Microbial Characterization and Optimization of Protease Production by Bacillus SP. Isolated from Soil

Shruti Gamya Dash  
*P.G. Department of BT and BI, Sambalpur University, At/po- Jyoti Vihar, Burla, Dist- Sambalpur, Pin- 768019, Odisha, India*

Jitesh Kumar Maharana  
*P.G. Department of BT and BI, Sambalpur University, At/po- Jyoti Vihar, Burla, Dist- Sambalpur, Pin- 768019, Odisha, India*

Amiya Kumar Patel  
*P.G. Department of BT and BI, Sambalpur University, At/po- Jyoti Vihar, Burla, Dist- Sambalpur, Pin- 768019, Odisha, India*

Abstract: Microbes can be exploited because of its ability to produce enzymes, which can be used for industrial applications. The bacterium isolated from the milkshed soil sample of dairy plant was found to be Gram negative and rod shaped structure. The study indicated that the bacterium is facultative heterotroph, which exhibited optimal growth and proliferation at 37°C and pH 9 (alkalophilic). Besides, the thermal death time of the bacterium was found to be 4 hrs at 60°C. Moreover, the culture dependent approach based on PCR amplified partial subunit 16S rDNA sequence analysis was used to identify the bacterium. Molecular phylogenetic analysis based on 16S rDNA sequence homology using neighbor-joining method suggested that the isolated bacterium belongs to the genus *Bacillus* with closer affiliation with other microorganisms that have the ability to thrive in such environment. All sequences used in the analyses were obtained from the GenBank. Besides, the alkaline protease sequence of *Bacillus thuringiensis* was used as the reference sequence for homology model building of alkaline protease in order to study its binding affinity with casein protein. The homology structure modeling was performed using molecular operating environment (MOE). Further, the final structure of alkaline protease was evaluated using ERRAT, Verify 3D and Ramchandran plot analysis.

Keywords: 16S rDNA; alkaline protease; *Bacillus* sp.; homology modeling; PCR amplification.

I. INTRODUCTION

Microbial enzymes due to their ecofriendly applications in industrial processes have become the focus of intense research [1]. Microbial proteases are one among the three largest groups of industrially and commercially produced enzymes. Bacteria are the most important producers of alkaline protease with genus *Bacillus* due to its ability to produce large amount of protease having proteolytic activity and can remain stable in high pH and temperature [2]. Microbes produce a wide variety of proteolytic enzymes, which are smaller in size and compact having spherical structure. The alkaline proteases (EC.3.4.21-24, 99) are hydrolytic group of enzymes that catalyze the cleavage of peptide bonds. Proteases (also called as peptidase or proteinase) perform proteolysis through protein catabolism by hydrolysis of peptide bonds that link amino acids through the polypeptide chain. Different classes of proteases can perform the same reaction through different catalytic mechanism [3]. Proteases can be classified into six groups such as aspartate, cysteine, glutamate, metallo, serine and threonine based on their characteristic mechanistic features consistent within each member of a group. Besides, proteases are classified into three major groups depending on their acid base behaviour and activities at different temperatures *i.e.* acid proteases, neutral proteases and alkaline proteases [4]. Moreover, they conduct high specific and selective modifications of protein *i.e.* zymogenic form of enzymes by limited proteolysis, blood clotting and lysis of fibrin clots, processing and transport of the secretory proteins across the membranes. Besides, they also play an important role in the proteolytic steps for tumour invasion or in infection cycle of a number of pathogenic microbes. Their involvement in the life cycle of disease causing organisms has led to become a pharmaceutical agent against fatal diseases such as cancer and AIDS [2].

Microbial proteases are preferred over the enzymes produced either from plant or animal sources [5]. Although several microbes produce proteases, but the immense proportion of commercial alkaline proteases is...
mostly derived from Bacillus sp., because of its ability to secrete large amount of alkaline proteases having significant proteolytic activity, stability at high temperature and wide range of pH [6]. Several investigations have suggested that the alkaline protease producers mostly include the bacterial genera such as Bacillus [7-9] and Pseudomonas [10-11], Streptomyces sp. [12], Conidiobolus sp., Rhizopus sp. [13] and Aspergillus sp. [14]. Therefore, the initial screening for protease production by the microbes is of utmost importance [15]. There is always a chance of finding novel microbes engendering enzyme synthesis with better properties and congruous for the commercial exploitation. However, the multitudes of diverse habitats have triggered the development of equipollently numerous molecular adaptations in microbes leading to microbial diversity. Molecular characterization by 16S rDNA sequence based phylogenetic analysis is adscitiously used to study the microbial diversity with reverence to their ecological distribution [16]. The 16S rDNA gene is the most conserved (least variable) DNA in all living cells, which are utilized in reconstructing phylogenies due to their slow rates of evolution of the gene. Thus, the 16S rDNA sequences from distantly related microorganisms can be precisely aligned to determine the taxonomy, phylogeny (evolutionary relatedness) as well as used to estimate the rate of species divergence among microbes from the diverse habitats.

