https://doi.org/10.31436/iiumej.v25i2.2024

FUNCTIONAL METAGENOMICS APPROACH FOR DISCOVERY OF NOVEL COLD-ACTIVE PROTEASE FROM ANTARCTIC REGION

MUHAMMAD ASYRAF ABD LATIP[1,2], NOOR FAIZUL HADRY NORDIN[3*], SITI AISYAH ALIAS[4,5], JERZY SMYKLA[6], FARIDAH YUSOF[2], MOHD AZRUL NAIM MOHAMAD[7]

[1]Langkawi Mariculture Research Center, Kompleks Perikanan Bukit Malut, 07000 Langkawi, Kedah, Malaysia
[2]Department of Chemical Engineering & Sustainability, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur
[3]International Institute for Halal Research and Training (INHART), Block A, Level 3, KICT Building, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur
[4]Institute of Ocean and Earth and Sciences, C308, Level 3, Block C, Institute for Advanced Studies Building, 50603 Kuala Lumpur, Malaysia
[5]National Antarctic Research Centre, B303, Level 3, Block C, Institute for Advanced Studies Building, 50603 Kuala Lumpur, Malaysia
[6]Institute of Nature Conservation, Polish Academy of Sciences, al. A. Mickiewicza 33 PL-31-120 Krakow, Poland
[7]Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia

**Corresponding author: faizul@iium.edu.my*

(Received: 5 November 2023; Accepted: 12 March 2024; Published online: 15 July 2024)

ABSTRACT: The structural complexity of bacterial life makes most of it impossible to culture. Functional metagenomics approaches overcome the limitations of a culture-based approach in exploring and assessing the genetic materials of uncultured microbes. The objective of this study was to identify clones with cold-active proteases through functional metagenomics. In this work, the environmental DNA (eDNA) isolated directly from Antarctic soils was ligated into the pCC1FOS fosmid vector, transformed into EPI300-T1R *E. coli* host cells, and screened for proteolytic enzymes. Positive protease-producing clones were identified and isolated on skim milk agar supplemented with chloramphenicol and arabinose. This clone harbored a fosmid, pCC1FOS, which has a 48.5 kb insert that has been completely sequenced in both directions. Further analysis of the insert showed 70 NODEs. The NODE_42 encoded hypothetical protein of 297 amino acids showed a significant match to Peptidase M23 and PG-binding 1 proteins families. A three-dimensional model of the predicted protease was generated based on the known mesophilic protease of *Neisseria meningitides* (PDB: 3SLU). The structural alignment showed 27.07 % similarity with RMSD value of 0.402 Å based on 58 aligned residues. The active site residues were identical, but major deletions were observed in the predicted proteases. This predicted protease showed higher activity at -20 °C and 20 °C than the positive control (protease from bovine pancreas). Functional metagenomics is a promising approach in the discovery of cold-active protease with low homology to the known sequences and expressed in the host cell that has the potential for bioprospecting in low-temperature applications.

ABSTRAK: Kesukaran struktur kehidupan bakteria menyebabkan kebanyakan bakteria sukar dikultur. Saringan fungsi metagenomik dapat mengatasi kekangan saringan berasaskan kultur dalam meneroka dan menilai bahan genetik mikrob tidak kultur. Objektif kajian ini adalah

bagi mengenal pasti klon protease aktif sejuk melalui saringan fungsi metagenomik. Kajian ini mencadangkan DNA alam sekitar (eDNA) yang diasingkan secara langsung dari tanah Antartika, dan dimasukkan ke dalam vektor fosmid pCC1FOS, ditransformasikan ke dalam sel hos EPI300-T1R E. coli, dan disaring untuk enzim proteolitik. Klon yang menghasilkan protease ini dikenal pasti dan diasingkan daripada agar susu skim. Agar telah ditambah dengan kloramfenikol dan arabinosa. Klon ini mengandungi fosmid pCC1FOS, di mana sisipan sebesar 48.5 kb telah disusun berturut-turut sepenuhnya dari kedua-dua arah. Analisis lanjut pada sisipan ini menunjukkan terdapat 70 NODE. NODE_42 mengandungi 297 asid amino protein hipotetis, iaitu sangat serasi dengan protein Peptidase M23 dan PG-binding 1. Model tiga dimensi protease yang diramalkan ini dihasilkan berdasarkan protease mesofilik Neisseria meningitides (PDB: 3SLU). Susunan struktur menunjukkan 27.07% persamaan nilai RMSD sebanyak 0.402 Å berdasarkan 58 sisa sejajar. Sisa tapak aktif kedua-duanya adalah sama, tetapi terdapat pemotongan besar telah berlaku dalam protease ramalan. Protease ini menunjukkan aktiviti enzim lebih tinggi pada suhu -20 °C dan 20 °C berbanding kawalan positif (protease dari pankreas lembu). Saringan fungsi metagenomik berhasil menemukan protease aktif sejuk berhomologi rendah terhadap susunan yang diketahui dan diekspresi dalam sel hos. Ia berpotensi bagi bioprospeksi dalam aplikasi bersuhu rendah.

KEYWORDS: *metagenomics; cold-active; enzyme; protease; Antarctic*

1. INTRODUCTION

Metagenomics study was encouraged with the escalation of next-generation sequencing (NGS) technologies such as 454 pyrosequencing, Illumina, Sequencing by Oligonucleotide Ligation and Detection (SOLiD), Helicos Genetic Analysis Platform, Ion Torrent, PacBio (SMRT) and Oxford Nanopore sequencing. First-generation sequencers, Sanger or Maxam-Gilbert sequencing, were not capable of sequencing this eDNA due to its huge size. NGS will generate massive post-sequencing data and require bioinformatic skills for processing and analyzing. This is an approach of a direct route for analyzing the DNA from the sample. The DNA or genetic materials are extracted directly from the environment sample, and this DNA is called environmental DNA (eDNA) [1]. This powerful study has the capability for the discovery of specific genes from both culturable and non-culturable bacteria [2,3]. The limitation of culture-based screening restricts the exploration of the potential enzymes or genes because less than 1% of the total bacterial community can be cultivated in standard media [4]. Most of the bacteria are uncultivable because of their natural complexity.

