, 2012.

23 FRY, S. Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. **New Phytologist**, v. 161, n. 3, p. 641-675, 2004.

24 MCCANN, M.; CARPITA, N. Designing the deconstruction of plant cell walls. **Current Opinion in Plant Biology**, v. 11, n. 3, p. 314-320, 2008.

25 SILVA, G. et al. Cell wall polysaccharides from fern leaves: evidence for a mannan-rich type III cell wall in *Adiantum raddianum*. **Phytochemistry**, v. 72, n. 18, p. 2352-2360, 2011.

26 HARRIS, P.; SMITH, B. Plant cell walls and cell-wall polysaccharides: structures, properties and uses in food products. **International Journal of Food Science and Technology**, v. 41, p. 129-143, 2006.

27 PAULY, M.; KEEGSTRA, K. Plant cell wall polymers as precursors for biofuels. **Current Opinion in Plant Biology**, v. 13, n. 3, p. 305-312, 2010.

28 WANG, H.; DIXON, R. On-off switches for secondary cell wall biosynthesis. **Molecular Plant**, v. 5, n. 2, p. 297-303, 2012.

29 PLOMION, C.; LÉPROVOST, G.; STOKES, A. Wood formation in trees. **Plant Physiology**, v. 127, n. 4, p. 1513-1523, 2001.

30 O'NEILL, M.; YORK, W. The composition and structure of plant primary cell walls. In: ROSE, J. (Ed.). **The plant cell wall**. Ithaca, New York: Blackwell Publishing/CRC Press, 2003. p.1-54.

31 VOGEL, J. Unique aspects of the grass cell wall. **Current Opinion in Plant Biology**, v. 11, n. 3, p. 301-307, 2008.

32 SOMERVILLE, C. Cellulose synthesis in higher plants. **Annual Review of Cell and Developmental Biology**, v. 22, p. 53-78, 2006.

33 MCCANN, M.; WELLS, B.; ROBERTS, K. Direct visualization of cross-links in the primary plant-cell wall. **Journal of Cell Science**, v. 96, p. 323-334, 1990.

34 PAULY, M.; KEEGSTRA, K. Cell-wall carbohydrates and their modification as a resource for biofuels. **Plant Journal**, v. 54, n. 4, p. 559-568, 2008.

35 MOORE, P.; STAEHELIN, L. Immunogold localization of the cell-wall-matrix polysaccharides rhamnogalacturonan-i and xyloglucan during cell expansion and cytokinesis in trifolium-pratense l - implication for secretory pathways. **Planta**, v. 174, n. 4, p. 433-445, 1988.

36 ZHANG, G.; STAEHELIN, L. Functional compartmentation of the golgi-apparatus of plant-cells - immunocytochemical analysis of high-pressure frozen-substituted and freeze-substituted sycamore maple suspension-culture cells. **Plant Physiology**, v. 99, n. 3, p. 1070-1083, 1992.

37 FRY, S. The structure and functions of xyloglucan. **Journal of Experimental Botany**, v. 40, n. 210, p. 1-11, 1989.

38 DARLEY, C.; FORRESTER, A.; MCQUEEN-MASON, S. The molecular basis of plant cell wall extension. **Plant Molecular Biology**, v. 47, n. 1-2, p. 179-195, 2001.

39 O'NEILL, M.; ALBERSHEIM, P.; DARVILL, A. The pectic polysaccharides of primary cell walls. In: DEY, P. (Ed.). **Methods in plant biochemistry**. London: Academic Press, 1990. p.415-441.

40 LI, X.; CHAPPLE, C. Understanding lignification: challenges beyond monolignol biosynthesis. **Plant Physiology**, v. 154, n. 2, p. 449-452, 2010.

41 GRIMA-PETTENATI, J.; GOFFNER, D. Lignin genetic engineering revisited. **Plant Science**, v. 145, n. 2, p. 51-65, 1999.

42 ASP, N. Dietary fiber - definition, chemistry and analytical determination. **Molecular Aspects of Medicine**, v. 9, n. 1, p. 17-29, 1987.

43 SHOWALTER, A. Introduction: plant cell wall proteins. **Cellular and Molecular Life Sciences**, v. 58, n. 10, p. 1361-1362, 2001.

44 SHOWALTER, A. Structure and function of plant-cell wall proteins. **Plant Cell**, v. 5, n. 1, p. 9-23, 1993.

45 RINGLI, C.; KELLER, B.; RYSER, U. Glycine-rich proteins as structural components of plant cell walls. **Cellular and Molecular Life Sciences**, v. 58, n. 10, p. 1430-1441, 2001.