Besides, the computational study of genes is pre-requisite to determine their structural and physiochemical properties, phylogenetic analyses and simulations to understand the molecular interactions among them. The in silico analyses can facilitate to reach a conclusive decision making in order to conduct the experiments, which are extravagant, time consuming and tedious [17]. The protein molecules as shapeless amino acid strings ultimately fold up into a correct topology of 3D structure that is pre-requisite to become biologically active. Consequently, the 3D structure of a protein provides insight into its functional attributes. Bioinformatics approaches can determine the secondary structure (helices, sheets, domains, strands and coils) in protein sequences, which prognosticate the 3D structure of a protein through comparative modelling [18]. Comparative modelling of protein refers to the construction of the atomic-resolution model of the target protein from its amino acid sequence and an experimental three dimensional structure of a cognate homologous protein (template). The homology modelling relies on the identification of one or more kenned protein structures liable to resemble the structure of query sequence and on the production of alignment that maps residues in the query sequence to residues in template sequence.

The ExPASy (Expert Protein Analysis System) powered by Swiss Institute of Bioinformatics (SIB) is one of the most paramount tool used to determine protein similarity, pattern identification and to study post-translational modifications [19]. In addition, the structure predicted approach was acclimated to predict the novel inhibitors of the protein. The in silico structure prediction based on the comparative modeling of protein molecules is used to understand the molecular basis of the ligand-receptor interactions and to rationalize how the different compounds can promote their biological activities. Keeping in view, the present study was designed to isolate the bacterium from milkshed soil sample of dairy plant followed by the microbial characterization including its growth response. The optimized culture condition for the bacterium to produce higher protease production was determined. Besides, the homology modelling of alkaline protease from the isolated bacterium and the protein-protein docking for prognostication of binding affinity of the modelled structure of alkaline protease with casein protein was performed.

II. MATERIALS AND METHODS

A. Study site and Sampling

The present study was carried out using the milkshed soil sample for bacterial isolation, which was collected from the dairy plant (OMFED) located at Goshala, Sambalpur, Odisha, India (Geographical location: 21°25'56" North latitude and 83°54'7" East longitude). The area receives the wash dairy waste regularly throughout the year. The district experiences temperate climate throughout the year and experiences semi-arid climate (1790 mm rainfall y−1, annual average temp 28°C) with three distinct seasonality i.e. summer (April to June), rainy (July to September) and winter (December to February). Sampling was done from (0-15) cm soil depth by digging pits of 15 cm2 size from five different blocks (in triplicates), which were mixed together to form a ‘composite sample’, aseptically packed in sterilized polypropylene vials and brought to the laboratory. Soil samples were stored at 4°C until analyzed.

B. Isolation of bacteria

The bacterium was isolated using skimmed milk broth (SMB) [Glucose: 10g, Peptone: 5g, Yeast Extract: 5g, KH2PO4: 1g, MgSO4: 0.2g, NaHCO3: 1g, NaCl: 10g, Skimmed milk: 28g, Casein: 5g per litre], which was used for isolation, cultivation and maintenance of protease producing bacteria [16]. When the absorbance reached to ~ 0.2, about 100μl of the culture was streaked on skimmed milk agar (SMA) [Skimmed milk powder: 28g, Casein: 5g, Yeast Extract: 2.05g, Dextrose: 1g, Agar: 15g per litre] and incubated at 37°C for 24hr for the development of colonies. These steps have been repeated in order to obtain the pure culture.
C. Screening of bacterial isolates
Bacterial isolates were screened for protease activity through skimmed milk hydrolysis using skimmed milk agar [16] and incubated at 37°C for 24hr. The protease producing bacterial isolates were screened through the occurrence of clear zone of hydrolysis revealed around the bacterial colony for subsequent microbial characterization.

D. Microbial characterization
The morphological characterization of the isolated bacterium was performed through Gram staining. Bacterial inoculum was taken in a clean sterilized glass slide with the help of an inoculation needle, smeared and heat fixed. Then, one or two drops of crystal violet followed by Gram’s iodine were added. The slide was washed with alcohol and dried followed by the addition of 1/2 drops of saffranin strain. The slide was washed with water, air dried and observed under the microscope. Besides, different biochemical tests were performed such as catalase test, oxidase test, carbohydrate fermentation test, indole production test, methyl red test, citrate utilization test, Voges Proskauer test, H2S production test, starch hydrolysis test, urease production test, nitrate reduction test [20]. Thermal death time (TDT) of the isolated bacterium was determined by inoculating 100µl culture in 5ml of SMB medium and incubated at 60°C for different time intervals. The culture was streaked onto the solidified SMA agar individually and incubated at 37°C for 24hr for the development of colonies.