Functional metagenomics is an approach to analyzing the whole genome for novel bioactivity. The main objective of functional metagenomics sequencing is to screen and identify novel enzymes or genes for a specific reaction. The advantages of this technique are that the researchers do not need to have prior knowledge about the gene sequence and do not have to depend on the known gene sequence from previous studies to identify the targeted enzyme. Thus, the culture-based method is important in the process to minimize false-positive errors. Prior to this approach, a more novel gene that is responsible for functional enzymatic expression was discovered. Many novel enzymes have been discovered via functional metagenomics, such as beta-lactamase [5], β-galactosidases [6], serine protease [7], cellulases [8], endoglucanases [9], phosphatases and phytases [10]. The novel enzymes can be defined when the similarity of gene homology is lesser (< 70%) compared to the known sequence [11]. A new characteristic of the enzyme can be identified as to whether it has the valuable potential for industrial and biotechnological application.

In addition, there are limited functional metagenomics studies that have been performed on the Antarctic regions. Although these cold-active enzymes have been discovered previously

within other extreme cool regions or the Arctic, unique characteristics and behaviors of the enzyme under certain conditions are important to be analyzed and applied in a real-life situation. Based on the findings, only five novel enzymes have been discovered from the Antarctic region via the functional metagenomics approach: lipase and esterase [12], esterase [13], lipase [14], chitinase [15] and alkane monooxygenase [16]. There are still a lot more opportunities to be discovered in this region, such as new species and novel enzymes. The objective of this study is to discover cold-active protease from Antarctic soil samples through the functional metagenomics approach and the scope of this study focused on the computational and prediction level.

2. MATERIALS AND METHODS

2.1. Environmental DNA (eDNA) Extraction

Soil samples obtained from the previous study were collected along with the coastal environments of the Southern Victoria Land during the Antarctic summer 2011/2012 [17]. Five grams of soil sample were weighed and transferred into the 10 mL elution buffer (4 % cetyltrimethylammonium bromide (CTAB), 1.5 M sodium chloride (NaCl), 0.1 M tris, 25 mM ethylenediaminetetraacetic acid (EDTA) and 1 % polyvinylpyrrolidone (PVP)). Subsequently, 0.03 mL of β-mercaptoethanol was added, and the soil was incubated at 65 °C for 30 minutes while being inversely mixed a few times. The sample was centrifuged for 5 minutes at 5,000 x g to remove the foam. The clear supernatant was transferred into a new tube, and one volume of chloroform phenol: chloroform: isoamyl alcohol was added. The sample was centrifuged at 5,000 x g for 10 minutes after inversely mixing for 5 minutes. The aqueous layer, or upper layer, was transferred into the new tube, and another volume of phenol, chloroform, or isoamyl alcohol, was added. The solution was inversely mixed for 5 minutes and centrifuged for another 10 minutes at 5,000 x g. The aqueous layer was carefully transferred to avoid contamination from the bottom layer and entering the new tube. The half volume of 5 M NaCl was added and inversely mixed. Three volumes of cold 95 % ethanol were added and inversely mixed. The sample was incubated at – 20°C for one hour. The sample was then centrifuged at the highest speed for 10 minutes. The supernatant was discarded, and the pellet DNA was washed with 3 mL of 70 % ethanol. The sample was centrifuged at 5,000 x g for 15 minutes. After discarding the supernatant, pelleted DNA was air-dried at room temperature. The air-dried DNA was suspended with 0.1 mL of TE buffer and stored at – 20°C.

2.2. Screening of Functional Expression Library

The extracted eDNA was ligated to the fosmid pCC1FOS and transfected into the host cells EPI300-T1R via a phage system. The host cell *E. coli* EPI300-T1R cell comes with the commercial kit CopyControl Fosmid Library Production Kit with pCC1FOS Vector (Epicentre, Madison, WI, USA). According to the manufacturer's protocol, the host cell or *E. coli* EPI300-T1R cell provides a mutant *trfA* gene whose gene product is required for initiation of replication from *oriV*. The cells have been engineered so that the *trfA* gene is under tight, regulated control of an inducible promoter. For this process, the CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector (Epicentre) was utilized according to the manufacturer's instructions. The infected EPI300-T1R cells were screened for functional gene expression on the 5 % (w/v) skim milk agar plates supplemented with 12.5 µg/mL chloramphenicol and 0.01 % arabinose. Arabinose induces the expression of the *trfA* gene product and subsequent amplification of the clones to a high copy number for direct function-based screening assays of the clone library (Epicentre, Madison, WI, USA). The clones were cultured for a week at 20 °C. After incubation, the formation of a clearing halo zone formation indicated the activity of the protease. A clone with positive proteolytic activity was isolated,

and the fosmid extraction process proceeded. The isolated clone was stored in 25 % glycerol at -80 °C. Fosmid extraction was performed using the FosmidMAX™ DNA Purification Kit (Lucigen). The purified fosmid was analyzed by 1 % agarose gel electrophoresis. Then, 50 µL of the suspended DNA was sent for sequencing services through Apical Scientific Sdn Bhd.

2.3. Metagenomics Data Analysis

Paired-end Illumina sequences were first removed from sequence adaptors and reads with low-quality scores using bbduk of the BBTools Packages. All QC-ed reads were filtered against the fosmid sequence (pCC1FOS). Unmapped reads (both paired-end and unpaired reads) were then assembled using SPAdes 3.11.1 [18]. We employed a two-prong approach to identify potential contigs associated with the fosmid insert sequence. First, all contigs >5000 bp were subjected to Principal Component Analysis of their tetranucleotide frequency [19]. Next, all contigs were filtered using BLASTN (maximum E value of 10) against a database of *Escherichia coli* consisting of AP021894, CP025520, NC_000913. The contigs within NODE_42 were subsequently annotated using Rapid Annotations using Subsystems Technology (RAST) [20].

2.4. *In silico* Analysis

The percentage of amino acid proportion was calculated using the GPMAW Lite program (Alphalyse, Palo Alto, CA, USA). Domains and active sites of the enzymes were searched through the Pfam database and NCBI Conserved Domain Search [21]. The protein sequences were aligned with the program ENDscript, and the secondary structures were compared with those of the program ESPript [22]. PyMOL calculated the RMSD values of psychrophilic and mesophilic enzymes after performing structure superimposition [23]. The structural differences between both enzymes were analyzed when aligning both enzymes together.

