doi.org/10.36721/PJPS.2020.33.5.SUP.2361-2369.1

Isolation and characterization of *nprB*, a novel protease from *Streptomyces thermovulgaris*

Amna Mushtaq[1], Tariq Mahmood Ansari[1], Ghulam Mustafa[2], Muhammad Aslam Shad[3], Jorge Cruz-Reyes[4] and Amer Jamil[5*]

[1]Institute of Chemical Sciences, Bahauddin Zakariya University, Multan, Pakistan
[2]Department of Biochemistry, Government College University, Faisalabad, Pakistan
[3]Department of Biochemistry, Bahauddin Zakariya University, Multan, Pakistan
[4]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA
[5]Department of Biochemistry, University of Agriculture, Faisalabad, Pakistan

Abstract: Bacterial proteases are of great pharmaceutical importance and have a key role in various biological processes and in life cycle of several pathogens. New technology used for rational protein engineering as well improved delivery options will expand the potential pharmaceutical applications of proteases. The catalytic proteases belong to metalloproteases (EC.3.4.24) that comprise thermo lysine. The metalloproteases and their homologs have many important biotechnological and therapeutic applications. In the present study, a novel protease gene *npr*B was isolated from a thermophilic bacterium *Streptomyces thermovulgaris* and bioinformatics analyses were performed. PCR amplification and sequencing of *npr*B gene indicated an open reading frame of 178 aa (20191.18 Dalton). Based on protein sequence homology as well as conserved motifs and PTF domain the protein is characterized as a thermo lysine-like protease and is a member of M4 family of metalloproteases. Different bioinformatics tools such as ProtParam, SOPMA, signalP4.1 and ProDom from the ExPAsy server were used for structural and functional analyses. A phylogram was also reconstructed to reveal evolutionary relationships of nprB with its various homologs. The provided data will serve as a background to further reveal pharmaceutical and biotechnological importance of this novel protease gene from *S. thermovulgaris* in future.

Keywords: Protease; thermophile; *Streptomyces thermovulgaris*; proteolytic enzyme.

INTRODUCTION

The pharmaceutical applications of proteases have been grown rapidly. These are an expanding class of drugs with great promise. The predominant uses of proteases which have been applied include tumour cell killers, against digestive disorders, useful agents for the treatment of skin problems, inflammation and other diseases (Abkouhi *et al*., 2019). The proteases available in market are of the microbial origins because they are of high yield, less consumption time, less space requirement, genetic manipulation and cost-effective which have made them suitable for various industrial, biotechnological and pharmaceutical applications (Ali *et al*., 2016).

Proteolytic enzymes are ubiquitous and essential for various biological processes such as cell growth, cell signalling, proliferation, differentiation and immune response (Patil and Jadhav, 2017). Proteases have also industrial importance and approximately 65% of total enzyme market is contributed by this class of enzymes (de Oliveira *et al*., 2018). These enzymes are distributed in nature and contribute in a wide range of biological functions. Proteases have a great number of applications in various industries including food, textile, detergent, silk, pharmaceutical and leather industries (de Oliveira *et al*., 2018).

A variety of organisms including invertebrates, vertebrates, fungi, plants and micro organisms have been utilized for the production of various products with pharmaceutical applications (Arif *et al*., 2020; Bukhari *et al*., 2020; Munir *et al*., 2020). Micro organisms contribute for two third commercial productions of proteases worldwide. According to an estimate, microbial proteases constitute around 40% of the total sale in the world enzyme market (Ibrahim *et al*., 2015). Microbial sources are predominately commercialized because they are not vulnerable to temperature and pH fluctuations and show tolerance towards different organic solvents and detergents (Kousar *et al*., 2013; Mustafa *et al*., 2016).

Thermo-stable enzymes are getting more and more attention nowadays as they show resistance to EDTA, pH, temperature, organic solvents and oxidizing agents (Shaheen *et al*., 2017). Inactivation of enzymes is commonly caused by thermal instability. Commercial applications of proteases would be increased by increasing their stability. Proteases with thermo-active properties have advantages in many ways, for example, a higher temperature could be applied during the process of acceleration of reaction rate, to increase nongaseous reactants solubility and also to minimize the chances of

*Corresponding author: e-mail: profamerjamil@yahoo.com

microbial contamination due to mesophilic organisms during the production of proteases (Ifandi and Alwi, 2018). *Bacillus* species which are industrially very important microorganisms can produce varieties of extra-cellular enzymes including proteases (Josephine *et al*., 2012). Several species of *Bacillus* for example *B. stearothermophilus, B. mojavensis*, *B. pumilus*, *B. licheniformis*, *B. lehnsis*, *B. subtilis*, *B. pseudofirmus* and *B. vallismortis* have been well characterized (Qureshi *et al*., 2018).

Streptomyces belong to actinomycete class and are important micro organisms because of their abilities of producing numerous molecules and growing in variety of habitats with different environmental conditions such as temperature and pH (Prasad *et al*., 2013; Mustafa and Jamil, 2017). However reports on thermophilic species capable of producing protease is lacking in literature. Therefore, in current study, isolation and characterisation of a novel protease gene *nprB* from a thermophilic species *Streptomyces thermovulgaris* were performed. The gene was analysed for its various properties that would be useful to reveal further uses of protease enzyme in future.

