Biological Forum – An International Journal **11(2): 14-20(2019)**

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Application of Bioinformatic Tools for Prediction of Active pH and Temperature Stability of Endoglucanases Based on Coding Sequences from Metagenomic DNA Data

Khanh Hoang Viet Nguyen[1,2,3], Thi Thao Nguyen[4], Nam Hai Truong[1,2] and Thi Huyen Do[1,2*]

[1]Institute of Biotechnology, Vietnam Academy of Science and Technology, 18-Hoang Quoc Viet, Cau Giay, Ha Noi, Vietnam.

[2]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18-Hoang Quoc Viet, Cau Giay, Ha Noi, Vietnam.

[3]Institute of New Technology/Academy of Military Science and Technology, 17 Hoang Sam Street, Nghia Do, Cau Giay, Ha Noi, Vietnam.

[4]Vinh University, 182 Le Duan, Vinh, Nghe An, Vietnam.

(Corresponding author: Thi Huyen Do)
(Received 08 March 2019, Accepted 18 June, 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The active pH and temperature stability are important features of enzyme for application. In this study, AcalPred and Tm prediction tools at TBI software were applied for prediction of active pH and temperature stability of two sequences that mined from metagenomic DNA data, coding for endoglucanases (EGC). The predicted results showed that both EGC1, EGC2 were acidic enzymes and Tm of EGC1 was 55-65^{o}C, of EGC2 was lower than 55^{o}C. The EGC1 and EGC2 were expressed in *E. coli* under infusions form (EGC1) and chimeric form (SUMOEGC2). Both purified EGC1, SUMOEGC2 were active at acidic condition with optimal pH of 6 and 4 respectively. Both enzymes were stable at the temperature lower than 50^{o}C and EGC1 was more stable than SUMOEGC2. These results confirmed the feasibility of AcalPred and the Tm prediction tool for estimating active pH and temperature stability of enzymes based on the coding sequences.

Keywords: AcalPred, Tm prediction tool, *E. coli*, endoglucanase, goat rumen, metagenomic DNA data, recombinant enzymes, termite gut.

How to cite this article: Nguyen, Khanh Hoang Viet, Nguyen, Thi Thao, Truong, Nam Hai and Do, Thi Huyen (2019). Application of Bioinformatic tools for Prediction of Active pH and Temperature Stability of Endoglucanases Based on Coding Sequences from Metagenomic DNA data. *Biological Forum – An International Journal*, **11**(2): 14-20.

INTRODUCTION

Nowadays, the using of enzymes and proteins has become increasingly popular for industrial application such us food, agriculture, chemical, pharmacy by dint of reducing processing time, energy, cost as well as having low-environmental impact (Singh *et al.*, 2016). In particular, glycoside hydrolases are enzymes playing an important role in converting lignocellulose biomass into sugars for bioethanol production and other biological polymers (Matsumoto *et al.*, 2013; Zverlov and Schwarz 2008). However, in order to degrade lignocellulose consisting of complex components, it is necessary to use a cocktail enzymes for effective reaction (Percival-Zhang *et al.*, 2006; Koeck *et al.*, 2014). The enzymes in the cocktail must be active in the same condition, especially pH and temperature. At the optimal pH and temperature, enzymes express the highest activities leading to a decrease in the amount of enzymes for hydrolysis process (Antonov *et al.*, 2017; Khare *et al.*, 2015; Klein-Marcuschamer *et al.*, 2012). Thus, the determination of active temperature and pH ranges of each enzyme is needed for application in biological process and for combination of many different enzymes in a one-pot reaction to accomplish the highest hydrolysis efficiency (Antonov *et al.*, 2017; Khare *et al.*, 2015; Klein-Marcuschamer *et al.*, 2012). The significant portion of industrial enzymes is produced by recombinant technology based on the coding genes (Adrio and Demain 2010). Nowadays, the rapid progress in next generation sequencing technology and the simultaneous development of bioinformatics tools have allowed to explore novel genes coding for valuable enzymes/proteins (Almaabadi *et al.*, 2015; Kennedy *et al.*, 2013; Rochat *et al.*, 2019).

With purpose of mining lignocellulolytic genes, we sequenced metagenomic DNA of bacteria freely living in the gut of lower termite *Coptotermes gestroi* and in rumen of Vietnamese goats and obtained thousands lignocellulolytic genes (Do *et al.,* 2014, 2018). However, for producing recombinant enzymes to make an effective enzyme cocktail, we have to screen from thousands genes to select some candidates which can act effectively in a required pH and temperature. To choose tools for the prediction, in this study, we examined feasibility of Tm prediction tool produced by Taiwan Bioinformatics Institute and AcalPred tools for active pH range prediction designed by Lin *et al.,* (2013) in two genes candidates coding for endoglucanase which were mined from the metagenomic data. The temperature stability and optimal pH of the enzymes were confirmed by experiment on the recombinant enzymes expressed in *Escherichia coli*.