E. Antibiotic sensitivity test
Antimicrobial sensitivity test was performed by disc diffusion technique following Kirby-Bauer’s method [21]. About 100µl bacterial culture was spread on the solidified SMA agar. Disc of different antibiotics such as Amikacin, Amoxycillin, Azithromycin, Cefixime, Cefotaxime, Chloramphenicol, Ciprofloxacin, Erythromycin, Gentamycin, Kanamycin, Levofloxacin, Norfloxacin, Ofloxacin, Rifampicin, Roxithromycin, Streptomycin and Tetracycline having the same concentration (0.5mg/ml) were placed (in triplicates) individually in each petridish and were incubated at 37°C for 24hr. The degrees of sensitivity contributed by different antibiotics against the isolated bacterium were determined by measuring the diameter of zone of inhibition.

F. Effect of temp and pH on bacterial growth
The growth analysis was performed by inoculating 100µl of 24hr grown bacterial culture in 50ml of SMB medium and incubated at different temperatures (30°C, 37°C, 45°C, 50°C and 60°C) for 48hr in order to determine the optimal temperature required by the bacterium for growth. Growth pattern of the bacterium with respect to different temperatures was determined by taking absorbance at 640nm at different time of incubation. Similarly, the effect of pH on bacterial growth was determined by keeping the optimized temperature constant required for growth. About 100µl of 24hr grown culture was inoculated in 50ml of SMB medium with different pH (7, 8, 8.5, 9 and 10) and bacterial growth was measured by taking absorbance at 640nm at different time intervals.

G. Isolation of Genomic DNA
Genomic DNA was isolated using bacterial genomic DNA isolation kit (Chromous bacterial genomic DNA Spin-50). About 750µl of 1X suspension buffer was mixed with 100mg of bacterial pellet followed by addition of 5µl of RNaseA with intermittent mixing for 5-6 times, and kept at 65°C for 10min. Then, 1ml of lysis buffer was added with intermittent mixing for 5-6 times and kept at 65°C for 15min. The mixture was centrifuged at 13000g and the supernatant was collected in a 2ml vial. Supernatant was loaded onto the spin column (600µl each time) and centrifuged at 13000g for 1 min at room temperature. The content of the collection tube was discarded. Then, 500µl of 1X wash buffer was added to the column and centrifuged at 13000g for 3 min at room temperature. Then, the spin column was placed in a fresh 1.5ml vial followed by the addition of 50µl of warm elution buffer (kept at 65°C) and centrifuged at 13000g for 1min at room temperature. The eluted DNA sample was collected and resolved by 1% agarose gel electrophoresis to estimate the quality and quantity of the template used subsequently for PCR amplification.

H. PCR amplification
PCR amplification of 16S rDNA was performed in combination with the following sets of universal primers to conform the size and approximate quantity of the generated amplicons [22]. The reaction mixture (100µl) was prepared (template DNA: 1µl; forward primer: 400ng; reverse primer: 400ng; dNTPs (2.5mM each): 4µl; 10X Taq DNA polymerase assay buffer: 10µl; Taq DNA polymerase enzyme (3U/µl): 1µl; make up the volume to 100µl with Milli Q water). Universal primers used for PCR amplification are forward primer: 5’-CGGCGTTCACACAWGCHAGYCG-3’ and reverse primer: 5’-GGCGSMTGWGTNCAAGSV-3’. Amplification was performed using MyCycler (Bio-Rad, USA) with initial denaturation at 95°C for 5 min, 35
cyles denaturation at 94°C for 30 sec, annealing at 50°C for 30 sec, elongation at 72°C for 1.30 min and final extension at 72°C for 7 min. The PCR products were stored at 4°C till further analysis. The amplification products were electrophoresed in 1% agarose gel containing ethidium bromide @ 0.5µg ml⁻¹ in TAE buffer for 2hr at 50 Volts. A total of 2.5µl loading buffer (1X TAE, 50% glycerol, 0.25% xylene cyanol) was added to each reaction. After electrophoresis, the gel was observed under UV-transilluminator, documented using GelDoc XR (Bio-Rad) and photographed. The size of the amplicon was determined using 250bp DNA ladder (Bangalore Genei Pvt. Ltd., Bangalore, India) and Quantity One software. To test the reproducibility, the reactions were repeated twice.