46 VAAJE-KOLSTAD, G. et al. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. **Science**, v. 330, n. 6001, p. 219-222, 2010.

47 BUCHHOLZ, K.; SEIBEL, J. Industrial carbohydrate biotransformations. **Carbohydrate Research**, v. 343, n. 12, p. 1966-1979, 2008.

48 SINGH, B. Exploring microbial diversity for biotechnology: the way forward. **Trends in Biotechnology**, v. 28, n. 3, p. 111-116, 2010.

49 WILSON, D. Cellulases and biofuels. **Current Opinion in Biotechnology**, v. 20, n. 3, p. 295-299, 2009.

50 BHAT, M. Cellulases and related enzymes in biotechnology. **Biotechnology Advances**, v. 18, n. 5, p. 355-383, 2000.

51 BHAT, M.; BHAT, S. Cellulose degrading enzymes and their potential industrial applications. **Biotechnology Advances**, v. 15, n. 3-4, p. 583-620, 1997.

52 JORGENSEN, H.; KRISTENSEN, J.; FELBY, C. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. **Biofuels Bioproducts & Biorefining-Biofpr**, v. 1, n. 2, p. 119-134, 2007.

53 WILSON, D. Microbial diversity of cellulose hydrolysis. **Current Opinion in Microbiology**, v. 14, n. 3, p. 259-263, 2011.

54 SHOSEYOV, O.; SHANI, Z.; LEVY, I. Carbohydrate binding modules: biochemical properties and novel applications. **Microbiology and Molecular Biology Reviews**, v. 70, n. 2, p. 283-295, 2006.

55 VARNAI, A.; SIIKA-AHO, M.; VIKARI, L. Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. **Biotechnology For Biofuels**, v. 6, 2013. doi: 10.1186/1754-6834-6-30.

56 WATANABE, H.; TOKUDA, G. Cellulolytic systems in insects. **Annual Review of Entomology**, v. 55, p. 609-632, 2010.

57 SANCHEZ, C. Lignocellulosic residues: biodegradation and bioconversion by fungi. **Biotechnology Advances**, v. 27, n. 2, p. 185-194, 2009.

58 BAYER, E.; LAMED, R.; HIMMEL, M. The potential of cellulases and cellulosomes for cellulosic waste management. **Current Opinion in Biotechnology**, v. 18, n. 3, p. 237-245, 2007.

59 DING, S. et al. A biophysical perspective on the cellulosome: new opportunities for biomass conversion. **Current Opinion in Biotechnology**, v. 19, n. 3, p. 218-227, 2008.

60 TACHAAPAIKON, C. et al. Isolation and characterization of a new cellulosome-producing clostridium thermocellum strain. **Biodegradation**, v. 23, n. 1, p. 57-68, 2012.

61 KUMAR, R.; SINGH, S.; SINGH, O. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. **Journal of Industrial Microbiology & Biotechnology**, v. 35, n. 5, p. 377-391, 2008.

62 GUSAKOV, A. Alternatives to *Trichoderma reesei* in biofuel production. **Trends in Biotechnology**, v. 29, n. 9, p. 419-425, 2011.

63 SUKUMARAN, R.; SINGHANIA, R.; PANDEY, A. Microbial cellulases - production, applications and challenges. **Journal of Scientific & Industrial Research**, v. 64, n. 11, p. 832-844, 2005.

64 CANTAREL, B. et al. The carbohydrate-active enZymes database (CAZy): an expert resource for glycogenomics. **Nucleic Acids Research**, v. 37, p. 233-238, 2009.

65 HENRISSAT, B. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. **Biochemical Journal**, v. 280, p. 309-316, 1991.

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

67 DUAN, C. J. et al. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. **Journal of Applied Microbiology**, v. 107, n. 1, p. 245-256, 2009.

68 ELIFANTZ, H. et al. Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean. **Fems Microbiology Ecology**, v. 63, n. 3, p. 316-327, 2008.

69 HESS, M. et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. **Science**, v. 331, n. 6016, p. 463-467, 2011.

70 CANTAREL, B. L. et al. The carbohydrate-active enzymes database (CAZy): an expert resource for glycomics. **Nucleic Acids Research**, v. 37, p. 233-238, 2009.

71 ST JOHN, F. J.; GONZALEZ, J. M.; POZHARSKI, E. Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. **Febs Letters**, v. 584, n. 21, p. 4435-4441, 2010.