2.5. Protease Assay of Crude Enzyme

For the enzymatic assay, 250 µL of the crude enzyme, 250 µL of 1% Azocasein and 250 µL of 0.5 M Tris were incubated at -20 °C, 20 °C and 60 °C. The clone was incubated in 10 mL of skim milk broth at 200 rpm until the media became clear. Then, the culture was centrifuged at 15,000 x g for 15 minutes and the supernatant was considered as the crude enzyme. Before adding the crude enzyme, the assay was pre-incubated at mentioned temperature for 10 minutes. For positive control, 0.2 units/mL of protease from the bovine pancreas was used. After 20 minutes, 500 µL of 25% trichloroacetic acid (TCA) was added to stop the reaction and centrifuged at 11,000 x g for 30 minutes. The absorbance of the supernatant was read at 340 nm.

3. RESULTS AND DISCUSSION

3.1. Clone with Positive Proteolytic Activity

In this experiment, only one clone showed the formation of a halo zone (Fig. 1 (A)). This clone was isolated, and the fosmid was extracted and purified. Fig. 1 (B) shows the gel electrophoresis results of the digested and the undigested fosmid with DNA insertion. After the fosmid was treated with the restriction enzyme, *EcoRI*, two separated bands appeared as the upper band was the insert DNA (± 40 kb) and the lower band was the fosmid DNA (8 kb). The undigested fosmid showed a single band with a high molecular weight of more than 48.5 kb.

https://doi.org/10.31436/iiumej.v25i2.2024

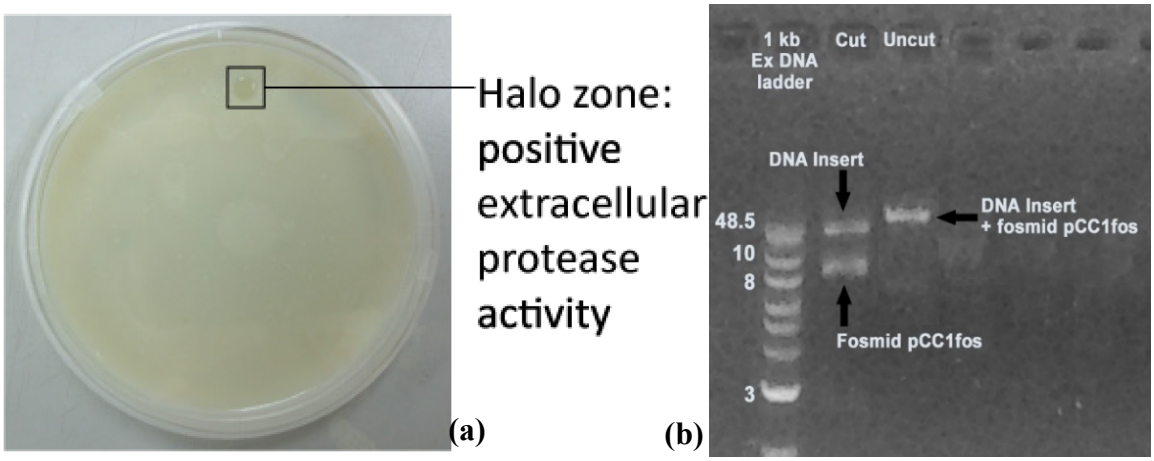

Figure 1. (a) A clone with halo zone formation on skim milk agar plate supplemented with chloramphenicol and arabinose. (b) The fosmid was analyzed by 1 % agarose gel electrophoresis. The undigested fosmid showed a single band in lane 1, while the digested fosmid with EcoR1 showed two separated bands in lane 2.

Other studies also showed that the hit rate for functional enzyme discovery through functional metagenomics was very low, as tabulated in Table 1[24]. This is because there were major bottlenecks for the unknown origin of the genes to be expressed in the host cell. Ekkers et al. (2012) have listed the causal factors for the lack of gene expression, including variations in the use of codons, inappropriate identification of promoters, lack of proper activation factors, ribosomal entry, improper folding of proteins, lack of important co-factors, rapid enzymatic degradation of the gene product, body development of incorporation, gene product toxicity, or the host's failure to secrete the product of gene expression. However, the advantage of this screening technique was that when the clone showed positive activity, it indirectly established a suitable host system for enzyme expression. Up to date, several studies successfully discovered protease through this approach, such as oxidant-stable serine protease from forest soil [7], putative proteases from slaughterhouse biofilm [25], alkaline protease from tannery-activated sludge [26] and mangrove-sediment [27]. However, some researchers hit zero rate frequency for protease activity through functional metagenomics approaches [28,29].

Table 1: Protease activity screening through functional metagenomics approach.

No. of positive clones	Vector	Total clones	Hit rate (%)	Sample	References
1	Plasmid	35,000	0.003	Forest soil	[7]
5	Plasmid	170,000	0.003	Slaughterhouse biofilm	[25]
1	Plasmid	10,000	0.010	Tannery sludge	[26]
4	Fosmid	1,824	0.219	Mangrove soil	[27]
0	Plasmid	32,000	0.000	Garden soils and compose	[28]
0	BAC	2,843	0.000	Ikaite columns	[29]

3.2. Metagenomics Data Analysis

From the shotgun metagenomic sequencing, the total number of reads for the raw data obtained for both forward and reverse sequences was 55,923,089 reads. After removing the sequence adaptors and low-quality reads, the total high-quality reads were reduced to 55,673,161 for both forward and reverse sequences. After assembling the reads using SPAdes, 70 NODEs were identified in the sample. NODE or scaffolds reconstructed the DNA sequences

https://doi.org/10.31436/iiumej.v25i2.2024

based on the contigs and bridged via the gap [30]. Contigs were generated after the DNA reads (k-mers) were assembled and sequenced according to the adjacent overlapping of the DNA base. Initially, all the scaffolds were analyzed using the Principal Component of Analysis (PCoA) of their tetranucleotide frequency. Accumulation of the NODEs at the center of the graph meant that the NODEs have similarities in their tetranucleotide pattern between each other. These scaffolds were majorly subjected to the host cell, *E. coli*. The targeted scaffolds were outlying the eclipse area where these were potential candidates for fosmid with the insertion. From the graphs in Fig. 2, the selected scaffolds were NODE_42, NODE_55, NODE_64, NODE_67 and NODE_69.

