1874-2858/22

Send Orders for Reprints to reprints@benthamscience.net

The Open Microbiology Journal

Content list available at: https://openmicrobiologyjournal.com

RESEARCH ARTICLE

Expression and Characterization of a Small, Xylan/Cellulose-degrading GH43 Protein Derived from Biofertilizer Metagenome

Atcha Oraintara[1,2,*] and Pitak Bhunaonin[1]

[1]*Department of Microbiology, Faculty of Science, Khon Kean University, Khon Kaen, Thailand*
[2]*Protein and Proteomics Research Center for Commercial and Industrial Purposes (ProCCI), Khon Kaen University, Khon Kaen, Thailand*

Abstract:

Background:

A putative glycosyl hydrolase gene *biof1_09* was identified from a metagenomic fosmid library of local biofertilizers in previous report [1]. The gene is renamed as *gh43kk* in this study.

Methods:

The gene *gh43kk*, encoding a putative β-D-xylosidase was amplified by polymerase chain reaction (PCR) and successfully cloned and expressed in *Escherichia coli*. The expressed recombinant protein was purified by metal affinity chromatography. Its properties were initially verified by enzyme assay and thin layer chromatography (TLC).

Results:

The purified recombinant protein showed the highest catalytic activities at acidic pH 4 and 50°C toward beechwood xylan, followed by carboxymethylcellulose (CMC). TLC analysis indicated a release of xylose and glucose when xylan and CMC were treated with Gh43kk protein, respectively, whereas glucose and cellobiose were detected when avicel, cellulose and filter paper were used as substrates, suggesting its dual function as xylanase with cellulase activity. The enzyme indicated great stability in a temperature between 10 to 50 °C and a wide range of pH from 4 to 8. Enzyme activity of Gh43kk was enhanced in the presence of magnesium and manganese ions, while calcium ions, Ethylenediaminetetraacetic acid (EDTA) and sodium dodecyl sulfate (SDS) inhibited the enzyme activity.

Conclusion:

These results suggest that Gh43kk could be a potential candidate for application in various bioconversion processes.

Keywords: Cellulase, GH43, Metagenome, Metagenomics library, Recombinant protein, Xylanase.

Article History	Received: May 11, 2022	Revised: September 8, 2022	Accepted: September 15, 2022

1. INTRODUCTION

Lignocellulosic material is one of the potential sources for biomass energy production. It is the main basic component found in all plant cell walls and is mostly composed of three polymers: cellulose, hemicellulose and lignin, together with small amounts of other components, like acetyl groups, minerals and phenolic substituents [2]. Cellulose is a linear, unbranched homopolysaccharide of 10,000 to 15,000 D-glucose units connected by (β1→4) glycosidic linkage [3]. Hemicelluloses are heterogeneous polymers of pentoses (xylose and L-arabinose), hexoses (mannose, glucose and galactose), and sugar acids [2]. Lignin mostly consists of three different monomers, *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) arranged in branches with no constant organization [4].

In order to effectively utilize lignocellulosic resources as a substrate, cellulose and hemicellulose have to be broken down *via* hydrolysis to release sugar monomers, such as glucose and xylose. The conversion of lignocellulose to sugars could be accomplished by physical, chemical or enzymatic hydrolysis. However, high temperature and strong acids are mostly involved in the physical and chemical hydrolysis approaches, by which expensive, corrosion-resistant equipment are needed. Compared with those mentioned methods, the cost of enzymatic hydrolysis is rather low compared to acid or alkaline

* Address correspondence to this author at the Department of Microbiology, Faculty of Science, Khon Kean University, Khon Kaen, Thailand; E-mail:atcha@kku.ac.th

hydrolysis because enzyme hydrolysis is usually conducted at mild conditions (pH 4.8 and temperature 45–50°C) and does not have a corrosion problem [5, 6].

Cellulolytic enzymes found in nature are produced by bacteria, fungi, protozoans, plants and animals. However, the majority of environmental microorganisms, with an estimation of about 85-99%, cannot be cultured by conventional laboratory cultivation methods [7]. In order to avoid the limitation of culturing method, the metagenomic approach has been introduced. Metagenomics is the genomic analysis of microorganisms by direct extraction and cloning of environmental DNA [8]. This method leads to the discovery of many novel cellulolytic enzymes such as *RmCel12A* glycoside hydrolase family 12 cellulase (LC-CelA) [9] and *Cel5R,* a novel halostable endoglucanase obtained from soil metagenomic library [10], *Zfyn184* a cellulase/ hemicellulase belonging to glycoside hydrolase family 44 [11].

In our previous work, we reported a clone from metagenomics library constructs showing activities toward both cellulose and xylan [1]. After sequencing the positive library construct, an open reading frame of a glycosyl hydrolase gene was predicted. The amino acid sequence of the translated nucleotides shared similarity with the conserved domain of glycosyl hydrolase family 43 (GH43) [1]. In this work, the above mentioned open reading frame, renamed as *gh43kk* gene, was cloned, expressed, and the encoded protein was purified, and its biochemical properties were characterized.

2. MATERIALS AND METHODS

2.1. Cloning of *gh43kk* Gene

The *gh43kk* gene was obtained from a nucleotide sequence of a metagenomic subclone Biof1_09 (Genbank JQ581599 [1]. The gene was amplified by polymerase chain reaction (PCR) using specific primers containing *Bam*HI and *Eco*RI restriction site (5'-GGGCGCGGATCCATGTACGCCAT GAAAAA CATTTTTAAATATG-3') and (5'-GGGCCGCCGGAATTC TGTAGAAGTGCCACCTTTCAAC -3'). The amplification protocol was according to the *Taq* DNA polymerase manual (Thermo Scienctific, USA), starting with an initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 95°C for 30 s, annealing at 42 °C for 1 min 30 s and extension at 72°C for 1 min 30 s, final extension at 72°C for 10 min. The PCR product was purified by QIAquick® PCR Purification Kit (Qiagen, Germany) and was digested with FastDigest *Bam*HI and *Eco*RI (Thermo Scienctific, USA). Afterward, the *gh43kk* fragment was inserted into pET28a(+) expression vector (Novagen, USA) by T4 DNA ligase (New England Biolabs, USA). The recombinant expression vector (pET/*gh43kk*) was transformed into *E. coli* BL21 by heat-shock technique. The transformants were selected by colony PCR and then also confirmed by sequencing (Macrogen, Korea).

