Bioinformatics analysis of extracellular subtilisin E from *Bacillus subtilis*

Reza Zolfaghari Emameh\(^a\), Justina Kazokaite\(^b\)\(^c\) and Bagher Yakhchali\(^a\)

\(^a\)Department of Energy and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran; \(^b\)Department of Biothermodynamics and Drug Design, Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

Communicated by Ramaswamy H. Sarma

**ABSTRACT**

_Bacillus_ spp. are the main sources of subtilisin E, which has several applications in biotechnology. The 3D structure of subtilisin E has a significant impact on its efficacy. In this study, we evaluated subtilisin E from _Bacillus subtilis_ subsp. _subtilis_ str. 168 by bioinformatic methods. The results revealed that the subtilisin E sequence from _B. subtilis_ contains highly conserved amino acids, including histidine (H), aspartic acid (D) and serine (S). Subtilisin E cleaves the bonds between hydrophobic and polar amino acids in keratin-associated proteins. The effects of point mutations on the crystal structure of subtilisin E (PDB ID: 1SCJ) showed that changes of asparagine 123 (N123) to valine (V) and serine 331 (S331) to leucine (L) respectively, were the most stabilizing. Genomic analysis of the subtilisin E-coding gene _aprE_ indicated that this gene and the _yhfN_ gene are expressed through a _σA_ promoter. The analysis of _TBFs_ revealed AbrB, ScoC, DegU, Hpr, σA, SinR, TenA, and DegU as relevant regulators of _aprE_ expression. Phylogenetic analysis showed that subtilisin Es have highly conserved structures among _Bacillus_ spp., sharing a common ancestor, where their coding genes were duplicated and evolved within the _Bacillus_ spp. As the conclusion, our _in silico_ study demonstrated that the overexpression of the _aprE_ gene and stability of the produced subtilisin E can be improved though system biology methods such as point mutations and identifying the involved transcription factors (TFs) or/and TBFs.

**Abbreviations:** 
\(A\): alanine; \(B\): basic local alignment search tool; \(C\): cysteine; \(D\): aspartic acid; \(G\): glycine; \(H\): histidine; \(I\): isoleucine; \(JPGV\): jena prokaryote genome viewer; \(L\): leucine; \(M\): methionine; \(MSA\): multiple sequence alignment; \(N\): asparagine; \(P\): phenyl alanine; \(PDB\): protein data bank; \(S\): serine; \(σA\): Sigma A; \(σH\): sigma H; \(SOC\): sporulation control; \(SPO\): sporulation; \(T\): threonine; \(TBF\): transcription binding factor; \(TF\): transcription factor; \(V\): valine; \(W\): tryptophan

**Introduction**

Subtilisin E, an alkaline serine protease belonging to the subtilase family of serine proteases, is produced by many prokaryotic and eukaryotic organisms as an extracellular enzyme with a molecular weight of 18 to 90 kDa (average \(\approx 40\) kDa) (Bouacem et al., 2015). Currently, over 60% of the enzyme market belongs to proteases with annual sales of $1.5–1.8 billion (Goncalves et al., 2016; Hirata et al., 2013). Subtilisin E is an enzyme applicable to many fields of industry, such as production of detergents, wool and leather, pharmaceuticals, cosmetics, and food (de Souza et al., 2015). The major application of subtilisin E in the food industry belongs to dairy, meat, and fish products (Vojcic et al., 2015). Moreover, the main usage of this enzyme in the wool and leather industries is for soaking and de-hairing. In addition, wound-healing and removal of dead tissues are the main applications of subtilisin E in medicine and the pharmaceutical industries.