72 ASPEBORG, H. et al. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). **BMC Evolutionary Biology**, v. 12, 2012. doi: 10.1186/1471-2148-12-186.

73 KOSHLAND, D. E. Stereochemistry and the mechanism of enzymatic reactions. **Biological Reviews of the Cambridge Philosophical Society**, v. 28, n. 4, p. 416-436, 1953.

74 VUONG, T. V.; WILSON, D. B. Glycoside hydrolases: catalytic base/nucleophile diversity. **Biotechnology and Bioengineering**, v. 107, n. 2, p. 195-205, 2010.

75 YUAN, J. S. et al. The endo-beta-mannanase gene families in Arabidopsis, rice, and poplar. **Functional & Integrative Genomics**, v. 7, n. 1, p. 1-16, 2007.

76 DOMINGUEZ, R. et al. A common protein fold and similar active-site in 2 distinct families of beta-glycanases. **Nature Structural Biology**, v. 2, n. 7, p. 569-576, 1995.

77 CAIRNS, J. R. K.; ESEN, A. beta-Glucosidases. **Cellular and Molecular Life Sciences**, v. 67, n. 20, p. 3389-3405 2010.

78 SANZ-APARICIO, J. et al. Crystal structure of beta-glucosidase a from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosyl hydrolases. **Journal of Molecular Biology**, v. 275, n. 3, p. 491-502, 1998.

79 KUNTOTHOM, T. et al. Rice family GH1 glycoside hydrolases with beta-D-glucosidase and beta-D-mannosidase activities. **Archives of Biochemistry and Biophysics**, v. 491, n. 1-2, p. 85-95, 2009.

80 CZJZEK, M. et al. The mechanism of substrate (aglycone) specificity in beta-glucosidases is revealed by crystal structures of mutant maize beta-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 25, p. 13555-13560, 2000.

81 CHUENCHOR, W. et al. Structural insights into rice BGlu1 beta-glucosidase oligosaccharide hydrolysis and transglycosylation. **Journal of Molecular Biology**, v. 377, n. 4, p. 1200-1215, 2008.

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

83 GUILLEN, D.; SANCHEZ, S.; RODRIGUEZ-SANOJA, R. Carbohydrate-binding domains: multiplicity of biological roles. **Applied Microbiology and Biotechnology**, v. 85, n. 5, p. 1241-1249, 2010.

84 VAAJE-KOLSTAD, G. et al. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. **Journal of Biological Chemistry**, v. 280, n. 31, p. 28492-28497, 2005.

85 MOSER, F. et al. Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. **Biotechnology and Bioengineering**, v. 100, n. 6, p. 1066-1077, 2008.

86 ABE, A. et al. Complex structures of *Thermoactinomyces vulgaris* R-47 alpha-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. **Journal of Molecular Biology**, v. 335, n. 3, p. 811-822, 2004.

87 KAWAZU, T. et al. Cloning and nucleotide-sequence of the gene coding for enzymatically active fragments of the bacillus-polymyxa beta-amylase. **Journal of Bacteriology**, v. 169, n. 4, p. 1564-1570, 1987.

88 MORLON-GUYOT, J. et al. Characterization of the *L-manihotivorans* alpha-amylase gene. **DNA Sequence**, v. 12, n. 1, p. 27-37, 2001.

89 BORASTON, A. B.; FICKO-BLEAN, E.; HEALEYT, M. Carbohydrate recognition by a large sialidase toxin from *Clostridium perfringens*. **Biochemistry**, v. 46, n. 40, p. 11352-11360, 2007.

90 ITO, Y. et al. Cloning, expression, and cell surface localization of *Paenibacillus* sp strain W-61 xylanase 5, a multidomain xylanase. **Applied and Environmental Microbiology**, v. 69, n. 12, p. 6969-6978 2003.

91 HASHIMOTO, H. Recent structural studies of carbohydrate-binding modules. **Cellular and Molecular Life Sciences**, v. 63, n. 24, p. 2954-2967, 2006.

92 KRAULIS, P. J. et al. Determination of the 3-dimensional solution structure of the c-terminal domain of cellobiohydrolase-i from *trichoderma-reesei* - a study using nuclear magnetic-resonance and hybrid distance geometry dynamical simulated annealing. **Biochemistry**, v. 28, n. 18, p. 7241-7257, 1989.

93 BORASTON, A. B. et al. Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules. **Journal of Molecular Biology**, v. 319, n. 5, p. 1143-1156, 2002.