2.2. Bioinformatics Analysis and Three-dimensional Structure Modeling

The molecular weight and isoelectric point of the translated nucleotides of the gene *gh43kk* were calculated by Expasy software (https://web.expasy.org/compute.pi/). The patterns of amino acid residue conservation and divergence in a protein family were analyzed by searching in the Conserved Domain Database (CDD, https://www.ncbi.nlm.nih.gov/Structure/cdd /cdd.shtml).

A three-dimensional model of Gh43kk was generated by RaptorX (http://raptorx.uchicago.edu/StructurePrediction/ myjobs/62894490_578596/) with a score of 32. The P-value for the relative global quality was 0.0365. The GDT (global distance test) and uGDT (un-normalized GDT) values of the model were 30 and 33, respectively.

2.3. Expression and Purification of Gh43kk Protein

The expression of the recombinant protein was done under low temperature as follows: OD_{600} of overnight bacterial culture was adjusted to 0.2 and the incubation was continued at 37°C until the OD_{600} reached 0.4. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM, and the cell suspension was further incubated for 20 h at 16°C. Lysates were prepared from 100 ml *E. coli* culture. Bacterial cells were harvested by centrifugation at the speed of 7,500 rpm, 4 °C for 10 min. Each gram of the resulting pellet was lysed in 5 ml of 50 mM sodium acetate pH 5, 300 mM NaCl, 20 mM imidazole and 90% (v/v) B-PER™ Bacterial Protein Extraction Reagent (Thermo Scientific, USA). Lysozyme (final concentration 1 mg/ml (w/v) and cOmplete™ Protease Inhibitor Cocktail (Sigma-Aldrich, USA) was further added to the lysate, then incubated at room temperature for 30 min prior to sonication. Sonication (Vibra-Cell™, Sonics, was done at 35% amplitude (pulse duration of 5 s on and 10 s off) on ice. The lysate was centrifuged at 18,000 g for 30 min at 4°C and the supernatant was sterile filtered through 0.22 µm membrane filter. Protino® Ni-NTA Agarose (Macherey Nagel, Germany) was used for purification according to the manufacturer's instruction. Briefly, 5 ml of HisPur™ Ni-NTA Resin (Thermo Fisher Scientific, USA) was loaded into an empty gravity flow column with filter frits (Protino® column 14 ml, Macherey-Nagel, Germany). The agarose was equilibrated with lysis buffer. The clear lysate was then added into the column, and incubated with the Ni-NTA agarose with gentle shaking on an orbital shaker for 60 min at 4°C. After letting the lysate flow through the column, the agarose was washed twice with 50 mM sodium acetate pH 5, 300 mM NaCl, and 30 mM imidazole. The recombinant protein was eluted with 50 mM sodium acetate pH 5, 300 mM NaCl, and 300 mM imidazole. The eluate was then brought through the Vivaspin® 6 (cut off 10 kDa; GE Healthcare) to change the buffer to 20 mM Sodium acetate buffer pH 5.4. The purified protein was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using 13.5% acrylamide-separating gel. Western blot analysis using anti-his tag monoclonal antibody was done to confirm the recombinant protein. Concentration and yields of the purified protein were calculated by Bradford protein assay. Purity and recovery of the recombinant proteins were estimated from protein band intensity in SDS-PAGE quantified by ImageJ software (National Institutes of Health, USA).

2.4. Agar Well Diffusion Test

Azurine cross-linked cellulose (AZCL-HE-cellulose) (Megazyme, Ireland) or beechwood xylan (Megazyme, Ireland) was mixed with molten sterile agar (prepared with 50 mM sodium acetate buffer pH 5) to a final concentration of 1% (w/v). Various amounts of the purified protein, as well as the commercial cellulose mixture, Celluclast™ (Novozymes, Denmark) and Cellubrix® (Novozymes, Denmark) as positive control and protein elution buffer (0.1 M sodium acetate buffer pH 5) containing imidazole as negative control were dropped into agar wells. After incubation at 37°C for 2 days, the plate containing beechwood xylan was stained with 0.33% Gram's iodine and clear zones around the agar wells were observed. When the plate containing AZCL-HE-cellulose was used, smooth blue zones around the agar well indicated the activity of cellulose-degrading enzymes.

2.5. Enzyme Activity Assay and Kinetics

Enzyme activity was determined by incubating substrates with the purified enzyme in 100 mM sodium acetate buffer pH 5 at 50°C for 30 min, which will be used as a standard condition for the rest of the enzyme activity experiment. Reducing sugars were assayed by the DNS method with glucose as a standard. The concentration of pulverized substrates (avicel, beechwood xylan, cellobiose, cellulose, carboxymethyl cellulose) was 1% (w/v), while the size of filter paper was 0.5×1.5 cm. Negative controls were done by incubating the substrates without the enzyme (using elution buffer instead). The values of colorimetric detection (in reducing sugars) of controls were subtracted from the results of the actual reactions. All the experiments were done in three replicates. One unit of enzyme activity was defined as the amount of enzyme that releases 1 µmol of glucose equivalents per minute. Specific activity was exhibited in the unit of activity enzyme in one mg of protein. Protein concentration was determined by using the method as described by Bradford [12] with BSA as a standard.

The kinetic studies were determined using beechwood xylan concentrations from 0.5 to 5 mM, respectively. Reducing sugars were measured after the reactions were incubated at 50°C for 30 min. The Michaelis constant (K_m) and maximum velocity (V_{max}) were determined by means of Lineweaver–Burk plots [13]

2.6. Thin Layer Chromatography (TLC)

TLC was performed as modified by Zhou and coworkers [14]. One percent (w/v) of substrates were incubated with Gh43kk protein under optimal conditions. Subsequently, aliquots of 2 µl each were spotted on thin layer chromatography (TLC) silica gel 60 plates (Merck, USA), along with standards of glucose, cellobiose, xylose, and arabinose, respectively. Separation was carried out by TLC with a mobile phase (ethanol/1-butanol/water at a ratio of 3:5:2) at room temperature. Sugars were detected by spraying the plates with ethanol and sulfuric acid (ratio 95%:5% v/v), followed by heating at 100°C for 10 min. The experiment was done in triplicates. Relative rates of flow (R_f) were calculated as $R_f = X/ Y$, by which X = migration distance of the sample; and Y = migration distance of the solvent.

