Determining the main factors contributing to xylanase resistance to different pH by applying the asymptotic distribution of a transformation on the Pearson correlation coefficient

E. Ebrahimie, M. Ebrahimi, M. Ebrahimi, Z. Zinati, A. Delavari, M. Mohammadi-dehcheshmeh

Abstract—The main role of xylanolytic enzyme, β-endoxylanase (1,4-β-D-xylan xylanohydrolase, EC 3.2.1.8) is to convert the polymer xylan to xylosaccharides. Many xylanase and xylosidase genes from fungi and bacteria have been analyzed and their encoded enzymes have been isolated and characterized. It has already been shown that those enzymes are active at a limited range of pH (with maximum, minimum and mean ± SE, 8, 2 and 5.47 ± 0.286, respectively). In some industries need for active enzymes in both alkaline and acidic pH have been elaborated and understanding the main factors contributing in pH resistance of theses enzymes are very important. So we looked at more than seventy properties of 30 xylanase proteins (active in different pH) by applying a feature selection algorithm which assigned a *p value* to each attribute based on the asymptotic distribution of a transformation on the Pearson correlation coefficient. The attributes were then sorted in a descending order of their importance to xylanase pH resistance based on calculated *p values*.

The results showed that the frequency of Arg, Ser, Pro, Tyr, the count of Arg, Pro, Trp, Gly, Leu, Gln, the frequency of positively charged residues, the count of hydrophobic residues, the count of positively charged residues, non-reduced cycteins extinctoin coefficient and reduced cycteins extinctoin coefficient were the most important features contributing to the resistance of xylanases at different pH, and thirteen other attributes were considered to have a marginal contribution to this function, while the other features were revealed to be unimportant. The significance of "important" and "marginal" properties in xylanase activity in both alkaline and acidic pH has been discussed in this paper.

***Index Terms*: pH resistance, Xylanase, Modeling, Bioinformatics.**

E. Ebrahimie, Department of Crop Production & Plant Breeding, School of Agriculture, Shiraz University, Shiraz, Iran. (corresponding author; e-mail: ebrahimiet@gmail.com)

M. Ebrahimi, Green Research Center, Qom University, QOM, IRAN

M. Ebrahimi, Department of Computer Science & Engineering, School of Engineering, Shiraz University, Shiraz, Iran.

Z. Zinati, Department of Crop Production & Plant Breeding, School of Agriculture, Shiraz University, Shiraz, Iran.

A. Delavari, Plant Biotechnology Center, School of Agriculture, Shiraz University, Shiraz, Iran.

M. Mohammadi-dehcheshmeh, Plant Research Center, Discipline of Wine and Horticulture, School of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia

M. Chetty, S. Ahmad, A. Ngom, S. W. Teng (Eds.): PRIB2008, Supplementary Conf. Proc., pp.268-271, 2008. (c) PRIB 2008

I. Introduction

XYLANASES (endo-β-1,4-xylanhydrolases; EC 3.2.1.8) catalyse the hydrolysis of xylan, which is the main component of hemicellulose [1]. Classified as families 10 and 11 of the glycoside hydrolase superfamily [2], these enzymes have great potential for application in food, feed, paper and pulp industries to reduce the amount of bleaching agents (often chlorine compounds) used, and so reduce the negative impact on the environment [3, 4]. Conventionally, the wood used for the production of pulp is treated at high temperature and basic pH (Kraft pulping). Increasing concern about waste and pollution with this traditional technology has led the industry to introduce processes such as biobleaching, which is treatment of pulp with enzymes [5, 6]. However, biobleaching with the current generation of xylanases leads to other serious problems: corrosion, due to high chloride content remaining in the treatment reactor, the need to use large volumes of acid for pH adjustment to facilitate the xylanase reaction and significant waste generation [7]. These problems could be solved by operating at a higher temperature and by reducing acid usage. However, this implies that enzymatic procedures using xylanases require proteins that exhibit high thermostability and activity over a broad pH range [8]. Xylanases for this purpose must also (a) lack cellulolytic activity to avoid hydrolysis of cellulose fibers, (b) be of relatively low molecular mass to facilitate diffusion into pulp fibers, and (c) be available in sufficient quantities at a low enough cost to be attractive to the industry [9]. The utilization of highly stabilized hyperthermophilic and alkalophilic xylanases would be very attractive for biobleaching processes. On the other hand, expression of those genes to produce thermophilic xylanases requires activity in the broad range of pH to accomplish their complete actions. That is why the focus of many projects is on the generation of molecularly engineered, bioactive xylanases expressed in alternative industrially appropriate host organisms [10, 11]. To induce genetically modified xylanases active in broad ranges of pH, understanding the features contributing to pH activity is inevitable. Although different random approaches such as