I. Gel extraction of PCR products

The amplified product was cut from the agarose gel and kept in a 2ml microcentrifuge tube and weighted. About 3 volumes of gel extraction buffer to 1 volume of gel was added and incubated at 55°C for 5-10 min with intermittent mixing till the complete solubilization of agarose. Then, 1 volume of isopropanol was mixed with the gel extracted solution and loaded on to the spin column (600µl each time) and centrifuged at 13000g for 1 min at room temperature. Then, 500µl of wash buffer was added to the column and centrifuged at 13000g for 3 min at room temperature. The content of the collection tube was discarded after each step. The spin column was placed in a fresh 1.5ml microcentrifuge tube and 15µl of elution buffer was added followed by centrifugation at 13000g for 1 min at room temperature. The purified DNA was collected for further analysis.

J. 16S rDNA sequencing and analysis

The purified amplified 16S rDNA fragment was subjected to sequencing based on the chain termination reaction [23] using ‘BigDye terminator (version 3.1) sequencing Ready Reaction kit’ (PE Applied Biosystems) with automated ABI 3500 XL genetic analyzer (PE Applied Biosystems) using PCR amplification forward primer: 5’- CMGSCVTDACACAWGCHAGYC -3’ and reverse primer: 5’- GGCGSMTGWGTNCAAGSV -3’. The sequencing mixture (10µl) included BigDye terminator ready reaction mix: 4µl, template (100ng/µl): 1µl, primer (10pmol/µl): 2µl, Milli Q water: 3µl. The conditions involved in PCR amplification includes 25 cycles with initial denaturation at 96°C for 5 min, denaturation at 96°C for 30 sec, hybridization at 50°C for 30 sec followed by the final elongation at 60°C for 1.30 min. The 16S rDNA gene sequence of the isolated bacterium was aligned using the ‘JustBio online bioinformatics tool’ (http://www.justbio.com) and assembled into a contiguous chain. Thereafter, the 16S rDNA gene sequence was subjected to homology search using ‘BLAST’ search at NCBI (http://www.ncbi.nlm.nih.gov/) [24]. The representative sequences were retrieved from the biological database and aligned using CLUSTAL W for multiple sequence alignment. The computed alignment was then manually checked and corrected. Subsequently, the resulting sequences were analyzed for chimera using QIIME (Version 1.5) software (http://www.qiime.org). The final sequence of 16S rDNA was deposited in GenBank using BankIt submission tool. Further, the evolutionary distances were computed by using MEGA (Version-7.0) packages [25] with p-distance using Neighbor-joining method [26]. The bootstrap values were calculated from 1000 replications to represent the evolutionary history of the taxa [27] using MEGA (Version-7.0).

K. Protease activity

About 200µl of 24hr grown bacterial culture was inoculated with 60ml of SMB medium and incubated at 37°C for different time intervals. The growth of the bacterium was determined by taking absorbance at 640nm. Then, the bacterial culture was centrifuged at 2500 rpm for 10 min and the supernatant was used for enzyme assay [28]. Protease activity was determined by addition of 0.2 ml of supernatant with 3ml Na₂CO₃ (1.4%) and incubated at room temperature for 30 min. The liberated reducing sugars were estimated by adding 1ml of folin’s reagent (33.3%) till the colour of solution changes to pale blue [29] by taking absorbance at 700 nm against control (tyrosine as standard).

L. Optimization of protease production

Protease production by the bacterium was determined in terms of protease activity using SMB medium at different time of incubation (24hr, 36hr, 48hr, 72hr and 96hr). The effect of pH on protease production was determined by varying pH of the culture medium (7, 8, 8.5, 9 and 10) keeping the optimized incubation time constant. Similarly, the optimal temperature required for maximal protease production by the bacterium was determined by varying incubation temperature (30°C, 37°C, 45°C, 50°C and 60°C) keeping the optimized incubation time and pH of the culture medium constant.