2.7. Effect of Temperature, pH, Metal Ions and other chemicals on the Recombinant Enzyme

The optimal temperature for enzyme activity was performed at various temperatures ranging from 10 to 90°C. To determine temperature stability of the enzyme, the purified enzyme was incubated at different temperatures for 30 min prior to the enzyme activity assay. Determination of optimum pH was done in 100 mM sodium acetate buffer (pH 3-6), 100 mM phosphate buffer (pH 6-8) and 100 mM tris-HCl buffer (pH 8-10). For the test of pH stability, the protein was pre-incubated in buffers with different pH for 1 h at 4°C, then the activity was assayed under the same condition as mentioned in Enzyme activity assay section. The effects of metal ions compounds (Ba^{2+} Ca^{2+} $Co^{2+}Cu^{2+}$, Fe^{2+}, Li^{+}, Mg^{2+}, Mn^{2+}, Ni^{2+}, Sr^{2+} and Zn^{2+}, salts (KCl, NaCl), chelating agent (Ethylenediaminetetraacetic acid; EDTA), and 1% (w/v) detergent (sodium dodecyl sulfate; SDS) were determined under standard condition. The concentrations of metal ions, salts and EDTA were 5 mM. Control was buffer without addition of any tested chemicals or metal ions. Beechwood xylan was used as substrate for all experiments.

2.8. Statistical Analysis

The data was analyzed by using the two-sample t-test (Satterthwaite's method), by which a p-value of ≤ 0.05 is considered significant.

3. RESULTS

3.1. Bioinformatic Analysis and Modular Structure of the Gh43kk Protein

The gene *gh43kk* contains putative nucleotides of 330 bp, which encode a polypeptide of 110 amino acid residues with a theoretical molecular mass of 12.81 kDa and an isoelectric point of 4.5. The protein Gh43kk shared conserved regions within the amino acid sequences of the glycosyl hydrolase family 43. Among those are amino acid residues which were reported to be found on the active site of the enzyme, namely aspartic acid (Asp43), glutamine (Gln93), glycine (Gly107), and Threonine (Thr108) (Graciano 2012) The multiple amino acid sequence alignment in Fig. (**1**) shows that these key residues are well conserved in Gh43kk protein.

As shown in the three-dimensional model (Fig. **2**), the N-terminal region of Gh43kk displays a β-meander motif containing 4 consecutive antiparallel β-strands linked together by hairpin loops. The C-terminal region shows a v-shaped helix-break-helix fold [15]. The conserved catalytic aspartate (Asp43) is located in the hairpin loops linking the first pair of the antiparallel β-sheets.

```
10                  --------------myamkinfkydketfllidndvlkpddqgnyd-----ipdgwtdis
CBK93970.1          drpsqlcveneeisrgifaptiryhegtfymittnvshg---gnfivtaqdpagpwsdpy
ACF39706.1          drpsqlnvadrkssdgiyaptlryhdgtfymittdvrgi---gnfyvtatnpagpwsdpi
EHB68557.1          drdsqldlrarrssdgiyaptlryhngtfymittdvygi---gnfyvttqdpagpwsdpi
                                          ::*.. ** :* .:*     **:       . *:*

10                  f--------dpglylpkfypkekvwketatqeyieslqpsete----------tneldll
CBK93970.1          ylgeeavgidpslffd----d-------dgrcyycgtrpnpegvryngdweiwiqeldlg
ACF39706.1          lvp--yggidpslffd----d-------dgkvyvtaqq----gadydsh--aiqyeidia
EHB68557.1          lvp--yggidpslffd----d-------dgkvyvtaqq----gadydsh--aiqyeidia
                            **.*::     .        : *  . :                 *:*:

10                  kkqn---avl----------------syqiaqlqkqieqlkggtst--------------
CBK93970.1          tmklkgrsmaiwkgavkgciwpegphiykingyy-ylmhaeggtgpehsitiarstelfk
ACF39706.1          tgkalsepvviwrgd--ggpwtegphlykiggky-yimtasggtakehreiigrssspyg
EHB68557.1          tgkalsepvviwrgd--ggpwtegphlykingmy-yimsasggtakdhreiigrsaspyg
                    . :     :                 *:*      :   .***.
```

Fig. (1). Multiple amino acid sequence alignment of GH43KK and other similar enzymes revealing the conserved regions within the amino acid sequences of glycosyl hydrolase family 43. Significant identity or similarity of amino acid residues is marked with an asterisk (similar), colon (highly identical), or dot (moderately identical). Amino acid residues, which were reported to be found on the active site, are indicated with red arrows. GH43KK was aligned with beta-xylosidase/alpha-L-arabinofuranosidase from following microorganisms: *Eubacterium rectale* M104/1 (CBK93970.1), *Caulobacter crescentus* NA1000 (ACF39706.1), *Paenibacillus lactis* 154 (EHB68557.1)

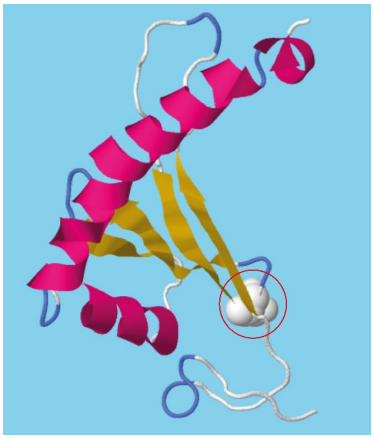


Fig. (2). Overall structure of the putative Gh43kk. The N-terminal, yellow colored antiparallel β-strands were linked together by hairpin loops. The catalytic amino acid, aspartate (Asp43) residue, is located at the hairpin loop of the peripheral β-sheet (red circle). The helical C-terminal region is shown in pink.

3.2. Cloning, Expression and Purification of the Recombinant Gh43kk Protein

The gene *gh43kk* was successfully cloned, and protein expression was optimized. The recombinant Gh43kk protein was effectively overexpressed at 37°C by using IPTG as an inducer, however, a large amount of the expressed protein remained as an inclusion body (Fig. **S1**), which obstructs the purification. The recombinant protein was soluble when the temperature was shifted to 16°C, therefore, the recombinant protein Gh43kk was successfully expressed and purified at that temperature (Fig. **3**).