Error-Prone PCR have been employed to modify those genes, the successes have been low [12, 13]. Using non-random tools such as site-directed mutagenesis require more relevant information on the factors directly contributing to the enzymes pH activity [14].

There are usually a lot of attributes determining different characteristics of a protein molecule. As a result, the majority of time and effort spent in the model-building process involves examining which variables to include in the model. Feature selection allows the variable set to be reduced in size, creating a more manageable set of attributes for modeling. Adding feature selection to the analytical process has several benefits: 1. Simplifies and narrows the scope of the features that is essential in building a predictive model. 2. Minimizes the computational time and memory requirements for building a predictive model because focus can be directed to the subset of predictors that is most essential. 3. Leads to more accurate and/or more parsimonious models. 4. Reduces the time for generating scores because the predictive model is based upon only a subset of predictors.

It have been shown many different features of enzymes may be contributing to pH activity in broad ranges and here we are trying to use the mentioned feature selection algorithm on seventy four common xylanase protein features to determine the most relevant features contributing to pH activity of selected xylanases.

II. Materials and Methods

From Swiss-Prot/TrEMBL database, all xylanase proteins and nucleotides sequences were extracted and those whom had been approved by experts were selected (76 xylanases). Then the maximum temperature and optimum pH of xylanases were determined as they were mentioned by the authors who had registered the protein or gene sequences. From 76 xylanases studied here, in available literature, temperature and pH of only 30 xylanases had been reported by the authors, so the feature selection algorithm was just applied to these proteins with 74 protein attributes or features describing each of them. All features were classified as continuous variables except N-terminal amino acid which was classified as categorical. The analysis was executed in two phases, as follows: 1- Screening: In this phase, variables were examined against a set of statistical tests, and those which were recognized not to contain useful information were removed. Going through this phase, half-life E.Coli, frequency of carbon, frequency of nitrogen, frequency of oxygen, frequency of hydrogen, frequency of hydrophobic residues and frequency of other charged residues were excluded because their coefficient of variation were below the minimum accepted threshold (<0.1). Furthermore, N-terminal amino acid variable was eliminated because the percentage of records of this variable which were in a single category was above the maximum accepted threshold (>90%). 2- Ranking Predictors: At this step, one predictor was considered at a time only to determine how well each one predicts the target variable (optimum pH). The importance of each variable at this phase was calculated as $(1-p)$ where p was the $p\ value$ based on the asymptotic t distribution of a transformation t on the Pearson correlation coefficient r.

The Pearson correlation coefficient r was defined as:

$$r = \frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)/(N-1)}{\sqrt{s(x)^2 s(y)^2}}$$

Where:

$$\bar{x} = \sum_{i=1}^{N} x_i \Big/ N,$$

$$\bar{y} = \sum_{i=1}^{N} y_i \Big/ N \text{ and}$$

$s(x)^2$ and $s(y)^2$ were the sample variance of X and Y. The transformation t on r was given by:

$$t = r\sqrt{\frac{N-2}{1-r^2}}$$

And finally, under the null hypothesis that the population Pearson correlation coefficient $\rho = 0$, the p value was calculated as:

$$p\ value = \begin{cases} 0 & if\ r^2 = 1, \\ 2\Pr ob\{T > |t|\} & else. \end{cases}$$

when T was a random variable that followed a t distribution with $N-2$ degrees of freedom. However, the $p\ value$ calculated this way was a test of linear relationship between X and Y. If there were some nonlinear relationship, the test might fail to catch it.