MATERIALS AND METHODS

Growth conditions and DNA extraction

Thermophilic bacterium *Streptomyces thermovulgaris* NBRC 12383 was maintained on Bennett's agar medium. The growth medium consisted of yeast extract 1 gL^{-1}, beef extract 1 gL^{-1}, N-Z-amine or peptone from casein 2 gL^{-1} and glucose 10 gL^{-1}. The pH of the medium was adjusted to 7.3. The cultivation was carried out aerobically at 45^{o}C and culture was harvested after 24 hrs. To extract genomic DNA, the bacterial strain was grown on malt extract or yeast extract medium. The DNA was isolated using the method described by Griffiths *et al*. (2000).

Amplification of protease gene

Two sets of primer pairs were designed using multiple NCBI primer designing tools. The sequences of primer pairs used are given in table 1. The PCR was run in a thermocycler (Bio Rad) for 30 cycles with following conditions: denaturation for 30 sec at 94^{o}C, annealing for 1 min at 56^{o}C, extension for 1 min at 72^{o}C followed by one step of extension for 10 min at 72^{o}C. Agarose gel (1%) was used to visualize the amplified products and the amplicons were stored at -20^{o}C for further use.

Sequencing of nprB

The sequence of PCR amplified protease gene was determined using the facility of DNA sequencing from the Centre for Applied Molecular Biology (CAMB), University of the Punjab, Lahore Pakistan.

Characterisation of nprB

In silico analyses were done using nucleotide and protein sequences of nprB to explore its structural, functional and phylogenetic relationships with its homologs from various bacterial species. Protein sequence was analysed using PSI-BLAST search program (Altschul *et al*., 1990) provided by NCBI. The Geneious was used to perform multiple sequence alignment (Josephine *et al*., 2012).

Physicochemical properties of nprB

The nucleotide sequence of *nprB* gene was translated into protein using online translator tools. The physicochemical properties of *nprB* such as molecular mass, instability index, theoretical pI, aliphatic index, grand average of hydropathicity (GRAVY) and extinction coefficient were studied using Prot Param from EXPASy server (Gasteiger *et al*., 2003). Predict Protein server was also used to predict protein binding regions (Yachdav *et al*., 2014).

Phylogenetic analysis

Along with nprB protein sequences of *S. thermovulgaris* thirty two homologous sequences from different bacterial species were also retrieved from protein database for phylogenetic systematics. All sequences were aligned using MUSCLE and imported into the MEGA5 program (Tamura *et al*., 2011) for manual alignment. A phylogenetic tree was reconstructed through Maximum Likelihood (ML) method using MEGA5 with 100 bootstrap replicates.

3D model prediction of nprB

Homology modeling approach was used to predict 3D structure of novel nprB. No suitable template was found from protein databank (PDB) therefore, total five templates were tried to predict 3D structure of nprB using Swiss-Model server (Waterhouse *et al*., 2018). The predicted 3D model was further evaluated using ProSA-web (Wiederstein and Sippl, 2007), ERRAT (Messaoudi *et al*., 2013) and PROCHECK Ramachandran plots (Lovell *et al*., 2002). Furthermore, RMSD, superimposition of template and predicted model and the visualization of 3D model were done using UCSF Chimera 1.5.3 workbench (Pettersen *et al*., 2004).

STATISTICAL ANALYSIS

Different online tools and software were used for statistical analyses. Physicochemical properties were predicted using ProtParam, PSI-BLAST was used to find out homologs of query protein, MSA was performed using Geneious, and MEGA was used for phylogram reconstruction. The 3D structure was predicted using Swiss-Model and the evaluation of the predicted model was done using ProSA-web, ERRAT and PROCHECK Ramachandran plot. The UCSF Chimera was used for model visualization and RMSD prediction.

RESULTS

Isolation and quantification of genomic DNA

The genomic DNA was isolated from a thermophilic bacterium *Streptomyces thermovulgaris* and quantified

Table 1: Nucleotide sequence of the primer pairs used to amplify *nprB* gene

Primer designation	Sequence (5′– 3′)
P1- F1	GATCGGCTGGCCAGAATAG
P1- R1	AGGCGAGATCATATTCACCG
P2- F2	GCAAAGAAATAGTCAGCGGC
P2- R2	CATGCCGGTCCAGGTACA

Table 2: Physicochemical properties of the protease predicted by Prot Param

Sr. No.	Property	Value
1	Molecular mass	20191.18 Dalton
2	Theoretical pI	9.42
3	Instability index	25.97
4	Aliphatic index	72.92
5	Extinction coefficient	11,920 M^{-1} cm^{-1} at 280 nm
6	Grand average of hydropathicity (GRAVY)	-0.551

Table 3: List of templates used to predict 3D structure of protease from *Streptomyces thermovulgaris*

S. No.	Template	Sequence identity	Coverage	Description
1	4k90.1.B	12.24%	55-162	Extracellular metalloproteinase mep
2	4exr.1.A	9.47%	56-166	Putative lipoprotein
3	6cp0.1.A	9.72%	54-134	SdcA
4	3rfr.1.A	16.36%	82-139	PmoB
5	1ups.1.A	18.37%	18-52	GlcNAc-α-1,4-gal-releasing endo-β-galactosidase

using a spectrophotometer. The DNA was in very good condition.

nprB gene manipulation through PCR

The quantified genomic DNA was used for the amplification of *nprB* gene using two sets of primers. Both the primer sets amplified the protease gene in appreciable amount (fig. 1). The amplified protease gene was sequenced which showed 792 base pairs. The amplified and sequenced gene was deposited in GenBank as neutral protease B (*nprB*) gene and got an accession number (i.e. Gen Bank: KX879552).

