

BIOINFORMÁTICA ESTRUTURAL E EXPRESSÃO RECOMBINANTE COMO FERRAMENTAS DE ESTUDO DE UMA NOVA ENDOGLUCANASE

ALESSANDRA NEIS¹; FREDERICO SCHMITT KREMER²; RAFAEL WOLOSKI DOS SANTOS³; LUCIANO DA SILVA PINTO⁴

¹Universidade Federal de Pelotas – alessandra.neis@ufpel.edu.br

²Universidade Federal de Pelotas – fred.s.kremer@gmail.com

³Universidade Federal de Pelotas – rafaelwoloski@gmail.com

⁴Universidade Federal de Pelotas – ls_pinto@hotmail.com

1. INTRODUÇÃO

As enzimas estão amplamente envolvidas em uma variedade de atividades industriais e contribuem para a utilização de resíduos, que estes e os efluentes sejam menos tóxicos e na diminuição de custos operacionais, comparado ao uso de produtos químicos (ABDELRAHEEM et al., 2019; CHAPMAN; ISMAIL; DINU, 2018). As celulasas microbianas são enzimas ativas em polímeros orgânicos presentes na natureza, como a celulose, e essa vantagem é amplamente utilizada para a desconstrução de biomassa (HIMMEL et al., 2007). As endoglucanases fúngicas, em específico, clivam as ligações glicosídicas β -1,4 no interior da cadeia de celulose e são componentes importantes para a indústria do etanol de segunda geração, têxtil, de papel e celulose, e detergentes (BEHERA et al., 2017).

As endoglucanases da família 5 (GH5) correspondem a 55% das enzimas secretadas pelo principal fungo utilizado na indústria, *Trichoderma reesei* (SUOMINEN et al., 1993), porém, mesmo sendo a família com maior quantidade de dados estruturais reportados, grande parte ainda não é bem caracterizada. A grande quantidade de dados gerados a partir de técnicas de sequenciamento fez da bioinformática uma ferramenta importante na elucidação de mecanismos e função, através de estudos estruturais e funcionais. Estas análises fornecem informações complementares sobre o comportamento das moléculas, prevendo características esperadas ou observando em detalhes processos difíceis de observar *in vitro*. Além disso, a estrutura tridimensional permite observar a organização dos sítios ativos e alostéricos, conformação geral e organização da estrutura secundária (ZHANG; ZHANG, 2013), o que é essencial para entender o funcionamento da molécula. Entretanto, algumas ferramentas bastante difundidas ainda são insuficientemente aplicadas para endoglucanases GH5, em que poderiam fornecer dados de relevância para o estudo de moléculas dessa família ainda não caracterizadas experimentalmente.

Assim, o objetivo deste estudo foi empregar ferramentas de bioinformática, aliadas à expressão heteróloga em *Escherichia coli* de uma nova endoglucanase da família 5, subfamília 5 (GH5_5). O estudo da enzima ocorreu através de alinhamento múltiplo de sequências, modelagem molecular, *docking* com a celotetraose, dinâmica molecular, análise das características bioquímicas e obtenção da molécula recombinante.

2. METODOLOGIA

Seis candidatas foram analisadas quanto a suas características e potencial celulolítico, provindas de um estudo prévio de nosso grupo (CAGLIARI, 2018). Os códigos de acesso do banco de dados foram alterados devido ao potencial de aplicação biotecnológica, sendo as candidatas denominadas “EGs”. As sequências das enzimas foram recuperadas do banco de dados Uniprot e o *software* Clustal Omega possibilitou observar a conservação de sequências através de alinhamento

múltiplo (MSA). Análises estruturais incluíram: modelagem molecular pelo I-TASSER, refinamento e análises de qualidade (QMEAN, FG-MD, gráfico de Ramachandran, *docking* molecular (AutoDock Vina e dinâmica molecular por 200 ns pelo GROMACS (v 5.0). As análises de características bioquímicas foram feitas nos *softwares* NetNGlyc, NetOGlyc, Protein-Sol, CCSol e SolPro e ProtParam.