M. Homology model building of alkaline protease

The 16S rDNA gene sequence of the isolated bacterium was subjected to BLAST with NCBI database to find out the highest homologous sequence, which showed 99% sequence similarity with the 16S rDNA sequence of Bacillus thuringiensis. Thus, the 16S rDNA sequence of alkaline protease of Bacillus thuringiensis
was used as the reference sequence for modelling of its structure and binding affinity study with the casein protein. Alkaline protease sequence of *Bacillus thuringiensis* was again put to BLAST analysis with PDB database in order to obtain a template structure. The X-ray crystal structure of alkaline protease (PDB ID: 1KAP, resolution 1.64 Å) from *Pseudomonas aeruginosa*, which showed maximum sequence similarity was used as template for homology structure modelling using MOE (Molecular Operating Environment) software (Version 2008.10). About 10 different structures of alkaline protease were modelled using OPLS 2005 force field and the average of them was taken as the homology modelled structure. The structure was validated using MOE that revealed only 1.39% of amino acid reside outside the allowed region. Further, the ERRAT score (86.617%) and Verify 3D score (greater than 0.2) demonstrated that the modelled structure is acceptable. For the prediction of binding affinity of modelled structure of alkaline protease with the casein protein, the protein-protein docking was performed using ZDOCK. For the purpose, the crystal structure of bovine casein protein was downloaded from the PDB database (PDB ID: 1q6p.1) and used for molecular docking. The topmost complex predicted by ZDOCK was taken and submitted to PATCH DOCK for the prediction of interface region and the binding affinity based on PATH DOCK score.

### III. RESULTS AND DISCUSSION

#### A. Isolation and screening of bacterial isolates

SMB medium was used for the isolation, cultivation and maintenance of protease producing bacteria. The isolated bacterium derives energy by utilizing the organic compounds from the SMB medium for its growth. Bacterial colonies appeared on the SMA agar plates were observed to be 2-3 mm in diameter, lobate, flat and white. About 15 bacterial isolates were subjected to screening for protease production. Out of which 5 isolates were selected for their protease activity on SMA supplemented with casein protein. The bacterial isolate exhibiting relatively highest clear zone of hydrolysis around the colony was selected (Figure 1).

![Figure 1. Isolated bacterial colonies on SMA agar showed clear zone of hydrolysis.](image)

![Figure 2. Gram negative, rod shaped bacterium isolated from milkshed soil of dairy plant.](image)

#### B. Microbial characterization

Microscopic studies revealed that majority of the bacterial isolates in milkshed soil sample were found to be Gram negative, which appeared to be reddish pink in color having rod shaped structure (Figure 2). The biochemical test of the isolated bacterium showed negative response to indole production test, methyl red test, H₂S production test, oxidase test and gas production test from glucose where as it exhibited positive response to Voges Proskauer test, citrate utilization test, urease test, catalase test, starch hydrolysis test and nitrate reduction test. Besides, it is evident from the study that there was an increasing trend of death with the increase in exposure time at 60°C. The study indicated that the thermal death time of the isolated bacterium was found to be 4hr at 60°C, which was found to be moderately thermophilic.

#### C. Antibiotics sensitivity test

The antibiotic sensitivity test reveled that the isolated bacterium was sensitive to amikacin, amoxicillin, azithromycin, cefixime, cefpodoxime, cefotaxime, cephibid, ciprofloxacin, erythromycin, gentamicin, kanamycin, levofloxacin, norfloxacin, norphin, ofloxacin, rifampicin, roxithromycin, streptomycin and tetracycline (Figure 3).
D. Effect of temperature and pH on growth

Growth analysis revealed that the bacterium exhibited relatively higher growth rate at 37°C (Figure 4). The lag phase of the bacterium was continued up to 2nd hr of incubation followed by log phase till 93rd hr of incubation at 37°C. Thereafter, the stationary phase was initiated.

The bacterium exhibited minimal growth rate at 60°C, which may be due to the inhibition in protease production [30-32]. Microbes exhibited higher growth rate at optimum pH. The decrease or increase in medium pH beyond the optimum resulted decrease in microbial growth rate. The study suggested that the isolated bacterium exhibited relatively higher growth rate with medium pH 9 at 37°C (Figure 5).

E. Optimization of incubation time for protease activity

The isolated bacterium was allowed to grow in SMB medium with pH 9 at 37°C for different incubation time. The bacterium exhibited relatively higher bacterial growth rate as well as protease activity at 72 hr of incubation.
The study indicated a gradual increase in protease activity from 24 hr (0.055 ± 0.009 μg tyrosine/g soil/hr) to 72 hr (0.157 ± 0.028 μg tyrosine/g soil/hr) of incubation (Figure 6). The protease activity showed a decline trend at 96 hr of incubation (0.092 ± 0.021 μg tyrosine/g soil/hr) (Figure 6). It is evident from the study that the incubation period beyond the optimum was accompanied by the decrease in growth rate and enzyme productivity, which may be due to the fact that the bacterial cell may reach the decline phase and exhibited less protease synthesis [33-34]. In addition, the decrease in protease activity in later phase of bacterial growth may be due to the catabolic repression mediated by readily metabolizable substrate glucose [35]. The protease production by Bacillus sp., which exhibited maximum activity at 72 hr of incubation, was substantiated by several investigations [36-38].