Bacterial species and specifically *Bacillus* spp., such as _Bacillus subtilis_ (Wang et al., 2016), _Bacillus cereus_ (Nilegaonkar et al., 2007), _Bacillus licheniformis_ (Fakhfakh et al., 2009), and _Bacillus pumilus_ (Özçelik et al., 2014), have been recognized as the main sources of subtilisin E, which was produced for the first time in 1967 (Levison & Aronson, 1967). Subtilisin E exhibits a number of important features, such as retaining enzymatic activity upon changes of pH or temperature, stability in organic solvents, and resistant to toxic compounds, thereby making it the highest selling industrial enzyme globally (Chittoor et al., 2016; Contesini et al., 2017). This secretory enzyme has the highest catalytic activity at neutral to alkaline pH range (pH 7.0–11.0) and contains conserved histidine (H), serine (S), and aspartic acid (D) amino acids at its active site (Tjalsma et al., 2000). This enzyme is stabilized by calcium (Ca\(^{2+}\)) ions in some thermostables, while its decreased thermostability is associated with weak affinity for Ca\(^{2+}\) ions (Almog et al., 2003).
Some studies have shown that there is a significant association between subtilisin E production and secondary metabolites during the sporulation phase (Paetzol et al., 1998). A combination of UV radiation and N-methyl-N-nitro-N-nitrosoguanidine-yielded B. subtilis and B. pumilus mutants, which were able to secrete larger amounts of subtilisin E with higher hydrolytic activity (Wang et al., 2007, 2016). Furthermore, studies on genetic manipulation of B. pumilus revealed that the bacterium could produce 114% and 92% more subtilisin E after optimization of its growth conditions (Krug et al., 2014). More recently, optimization of the growth conditions and the alkaline protease promoter, respectively (Krug et al., 2016). Moreover, prediction of ligand-binding residues through comparisons with the 3D models of proteins was carried out using FunFOLD (http://www.reading.ac.uk/bioinf/) (Roche et al., 2011).

Substrate cleavage sites of subtilisin E

Prediction of substrate cleavage sites of subtilisin E from 8. subtilis subsp. subtilis str. 168 (UniProt ID: P04189) was conducted using the protease specificity prediction server (PROSPER), (https://prosper.erc.monash.edu.au/home.html) (Song et al., 2012). PROSPER is an integrated feature-based server for prediction of subtilisin Es’ cleaving properties regarding various proteinaceous substrates. The prediction mechanism of PROSPER is based on multiple factors, such as the secondary structure of substrate and solvent accessibility. For this prediction, bi-profile Bayesian features were applied to enable PROSPER to predict subtilisin E cleavage sites through four different combinations, including bi-profile Bayesian amino acid, disordered, secondary structure, and solvent accessibility profiles.

To perform subtilisin E substrate cleavage site predictions using PROSPER, four feather keratin protein sequences and four keratin-associated protein sequences from birds and mammals were extracted from the UniProt database, which included P02450 (Gallus gallus), P07521 (Columba livia), P08335 (Anas platyrhynchos), P02449 (Dromaius novaehollandiae), P02447 (Capra hircus), P02446 (Ovis aries), G3MYE1 (Bos taurus), and Q9BYR7 (Homo sapiens). Furthermore, the protease cleavage sites were predicted using the enzyme serine protease.

Effect of point mutations on the stability of subtilisin E

Predicting the stability of the crystal and 3D-visualized structure of subtilisin E (PDB ID: 1SCJ) revealed that the subtilisin E-propeptide complex includes two chains or asymmetric units: subtilisin E domain (A) and E-propeptide (B), which are autoprocessed by the enzyme cleavage activity (Jain et al., 1998; Sehnal et al., 2017) (Figure S1, supplementary material). In the present analysis, the sensitivity of point mutations and their stabilization were predicted at pH 7.0 using the MAESTRO (https://biwww.che.sbg.ac.at/maestro/web) web tool (Laimer et al., 2015). We performed this prediction to evaluate the effect of point mutations on the stability of subtilisin E. MAESTRO, a multi-agent machine learning system provides the following: (1) predicted ΔΔG (ΔΔG_{pre}) (kcal/mol): values of changes in standard binding free energy upon mutations, where ΔΔG_{pre} < 0.0 and ΔΔG_{pre} > 0.0 reveal the stabilizing and destabilizing mutations, respectively; (2) confidence estimation (C_{pre}); and (3) disulfide bond score (S_{s}).