A expressão heteróloga se deu apenas com a melhor enzima candidata, produzida a partir de gene sintético (Epoch Life Science Inc.) otimizado para *E. Coli*. As cepas STAR, Rosetta e pLysS foram transformadas por choque térmico. A expressão foi induzida em 500 mL de meio Luria-Bertani, acrescido de ampicilina (100 mg/mL), através de β -D-1-tiogalactopiranosida (IPTG) (100 mM) a 1%. A purificação das porções solúvel e insolúvel se deu em coluna de matriz de sefarose carregada com níquel His Trap 1 mL (Cytiva Life Sciences Inc.) em purificador GE ÄKTA Purifier 100 FPLC System (GE Healthcare Inc.). Os resultados foram visualizados através da técnica de SDS-PAGE em gel de poliacrilamida 12% e *Western Blotting*.

3. RESULTADOS E DISCUSSÃO

As seis novas endoglucanases GH5_5 (EG2, EG3, EG4, EG5, EG6 e EG8) foram avaliadas quanto ao seu potencial *in silico* e viabilidade de obtenção *in vitro*. Nenhuma apresentou anotação a nível experimental no banco de dados de sequências de proteínas Uniprot e os resíduos conservados foram identificados por MSA em relação à *T. Reesei* (PDB: 3QR3). Além disso, as enzimas EG4, EG5 e EG8 apresentaram módulo de ligação à celulose, enquanto EG2, EG3 e EG6 continham apenas o módulo catalítico. A identidade entre as sequências de aminoácidos em relação à TrCel5A foi de 68,75 (EG2), 67,39% (EG3), 68,05% (EG4), 53,40% (EG5), 62,62% (EG6) e 28,65% (EG8).

As sequências das seis moléculas foram usadas para obter seu modelo tridimensional (3D) e a análise de qualidade e refinamento demonstrou que poderiam ser utilizados em posteriores análises (QMEAN = 0,87). O ideal é que 90% dos aminoácidos estejam nas regiões mais favorecidas do gráfico de Ramachandran, porém, a enzima EG2 foi a que atingiu a maior taxa de aminoácidos corretamente posicionados nas regiões mais favoráveis (76,5%) e favoráveis (22,8%). Em conjunto com QMEAN e análise visual, foi possível verificar que as enzimas estavam bem modeladas e empacotadas.

A afinidade enzima-substrato (kcal/mol) foi avaliada através de *docking* molecular. Como controle, foi feito o *re-docking* da molécula AnCel5A (PDB ID: 5I79), estruturalmente muito similar à *T. reesei* (RMSD = 1.024), com a celotetraose. Este método, aplicado à proteínas modeladas, é comparável ao de estruturas resolvidas e o refinamento com FG-MD melhora o empacotamento da estrutura (DU et al., 2015). Outros estudos com celulasas são difíceis de serem comparados, pois cada programa utiliza parâmetros diferentes, além de não existir um *threshold* para a energia de ligação (SELVAM et al., 2017). A enzima que mais se destacou foi EG2, por formar pontes de hidrogênio corretamente com os resíduos, o ligante estar corretamente posicionado no sítio ativo e apresentar energia de ligação superior às restantes (-7.2 kcal/mol). Esta foi escolhida para análises de dinâmica molecular e expressão heteróloga, uma vez que tem o

tamanho adequado para ser obtida de forma recombinante e apresentou bons resultados na modelagem e *docking* molecular com o substrato celulósico.