F. Optimization of pH and temperature for protease activity

Enzyme activity is markedly affected by pH, may be due to binding affinity for substrate and catalysis that are often dependent on charge distribution on both substrate and enzyme molecules [39]. Besides, the medium pH plays an important role in enzyme production and morphological changes in microbes [40].

In the present study, the protease activity showed an increasing trend from pH 7 (0.135 ± 0.014 μg tyrosine/g soil/hr) to pH 9 (0.482 ± 0.027 μg tyrosine/g soil/hr) (Figure 7). The result suggested that the pH of medium induced enzyme synthesis and higher protease activity at pH 9, which may be due to the increased bacterial growth [41]. However, the protease activity showed a decline trend as the medium pH increases beyond pH 9 and was found to be minimum at pH 10 (0.438 ± 0.021 μg tyrosine/g soil/hr). Similar results were also reported for protease production at alkaline pH [42, 43, 44, 45]. The reduction in protease activity may be due to the decrease in bacterial growth rate when the pH of the medium is altered below or above the optimum [40]. Further, the bacterial proteases are synthesized at wide ranges of temperature by regulating not only the
bacterial growth rate but also the production of different metabolites [46, 47, 48, 49]. The Bacillus sp. (B. amyloliquefaciens, B. subtilis, B. licheniformis and B. stearothermophilus) were reported to produce protease at (37-60)°C [50]. Besides, the optimum growth rate and protease production in Bacillus sp. has also been reported at (35-80)°C [51]. In the present investigation, the optimum temperature required by the isolated bacterium for maximal protease activity was determined. The results showed positive correlation between the bacterial growth rate and enzyme activity up to 37°C (0.104 ± 0.0049 μg tyrosine/g soil/hr) followed by a decline trend with the increase in incubation temperature (Figure 8), which may be due to the inhibition in bacterial growth at higher temperature and thereby the enzyme production get inhibited [52]. Several studies have reported maximum protease activity at (37-40)°C in Bacillus sp. [50, 53].

G. PCR amplification of 16S rDNA

Isolation and identification of the bacterium isolated from the milkshed soil sample becomes a challenging mission for the microbial ecologists to unravel microbial diversity associated with their functional attributes. Such investigations provide insight into the microbial community structure and function. The PCR amplification using forward primer (5’-CMGSCVTDACACAWGCHAGYC -3’) and reverse primer (5’-GGCGSMTGWGTNCAAGSV -3’) generated a single band amplicon size of ~1.5 Kb on 1% agarose gel (Figure 9).

![Figure 9. PCR amplification of 16S rDNA gene of the isolated bacterium. (Lane M: marker DNA; Lane 1: represents the amplicon size ~1.5Kb of the bacterial isolate).](image)

<table>
<thead>
<tr>
<th>H. BLAST analysis and sequence homology</th>
</tr>
</thead>
</table>

The amplified 16S rDNA gene products was excised from the agarose gel and subjected to sequencing. The 16S rDNA nucleotide sequence of the isolated bacterium was subjected to homology search using BLAST. The highest degree of homology exhibited by the 16S rDNA sequence indicated by BLAST analysis were represented (Table 1).

The analysis suggested that the 16S rDNA sequence of protease producing bacteria shared 99% sequence identity with 16S rDNA sequence of Bacillus thuringiensis serovar konkukian str. 97-27 (NC_005957.1), Bacillus cereus ATCC 14579 (NC_004722.1) and Bacillus cereus Rock418 (NZ_CM000735.1); 95% sequence identity with 16S rDNA sequence of Bacillus shackletonii strain LMG 18435 (NZ_LIJC01000006.1), Bacillus sp. FJAT-18043 (NZ_LIJX01000003.1), Bacillus sp. FJAT-18043 (NZ_LIJX0100006.1) and Bacillus sp. FJAT-14578 (NZ_AYSE01000017.1). Besides, it exhibited 94% identity with 16S rDNA of Bacillus butanolivorans (NZ_LGYA01000001.1), Bacillus licheniformis ATCC 14580 (NC_006270.3), Bacillus pumilus NJ-M2 (NZ_CP012329.1), Bacillus amyloliquefaciens DSM7 (NC_014551.1) and Bacillus megaterium NBRC 15308 (NZ_CP009920.1). In addition, it shared 93% sequence identity with 16S rDNA sequence of Bacillus boroniphilus JCM 21738 (NZ_BAUW01000204.1), Bacillus sp. FJAT-18017 (NZ_CP012602.1), Lysinibacillus varians GY32 (NZ_CP006837.1) and Bacillus hemicellulosilyticus JCM (NZ_BAUU01000088.1) (Table 1).

