Heliyon 10 (2024) e33895

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Computational analysis to comprehend the structure-function properties of fibrinolytic enzymes from *Bacillus* spp for their efficient integration into industrial applications

Nitisha Boro [a], Pedro Alexandrino Fernandes [b], Ashis K. Mukherjee [a,c,*]

[a] *Microbial Biotechnology and Protein Research Laboratory, Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, 784028, Assam, India*

[b] *LAQV@REQUIMTE, Departamento de Química e Bioquímica, Faculdade De Ciências, Universidade do Porto, Rua Do Campo Alegre S/N, 4169-007, Porto, Portugal*

[c] *Microbial Biotechnology and Protein Research Laboratory, Division of Life Sciences, Institute of Advanced Studies in Science and Technology, Vigyan Path, Garchuk, Paschim Boragaon, Guwahati, 781035, Assam, India*

ARTICLE INFO

Keywords:
Fibrinolytic enzyme
Bacterial enzymes
Structure-function properties of enzyme
Enzyme therapeutics

ABSTRACT

Background: The fibrinolytic enzymes from *Bacillus* sp. are proposed as therapeutics in preventing thrombosis. Computational-based analyses of these enzymes' amino acid composition, basic physiological properties, presence of functional domain and motifs, and secondary and tertiary structure analyses can lead to developing a specific enzyme with improved catalytic activity and other properties that may increase their therapeutic potential.

Methods: The nucleotide sequences of fibrinolytic enzymes produced by the genus *Bacillus* and its corresponding protein sequences were retrieved from the NCBI database and aligned using the PRALINE programme. The varied physiochemical parameters and structural and functional analysis of the enzyme sequences were carried out with the ExPASy-ProtParam tool, MEME server, SOPMA, PDBsum tool, CYS-REC tool, SWISS-MODEL, SAVES servers, TMHMM program, GlobPlot, and peptide cutter software. The assessed *in-silico* data were compared with the published experimental results for validation.

Results: The alignment of sixty fibrinolytic serine protease enzymes (molecular mass 12–86 kDa) sequences showed 49 enzymes possess a conserved domain with a catalytic triad of Asp196, His242, and Ser569. The predicted instability and aliphatic indexes were 1.94–37.77, and 68.9–93.41, respectively, indicating high thermostability. The random coil means value suggested the predominance of this secondary structure in these proteases. A set of 50 amino acid residues representing motif 3 signifies the Peptidase S8/S53 domain that was invariably observed in 56 sequences. Additionally, 28 sequences have transmembrane helices, with two having the most disordered areas, and they pose 25 enzyme cleavage sites. A comparative analysis of the experimental work with the results of *in-silico* study put forward the characteristics of the enzyme sequences JF739176.1 and MF677779.1 to be considered when creating a potential mutant enzyme as these sequences are stable at high pH with thermostability and to exhibit αβ-fibrinogenase activity in both experimental and *in-silico* studies.

* Corresponding author. Institute of Advanced Studies in Science and Technology, Vigyan Path, Garchuk, Paschim Boragaon, Guwahati, Assam, 781035, India.

E-mail addresses: akm@tezu.ernet.in, ashmukh@yahoo.co.uk (A.K. Mukherjee).

https://doi.org/10.1016/j.heliyon.2024.e33895
Received 28 June 2023; Received in revised form 28 June 2024; Accepted 28 June 2024
Available online 1 July 2024

1. Introduction

In the past 100 years, enzyme replacement therapy has become a widespread practice in medicine. Proteases with a unique property known as fibrinolytic enzymes can break down the fibrin mesh that forms blood clots. The soluble protein precursor fibrinogen is the source of fibrin, which is then formed inside blood vessels damaged or subjected to disease as tiny strands that gradually harden, effectively encasing the blood components. While endogenous fibrin clots pose a significant risk for thrombotic or cardiovascular diseases (CVDs), there is no denying that exogenous fibrin clots are essential for repairing blood vessels [1]. The leading cause of death worldwide, particularly in Western nations, is cardiovascular disease. According to the World Health Organization, cardiovascular illness causes 17.9 million deaths annually, accounting for 31 % of global mortality; this severe issue requires urgent medical intervention [2].

When the vascular system is damaged, fibrinogen, a large soluble plasma glycoprotein generated and released by the liver, is transformed into polymeric fibrin by thrombin. When platelets gather at the location of the wound, a proteolytic cascade mechanism that converts fibrinogen into fibrin is initiated to commence the blood clotting process [3,4]. Under natural physiological conditions, plasmin hydrolyses these clots; however, due to numerous pathophysiological shambles, this constant dynamic equilibrium is disturbed, leading potentially fatal condition called hyperfibrinogenemia, which then causes several thrombotic diseases, including abdominal aortic aneurysm [5,6], stroke [7,8], pulmonary embolism [9], cardiovascular disease (CVD) [10,11], and peripheral vascular disease [12].

Heparin, fondaparinux, idraparinux, warfarin, and rivaroxaban are anticoagulants used to treat thrombosis. However, these medications can have side effects, including bleeding, the main issue with all anticoagulants, hemorrhages [13], and long-term adherence. Urokinase and plasminogen activators are two common thrombolytic medications used to treat thrombosis; however, in addition to being expensive, they can potentially cause internal bleeding in the gastrointestinal tract when taken orally [1,14]. Likewise, prasugrel, aspirin, and ticagrelor, antiplatelet medications intended to prevent blood clots, can also cause gastrointestinal bleeding, skin bruising, and intracranial bleeding hemorrhage [15]. The non-specific fibrin degradation and degradation of other hemostatic proteins, ultimately leading to hemorrhage, have prompted researchers to hunt for alternative or axillary treatment using natural drugs to improve their selectivity and specificity towards fibrin-rich thrombi while minimizing the risk of bleeding complications associated with systemic fibrinolysis, offering a more promising approach to treat thrombotic disorders by addressing these limitations.

The low thermostability of several fibrinolytic enzymes available today decreases activity or denaturation at a higher temperature. This constraint makes them less suitable for industrial processes where high temperatures are necessary for effective enzyme activity [16]. Moreover, certain fibrinolytic enzymes could not be selective enough for the substrates they are meant to target, which could result in side effects or decreased effectiveness in therapeutic applications. Increasing the substrate selectivity of these enzymes is essential to improve their therapeutic potential [17]. Furthermore, alkaline stability is crucial because fibrinolytic enzymes may encounter alkaline conditions during production. Under these circumstances, enzymes with low alkaline stability may experience reduced activity or structural integrity [18]. To improve the stability, specificity, and effectiveness of fibrinolytic enzymes, it is imperative that either novel fibrinolytic enzymes be discovered or that the evolution of already existing ones is regulated.

The fibrinolytic enzymes either catalyze the process of fibrin degradation or convert inactive plasminogen into active plasmin, restoring the typical blood vascular architecture [19]. They are primarily proteases, which can be produced by all living cells, particularly bacteria, and are involved in the complete hydrolysis of proteins. Based on their site of action, proteases can be categorized into two groups: endopeptidases and exopeptidases, and they can be further classified into five types based on the different functional groups found at the active site, including serine protease, metalloproteases, aspartic proteases, cysteine proteases and threonine proteases [14]. Numerous fibrinolytic enzymes have been discovered in a variety of organisms, including microorganisms [1,20–24], mammals [25], and plants [26]. Microbial fibrinolytic enzymes are the most advantageous among all of these sources because they can be mass-produced, are inexpensive, have a wide range of biochemical properties, are simple to manipulate genetically, and have the potential to be employed as medications for treating thrombosis and other related disorders since they can dissolve the thrombus by directly destroying the fibrin at a high rate [1].

Numerous microorganisms have been found to have fibrinolytic enzymes; however, researchers worldwide are particularly interested in proteases by the genus *Bacillus* since they are typically considered harmless [27–30]. The extensive application of fibrinolytic enzymes for industrial and therapeutic has piqued people's interest in learning more about their mechanism of action and structure-function relationships [31]. High protease effectiveness is typically required by industrial applications nowadays in non-physiological conditions, including high or low pH and temperature, chelating agents, and organic solvents. Modern methods include the addition of intra-molecular disulfide bonds, the construction of metal binding sites, and the alteration of amino acid groups in the active site are exploited to stabilize the protease architecture and enhance its heat (thermal) durability. There are primarily two approaches to protein engineering used to create novel enzymes for commercial or therapeutic applications: rational redesigning and direct evolution [32,33].

To achieve the above target, a thorough understanding of fibrinolytic enzyme structure-function properties is essential for their efficient integration into large-scale industrial processes and therapeutic applications. Furthermore, a theoretical overview study will aid researchers in gaining a detailed understanding of protein assembly. It may also assist in designing enzymes with anticipated properties for use in industry or as therapeutic targets. However, only a few studies have been done to comprehend the structure-function relationship of fibrinolytic protease enzymes from *Bacillus* sp. by *in silico* analysis of their gene sequences. Research might aid in constructing an industrially important potent fibrinolytic protease enzyme.

Considering the above, therefore, the goal of this study was to use *in-silico* methods to characterize fibrinolytic enzymes from the *Bacillus* genus for comparing their physiochemical properties, primary, secondary, and tertiary structure, functional analysis, domains and motifs, and protein model analysis that will assist in the production of a potential mutant enzyme for its commercial application with appropriate properties by site-directed mutagenesis.

2. Materials and methods

2.1. Sequence retrieval and alignment

The keyword '*Bacillus* fibrinolytic enzyme' was used to search for submitted nucleotide sequences of fibrinolytic enzymes produced by the genus *Bacillus* from the NCBI database (https://www.ncbi.nlm.nih.gov/) covering the period from 2002 to May 31, 2022 (approximately 20 years). The selected fibrinolytic enzyme nucleotide sequences' corresponding protein sequences were retrieved from the NCBI database. The obtained protein sequences were aligned using multiple sequence alignments using the PRALINE programme (https://www.ibi.vu.nl/programs/pralinewww/), and the alignments were examined using the CLC sequence viewer 8.0 (http://www.clcbio.com).

2.2. Determination of physiochemical characteristics of the proteins

The ExPASy-ProtParam tool (https://web.expasy.org/protparam) was used to calculate the varied physiochemical parameters of the fibrinolytic enzymes [34]. The computed parameters in the ProtParam data include-molecular mass, isoelectric point (pI), extinction coefficient (EC-quantitative study of protein-protein and protein-ligand interactions), aliphatic index (AI-relative volume of protein occupied by aliphatic side chains), instability index (II-stability of proteins), and Grand Average of Hydropathicities (GRAVY-sum of all hydrophobicity values of all amino acids divided by the number of residues in a sequence).

2.3. Structure analyses of proteases

The ExPASy-ProtParam tool analyzed the primary structure, i.e., the number of amino acids in the polypeptide chain. The Pfam site was used for its domain search, and the MEME server was used for motif analysis [35]. Protein BLAST was used to conduct a functional biological study of MEME-deduced conserved protein motifs. The domains were predicted using the InterPro scan, which identified the best match based on the highest resemblance score.

SOPMA from the Network Protein Sequence Analysis (NPS@) server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) performed secondary structure analysis on the retrieved fibrinolytic enzymes, which comprised the number of α-helices, β-turn, extended strand, β-sheet, and random coils [34]. The PDBsum tool (https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) was used to generate the secondary motif map and topology diagram [36]. Each protein's anticipated disulfide patterns and consensus secondary structure contents were tallied. The CYS-REC tool (http://www.softberry.com/berry.phtml?topic=cys_rec) was used to determine the presence of disulfide bridges by predicting the most likely bonding patterns between accessible cysteine residues.

The SWISS-MODEL 3.1.0 (https://swissmodel.expasy.ord/) was used to create 3D models of all the obtained fibrinolytic enzymes for the tertiary structure analysis and validation [37]. The most crucial aspect of structure prediction was structure evaluation. The QMEAN (https://swissmodel.expasy.org/qmean/) and SAVES servers (https://saves.mbi.ucla.edu/) were used to test and verify the predicted protein models for all of the bacterial fibrinolytic enzymes. SAVES server was used to analyze Verify3D [38] and ERRAT [39]. Verify3D software testifies whether an atomic model (3D) was well-matched with its amino acid sequence (1D) [40], whereas ERRAT verifies the crystallographic structure of proteins.

2.4. Functional analysis

Membrane protein topology and proteolytic cleavage sites were used to predict function. The TMHMM 2.0 program (www.cbs.dtu.dk/services/TMHMM) was used to determine the topology of membrane proteins, precisely whether they were membrane-spanning or extracellular [41]. The protein (protease) sequences were examined for any areas of globularity and disorder using the GlobPlot 2.3 (http://globplot.embl.de/). This online bioinformatics tool searches the domain databases and sets of disordered proteins for order/globularity or disorder tendency in the query protein based on a running total of the propensity for an amino acid [42]. Proteolytic cleavage sites in the fibrinolytic enzymes were predicted using peptide cutter software (http://web.expasy.org/peptide_cutter) [34], a web-based application that indicates proteolytic cleavage sites and chemically cleaved sites in a given protein sequence.

2.5. Validation of the in-silico data analysis

To validate the *in-silico work,* published research articles on the selected fibrinolytic enzyme sequences were downloaded, and the experimental results were compared with the assessed *in-silico* analysis data.

Table 1
List of fibrinolytic enzyme sequences retrieved from NCBI databases.