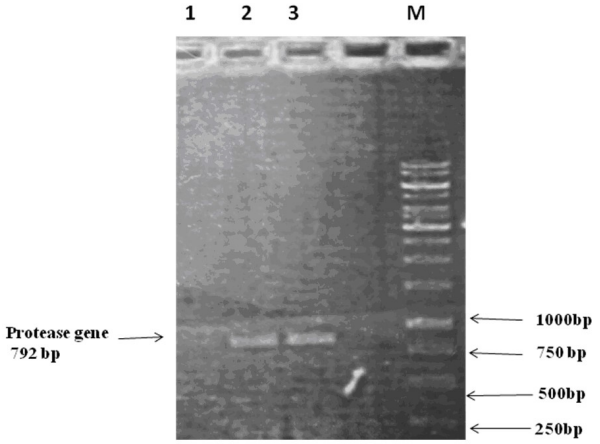

Fig. 1: PCR amplification showed about 800 nucleotide base pairs sequence of *nprB* gene.

In silico analyses of nprB protein

Physicochemical properties of protease protein by ProtParam tool are given in Table 2. It was found that the protease had a molecular mass of 20191.18 Daltons and pI 9.42 which showed that it was a basic protein. The value of instability index was found to be 25.97 which showed that protease is a stable protein. GRAVY value of protease was found to be negative that means it is hydrophilic in nature. The half-life of protease was estimated in mammalian reticulocytes (*in vitro*) 30 hours, in yeast (*in vivo*) >20 hours and in *Escherichia coli* (*in vivo*) >10 hours. Some more important structural annotations of the protease were predicted using Predict Protein (fig. 2). Seven protein binding regions were predicted in the protease by Predict Protein.

Structural annotation of protease

Complete structural annotation of protease protein is given in fig. 2. Protein binding regions are shown by pentagons. Coloured rectangle is showing FTP domain. Start codon is shown in red. Location of important residues has been shown below one letter codes of amino acids.

Protease phylogeny

For the identification of similar local regions and statistically significant protein sequences of protease the PSI-BLAST was used. Multiple sequence alignment was also performed through Geneious (Kearse *et al*., 2012). The truncated sequences were deleted during multiple