MATERIALS AND METHODS

A. Research conducted

Research conducted at Genetic Engineering Lab, Institute of Biotechnology, VAST, 18, Hoang Quoc Viet, Ha Noi, Vietnam.

B. Materials

The *egc1* gene of 1035 bps (NCBI code KP641169.1) coding for mature endogulucanase GH8 was screened from the metagenomic DNA data of Vietnamese termite's gut (Do *et al.,* 2014) (Fig. 1) and a gen *egc2* (GL13359) of 1545 bps coding for mature endoglucanase GH5 was selected from metagenomic DNA data of goat rumen (Do *et al.,* 2018) (Fig. 2).

```
tgtcgctggccagcatgggatcagtttaaaaaggactatatcagcgacggtgggcgcgtcattgaccccagtgacgcgcgcaaaattacc
 C  R  W  P  A  W  D  Q  F  K  K  D  Y  I  S  D  G  G  R  V  I  D  P  S  D  A  R  K  I  T
acctcagaagggcaaagctatgccatgttcttcgcgctggtggcgaatgaccgtagcgcttttgccacgttgtatgactggacacaagac
 T  S  E  G  Q  S  Y  A  M  F  F  A  L  V  A  N  D  R  S  A  F  A  T  L  Y  D  W  T  Q  D
aatctcgcccagggtgatctgagcgcgaatctgcccgcctggctgtggggcaaaaacgaagataaatgggcgatacttgacagcaactcc
 N  L  A  Q  G  D  L  S  A  N  L  P  A  W  L  W  G  K  N  E  D  K  W  A  I  L  D  S  N  S
gcctctgacgccgatctgtggatcgcctggtcgctgatggaagccgggcgcttatggaaggagccacgctacagcgacgccgggaaaggg
 A  S  D  A  D  L  W  I  A  W  S  L  M  E  A  G  R  L  W  K  E  P  R  Y  S  D  A  G  K  G
ctgctgtcgcggattgctaaagatgaagtggtcactgtgccggggcttggctcgatgctgttgccggggaaagtcggttttgtcgacgac
 L  L  S  R  I  A  K  D  E  V  V  T  V  P  G  L  G  S  M  L  L  P  G  K  V  G  F  V  D  D
aaaagctggcgtttcaaccccagctatttcccgccgcaactgaccagttactttgcccgtcttggcgcgccgtggagcacattgcgggaa
 K  S  W  R  F  N  P  S  Y  F  P  P  Q  L  T  S  Y  F  A  R  L  G  A  P  W  S  T  L  R  E
actaacctgcgcctgctgctggaaactgcgccaaaaggtttttcgccggactgggtgcgctatgaaaaagacaaaggctggcagttaaag
 T  N  L  R  L  L  L  E  T  A  P  K  G  F  S  P  D  W  V  R  Y  E  K  D  K  G  W  Q  L  K
caggataagacgctggttggcagctatgacgcgattcgcgtctatctgtgggcgggcatgctgcacgacagcgatccgcaaaaagcgcgc
 Q  D  K  T  L  V  G  S  Y  D  A  I  R  V  Y  L  W  A  G  M  L  H  D  S  D  P  Q  K  A  R
ctgctgaagcgctttaacccgatggcagtgcaaaccacccgtgaaggattgccgccggagaaagttgatatcgcgaccggcaaggtcacc
 L  L  K  R  F  N  P  M  A  V  Q  T  T  R  E  G  L  P  P  E  K  V  D  I  A  T  G  K  V  T
aataccggcccggtcgggttctcggcctcgctgctgcctttcttacaaaatcgcgatgcccaggctgttcagcgccagcgcgtggcggat
 N  T  G  P  V  G  F  S  A  S  L  L  P  F  L  Q  N  R  D  A  Q  A  V  Q  R  Q  R  V  A  D
aacttccctggcgcggacgcctattttaactatgtgctaaccctgtttggacaaggctgggatcaacaccgttttcgtttcaccgcccgc
 N  F  P  G  A  D  A  Y  F  N  Y  V  L  T  L  F  G  Q  G  W  D  Q  H  R  F  R  F  T  A  R
                     ggtgaattactacctgactggggccaggaatgcgcaagttcacag
                      G  E  L  L  P  D  W  G  Q  E  C  A  S  S  Q
```

Fig. 1. The *egc1* sequence coding for mature endoglucanase GH8 (EGC1) was mined in metegenomic DNA data from bacteria free-living in termite gut.