Table 1. 16S rDNA sequence homology of the query sequence (KX034406) with closely related 12 subject sequences retrieved from the microbial databases using BLAST analysis.

<table>
<thead>
<tr>
<th>Microbial strain(s)</th>
<th>NCBI accession number</th>
<th>Query coverage</th>
<th>E-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus thuringiensis serovar konkukian str. 97-27</td>
<td>NC_005957.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus cereus ATCC 14579</td>
<td>NC_004722.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus cereus Rock418</td>
<td>NZ_CM000735.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus shackletonii strain LMG 18435</td>
<td>NZ_LIJC01000006.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus sp. FJAT-18043</td>
<td>NZ_LIJX01000003.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus sp. FJAT-18043</td>
<td>NZ_LIJX0100006.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus sp. FJAT-14578</td>
<td>NZ_AYSE01000017.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>Bacillus butanolivorans DSM 18926</td>
<td>NZ_LGYA01000001.1</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
</tr>
</tbody>
</table>
I. Molecular phylogenetic analysis

The 16S rDNA gene sequence obtained from the isolated bacterium after sequencing was subjected to phylogenetic analysis conducted in MEGA 7.0 [25]. The evolutionary history was inferred using the Neighbor-Joining method [26]. The optimal tree with the sum of branch length = 1.20230786 is shown (Figure 10).

![Figure 10. Neighbor-joining phylogenetic tree showing the relatedness between the 16S rDNA sequence of the isolated bacterium and their closest relative sequences retrieved from microbial databases. The numbers in parentheses correspond to the accession number.](image)

The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches [27]. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree (Figure 10). The evolutionary distances were computed using the p-distance method [54] and were expressed in the units of the number of base differences per site. The present study involved 17 nucleotide sequences for phylogenetic analysis of the isolated bacterium. For the purpose, all positions containing the gaps and missing data were eliminated. It is evident from the analysis that there were a total of 1121 positions in the final datasets. The 16S rDNA sequence of the isolated Bacillus sp. from the milkshed soil sample of diary plant has been submitted to NCBI GenBank with NCBI accession No. KX034406.

J. Molecular modeling and binding affinity of alkaline protease

The 16S rDNA sequence of the isolated bacterium showed 99% sequence identity with the 16S rDNA sequence of Bacillus thuringiensis. Therefore, the alkaline protease sequence of Bacillus thuringiensis was used as the reference sequence of isolated bacterium for the homology model building of alkaline protease in order to study the binding affinity with casein protein. The homology structure modeling was performed using MOE software (Version 2008.10). About 10 different structures of alkaline proteases have been modeled using OPLS 2005 force field and the average was taken for homology model structure. The 10 homology modeled structures of alkaline proteases were included along with their RMSD value that varies from 0.3501 to 0.4373 (Table 2), which suggested that the developed model structure is quite robust. The average of 10 predicted structures was generated and used for molecular docking.

Table 2. Database of the homology model of alkaline protease showing 10 intermediates and one final refined model molecule.
<table>
<thead>
<tr>
<th></th>
<th>TrQ6HIZ4/Q6HIZ</th>
<th>Model#</th>
<th>3D-1D Score</th>
<th>2D-2D Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#4</td>
<td>0.4373</td>
<td>0.3124</td>
</tr>
<tr>
<td>5</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#5</td>
<td>0.3501</td>
<td>0.2472</td>
</tr>
<tr>
<td>6</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#6</td>
<td>0.4175</td>
<td>0.3058</td>
</tr>
<tr>
<td>7</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#7</td>
<td>0.3668</td>
<td>0.2575</td>
</tr>
<tr>
<td>8</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#8</td>
<td>0.3732</td>
<td>0.2630</td>
</tr>
<tr>
<td>9</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#9</td>
<td>0.3882</td>
<td>0.2753</td>
</tr>
<tr>
<td>10</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Model#10</td>
<td>0.3780</td>
<td>0.2763</td>
</tr>
<tr>
<td>11</td>
<td>TrQ6HIZ4/Q6HIZ</td>
<td>Homology model</td>
<td>************</td>
<td>************</td>
</tr>
</tbody>
</table>

The final refined model structure of alkaline protease was considered for the evaluation of homology model, which revealed that 90.61% of the residues were in the core region in the Ramachandran map performed by MOE for the analysis of protein geometry structure. It contains the core score > 0.02 and the outlier score < 0.0005. However, only 1.39% of amino acid residues were revealed outside the allowed region (Figure 11).