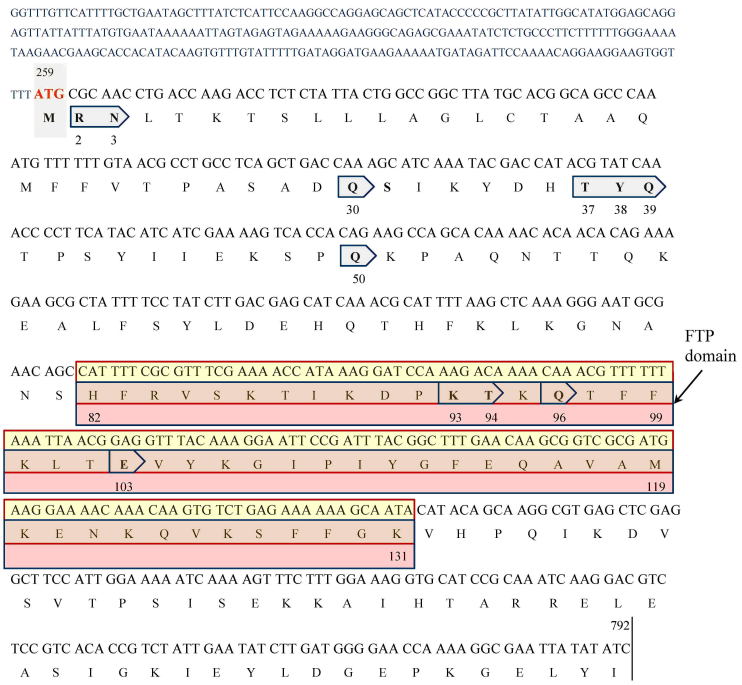

Fig. 2: Structural annotation of nprB protein sequence

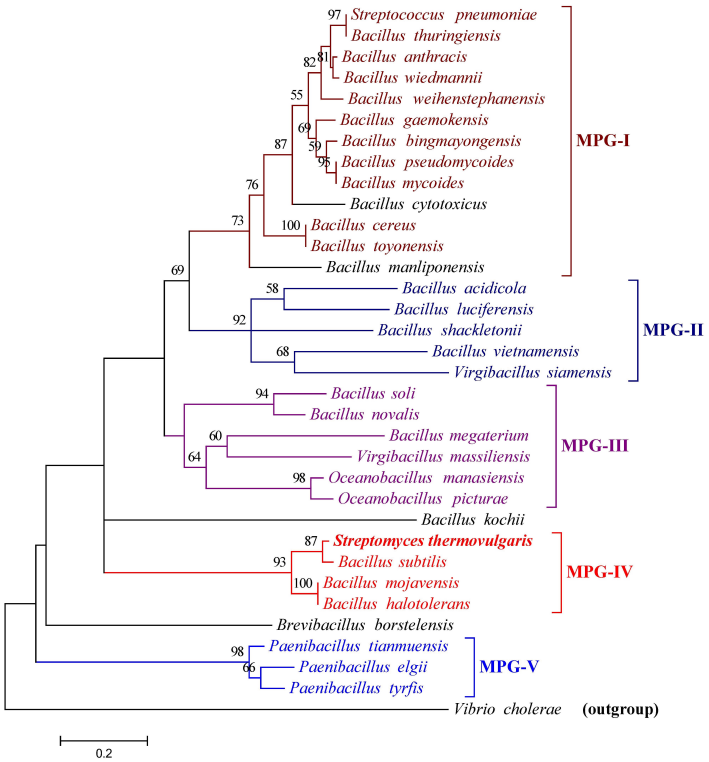

Fig. 3: Phylogeny of proteases from selected organisms using Maximum Likelihood method. Poisson correction model (Zuckerkandl *et al*., 1965) was used to generate the phylogram. To represent evolutionary history of selected taxa, the bootstrap consensus phylogram has been inferred from 100 replicates (Felsenstein 1985). All the branches corresponding to partitions which replicated less than 50% were collapsed. Bootstrap values (100 replicates) have been shown above the nodes. The tree is drawn to scale with branch lengths measured in the number of substitutions per site. Total 34 amino acid sequences were taken for analysis in which all positions that contain gaps and missing data have been removed. The analysis was done in MEGA5 software (Tamura *et al*., 2011).

sequence alignment and shorted the longer sequences in order to make all the sequences equal in length. Molecular phylogenetic analysis was performed to show evolutionary relationships of protease from selected organisms (fig. 3). A protease from *Vibrio cholerae* was used to root the ML tree.

Phylogenetic tree was divided into five different monophyletic groups (MPG-I to MPG-V) (fig. 3). The protein sequence of protease from *S. thermovulgaris* has appeared in MPG-IV and showing its evolutionary closeness with *B. subtilis*, *B. mojavensis* and *B. halotolerans*. Protease of *Streptococcus pneumoniae* has appeared in MPG-I with *Bacillus* species and revealed that protease of *S. pneumoniae* has wide evolutionary divergence from *S. thermovulgaris*. Two *Bacillus* species (i.e. *B. shackletonii* and *B. kochii*) could not be resolved completely in the ML tree and both have been generated from multifurcating nodes. The protease of *Brevibacillus borstelensis* has not joined any monophyletic group and appeared as a separate branch showing that it is very distantly related with other bacterial species.

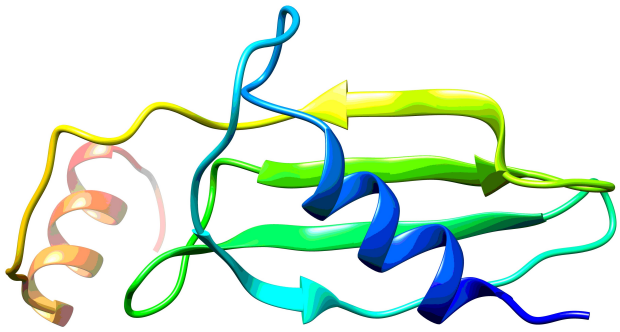

Fig. 4: Predicted 3D structure of novel protease from *Streptomyces thermovulgaris*

Poisson correction model (Zuckerkandl *et al*., 1965) was used to generate the phylogram. To represent evolutionary history of selected taxa, the bootstrap consensus phylogram has been inferred from 100 replicates (Felsenstein 1985). All the branches corresponding to partitions which replicated less than 50% were collapsed. Bootstrap values (100 replicates) have been shown above the nodes. The tree is drawn to scale with branch lengths measured in the number of substitutions per site. Total 34 amino acid sequences were taken for analysis in which all positions that contain gaps and missing data have been removed. The analysis was done in MEGA5 software (Tamura *et al*., 2011).

Homology modelling of novel protease

Being a novel protein from *Streptomyces thermovulgaris*, no suitable template was found from PDB therefore, total five templates were tried to predict 3D structure of protease protein (table 3). The percentage of sequence identity between target and templates was ranged from 9.47% to 18.37% only. The selected templates have been arranged according to query coverage in descending order. The description of each template is also given in the table.

Structure evaluation and superimposition

The evaluation of the predicted model is an important step to guess the accuracy levels and reliability of the model for further applications (Mustafa *et al*., 2018; Hameed *et al*., 2020). The Z-score of predicted protease model was found to be -2.85 (fig. 5a) which is in the range of values of proteins of related sizes. The overall quality factor predicted by ERRAT was 87.097 (fig. 5c). Generally accepted range for a predicted model is >50 for high quality models. ProCheck was used to evaluate the stereochemistry of backbone Psi and Phi dihedral angles through Ramachandran's plot calculations. The percentage of residues of protease occupying most favoured regions (A,B,L) was found to be 86.7% while 12.2% occupied additional allowed regions (a,b,l,p), 1.0% and 0.0% resides were found in generously allowed (~a,~b,~l,~p) and disallowed regions, respectively (Fig. 5d). On the basis of these results it was revealed that the predicted protease model is of good quality.