Materials and methods

Structural analysis of the subtilisin E

A subtilisin E protein sequence (UniProt ID: ARW53276.1) from B. licheniformis (Ekici et al., 2008) was used for the initial UniProt BLAST protein homology search (http://www.uniprot.org/blast/) (UniProt, 2010). In the next step, 33 subtilisin E protein sequences (the first three amino acids prior to histidine (H) as the first highly conserved amino acid), including subtilisin E from B. subtilis subsp. subtilis str. 168 (UniProt ID: P04189), were aligned using the Clustal Omega algorithm to create a multiple sequence alignment (MSA) within the Jalview open source program (version 2.10.3) (http://www.jalview.org/) (Waterhouse et al., 2009; Zolfaghari Emameh et al., 2014). Subtilisin E from the fungus Dactylellina haptotyla (UniProt ID: B1NNU1) was applied as the query in the analysis.

The structural analysis of the subtilisin E protein sequence from B. subtilis subsp. subtilis str. 168 (P04189) was performed using the Swiss-Model-ExPASy protein modelling database (https://swissmodel.expasy.org/) (Biasini et al., 2014). Moreover, prediction of ligand-binding residues through comparisons with the 3D models of proteins was carried out using FunFOLD (http://www.reading.ac.uk/bioinf/) (Roche et al., 2011).

Effect of point mutations on the stability of subtilisin E

Predicting the stability of the crystal and 3D-visualized structure of subtilisin E (PDB ID: 1SCJ) revealed that the subtilisin E-propeptide complex includes two chains or asymmetric units: subtilisin E domain (A) and E-propeptide (B), which are autoprocessed by the enzyme cleavage activity (Jain et al., 1998; Sehnal et al., 2017) (Figure S1, supplementary material). In the present analysis, the sensitivity of point mutations and their stabilization were predicted at pH 7.0 using the MAESTRO (https://biwww.che.sbg.ac.at/maestro/web) web tool (Laimer et al., 2015). We performed this prediction to evaluate the effect of point mutations on the stability of subtilisin E. MAESTRO, a multi-agent machine learning system provides the following: (1) predicted ΔΔG (ΔΔG_{pre}) (kcal/mol): values of changes in standard binding free energy upon mutations, where ΔΔG_{pre} < 0.0 and ΔΔG_{pre} > 0.0 reveal the stabilizing and destabilizing mutations, respectively; (2) confidence estimation (C_{pre}); and (3) disulfide bond score (S_{s}).
Genomic location of subtilisin E (aprE) gene

The genomic location of *B. subtilis* subsp. *subtilis* str. 168 subtilisin E (*aprE*) gene was identified using the Jena Prokaryote Genome Viewer (JPGV) (Romualdi et al., 2007). JPGV is a freely accessible web tool which provides various informative data on sequenced prokaryotic genomes and allows visualization of genome plots with the exact location of individual genes. Moreover, this web tool can be used in horizontal gene transfer (HGT) and evolutionary studies to identify the location of genes on mobile genetic elements (MGEs) (Zolfaghari Emameh et al., 2018; Zolfaghari et al., 2016).

Phylogenetic analysis

A total of 33 subtilisin E sequences were retrieved from the UniProt database (Zolfaghari Emameh et al., 2020). The phylogenetic tree was constructed by using the Molecular Evolutionary Genetics Analysis (MEGA) software (version 7.0) (Kumar et al., 2016) through the maximum likelihood method. This method applies Bayesian inference to predict relative branch-length differences and presents a high-quality tree. In the current analysis, the alignment transformation environment (ALTER) tool (http://sing.ei.uvigo.es/ALTER/) (Glez-Pena et al., 2010) was used to convert MSA of the coding DNA sequences (CDSs) into an aligned codon. Subtilisin E from *D. haptotyla* (UniProt ID: B1NNU1) was included in the analysis as an outgroup.