Strains: *E. coli* DH10B - Invitrogen (USA) and *E. coli* BL21 (DE3) were used respectively for cloning and expression of the genes.

B. Prediction of pH for enzyme activities

For prediction of acidic or alkaline enzymes we used a sequence-based tool AcalPred (http://lin-group.cn/server/AcalPred) (Lin *et al.,* 2013). A feature selection technique was used to pick out a number of informative features. On the basis of these features, the support vector machine (SVM) was performed to establish prediction model. The total score is one, equivalent to acidic score added to alkaline score.

C. Prediction of thermal stability of enzymes

From the amino acid sequences, thermal stability of enzymes was predicted based on the bonds between amino acid such as hydrogen, Van der waals and hydrophobic interaction (Ebrahimi and Ebrahimie 2010). In this study, Tm prediction in TBI software (http://tm.life.nthu.edu.tw/) was used. The thermal stability of an enzyme is demonstrated at three levels: above 65^{o}C, 55-65^{o}C and below 55^{o}C.

D. Genes expression and purification

The genes coding for mature enzymes (Fig. 1, Fig. 2) were artifical synthezied in Genscript (HongKong). The *egc1* gene was inserted into *Nco*I-*Xho*I sites in pET22b(+) to generate pET22egc1. Because protein EGC2 was predicted to be too alkaline (pI of 9.64) to be expressed in *E. coli*, thus the gene was fused with SUMO gene for expression in *E. coli*.

```
gcggtgccggttggtagcggtgacaacgcggtgattaacagcattaccgcggaggcggcgccgtgagcgttccgaaggtgaccggtctg
 A  V  P  V  G  S  G  D  N  A  V  I  N  S  I  T  A  E  A  A  T  V  S  V  P  K  V  T  G  L
aaaagccagaccctgggcaccgataaggttaaactgacctggaacaagattagcaacgcgaccggttaccaggtgttcatgcagaagaac
 K  S  Q  T  L  G  T  D  K  V  K  L  T  W  N  K  I  S  N  A  T  G  Y  Q  V  F  M  Q  K  N
ggcgtttatacccgtctgattagcatctacagcaacaccaacagctataccgtgagcaacctgccggcggcgcgtatcaccaacctgaaa
 G  V  Y  T  R  L  I  S  I  Y  S  N  T  N  S  Y  T  V  S  N  L  P  A  A  R  I  T  N  L  K
gtgcgtgcgtggcgtaccgttggtggcaaaaactactttggtccgctgagcgcggcgcgtaccaccgcgaccaacccggcgaaggttacc
 V  R  A  W  R  T  V  G  G  K  N  Y  F  G  P  L  S  A  A  R  T  T  A  T  N  P  A  K  V  T
aacattcagagcaccgcgggcagcaccagcatcaagtttacctggaacaaagtgaactgcagctactatcgtatttacgttaagcaagcg
 N  I  Q  S  T  A  G  S  T  S  I  K  F  T  W  N  K  V  N  C  S  Y  Y  R  I  Y  V  K  Q  A
gacggcaagtggaaaggcctgaaaaacagcaccaccaacagcgtgaccgttaccggtctgaacgcgaacaccagctatacctttgcgtg
 D  G  K  W  K  G  L  K  N  S  T  T  N  S  V  T  V  T  G  L  N  A  N  T  S  Y  T  F  C  V
cgtgcgtggcaaaccgactggagcaagaaagatcacggcagcgttctgagcgataacgtgaccgttaagaccaaaaccgcggcgaccccg
 R  A  W  Q  T  D  W  S  K  K  D  H  G  S  V  L  S  D  N  V  T  V  K  T  K  T  A  A  T  P
gtggcggcgaacggtcgtctgagcgtgaagggcgcgaacatcgttaaccagagcggtcaaattttcaagatcaagggtatgagcacccac
 V  A  A  N  G  R  L  S  V  K  G  A  N  I  V  N  Q  S  G  Q  I  F  K  I  K  G  M  S  T  H
ggcattatgtgggaggactttcgtaacatcctgagctacaacagcctgaaggttctgcgtgacgattggaaatgcaacaccattcgtatc
 G  I  M  W  E  D  F  R  N  I  L  S  Y  N  S  L  K  V  L  R  D  D  W  K  C  N  T  I  R  I
gcgatgtataccgcggagtggggtggctacaccaccggtagcacctatgcggcgcaggcgaaggcgcgtgtgaaagcgggtgttgaaaac
 A  M  Y  T  A  E  W  G  G  Y  T  T  G  S  T  Y  A  A  Q  A  K  A  R  V  K  A  G  V  E  N
gcgaaaagcgcgggcatgtacgcgatcattgactggcacatcctgagcgacggtgatccgcgtacccaccagagccaagcggttgcgttc
 A  K  S  A  G  M  Y  A  I  I  D  W  H  I  L  S  D  G  D  P  R  T  H  Q  S  Q  A  V  A  F
tttaaggaaatggcgaacacctacaagaactacgataacgtgatctacgaaatctgcaacgaaccgaacggtggcgttacctggaacggt
 F  K  E  M  A  N  T  Y  K  N  Y  D  N  V  I  Y  E  I  C  N  E  P  N  G  G  V  T  W  N  G
ggcattaagagctactgccaggcggtggttaacgcgatccgtcaatatgataaaaacgcgatcattatctgcggcaccggcacctggagc
 G  I  K  S  Y  C  Q  A  V  V  N  A  I  R  Q  Y  D  K  N  A  I  I  I  C  G  T  G  T  W  S
caagacatcgataaggtgctgggtaaccgtctgagcgacaaaaactgcgtttacgcgctgcacttctatgcgaacacccacaccgattgg
 Q  D  I  D  K  V  L  G  N  R  L  S  D  K  N  C  V  Y  A  L  H  F  Y  A  N  T  H  T  D  W
ctgcgtagccgtctgaagagctgctacaacagcggcctgccggtgctggttagcgagttcggcacctgcgatgcgagcggtaacggtggc
 L  R  S  R  L  K  S  C  Y  N  S  G  L  P  V  L  V  S  E  F  G  T  C  D  A  S  G  N  G  G
tttaacaagtaccagacccaagaatggctgaagctgtgcgatagcctgaaagtgggttatattaactggagcgcggcgaacaaaagcgaa
 F  N  K  Y  Q  T  Q  E  W  L  K  L  C  D  S  L  K  V  G  Y  I  N  W  S  A  A  N  K  S  E
accgcgagcgcgtttaagaccggcaccgacctgaaaagcattaccgcgggcaccagccagctgaccgaaagcggcaaactggttcgtgat
 T  A  S  A  F  K  T  G  T  D  L  K  S  I  T  A  G  T  S  Q  L  T  E  S  G  K  L  V  R  D
                                    tggtatcgtgcgcac
                                     W  Y  R  A  H
```