Sl. No.	*Bacillus* Species	Sequence Description	Nucleotide GenBank Accession No.	UniProtKB	Number of Amino Acids
1.	*Bacillus amyloliquefaciens*	Strain An6 fibrinolytic enzyme F1 gene	FJ517583.1	ACL37471.1	382
2.	*Bacillus amyloliquefaciens*	Strain Ba–32732C subtilisin gene	GQ240809.1	ACT33949.1	382
3.	*Bacillus amyloliquefaciens*	Strain CB1 protease gene	KM086575.1	AIR72259.1	382
4.	*Bacillus amyloliquefaciens*	Strain CFR15 fibrinolytic enzyme CFR15.1 gene	KX550274.1	ANW10833.1	177
5.	*Bacillus amyloliquefaciens*	Strain CFR15 fibrinolytic enzyme gene	KX527848.1	ANW10832.1	178
6.	*Bacillus amyloliquefaciens*	Strain CH51 AprE51 precursor gene	EU414203.1	ACA34903.1	382
7.	*Bacillus amyloliquefaciens*	Strain CH86-1 major fibrinolytic enzyme gene	FJ882063.1	ACS45325.1	382
8.	*Bacillus amyloliquefaciens*	Strain LSSE-62 fibrinolytic enzyme gene	HQ419279.1	ADR32141.1	382
9.	*Bacillus amyloliquefaciens*	Strain MH18B1 fibrinolytic enzyme F53 gene	KJ470769.1	AHZ65109.1	301
10.	*Bacillus amyloliquefaciens*	Strain MJ5-41 fibrinolytic enzyme gene	JF739176.1	AEE81297.1	382
11.	*Bacillus amyloliquefaciens*	Subtilisin DFE precursor gene	DQ132806.1	AAZ66858.1	382
12.	*Bacillus licheniformis*	Strain CH3-17 AprE3-17 gene	GQ337861.1	ACU32756.1	382
13.	*Bacillus licheniformis*	Strain DCHI27 fibrinolytic enzyme BL27 gene	KX668266.1	ARX70409.1	167
14.	*Bacillus licheniformis*	Strain NM41 fibrinolytic enzyme (NFE-1) gene	MK695180.1	QCW95718.1	353
15.	*Bacillus licheniformis*	Strain NM74 fibrinolytic enzyme (NFE-1) gene	MK695181.1	QCW95719.1	176
16.	*Bacillus pumilus*	Strain ASM5 CF3 gene	MT743006.1	QNJ60183.1	274
17.	*Bacillus pumilus*	Strain BS15 fibrinolytic enzyme AprE (aprE) gene	MF943247.1	ATD12229.1	382
18.	*Bacillus* sp.	Strain AMS7 CF1 gene	MT743004.1	QNJ60181.1	400
19.	*Bacillus* sp.	Strain CRK CF2 gene	MT743005.1	QNJ60182.1	410
20.	*Bacillus* sp.	Ace02 fibrinolytic enzyme gene	DQ885476.1	ABI35684.1	275
21.	*Bacillus* sp.	CN subtilisin gene	EF674549.1	ABU93240.1	352
22.	*Bacillus* sp.	DJ-4 pro-subtilisin DJ-4 gene	AY627764.1	AAT45900.1	382
23.	*Bacillus* sp.	Novel Bacillus sp. isolated from Korean Traditional Soy bean paste with high fibrinolytic enzyme productivity	OF014327.1	NA	362
24.	*Bacillus* sp.	SJ subtilisin gene	EF674550.1	ABU93241.1	352
25.	*Bacillus* sp.	ZLW-2 fibrinolytic enzyme precursor	EU734749.1	ACE63521.1	381
26.	*Bacillus subtilis*	aprN gene for thermostable fibrinolytic enzyme, isolate BY25	HE717022.1	CCG39575.1	381
27.	*Bacillus subtilis*	Clone YF038 nattokinase precursor gene	AY219901.1	AAO65246.1	381
28.	*Bacillus subtilis*	DC-2 nattokinase precursor gene	DQ178658.1	ABA29609.1	275
29.	*Bacillus subtilis*	Fibrinolytic enzyme AprE2 gene	DQ997812.1	ABJ98765.1	381
30.	*Bacillus subtilis*	Fibrinolytic enzyme AprE8 gene	DQ997813.1	ABJ98766.1	374
31.	*Bacillus subtilis*	Bacillus subtilis IDCC 9204, which has a high production of proteolytic enzyme having fibrinolytic activity	DI182725.1	NA	362
32.	*Bacillus subtilis*	Strain A26 fibrinolytic enzyme F1 gene	FJ517584.1	ACL37472.1	381
33.	*Bacillus subtilis*	Strain ASM1 CF4 gene	MT743007.1	QNJ60184.1	416
34.	*Bacillus subtilis*	Strain AZ01 subtilisin (qk) gene	JQ927217.1	AFN10777.1	363
35.	*Bacillus subtilis*	Strain booming nattokinase gene	FJ407060.1	ACJ48969.1	362
36.	*Bacillus subtilis*	Strain BR21 fibrinolytic enzyme gene	KX527847.1	ANW10831.1	150
37.	*Bacillus subtilis*	Strain DC10 fibrinolytic enzyme gene	EF474344.1	ABO77900.1	362
38.	*Bacillus subtilis*	Strain HK176 subtilisin-like protease (aprE) gene	KJ572414.1	AHN52401.1	382
39.	*Bacillus subtilis*	Strain HN03 serine protease (DC1) gene	FJ211593.1	ACI32816.1	274
40.	*Bacillus subtilis*	Strain JS2 fibrinolytic enzyme AprE (aprE) gene	MF677779.1	ATA67131.1	382
41.	*Bacillus subtilis*	Strain JS2 Vpr (vpr) gene	MN055600.1	QEG98964.1	810
42.	*Bacillus subtilis*	Strain K3 alkaline serine protease (aprEK3) gene	MT093822.1	QIN90910.1	382
43.	*Bacillus subtilis*	Strain KCTC 3014 subtilisin precursor gene	DQ007296.1	AAY23643.1	275
44.	*Bacillus subtilis*	Strain LSSE-22 nattokinase (aprN) gene	JN392072.1	AEV91244.1	381
45.	*Bacillus subtilis*	Strain lwo fibrinolytic enzyme gene	KU640166.1	ANY30161.1	381
46.	*Bacillus subtilis*	Strain MH10B5 fibrinolytic enzyme F315 gene	KJ470771.1	AHZ65111.1	225
47.	*Bacillus subtilis*	Strain MH12B1 fibrinolytic enzyme C1-BS-F24 gene	KJ470773.1	AHZ65113.1	126
48.	*Bacillus subtilis*	Strain MH12B1 fibrinolytic enzyme F24 gene	KJ470768.1	AHZ65108.1	297
49.	*Bacillus subtilis*	Strain MH12B1 fibrinolytic enzyme H1-BS-F24 gene	KJ470772.1	AHZ65112.1	196
50.	*Bacillus subtilis*	Strain MH12B3 fibrinolytic enzyme F209 gene	KJ470770.1	AHZ65110.1	230
51.	*Bacillus subtilis*	Strain N2-10 subtilisin (qk) gene	KX943525.1	APB87957.1	230
52.	*Bacillus subtilis*	Strain natto-89 subtilisin BRC gene	MH259324.1	QBA82440.1	352
53.	*Bacillus subtilis*	Strain PTCC 1023 subtilisin gene	HQ699519.1	ADY38664.1	381
54.	*Bacillus subtilis*	Strain SJ4 fibrinolytic enzyme AprE (aprE) gene	MK796246.1	QEG98963.1	381

(*continued on next page*)

Table 1 (*continued*)

Sl. No.	*Bacillus* Species	Sequence Description	Nucleotide GenBank Accession No.	UniProtKB	Number of Amino Acids
55.	*Bacillus subtilis*	Strain SJ4 Vpr (vpr) gene	MN055601.1	QEG98965.1	810
56.	*Bacillus subtilis* subsp.	Natto strain jap10 nattokinase gene	KF734090.1	AHD46009.1	362
57.	*Bacillus subtilis*	Natto strain Tangshan nattokinase gene	FJ376817.1	ACJ06132.1	362
58.	*Bacillus subtilis*	Thermostable fibrinolytic enzyme Nk1 gene	AY940162.1	AAX35771.1	381
59.	*Bacillus subtilis*	Thermostable fibrinolytic enzyme Nk2 gene	AY940167.1	AAX35772.1	381
60.	*Bacillus velezensis*	Strain K208 alkaline serine protease (aprEK208) gene	MT093821.1	QIN90909.1	382

3. Results

3.1. Sequence retrieval and alignment

Two hundred seventy-eight nucleotide sequences of the fibrinolytic enzyme gene from *the Bacillus* genus were submitted to the NCBI database over 20 years. Of these, 60 sequences with complete and partial coding sequences of the fibrinolytic enzyme gene were selected. The FASTA format for its corresponding protein sequences of 58 selected sequences was retrieved from NCBI databases. Two sequences, i.e., OF014327.1 and DI182725.1, were translated using the ExPASy translate tool (https://web.expasy.org/translate), which showed that their primary structure contains 126–810 amino acids (Table 1). The multiple sequence alignment revealed that 49 fibrinolytic proteases show a conserved catalytic domain consisting of amino acids Asp196, His242, and Ser 569 (Fig. 1). However, due to partial protein sequence deposition, we did not find the catalytic triad in the remaining 11 protein sequences, despite being categorized as fibrinolytic enzymes (Table 1).

3.2. In silico prediction of physicochemical properties of the fibrinolytic enzymes

Characterization of the biochemical properties of the enzymes is the first step in validating the uniqueness of any protein or enzyme molecule [34]. The computed physicochemical characteristics of fibrinolytic enzyme sequences obtained from the ExPASy ProtParam tool are listed in Supplementary Table S1. The physicochemical characterization has shown that MN055600.1 (86069.66 Da, 810 residues) has the highest molecular weight, and KJ470773.1 has the lowest mass (12297.63 Da, 126 residues). Low molecular mass enzymes may penetrate the tissues more rapidly, giving low molecular weight enzymes easier access to fibrin clots and possibly increasing thrombolytic efficiency, which can be vital in treating conditions like acute ischemic stroke. Furthermore, lower molecular weight enzymes frequently have lower immunogenicity, which minimizes the possibility of immunological reactions and enhances patient safety.

Moreover, expressing smaller enzymes in simpler expression systems like bacteria or yeast may be more manageable. Because of their reduced size, fewer complex folding and post-translational modification procedures might be necessary [43]. The pH level at which a protein's surface is eventually coated with charge, yet its net charge is zero, is known as the isoelectric point (pI). Higher pH ranges can be tolerated by proteases from *Bacillus* sp. with extremely high isoelectric points [44]. The isoelectric points of the enzymes KX550274.4, KJ470769.1, KX668266.1, MK695181.1, MT743006.1, MT743004.1, MT743005.1, DQ885476.1, DQ178658.1, MT743007.1, FJ211593.1, MN055600.1, DQ007296.1, KJ470768.1, KJ470772.1, KX943525.1, MN055601.1 were predicted at 5.19, 6.65, 5.58, 6.51, 5.44, 5.12, 5.12, 6.65, 6.65, 4.98, 6.64, 5.55, 6.65, 6.72, 6.31, 6.05, and 6.13, respectively. The isoelectric point of the remaining protein sequences was higher than 7, indicating that the proteins are stable at high pH. Different fibrinolytic enzymes used as therapeutic drugs have different optimal pH ranges depending on the specific enzyme. Nonetheless, these enzymes usually have a

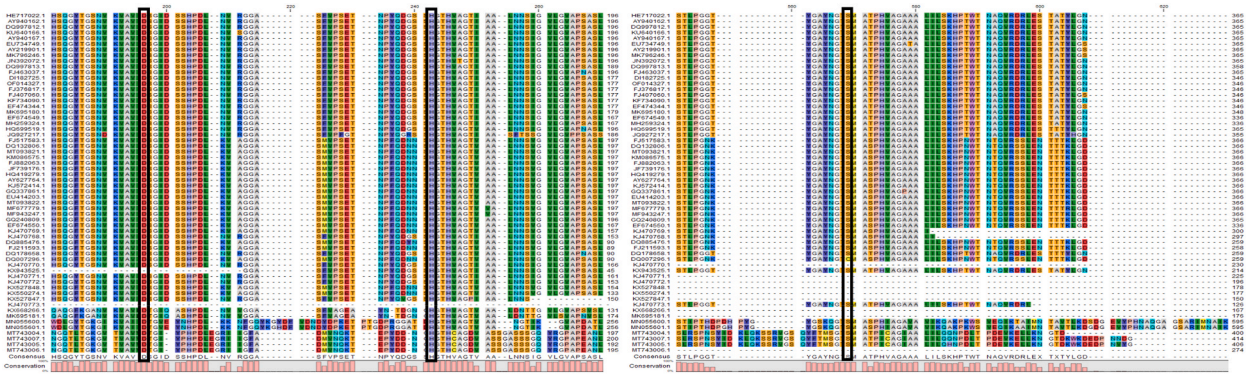

Fig. 1. The catalytic triad (Asp, His, and Ser) is shown in the black box in multiple sequence alignments of fibrinolytic enzyme amino acid sequences.

pH range of 7.4–8.0, corresponding to human blood's physiological pH. One well-known fibrinolytic enzyme employed in thrombolytic therapy for acute ischemia stroke is tissue plasminogen activator (tPA), which has an ideal pH range of 7.4. These enzymes may break down fibrin clots effectively within this pH range without becoming denatured or losing their activity [45].

Determining a purified protein's concentration or identifying protein-containing fractions in numerous applications involving peptides or proteins is essential. Tryptophan, phenylalanine, and tyrosine are amino acids with aromatic side chains that strongly absorb UV radiation. As a result, proteins and peptides absorb UV radiation in a manner determined by the concentration and composition of aromatic amino acids. Protein concentration in solution can be computed from absorbance once an absorptivity coefficient for a given protein (with its fixed amino acid content) has been determined [46]. The extinction coefficient (EC) shows how much light at a specific wavelength is absorbed by a protein. The value of EC was found to be minimum for MK695181.1 (7450) and maximum for MN055601.1 (84,690), which is beneficial in estimating the protein concentration in solution. The low or high extinction coefficient indicates the presence of low and high concentrations of aromatic amino acids in the query protein [47]. A study conducted by Kapoor et al. [48], explored the Nattokinase (NK) producing capacity of five *Bacillus subtilis* strains (R1-R5), where they calculated the extinction coefficient of the five variant by computational analysis and out of the five variant, R2, R3 and R5 exhibited the maximum value of 34,380 respectively. Moreover, the NK R3 variant had the lowest significant Km value, suggesting that it has the highest specificity among the five strains against fibrin as a substrate and could serve as a viable candidate for the treatment of cardiovascular disease. Protease concentrations and anticipated extinction coefficients may be helpful in quantitative investigations of protein-ligand and protein-protein interactions in various solutions.

According to secondary structure analysis, the instability index determines a protein's in vivo half-life [49]. Proteins with in vivo half-lives of less than 5 h showed an instability score more significant than 40, whereas those with in vivo half-lives of more than 16 h [50] had an instability value of less than 40. All protein sequences, except DQ997813.1, had an instability index of less than 40, meaning they are pretty stable in vivo. Based on the content of a protein's amino acids, the instability index is a metric used to predict the stability of that protein. It determines a protein's susceptibility to experience spontaneous protein breakdown, denaturation, or aggregation. Several areas of protein study, including protein engineering, protein expression, and protein folding, have used the instability index. The instability index has been used to predict the stability of pharmacological targets like enzymes or receptors. For instance, the stability of a group of G protein-coupled receptors (GPCRs) was predicted using the instability index in a study published in the journal Amino Acids. The authors found that GPCRs with higher instability indexes were less stable and more prone to misfolding and aggregation. Drugs that target unstable GPCRs may be less effective or have more side effects. Therefore, this information can be helpful in the development of new medications [51].