Fig. (**3**) shows the purification step with the purified protein band of about 23 kDa. The purified recombinant protein was more than 95% pure, with around 75% recovery. Approximately 10 milligrams of recombinant protein was obtained from 1 liter of bacterial expression culture (1-1.5 g wet weight). The recombinant Gh43kk was confirmed by Western blot analysis and a protein band of about 23 kDa was visualized (Fig. **3B**).

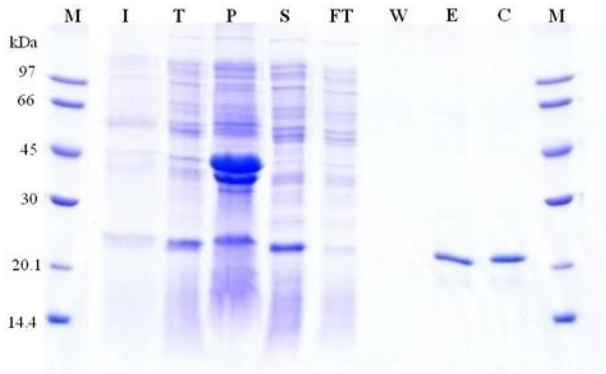


Fig. (3). SDS-PAGE showing purification step of the recombinant protein (**A**) I = uninduced protein; T = total cell lysate after induction; P = pellet; S = supernatant; Ft = flow-through; W = washing; E = elution; C = concentrated protein by Vivaspin® column; M = protein standard maker.

3.3. Substrate Specificity of the Purified Recombinant Protein Gh43kk

Activities of the purified Gh43kk protein were detected both on agar and in liquid suspension using various substrates. Smooth blue-colored zone or clear zones were observed on agar plates containing AZCL-Cellulose and xylan as substrates, respectively (Fig. **4**). The same result was observed by spectrophotometric experiments, by which the enzyme showed the highest hydrolytic activity towards the beechwood xylan with a specific activity of 286.37 ± 7.37 mU·mg^{-1} protein followed by CMC with 129.78 ± 7.91 mU·mg^{-1} protein, respectively (Table **1**). The K_m for Gh43kk on beechwood xylan was 0.153 µM, while the Turnover number (K_{cat}) was

0.0213, and Kcat/Km of the enzyme was 0.143.

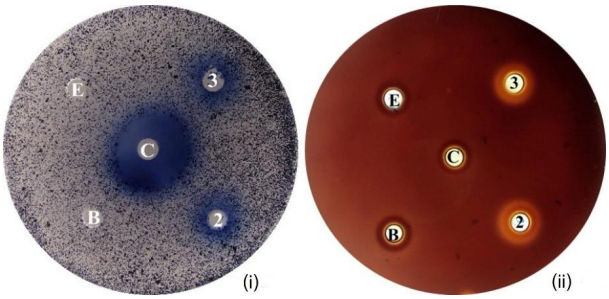


Fig. (4). Agar well diffusion test of the purified recombinant protein Gh43kk on AZCL-Cellulose **(i)** and beechwood xylan **(ii)** agar plates. **B** = 0.1 M sodium acetate buffer pH 5, **C** = mixture of diluted Celluclast™ and Cellubrix®, **E** = 0.1 M sodium acetate buffer pH 5 with 300 mM imidazole, **2** = 17.27 mg purified protein and **3** = 25.91 mg purified protein.

Thin Layer Chromatography (TLC) analysis of hydrolysate of different substrates treated by the recombinant protein Gh43kk revealed several patterns of sugar products (Fig. **5**). Both glucose and cellobiose were detected when avicel, cellulose and filter paper were hydrolyzed with Gh43KK. Only glucose was observed when CMC and rice straw were degraded with the recombinant enzyme. Gh43KK also exhibits its xylanase activity since xylan could be hydrolyzed as well, yielding xylose as a product.

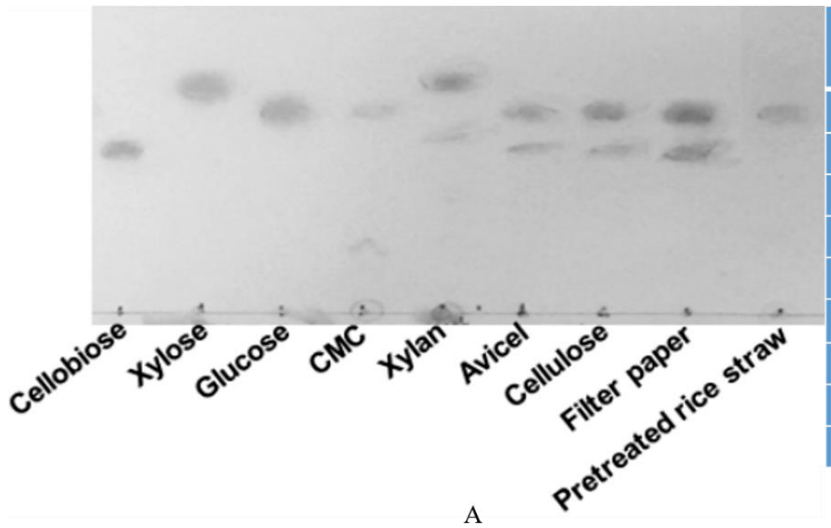
3.4. Effect of Temperature, pH, Metal Ion, and other Chemicals on Activity and Stability of the Recombinant Protein Gh43kk

The purified recombinant Gh43kk could work well (with an activity of more than 80%) in pH ranging from 4 to 8, with an optimum pH of 4 (Fig. **6A**). However, the enzyme activity of about 60% was still observed at pH 3 and 9, suggesting its capability in working at a very wide range of pH (Fig. **6A**). The enzyme maintained its activity of more than 60% when pre-incubated overnight in buffer with pH ranged from 4 to 9, where pH 5 to 7 was shown to have the least effect on its activity (Fig. **6B**). The recombinant protein Gh43kk showed increased activity when the temperature rose toward 50°C, which was also the optimum temperature for the enzyme (Fig. **6C**). Its activity decreased rapidly when the temperature exceeded 50°C (Fig. **6C**). The recombinant protein maintained more than 80% of its activity after pre-incubation at the temperature under 50°C (Fig. **6D**).