Fig. 2. The *egc2* sequence coding for mature endoglucanase GH5 (EGC2) was mined in metegenomic DNA data from bacteria in goats rumen.

Accordingly, the SUMO gene from pETSUMO3 was amplified and inserted into *Nde*I-*Nco*I in pET22b(+) to generate pET22SUMO. And then, the *egc2* was inserted into pET22SUMO to generate pET22SUMOegc2. The obtained vector pET22egc1, pET22SUMOegc2 was transformed into *E. coli* Rossetta for the genes expression.

The recombinant *E. coli* Rosetta harboring pET22egc1 or pET22SUMOegc2 were inoculated overnight in 5 ml LB medium containing 100 µg/ml ampicilin (LBA) at 37°C, 200 rpm. The culture then was diluted into OD_{600} of 0,1 in LBA medium and continued incubating in the previous conditions until OD_{600} of 0.4 - 0.6 before inducing with IPTG (up to 0,5 mM) and inoculating at 25^{o}C, for 4 hours, 200 rpm. Afterward, the bacterial cultures were centrifuged at 6000 rpm for 10 min and the cells were diluted in PBS to OD_{600} of 10. The cells were disrupted by sonication at 20kHz (10 pulses, 30 s each at 100 W with 20 s intermission). Soluble protein was harvested by centrifugation and analized by SDS-PAGE (Laemmli, 1970).

For recombinant enzyme purification, the soluble fraction was applied to a 5 ml column containing Ni-NTA resin. Contaminant proteins in the column was washed with 10 column volumes of 50 mM phosphate buffer, pH 6.8 containing 0.1 M imidazole. The target proteins were eluted in 50 mM phosphate buffer pH 6.8 containing 0.5 M imidazole. The elute fractions of 2 ml were collected and were analyzed in polyacrylamide gel 12.6% by SDS-PAGE. The final purified protein solution was desalted using a PD-10 ultrafiltration column (GE Healthcare, Mississauga, ON, Canada) by gravity flow according to the manufacturer's direction. The desalted enzymes were used for determination of endoglucanase activity.