Prediction of transcription binding factors (TBFs)

The subtilisin E (*aprE*) gene from *B. subtilis* subsp. *subtilis* str. 168 was evaluated by DBTBS (http://dbtbs.hgc.jp/) (Tjalsma et al., 2000), which is a reference database used for the study of transcriptional regulation in *B. subtilis* based on characterized TBFs, targeted genes for regulation, and motifs recognized by transcription factors (TFs). The latest DBTBS release (release 5.0) is capable of providing some updated information, including number and name of TFs, position-specific scoring matrices, promoters, regulated operons, and terminators. The release 5.0 of the database predicted TBFs for the *aprE* gene among 44 binding factor candidates. In addition, predictions were evaluated by the Bacillus gene Regulatory Network (BacillusRegNet) (http://vm-bacillusregnet.bioswarm.net/v1/index.html) (Misirli et al., 2014). The database is supported with a version containing experimental results, which provides a platform for validation of TBF predictions.

Results

Structural analysis of the subtilisin E

The MSA analysis of 33 subtilisin E protein sequences, including that from *B. subtilis* subsp. *subtilis* str. 168, showed that all sequences contain three highly conserved amino acids: histidine (H125), aspartic acid (D146), and serine (S231) (Figure 1).

Cleavage sites in the substrates of subtilisin E

Prediction of cleavage sites of subtilisin E from *B. subtilis* subsp. *subtilis* str. 168 in keratin and keratin-associated proteins from different hosts are shown in Figure 2 and Table 1.

Effect of point mutations on the stability of subtilisin E

The stability prediction of the crystal structure of subtilisin E (PDB ID: 1SCJ) revealed that point mutations can increase...
the stability of subtilisin E protein structure. The stabilizing mutations ($\Delta \Delta G_{\text{pred}} < 0.0$) and confidence estimation ($C_{\text{pred}}$) for both chains A and B of 1SCJ are shown in Supplementary data Tables S1 and S2, respectively. This evaluation revealed that the point mutations leading to changes of asparagine 123 (N123) to valine (V), isoleucine (I), and leucine (L) with $\Delta \Delta G_{\text{pred}} = -3.059$, $-2.978$, and $-2.967$, respectively, are the most stabilizing point mutations in chain A of subtilisin E. In addition, the point mutations causing the change of serine 331 (S331) to leucine (L), methionine (M), and isoleucine (I) with $\Delta \Delta G_{\text{pred}} = -1.697$, $-1.607$, and $-1.505$, respectively, are the most stabilizing point mutations in chain B of subtilisin E.

Genomic location of subtilisin E (aprE) gene

Analysis of the genomic location of the subtilisin E-coding gene (aprE) revealed that the aprE gene is located on the genome of B. subtilis subsp. subtilis str. 168 between 1,104,423 and 1,105,568 nucleotides (1,146 bp) (Figure 3). This genomic visualization showed that the aprE gene is expressed from a SigA ($\sigma$A) promoter. Moreover, the sporulation genes, including yhfN and yhfM, are located upstream of the aprE gene, whereas yhfO, yhfP, and yhfQ are downstream of the aprE gene. Both aprE and yhfN are located in a single operon, so the yhfN gene is expressed through the $\sigma$A promoter, similarly as the aprE gene.

Phylogenetic analysis

The phylogenetic results of this study with 100% data coverage for internal nodes revealed that the structure of subtilisin E from B. subtilis is highly conserved among various Bacillus spp. It is most likely due to aprE gene duplications, which have evolved in Bacillus spp. during the evolution of subtilisin E. Therefore, a high degree of similarity is observed between subtilisin E of B. subtilis subsp. subtilis and subtilisin E of Bacillus spp. from other branches of the phylogenetic tree. The analysis showed that aprE genes from B. subtilis subsp. subtilis str. 168 and B. mojavensis have a common ancestor (Figure 4).

Prediction of TBFs

Prediction of TBFs of the aprE gene identified five binding factors, namely AbrB, DegU, Hpr, $\sigma$A, and SinR, which actively bind to the relevant positions and act as negative or positive regulators, or recognize the aprE gene promoter (Table 2). Among these TBFs, $\sigma$A and DegU bind to the promoter regions as positive regulators, while phosphorylated
Spo0A acts as a negative regulator and dephosphorylated DegU acts as a positive regulator of AbrB, ScoC, and SinR (Figure 5). The validation of TBF predictions revealed that in addition to AbrB, ScoC, and SinR, other TBFs including TenA, rA, and DegU are involved in the regulation of the aprE gene (Table 3).