Figure 11. Ramachandran map of the alkaline protease generated by MOE (Version 2008.10).

Further, the results obtained from Verify 3D showed that about 96.40% of the residues had an average 3D-1D score ≥ 0.2, which revealed that the protein structure is acceptable. Besides, the ERRAT2 score values were used to check the quality of the modulated protein, which estimated the overall quality factor of about 86.617 (Figure 12). The overall quality factor is expressed as the percentage of protein for which the calculated error value falls below 95% rejection limit. The good high resolution structures generally produce the values around 95% or higher. For lower resolution (2.5 to 3Å), the average overall quality factor is around 91%. In the present study, the ERRAT score of 86.617% revealed that the predicted protein structure of alkaline protease is well accepted. Similarly, more than 96.4% of the amino acids with the Verify 3D score ≥ 0.2 revealed that the modeled structure is highly accurate and could be used for the molecular docking experiment.
The recent advances in bioinformatics with the rapid accumulation of 3D structures of protein complexes through X-ray crystallization provide insights into the protein-protein interactions facilitating rational drug design and treatment of diseases. However, all the protein complexes have not been crystallized and therefore various computational techniques have been developed to address such situation.

One of the promising approaches is protein docking where the structure of a complex between two proteins is predicted based on the independently crystallized structures of the components. The Z-DOCK program performs full rigid-body search of the docking orientations between the two proteins, which includes performance optimization and a novel pair-wise statistical energy potential. The ZDOCK searches all the possible binding modes in the translational and rotational space between the two proteins and evaluates each pose using energy based scoring function. The result of the ZDOCK server was then put into PatchDock to prognosticate the binding affinity with casein protein using PatchDock. PatchDock is geometry based molecular docking based on the shape complementarity principles, which aimed at finding the docking transformations that yield good molecular shape complementarity [55]. Such transformations when applied induce both wide interface areas and small amounts of steric clashes. A wide interface is ascertained to include several matched local features of the docked molecules that have complementary characteristics (Figure 13).

**Figure 12.** Overall quality factor of the protein generated by the ERRAT server.

**Figure 13.** PATCHDOCK representing the geometric score, interface area size and desolvation energy of the top 20 scoring solutions. PatchDock algorithm divides the Connolly dot surface representation of the molecules into concave, convex and flat patches [56]. Then, the complementary patches are matched in order to generate candidate transformations. Each candidate transformation is further evaluated by scoring function that considers both geometric fit and atomic desolvation energy [57]. Finally, RMSD (root mean square deviation) clustering is applied to candidate solutions to discard the redundant solutions. The main reason behind the PatchDock's high efficiency is its expeditious transformational search, which is driven by the local features matching rather than the brute force probing of the six-dimensional transformation space. Further, it expedites the computational processing time by utilizing the advanced data structures and spatial pattern detection techniques such as geometric hashing and pose clustering that were originally developed in the field of computer vision [55].
The protein complex with PatchDock score of 13868 with interface area of 2088.8 Å was considered as the probable site of interaction in absence of co-crystal structure between alkaline protease and casein (Figure 15). Conclusively, a complex structure was found, which showed that the alkaline protease binds with the casein protein to facilitate its degradation (Figure 14).

III. CONCLUSION

It is evident from the study that the bacterium isolated from milkshed soil sample is considered to be the good source of alkaline protease production. The isolated bacterium is identified as *Bacillus* sp. based on the morphological, biochemical and molecular characterization. The isolated bacterium was found to exhibit closer resemblance with *Bacillus thuringiensis* based on 16S rDNA sequence comparison. The study revealed that the isolated bacterium exhibited relatively higher growth rate in alkaline medium (pH 9) at 37°C, which was considered to be the favorable conditions for higher protease activity. The sensitivity of bacterium was analyzed through antibiotics sensitivity test. Besides, the thermal death time of the bacterium was determined to be 4 hr at 60°C. Moreover, the molecular modeling suggested that the alkaline protease was being produced by the isolated bacterium has higher binding affinity with casein protein, which facilitate its degradation in soil. The study suggested that the isolated bacterium could be used as an alternative source for the commercial production of alkaline protease for industrial use.

V. ACKNOWLEDGEMENTS

The authors were thankful to the Head, P.G. Department of Biotechnology and Bioinformatics, Sambalpur University, Odisha for providing the necessary laboratory facilities during the study. Further, the help rendered by several persons during sampling, data analysis and interpretation of data in several counts were duly acknowledged.

REFERENCES

K. R. Aneja. “Microbiology: A laboratory manual”. 5th (Eds) Wishwa Prakashan, New Age International Pvt Ltd., New Delhi, pp. 268-