The predicted protease model was also superimposed with the best template (i.e. 4k90) on the basis of their coverage (fig. 6). The root-mean-square deviation plays an important role in structural comparisons and therefore it is the most popular estimator of structural similarity (Mustafa *et al*., 2020). The overall root-mean-square deviation (RMSD) between protease and its template was found to be 0.996 that has come well under the expected range of <2 Å and showing the accuracy and reliability of the predicted model.

DISCUSSION

Scientists have successfully found wide uses of proteases in the medical field with the passage of time. In medicines, different formulas such as non-woven tissues, gauze and of ointment composition alkaline proteases produced by *B. subtilis* have showed promising therapeutic properties. The deficiency of certain lytic enzyme syndromes has been diagnosed that aided by the oral administration of alkaline proteases (Awad *et al*., 2013). Fibrin degradation by fibrinolytic proteases has been achieved that suggests its future applications as anticancer drug and also as a thrombolytic therapy. The collagen hydrolysis by proteases liberates peptides of low molecular weight without release of any amino acid for the therapeutic use. Similarly, proteases have been used for the treatment of various diseases including carbuncles, burns, wounds, furuncles and for the preparation of elastoterase that is immobilized on the bandage (Abkouhi *et al*., 2019).

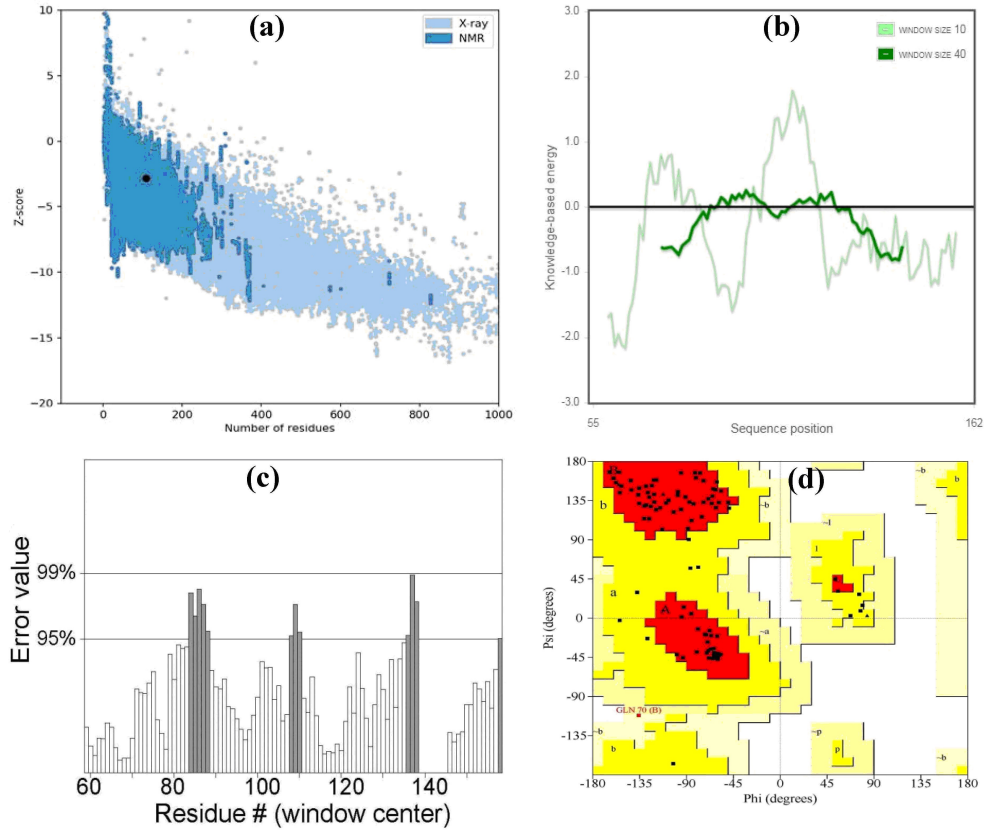

Fig. 5: Evaluation of predicted 3D model of protease from *Streptomyces thermovulgaris.* (a) Z-score plot showing the quality of predicted model, (b) Energy plot showing the positions of all residues of predicted model, the plot is presented with window size 10 and 40, (c) ERRAT plot for the protease model, (d) Ramachandran plot analysis of the modeled structure.