**Discussion**

Suitable production of recombinant subtilisin E needs identification of genomic and proteomic properties of the coding gene and amino acid structure of the protein through various omics approaches.

The MSA analysis of subtilisin E protein sequences revealed that three amino acids including histidine (H125), aspartic acid (D146), and serine (S231) are highly conserved in all Bacillus spp. as well as B. subtilis subsp. subtilis str. 168. Moreover, the results from the prediction of cleavage sites in the keratin and keratin-associated proteins indicated that this enzyme cleaves the binding sites between hydrophobic [valine (V), isoleucine (I), glycine (G), and proline (P)] and polar [glutamine (G), asparagine (N), serine (S), threonine (T), cysteine (C), and tryptophan (W)] amino acids. In addition, this analysis revealed that subtilisin E cleaves the hair keratins of Homo sapiens (human) and Bos taurus (bovine) as well as wool keratin of Capra hircus (goat) to further segments. Therefore, the hair and wool from human, bovine, and goat sources can be more suitable substrates for subtilisin E.

Furthermore, the present study on point mutations demonstrated that the most stabilizing mutations occurred by changing asparagine 123 (N123) to valine (V), lysine (L), and isoleucine (I) in chain A, and serine 331(S331) to lysine (L), methionine (M), and isoleucine (I) in chain B. This result indicated that the bond between two hydrophobic residues translocates the enzyme cleavage site from the surface to the core and generates a more stabilized protein structure. Moreover, the two chains of subtilisin E differ by biological function. Chain A of the subtilisin E protein exhibits enzymatic activity, whereas chain B functions as an intramolecular chaperon (Jain et al., 1998). Thus, we hypothesize that chain A could be the relevant target of further studies involving cleavage and site-directed mutagenesis leading to improved protein stability and higher yield.

Based on the genomic location of the sporulation genes, such as yhfN, yhfM, yhfO, yhfP, and yhfQ, downstream and upstream of the aprE gene, the σA promoter can be employed for both aprE and yhfN genes, and the intergenic region between the aprE and yhfN genes. Therefore, an association between the expression of the aprE gene and sporulation genes has been speculated, emphasizing the regulatory role of sporulation genes in the expression of the aprE gene (Hertel et al., 2017). This result is consistent with our previous experiments, where the expression of protease in an asporogenic B. cereus strain was activated by the induction of the sucrose promoter of the sacPA operon and expression of the sigma-H and spo0B gene products (Ahmadpour & Yakhchali, 2017). The present phylogenetic analysis showed the same origin for the coding genes of