E. Determination of endoglucanase activity

Endoglucanase activity was measured by incubating 0,075 ml of assay mixture containing 0.025 ml of 1% (w/v) of carboxymethyl cellulose (CMC) in PBS pH6 (1 L of PBS contained 8 g NaCl; 0.2 g KCl; 0.24 g KH_2PO_4; 1.42 g Na_2HPO_4, adjusted to pH 6) and 0,045 ml of diluted enzyme in PBS for 60 min at 40^{o}C. Reducing sugar content released from the reaction was measured by DNS method described by Miller (1959) using glucose as standard sugar.

Accordingly, the reaction was stoped by adding 0,075 ml of DNS reagen (1% DNS; 1% NaOH; 18,2% K-Na-tartrate) and boiled 15 min. Afterward, 25 µl K-Na-Tartrate 40% was added and kept for cool at room temperature and measured at 540 nm by ELISA reader. The enzyme activity (CMCase) is expressed as amount of reducing sugar released equal to mM of glucose.

F. Effect of pH and temperature on endoglucanase activity and stability

The effect of pH and temperature was evaluated in the range of pH 2.0–9.0 (pH interval 1.0) and temperature of 40–60°C (temperature interval 10°C) by performing the CMCase activity assay. Buffer PBS was used for pH profile determination using HCl or NaOH to adjust to desired pH. Thermal stabilities were determined by pre-incubating the enzymes temperatures: 40°C, 50°C and 60°C, at optimal pH for up to 90 min and the CMCase activity was measured.

RESULTS AND DISCUSSION

A. Prediction of the Tm and pH for enzymes activity

Based on amino acid sequences, EGC1 was predicted to be acidic enzyme and stable in temperature range of 55-65^{o}C (Table 1). The EGC2 also was predicted to be acidic enzyme and stable at the temperature lower than 55^{o}C. After fusion with SUMO, temperature stability of enzyme was not improved and still was acidic enzyme (Table 1).

Table 1: Temperature stability and active pH of enzymes were predicted by AcalPred and Tm prediction tools.

Protein	pI	Mw (kDa)	Tm	Acid	Alkaline
EGC1	6.95	38.980	55-65°C	0.9700	0.0300
EGC2	9.58	56.525	< 55°C	0.9238	0.0762
SUMOEGC2	9.37	69.478	< 55°C	0.946119	0.053881

B. Expression and purification of recombinant endoglucanase in E. coli

After 5 hours of culturing in LBA containing inducer IPTG at 25^{o}C, EGC1 of ~40 kDa and SUMOEGC2 of 69 kDa were expressed in *E. coli* BL21 (DE3) (Fig. 3). However, a half of EGC1 was in soluble fraction and a half was insoluble. Almost SUMOEGC2 was soluble (Fig. 3). We had inserted *egc2* directly into pET22b(+) and express in *E. coli* BL21 (DE3), but EGC2 was not expressed. Thus the fusion of *egc2* with SUMO really improved the expression ability of *egc2* in *E. coli*. EGC1 and SUMOEGC2 in soluble fractions were purifed by His-tag affinity chromatography (Fig. 4). These enzymes were desalted for enzyme activity measurement.