Thermostability is the capacity of a material to resist permanent changes in chemical or physical properties brought on by a rise in temperature. Thus, protein thermostability refers to the ability of polypeptide chains to maintain their unique structure and chemical composition at high temperatures [52]. The relative volume occupied by the amino acids' aliphatic side chains (Alanine, Valine, Leucine, and Isoleucine) is measured by the protein's aliphatic index. Globular proteins exhibiting a high aliphatic index have high thermostability, and protein thermostability rises as the aliphatic index increases [53,54]. All of the sequences have a high aliphatic index, ranging from 68.9 for sequence JQ927217.1 to 93.41 for sequence KX668266.1, which suggests that all of the studied enzymes are presumably thermostable. Thermostable enzymes have been shown to be more effective and practical foe a wider range of applications [55]. A vital feature that increases industrial productivity is thermostability, or a protein's ability to maintain its structural integrity and function at high temperatures, where more reagents and chemicals are available [56,57]. The ability of proteins or enzymes to withstand high temperatures suggests an evolved metabolic network system and the use of thermostable proteins and enzymes, which accelerate vital cellular biochemical processes for survival. Because these enzymes are thermostable and resistant to chemical reagents, salt concentrations, high pressure, and acidic and alkaline conditions, they are even more perfect for biotechnological applications [58]. In a study, the fibrinolytic enzyme produced by *Bacillus cereus* RSA1 exhibited thermostability over a broad temperature range of 20–80 °C. After 2 h of incubation, the enzyme maintained 76.59 % of its activity up to 80 °C and reached its maximum stability (100 %) at 20 °C [59]. In another study, subtilisin BSF1, a novel fibrinolytic enzyme, isolated from *B. subtilis* A26 exhibited maximum activity at 60 °C [60]. Additionally, quick mixing, improved substrate solubility, high mass transfer rate, and a decreased danger of contamination are some benefits of thermostable enzymes [61].

Additionally, applying thermostability can improve the stability of biological drugs like monoclonal antibodies. In another study, researchers constructed a monoclonal antibody with enhanced thermostability. The modified antibody exhibited increased stability and could continue functioning even after exposure to high temperatures, extending its shelf life [62]. The same may be true for the fibrinolytic enzymes; however, in-depth studies are warranted.

Grand Average Hydropathy (GRAVY) indices for all the sequences except KJ470769.1, KX668266.1, DQ885476.1, DQ178658.1, FJ211593.1, DQ007296.1, KJ470773.1, and KX943525.1 ranged from −0.442 to −0.014. This low range of values indicates the likelihood of a better interaction with water, thus indicating its good water solubility [63]. Therefore, extracting these proteases in the industrial sector is simple because they do not bind to hydrophobic membranes. Since water solubility is a critical factor for determining the bioavailability of the drug, therapeutic molecules should have good water solubility. The ability of a drug molecule to dissolve in water affects its absorption, distribution, metabolism, and excretion (ADME) profile. Highly water-soluble drug molecules can quickly be absorbed in the gastrointestinal tract and rapidly reach the bloodstream, increasing bioavailability. Numerous studies have demonstrated a beneficial correlation between high water solubility with enhanced pharmacokinetics and bioavailability of drugs. For instance, Amidon et al.'s study revealed that drugs with high solubility and high permeability have the highest probability of being absorbed orally. Another survey by Lipinski et al. reported that drugs with good water solubility have a higher likelihood of oral bioavailability [64,65].

Additionally, according to the U.S. Food and Drug Administration (FDA) standards, novel drug candidates should have acceptable

Table 2
Biochemical characteristic summary of the retrieved fibrinolytic enzymes determined by computational analysis.

Sl. No.	*Bacillus* Species	Nucleotide GenBank Accession No.	UniProtKB	pI (Stable at high pH)	II (Stable in vivo)	AI (Thermostability)	GRAVY (Good water solubility)
1.	*Bacillus amyloliquefaciens*	FJ517583.1	ACL37471.1	✓	✓	✓	✓
2.	*Bacillus amyloliquefaciens*	GQ240809.1	ACT33949.1	✓	✓	✓	✓
3.	*Bacillus amyloliquefaciens*	KM086575.1	AIR72259.1	✓	✓	✓	✓
4.	*Bacillus amyloliquefaciens*	KX550274.1	ANW10833.1	X	✓	✓	✓
5.	*Bacillus amyloliquefaciens*	KX527848.1	ANW10832.1	✓	✓	✓	✓
6.	*Bacillus amyloliquefaciens*	EU414203.1	ACA34903.1	✓	✓	✓	✓
7.	*Bacillus amyloliquefaciens*	FJ882063.1	ACS45325.1	✓	✓	✓	✓
8.	*Bacillus amyloliquefaciens*	HQ419279.1	ADR32141.1	✓	✓	✓	✓
9.	*Bacillus amyloliquefaciens*	KJ470769.1	AHZ65109.1	X	✓	✓	X
10.	*Bacillus amyloliquefaciens*	JF739176.1	AEE81297.1	✓	✓	✓	✓
11.	*Bacillus amyloliquefaciens*	DQ132806.1	AAZ66858.1	✓	✓	✓	✓
12.	*Bacillus licheniformis*	GQ337861.1	ACU32756.1	✓	✓	✓	✓
13.	*Bacillus licheniformis*	KX668266.1	ARX70409.1	X	✓	✓	X
14.	*Bacillus licheniformis*	MK695180.1	QCW95718.1	✓	✓	✓	✓
15.	*Bacillus licheniformis*	MK695181.1	QCW95719.1	X	✓	✓	✓
16.	*Bacillus pumilus*	MT743006.1	QNJ60183.1	X	✓	✓	✓
17.	*Bacillus pumilus*	MF943247.1	ATD12229.1	✓	✓	✓	✓
18.	*Bacillus sp.*	MT743004.1	QNJ60181.1	X	✓	✓	✓
19.	*Bacillus sp.*	MT743005.1	QNJ60182.1	X	✓	✓	✓
20.	*Bacillus sp.*	DQ885476.1	ABI35684.1	X	✓	✓	X
21.	*Bacillus sp.*	EF674549.1	ABU93240.1	✓	✓	✓	✓
22.	*Bacillus sp.*	AY627764.1	AAT45900.1	✓	✓	✓	✓
23.	*Bacillus sp.*	OF014327.1	NA	✓	✓	✓	✓
24.	*Bacillus sp.*	EF674550.1	ABU93241.1	✓	✓	✓	✓
25.	*Bacillus sp.*	EU734749.1	ACE63521.1	✓	✓	✓	✓
26.	*Bacillus subtilis*	HE717022.1	CCG39575.1	✓	✓	✓	✓
27.	*Bacillus subtilis*	AY219901.1	AAO65246.1	✓	✓	✓	✓
28.	*Bacillus subtilis*	DQ178658.1	ABA29609.1	X	✓	✓	X
29.	*Bacillus subtilis*	DQ997812.1	ABJ98765.1	✓	✓	✓	✓

water solubility and permeability to be considered for development. The FDA has established a minimum solubility requirement of 0.1 mg/mL for a drug to be deemed highly soluble [66]. The different biochemical characteristics of the enzymes are summarized in Table 2.

3.2.1. Primary sequence analysis

The properties of the retrieved sequences' were also revealed through primary structure analysis. A heat map analysis of amino acid distribution among the retrieved sequences shows that Ala is the most abundant amino acid, accounting for 12.53 % of the enzyme's primary structure; in contrast, Cys is the least abundant amino acid (Fig. 2). Further, the five predominant amino acids are found in the following order; Ser (12.06 %) > Gly (10.25 %) > Val (9.44 %) > Lys (6.82 %) > Thr (5.6 %) [67]. In comparison with the general trend of average amino acid compositions that have been calculated for a large number of proteins from diverse taxa, it was observed that the estimated average percentage of amino acids of the retrieved sequences was constant as in the general trend, the average percentage of Ala, Ser, Gly, Val, Lys, Thr, and Cys accounts for 8.7 %, 7.0 %, 6.8 %, 6.5 %, 5.3 %, 5.3 %, and 1.5 %, respectively. Hydrophilic amino acids are those that seek the aqueous phase. The presence of many serine residues, a hydrophilic amino acid, in all these sequences indicates their extracellular nature, i.e., they are secreted from the cell (extracellular enzyme). Proteins' extracellular nature has significant applicability in the pharmaceutical industry, particularly in developing biological products. Proteins secreted or localized to the extracellular matrix play essential roles in signalling, immune response, and tissue repair and can be targeted by therapeutic agents to modulate these processes.

A method of treatment for genetic disorders caused by the deficiency of an extracellular enzyme is known as enzyme replacement therapy (ERT). In ERT, a recombinant enzyme is administered to replace the deficient enzyme and restore normal function. For instance, Pompe disease, a condition brought on by a lack of the extracellular enzyme acid alpha-glucosidase, is treated with the ERT

30.	*Bacillus subtilis*	DQ997813.1	ABJ98766.1	✓	X	✓	✓
31.	*Bacillus subtilis*	DI182725.1	NA	✓	✓	✓	✓
32.	*Bacillus subtilis*	FJ517584.1	ACL37472.1	✓	✓	✓	✓
33.	*Bacillus subtilis*	MT743007.1	QNJ60184.1	X	✓	✓	✓
34.	*Bacillus subtilis*	JQ927217.1	AFN10777.1	✓	✓	✓	✓
35.	*Bacillus subtilis*	FJ407060.1	ACJ48969.1	✓	✓	✓	✓
36.	*Bacillus subtilis*	KX527847.1	ANW10831.1	✓	✓	✓	✓
37.	*Bacillus subtilis*	EF474344.1	ABO77900.1	✓	✓	✓	✓
38.	*Bacillus subtilis*	KJ572414.1	AHN52401.1	✓	✓	✓	✓
39.	*Bacillus subtilis*	FJ211593.1	ACI32816.1	X	✓	✓	X
40.	*Bacillus subtilis*	MF677779.1	ATA67131.1	✓	✓	✓	✓
41.	*Bacillus subtilis*	MN055600.1	QEG98964.1	X	✓	✓	✓
42.	*Bacillus subtilis*	MT093822.1	QIN90910.1	✓	✓	✓	✓
43.	*Bacillus subtilis*	DQ007296.1	AAY23643.1	X	✓	✓	X
44.	*Bacillus subtilis*	JN392072.1	AEV91244.1	✓	✓	✓	✓
45.	*Bacillus subtilis*	KU640166.1	ANY30161.1	✓	✓	✓	✓
46.	*Bacillus subtilis*	KJ470771.1	AHZ65111.1	✓	✓	✓	✓
47.	*Bacillus subtilis*	KJ470773.1	AHZ65113.1	✓	✓	✓	X
48.	*Bacillus subtilis*	KJ470768.1	AHZ65108.1	X	✓	✓	✓
49.	*Bacillus subtilis*	KJ470772.1	AHZ65112.1	X	✓	✓	✓
50.	*Bacillus subtilis*	KJ470770.1	AHZ65110.1	✓	✓	✓	✓
51.	*Bacillus subtilis*	KX943525.1	APB87957.1	X	✓	✓	X
52.	*Bacillus subtilis*	MH259324.1	QBA82440.1	✓	✓	✓	✓
53.	*Bacillus subtilis*	HQ699519.1	ADY38664.1	✓	✓	✓	✓
54.	*Bacillus subtilis*	MK796246.1	QEG98963.1	✓	✓	✓	✓
55.	*Bacillus subtilis*	MN055601.1	QEG98965.1	X	✓	✓	✓
56.	*Bacillus subtilis subsp.*	KF734090.1	AHD46009.1	✓	✓	✓	✓
57.	*Bacillus subtilis*	FJ376817.1	ACJ06132.1	✓	✓	✓	✓
58.	*Bacillus subtilis*	AY940162.1	AAX35771.1	✓	✓	✓	✓
59.	*Bacillus subtilis*	AY940167.1	AAX35772.1	✓	✓	✓	✓
60.	*Bacillus velezensis*	MT093821.1	QIN90909.1	✓	✓	✓	✓

drug alglucosidase alfa. Alglucosidase alfa improved survival and motor function in patients with infantile-onset Pompe illness, according to research by Kishnani et al. [68]. Extracellular signalling proteins called cytokines are essential for immune system responses and inflammation. Inflammatory diseases like rheumatoid arthritis can be treated with cytokine inhibitors like tumor necrosis factor-alpha (TNF-α). Adalimumab, a TNF inhibitor, has been demonstrated to improve symptoms and reduce disease activity in patients with rheumatoid arthritis. Adalimumab was more efficient than placebo at reducing the signs and symptoms of rheumatoid arthritis, according to a study by Keystone et al. [69]. Hydrophobic interactions inside the protein core can be prompted by many hydrophobic amino acids, leading to more excellent protein stability. This property may strengthen the protein's ability to withstand denaturation and maintain its structural integrity [70]. A low proportion of charged amino acids may impact the protein's net charge, influencing its stability, solubility, and interactions with other molecules. For instance, because of electrostatic repulsion, a high concentration of positively charged amino acids might increase the solubility of proteins [71]. For protein engineering and design, it is crucial to comprehend how amino acid arrangement affects proteins' properties. This fact is because amino acid composition can direct the alteration of proteins to produce desired structural and functional qualities.

The MEME suite is a tool for eliciting motifs in a specific group of proteins. Any motif present for a typical protein family may be used as a signature sequence to aid in the initial *in silico* identification of any protein. Six motifs were discovered when the sixty retrieved sequences were processed to MEME (Supplementary Fig. S1). Analysis shows the motifs with the widest width and the best possible amino acid sequence matches (Table 3). A set of 50 amino acid residues, i.e., LNNSIGVLGVAPSASLYAVKVLDSTGSGQYSWIINGIEWAISNNMDVINM, representative of motif 3, consistently occurred in 56 protease sequences, demonstrating its identity with the Peptidase S8/S53 domain (Supplementary Fig. S1). Motifs 4 and 6 with 41 and 50 amino acid residues belonging to the peptidase S8/S53 domain were observed for sequences KX527848.1, KX527847.1, and KJ470773.1. The peptidase S8/S53 domain function is reported to be associated with the catalysis of peptide bonds by a catalytic triad mechanism [72]. Motif 1 with 50 amino acid residues, respectively belonging to the peptidase S8 propeptide/proteinase inhibitor I9 superfamily, responsible for the modulation of folding and activity of the pro-enzyme [73,74], was observed in MK695181.1.