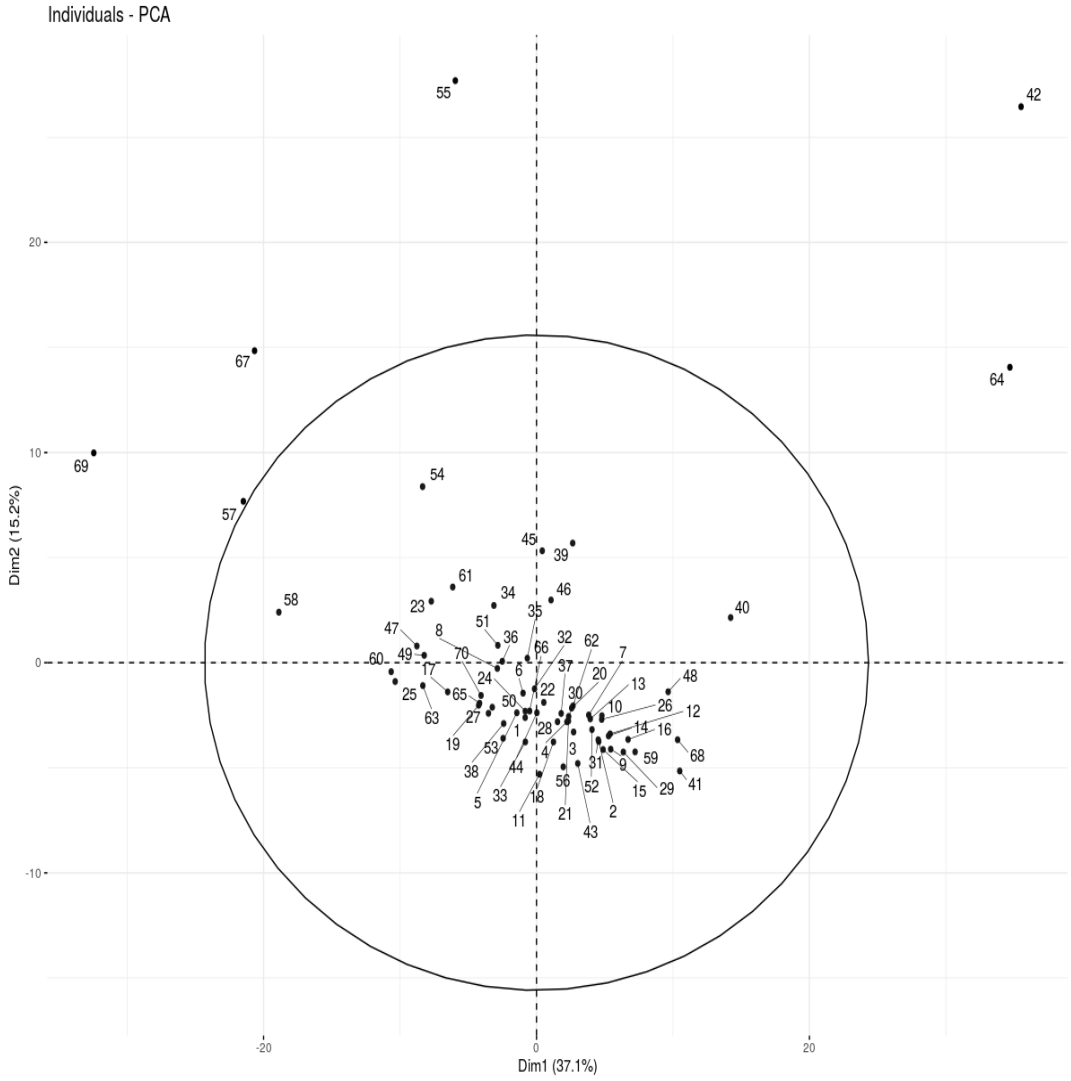

Fig. 2. The PCoA-tetranucleotide frequency of the assembled contigs. Contigs (NODE) outside of the eclipse were potential candidates for the fosmid insert.

Based on the BLASTN results, all the nodes showed high similarity to the *E. coli* except for NODE_42. Ambiguously, NODE_55, NODE_64, NODE_67 and NODE_69 have different tetranucleotide patterns to the *E. coli* genome based on the PCoA result because these scaffolds were part of mobile genetic elements (MGEs) and were expected to have differential patterns in their nucleotide composition as illustrated in Fig. 3. MGEs were genetic material that can move around the genome and also can be transferred among bacterial genomes. MEGs carried noncore genes and provided important contributions to bacterial adaptation to the environment,

such as antibiotic-resistant genes in *E. coli* [31]. The candidate NODE_42 showed similarity to the cloning vector pCC1FOS, but with less percentage query coverage, 1.04 %.

Figure 3. Alignment of NODE_67 to a fragment of the E. coli reference genome where mobile genetic elements (MGEs) were detected in this region. NODE_55, NODE_64 and NODE_6 also showed the same results.

3.3. *In silico* Analysis and Enzymatic Assay of Predicted Protease

Further analysis was focused on NODE_42 as this was the only candidate for DNA insertion. After trimming the fosmid sequence, the total length of the insertion was 34493 bp. As expected, the DNA insert size length should be between 20 kb to 40 kb as these were the range sizes of gDNA extracted from agarose gel. This scaffold NODE_42 contained 56 contigs of DNA with an average size of 572 bp with the lowest and the highest size was 104 bp and 2648 bp, respectively. From the RAST results, the majority of the contigs were hypothetical proteins. In the Pfam database, the function of the contigs was predicted through the conserved part of the protein sequence and tertiary structure called domain. In this experiment, contigs of NODE 42 (893 bp) showed significant match to Peptidase M23 and PG binding 1 protein families based on both Pfam and NCBI databases as illustrated in Fig. 4 (A) and (B), respectively. According to on Pfam database, the family members of peptidase M23 (PF01551) were generally categorized as endopeptidases. Specifically, zinc metallopeptidases with several specificities and Gly-Gly endopeptidases were included in this family. Based on the MEROPS, the activities of this peptidase family were cleaving the cross-linking peptide bonds either between or within the cell wall of peptidoglycan [32]. Besides, this family member has a propensity to cleave glycine-rich substrate, such as elastin.

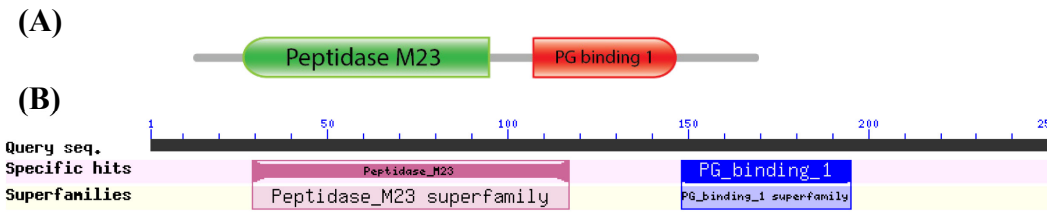

Figure 4: (A) Pfam result analysis showed two domain sections, Peptidase_M23 (green) and PG_binding_1 (red). (B) Analysis of the conserved domain through NCBI showed that the selected contig's protein sequence hit specific domain Peptidase M23 and PG binding 1 superfamily.