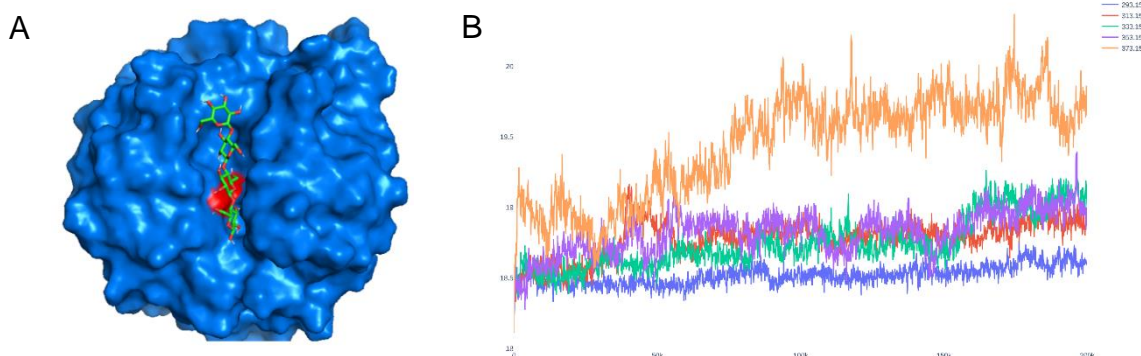


Figura 1. Docking e dinâmica molecular da enzima EG2. A: encaixe da celotetraose no sítio ativo. B: padrão do raio de giro da dinâmica molecular (200 ns). Em azul: 20 °C; vermelho: 40°C; verde: 60 °C; roxo: 80 °C; laranja: 100 °C.

Simulações de dinâmica molecular permitem analisar o comportamento molecular em diferentes temperaturas e tem sido empregadas para outras celulasas. Avaliou-se o padrão de raio de giro (Gyr) e a variação da posição dos átomos (RMSD) em 20, 40, 60, 80 e 100 °C, sendo o último indicativo de desnaturação proteica. A menor variação no Gyr ocorreu na temperatura de 20 °C, enquanto nas temperaturas de 40, 60 e 80 °C a enzima demonstrou-se razoavelmente estável, a não ser por alguns picos que podem representar instabilidade. Os dados de RMSD (não mostrados) corroboraram com Gyr.

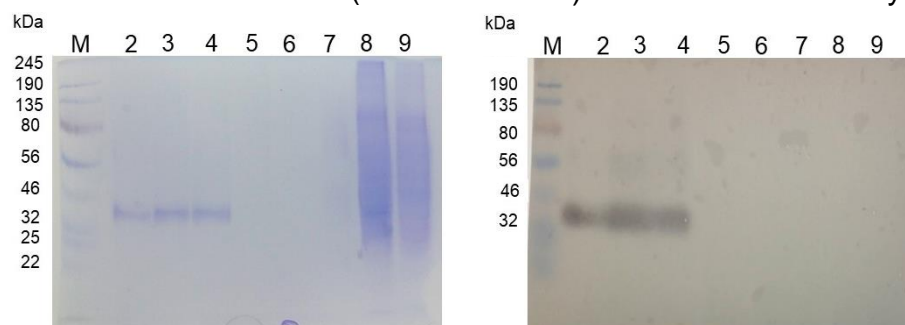


Figura 2. SDS-PAGE e *Western Blot* das porções purificadas solúvel e insolúvel produzidas em *E. coli* STAR. M: marcador de peso molecular de proteínas (New England BioLabs Inc.); 2, 3 e 4: fração insolúvel purificada; 5, 6 e 7: porção solúvel purificada; 8: extrato total de *E. coli* STAR induzida; 9: extrato total de *E. coli* STAR não induzida.