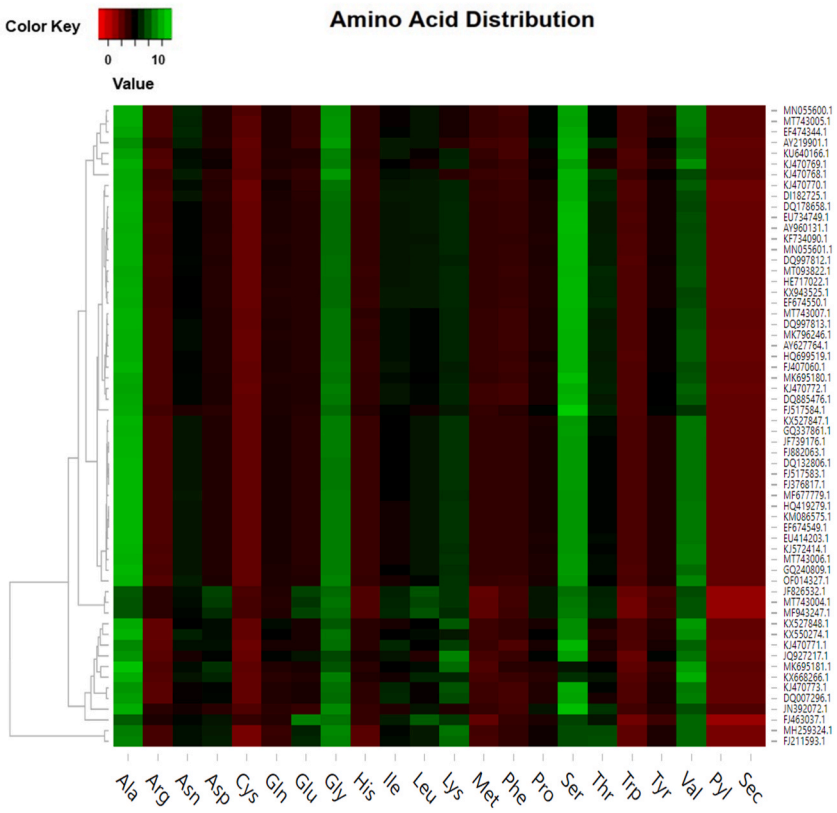

Fig. 2. Amino acid distribution in sixty fibrinolytic enzymes.

3.2.2. Secondary structure analysis

Depending on the secondary structure, an amino acid may be found in a coil, strand, or helix [75,76]. The random coil was predicted to dominate all other protease sequences except KX668266.1, MK695181.1, MT743006.1, and KJ470771.1, in which α-helix

Table 3
Prediction of various motifs showing the best potential amino acid sequence matches, domain, and molecular activity.

Motif No.	Width	Best Matching Amino Acid Sequence	Predicted Domain	Molecular Function	Reference
1.	50	TLBEKAVKELKKDPSVAYVEEDHIAHEYAQSVPYGISQIKAPALHSQGYT	Peptidase S8 propeptide/ proteinase inhibitor I9	Accountable for the activity and modulation of folding of the pro-enzyme	Tangrea et al., 2002, Jain et al., 1998.
2.	50	PGVSIQSTLPGGTYGAYNGTSMATPHVAGAAALILSKHPTWTNAQVRDRL	Peptidase S8/ S53 domain	Catalytic mechanism involving a catalytic trio catalyses peptide bonds in a polypeptide chain	Siezen et al., 1997
3.	50	LNNSIGVLGVAPSASLYAVKVLDSTGSGQYSWIINGIEWAISNNMDVINM	Peptidase S8/ S53 domain	Catalytic mechanism involving a catalytic trio catalyses peptide bonds in a polypeptide chain	Siezen et al., 1997
4.	41	IDSGIDSSHPDLNVRGGASFVPSETNPYQDGSSHGTHVAGT	Peptidase S8/ S53 domain	Catalytic mechanism involving a catalytic trio catalyses peptide bonds in a polypeptide chain	Siezen et al., 1997
5.	41	EKKYIVGFKQTMSAMSSAKKKDVISEKGGKVQKQFKYVBAA	Peptidase S8 propeptide/ proteinase inhibitor I9	Accountable for the activity and modulation of folding of the pro-enzyme	Tangrea et al., 2002, Jain et al., 1998.
6.	50	AAAGNEGTSGSSSTVGYPAKYPSTIAVGAVNSSNQRASFSSVGSELDVMA	Peptidase S8/ S53 domain	Catalytic mechanism involving a catalytic trio catalyses peptide bonds in a polypeptide chain	Siezen et al., 1997

was dominant (Supplementary Table S2). It was noted that β-turns displayed a lower percentage of confirmation (below 16 %) in the case of all protease sequences except for KX668266.1. All the protease sequences had extended strands ranging from 16.35 to 25.97 % (Supplementary Table S2).

Helix-helix interaction data is beneficial for reconstructing the structure of membrane proteins [77]. Since most membrane proteins' transmembrane (TM) regions are folded into helices, the interactions among these TM helices play a crucial role in determining the folding and stabilization of the membrane proteins [78,79]. Interactions between transmembrane α-helices, the main structural component of integral membrane proteins, have been the subject of numerous research on the stability of membrane proteins [80]. According to one current hypothesis, the stability of membrane protein oligomers is greatly influenced by helix-helix interactions within the membrane [81]. Because they stabilize many protein structures, helix-helix interactions are of interest for the scientists. Helix interactions are fundamental in membrane proteins, where transmembrane helical regions frequently control protein orientation about the lipid bilayer [82]. Many proteins must fold correctly as monomers and assemble or interact with other proteins permanently or temporarily to carry out their tasks. Helixes are often involved in both structural and functional protein-protein interactions. [83], even though the intra and inter-protein contacts that characterize these phenomena may result from β-sheet structures found in bacterial outer membranes and mitochondria [84] and/or binding to intrinsically disordered proteins [85].

Seven different motif maps and topology diagrams were obtained when a more detailed secondary structure prediction was made using the PDBsum tool. The sequences that shared a similar motif map and topology diagrams were clustered into Groups 1 to 7. Group 1 sequences involved nine helices in 10 helix-helix interactions. At the same time, three β-sheet motifs with ten strands were found in the secondary structure, depicted in the schematic diagram (Fig. 3a). The topological diagram of group 1 sequences shows the helices and β-strands with cylinders and arrows (Fig. 3b). The secondary structure schematic diagram of group 2 sequences revealed nine helices involved in 10 helix-helix interactions and two β-sheet motifs made up of 9 strands (Supplementary Fig. S2a). Its topological diagram is depicted in Supplementary Fig. S2b. Twenty-three helices in Group 3 secondary structure. Topological graphs involved 12 helix-helix interactions and twenty-one β-sheet motifs with eighty-two strands (Supplementary Figs. S3a and b).

The secondary structure and its topological diagram of Group 4 sequences revealed nine helices with 11 helix-helix contacts and two β-sheet motifs with nine β-strands (Supplementary Figs. S4a and b). The topological graph and its secondary structure of Group 5 sequences revealed nine helices involved in 11 helix-helix interactions and two β-sheet motifs with nine β-strands (Supplementary Figs. S5a and b). Group 6 sequences involved eleven helices in 14 helix-helix interactions, while four β-sheet motifs with seventeen strands were found in the secondary structure (Supplementary Figs. S6a–c). Finally, the secondary structure of Group 7 had 14 helices involved in 9 helix-helix interactions and six β-sheet motifs with 25 β-strands (Supplementary Figs. S7a–d).

Disulfide linkages lower the entropy of the unfolded form of the protein, which helps it fold and stabilize the secondary structure from misfolding. The study of disulfide patterns for each sequence revealed that disulfide bridges were absent in most sequences except for five, where the incidence of cysteine residues was extremely low (Supplementary Table S2). Disulfide bridges in proteins have significant applicability in drug development, particularly when it comes to the creation of biologics. Disulfide bridges can be explored

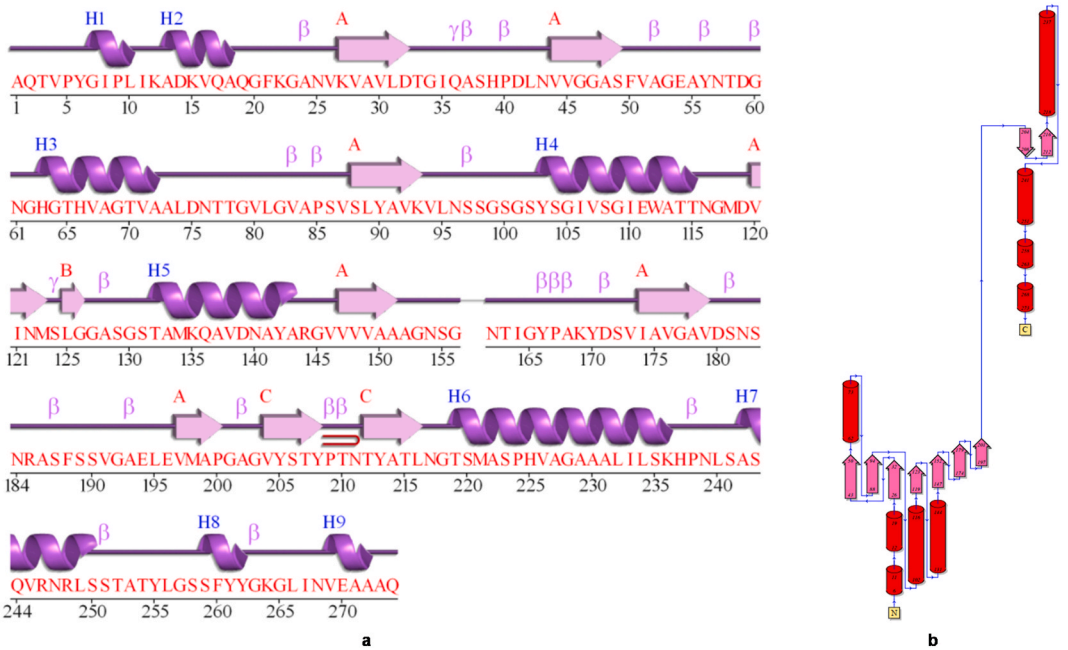

Fig. 3. The PDBsum tool generated schematic and topological diagrams depicting the secondary structural components in the group 1 protein category. **a** α-Helices labelled with the letter 'H', and β-strands lettered in uppercase. B, γ, as well as hairpin twists. **b** Helices are depicted as cylinders, while β-strands are represented with arrows.

as a target for drug designing because they are crucial in stabilizing protein structure and function. Disulfide bonds can be utilized in protein engineering to alter proteins' stability, activity, and pharmacokinetics. In protein structures, disulfide linkages are sulfur atoms from cysteine pairs that are covalently connected. Artificial disulfide links are frequently created via cysteine modification to improve protein structural stability because disulfide linkages are crucial for protein folding and structural strength. Machine learning and other approaches help predict amino acid pairs for cysteine mutations that generate manufactured disulfide bonds, which helps to facilitate the experimental design [86]. In a study conducted by Takagi et al. [87], the structure of aqualysin I of *Thermus aquaticus* YT-1 (a thermophilic subtilisin-type protease containing two disulfide links) was used to choose sites for Cys replacements to produce a disulfide bond in subtilisin E, a cysteine-free bacterial serine protease. The Cys-61/Cys-98 mutant was found to have a half-life that was 2–3 times longer than the wild-type enzyme's, and it also displayed a Tm of 63.0 °C, which was 4.5 °C higher than the enzyme's observed Tm. Another study conducted by Pantoliana et al. [88] discovered that adding a disulfide bond by site-directed mutagenesis improved subtilisin BPN' stability in a range of scenarios. Therefore, the same theory may be applied to fibrinolytic enzymes to augment their stability under physiological conditions, thus improving their biomedical application.

Through site-directed mutagenesis, a subtilisn variation with cysteine residues at positions 22 and 87 was produced, and it was demonstrated to have an activity nearly identical to the wild-type enzyme. For instance, the cytokine interferon-alpha has disulfide bonds added to promote stability and enhance pharmacokinetics. A study by Harris et al. reported that the disulfide-linked interferon-alpha was more stable and had a longer half-life in circulation than the wild-type protein [89]. Short chains of amino acids called peptides can be employed as therapeutics. Disulfide linkages can strengthen the peptide's pharmacological qualities, including potency and stability while stabilizing its structural integrity. Exenatide (Byetta), for instance, is a peptide-based drug used to treat type 2 diabetes. Exenatide has two disulfide bonds in its structure, stabilizing it and increasing its half-life in circulation. A study by Buse et al. showed that exenatide improved glycemic control and decreased body weight in type 2 diabetic individuals [90].