94 NOTENBOOM, V. et al. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. **Biochemistry**, v. 40, n. 21, p. 6248-6256, 2001.

95 GOTO, M. et al. Analysis of the raw starch-binding domain by mutation of a glucoamylase from *aspergillus-awamori* var. *Kawachi* expressed in *saccharomyces-cerevisiae*. **Applied and Environmental Microbiology**, v. 60, n. 11, p. 3926-3930, 1994.

96 ITOH, Y. et al. Importance of Trp59 and Trp60 in chitin-binding, hydrolytic, and antifungal activities of *Streptomyces griseus* chitinase C. **Applied Microbiology and Biotechnology**, v. 72, n. 6, p. 1176-1184, 2006.

97 NEWSTEAD, S. L. et al. Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. **Acta Crystallographica Section D-biological crystallography**, v. 61, p. 1483-1491, 2005.

98 RODRIGUEZ-SANOJA, R. et al. A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus* alpha-amylase. **Journal of Industrial Microbiology & Biotechnology**, v. 36, n. 3, p. 341-346, 2009.

99 VIEGAS, A. et al. Molecular determinants of ligand specificity in family 11 carbohydrate binding modules - an NMR, X-ray crystallography and computational chemistry approach. **Febs Journal**, v. 275, n. 10, p. 2524-2535, 2008.

100 KLEYWEGT, G. et al. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 angstrom resolution, and a comparison with related enzymes. **Journal of Molecular Biology**, v. 272, n. 3, p. 383-397, 1997.

101 NAKAZAWA, H. et al. Characterization of the catalytic domains of *Trichoderma reesei* endoglucanase I, II, and III, expressed in *Escherichia coli*. **Applied Microbiology and Biotechnology**, v. 81, n. 4, p. 681-689, 2008.

102 QIN, Y. et al. Purification and characterization of recombinant endoglucanase of *Trichoderma reesei* expressed in *Saccharomyces cerevisiae* with higher glycosylation and stability. **Protein Expression and Purification**, v. 58, n. 1, p. 162-167, 2008.

103 DOYLE, S. A. High-throughput cloning for proteomics research. **Methods in Molecular Biology**, v. 310, p. 107-113, 2005.

104 ASLANIDIS, C.; DEJONG, P. Ligation-independent cloning of pcr products (lic-pcr). **Nucleic Acids Research**, v. 18, n. 20, p. 6069-6074, 1990.

105 ERICSSON, U. et al. Thermofluor-based high-throughput stability optimization of proteins for structural studies. **Analytical Biochemistry**, v. 357, n. 2, p. 289-298, 2006.

106 DUPEUX, F. et al. A thermal stability assay can help to estimate the crystallization likelihood of biological samples. **Acta Crystallographica Section D-biological crystallography**, v. 67, p. 915-919, 2011.

107 MANDEL, M.; WEBER, J. Exoglucanase activity by microorganisms. **Advances in Chemistry**, v. 95, p. 391-414, 1969.

108 MILLER, G. Use of dinitrosalicylic acid reagent for determination of reducing sugar. **Analytical Chemistry**, v. 31, n. 3, p. 426-428, 1959.

109 KABSCH, W. XDS. **Acta Crystallographica Section D-biological crystallography**, v. 66, p. 125-132, 2010. doi:10.1107/S0907444909047337.

110 EVANS, P. Scaling and assessment of data quality. **Acta Crystallographica Section D-biological crystallography**, v. 62, p. 72-82, 2006. doi:10.1107/S0907444905036693.

111 EVANS, P. An introduction to data reduction: space-group determination, scaling and intensity statistics. **Acta Crystallographica Section D**-biological crystallography, v. 67, p. 282-292, 2011. doi:10.1107/S090744491003982X.

112 MATTHEWS, B. Solvent content of protein crystals. **Journal of Molecular Biology**, v. 33, n. 2, p. 491-497, 1968.

113 KANTARDJIEFF, K.; RUPP, B. Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. **Protein Science**, v. 12, n. 9, p. 1865-1871, 2003.

114 MCCOY, A. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, p. 658-674, 2007. doi:10.1107/S0021889807021206.

115 ADAMS, P. et al. PHENIX: building new software for automated crystallographic structure determination. **Acta Crystallographica Section D**-biological crystallography, v. 58, p. 1948-1954, 2002. doi:10.1107/S0907444902016657.

116 EMSLEY, P.; COWTAN, K. Coot: model-building tools for molecular graphics. **Acta Crystallographica Section D**-biological crystallography, v. 60, p. 2126-2132, 2004. doi:10.1107/S0907444904019158.