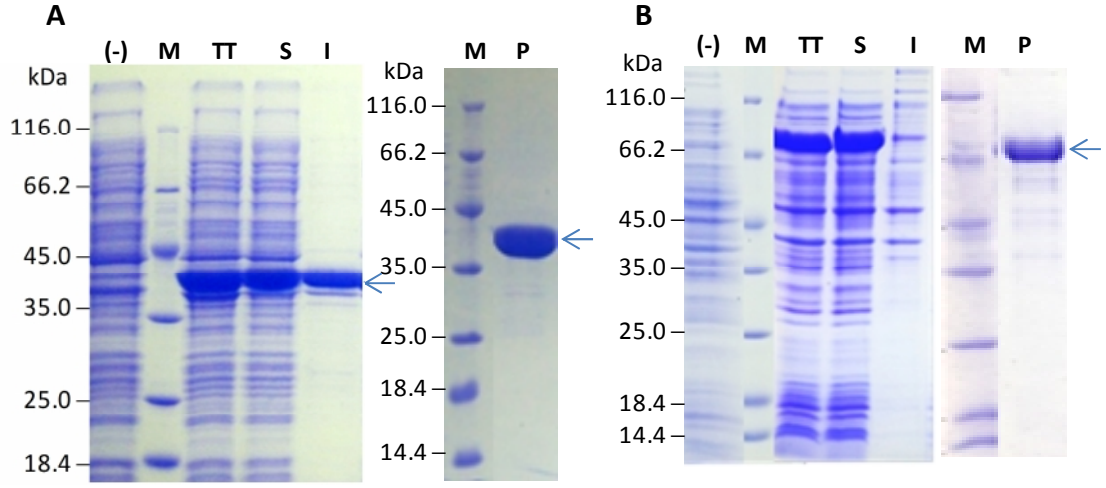

Fig. 3. SDS-PAGE analysis of proteins expressed in recombinant *E. coli* cells harboring pET22egc1 (A), pETSUMOegc2 (B) and purified EGC1, SUMOEGC2. (M: Protein marker (Thermo scientific); (-): total protein from *E. coli* BL21(DE3) harboring pET22b(+); TT: Total protein; S: Soluble fraction; I: Insoluble fraction; P: Purified EGC1 or SUMOEGC2; arrows indicate the target proteins).

C. The effect of temperature on enzyme activity

As the prediction results, Tm of EGC1 was 55-65^{o}C and SUMOEGC2 was lower than 55^{o}C. Thus in this experiment, the range of temperature between 30^{o}C and 55^{o}C were selected for enzymes activity. As a result, both enzymes expressed CMCase in the range of investigated temperatures. From 30°C to 40°C, the activity of two endoglucanses increased to reach the highest activities at 40^{o}C then gradually decreased at 45, 50, 55^{o}C. At 55^{o}C, enzymes remained 50% activities. Thus experimental result agreed with the predicted result. Tm is melting temperature of enzymes, reflects the stability of enzymes in temperature. In theory, the active temperature and optimal temperature of enzymes is lower than Tm. In this study, optimal temperatures of EGC1 and SUMOEGC2 were 40^{o}C, lower than predicted Tm. However, EGC1 was expected to have optimal temperature higher than SUMOEGC2 because the predicted Tm of EGC1 was higher than SUMOEGC2 but they have the same optimal temperature (Fig. 4). The result from experiment checking the temperature stability of EGC1, SUMOEGC2 showed that EGC1 was more stable than SUMOEGC2 (Fig. 5). This result agreed with the predicted Tm of EGC1, SUMOEGC2 by the Tm predion tool. EGC1 was stable at the temperature up to 50^{o}C and reduced at 60^{o}C. But SUMOEGC2 activity reduced steadily at 50^{o}C and rapidly at 60^{o}C. In this experiment we also found that temperature have ability to activate EGC1, thus the incubation of enzyme at 50^{o}C encreased enzyme activity.

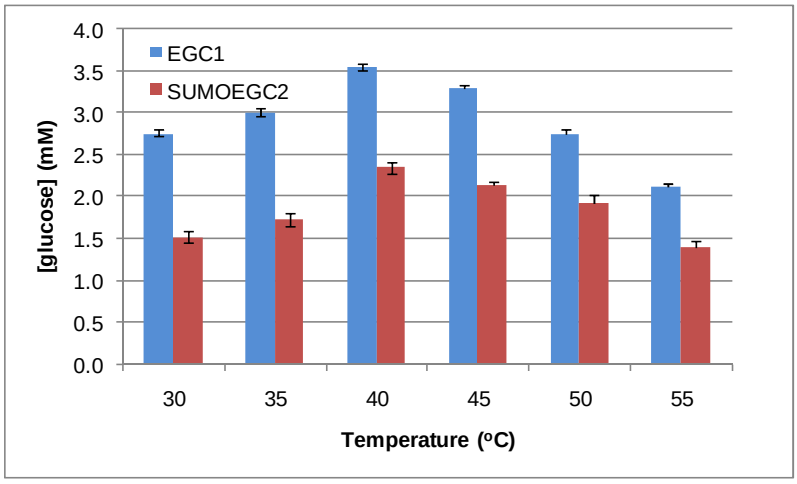

Fig. 4. The effect of temperatue on enzymes activity.