Applying this model to the set of mentioned features categorized them into three groups, namely: “important” with a level of importance greater than 0.95, “marginal” with a level of importance between 0.90 and 0.95 and “unimportant” with a level of importance less than 0.90.

To evaluate the efficiency of proposed model in recognizing different enzyme’s features, we applied the model on the same number of features of cellulose enzyme.

III. Results and Discussion

The average length of xylanases protein studied here was 475 amino acids with maximum and minimum length of 1157 and 194, respectively. The average weight of proteins was 52 Kda with isoelectric point 6. In 93.1% of proteins the N-terminal amino acid was Methionine, in 3.4% of them the same amino acid was Glycine and in the rest it was Glutamine. The average count of carbon, nitrogen, oxygen and hydrogen were 34, 430, 457 and 4 respectively, while the average count of hydrophilic, hydrophobic and other residues were 150, 225 and 99. The average count of negatively and positively charged residues were nearly the same (50 and 41) but the average count of other charged residues were 384. The highest average count of different amino acids belonged to Ser (39) and the lowest average accounted for Cys (3). The highest and the lowest

average of frequency of different amino acids belonged to Gly (0.089) and Met (0.017). The minimum and maximum optimum pH of xylanases were 2 (AC: P33557 and P48824) and 7.5 (AC: P54865), respectively with average optimum pH of 5. The xylanases studied here were active between 30°C and 102°C and the average active temperature was 65°C.

By applying the mentioned algorithm, these features were recognized as "important", and were sorted in descending order of their importance; frequency of Arg (0.999), count of Arg (0.997), frequency of Ser (0.989), frequency of Pro (0.985), count of Pro (0.98), count of Trp (0.969), frequency of positively charged residues (0.967), count of Gly (0.963), count of Leu (0.962), frequency of Tyr (0.962), count of hydrophobic residues (0.96), count of Gln (0.96), count of positively charged residues (0.954), Non-reduced cycteins extinction coefficient (0.951) and reduced cycteins extinction coefficient (0.951). While the count of sulfur (0.949), count of other charged residues (0.947), length (0.946), count of hydrogen (0.946), weight (Kda) (0.945), count of His (0.945), frequency of hydrophilic residues (0.94), count of Ala (0.937), count of Phe (0.928), count of Met (0.926), count of other residues (0.924), count of Thr (0.919) and count of Asn (0.906) were "marginal". Other features showed to be "unimportant", the value less than 0.90. As mentioned in previous section, applying the same model on another enzyme, cellulose, showed that frequency of hydrophilic residues (0.994), Aliphatic index (0.994), frequency of Try and Leu (0.981), count of nitrogen (0.972), isoeleric point (0.959), frequency of Ala (0.953), count of other charged residues (0.951) and frequency of sulfur (0.951) features were "important" and eight more features were "marginal" and others were "unimportant".

Xylanases enzymes are gaining wide industrial and biotechnical interest due to their wide applications in many industries. So isolating or genetically manipulating genes encoding the xylanases enzymes active in both alkaline and acidic harsh conditions has gained a lot of interests and researches [15]. Many applications have already been proposed for a xylanase enzyme with activity in broad pH and also at the same time with thermal stability. Many methods so far have been practiced to genetically modify the genes encoding the enzyme to make xylanase enzymes active in both alkaline and acidic pH but it has already been elaborated the need for defining the most active features of enzymes contributing to enzyme activity in different pH.Here we applied an algorithm on more than seventy features of xylanase enzymes regarding their ability to stand at both alkaline and acidic pH which showed the frequency and count of Arg are the most important features of xylanases enzymes in pH activity. Frequency of Ser and Pro are the next most important features in this regard. Also count of Pro, Trp, Gly, Leu and Gln were the other most important features contributing to pH activity of xylanases. The model described here can be applied to other enzymes to predict the behavior of other enzymes in different pH ranges.

ACKNOWLEDGMENT

The authors greatly appreciate and acknowledge the support of Green Research Center and Shiraz University.

REFERENCES

[1] A. Sunna and G. Antranikian, "Xylanolytic enzymes from fungi and bacteria," *Crit Rev Biotechnol*, 17, pp. 39-67. 1997.