This contig with 893 bp was translated into a 297 amino acid sequence for further analysis. Based on this protein sequence, the Swiss-model tool constructed the structure of the predictive homology model based on the M23 peptidase domain template with PDB ID 3SLU. This template protein was characterized as a proteolytic enzyme located at the outer membrane of *Neisseria meningitidis* [33]. The percentage composition of an amino acid between the predictive model and the template was calculated in Table 2. In this experiment, the predictive model contained a lower arginine residue proportion, 6.06%, compared to the template enzyme, 7.82%. In contrast, the proportion of proline for the predictive enzyme was 4.71% higher than

template enzymes, 3.77%, but the count numbers were the same, 14 residues. From this finding, it appeared that arginine residue proportion was more crucial than proline for cold adaptation in enhancing enzyme structure flexibility. In addition, there were also reports that proline residue appeared to be increased in some psychrophilic enzymes [34–36]. These psychrophilic enzymes showed thermostable characteristics, and they can work at lower to higher temperatures [37]. The flexibility of the enzyme structure could affect the stability and function of the enzyme. When the enzyme became more flexible, the conformation of the active site became easier to change and form. This allows the enzyme to bind tighter with various forms of substrates [38,39]. In this situation, the induced fit model was much more relevant compared to the lock and key model [40]. When the substrate was bound to the enzyme, the active site changed its conformation to align perfectly with the substrate's structure. This optimal orientation strengthens the binding of enzyme and substrate complex for catalytic activity.

Table 2: Amino acid compositions between predictive and template model.

Amino acids	3-Letter	1-Letter	Predictive model		Template model (3SLU)	
			Count	%	Count	%
Alanine	Ala	A	32	10.77	41	11.05
Cysteine	Cys	C	3	1.01	0	0
Aspartic Acid	Asp	D	26	8.75	20	5.39
Glutamic Acid	Glu	E	10	3.37	22	5.93
Phenylalanine	Phe	F	11	3.7	12	3.23
Glycine	Gly	G	31	10.44	50	13.48
Histidine	His	H	6	2.02	11	2.96
Isoleucine	Ile	I	9	3.03	15	4.04
Lysine	Lys	K	7	2.36	14	3.77
Leucine	Leu	L	19	6.4	28	7.55
Methionine	Met	M	3	1.01	4	1.08
Asparagine	Asn	N	9	3.03	8	2.16
Proline	Pro	P	14	4.71	14	3.77
Glutamine	Gln	Q	12	4.04	16	4.31
Arginine	Arg	R	18	6.06	29	7.82
Serine	Ser	S	18	6.06	23	6.2
Threonine	Thr	T	24	8.08	16	4.31
Valine	Val	V	26	8.75	31	8.36
Tryptophan	Trp	W	8	2.69	3	0.81
Tyrosine	Tyr	Y	11	3.7	14	3.77

In overall secondary structures in Fig. 5 (A), there were fewer differences that could be observed between the predicted protease and the template enzyme. The active site residue between the predicted protease and the template enzyme was identical. The catalytic residue for the predicted protease and the template enzyme was His111 and His373, respectively. Protein members in this M23 family were characterized by the conserved histidines that occurred at the HXH motif, as shown in Fig. 5 (a) [41,42]. Based on the structural alignment in Fig. 5 (b), the predicted structure showed 27.07% similarity with the template enzyme. These two structures were aligned according to rigid body superstition. Based on the superposed analysis, the predicted enzyme showed a root mean square deviation (RMSD) value of 0.402 Å based on aligned 58 residues. A lower RMSD value meant higher structure similarity between proteins, and the model was considered very close to the template when the RMSD value was less than 3 Å [43].