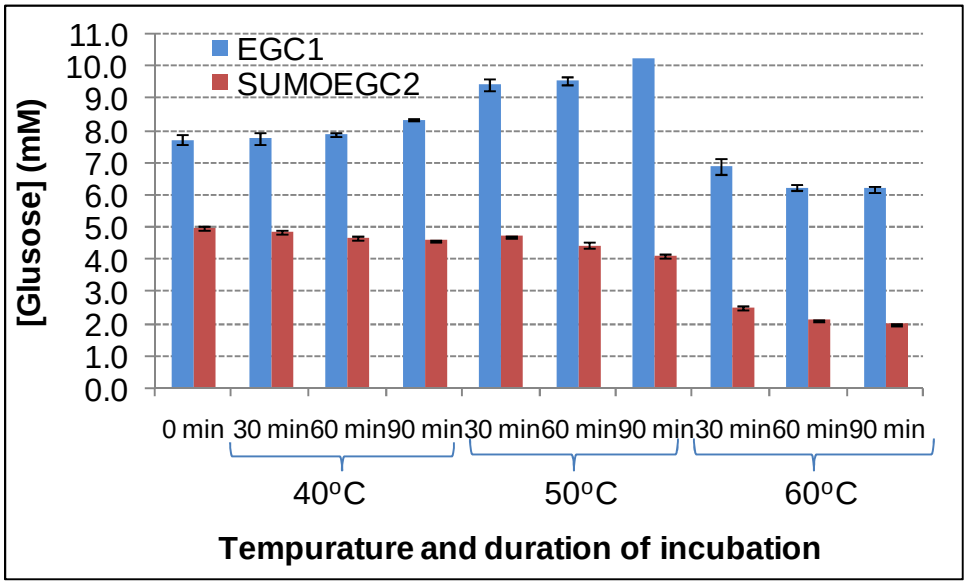

Fig. 5. Temperature stability of enzymes.

The activation of enzyme by temperature is also seen in the StayBrite™ Highly Stable Luciferase (BioVision) or in the experiment measuring optimal pH and temperature of recombinant endoglucanase expressed in *Bacillus subtilis* (Zafar *et al.*, 2014).

C. The effect of pH on enzyme activity

In this study, eight pH values were selected from the range of pH from 2.0 to 9.0 to investigate the active pH values of EGC1 and SUMOEGC2 protein. In aggreement with the predicted results by AcalPred tool, both EGC1 and SUMOEGC2 were acidic enzymes. Optimal pH of EGC1 and SUMOEGC2 was 6 and 4 respectively. But active pH of EGC1 trended toward neutral. EGC1 activity was impacted by at acidic condition but lightly reduced from pH 6 to 9. Whereas, active pH range of SUMOEGC2 was large from 2-9 but optimal pH was at acidic (Fig. 6).

The predicted results showed that the acidic scores of EGC1, SUMOEGC2 were 0.97 and 0.94 respectively. However, the experiment result showed that SUMOEGC2 was more acidic than EGC1 and more stable in large pH range than EGC1. This result maybe explained that the difference in the scores is not large enough to distinguish which enzyme is more acidic or alkaline.

AcalPred program was developed for predicting pH values that would be optimal for enzyme functioning based on its sequence, but was successfully applied for pH-sensitive receptors (Deyev *et al.*, 2015). It have limited researches investigating the feasibility of application of AcalPred and Tm prediction tools for estimating active pH and temperature for enzymes, thus this study is valuable on the application of bioinformatic tools on prediction of enzyme properties.

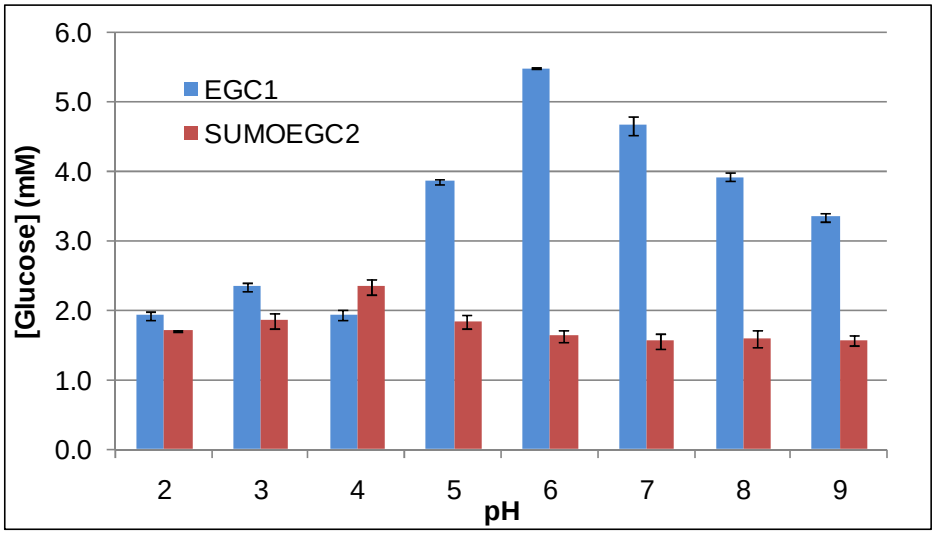

Fig. 6. The effect of pH on enzymes activity

CONCLUSION

In this study, the AcalPred and the Tm prediction tool expressed to be the powerful tools for prediction of active pH and temperature stability of endoglucanases based on the sequences mined from metagenomic DNA data. The results from bioinformatic prediction agreed with experiment results.