117 GASTEIGER, E. et al. Protein identification and analysis tools on the ExPASy server. In: WALKER, J. M. (Ed.). **The Proteomics Protocols Handbook**. Totowa, NJ: Humana Press, 2005. p.571-607.

118 PREDICTION of N-glycosylation sites in human proteins. 2004. Disponível em <<http://www.cbs.dtu.dk/services/NetNGlyc/>>. Acesso em 20 ago. 2013.

119 STEENTOFT, C. et al. Precision mapping of the human O-GalNAc glycoproteome through simplecell technology. **Embo Journal**, v. 32, n. 10, p. 1478-1488, 2013.

120 GELLISSEN, G. et al. New yeast expression platforms based on methylotrophic hansenula polymorpha and pichia pastoris and on dimorphic arxula adenivorans and yarrowia lipolytica - a comparison. **Fems Yeast Research**, v. 5, n. 11, p. 1079-1096, 2005.

121 CREGG, J. et al. Functional-characterization of the 2 alcohol oxidase genes from the yeast pichia-pastoris. **Molecular and Cellular Biology**, v. 9, n. 3, p. 1316-1323, 1989.

122 CREGG, J. et al. Recombinant protein expression in pichia pastoris. **Molecular Biotechnology**, v. 16, n. 1, p. 23-52, 2000.

123 BRETTHAUER, R.; CASTELLINO, F. Glycosylation of pichia pastoris-derived proteins. **Biotechnology and Applied Biochemistry**, v. 30, n. 3, p. 193-200, 1999.

124 WU, J.; WANG, S.; FU, W. Lower temperature cultures enlarge the effects of vitreoscilla hemoglobin expression on recombinant pichia pastoris. **International Journal of Molecular Sciences**, v. 13, n. 10, p. 13212-13226, 2012.

125 LING, L.; ITHOI, I.; FONG, M. Optimization for high-level expression in pichia pastoris and purification of truncated and full length recombinant sag2 of toxoplasma gondii for diagnostic use. **Southeast Asian Journal of Tropical Medicine and Public Health**, v. 41, n. 3, p. 507-513, 2010.

126 LEE, T. M. et al. A structural study of hypocrea jecorina Cel5A. **Protein Science**, v. 20, n. 11, p. 1935-1940, 2011.

127 YOON, J. et al. Identification and purification of a soluble region of BubR1: a critical component of the mitotic checkpoint complex. **Protein Expression and Purification**, v. 44, n. 1, p. 1-9, 2005.

128 YE, H.; CIRILLI, M.; WU, H. The use of construct variation and diffraction data analysis in the crystallization of the TRAF domain of human tumor necrosis factor receptor associated factor 6. **Acta Crystallographica Section D-biological crystallography**, v. 58, p. 1886-1888, 2002. doi:10.1107/S0907444902013318.

129 QIN, Y. et al. Engineering endoglucanase II from Trichoderma reesei to improve the catalytic efficiency at a higher pH optimum. **Journal of Biotechnology**, v. 135, n. 2, p. 190-195, 2008.

130 LASKOWSKI, R. et al. Procheck - a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, p. 283-291, 1993. doi:10.1107/S0021889892009944.

131 CHEN, V. et al. MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D-biological crystallography**, v. 66, p. 12-21, 2010. doi: 10.1107/S0907444909042073.

132 GASKELL, A.; CRENNELL, S.; TAYLOR, G. The 3 domains of a bacterial sialidase - a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. **Structure**, v. 3, n. 11, p. 1197-1205, 1995.

133 WIMMEROVA, M. et al. Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods. **Plos One**, v. 7, n. 10, 2012. doi: 10.1371/journal.pone.0046032.

134 SPIWOK, V. et al. Role of CH/pi interactions in substrate binding by Escherichia coli beta-galactosidase. **Carbohydrate Research**, v. 339, n. 13, p. 2275-2280, 2004.

135 BRANDL, M. et al. C-H center dot center dot center dot pi-interactions in proteins. **Journal of Molecular Biology**, v. 307, n. 1, p. 357-377, 2001.

136 MOSBAH, A. et al. Solution structure of the module X2_1 of unknown function of the cellulosomal scaffolding protein CipC of Clostridium cellulolyticum. **Journal of Molecular Biology**, v. 304, n. 2, p. 201-217, 2000.