https://doi.org/10.31436/iiumej.v25i2.2024

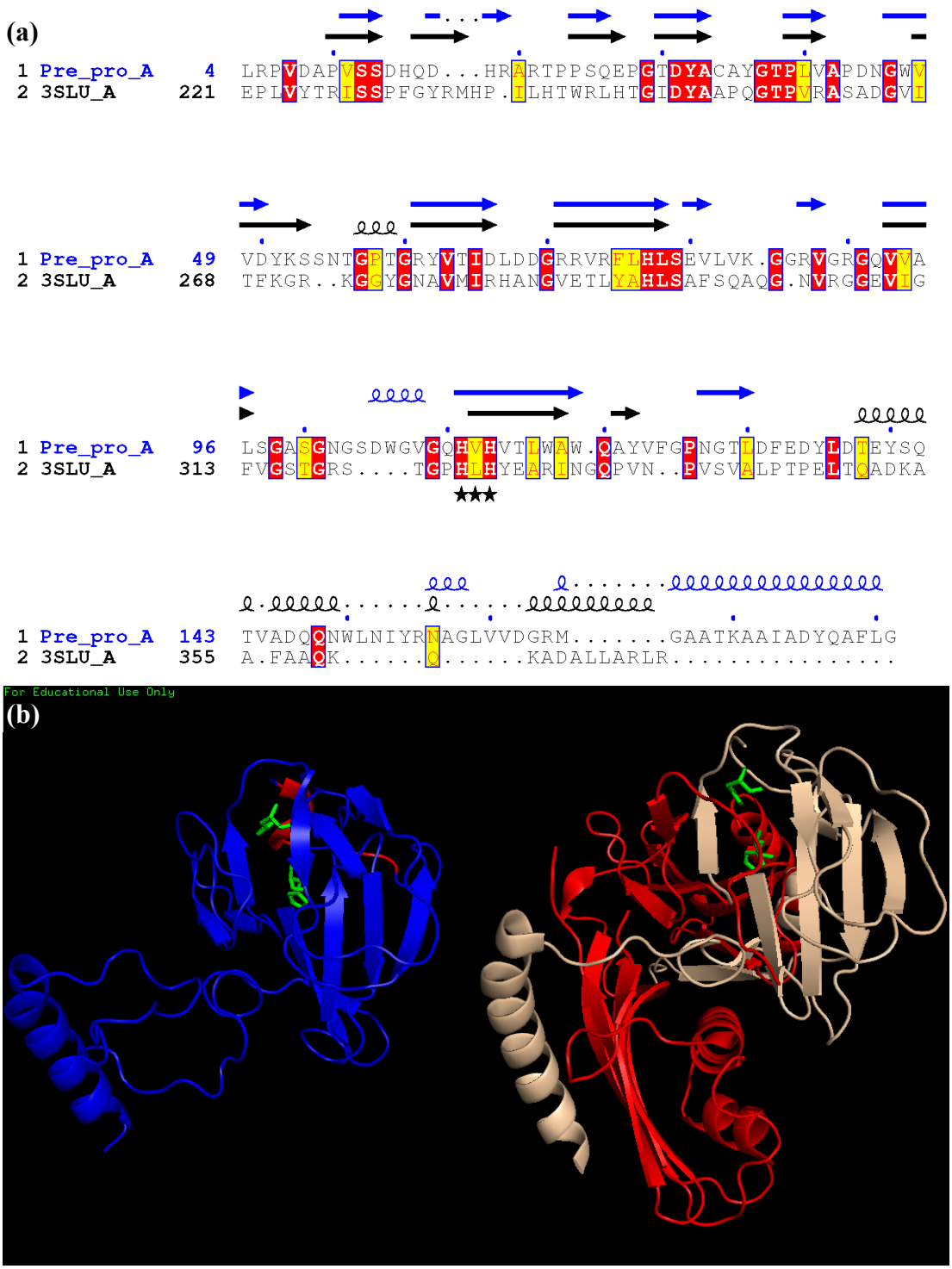

Figure 5. (a) Sequence alignment of predicted protease (Pre_pro) and the template enzyme (3SLU). There was not much difference that can be observed in the overall secondary structure of both enzymes (*** is conserved HXH motif). (b) Predicted protease homology structure (blue) and the template enzyme (3SLU) (brown). Large deletion (red) was observed when both enzymes aligned together.

Based on this computational analysis, the predicted enzyme sequence was shorter and the size was smaller than the template enzyme. The most noticeable variance was the deletion of a large section at the beginning of the predicted enzyme. In the majority, deletion of this section

was assumed to provide the more accessible active site for substrate binding of psychrophilic enzymes [44,45]. Besides, the function of the loop in assisting the enzyme-substrate binding required energy [46,47]. As psychrophiles were surrounded by a lower-temperature environment and exerted lower kinetic energy, the loop formation appeared to be functionless. To compensate for these deleted loops, the psychrophilic enzyme was adapted with a higher flexibility structure, as mentioned earlier. Higher flexibility and more accessible active site of enzyme reduced the substrate specificity and minimized the energy for substrate approachability [48]. Due to the scope of the study at the computational and predictive level, the determination of enzyme size will proceed later through SDS-PAGE.

Moreover, these structural features of the cold-adapted enzymes fitted their thermo-liability characteristic. When the temperature was raised, heat disrupted hydrogen bonds and non-polar hydrophobic interaction formation within the secondary and tertiary structure of the enzyme [49]. As previously discussed, the flexible structure of the cold-active enzyme was increased when these bond formations were decreased by lowering the proportion of arginine and proline residues. Therefore, the fewer bond formations within the enzyme, the easier the enzyme could be denatured by heat. In addition, loop formation also provided extra support for the enzyme to be structurally maintained in high temperatures [50]. When these loops were deleted in a cold-adapted enzyme, the enzyme could not withstand high temperatures, and it unfolded or denatured.

3.4. Enzymatic Assay

For the enzymatic assay, the positive control protease (mesophilic enzyme from bovine pancreases) is used to compare the activity of the predicted protease to determine whether the enzyme is categorized as a cold-active enzyme or not.

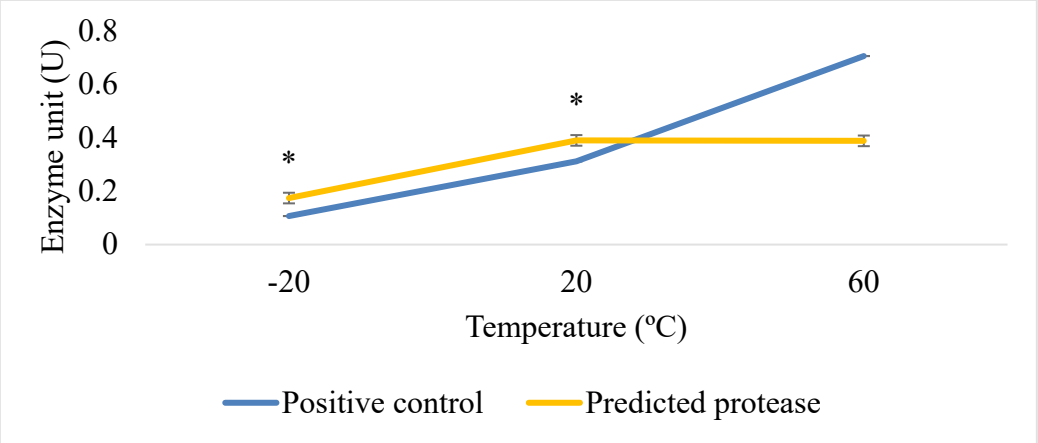

Figure 6. The crude enzyme activity of the predicted protease showed significantly higher at -20 and 20 ºC than the positive control (protease from bovine pancreases). Results were expressed as mean ± SEM. * $p < 0.05$.

Based on the graph in Fig. 6, the predicted protease showed significantly higher proteolytic activity than the positive control at -20 and 20 °C. Regarding cold-active protease, this enzyme has been extracted from psychrophiles inhabiting extreme cold environments, and the majority showed optimal activity at a temperature range between 30 – 40 ºC [51–59]. Researchers' interest in cold-active enzymes has led to the discovery of many novel cold-active proteases and reviewed their bioprospecting potential. Generally, cold-active enzymes are eco-friendly and economical due to their saving energy and cost processing. The use of cold-active protease

in the detergent is beneficial because washing with cold water protects the color and structure of the fabric [60]. In the leather industry, the peeling process with cold-active protease can be performed using tap water only [61]. In the textile industry, cold-active protease could be used in the treatment processes on synthetic fibers that are sensitive to higher temperatures to provide a unique finishing surface [62]. Furthermore, in the food industry, processing at a low temperature preserves the nutritional value and taste in fish and soy sauce production [63]. Additionally, cold-active protease can be used as a meat tenderizer while stored in the freezer [64].

4. CONCLUSION

This preliminary research is vital to unearth all potential enzymes for bioprospecting purposes. The discovery of cold-active proteases in this current research has the potential to be applied in the various industrial field and biotechnological applications. In future, purification of this enzyme is essential to inspect precisely the optimum conditions for enzyme activity.