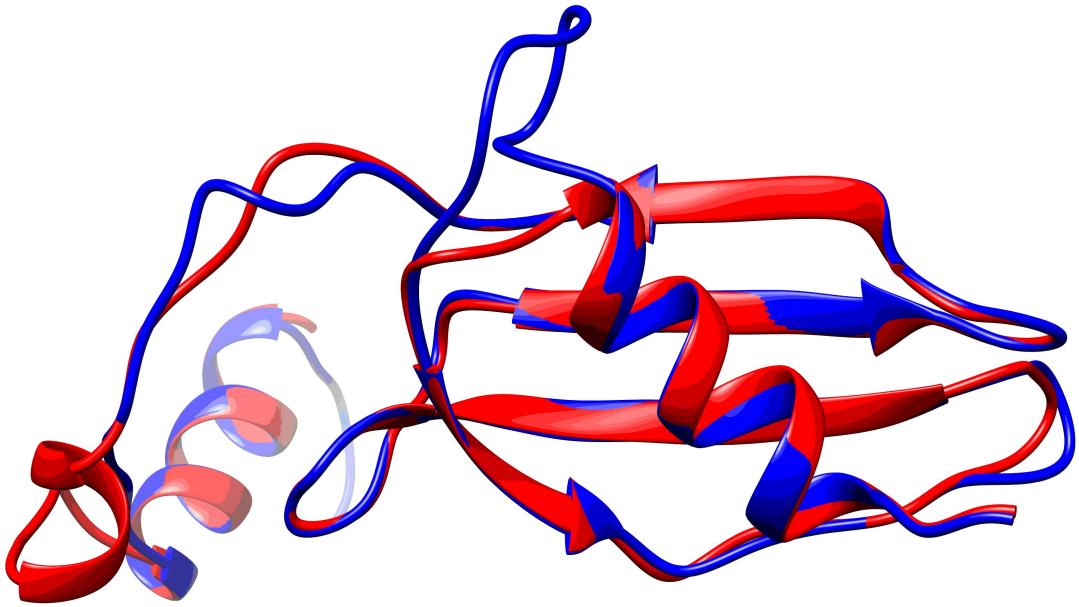

Fig. 6: Structural superimposition of protease (red) with its template (blue). Most of the regions of protease have perfectly superimposed with the template (4k90)

The fundamental properties of enzymes are determined by their amino acid sequences. The region within a single protein possesses a special biological interest is termed as a domain. A fungalysin/ thermolysin propeptide (FTP) domain was found in nprB protein. The FTP domain is found in both fungal M36 propeptide and bacterial M4 peptidase (Tang *et al*., 2003). The exact functioning of this domain is not much clear but it may prevent premature activation either to inhibit peptidase or has chaperone-like activity. Both roles have been attributed to the M4 and M36 propeptides (Markaryan *et al*., 1996).

The isoelectric point (pI) is important to separate proteins on polyacrylamide gel through isoelectric focusing. The values of aliphatic index, instability index and grand average of hydropathicity (GRAVY) play important roles in the stability of proteins (Mustafa *et al*., 2017). The pI value of novel protease was found to be 9.42 which revealed that the query protein is basic in nature. In a similar study, Jabalia *et al*. (2015) also showed that proteases from *Pandoravirus inopinatum* and *Ancylostoma duodenale* are basic in nature as their pI values were found to be more than 7. The values of pI of proteases from *Bacillus cereus*, *Legionella pneumophila*, *Gregarina niphandrodes*, *Dictyocaulus viviparous*, *Ancylostoma duodenale*, *Pandoravirus inopinatum* and *Capsaspora owczarzaki* were found to be less than 7 therefore these proteases were characterized as acidic in nature. The *in silico* predicted pI would be useful to develop buffer systems for the purification of those enzymes using isoelectric focusing method (Pergande and Cologna, 2017). The ProtParam available at ExPASy server was also used to predict extinction coefficient of novel protease and it was found to be 11,920 M^{-1} cm^{-1} at 280 nm with respect to Cys and Tyr concentrations. The low extinction coefficient indicates presence of low concentration of aromatic amino acids in the query protein. In a study, the extinction coefficient of *Capsaspora owczarzaki* protease was found to be high that showed high concentration of aromatic amino acids (Jabalia *et al*., 2015). The predicted extinction coefficients and protease concentrations could serve in the quantitative studies of protein-ligand and protein-protein interactions in different solutions.

The relative volume of a protein which is occupied by aliphatic side chains is termed as aliphatic index and is considered as a positive factor for an enhanced thermal stability of globular proteins. The value of aliphatic index of novel protease sequence was found to be 72.92, and this high value indicates that the query protein would be stable over a wide range of temperature (Verma and Singh, 2013). The value of grand average hydropathy (GRAVY) for a protein or peptide is obtained as a sum of hydropathy values of all amino acid residues divided by the total number of amino acids found in the sequence of that protein/peptide (Bansal *et al*., 2014). The GRAVY value for the query protease sequence was found to be -0.551 and the low value exhibited better interactions with water.

Phylogenetic methods have been getting more and more intention for comparative analysis of DNA and protein sequences with rapid accumulation of molecular sequences (Jabbir *et al*., 2019). The evolutionary rates of genes and proteins are being estimated using phylogenetic methods to follow footprints of natural selection (Bukhari *et al*., 2018). All the sequences of different proteases used in the reconstruction of a phylogram showed that they had originated from a common ancestor and then divided into various monophyletic groups. In a similar study (Jabalia *et al*., 2015), the sequences of papain family cysteine proteases were used to reconstruct a phylogenetic tree and revealed that all proteases had evolved from a common ancestor and also during the time of evolution they were diverged further into different sub groups.