Acknowledgement

This study was carried out with the financial support of the Project code VAST02.05/18-19 "Investigation of the role of the unknown functional region in module structure of cellulase" and equipments of National Key Laboratory of Gene Technology, Institute of Biotechnology, VAST, Vietnam.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

Adrio, J.L. and Demain, A.L. (2010). Recombinant organisms for production of industrial products. *Bioeng. Bugs.*, **1**(2): 116–131.

Almaabadi, A., Gojobori, T., and Mineta, K. (2015). Marine metagenome as a resource for novel enzymes. *Genom. Proteom. Bioinf.*, **13**(5): 290–295.

Antonov, E., Schlembach, I., Regestein, L., Rosenbaum, M. A., & Büchs, J. (2017). Process relevant screening of cellulolytic organisms for consolidated bioprocessing. *Biotechnology for biofuels*, **10**(1), 106.

Do, T.H., Le, N.G., Dao, T.K., Nguyen, T.M.P., Le, T.L., Luu, H.L., Nguyen, K.H.V., Nguyen, V.L., Le, L.A., Phung, T.N., van Straalen, N.M., Roelofs, D., and Truong, N.H. (2018). Metagenomic insights into lignocellulose-degrading genes through Illumina-based de novo sequencing of the microbiome in Vietnamese native goats' rumen. *J. Gen. Appl. Microbiol.*, **64**(3): 108–116.

Do, T.H., Nguyen, T.T., Nguyen, T.N., Le, Q.G., Nguyen, C., Kimura, K., and Truong, N.H. (2014). Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. *J. Biosci. Bioeng.*, **118**(6): 665–671.

Ebrahimi, M. and Ebrahimie, E. (2010). Sequence-based prediction of enzyme thermostability through bioinformatics algorithms. *Curr. Bioinform.,* **5**(3): 195-203.

Deyev, I.E., Popova, N.V., Petrenko, A.G. (2015). Determination of alkali-sensing parts of the insulin receptor-related receptor using the bioinformatic approach. *Acta Naturae,* **7**(2): 80-86.

Kennedy, J., Margassery, L.M., Morrissey, J.P., O'Gara, F., & Dobson, A.D. (2013). Metagenomic strategies for the discovery of novel enzymes with biotechnological application from marine ecosystems. In *Marine Enzymes for Biocatalysis*(pp. 109-130). Woodhead Publishing.

Khare, S.K., Pandey, A., and Larroche, C. (2015). Current perspectives in enzymatic saccharification of lignocellulosic biomass. *Biochem. Engin. J.*, **102**(15): 38-44.

Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B.A., and Blanch, H.W. (2012). The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol. Bioeng.*, **109**(4): 1083-1087.

Koeck, D.E., Pechtl, A., Zverlov, V.V., and Schwarz, W.H. (2014). Genomics of cellulolytic bacteria. *Curr. Opin. Biotechnol.*, **29**: 171-83.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227**(5259): 680-685.

Lin, H., Chen, W., and Ding, H. (2013). AcalPred: a sequence-based tool for discriminating between acidic and alkaline enzymes. *PLOS ONE*, **8**(10): e75726.

Matsumoto, N.M., González-Toro, D.C., Chacko, R.T., Maynard, H.D., and Thayumanavan, S. (2013). Synthesis of nanogel-protein conjugates. *Polym. Chem.,* **4**(8): 2464-2469.

Miller, G.L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.,* **31** (3): 426–428.

Percival Zhang, Y.H., Himmel, M.E., and Mielenz, J.R. (2006). Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol. Adv.,* **24**(5): 452-81.

Rochat, T., Pérez-Pascual, D., Nilsen, H., Carpentier, M., Bridel, S., Bernardet, J., and Duchaud, E. (2019). Identification of a novel elastin-degrading enzyme from the fish pathogen *Flavobacterium psychrophilum. Appl. Environ. Microbiol.,* **85**(6): e02535-18.

Singh, R., Kumar, M., Mittal, A., and Mehta, P.K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech,* **6**(2): 174.

Zafar, M., Ahmed, S., Khan, M.I.M., and Jamil, A. (2014). Recombinant expression and characterization of a novel endoglucanase from *Bacillus subtilis* in *Escherichia coli. Mol. Biol. Rep.*, **41**(5): 3295–3302.

Zverlov, V. V. and Schwarz, W.H. (2008). Bacterial cellulose hydrolysis in anaerobic environmental subsystems - *Clostridium thermocellum* and *Clostridium stercorarium*, thermophilic plant-fiber degraders. *Ann. N Y Acad. Sci.,* **1125**: 298-307.