ACKNOWLEDGEMENT

The research leading to these results received funding from Yayasan Penyelidikan Antartika Sultan Mizan (2015) and International Islamic University Malaysia under grant number RIGS 16-332-0496. The field survey and sampling were made possible with funding to J. Smykla from the Polish National Science Center within grant number NN305376438

REFERENCES

[1] Bálint M, Pfenninger M, Grossart HP, Taberlet P, Vellend M, Leibold MA, et al. (2018) Environmental DNA time series in ecology. Trends Ecol Evol., 33(12):945–57.

[2] Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, et al. (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl Environ Microbiol., 72(5):3291–301.

[3] Pearce DA, Newsham KK, Thorne MAS, Calvo-Bado L, Krsek M, Laskaris P, et al. (2012) Metagenomic analysis of a southern maritime Antarctic soil. Front Microbiol., 3(DEC):1–13.

[4] Bodor A, Bounedjoum N, Vincze GE, Erdeiné Kis Á, Laczi K, Bende G, et al. (2020) Challenges of unculturable bacteria: environmental perspectives. Reviews in Environmental Science and Biotechnology.

[5] Marathe NP, Janzon A, Kotsakis SD, Flach CF, Razavi M, Berglund F, et al. (2018) Functional metagenomics reveals a novel carbapenem-hydrolyzing mobile beta-lactamase from Indian river sediments contaminated with antibiotic production waste. Environ Int.

[6] Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD, et al. (2017) Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. PLoS One.

[7] Biver S, Portetelle D, Vandenbol M. (2013) Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. Springerplus, 2(1):1–10.

[8] Maruthamuthu M, Jiménez DJ, Stevens P, van Elsas JD. (2016) A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveilsnovel thermoalkaliphilic enzymes. BMC Genomics.

[9] Pabbathi NPP, Velidandi A, Tavarna T, Gupta S, Raj RS, Gandam PK, et al. (2021) Role of metagenomics in prospecting novel endoglucanases, accentuating functional metagenomics approach in second-generation biofuel production: a review. Biomass Conversion and Biorefinery.

[10] Castillo Villamizar GA, Nacke H, Boehning M, Herz K, Daniel R. (2019) Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. MBio.

[11] Ngara TR, Zhang H. (2018) Recent advances in function-based metagenomic screening. Genomics, Proteomics Bioinforma., 16(6):405–15.

[12] Berlemont R, Pipers D, Delsaute M, Angiono F, Feller G, Galleni M, et al. (2011) Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements. Rev Argent Microbiol., 43(2):94–103.

[13] Heath C, Xiao PH, Cary SC, Cowan D. (2009) Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. Appl Environ Microbiol., 75(13):4657–9.

[14] Cieśliński H, Białkowska A, Tkaczuk K, Długołęcka A, Kur J, Turkiewicz M. (2009) Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library. Polish J Microbiol., 58(3):199–204.

[15] Xiao X, Yin X, Lin J, Sun L, You Z, Wang P, et al. (2005) Chitinase genes in lake sediments of Ardley Island, Antarctica. Appl Environ Microbiol., 71(12):7904–9.

[16] Kuhn E, Bellicanta GS, Pellizari VH. (2009) New alk genes detected in Antarctic marine sediments. Environ Microbiol., 11(3):669–73.

[17] Latip MAA, Alias SA, Smykla J, Yusof F, Mohamad MAN, Nordin NFH. (2020) Discovery of cold-active protease from psychrophilic bacteria isolated from Antarctic region for bio-prospecting. Malays Appl Biol., 49:55–60.

[18] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol., 19(5):455–77.

[19] Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, et al. (2009) Community-wide analysis of microbial genome sequence signatures. Genome Biol., 10(8).

[20] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res., 42(D1).

[21] Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. (2017) CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res., 45(D1):D200–3.

[22] Robert X, Gouet P. (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res., 42(W1).

[23] De Lano WL. (2020) The PyMOL molecular graphics system, version 2.3. Schrödinger LLC.

[24] Ekkers DM, Cretoiu MS, Kielak AM, Van Elsas JD. (2012) The great screen anomaly-a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol., 93(3):1005–20.

[25] Thies S, Rausch SC, Kovacic F, Schmidt-Thaler A, Wilhelm S, Rosenau F, et al. (2016) Metagenomic discovery of novel enzymes and biosurfactants in a slaughterhouse biofilm microbial community. Sci Rep., 6.

[26] Devi SG, Fathima AA, Sanitha M, Iyappan S, Curtis WR, Ramya M. (2016) Expression and characterization of alkaline protease from the metagenomic library of tannery activated sludge. J Biosci Bioeng., 122(6):694–700.

[27] Pessoa TBA, Rezende RP, Marques E de LS, Pirovani CP, dos Santos TF, dos Santos Gonçalves AC, et al. (2017) Metagenomic alkaline protease from mangrove sediment. J Basic Microbiol., 57(11):962–73.

[28] Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, et al. (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol., 127(4):575–92.

[29] Vester JK, Glaring MA, Stougaard P. (2014) Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing. Microb Cell Fact., 13(1):72.

[30] Douglass A, O'Brien C, Offei B, Coughlan A, Ortiz-Merino R, Butler G, et al. (2018) Coverage-versus-Length plots, a simple quality control step for de novo yeast genome sequence assemblies. bioRxiv.

[31] Jarocki VM, Reid CJ, Chapman TA, Djordjevic SP. (2020) Escherichia coli ST302: Genomic analysis of virulence potential and antimicrobial resistance mediated by mobile genetic elements. Front Microbiol.

[32] Rawlings ND, Salvesen G. (2013) Handbook of proteolytic enzymes. In: Handbook of Proteolytic Enzymes. Academic Press.

[33] Wang X, Yang X, Yang C, Wu Z, Xu H, Shen Y. (2011) Crystal structure of outer membrane protein NMB0315 from Neisseria meningitidis. PLoS One. 6(10).

[34] Kumar V, Sharma N, Bhalla TC. (2014) In silico analysis of β -galactosidases primary and secondary structure in relation to temperature adaptation. J Amino Acids.

[35] Goldstein RA. (2007) Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: Insights from the quasi-chemical approximation. Protein Sci., 16(9):1887–95.

[36] Metpally RPR, Reddy BVB. (2009) Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. BMC Genomics. 10.

