Structural pattern detection for engineering more efficient enzymes for second-generation biofuel production

DCB Mariano[1], TS Correia[1], JRPM Barroso[1], RC de Melo-Minardi[1]

[1] LBS - Laboratory of Bioinformatics and Systems. Department of Computer Science. Federal University of Minas Gerais. Belo Horizonte, Brazil

β-glucosidase (E.C. 3.2.1.21) is the main enzyme in the process of the second-generation biofuel production. It acts synergically with endoglucanases and exoglucanases to degrade cellulose in glucose, which will be used for bioethanol production. β-glucosidase plays a key role in the last step of this enzymatic system, converting cellobiose, a disaccharide that inhibits both endo- and exoglucanases, in glucose. However, most of the β-glucosidases known are inhibited by glucose. Recently, our group has preceded a systematic literature review (SLR) to evaluate the state-of-art of β-glucosidase researches. In this SLR, we collected 23 sequences and three-dimensional structures of β-glucosidases with high tolerance to glucose inhibition. In this work, we propose an analysis of these collected structures to detect patterns that can be used to engineering of β-glucosidases with high catalytic efficiency, and also, to detect possible glucose-tolerant β-glucosidases in data obtained by high-throughput platforms of sequencing. We collected 3,991 β-glucosidases sequences of the GH1 family (described in the literature as more efficient for biofuel production) from UniProt and performed homology modeling. In the first step, we analyzed patterns in the primary structure using amino acid k-mer frequency and singular value decomposition (SVD). However, the sequences were not sufficient to cluster the glucose-tolerant β-glucosidases. To investigate whether the patterns appear in other structural levels, we performed structural alignment of 21 three-dimensional GH1 β-glucosidases structures with the β-glucosidase from termite *Neotermes koshunensis* in complex with cellobiose (PDB: 3VIK). Then, we collected the amino acids at a distance of 6 Å, 6.5 Å, 7 Å, 7.5 Å, and 8 Å of the substrate, and submitted them to the software aCSM-ALL to detect contact frequencies based on cutoff of atomic distances and on physicochemical information. We reduced the noise of the atomic distances collected with SVD, and with this information we constructed a glucose tolerance signature to identify high efficient β-glucosidases for biofuel production. We also characterized an ideal active site based on multiple alignments of high tolerant β-glucosidases. The glucose tolerance signature can be used to detect proteins potential targets for cellulose degradation. Also the characterization of an ideal active site based on glucose-tolerant β-glicosidase data can be useful for enzyme engineering with high catalytic efficiency and may help shed light on the second-generation biofuel production.

Supported by: FAPEMIG, CNPq, and CAPES (51/2013 - 23038.004007/2014-82).