137 ALTSCHUL, S. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, 1990.

138 PUNTA, M. et al. The Pfam protein families database. **Nucleic Acids Research**, v. 40, n. 1, p. 290-301, 2012.

139 TOTIR, M. et al. Macro-to-micro structural proteomics: native source proteins for high-throughput crystallization. **Plos One**, v. 7, n. 2, 2012. doi: 10.1371/journal.pone.0032498.

140 CAIRNS, J.; ESEN, A. Beta-glucosidases. **Cellular and Molecular Life Sciences**, v. 67, n. 20, p. 3389-3405, 2010.

141 GLOSTER, T. et al. Glycosidase inhibition: an assessment of the binding of 18 putative transition-state mimics. **Journal of the American Chemical Society**, v. 129, n. 8, p. 2345-2354, 2007.

142 STEPPER, J. et al. Structure and activity of the Streptococcus pyogenes family GH1 6-phospho- β -glucosidase SPy1599. **Acta Crystallographica Section D-biological crystallography**, v. 69, n. 1, p. 16-23, 2013.

143 CHUENCHOR, W. et al. The structural basis of oligosaccharide binding by rice BGluc1 beta-glucosidase. **Journal of Structural Biology**, v. 173, n. 1, p. 169-179, 2011.

144 HUA, Y. et al. Enzymatic and structural characterization of hydrolysis of gibberellin A4 glucosyl ester by a rice beta-D-glucosidase. **Archives of Biochemistry and Biophysics**, v. 537, n. 1, p. 39-48, 2013.

145 HENRISSAT, B.; BAIROCH, A. New families in the classification of glycosyl hydrolases based on amino-acid-sequence similarities. **Biochemical Journal**, v. 293, n. 3, p. 781-788, 1993.

146 HENRISSAT, B.; BAIROCH, A. Updating the sequence-based classification of glycosyl hydrolases. **Biochemical Journal**, v. 316, n. 2, p. 695-696, 1996.

147 VINCENNELLI, R. et al. High-throughput protein expression screening and purification in *Escherichia coli*. **Methods**, v. 55, n. 1, p. 65-72, 2011.

148 SAKON, J. et al. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. **Nature Structural Biology**, v. 4, n. 10, p. 810-818, 1997.

149 ARAI, T. et al. Sequence of celQ and properties of CelQ, a component of the *Clostridium thermocellum* cellulosome. **Applied Microbiology and Biotechnology**, v. 57, n. 5-6, p. 660-666, 2001.

150 RAGHOTHAMA, S. et al. Solution structure of the CBM10 cellulose binding module from *Pseudomonas xylanase A*. **Biochemistry**, v. 39, n. 5, p. 978-984, 2000.

151 ZIEGELHOFFER, T.; RAASCH, J.; AUSTIN-PHILLIPS, S. Dramatic effects of truncation and sub-cellular targeting on the accumulation of recombinant microbial cellulase in tobacco. **Molecular Breeding**, v. 8, n. 2, p. 147-158, 2001.

152 GAL, L. et al. CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. **Journal of Bacteriology**, v. 179, n. 21, p. 6595-6601, 1997.

153 ARAI, T. et al. Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CelJ: importance of the CBM to cellulose hydrolysis. **Journal of Bacteriology**, v. 185, n. 2, p. 504-512, 2003.

154 MINGARDON, F. et al. Comparison of family 9 cellulases from mesophilic and thermophilic bacteria. **Applied and Environmental Microbiology**, v. 77, n. 4, p. 1436-1442, 2011.

155 ANBARASAN, S. et al. Effect of glycosylation and additional domains on the thermostability of a family 10 xylanase produced by thermopolyspora flexuosa. **Applied and Environmental Microbiology**, v. 76, n. 1, p. 356-360, 2010.

156 HOLM, L.; ROSENSTROM, P. Dali server: conservation mapping in 3D. **Nucleic Acids Research**, v. 38, p. 545-549, 2010.

157 TORMO, J. et al. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. **Embo Journal**, v. 15, n. 21, p. 5739-5751, 1996.

158 SULZENBACHER, G. et al. Crystal structure of thermotoga maritima alpha-L-fucosidase - insights into the catalytic mechanism and the molecular basis for fucosidosis. **Journal of Biological Chemistry**, v. 279, n. 13, p. 13119-13128, 2004.

159 CAINES, M. et al. Structural and mechanistic analyses of endo-glycoceramidase II, a membrane-associated family 5 glycosidase in the Apo and G(M3) ganglioside-bound forms. **Journal of Biological Chemistry**, v. 282, n. 19, p. 14300-14308, 2007.