3.2.3. Tertiary structure analysis

The QMEAN4 score evaluated the protein 3D model of all the sequences gained from SWISS-MODEL. Fig. 4a and b depicts the QMEAN PDB 3D model of *Bacillus* sp. (UniProtKB: ABI35684.1) protein structure and graphical representation of estimation of local quality as a representative example of the protein sequences. The QMEAN4 z–score for query model FJ517583.1 (*B. amyloliquefaciens*), DQ132806.1 (*B. amyloliquefaciens*), DQ885476.1 (*Bacillus* sp.), EF674549.1 (*Bacillus* sp.), OF014327.1 (*Bacillus* sp.), EU734749.1 (*Bacillus* sp.), HE717022.1 (*B. subtilis*.), AY219901.1 (*B. subtilis*), DQ997812.1 (*B. subtilis*), DQ997813.1 (*B. subtilis*), DI182725.1 (*B. subtilis*), FJ517584.1 (*B. subtilis*), FJ407060.1 (*B. subtilis*), KX527847.1 (*B. subtilis*), JN392072.1 (*B. subtilis*), KJ470773.1 (*B. subtilis*), KX943525.1 (*B. subtilis*), MH259324.1 (*B. subtilis*), MK796246.1 (*B. subtilis*), FJ376817.1 (*B. subtilis*), AY940162.1 (*B. subtilis*), AY940167.1 (*B. subtilis*), and MT093821.1 (*B. subtilis*) was determined at −0.96, −0.96, −0.05, −0.88, −0.79, −0.84, −0.88, −0.73, −0.86, −0.88, −0.78, −0.7, −0.8, −0.9, −0.72, −0.27, −0.73, −0.79, −0.85, −0.74, −0.78, −0.8 and −0.96 respectively, which was lower than the standard deviation '1' from the mean value '0' of good models, indicating that the estimated

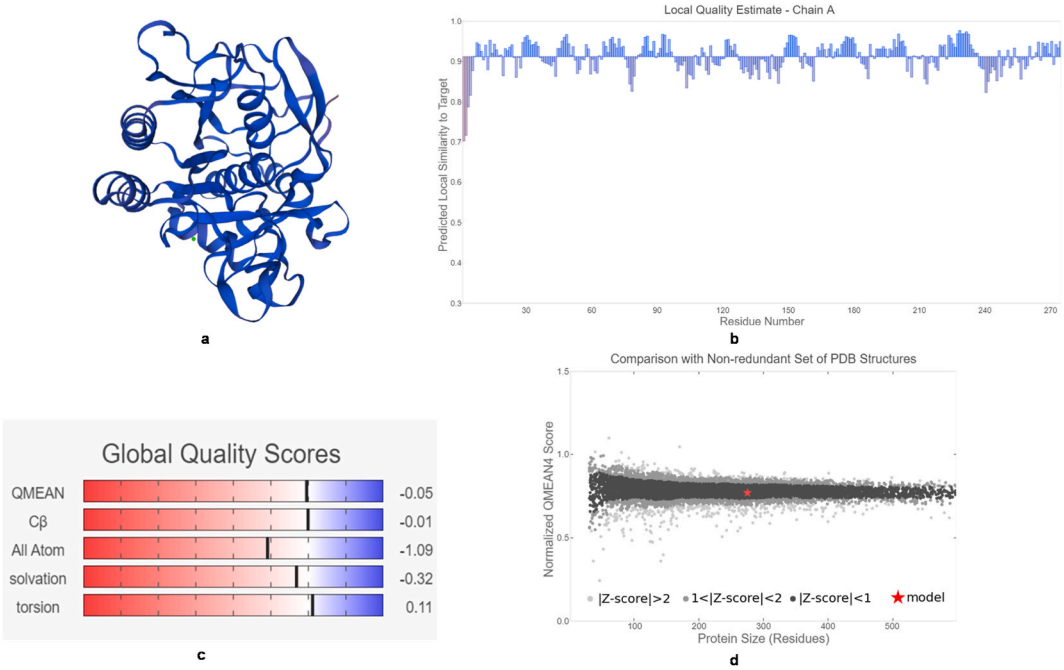

Fig. 4. Quality analysis of the predicted model of *Bacillus* sp. (UniProtKB: ABI35684.1). **a** QMEAN PDB 3D model of *Bacillus* sp. protein (PDB ID: 1sup.A) structure. **b** Estimation of local quality as a graphical representation. **c** The z-scores of the QMEAN terms of the protein model PDB ID: 1sup. A **d** Graphical representation of the model's absolute quality estimation (PDB ID: 1sup.A).

models were comparable to high-resolution structures with the same size. The five different structure descriptors of all the sequences obtained from the QMEAN4 server are summarized in Supplementary Table S3. The 3D and ERRAT evaluations from SAVES were done to further verify the structures by understanding 3D protein models and estimating their accuracy. Except for five sequences, all recovered sequences' 3D models passed Verify 3D with a maximum score of 100, indicating that 100 % of their residues had an average 3D-1D score of 0.2.

The structural verification process in ERRAT interpreted the overall quality of the models with a maximum result score of 100 %. The Verify3D and ERRAT-verified protein structure of *Bacillus* sp. (UniProtKB: ABI35684.1), an example of the protein sequences, is represented in Supplementary Figs. S8a and b. The above investigations confirmed the predicted structure of the proteins to be good, stable, dependable, and consistent. The Verify3D score data and the overall quality factor from ERRAT for all sequences are shown in Supplementary Table S4.

The architecture of proteins is thought to be organised and stabilised primarily by hydrophobic interactions [91]. The hydrophobic effect is the most widely considered important factor which causes the protein's hierarchical structure and three-dimensional stability [92]. Many aspects of protein structure-function dependencies, such as the presence of amphipathic structures induced in peptides or membrane proteins at lipid boundaries, stabilization of the folded conformity of globular proteins in solution, protein-receptor binding, protein-protein interactions related to protein subunit assembly, and other intermolecular biorecognition processes are the pieces of evidence that exhibit hydrophobic effect [93]. When proteins are separated from polar solvents, their hydrophobic core often comprises the residues of hydrophobic amino acids. van der Waals interactions, which are essential for stabilizing the structure, are made possible by the closely packed side chains in the core. Moreover, disulfide (*S*–*S*) bridges, which can connect separate subunits within a complex or different secondary structure elements, are formed by Cys residues and stabilize three-dimensional structures. Another essential function of Cys is metal binding, which can occur in metal centers that stabilize structures or enzymes' active sites [94].

3.3. Functional analysis of fibrinolytic enzymes

TMHMM server 2.0 analysis of transmembrane helix estimation revealed that 28 of the total sequences possess a transmembrane helix (Supplementary Table S5). Most of these proteins have 19–22 residues in the transmembrane helixes. Almost the majority belong to the first 60 residues, predicting the transmembrane helix in the N-terminal region could be a signal peptide. The TMHMM result of *Bacillus* sp. (UniProtKB: ABI35684.1) as an illustrative example is depicted in Supplementary Fig. S9. Membrane proteins frequently contain transmembrane helices, crucial for cell signalling, transport, and other biological functions. Targeting membrane proteins with therapies can be an effective strategy for treating various disorders because many diseases are associated with the dysfunction of these proteins. Transmembrane helices may help develop therapeutics that target membrane proteins [95]. Transmembrane helices can confer stability and specificity to a protein structure, which may be advantageous for some treatments. For instance, synthetic peptides targeting cancer cells or harmful bacteria have been made using transmembrane helices [96]. Due to their usual hydrophobicity and low solubility in aqueous solutions, transmembrane helices can be challenging to deal with in the lab and have limited therapeutic potential [95].

Additionally, transmembrane helices may be more immunogenic than other protein domains, causing the body to produce an immunological response. This effect may result in adverse effects or reduced effectiveness of therapy [97]. Finally, transmembrane helices might provide difficulties for drug delivery since they are frequently challenging to distribute across cell membranes and may be susceptible to degradation or clearance by the body's immune system. However, alternative methods for overcoming these challenges exist, such as employing delivery systems based on nanotechnology or altering the protein's structure to enhance its stability and solubility [97].

The GlobPlot tool analysis showed that sequences MN055600.1 (*B. subtilis*) and MN055601.1 (*B. subtilis*) had a maximum number of disordered regions, i.e., 11 with three ordered domains. The remaining 58 sequences had 6-1 disordered regions with a 3-0 number of ordered domains (Supplementary Table S6). The presence of disordered regions and ordered domains in a protein may affect the design and creation of protein therapies. Proteins' disordered regions can be crucial to signal transduction, enzyme catalysis, and protein-protein interactions. Disordered areas can occasionally be targeted for drug development. For instance, the p53 protein has a disordered region that interacts with other proteins, such as MDM2, a negative regulator of p53. Increased p53 activity and death in cancer cells can result from peptides that mimic the disordered area of p53 and block the interaction between p53 and MDM2 [98]. Alternatively, proteins with ordered domains may provide structural stability and particular binding sites for ligands or other proteins. Many protein therapeutics, such as monoclonal antibodies, utilize ordered domains for their specific binding properties. For instance, rheumatoid arthritis is treated with monoclonal adalimumab, which binds to tumor necrosis factor-alpha (TNF-α) [99].

In some cases, protein engineering can modify the balance of disordered regions and ordered domains in a protein for therapeutic purposes. For instance, in a current study, a protein was created to have an ordered domain and a disordered region that served as a degradation signal. The proteasome could selectively degrade the resulting protein, offering a potential therapeutic approach for diseases caused by the accumulation of misfolded proteins [100].

The peptide cutter software results show that all recovered amino acid sequences feature numerous cleavage sites for 25 diverse digestive enzymes. The total number of cleavages was found to be maximum for MN055600.1 (*B. subtilis*), i.e., 1674, and minimum for KJ470773.1 (*B. subtilis*), i.e., 193 (Supplementary Table S7). Protein cleavage sites for particular digestive enzymes can be added or removed using protein engineering techniques, which may have therapeutic implications for several diseases. A specific example is the treatment of celiac disease, an autoimmune condition brought on by gluten ingestion. A protein called gluten, found in wheat, barley, and rye, cannot be broken down by human enzymes. Gluten fragments that contain specific amino acid sequences (gluten peptides) can

cause an immunological reaction in persons with celiac disease that harms the small intestine. Researchers have modified gluten peptides via protein engineering by adding or eliminating cleavage sites for digestive enzymes like pepsin and trypsin. Therefore, individuals with celiac disease are less likely to develop an immunological reaction to these modified peptides. For instance, a recent study [101] used computational modelling and experimental validation to design a gluten peptide variant that was digested by pepsin and trypsin more effectively, resulting in a 90 % reduction in immune response compared to the original peptide in cells from celiac patients [101]. The development of protein therapies serves as another illustration. To increase their stability, bioavailability, or therapeutic activity, therapeutically given proteins like enzymes or antibodies might be altered to incorporate cleavage sites for particular proteases [102].

3.4. Validation of in-silico data

Out of the 60 fibrinolytic enzyme sequences, only 16 sequences were from published research articles. According to the published data, the molecular mass of the sequences ranged from 27 kDa to 38 kDa, which was relatively similar to the predicted in-silico data analysis. The sequence with HQ699519.1 [103] accession no. has the highest molecular weight of 38,000.00 Da size, while the sequence with KM086575.1 [104] and EU414203.1 [105] accession no. has the lowest molecular weight of 27446.57 Da. As per the physiochemical characteristic analysis, the molecular weight of HQ699519.1 was predicted to be 39554.59 Da, while for KM086575.1 and EU414203.1, it was 39095.13 Da and 39099.12 Da. The divergence in the experimentally predicted molecular weight and the *in-silico* analyzed molecular weight of KM086575.1 and EU414203.1 may have variations because of initial signal peptide sequences in the submitted database sequence.

The optimum pH at which the enzymes showed the highest fibrinolytic activity ranged from 3.0 to 9.0, with the highest value for DQ132806.1 [106] and FJ517584.1 [60] at pH 9.0 and the lowest for AY627764.1 [107] at pH 3.0. While according to the calculated isoelectric point by the ExPASy ProtParam tool, the predicted isoelectric point (pI) for DQ132806.1 and FJ517584.1 was 9.23 and 9.05 and for AY627764.1, it was 9.23. The experimental results specified that the enzymes' stability ranged from 40 to 60 °C. The enzymes' fibrinolytic activity was relatively stable at 40 °C for the sequences with accession no. KM086575.1 [104] and MF677779.1 [108], while the sequences with AY627764.1 [107] and FJ517584.1 [60] accession no., the fibrinolytic activity was stable at a higher temperature of 60 °C. However, the evaluated aliphatic index (AI) by the ExPASy ProtParam tool was found to be 82.02 for KM086575.1, 82.51 for MF677779.1, 82.02 for AY627764.1 and 81.23 for FJ517584.1. The sequences with accession numbers KM086575.1 [104], JF739176.1 [109], and DQ132806.1 [106] could efficiently degrade N-succinyl-Ala-Ala-Pro-Phe-*p*NA, which is a substrate for chymotrypsin and subtilisin, within 10 min of incubation at 37 °C. It has also been experimentally proven that the enzymes with JF739176.1 [109] and MF677779.1 [108] accession no. can degrade the α-chain and β-chain of fibrinogen within 10 min and 60 min, thus indicating them to be considered as a potent fibrinolytic enzyme. A comparative analysis of the *in-silico* data and the experimentally proven data is represented in Table 4.

4. Discussion

This study aimed to predict the structure-function and compare the properties of fibrinolytic enzymes from the *Bacillus* genus,

Table 4
Comparative analysis of the *in-silico* data and experimental data.

Sl. No.	*Bacillus* Species	Nucleotide GenBank Accession No.	UniProtKB	*In-silico* value			Experimental value		
				Molecular Mass (Da)	pI	AI	Molecular Mass (Da)	pI	AI
1.	*Bacillus amyloliquefaciens*	KM086575.1	AIR72259.1	39095.13	9.23	82.02	27446.57	6.0	40 °C
2.	*Bacillus amyloliquefaciens*	EU414203.1	ACA34903.1	39099.12	9.23	82.02	27446.57	6.0	45 °C
3.	*Bacillus amyloliquefaciens*	FJ882063.1	ACS45325.1	39125.16	9.23	82.02	27476.60	NA	NA
4.	*Bacillus amyloliquefaciens*	JF739176.1	AEE81297.1	39125.16	9.23	82.02	27476.60	7.0	45 °C
5.	*Bacillus amyloliquefaciens*	DQ132806.1	AAZ66858.1	39139.19	9.23	82.28	27700.00	9.0	48 °C
6.	*Bacillus licheniformis*	MK695180.1	QCW95718.1	36341.52	7.96	75.78	NA	NA	NA
7.	*Bacillus licheniformis*	MK695181.1	QCW95719.1	18244.58	6.51	89.26	NA	NA	NA
8.	*Bacillus* sp.	DQ885476.1	ABI35684.1	27495.64	6.65	84.8	28000.00	NA	NA
9.	*Bacillus* sp.	AY627764.1	AAT45900.1	39095.13	9.23	82.02	29000.00	3.0–4.0	60 °C
10.	*Bacillus* sp.	EU734749.1	ACE63521.1	39496.46	9.04	81.23	28000.00	NA	NA
11.	*Bacillus subtilis*	AY219901.1	AAO65246.1	39480.46	9.04	81.73	28000.00	8.0	50 °C
12.	*Bacillus subtilis*	DQ178658.1	ABA29609.1	27832.92	6.65	84.07	28000.00	8.0	42 °C
13.	*Bacillus subtilis*	FJ517584.1	ACL37472.1	39572.52	9.05	81.23	27720.00	9.0	60 °C
14.	*Bacillus subtilis*	MF677779.1	ATA67131.1	39151.2	9.25	82.51	27502.64	8.0	40 °C
15.	*Bacillus subtilis*	JN392072.1	AEV91244.1	39537.51	9.04	81.47	NA	NA	NA
16.	*Bacillus subtilis*	HQ699519.1	ADY38664.1	39554.59	9.25	82.23	38000.00	NA	NA

which can shed light on their active site, substrate selectivity, and catalytic residues, thus revealing the mechanism(s) by which these enzymes break down fibrin clots. Further, structure-function studies of these enzymes can discover new targets for drug development [110]. Various mutagenesis and engineering investigations show that understanding the relationship between the structure and function of fibrinolytic enzymes from *Bacillus* spp helps direct the logical design of enzyme variants with increased activity, stability, and specificity [111,112]. It can also provide insights into the ideal conditions of fibrinolytic enzymes, including the utilization of recombinant expression systems and bioprocessing techniques, which have been extensively explored in recent years [113,114]. Overall, structure-function studies of fibrinolytic enzymes from *Bacillus* bacteria are critical for developing new therapeutic strategies for thrombotic and cardiovascular conditions and expanding our knowledge of enzyme function in general.