The highest activity (with an activity of about 120%) of the recombinant enzyme was observed when manganese ions were present. Slightly higher activity of the enzyme was also observed with the presence of magnesium ion, however it did not significantly enhance the activity of the enzyme (p-value of 0.21). Moderate inhibition (activity of about 76-79%) of the enzyme Gh43kk was observed when copper, barium, potassium and sodium ions were present. Metal ions of zinc, calcium and strontium, and sodium dodecyl sulfate (SDS) were shown to strongly inhibit the activity of the recombinant enzyme. Gh43kk protein almost lost its activity when incubated with Ethylenediaminetetraacetic acid (EDTA) (with an activity of about 4.68%) (Table **2**).

Table 1. Substrate specificity of the purified recombinant Gh43kk.

Substrate	Specific Activity (mU-mg^{-1})	Relative Activity (%)
Avicel	0.73± 0.20	0.25
Beechwood xylan	286.37± 7.37	100.00
CMC	129.78 ± 7.91	45.00
Cellobiose	17.37± 1.86	6.07
Cellulose	3.80± 0.94	1.33
Filter paper	0.31± 0.08	0.11



Substrates	Rf value	
	Rf_1	Rf_2
Cellobiose	0.594 ± 0.016	
Xylose	0.842 ± 0.022	
Glucose	0.743 ± 0.023	
CMC	0.743 ± 0.020	
Xylan	0.842 ± 0.021	0.644 ± 0.017
Avicel	0.743 ± 0.024	0.604 ± 0.021
Cellulose	0.743 ± 0.037	0.594 ± 0.016
Filter paper	0.733 ± 0.028	0.594 ± 0.022
Rice straw	0.743 ± 0.029	

B

Fig. (5). (A) Thin Layer Chromatography of hydrolysis result after treatment of different substrates with the recombinant protein Gh43kk under optimal condition (**B**). Rf value of each substrate observed by TLC in this experiment.

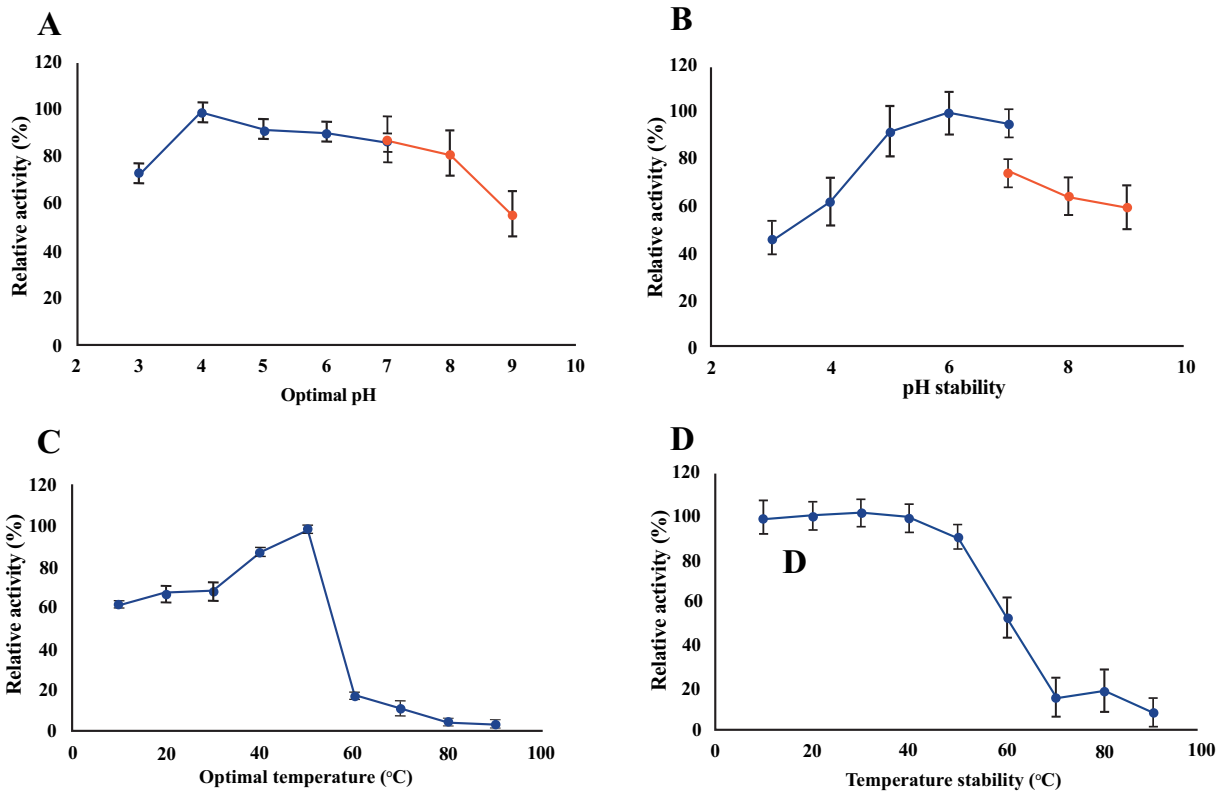


Fig. (6). Effect of temperature and pH on activity and stability of the recombinant protein Gh43kk. (**A**) Effect of pH, (**B**) pH stability, (**C**) Effect of temperature, (**D**) Temperature stability. Blue lines indicate experiments conducted with sodium acetate buffer. Orange lines indicate experiment conducted with Tris buffer. Beechwood xylan was used as substrate.

Table 2. Effects of various metal ions, chelating agents and surfactants on Gh43kk activity.