### Table 1. Subtilisin E cleavage sites on the keratin and keratin-associated proteins.

| Host                  | Keratin or keratin-associated protein IDs (UniProt) | Recommended name | Position | Segment         | Score |
|-----------------------|---------------------------------------------------|-----------------|----------|-----------------|-------|
| Gallus gallus (Chicken) | P02450                                            | Feather keratin 1 | 41       | PVVV TLPG       | 1.15  |
|                        |                                                   |                 | 76       | GPVS SSGG       | 1.12  |
|                        |                                                   |                 | 34       | RVVQ QPSG       | 1.10  |
| Columba livia (Rock dove) | P07521                                           | Feather keratin B-4 | 40      | PVVV TLPG       | 1.15  |
|                        |                                                   |                 | 74       | GPVS SSGG       | 1.12  |
|                        |                                                   |                 | 33       | RVVQ QPSG       | 1.10  |
| Anas platyrhynchos (Mallard) | P08335                                         | Feather keratin B-4 | 40      | PVVV TLPG       | 1.15  |
|                        |                                                   |                 | 74       | GPVS SSGG       | 1.12  |
|                        |                                                   |                 | 33       | RVVQ QPSG       | 1.10  |
| Dromaius novaehollandiae (Ostrich) | P02449                     | Feather keratin | 60       | NTWV GGSS       | 1.33  |
|                        |                                                   |                 | 44       | PVVV TLPG       | 1.15  |
|                        |                                                   |                 | 80       | GPVS SSGG       | 1.12  |
| Capra hircus (Goat)    | P02447                                            | Wool keratin    | 80       | LGSL NLTT       | 1.23  |
|                        |                                                   |                 | 19       | ATTI CSSD       | 1.10  |
|                        |                                                   |                 | 30       | RCVL CLPS       | 1.03  |
|                        |                                                   |                 | 53       | NSPV PYHV       | 1.02  |
| Ovis aries (Sheep)     | P02446                                            | Wool keratin    | 80       | LGSL NLTT       | 1.23  |
|                        |                                                   |                 | 19       | ATTI CSSD       | 1.10  |
|                        |                                                   |                 | 30       | RCVL CLPS       | 1.03  |
| Bos taurus (Bovine)    | G3MYE1                                            | Hair keratin    | 80       | LGSL NLTT       | 1.23  |
|                        |                                                   |                 | 19       | ATTI CSSD       | 1.10  |
|                        |                                                   |                 | 64       | YVIP CWLL       | 1.10  |
|                        |                                                   |                 | 30       | RCVL CLPS       | 1.03  |
|                        |                                                   |                 | 86       | TFTP SPGC       | 1.00  |
| Homo sapiens (Human)   | Q9BYR7                                            | Hair keratin    | 19       | ATTI CSSD       | 1.10  |
|                        |                                                   |                 | 40       | PHTV            |       |
subtilisin E and revealed a conserved subtilisin E structure in all *Bacillus* spp. Although the subtilisin Es from *B. subtilis subsp. subtilis* str. 168 (UniProt ID: P04189) and *B. mojavensis* (UniProt ID: AGL76445.1) have a common ancestor, the enzyme from the former species has undergone further evolutionary changes. Thus, the identification of TBFs could...
provide an in vitro opportunity to overexpress the aprE gene through activation of DegU and σA, and thereafter increase the production yield of subtilisin E.

Conclusions

Several genomic and structural factors, including single mutations and TBFs, may have crucial roles in the expression and stability of the recombinant subtilisin E. Hence, the application of genetic and protein engineering as well as system biology techniques can potentially improve the stability of Bacillus spp. in both bench and large-scale production.

Acknowledgements

To perform this study, RZE and BY received a research grant from the National Institute of Genetic Engineering and Biotechnology (NIGEB) of Islamic Republic of Iran. We would like to thank Dr. Parvin Shariati (NIGEB, Tehran, Iran) for the English editing of the manuscript.

Disclosure statement

The authors declare that they have no conflict of interests.

Funding

To perform this study, RZE received a research grant support from the National Institute of Genetic Engineering and Biotechnology (NIGEB) of the Islamic Republic of Iran.