Análises das características bioquímicas mostraram 3 possíveis sítios de glicosilação e resultados discrepantes para a solubilidade, provavelmente devido aos diferentes algoritmos utilizados. Protein-Sol detectou uma probabilidade de 32,9%, CCSol 61% e SolPro 61,3%. Durante a expressão heteróloga, confirmou-se que na cepa STAR a proteína encontrava-se insolúvel (Figura 2). Nas cepas Rosetta e pLysS não foi produzida proteína recombinante (dados não mostrados). Isso está de acordo com outros estudos, que relatam a dificuldade de obter endoglucanases fúngicas em procariotos (LAMBERTZ et al., 2014). Além disso, a EG2 possui 3 pontes dissulfeto, mas *E. Coli* não tem a maquinaria necessária para esta e outras modificações, como glicosilações, que podem ter influenciado na obtenção da molécula (BERKMEN, 2012; FRANCIS; PAGE, 2010). É comum que proteínas não sejam expressas em *E. coli*; assim, outras estratégias de

recuperação incluem testar outras cepas, como Origami, e alterar as condições de indução da expressão, como a temperatura (JUTURU; WU, 2014).

4. CONCLUSÕES

A partir deste estudo, foi possível avaliar a aplicabilidade de ferramentas bioinformáticas para a caracterização de seis endoglucanases GH5_5 não descritas anteriormente. Concluiu-se que as estratégias computacionais auxiliam na prospecção e predição de características essenciais da funcionalidade dessas moléculas, mas ainda existem divergências entre os resultados de bioinformática e os experimentos. Não é possível afirmar que os métodos *in silico* não funcionaram, uma vez que é comum cepas não expressarem.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELRAHEEM, E. M. M. et al. Biocatalysis explained: From pharmaceutical to bulk chemical production. **Reaction Chemistry and Engineering**, v. 4, n. 11, p. 1878–1894, 2019.
- BEHERA, B. C. et al. Microbial cellulases – Diversity & biotechnology with reference to mangrove environment : A review. **Journal of Genetic Engineering and Biotechnology**, v. 15, n. 1, p. 197–210, 2017.
- BERKMEN, M. **Production of disulfide-bonded proteins in Escherichia coli Protein Expression and Purification** Academic Press, , 1 mar. 2012.
- CAGLIARI, R. **Prospecção e análise in vitro de enzimas celulasas de espécies de fungos fitopatogênicos com potencial aplicação na produção de biocombustíveis**. [s.l.] Universidade Federal de Pelotas, 2018.
- CHAPMAN, J.; ISMAIL, A. E.; DINU, C. Z. Industrial applications of enzymes: Recent advances, techniques, and outlooks. **Catalysts**, v. 8, n. 6, p. 20–29, 2018.
- DU, H. et al. Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. **Methods (San Diego, Calif.)**, v. 71, p. 77–84, jan. 2015.
- FRANCIS, D. M.; PAGE, R. Strategies to optimize protein expression in E. coli. **Current Protocols in Protein Science**, n. SUPPL. 61, p. 1–29, 2010.
- HIMMEL, M. E. et al. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. **Science**, v. 315, n. 5813, p. 804–807, 2007.
- JUTURU, V.; WU, J. C. Microbial cellulases: Engineering, production and applications. **Renewable and Sustainable Energy Reviews**, v. 33, p. 188–203, 1 maio 2014.
- LAMBERTZ, C. et al. Challenges and advances in the heterologous expression of cellulolytic enzymes: A review. **Biotechnology for Biofuels**, v. 7, n. 1, p. 135, 18 dez. 2014.
- SELVAM, K. et al. Cellulase enzyme: Homology modeling, binding site identification and molecular docking. **Journal of Molecular Structure**, v. 1150, p. 61–67, 2017.
- SUOMINEN, P. L. et al. High frequency one-step gene replacement in *Trichoderma reesei*. II. Effects of deletions of individual cellulase genes. **MGG Molecular & General Genetics**, v. 241, n. 5–6, p. 523–530, dez. 1993.
- ZHANG, X.; ZHANG, Y. P. CELLULASES: CHARACTERISTICS, SOURCES, PRODUCTION, AND APPLICATIONS. In: YANG, S.-T.; EL-ENSHASY, H. A.; THONGCHUL, N. (Eds.). **Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals, and Polymers**. 1. ed. [s.l.] John Wiley & Sons, Inc., 2013. p. 131–146.