Agent	Relative Activity[a] (%)	Agent	Relative Activity [a] (%)
None*	100± 6.48	Mn^{2+}	120.52 ± 6.91
Ba^{2+}	76.12 ± 5.33	Na^{+}	79.25 ± 3.19
Ca^{2+}	38.07 ± 3.80	Ni^{2+}	48.57 ± 3.49
Co^{2+}	64.97 ± 3.56	Sr^{2+}	25.16 ± 2.61
Cu^{2+}	61.04 ± 9.32	Zn^{2+}	42.86 ± 7.71
Fe^{2+}	96.26 ± 3.56	EDTA	4.68 ± 3.85
K^{+}	78.44 ± 3.06	SDS	34.22 ± 6.41
Li^{+}	59.97 ± 3.30	TritonX-100	69.61 ± 2.04
Mg^{2+}	108.57 ± 6.98	-	-

4. DISCUSSION

Small proteins (SPs) are loosely defined as polypeptides that contain ≤100 amino acids, while some works suggested greater thresholds for SPs of ≤200 amino acids [16, 17]. In general, SPs contain at least a basic domain, which is sufficient to perform or assist a biological process. Previous studies suggested that the length of a protein sequence is connected to its specific functions and that SPs might show not many remarkable functions compared with large proteins [18]. However, there seems to be an important tendency in evolution favoring shorter rather than longer proteins for specific functions [19].

In this work, we demonstrated a small protein with a complete bifunctional activity as xylanase and cellulase despite its small size. As shown by the multiple alignments of the amino acid sequence of Gh43kk, the protein contains amino acid residues which were reported to be found on the active site of enzyme members in the glycosyl hydrolase family 43, namely Asp42, Gln93, Gly107, and Thr108. Some of these amino acid residues play critical roles in enzyme activity, such as Asp42, which was reported to play a major role in modulating pKa and in retaining the right position of the active site residues [20, 21].

A bifunctional enzyme is an enzyme that contains two different catalytic properties in the same polypeptide chain [22]. Utilization of both hexoses and pentoses sugars present in lignocellulosic biomass could be an important step to decrease production costs for many bioconversion processes. Therefore, a bifunctional xylanase/ cellulase is certainly more valuable than single xylanase or cellulase since it can degrade both the hemicellulose and cellulose fraction of the biomass simultaneously. Furthermore, the problem in the mashing process caused by viscous pentosans can be solved when both xylanase and cellulase are present in the fermentation mixture [23].

Several conserved amino acids indicating the predicted active site were found in the primary structure of the Gh43kk modules. This might indicate a minimal requirement for the catalytic activity of enzymes in this family. In general, GH43 enzymes exhibit a five-bladed β-propeller fold (clan F) [21]. The active site is found extending far down in the cavity at the

core of the β-propeller. In our work, the β-meander motif found in the structural model of Gh43kk would possibly represent one single blade out of the five blades of the β-propeller module of GH43 enzymes [21]. One important conserved residue, an aspartate (Asp43) might play a crucial role as a catalytic acid for the hydrolysis activity of this small enzyme. This residue is found at the hairpin loop of the peripheral β-sheet. The position of the Asp43 residue at the hairpin loop seems to share a similar position of the catalytic aspartate (Asp46) at the center of the active site of the GH43 five-bladed β-propeller module [21]. Hence, this result might suggest the most functional domain which is sufficient for the catalytic ability of the GH43 enzyme. However, the activity of Gh43kk is low when compared to other enzymes in the same categories, as cited in Table **3**. This might be due to its small size, which does not fulfill the complete configuration of the active site of the GH43 enzyme reported earlier. Nevertheless, this result might indicate a minimal domain/residue requirement for the functionality of the GH43 enzyme.

Inclusion bodies have been frequently reported when recombinant protein is expressed at a high level in *Escherichia coli*, which results in aggregation of the highly expressed protein molecules [24]. Causes of high expression of recombinant protein include using high temperature during protein expression, high concentration of inducer, and expression under strong promoter systems, which result in expression of the target protein at a high translational rate [25]. Expression of recombinant protein at low temperature has been considered to support the formation of "non-classical" inclusion bodies [26]. These non-classical inclusion bodies are soft, easily extractable by using non-denaturing solvents [27]. In general, overall cell processes slow down at low temperature, leading to a reduction rate of transcription, translation; hence recombinant proteins are produced in a lower amount. Furthermore, the degradation of proteolytically sensitive proteins also declines at low temperature. Lowering the concentration of the induction agent also improves the solubility of the recombinant protein by reducing the transcription rate as well.

The observed size of about 19 kDa of the recombinant protein in SDS-PAGE is slightly larger than the predicted size (12.81 kDa). These might be due to a decrease in protein mobility, which can be a result of post-translational modification in the possible SDS-binding sites, such as glycosylation, phosphorylation of serine, threonine and tyrosine, or sulfation of tyrosine [28]. Those alterations could change the internal hydrophobicity of the protein and hence slow down the protein migration observed in SDS-PAGE. Furthermore, the presence of additional chemicals in protein preparation, such as salts, detergents, or organic solutions, can result in the same consequence since they can alter or distort the protein or many of them might take part or obstruct the binding of SDS with the protein.

Substrate specificity of the recombinant enzyme Gh43kk was verified using three different methods, namely spectrophotometric analysis of released reducing sugar, agar well assay, and TLC analysis. The protein Gh43kk shows activities toward CMC and xylan, thereby confirming the previous finding about the dual activities of the Biof1_09 crude cell-free extract [1]. However, the treatment condition of rice straw has to be further investigated since xylose was not detected when rice straw was hydrolyzed under the given condition. This might be due to harsh treatment with acid and high temperature in the pretreatment condition that could degrade the loose hemicellulose structure (including xylose), thereby leaving only the rigid cellulose after pretreatment.

Gh43kk shows comparable properties with the Biof1_09 crude cell-free extract, including pH optimum (pH 4), optimal temperature (50°C), a wide range of pH stability (from pH 3–9) and stability at a temperature below 60°C, as reported earlier by the authors [1]. Some enzymatic properties of the recombinant Gh43kk protein were compared with other GH43 enzymes derived from metagenomics (Table **3**). Most of them work effectively under relatively high temperatures (40-55°C) and near-neutral pH (pH 5.5-7). However, it is notable that Gh43kk has an optimal pH for its activity in acidic conditions (pH 4).

Differences in metal ion sensitivities among the GH43 enzymes have been reported earlier [29]. To determine further properties of the enzyme Gh43kk in this study, several chemicals and metal ions were applied with the enzyme and its activity was monitored. Similar observation was made with CoXyl43, a GH43 β-xylosidase/α-arabinofuranosidase, which was inactivated by zinc and copper ions but was activated by manganese ions [30]. Chelating reagents such as EDTA had a negative impact on Gh43kk activity, which leads us to suggest that Gh43kk might be a metFalloenzyme. However, other GH43 enzymes, such as HiAbf43 and $XynA_{MG1}$, do not require metal ions for their activities [31].