The isolation and characterization of a novel protease gene, *nprB* from a thermophilic bacterium *Streptomyces thermovulgaris* NBRC 12383 have been reported in this study. *Streptomyces* species mostly have the ability to produce multiple proteases with specific substrates however the proteases derived from *Streptomyces* species are not well studied. Interestingly, most of the reported extracellular proteases of thermophilic bacteria are serine or neutral metallo proteases (Da Silva, 2017). *S. thermovulgaris* produces both metallo and serine type of enzymes. The metallo proteases of this species have not been well characterized but on the basis of collected data it can be suggested that this protease is similar to metallo proteases that have been isolated from *Streptomyces* species. Metallo proteases of M4 family have great similarity with various Zn^{2+} containing metallo proteases with two histidine residues of highly conserved zinc binding motif HEXXH. The active site residue was glutamic acid (Seddigh *et al*., 2016).

CONCLUSION

A novel metallo protease, *nprB* was isolated from a thermophilic bacterium *Streptomyces thermovulgaris* NBRC 12383. The amino acid sequence of novel nprB protein was characterised to exhibit its functional and physicochemical properties and phylogenetic analysis using different *in silico* approaches. The analysis of primary structure showed that nprB is hydrophilic in nature and expected to be stable over a wide temperature range. The phylogram revealed that different bacterial proteases have been evolved from a common ancestor and they all are evolutionary related proteins. The biochemical properties showed that it is a unique thermostable protease from thermolysine M4 family. The properties of this enzyme make it promising for basic research and industrial applications. The current study would provide

an insight into various functions of this novel protease through its physicochemical properties which would further help in the formulation of its uses not only in various industries but also in academics.

ACKNOWLEDEMENT

This research was supported by Higher Education Commission (HEC), Government of Pakistan.

REFERENCES

Abkouhi TN, Yazdi MT, Mortazavi FSA and Edalatian MR (2019). Identification of alkaline protease producing Bacilli from sludge of bactofuge and separator using culture–based and molecular techniques. *J. Agric Sci Technol*., 21(7): 1819-1830.

Ali N, Ullah N, Qasim M, Rahman H, Khan SN, Sadiq A and Adnan M (2016). Molecular characterization and growth optimization of halo-tolerant protease producing *Bacillus subtilis* Strain BLK-1.5 isolated from salt mines of Karak, Pakistan. *Extremophiles*, **20**(4): 395-402.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215**(3): 403-410.

Arif R, Ahmed S and Mustafa G (2020). *In silico* study to reveal annotation and significant interactions of human defensin with its isoforms and their phylogeny. *Indian J. Pharm. Sci*., **82**(3): 465-471.

Awad HM, Mostafa ESE, Saad MM, Selim MH and Hassan HM (2013). Partial purification and characterization of extracellular protease from a halophilic and thermotolerant strain *Streptomyces pseudogrisiolus* NRC-15. *Indian. J. Biochem. Biophys*., **50**(4): 305-311.

Bansal H, Srivastava S, Chaurasia A and Jabalia N (2014). A comparative study of antifreeze proteins from *Antarctomyces psychrotrophicus* and *Typhula ishikariensis* using computational tools and servers. *VIVECHAN Int. J. Res*., **5**(1): 21-28.

Bukhari SA, Salman M, Numan M, Javed MR, Zubair M, Mustafa G (2020). Characterization of antifungal metabolites produced by *Lactobacillus plantarum* and *Lactobacillus coryniformis* isolated from rice rinsed water. *Mol. Biol. Rep*., **47**(3): 1871-1881.

Bukhari SA, Tahir MF, Akhter N, Anjum F, Anwar H and Mustafa G (2018). Phylogeny and comparative modeling of phytochelatin synthase from *Chlorella* sp. as an efficient bioagent for detoxification of heavy metals. *J. Biol. Reg. Homeos. Ag*., **32**(5): 1191-1197.

Da Silva RR (2017). Bacterial and fungal proteolytic enzymes: production, catalysis and potential applications. *Appl Biochem Biotechnol*., 183(1): 1-19.

de Oliveira TB, Gostincar C, Cimerman NG and Rodrigues A (2018). Genome mining for peptidases in heat-tolerant and mesophilic fungi and putative adaptations for thermostability. *BMC Genomics*, **19**(1): 152.

Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**(4): 783-791.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*., **31**(13): 3784-3488.

Griffiths RI, Whiteley AS, O'Donnell AG and Bailey MJ (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA and rRNA based microbial community composition. *Appl. Environ. Microbiol*., **66**(12): 5488-5491.

Hameed U, Khalid S, Javed S and Mustafa G (2020). Importance of cytochrome oxidase subunit 1 in phylogeny reconstruction of various classes of subphylum vertebrata. *Pure Appl. Biol*., **9**(3): 1995-2006.

Ibrahim ASE, Al-Salma AA, Elbadawi YB, EI-Tayeb MA and Ibrahim SSS (2015). Production of extracellular alkaline protease by new halotolerant alkaliphilic *Bacillus* SP. NPST-AK15 isolateted from hyper saline soda lakes. *Electron. J. Biotechnol*., **18**(3): 236-241.

Ifandi S and Alwi M (2018). Isolation of thermophilic bacteria from Bora hot springs in central Sulawesi. Biosaintifika: J Biol Biol Educ., 10(2): 291-297.