[2] M. Vardakou, P. Katapodis, M. Samiotaki, D. Kekos, G. Panayotou, and P. Christakopoulos, "Mode of action of family 10 and 11 endoxylanases on water-unextractable arabinoxylan," *Int J Biol Macromol*, 33, pp. 129-34. 2003.

[3] T. Dutta, R. Sengupta, R. Sahoo, S. Sinha Ray, A. Bhattacharjee, and S. Ghosh, "A novel cellulase free alkaliphilic xylanase from alkali tolerant Penicillium citrinum: production, purification and characterization," *Lett Appl Microbiol*, 44, pp. 206-11. 2007.

[4] R. Khandeparkar and N. B. Bhosle, "Application of thermoalkalophilic xylanase from Arthrobacter sp. MTCC 5214 in biobleaching of kraft pulp," *Bioresour Technol*, 98, pp. 897-903. 2007.

[5] H. J. Bae, H. J. Kim, and Y. S. Kim, "Production of a recombinant xylanase in plants and its potential for pulp biobleaching applications," *Bioresour Technol*, 2007.

[6] M. Paloheimo, A. Mantyla, J. Kallio, T. Puranen, and P. Suominen, "Increased production of xylanase by expression of a truncated version of the xyn11A gene from Nonomuraea flexuosa in Trichoderma reesei," *Appl Environ Microbiol*, 73, pp. 3215-24. 2007.

[7] A. Mantyla, M. Paloheimo, S. Hakola, E. Lindberg, S. Leskinen, J. Kallio, J. Vehmaanpera, R. Lantto, and P. Suominen, "Production in Trichoderma reesei of three xylanases from Chaetomium thermophilum: a recombinant thermoxylanase for biobleaching of kraft pulp," *Appl Microbiol Biotechnol*, 76, pp. 377-86. 2007.

[8] D. Chantasingh, K. Pootanakit, V. Champreda, P. Kanokratana, and L. Eurwilaichitr, "Cloning, expression, and characterization of a xylanase 10 from Aspergillus terreus (BCC129) in Pichia pastoris," *Protein Expr Purif*, 46, pp. 143-9. 2006.

[9] H. Balakrishnan, B. Kamal Kumar, M. Dutta-Choudhury, and M. V. Rele, "Characterization of alkaline thermoactive cellulase-free xylanases from alkalophilic Bacillus (NCL 87-6-10)," *J Biochem Mol Biol Biophys*, 6, pp. 325-34. 2002.

[10] S. Singh, A. M. Madlala, and B. A. Prior, "Thermomyces lanuginosus: properties of strains and their hemicellulases," *FEMS Microbiol Rev*, 27, pp. 3-16. 2003.

[11] J. Lin, L. M. Ndlovu, S. Singh, and B. Pillay, "Purification and biochemical characteristics of beta-D-xylanase from a thermophilic fungus, Thermomyces lanuginosus-SSBP," *Biotechnol Appl Biochem*, 30 (Pt 1), pp. 73-9. 1999.

[12] R. Ruller, L. Deliberto, T. L. Ferreira, and R. J. Ward, "Thermostable variants of the recombinant xylanase a from bacillus subtilis produced by directed evolution show reduced heat capacity changes," *Proteins*, 2007.

[13] M. Diaz, S. Rodriguez, J. M. Fernandez-Abalos, J. De Las Rivas, A. Ruiz-Arribas, V. L. Shnyrov, and R. I. Santamaria, "Single mutations of residues outside the active center of the xylanase Xys1 Delta from Streptomyces halstedii JM8 affect its activity," *FEMS Microbiol Lett*, 240, pp. 237-43. 2004.

[14] F. Fenel, A. J. Zitting, and A. Kantelinen, "Increased alkali stability in Trichoderma reesei endo-1, 4-beta-xylanase II by site directed mutagenesis," *J Biotechnol*, 121, pp. 102-7. 2006.

[15] W. W. Wakarchuk, W. L. Sung, R. L. Campbell, A. Cunningham, D. C. Watson, and M. Yaguchi, "Thermostabilization of the Bacillus circulans xylanase by the introduction of disulfide bonds," *Protein Eng*, 7, pp. 1379-86. 1994.