Table 3. Summary of some GH43 enzymes derived from the metagenomics library.

Functions	Source	Substrate	Protein (kDa)	Optimum Temp (°C)	Optimum pH	Activity ($U \cdot mg^{-1}$)	References
Xylanase/ cellulase	biofertilizer	Beechwood xylan	23	50	4	0.29	This study
β-xylosidase	soil	pNPX	64.4	40	6	0.012 (12 $mU \cdot mg^{-1}$)	[31]
β-xylosidase/α-arabinofuranosidase	compost	birchwood xylan	36.19	55	7.5	22.0	[32]
β-xylosidase	gut bacterial genome	pNPX	61.07	50	6.5	2.83 ± 0.23	[33]
endo- β -1,4-xylanase	Cow rumen	beechwood xylan	99	50	6	709	[34]
β -D-xylosidase/α-L-arabinofuranosidase	rumen	pNPX	42	40	7	36.3	[35]

CONCLUSION

The enzyme in this study is one of the smallest GH43 enzymes ever reported so far. It displays a notable ability to work under acidic conditions, whereas other reported enzymes in the same family could work only at near-neutral pH. With its dual activities as xylanase and cellulase, and its ability to work under acidic condition and high temperature, Gh43kk enzyme could be an interesting candidate for large-scale expression for applications in industrial processes that require acidic conditions and high temperatures such as pulp and paper industries.

LIST OF ABBREVIATIONS

Asp43	=	Aspartate
SPs	=	Small Proteins
EDTA	=	Ethylenediaminetetraacetic Acid

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data that support the findings of this study are available within the article.

FUNDING

This study was funded by the National Research Council of Thailand (Grant number 570017).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors thank Mr. Pumi Eawpadung for laboratory assistance, and Dr. Piyanun Harnpicharnchai for the critical discussions. We also thank Assoc. Prof.Dr. Soontorn Oraintara for assistance in statistical analysis. We are also grateful to the Protein and Proteomics Research Center for Commercial and Industrial Purposes (ProCCI), Khon Kaen University, Khon Kaen, Thailand, for additional laboratory support.

REFERENCES

[1] Sae-Lee R, Boonmee A. Newly derived GH43 gene from compost metagenome showing dual xylanase and cellulase activities. Folia Microbiol 2014; 59(5): 409-17. [http://dx.doi.org/10.1007/s12223-014-0313-7] [PMID: 24737296]

[2] Isikgor FH, Becer CR. Lignocellulosic biomass: A sustainable platform for the production of bio-based chemicals and polymers. Polym Chem 2015; 6(25): 4497-559. [http://dx.doi.org/10.1039/C5PY00263J]

[3] Haghi AK, Gonzalez CNA, Thomas S, Praveen KM, Eds. Physical chemistry for engineering and applied sciences: theoretical and methodological implications. CRC Press 2018. [http://dx.doi.org/10.1201/9781315109725]

[4] Pelzer AW, Sturgeon MR, Yanez AJ, *et al.* Acidolysis of α-O-4 aryl-ether bonds in lignin model compounds: A modeling and experimental study. ACS Sustain Chem Eng 2015; 3(7): 1339-47. [http://dx.doi.org/10.1021/acssuschemeng.5b00070]

[5] Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. Bioresour Technol 2010; 101(13): 4851-61. [http://dx.doi.org/10.1016/j.biortech.2009.11.093] [PMID: 20042329]

[6] Balat M. Production of bioethanol from lignocellulosic materials *via* the biochemical pathway: A review. Energy Convers Manage 2011; 52(2): 858-75. [http://dx.doi.org/10.1016/j.enconman.2010.08.013]

[7] Lok C. Mining the microbial dark matter. Nature 2015; 522(7556): 270-3. [http://dx.doi.org/10.1038/522270a] [PMID: 26085253]

[8] Garrido-Cardenas JA, Manzano-Agugliaro F. The metagenomics worldwide research. Curr Genet 2017; 63(5): 819-29. [http://dx.doi.org/10.1007/s00294-017-0693-8] [PMID: 28401295]

[9] Okano H, Ozaki M, Kanaya E, *et al.* Structure and stability of metagenome-derived glycoside hydrolase family 12 cellulase (LC-CelA) a homolog of Cel12A from *Rhodothermus marinus*. FEBS Open Bio 2014; 4(1): 936-46. [http://dx.doi.org/10.1016/j.fob.2014.10.013] [PMID: 25426413]

[10] Kumar N, Sudan SK, Garg R, Sahni G. Enhanced production of novel halostable recombinant endoglucanase derived from the metagenomic library using fed-batch fermentation. Process Biochem 2019; 78: 1-7. [http://dx.doi.org/10.1016/j.procbio.2018.12.033]

[11] Chai S, Zhang X, Jia Z, *et al.* Identification and characterization of a novel bifunctional cellulase/hemicellulase from a soil metagenomic library. Appl Microbiol Biotechnol 2020; 104(17): 7563-72. [http://dx.doi.org/10.1007/s00253-020-10766-x] [PMID: 32666188]

[12] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976; 72(1-2): 248-54. [http://dx.doi.org/10.1016/0003-2697(76)90527-3] [PMID: 942051]

[13] Lineweaver H, Burk D. The determination of enzyme dissociation constants. J Am Chem Soc 1934; 56(3): 658-66. [http://dx.doi.org/10.1021/ja01318a036]

[14] Zhou J, Bao L, Chang L, Zhou Y, Lu H. Biochemical and kinetic characterization of GH43 β- d -xylosidase/α- l -arabinofuranosidase and GH30 α- l -arabinofuranosidase/β- d -xylosidase from rumen metagenome. J Ind Microbiol Biotechnol 2012; 39(1): 143-52. [http://dx.doi.org/10.1007/s10295-011-1009-5] [PMID: 21720773]

[15] Kozic M, Fox SJ, Thomas JM, Verma CS, Rigden DJ. Large scale ab initio modeling of structurally uncharacterized antimicrobial peptides reveals known and novel folds. Proteins 2018; 86(5): 548-65. [http://dx.doi.org/10.1002/prot.25473] [PMID: 29388242]