One of the objectives of this study was also to determine the similarities among the fibrinolytic enzyme sequences on a global level. Analysis of sequence similarities can provide insight by comparing newly found fibrinolytic enzymes from *Bacillus* spp to known enzymes with related sequences and activities and predicting the function of these enzymes [115]. Sequence comparison can identify common domains and regions across fibrinolytic enzymes, which can be utilized to direct experimental studies of enzyme function and structure [116]. Further, gene expression and regulation studies can be facilitated using sequence similarities in designing PCR primers to amplify and detect fibrinolytic enzyme genes [117]. Sequence comparison can identify prospective therapeutic targets among fibrinolytic enzymes based on conserved domains and regions essential for enzyme function and stability [24]. Therefore, uncovering global similarities between fibrinolytic enzyme sequences from *Bacillus* spp is vital to comprehend enzyme evolution, predict function, and discover new therapeutic targets.

Conserved amino acids are essential for protein and helix coil transition confirmation [54,118]. The conserved portion of the catalytic triad, Asp, His, and Ser, was found in 49 amino acid sequences among the 60 fibrinolytic enzymes. The serine proteases use the traditional Ser/His/Asp catalytic triad to exert their catalytic activity, where serine is used as the nucleophile, histidine as the general base and acid, and aspartate to help orient the histidine residue and neutralizes the charge that forms on the histidine during the transition states.

The physicochemical features of the fibrinolytic enzyme sequences revealed that all proteins have negative GRAVY scores except eight protein sequences, which attested to their solubility in hydrophilic solvents. In contrast, the remaining eight are polar [47,119, 120]. Theoretically, a protein's pI value of greater than 7 indicates it is alkaline (positively charged overall). In contrast, less than seven indicates that the overall charge is negative (acidic protein). The pI values of all the fibrinolytic enzymes in this investigation ranged from 4.98 to 9.25, showing that fibrinolytic enzymes produced by *Bacillus* spp possess many biochemical properties that make them functional under different environmental conditions. Amino acids that are positively charged, like lysine and arginine, can interact or form ion pairs with negatively charged residues. They can maintain the overall protein shape and reduce repulsive forces, stabilizing the enzyme structure in alkaline environments.

Additionally, alkaline stability can result in longevity and enhanced enzyme activity at higher pH levels. This property is crucial for applications requiring alkaline conditions for enzymes to perform at their best. With an instability index of less than 40, all selected bacterial fibrinolytic enzymes are predicted to be stable, except for DQ997813.1. *Bacillus* bacteria can adapt to changing environmental conditions with fibrinolytic (protease) enzymes by allowing them to survive and thrive in various ecological niches, establish infections, and evade host immune responses. The ability of *Bacillus* spp to endure harsh environmental conditions, such as high temperatures, pH variations, and high salt concentrations, is well documented. Therefore, they are perfect for creating enzymes that work in various environments. These enzymes can also be genetically modified to have increased enzymatic activity. They are typically considered safe for industrial enzyme synthesis as they don't create toxic by-products or pollutants.

A protein's aliphatic index is used to calculate the relative volume of a protein by amino acids in the aliphatic side chain [121]; a greater aliphatic index is considered a good indicator of increased thermostability. The aliphatic index of all the sequences is high, ranging from 68.9 to 93.41, indicating that all the enzymes under study are probably significantly thermostable [54]. Additionally, experimental results showed that the proteases from *Bacillus subtilis* were 100 % stable between 35 and 55 °C and at a pH of 7.4 [122]. At 60 °C, a protease from *Bacillus cereus* displayed 100 % activity, and between pH 8 and 11, it retained approximately 80 % of its initial activity [123]. In these conditions, where other enzymes may denature or lose their activity at alkaline pH levels, the proteases being able to function effectively in alkaline conditions is an advantage. This feature is especially useful in industrial processes requiring a high pH environment for optimal efficiency. For example, their application in detergent formulations requires the protease to tolerate an alkaline pH. These experimental data validate the *in silico* analysis results.

According to the amino acid dispersal, Ala is the most prevalent amino acid, accounting for 12.53 % of the enzymes' main structure, whereas cysteine was the least abundant amino acid. Studies have shown that proteins with fewer cysteine residues tend to be more thermostable than those with a higher cysteine content, demonstrating the vital link between the amount of cysteine in a protein and its thermostability. This discovery can be explained by the fact that cysteine residues can create disulfide linkages that maintain the protein structure but are also vulnerable to thermal denaturation, reducing the protein's stability at high temperatures [124]. Hence, fewer cysteine residues may minimize the chance of a disulfide bond breaking and enhance the protein's thermostability. Ser, Gly, Val, Lys, Thr, and Leu were the other predominant amino acids. The high quantity of serine, a hydrophilic amino acid, suggests that these fibrinolytic proteases are extracellular in origin. This phenomenon is because serine proteases are enzymes typically secreted by cells and used for extracellular functions [125]. The fibrinolytic proteases' high serine amino acid content shows that these enzymes have evolved to work in the extracellular environment, where they can interact with extracellular substrates effectively and are capable of cleaving peptide bonds in their substrate(s) for the survival and growth of the producing bacterium [126,127]. Concerning industrial biotechnology, extracellular proteases-producing microorganisms have industrial significance. These enzymes are utilized in various industrial processes, such as the production of pharmaceuticals, detergents, and leather products [128].

Different conserved sites in fibrinolytic enzymes from *the Bacillus* genus were discovered using domain analysis [116]. The

existence of standard and unique domains among the fibrinolytic enzyme sequences may confer structural flexibility, directly impacting protease functional activity. These conserved areas could be used to design primers for PCR-based multiplication and cloning of these *Bacillus* protease genes [128]. The domains of the gene are highly conserved across different species and strains, indicating that they are crucial for the protein's functionality. These conserved areas can be amplified using PCR-based methods, enabling the gene to be cloned and studied. For example, Kim et al. [129] developed PCR primers for cloning and expressing *B. subtilis* using conserved domains of the genes encoding an aminopeptidase from *Bacillus licheniformis*. The enzyme was successfully expressed and exhibited the same biochemical properties and apparent molecular mass as the *B. licheniformis* original enzyme. In general, creating PCR primers for the multiplication and cloning of the genes encoding the *Bacillus* protease using conserved domains is a valuable technique for understanding the function and properties of these enzymes.

The presence of coiled regions in the secondary structure prediction tool SOPMA results indicated that the models were significantly conserved and stable [130,131]. The PDBsum data helps establish the general structural arrangement of proteins and predict ligand-binding pockets. As a result, the protein's secondary structural understanding may aid in predicting tertiary structures. On the contrary, secondary structure element prediction may be able to predict the tertiary structure of proteins despite the limitations of NMR and X-ray crystallography. X-ray crystallography has difficulty crystallizing a few proteins, and NMR is limited to relatively tiny protein molecules. Furthermore, it has been shown that secondary structural components are critical for identifying conformational modifications within the protein of concern. Therefore, the prediction of structure-function properties of proteases by *in silico* analysis has some added advantages.

Based on the model structure obtained from SWISS-MODEL, the sequence similarity percentage of the template-target pair was minimal for *Bacillus subtilis* (MN055601.1) with 26.98 % and maximum for *Bacillus subtilis* (KJ470773.1) with 99.21 %. QMEAN4 and the SAVES server analyzed the 3D protein model obtained from SWISS-MODEL. The QMEAN output characterized the global arrangement of variable protease residues by estimating the geometrical characteristics of the protein structure. It also compares the query model's ideal quality to an illustrative set of high-resolution X-ray crystallography-solved reference structures [132]. The resulting QMEAN z-score measures the structure's degree of nativeness [133]. For high-resolution models, the average z-score is '0'. Here, 23 sequences' query model's QMEAN Z-score was lower than the standard deviation of '1' from the mean value '0' of good models, recommending that the estimated models were equivalent to the high-resolution models. The z-scores of the QMEAN terms of the protein model of *Bacillus* sp. (UniProtKB: ABI35684.1) in Fig. 4c were −0.01, −1.09, −0.32, and 0.11 for Cβ interaction energy, all atom energy, salvation energy, and torsion energy, respectively.

The dark zone in Fig. 4d, which illustrates *Bacillus* sp. (UniProtKB: ABI35684.1) as a representative example of the protein sequences, indicates the region where the model had a Z-score below 1 (absolute number), based on evaluating the modified protein's absolute quality by comparing with the score of high-resolution z-ray structures of the same size. The advantage of such a comparison is that it provides a quantitative measure of the model's accuracy, allowing researchers to identify potential inaccuracies or errors in the structure, resulting in more accurate structural predictions, which can be valuable for understanding the protein's function and designing likely enzymes having tremendous industrial significance.

The best-quality models are typically expected to situate themselves in this dark zone. Only 23 question models were deemed suitable in this scenario, based on their position in the dark zone, indicated by a red marker. The SAVES server double-checks the protein structures and also analyses 3D and ERRAT. Understanding 3D protein models and estimating their correctness required these methods. Except for five sequences, all of the retrieved sequences' 3D models passed Verify 3D with a maximum score of 100, indicating that 100 % of their residues had an average 3D-1D score of 0.2 in the 3D/1D profile, which was acceptable. The structural verification algorithm in ERRAT construed the overall quality of the models with a maximum result score of 100 %; this score represented the percentage of the protein that fell below the rejection limit of 95 % [134]. As a result, the ERRAT software also confirmed that the protein's 3D shape is acceptable. The predicted structure of proteins was found to be good, stable, dependable, and consistent due to the above analyses.

Since the estimated number of amino acids in transmembrane helixes in the first 60 amino acids of 28 of these protein sequences is 19–22, the TMHMM tool indicates that these sequences have a transmembrane helix present in the protein, confirming the transmembrane helix in the N-terminal as a signal peptide. Signal peptides in proteins facilitate their proper localization and secretion within cells. A signal peptide must be present for secreted proteins to fold, stabilize, and perform as intended. Therefore, a signal peptide in a protein is directly tied to its extracellular nature [135,136].

Polypeptide chains define protein disorder with high flexibility and a lack of regular secondary structure [137]. For example, the blue-colored areas shown in Supplementary Fig. S10, which illustrate *Bacillus* sp. (UniProtKB: ABI35684.1), were chaotic regions. In *in-vivo*, many proteins are fundamentally disordered. Disordered regions are essential since many inherently disordered proteins occur as unstructured and arranged once they are attached to another molecule [138,139]. The total number of cleavages was found to be highest for MN055600.1 (*B. subtilis*) at 1674 and lowest for KJ470773.1 (*B. subtilis*) at 193, which could help conduct experiments with a segment of a protein, separating domains in a protein, and removing a tagged protein when trying to express a fusion protein, according to the results from the peptide cutter tool.

In conclusion, it may be stated that although there has been a significant improvement in our understanding of proteases' sequences, as well as the emergence of modern technologies enabling the production of recombinant proteins to cater to the needs of the industry, there is still much to learn when constructing a protein with the desirable applications in the field of the biotech industry. The *in-silico* analysis of the gene for the fibrinolytic enzyme produced by *Bacillus* spp provides a theoretical framework for its physiochemical properties, primary, secondary, and tertiary structure, functional analysis, domains and motifs, and protein model analysis that will help in the development of a potential mutant enzyme with desirable features like increased catalysis or activity, stability, and expression for its commercial application by changing a specific amino acid to a particular location. The characteristics of the enzyme

sequence JF739176.1 and MF677779.1 can be taken into account when developing a potent mutant enzyme, as both sequences have demonstrated to be stable at high pH with thermostability and exhibit αβ-fibrinogenase activity both experimentally and through *in-silico* analysis.

With a mid-molecular weight of 39125.16 Da and 39095.09 Da, the enzyme sequences JF739176.1 and MF677779.1 can be advantageous in pharmaceutical applications by enhancing tissue penetration and reducing immunogenicity. They would also work well with simple expression systems. The enzymes' computed isoelectric point and aliphatic index values of 9.23 and 9.16, as well as 82.02 and 82.28, indicate that they are thermostable, which is a crucial property for drug development since it helps maintain the structural integrity of proteins at elevated temperatures. The industrial extraction of these enzymes will be facilitated by their excellent water solubility, as indicated by their low GRAVY indices of −0.025 and −0.017. Experimental evidence shows these two enzymes can be considered potent fibrinolytic enzymes since they can break down fibrinogen's α- and β-chains in 10 and 60 min, respectively. This conclusion can be drawn from a comparative analysis of the experimental work and the *in-silico* analyzed data.

CRediT authorship contribution statement

Nitisha Boro: Writing – original draft, Investigation, Formal analysis. **Pedro Alexandrino Fernandes:** Methodology, Data curation. **Ashis K. Mukherjee:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

NB received CSIR-JRF (Council of Scientific and Industrial Research-Junior Research Fellow). This study received partial funding from the DBT-NER project (Department of Biotechnology-North Eastern Region) (102/IFD/SAN/1809/2018–2019) to AKM.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33895.