[37] Óskarsson KR, Sævarsson AF, Kristjánsson MM. (2020) Thermostabilization of VPR, a kinetically stable cold adapted subtilase, via multiple proline substitutions into surface loops. Sci Rep., 10(1).

[38] Huber R, Bennett WS. (2013) Functional significance of flexibility in proteins. Pure Appl Chem., 54(12):2489–500.

[39] Janin J, Sternberg MJE. (2013) Protein flexibility, not disorder, is intrinsic to molecular recognition. F1000 Biol Rep., 5(1).

[40] Baer P, Rabe P, Fischer K, Citron CA, Klapschinski TA, Groll M, et al. (2014) Induced-fit mechanism in class I terpene cyclases. Angew Chemie - Int Ed., 53(29):7652–6.

[41] Grabowska M, Jagielska E, Czapinska H, Bochtler M, Sabala I. (2015) High resolution structure of an M23 peptidase with a substrate analogue. Sci Rep., 5.

[42] Odintsov SG, Sabala I, Marcyjaniak M, Bochtler M. (2004) Latent LytM at 1.3 Å resolution. J Mol Biol., 335(3):775–85.

[43] Reva BA, Finkelstein A V., Skolnick J. (1998) What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? Fold Des., 3(2):141–7.

[44] Russell RJM, Gerike U, Danson MJ, Hough DW, Taylor GL. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. Structure. 6(3):351–61.

[45] Aghajari N, Van Petegem F, Villeret V, Chessa JP, Gerday C, Haser R, et al. (2003) Crystal structures of a psychrophilic metalloprotease reveal new insights into catalysis by cold-adapted proteases. Proteins Struct Funct Genet., 50(4):636–47.

[46] Yüksel KÜ, Sun AQ, Gracy RW, Schnackerz KD. (1994) The hinged lid of yeast triose-phosphate isomerase: Determination of the energy barrier between the two conformations. J Biol Chem., 269(7):5005–8.

[47] Chesters C, Wilding M, Goodall M, Micklefield J. (2012) Thermal bifunctionality of bacterial phenylalanine aminomutase and ammonia lyase enzymes. Angew Chemie - Int Ed., 51(18):4344–8.

[48] Siddiqui KS, Poljak A, Guilhaus M, De Francisci D, Curmi PMG, Feller G, et al. (2006) Role of lysine versus arginine in enzyme cold-adaptation: Modifying lysine to homo-arginine stabilizes the cold-adapted α-amylase from Pseudoalteramonas haloplanktis. Proteins Struct Funct Genet., 64(2):486–501.

[49] Povarova OI, Kuznetsova IM, Turoverov KK. (2007) Different disturbances - one pathway of protein unfolding. Actin folding-unfolding and misfolding. Cell Biol Int., 31(4 SPEC. ISS.):405–12.

[50] Vemparala S, Mehrotra S, Balaram H. (2011) Role of loop dynamics in thermal stability of mesophilic and thermophilic adenylosuccinate synthetase: A molecular dynamics and normal mode analysis study. Biochim Biophys Acta - Proteins Proteomics. 1814(5):630–7.

[51] Furhan J, Salaria N, Jabeen M, Qadri J. (2019) Partial purification and characterisation of cold-active metalloprotease by Bacillus sp. Ap1 from Apharwat Peak, Kashmir. Pakistan J Biotechnol., 16(1):47–54.

[52] Leng W, Gao R, Wu X, Zhou J, Sun Q, Yuan L. (2021) Genome sequencing of cold-adapted Planococcus bacterium isolated from traditional shrimp paste and protease identification. J Sci Food Agric., 101(8):3225–36.

[53] Park HJ, Han SJ, Yim JH, Kim D. (2018) Characterization of an Antarctic alkaline protease, a cold-active enzyme for laundry detergents. Korean J Microbiol., 54(1):60–8.

[54] Yang XS, Chen XL, Xu XZ, Zeng RY. (2011) Cold-adaptive alkaline protease from the psychrophilic Planomicrobium sp . 547: enzyme characterization and gene cloning. Adv Polar Sci., 22(1):49–54.

[55] Perfumo A, Freiherr von Sass GJ, Nordmann EL, Budisa N, Wagner D. (2020) Discovery and characterization of a new cold-active protease from an extremophilic bacterium via comparative genome analysis and in vitro expression. Front Microbiol., 11.

[56] Ji XL, Taj MK, Lu XB, Lin LB, Zhang Q, Wei YL. (2014) Purification and characterization of an extracellular cold-active protease produced by the psychrotrophic bacterium Serratia Sp. WJ39. In: Applied Mechanics and Materials. 330–4.

[57] Tariq AL. (2011) Purification and characterization of 56 kDa cold active protease from Serratia marcescens. African J Microbiol Res., 5(32).

[58] Nor NM, Mohamed MS, Loh TC, Foo HL, Rahim RA, Tan JS, et al. (2017) Comparative analyses on medium optimization using one-factor-at-a-time, response surface methodology, and artificial neural network for lysine–methionine biosynthesis by Pediococcus pentosaceus RF-1. Biotechnol Biotechnol Equip., 31(5):935–47.

[59] Mohammed K, Pramod WR. (2011) Production optimization of an extracellular cold-active alkaline protease from Stenotrophomonas maltophilia MTCC 7528 and its application in detergent industry. African J Microbiol Res., 5(7):809–16.

[60] Salwan R, Sharma V, Kasana RC, Gulati A. (2020) Bioprospecting psychrotrophic bacteria for serine-type proteases from the cold areas of western Himalayas. Curr Microbiol., 77(5):795–806.

[61] Paul T, Jana A, Mandal AK, Mandal A, Das Mohpatra PK, Mondal KC. (2016) Bacterial keratinolytic protease, imminent starter for NextGen leather and detergent industries. Sustain Chem Pharm., 3:8–22.

[62] Kanelli M, Vasilakos S, Ladas S, Symianakis E, Christakopoulos P, Topakas E. (2017) Surface modification of polyamide 6.6 fibers by enzymatic hydrolysis. Process Biochem., 59:97–103.

[63] Kaewkrod A, Niamsiri N, Likitwattanasade T, Lertsiri S. (2018) Activities of macerating enzymes are useful for selection of soy sauce koji. LWT - Food Sci Technol., 89:735–9.

[64] Zhao GY, Zhou MY, Zhao HL, Chen XL, Xie B Bin, Zhang XY, et al. (2012) Tenderization effect of cold-adapted collagenolytic protease MCP-01 on beef meat at low temperature and its mechanism. Food Chem., 134(4):1738–44.