[16] Su M, Ling Y, Yu J, Wu J, Xiao J. Small proteins: Untapped area of potential biological importance. Front Genet 2013; 4: 286. [http://dx.doi.org/10.3389/fgene.2013.00286] [PMID: 24379829]

[17] Yang X, Tschaplinski TJ, Hurst GB, *et al.* Discovery and annotation of small proteins using genomics, proteomics, and computational approaches. Genome Res 2011; 21(4): 634-41. [http://dx.doi.org/10.1101/gr.109280.110] [PMID: 21367939]

[18] Wang F, Xiao J, Pan L, *et al.* A systematic survey of mini-proteins in bacteria and archaea. PLoS One 2008; 3(12): e4027. [http://dx.doi.org/10.1371/journal.pone.0004027] [PMID: 19107199]

[19] Lipman DJ, Souvorov A, Koonin EV, Panchenko AR, Tatusova TA. The relationship of protein conservation and sequence length. BMC Evol Biol 2002; 2(1): 20. [http://dx.doi.org/10.1186/1471-2148-2-20] [PMID: 12410938]

[20] Graciano L, Corrêa JM, Gandra RF, *et al.* The cloning, expression, purification, characterization and modeled structure of *Caulobacter crescentus* β-Xylosidase I. World J Microbiol Biotechnol 2012; 28(9): 2879-88. [http://dx.doi.org/10.1007/s11274-012-1099-x] [PMID: 22806729]

[21] Mroueh M, Aruanno M, Borne R, *et al.* The *xyl-doc* gene cluster of *Ruminiclostridium cellulolyticum* encodes GH43- and GH62-α-l-arabinofuranosidases with complementary modes of action. Biotechnol Biofuels 2019; 12(1): 144. [http://dx.doi.org/10.1186/s13068-019-1483-y] [PMID: 31198441]

[22] Khandeparker R, Numan MT. Bifunctional xylanases and their potential use in biotechnology. J Ind Microbiol Biotechnol 2008;

35(7): 635-44.
[http://dx.doi.org/10.1007/s10295-008-0342-9] [PMID: 18365260]

[23] Senn T, Pieper HJ, Roehr M, Eds. The biotechnology of ethanol: Classical and future applications. Wiley-Blackwell 2001.

[24] Humer D, Spadiut O. Wanted: More monitoring and control during inclusion body processing. World J Microbiol Biotechnol 2018; 34(11): 158.
[http://dx.doi.org/10.1007/s11274-018-2541-5] [PMID: 30341583]

[25] Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. Microb Cell Fact 2015; 14(1): 41.
[http://dx.doi.org/10.1186/s12934-015-0222-8] [PMID: 25889252]

[26] Peternel Š, Grdadolnik J, Gaberc-Porekar V, Komel R. Engineering inclusion bodies for non denaturing extraction of functional proteins. Microb Cell Fact 2008; 7(1): 34.
[http://dx.doi.org/10.1186/1475-2859-7-34] [PMID: 19046444]

[27] Singh A, Upadhyay V, Panda AK. Solubilization and refolding of inclusion body proteins. Methods Mol Biol 2015; 1258: 283-91.
[http://dx.doi.org/10.1007/978-1-4939-2205-5_15] [PMID: 25447870]

[28] Yang YS, Wang CC, Chen BH, Hou YH, Hung KS, Mao YC. Tyrosine sulfation as a protein post-translational modification. Molecules 2015; 20(2): 2138-64.
[http://dx.doi.org/10.3390/molecules20022138] [PMID: 25635379]

[29] Viborg AH, Sørensen KI, Gilad O, *et al.* Biochemical and kinetic characterisation of a novel xylooligosaccharide-upregulated GH43 β-d-xylosidase/α-l-arabinofuranosidase (BXA43) from the probiotic *Bifidobacterium animalis* subsp. lactis BB-12. AMB Express 2013; 3(1): 56.
[http://dx.doi.org/10.1186/2191-0855-3-56] [PMID: 24025736]

[30] Matsuzawa T, Kaneko S, Yaoi K. Screening, identification, and characterization of a GH43 family β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome. Appl Microbiol Biotechnol 2015; 99(21): 8943-54.
[http://dx.doi.org/10.1007/s00253-015-6647-5] [PMID: 25971196]

[31] AL-Darkazali H, Meevootisom V, Isarangkul D, Wiyakrutta S. Gene expression and molecular characterization of a xylanase from chicken cecum metagenome. Int J Microbiol 2017; 2017: 1-12.
[http://dx.doi.org/10.1155/2017/4018398] [PMID: 28751915]

[32] Campos E, Negro AMJ, Sabarís di Lorenzo G, *et al.* Purification and characterization of a GH43 β-xylosidase from *Enterobacter* sp. identified and cloned from forest soil bacteria. Microbiol Res 2014; 169(2-3): 213-20.
[http://dx.doi.org/10.1016/j.micres.2013.06.004] [PMID: 23838121]

[33] Xu B, Dai L, Zhang W, *et al.* Characterization of a novel salt-, xylose- and alkali-tolerant GH43 bifunctional β-xylosidase/α-l-arabinofuranosidase from the gut bacterial genome. J Biosci Bioeng 2019; 128(4): 429-37.
[http://dx.doi.org/10.1016/j.jbiosc.2019.03.018] [PMID: 31109875]

[34] Zhao S, Wang J, Bu D, *et al.* Novel glycoside hydrolases identified by screening a Chinese Holstein dairy cow rumen-derived metagenome library. Appl Environ Microbiol 2010; 76(19): 6701-5.
[http://dx.doi.org/10.1128/AEM.00361-10] [PMID: 20709844]

[35] Yang X, Shi P, Ma R, *et al.* A new GH43 α-arabinofuranosidase from *Humicola insolens* Y1: Biochemical characterization and synergistic action with a xylanase on xylan degradation. Appl Biochem Biotechnol 2015; 175(4): 1960-70.
[http://dx.doi.org/10.1007/s12010-014-1416-y] [PMID: 25432346]

© 2022 Oraintara and Bhunaonin

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: https://creativecommons.org/licenses/by/4.0/legalcode. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.