References

[1] E. Kotb, E. Kotb, Fibrinolytic Bacterial Enzymes with Thrombolytic Activity, Springer, 2012.
[2] G.A. Roth, D. Abate, K.H. Abate, S.M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim, Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017, The lancet 392 (2018) 1736–1788.
[3] J.W. Weisel, R.I. Litvinov, Fibrin formation, structure and properties. Fibrous Proteins: Structures Mechanisms, 2017, pp. 405–456.
[4] R.F. Doolite, Fibrinogen and fibrin(ELS), in: Encyclopedia of Life Sciences (ELS), John Wiley & Sons, Chichester, 2010.
[5] D. Kapetanios, C.D. Karkos, I. Pliatsios, M. Mitka, I.T. Giagtzidis, K. Konstantinidis, K.O. Papazoglou, Association between perioperative fibrinogen levels and the midterm outcome in patients undergoing elective endovascular repair of abdominal aortic aneurysms, Ann. Vasc. Surg. 56 (2019) 202–208.
[6] D. Parry, H. Al-Barjas, L. Chappell, T. Rashid, R. Ariëns, D. Scott, Haemostatic and fibrinolytic factors in men with a small abdominal aortic aneurysm, Journal of British Surgery 96 (2009) 870–877.
[7] F.S. Collaboration, Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis, JAMA 294 (2005) 1799–1809.
[8] B. Chelluboina, R. Vemuganti, Chronic kidney disease in the pathogenesis of acute ischemic stroke, J. Cerebr. Blood Flow Metabol. 39 (2019) 1893–1905.
[9] J. Klovaite, B.G. Nordestgaard, A. Tybjærg-Hansen, M. Benn, Elevated fibrinogen levels are associated with risk of pulmonary embolism, but not with deep venous thrombosis, Am. J. Respir. Crit. Care Med. 187 (2013) 286–293.
[10] E. Tatli, F. Ozcelik, M. Aktoz, Plasma fibrinogen level may predict critical coronary artery stenosis in young adults with myocardial infarction, Cardiol. J. 16 (2009) 317–320.
[11] J. Liu, Y. Zhang, C.J. Lavie, F.K. Tabung, J. Xu, Q. Hu, L. He, Y. Zhang, Associations of C-reactive protein and fibrinogen with mortality from all-causes, cardiovascular disease and cancer among US adults, Prev. Med. 139 (2020) 106044.
[12] P. Altes, P. Perez, C. Esteban, J.F. Sánchez Muñoz-Torrero, E. Aguilar, A.M. García-Díaz, L.R. Álvarez, P.E. Jiménez, J.C. Sahuquillo, M. Monreal, Raised fibrinogen levels and outcome in outpatients with peripheral artery disease, Angiology 69 (2018) 507–512.
[13] J. Stam, S. de Bruijn, G. deVeber, Anticoagulation for cerebral sinus thrombosis, Stroke 34 (2003) 1054–1055.
[14] M. Flemmig, M.F. Melzig, Serine-proteases as plasminogen activators in terms of fibrinolysis, J. Pharm. Pharmacol. 64 (2012) 1025–1039.
[15] D. Swan, N. Loughran, M. Makris, J. Thachil, Management of bleeding and procedures in patients on antiplatelet therapy, Blood Rev. 39 (2020) 100619.
[16] N.T. Thu, N. Khue, N.D. Huy, N.Q. Tien, N.H. Loc, Characterizations and fibrinolytic activity of serine protease from Bacillus subtilis C10, Curr. Pharmaceut. Biotechnol. 21 (2020) 110–116.
[17] D. Collen, Fibrin-selective thrombolytic therapy for acute myocardial infarction, Circulation 93 (1996) 857–865.
[18] M. Zaman, T. Akhtar, A. Azam, M.A. Al Mamun, M.M. Hoq, M.A. Mazid, Thrombolytic activity of alkaline protease purified from a mutant strain Bacillus licheniformis MZK05M9, Bangladesh Pharmaceut J 21 (2018) 63–70.
[19] A. Krishnamurthy, P.D. Belur, S.B. Subramanya, Methods available to assess therapeutic potential of fibrinolytic enzymes of microbial origin: a review, Journal of Analytical Science Technology 9 (2018) 1–11.
[20] E. Kotb, Activity assessment of microbial fibrinolytic enzymes, Appl. Microbiol. Biotechnol. 97 (2013) 6647–6665.
[21] E. Kotb, The biotechnological potential of fibrinolytic enzymes in the dissolution of endogenous blood thrombi, Biotechnol. Prog. 30 (2014) 656–672.

[22] E. Kotb, Purification and partial characterization of a chymotrypsin-like serine fibrinolytic enzyme from Bacillus amyloliquefaciens FCF-11 using corn husk as a novel substrate, World J. Microbiol. Biotechnol. 30 (2014) 2071–2080.
[23] E. Kotb, Purification and partial characterization of serine fibrinolytic enzyme from Bacillus megaterium KSK-07 isolated from kishk, a traditional Egyptian fermented food, Appl. Biochem. Microbiol. 51 (2015) 34–43.
[24] E. Kotb, The biotechnological potential of subtilisin-like fibrinolytic enzyme from a newly isolated Lactobacillus plantarum KSK-II in blood destaining and antimicrobials, Biotechnol. Prog. 31 (2015) 316–324.
[25] Y. Zhang, A. Wisner, Y. Xiong, C. Bon, A novel plasminogen activator from snake venom: purification, characterization, and molecular cloning (*), J. Biol. Chem. 270 (1995) 10246–10255.
[26] D.-M. Chung, N.-S. Choi, P.J. Maeng, H.K. Chun, S.-H. Kim, Purification and characterization of a novel fibrinolytic enzyme from chive (Allium tuberosum), Food Sci. Biotechnol. 19 (2010) 697–702.
[27] S. Majumdar, B. Sarmah, D. Gogoi, S. Banerjee, S.S. Ghosh, S. Banerjee, P. Chattopadhyay, A.K. Mukherjee, Characterization, mechanism of anticoagulant action, and assessment of therapeutic potential of a fibrinolytic serine protease (Brevithrombolase) purified from Brevibacillus brevis strain FF02B, Biochimie 103 (2014) 50–60.
[28] S. Majumdar, P. Chattopadhyay, A.K. Mukherjee, In vivo anticoagulant and thrombolytic activities of a fibrinolytic serine protease (Brevithrombolase) with the k-carrageenan-induced rat tail thrombosis model, Clinical Applied Thrombosis/Hemostasis 22 (2016) 594–598.
[29] A.K. Mukherjee, S.K. Rai, R. Thakur, P. Chattopadhyay, S.K. Kar, Bafibrinase: a non-toxic, non-hemorrhagic, direct-acting fibrinolytic serine protease from Bacillus sp. strain AS-S20-I exhibits in vivo anticoagulant activity and thrombolytic potency, Biochimie 94 (2012) 1300–1308.
[30] J. Yuan, J. Yang, Z. Zhuang, Y. Yang, L. Lin, S. Wang, Thrombolytic effects of Douchi Fibrinolytic enzyme from Bacillus subtilis LD-8547 in vitro and in vivo, BMC Biotechnol. 12 (2012) 1–9.
[31] B. Bora, A.D. Biswas, A.B. Gurung, A. Bhattacharjee, V.S.K. Mattaparthi, A.K. Mukherjee, An in silico approach to understand the structure–function properties of a serine protease (Bacifrinase) from Bacillus cereus and experimental evidence to support the interaction of Bacifrinase with fibrinogen and thrombin, J. Biomol. Struct. Dynam. 35 (2017) 622–644.
[32] M. Baweja, L. Nain, Y. Kawarabayasi, P. Shukla, Current technological improvements in enzymes toward their biotechnological applications, Front. Microbiol. 7 (2016) 206451.
[33] M. Baweja, R. Tiwari, P.K. Singh, L. Nain, P. Shukla, An alkaline protease from Bacillus pumilus MP 27: functional analysis of its binding model toward its applications as detergent additive, Front. Microbiol. 7 (2016) 1195.
[34] E. Gasteiger, C. Hoogland, A. Gattiker, S.e. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, Springer, 2005.
[35] T.L. Bailey, C. Elkan, Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers, 1994.
[36] R.A. Laskowski, V.V. Chistyakov, J.M. Thornton, PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids, Nucleic Acids Res. 33 (2005) D266–D268.
[37] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, Bioinformatics 22 (2006) 195–201.
[38] D. Eisenberg, R. Lüthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. (1997) 396–404. Elsevier.
[39] M.W. MacArthur, R.A. Laskowski, J.M. Thornton, Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy, Curr. Opin. Struct. Biol. 4 (1994) 731–737.
[40] J.U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (1991) 164–170.
[41] A. Krogh, B. Larsson, G. Von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[42] R. Linding, R.B. Russell, V. Neduva, T. Gibson, GlobPlot: Exploring protein sequences for globularity and disorder, Nucleic Acids Res. 31 (2003) 3701–3708.
[43] O. Matsuo, K. Okada, H. Fukao, Y. Tomioka, S. Ueshima, M. Watanuki, M. Sakai, Thrombolytic Properties of Staphylokinase, 1990.
[44] B.S. Mienda, A. Yahya, I. Galadima, M. Shamsir, An overview of microbial proteases for industrial applications, Res. J. Pharmaceut. Biol. Chem. Sci. 5 (2014) 388–396.
[45] F. Hamon, C. Masson-Lunven, B. Boutiere, C. Boyer-Neumann, M. Larrieu, E. Anglés-Cano, Factors affecting the stability of tPA and PAI-1 during storage and handling of human plasma for in vitro studies: implications in the determination of tPA and PAI-1 activities, Blood Coagul. Fibrinolysis 1 (1990) 393–400.
[46] E.C.A. Note, A guide to understanding extinction coefficients, with emphasis on spectrophotometric determination of protein concentration, Thermo (2013) 1–3. Scientific.
[47] A. Mushtaq, T.M. Ansari, G. Mustafa, M.A. Shad, J. Cruz-Reyes, A. Jamil, Isolation and characterization of nprB, a novel protease from Streptomyces thermovulgaris, Pak. J. Pharm. Sci. 33 (2020).
[48] R. Kapoor, S. Khowal, B.P. Panda, S. Wajid, Comparative genomic analyses of Bacillus subtilis strains to study the biochemical and molecular attributes of nattokinases, Biotechnol. Lett. 44 (2022) 485–502.
[49] K. Guruprasad, B.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Eng. Des. Sel. 4 (1990) 155–161.
[50] S. Rogers, R. Wells, M. Rechsteiner, Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis, Science 234 (1986) 364–368.
[51] M. Naveed, A.U. Khan, GPCR-MPredictor: multi-level prediction of G protein-coupled receptors using genetic ensemble, Amino Acids 42 (2012) 1809–1823.
[52] X.-X. Zhou, Y.-B. Wang, Y.-J. Pan, W.-F. Li, Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins, Amino Acids 34 (2008) 25–33.
[53] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.
[54] N.D. Rawlings, F.R. Morton, C.Y. Kok, J. Kong, A.J. Barrett, MEROPS: the peptidase database, Nucleic Acids Res. 36 (2007) D320–D325.
[55] J. Deacon, The Microbial World: Thermophilic Microorganisms, Institute of Cell and Molecular Biology. The University of Edinburgh, 2004.
[56] C. Leuschner, G. Antranikian, Heat-stable enzymes from extremely thermophilic and hyperthermophilic microorganisms, World J. Microbiol. Biotechnol. 11 (1995) 95–114.
[57] I. Lasa, J. Berenguer, Thermophilic enzymes and their biotechnological potential, Microbiologia 9 (1993) 77–89.
[58] F. Niehaus, C. Bertoldo, M. Kähler, G. Antranikian, Extremophiles as a source of novel enzymes for industrial application, Appl. Microbiol. Biotechnol. 51 (1999) 711–729.
[59] C. Sharma, G.E.M. Salem, N. Sharma, P. Gautam, R. Singh, Thrombolytic potential of novel thiol-dependent fibrinolytic protease from Bacillus cereus RSA1, Biomolecules 10 (2019) 3.
[60] R. Agrebi, A. Haddar, N. Hmidet, K. Jellouli, L. Manni, M. Nasri, BSF1 fibrinolytic enzyme from a marine bacterium Bacillus subtilis A26: purification, biochemical and molecular characterization, Process Biochem. 44 (2009) 1252–1259.
[61] F. Rigoldi, S. Donini, A. Redaelli, E. Parisini, A. Gautieri, Engineering of thermostable enzymes for industrial applications, APL Bioeng. 2 (2018).
[62] F. He, S. Hogan, R.F. Latypov, L.O. Narhi, V.I. Razinkov, High throughput thermostability screening of monoclonal antibody formulations, J. Pharmaceut. Sci. 99 (2010) 1707–1720.
[63] N. Pradeep, A. Anupama, K. Vidyashree, P. Lakshmi, In silico characterization of industrial important cellulases using computational tools, Adv. Life Sci. Technol. 4 (2012) 8–14.
[64] G.L. Amidon, H. Lennernäs, V.P. Shah, J.R. Crison, A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability, Pharmaceut. Res. 12 (1995) 413–420.
[65] C.A. Lipinski, F. Lombardo, B.W. Dominy, P. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 64 (2012) 4–17.