References

Abe, S., Yasumura, A., & Tanaka, T. (2009). Regulation of Bacillus subtilis aprE expression by glnA through inhibition of scoC and sigma(D)-dependent degR expression. Journal of Bacteriology, 191(9), 3050–3058. https://doi.org/10.1128/JB.00049-09
Ahmadpour, F., & Yakhchali, B. (2017). Development of an asporogenic Bacillus cereus strain to improve keratinase production in exponential phase by switching sigmaH on and sigmaF off. FEMS Microbiology Letters, 364(24). https://doi.org/10.1093/femsle/fnx216
Almog, O., Gonzalez, A., Klein, D., Greenblatt, H. M., Braun, S., & Shoham, G. (2003). The 0.93A crystal structure of sphericase: A calcium-loaded serine protease from Bacillus sphaericus. Journal of Molecular Biology, 332(5), 1071–1082. https://doi.org/10.1016/j.jmb.2003.07.011
Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res, 42(Web Server issue), W252–8. https://doi.org/10.1093/nar/gku340
Bouacem, K., Bouanane-Darenfed, A., Laribi-Habchi, H., Elhoul, M. B., Hmida-Sayari, A., Hacene, H., Ollivier, B., Fardeau, M.-L., Jaouadi, B., & Bejar, S. (2015). Biochemical characterization of a detergent-stable serine alkaline protease from Caldicoprobacter guelmiensis. International Journal of Biological Macromolecules, 81, 299–307. https://doi.org/10.1016/j.ijbiomac.2015.08.011
Chittoor, J. T., Balaji, L., & Jayaraman, G. (2016). Optimization of parameters that affect the activity of the alkaline protease from halotolerant bacterium, Bacillus acidarmanii VTP4, by the application of response surface methodology and evaluation of the storage stability of the enzyme. Iranian Journal of Biotechnology, 14(1), 23–32. https://doi.org/10.15171/ijb.1269
Contesini, F. J., Melo, R. R., & Sato, H. H. (2017). An overview of Bacillus spp. in both bench and large-scale production. Brazilian Journal of Microbiology, 46(2), 337–346. https://doi.org/10.1590/1517-838246220140359
Degering, C., Eggett, T., Pul, M., Bongaerts, J., Evers, S., Maurer, K.-H., & Jaeger, K.-E. (2010). Optimization of protease secretion in Bacillus subtilis and Bacillus licheniformis by screening of homologous and heterologous signal peptides. Applied and Environmental Microbiology, 76(19), 6370–6376. https://doi.org/10.1128/AEM.01146-10
Ekici, O. D., Paetzel, M., & Dalbey, R. E. (2008). Unconventional serine protease: Variations on the catalytic Ser/His/Asp triad configuration. Protein Science, 17(12), 2023–2037. https://doi.org/10.1110/ps.035436.108
Fakhfakh, N., Kanoun, S., Manni, L., & Nasri, M. (2009). Production and biochemical and molecular characterization of a keratinolytic serine protease from chicken feather-degrading Bacillus licheniformis RPK. Canadian Journal of Microbiology, 55(4), 427–436. https://doi.org/10.1139/W09-013
Glez-Peña, D., Gomez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., & Posada, D. (2010). ALTER: Program-oriented conversion of DNA and protein alignments. Nucleic Acids Res, 38(Web Server issue), W14–8. https://doi.org/10.1093/nar/gkq321
Gonçalves, R. N., Gozzini Barbosa, S. D., & da, S.-L. R. (2016). Proteases from Canavalia ensiformis: Active and thermostable enzymes with potential of application in biotechnology. Biotechnology Research International, 2016, 1–11. https://doi.org/10.1155/2016/3427098
Hertel, R., Meyerjurgens, S., Voigt, B., Liesegang, H., & Volland, S. (2017). Small RNA mediated repression of subtilisin production in Bacillus licheniformis. Scientific Reports, 7(1), 5699. https://doi.org/10.1038/s41598-017-05628-y
Hirata, A., Hori, Y., Koga, Y., Okada, J., Sakudo, A., Ikuta, K., Kanaya, S., & Takano, K. (2013). Enzymatic activity of a subtilisin homolog, Tk-SP, from Thermococcus kodakarenensis in detergents and its ability to degrade the abnormal prion protein. BMC Biotechnology, 13, 19. https://doi.org/10.1186/1472-6750-13-19
Jain, S. C., Shinde, U., Li, Y., Inouye, M., & Berman, H. M. (1998). The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 A resolution. Journal of Molecular Biology, 284(1), 137–144. https://doi.org/10.1006/jmbi.1998.2161
Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, 33(7), 1807–1817. https://doi.org/10.1093/molbev/msw054
Küppers, T., Steffen, V., Hellmuth, H., O’Connell, T., Bongaerts, J., Maurer, K.-H., & Wiechert, W. (2014). Developing a new production host from...
a blueprint: Bacillus pumilus as an industrial enzyme producer. *Microbial Cell Factories*, 13(1), 46. https://doi.org/10.1186/1475-2859-13-46

Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., & Lackner, P. (2015). MAESTRO-multi agent stability prediction upon point mutations. *BMC Bioinformatics*, 16, 116. https://doi.org/10.1186/s12859-015-0548-6