160 WU, T. et al. Diverse substrate recognition mechanism revealed by thermotoga maritima Cel5A structures in complex with cellotetraose, cellobiose and mannotriose. **Biochimica et Biophysica Acta-proteins and proteomics**, v. 1814, n. 12, p. 1832-1840, 2011.

161 MACKENZIE, L. et al. Crystal structure of the family 7 endoglucanase I (Cel7B) from Humicola insolens at 2.2 angstrom resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate. **Biochemical Journal**, v. 335, p. 409-416, 1998.

162 VUONG, T.; WILSON, D. Glycoside hydrolases: catalytic base/nucleophile diversity. **Biotechnology and Bioengineering**, v. 107, n. 2, p. 195-205, 2010.

163 REY, M. et al. Complete genome sequence of the industrial bacterium bacillus licheniformis and comparisons with closely related bacillus species. **Genome Biology**, v. 5, n. 10, 2004. doi:10.1186/gb-2004-5-10-r77.

164 SALKINOJA-SALONEN, M. S. et al. Toxigenic strains of *Bacillus licheniformis* related to food poisoning. **Applied and Environmental Microbiology**, v. 65, n. 10, p. 4637-4645, 1999.

165 SANZ-APARICIO, J. et al. Crystal structure of beta-glucosidase a from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosyl hydrolases. **Journal of Molecular Biology**, v. 275, n. 3, p. 491-502, 1998.

ANEXO A – Primeira página do artigo sobre a cristalização da enzima endoglucanase III de *T. harzianum*, publicado na revista Acta Crystallographica Section F em 2012.

crystallization communications

Acta Crystallographica Section F
Structural Biology
and Crystallization
Communications
ISSN 1744-3091

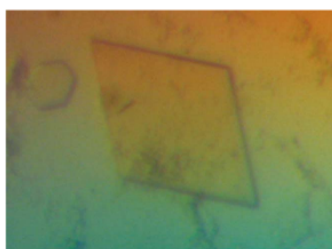
Marcelo Vizoná Liberato,^a
Wesley Cardoso Generoso,^b
Wilson Malagó Jr,^b Flávio
Henrique-Silva^b and Igor
Polikarpov^{a*}

^aInstituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador São-carlense 400, 13566-590 São Carlos-SP, Brazil, and

^bDepartamento de Genética e Evolução, Universidade Federal de São Carlos, Rodovia Washington Luís, km. 235, 13565-905 São Carlos-SP, Brazil

Correspondence e-mail:
ipolikarpov@ifsc.usp.br

Received 20 September 2011
Accepted 9 January 2012



© 2012 International Union of Crystallography
All rights reserved

Crystallization and preliminary X-ray diffraction analysis of endoglucanase III from *Trichoderma harzianum*

Endoglucanases are enzymes that hydrolyze cellulose and are important components of the cellulolytic complex. In contrast to other members of the complex, they cleave internal β -1,4-glycosidic bonds in the cellulose polymer, allowing cellulose to be used as an energy source. Since biomass is an important renewable source of energy, the structural and functional characterization of these enzymes is of interest. In this study, endoglucanase III from *Trichoderma harzianum* was produced in *Pichia pastoris* and purified. Crystals belonging to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 47.54$, $b = 55.57$, $c = 157.3$ Å, were obtained by the sitting-drop vapour-diffusion method and an X-ray diffraction data set was collected to 2.07 Å resolution.

1. Introduction

To date, humans have mostly relied on energy provided by fossil fuels, but the increasing energy needs of the growing human population and the recognition of the problems of fossil fuels (pollution, global warming and scarcity, to mention three) have provided a driver for looking at alternative sustainable sources of energy (Singh *et al.*, 2011).

The method of producing energy through the hydrolysis of biomass has been known for a long time (Antoni *et al.*, 2007). Biomass is a renewable resource and the final product of its transformation *via* the fermentation of depolymerized sugars, cellulosic ethanol, is less toxic, more readily biodegradable and produces fewer airborne pollutants than petroleum fuel (John *et al.*, 2011). Although it is not yet possible to fulfil the world's energy demands using renewable fuels, they are increasingly becoming a feasible option.