Jabalia N, Bansal H, Mishra PC and Chaudhary N (2015). *In silico* comparative analysis of papain family cysteine protease using computational tools and servers. *Int. J. Basic and Appl. Eng. Res*., **2**(5): 310-314.

Jabbir F, Irfan M, Mustafa G and Ahmad HI (2019). Bioinformatics approaches to explore the phylogeny and role of BRCA1 in breast cancer. *Crit. Rev. Eukaryot. Gene Expr*., **29**(6): 551-564.

Josephine FS, Ramya VS, Neelam D, Ganapa SB, Siddalingeshwara KG, Venugopal N and Vishwanatha T (2012). Isolation, production and characterization of protease from *Bacillus* sp. isolated from soil sample. *J. Microbiol. Biotech. Res*., **2**(1): 163-168.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C and Thierer T (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*., **28**(12): 1647-1649.

Kousar S, Mustafa G and Jamil A (2013). Microbial xylosidases: production and biochemical characterization (mini-review). *Pak. J. Life Soc. Sci*., **11**(2): 85-95.

Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS and Richardson DC (2002). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins,* **50**(3): 437-450.

Markaryan A, Lee JD, Sirakova TD and Kolattukudy PE (1996). Specific inhibition of mature fungal serine

proteinases and metalloproteinases by their propeptides. *J. Bacteriol*., **178**(8): 2211-2215.

Messaoudi A, Belguith H and Hamida JB (2013). Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. *Theor. Biol. Med. Model*., **10**(1): 22.

Munir N, Mahmood Z, Yameen M and Mustafa G (2020). Therapeutic response of *Epimedium grandiflorum*'s different doses to restore the antioxidant potential and reproductive hormones in male albino rats. *Dose-Response*, **18**(3): 1-13.

Mustafa G and Jamil A (2017). Comparative analyses of gene clusters and Ksα genes involved in the biosynthesis of chromomycin A_3 and mithramycin. *Indian J. Pharm. Sci*., **79**(5): 707-714.

Mustafa G, Arif R, Bukhari SA, Ali M, Sharif S and Atta A (2018). Structural and functional annotation of citrate synthase from *Aspergillus niger* ANJ-120. *Pak. J. Pharm. Sci*., **31**(2): 709-717.

Mustafa G, Iqbal MJ, Hassan M and Jamil A (2017). Bioinformatics characterization of growth differentiation factor 11 of *Oryctolagus cuniculus*. *J. Chem. Soc. Pak*., **39**(6): 1089-1095.

Mustafa G, Kousar S, Rajoka MI and Jamil A (2016). Molecular cloning and comparative sequence analysis of fungal β-xylosidases. *AMB Express*, **6**(1): 30.

Mustafa G, Majid M, Ghaffar A, Yameen M, Samad HA and Mahrosh HS (2020). Screening and molecular docking of selected phytochemicals against NS5B polymerase of hepatitis C virus. *Pak. J. Pharm. Sci*., **33**(5): 2317-2322.

Patil RC and Jadhav BL (2017). Isolation and characterization of protease producing *Bacillus* species from soil of dairy industry. *Int. J. Curr. Microbiol. App. Sci*., **6**(6): 853-860.

Pergande MR and Cologna SM (2017). Isoelectric point separations of peptides and proteins. *Proteomes*, **5**(1): 4.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem*., **25**(13): 1605-1612.

Prasad P, Singh T and Bedi S (2013). Characterization of the cellulolytic enzyme produced by *Streptomyces griseorubens* (Accession No. AB184139) isolated from Indian soil. *J. King Saud Univ. Sci*., **25**(3): 24-250.

Qureshi AS, Simair AA, Ali CH, Khusha I, Khokhar JA, Ahmad A, Danish M and Lu C (2018). Production, purification and partial characterization of organo-solvent tolerant protease from newly isolated *Bacillus* sp. BBXS-2. *Ferment. Technol*., **7**(1): 151-159.

Seddigh S and Darabi M (2016). Proteomics comparison of aspartic protease enzyme in insects. *Turk. J. Biol*., **40**(1): 69-83.

Shaheen S, Mustafa G and Jamil A (2017). Hyperexpression of xylanase from 2-deoxyglucose (2-DG) resistant mutant of *Chaetomium thermophilum*. *J. Chem. Soc. Pak*., **39**(4): 695-700.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol*., **28**(10): 2731-2739.

Tang B, Nirasawa S, Kitaoka M, Marie-Claire C and Hayashi K (2003). General function of N-terminal propeptide on assisting protein folding and inhibiting catalytic activity based on observations with a chimeric thermolysin-like protease. *Biochem. Biophys. Res. Commun*., **301**(4): 1093-1098.

Verma NK and Singh B (2013). Insight from the structural molecular model of cytidylate kinase from *Mycobacterium tuberculosis*. *Bioinformation*, **9**(13): 680-684.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R and Schwede T (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*., **46**(W1): W296-W303.

Wiederstein M and Sippl MJ (2007). Pro SA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*., **35**(2): W407-W410.

Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Honigschmid P, Schafferhans A, Roos M, Bernhofer M and Richter L (2014). Predict Protein an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res*., **42**(W1): W337-W343.

Zuckerkandl E and Pauling L (1965). Evolutionary divergence and convergence in proteins. In Evolving genes and proteins. Academic Press. pp.97-166.