[66] Guidance for Industry: Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System. U.S. Food and Drug Administration., (2000).
[67] O. Carugo, Amino acid composition and protein dimension, Protein Sci. 17 (2008) 2187–2191.
[68] P.S. Kishnani, D. Corzo, N.D. Leslie, D. Gruskin, A. Van der Ploeg, J.P. Clancy, R. Parini, G. Morin, M. Beck, M.S. Bauer, Early treatment with alglucosidase alfa prolongs long-term survival of infants with Pompe disease, Pediatr. Res. 66 (2009) 329–335.
[69] E.C. Keystone, A.F. Kavanaugh, J.T. Sharp, H. Tannenbaum, Y. Hua, L.S. Teoh, S.A. Fischkoff, E.K. Chartash, Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti–tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial, Arthritis Rheum. 50 (2004) 1400–1411.
[70] C.N. Pace, J.M. Scholtz, A helix propensity scale based on experimental studies of peptides and proteins, Biophys. J. 75 (1998) 422–427.
[71] S.R. Trevino, J.M. Scholtz, C.N. Pace, Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa, J. Mol. Biol. 366 (2007) 449–460.
[72] R.J. Siezen, J.A. Leunissen, Subtilases: the superfamily of subtilisin-like serine proteases, Protein Sci. 6 (1997) 501–523.
[73] M.A. Tangrea, P.N. Bryan, N. Sari, J. Orban, Solution structure of the pro-hormone convertase 1 pro-domain from Mus musculus, J. Mol. Biol. 320 (2002) 801–812.
[74] S.C. Jain, U. Shinde, Y. Li, M. Inouye, H.M. Berman, The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 Å resolution, J. Mol. Biol. 284 (1998) 137–144.
[75] C. Jyotsna, P. Ashish, G. Shailendra, M. Verma, Homology modeling and binding site identification of 1 deoxy d-xylulose 5 phosphate reductoisomerase of Plasmodium falciparum: new drug target for Plasmodium falciparum, Int. J. Eng. Sci. Technol. 2 (2010) 3468–3472.
[76] O.F. Ezomo, K. Takahashi, Y. Horie, M.S. Mustak, S. Meshitsuka, Circular dichroism studies on C-terminal zinc finger domain of transcription factor GATA-2, Yonago Acta Med. 53 (2010) 25–28.
[77] P.W. Hildebrand, S. Lorenzen, A. Goede, R. Preissner, Analysis and prediction of helix–helix interactions in membrane channels and transporters, Proteins: structure, Function, Bioinformatics 64 (2006) 253–262.
[78] W.F. DeGrado, H. Gratkowski, J.D. Lear, How do helix–helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles, Protein Sci. 12 (2003) 647–665.
[79] A. Lo, Y.-Y. Chiu, E.A. Rødland, P.-C. Lyu, T.-Y. Sung, W.-L. Hsu, Predicting helix–helix interactions from residue contacts in membrane proteins, Bioinformatics 25 (2009) 996–1003.
[80] M.A. Lemmon, K.R. MacKenzie, I.T. Arkin, D.M. Engelman, in: G. von Heijne (Ed.), Membrane Protein Assembly, 1997.
[81] J.L. Popot, D.M. Engelman, Biochemistry 29 (1990) 4031–4037.
[82] A.E.C. Burba, U. Lehnert, E.Z. Yu, M. Gerstein, Helix Interaction Tool (HIT): a web-based tool for analysis of helix-helix interactions in proteins, Bioinformatics 22 (2006) 2735–2738.
[83] C.M. Deber, D.P. Ng, Helix-helix interactions: is the medium the message? Structure 23 (2015) 437–438.
[84] E.G. Hutchinson, R.B. Sessions, J.M. Thornton, D.N. Woolfson, Determinants of strand register in antiparallel β-sheets of proteins, Protein Sci. 7 (1998) 2287–2300.
[85] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annu. Rev. Biophys. 37 (2008) 215–246.
[86] X. Gao, X. Dong, X. Li, Z. Liu, H. Liu, Prediction of disulfide bond engineering sites using a machine learning method, Sci. Rep. 10 (2020) 10330.
[87] H. Takagi, T. Takahashi, H. Momose, M. Inouye, Y. Maeda, H. Matsuzawa, T. Ohta, Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease, J. Biol. Chem. 265 (1990) 6874–6878.
[88] M.W. Pantoliano, R.C. Ladner, P.N. Bryan, M.L. Rollence, J.F. Wood, T.L. Poulos, Protein engineering of subtilisin BPN': enhanced stabilization through the introduction of two cysteines to form a disulfide bond, Biochemistry 26 (1987) 2077–2082.
[89] N. C Harris, M.G. Achen, The proteolytic activation of angiogenic and lymphangiogenic growth factors in cancer–its potential relevance for therapeutics and diagnostics, Curr. Med. Chem. 21 (2014) 1821–1842.
[90] J.B. Buse, R.R. Henry, J. Han, D.D. Kim, M.S. Fineman, A.D. Baron, E.-C.S. Group, Effects of exenatide (exendin-4) on glycemic control over 30 weeks in sulfonylurea-treated patients with type 2 diabetes, Diabetes Care 27 (2004) 2628–2635.
[91] W. Kauzmann, Some factors in the interpretation of protein denaturation, Adv. Protein Chem. (1959) 1–63. Elsevier.
[92] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[93] M.C. Wilce, M.-I. Aguilar, M.T. Hearn, Physicochemical basis of amino acid hydrophobicity scales: evaluation of four new scales of amino acid hydrophobicity coefficients derived from RP-HPLC of peptides, Anal. Chem. 67 (1995) 1210–1219.
[94] S.A. Karadaghi, The 20 Amino Acids and Their Role in Protein Structures, Basics of Protein Structure, 2015.
[95] E. Gulezian, C. Crivello, J. Bednenko, C. Zafra, Y. Zhang, P. Colussi, S. Hussain, Membrane protein production and formulation for drug discovery, Trends Pharmacol. Sci. 42 (2021) 657–674.
[96] L.C. Bock, L.C. Griffin, J.A. Latham, E.H. Vermaas, J.J. Toole, Selection of single-stranded DNA molecules that bind and inhibit human thrombin, Nature 355 (1992) 564–566.
[97] J.G. Almeida, A.J. Preto, P.I. Koukos, A.M. Bonvin, I.S. Moreira, Membrane proteins structures: a review on computational modeling tools, Biochim. Biophys. Acta Biomembr. 1859 (2017) 2021–2039.
[98] G.M. Popowicz, A. Dömling, T.A. Holak, The structure-based design of MDM2/MDMX–p53 inhibitors gets serious, Angew. Chem. Int. Ed. 50 (2011) 2680–2688.
[99] M.E. Weinblatt, E.C. Keystone, D.E. Furst, L.W. Moreland, M.H. Weisman, C.A. Birbara, L.A. Teoh, S.A. Fischkoff, E.K. Chartash, Adalimumab, a fully human anti–tumor necrosis factor α monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial, Arthritis Rheum. 48 (2003) 35–45.
[100] F. Wang, S. Ning, B. Yu, Y. Wang, USP14: structure, function, and target inhibition, Front. Pharmacol. 12 (2022) 801328.
[101] R. Dhenni, T.G. Phan, The geography of memory B cell reactivation in vaccine-induced immunity and in autoimmune disease relapses, Immunol. Rev. 296 (2020) 62–86.
[102] V. Schellenberger, C.-w. Wang, N.C. Geething, B.J. Spink, A. Campbell, W. To, M.D. Scholle, Y. Yin, Y. Yao, O. Bogin, A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner, Nat. Biotechnol. 27 (2009) 1186–1190.
[103] Y. Ghasemi, F. Dabbagh, A. Ghasemian, Cloning of a fibrinolytic enzyme (subtilisin) gene from Bacillus subtilis in Escherichia coli, Mol. Biotechnol. 52 (2012) 1–7.
[104] K. Heo, K.M. Cho, C.K. Lee, G.M. Kim, J.-H. Shin, J.S. Kim, J.H. Kim, Characterization of a fibrinolytic enzyme secreted by Bacillus amyloliquefaciens CB1 and its gene cloning, J. Microbiol. Biotechnol. 23 (2013) 974–983.
[105] G.-M. Kim, A.-R. Lee, K.-W. Lee, A.-Y. Park, J.-Y. Chun, J.-H. Cha, Y.-S. Song, J.-H. Kim, Characterization of a 27 kDa fibrinolytic enzyme from Bacillus amyloliquefaciens CH51 isolated from cheonggukjang, J. Microbiol. Biotechnol. 19 (2009) 997–1004.
[106] Y. Peng, Q. Huang, R.-h. Zhang, Y.-z. Zhang, Purification and characterization of a fibrinolytic enzyme produced by Bacillus amyloliquefaciens DC-4 screened from douchi, a traditional Chinese soybean food, Comparative biochemistry physiology part b, Biochem. Mol. Biol. 134 (2003) 45–52.
[107] N.-S. Choi, K.-T. Chang, P. Jae Maeng, S.-H. Kim, Cloning, expression, and fibrin (ogen) olytic properties of a subtilisin DJ-4 gene from Bacillus sp. DJ-4, FEMS Microbiol. Lett. 236 (2004) 325–331.
[108] Z. Yao, J.A. Kim, J.H. Kim, Properties of a fibrinolytic enzyme secreted by Bacillus subtilis JS2 isolated from saeu (small shrimp) jeotgal, Food Sci. Biotechnol. 27 (2018) 765–772.
[109] H.-D. Jo, H.-A. Lee, S.-J. Jeong, J.-H. Kim, Purification and characterization of a major fibrinolytic enzyme from Bacillus amyloliquefaciens MJ5-41 isolated from Meju, J. Microbiol. Biotechnol. 21 (2011) 1166–1173.

[110] X. Wei, Y. Zhou, J. Chen, D. Cai, D. Wang, G. Qi, S. Chen, Efficient expression of nattokinase in Bacillus licheniformis: host strain construction and signal peptide optimization, J. Ind. Microbiol. Biotechnol. 42 (2015) 287–295.
[111] W. Cui, F. Suo, J. Cheng, L. Han, W. Hao, J. Guo, Z. Zhou, Stepwise modifications of genetic parts reinforce the secretory production of nattokinase in Bacillus subtilis, Microb. Biotechnol. 11 (2018) 930–942.
[112] M. Weng, X. Deng, W. Bao, L. Zhu, J. Wu, Y. Cai, Y. Jia, Z. Zheng, G. Zou, Improving the activity of the subtilisin nattokinase by site-directed mutagenesis and molecular dynamics simulation, Biochem. Biophys. Res. Commun. 465 (2015) 580–586.
[113] X. Liang, S. Jia, Y. Sun, M. Chen, X. Chen, J. Zhong, L. Huan, Secretory expression of nattokinase from Bacillus subtilis YF38 in Escherichia coli, Mol. Biotechnol. 37 (2007) 187–194.
[114] Z. Liu, W. Zheng, C. Ge, W. Cui, L. Zhou, Z. Zhou, High-level extracellular production of recombinant nattokinase in Bacillus subtilis WB800 by multiple tandem promoters, BMC Microbiol. 19 (2019) 1–14.
[115] W. Chen, Y. Zeng, L. Zheng, W. Liu, Q. Lyu, Discovery and characterization of a novel protease from the Antarctic soil, Process Biochem. 111 (2021) 270–277.
[116] C. Sharma, A. Nigam, R. Singh, Computational-approach understanding the structure-function prophecy of Fibrinolytic Protease RFEA1 from Bacillus cereus RSA1, PeerJ 9 (2021) e11570.
[117] F. Irshad, Z. Mushtaq, S. Akhtar, Sequence analysis and comparative bioinformatics study of camelysin gene (calY) isolated from Bacillus thuringiensis, Biochem. Genet. 56 (2018) 103–115.
[118] N. Rawlings, F. Morton, A. Barrett, J. In Polaina, A.P. MacCabe (Eds.), Industrial Enzymes Structure, Function and Applications, Springer, The Netherlands, 2007, pp. 161–180.
[119] A. Verma, V.K. Singh, S. Gaur, Computational based functional analysis of Bacillus phytases, Comput. Biol. Chem. 60 (2016) 53–58.
[120] K. Pramanik, T. Soren, S. Mitra, T.K. Maiti, In silico structural and functional analysis of Mesorhizobium ACC deaminase, Comput. Biol. Chem. 68 (2017) 12–21.
[121] V.K. Morya, S. Yadav, E.-K. Kim, D. Yadav, In silico characterization of alkaline proteases from different species of Aspergillus, Appl. Biochem. Biotechnol. 166 (2012) 243–257.
[122] R.A. Abusham, R.N.Z.R. Rahman, A.B. Salleh, M. Basri, Optimization of physical factors affecting the production of thermo-stable organic solvent-tolerant protease from a newly isolated halo tolerant Bacillus subtilis strain Rand, Microb. Cell Factories 8 (2009) 1–9.
[123] G. Pant, A. Prakash, J. Pavani, S. Bera, G. Deviram, A. Kumar, M. Panchpuri, R.G. Prasuna, Production, optimization and partial purification of protease from Bacillus subtilis, J. Taibah Univ. Sci. 9 (2015) 50–55.
[124] H. Yang, L. Liu, J. Li, J. Chen, G. Du, Rational design to improve protein thermostability: recent advances and prospects, ChemBioEng Rev. 2 (2015) 87–94.
[125] N.D. Rawlings, A.J. Barrett, R. Finn, Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Res. 44 (2016) D343–D350.
[126] M. Kuddus, Cold-active enzymes in food biotechnology: an updated mini review, J. Appl. Biol. Biotechnol. 6 (2018) 58–63.
[127] A.J. Ratner, R. Bryan, A. Weber, S. Nguyen, D. Barnes, A. Pitt, S. Gelber, A. Cheung, A. Prince, Cystic fibrosis pathogens activate Ca2+-dependent mitogen-activated protein kinase signaling pathways in airway epithelial cells, J. Biol. Chem. 276 (2001) 19267–19275.
[128] J. Liu, F. Song, J. Zhang, R. Liu, K. He, J. Tan, D. Huang, Identification of vip3A-type genes from Bacillus thuringiensis strains and characterization of a novel vip3A-type gene, Lett. Appl. Microbiol. 45 (2007) 432–438.
[129] J.-S. Kim, I.-S. Lee, S.-W. Lee, Y.-P. Lee, C.-H. Jung, H.-C. Kim, S.-Y. Choi, Cloning and expression of the aminopeptidase gene from the Bacillus lichenformis in Bacillus subtilis, J. Microbiol. Biotechnol. 12 (2002) 773–779.
[130] A. Hasan, H.H. Mazumder, A. Khan, M.U. Hossain, H.K. Chowdhury, Molecular characterization of legionellosis drug target candidate enzyme phosphoglucosamine mutase from Legionella pneumophila (strain Paris): an in silico approach, Genom. Inf. 12 (2014) 268.
[131] C. Geourjon, G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Bioinformatics 11 (1995) 681–684.
[132] P. Benkert, M. Künzli, T. Schwede, QMEAN server for protein model quality estimation, Nucleic Acids Res. 37 (2009) W510–W514.
[133] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics 27 (2011) 343–350.
[134] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, J. Mol. Biol. 293 (1999) 321–331.
[135] J. Liu, C.-L. Hsieh, O. Gelincik, B. Devolder, S. Sei, S. Zhang, S.M. Lipkin, Y.-F. Chang, Proteomic characterization of outer membrane vesicles from gut mucosa-derived fusobacterium nucleatum, J. Proteonomics 195 (2019) 125–137.
[136] J.P. van den Hombergh, P.J. van de Vondervoort, L. Fraissinet-Tachet, J. Visser, Aspergillus as a host for heterologous protein production: the problem of proteases, Trends Biotechnol. 15 (1997) 256–263.
[137] M.A. Hasan, M.A. Khan, A. Datta, M.H.H. Mazumder, M.U. Hossain, A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment, Mol. Immunol. 65 (2015) 189–204.
[138] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci. 11 (2002) 739–756.
[139] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, Intrinsically disordered protein, J. Mol. Graph. Model. 19 (2001) 26–59.