Cellulose forms tightly interwoven crystalline fibres which are generally very resistant to degradation. However, a number of microorganisms are able to efficiently hydrolyze cellulose. For example, *Hypocrea* fungi in general and *H. jecorina* in particular are able to secrete large amounts of cellulases to hydrolyze cellulosic biomass. Another member of the same group, *H. lixii*, also known as *Trichoderma harzianum*, has been utilized in phytopathogen control (Monteiro *et al.*, 2010) and more recently has been described as a competent cellulolytic fungus which produces a balanced mixture of cellulases (de Castro, Ferreira *et al.*, 2010). It has been demonstrated that *T. harzianum* secretes large quantities of endoglucanases when cultivated on pretreated sugarcane bagasse (de Castro, Pedro *et al.*, 2010).

The cellulolytic complex produced by most fungal species is formed of at least three major components: (i) endoglucanases (EGs), which hydrolyze the internal β -1,4-glycosidic bonds in the cellulose polymer, thus introducing two new internal chain termini; (ii) cellobiohydrolases (CBHs), which act on the reduced and nonreduced termini of cellulose polymer chains, thus degrading cellulose into cellobiose; and (iii) β -glycosidases, which hydrolyze cellobiose to glucose (Medve *et al.*, 1998). A large number of endoglucanases have been identified and they have been classified into glycoside hydrolase families based on their amino-acid sequence similarity (Henrissat & Davies, 2000). Despite the differences in their three-dimensional structures, the catalytic mechanism attributed to all endoglucanases is

ANEXO B – Primeira página do Artigo sobre a estrutura cristalográfica e simulação da dinâmica molecular da enzima endoglucanase III de *T. harzianum*, publicado na revista PLOS ONE em 2013.

OPEN ACCESS Freely available online

PLOS ONE

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

Érica T. Prates¹, Ivana Stankovic¹, Rodrigo L. Silveira¹, Marcelo V. Liberato², Flávio Henrique-Silva³, Nei Pereira, Jr.⁴, Igor Polikarpov^{2*}, Munir S. Skaf^{1*}

1 Institute of Chemistry, State University of Campinas–UNICAMP, Cx.P. 6154, Campinas, São Paulo, Brazil, **2** Institute of Physics of São Carlos, University of São Paulo, São Carlos, São Paulo, Brazil, **3** Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo, Brazil, **4** Centro de Tecnologia, Escola de Química, Laboratório de Desenvolvimento de Bioprocessos (LaDeBio), Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Plant biomass holds a promise for the production of second-generation ethanol via enzymatic hydrolysis, but its utilization as a biofuel resource is currently limited to a large extent by the cost and low efficiency of the cellulolytic enzymes. Considerable efforts have been dedicated to elucidate the mechanisms of the enzymatic process. It is well known that most cellulases possess a catalytic core domain and a carbohydrate binding module (CBM), without which the enzymatic activity can be drastically reduced. However, Cel12A members of the glycosyl hydrolases family 12 (GHF12) do not bear a CBM and yet are able to hydrolyze amorphous cellulose quite efficiently. Here, we use X-ray crystallography and molecular dynamics simulations to unravel the molecular basis underlying the catalytic capability of endoglucanase 3 from *Trichoderma harzianum* (ThEG3), a member of the GHF12 enzymes that lacks a CBM. A comparative analysis with the *Cellulomonas fimi* CBM identifies important residues mediating interactions of EG3s with amorphous regions of the cellulose. For instance, three aromatic residues constitute a harboring wall of hydrophobic contacts with the substrate in both ThEG3 and CfCBM structures. Moreover, residues at the entrance of the active site cleft of ThEG3 are identified, which might hydrogen bond to the substrate. We advocate that the ThEG3 residues Asn152 and Glu201 interact with the substrate similarly to the corresponding CfCBM residues Asn81 and Arg75. Altogether, these results show that CBM motifs are incorporated within the ThEG3 catalytic domain and suggest that the enzymatic efficiency is associated with the length and position of the substrate chain, being higher when the substrate interact with the aromatic residues at the entrance of the cleft and the catalytic triad. Our results provide guidelines for rational protein engineering aiming to improve interactions of GHF12 enzymes with cellulosic substrates.

Citation: Prates ET, Stankovic I, Silveira RL, Liberato MV, Henrique-Silva F, et al. (2013) 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 8(3): e59069. doi:10.1371/journal.pone.0059069

Editor: Chandra Verma, Bioinformatics Institute, Singapore

Received: October 5, 2012; **Accepted:** February 11, 2013; **Published:** March 14, 2013

Copyright: © 2013 Prates et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors acknowledge the support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via grants 08/56255–9, 09/54035–4 and 10/08680–2; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via grants 490022/2009–0, 550985/2010–7. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ipolikarpov@ifsc.usp.br (IP); skaf@iqm.unicamp.br (MS)