Levisohn, S., & Aronson, A. I. (1967). Regulation of extracellular protease production in Bacillus cereus. *Journal of Bacteriology*, 93(3), 1023–1030. https://doi.org/10.1128/JB.93.3.1023-1030.1967

Misirli, G., Hallinan, J., Rottger, R., Baumbach, J., & Wipat, A. (2014). BacillusRegNet: A transcriptional regulation database and analysis platform for Bacillus species. *Journal of Integrative Bioinformatics*, 11(2), 244. https://doi.org/10.2390/ib-j-2014-2014

Nilegaonkar, S. S., Zambare, V. P., Kanekar, P. P., Dhakephalkar, P. K., & Sarnaik, S. S. (2007). Production and partial characterization of dehairing protease from Bacillus cereus MCM B-326. *Bioresource Technology*, 98(6), 1238–1245. https://doi.org/10.1016/j.biortech.2006.05.003

Özçelik, B., Aytar, P., Gedikli, S., Yardımcı, F., & Çabuk, A. (2014). Production of an alkaline protease using Bacillus pumilus D3 without inactivation by SDS, its characterization and purification. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 29(3), 388–396. https://doi.org/10.3109/14756366.2013.788503

Paetzel, M., Dalbey, R. E., & Strynadka, N. C. (1998). Crystal structure of a bacterial signal peptide in complex with a beta-lactam inhibitor. *Nature*, 396(6707), 186–190. https://doi.org/10.1038/24196

Roche, D. B., Tetchner, S. J., & McGuffin, L. J. (2011). FunFOLD: An improved automated method for the prediction of ligand binding residues using 3D models of proteins. *BMC Bioinformatics*, 12, 160. https://doi.org/10.1186/1471-2105-12-160

Romualdi, A., Felder, M., Rose, D., Gaissmann, U., Schilhabel, M., Glöckner, G., Platzer, M., & Sühnel, J. (2007). GenColors: Annotation and comparative genomics of prokaryotes made easy. *Methods in Molecular Biology (Clifton, N.J.*), 395, 75–96.

Sehnal, D., Deshpande, M., Vareković, R. S., Mir, S., Berka, K., Midlik, A., Pravda, L., Velankar, S., & Koca, J. (2017). LiteMol suite: Interactive web-based visualization of large-scale macromolecular structure data. *Nature Methods*, 14(12), 1112–1122. https://doi.org/10.1038/nmeth.4499

Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., Whiststock, J. C., & Pike, R. N. (2012). PROSPER: An integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One*, 7(11), e50300. https://doi.org/10.1371/journal.pone.0050300

Takagi, H., Morinaga, Y., Ikemura, H., & Inouye, M. (1988). Mutant subtilisin E with enhanced protease activity obtained by site-directed mutagenesis. *The Journal of Biological Chemistry*, 263(36), 19592–19596.

Takagi, H., Takahashi, T., Momose, H., Inouye, M., Maeda, Y., Matsuzawa, H., & Ohta, T. (1990). 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. *The Journal of Biological Chemistry*, 265(12), 6874–6878.

Tjalsma, H., Stover, A. G., Driks, A., Venema, G., Bron, S., & van Dijl, J. M. (2000). Conserved serine and histidine residues are critical for activity of the ER-type signal peptidase SipW of Bacillus subtilis. *The Journal of Biological Chemistry*, 275(33), 25102–25108. https://doi.org/10.1074/jbc.M002676200

UniProt, C. (2010). The universal protein resource (UniProt) in 2010. *Nucleic Acids Research*, 38(Database issue), D142–8.

Vojicic, L., Pitcler, C., Korfer, G., Jakob, F., Ronny, M., & Maurer, K. H. (2015). Advances in protease engineering for laundry detergents. *New Biotechnology*, 32(6), 629–634. https://doi.org/10.1016/j.nbt.2014.12.010

Wang, X. C., Zhao, H. Y., Liu, G., Cheng, X. J., & Feng, H. (2016). Improving production of extracellular proteases by random mutagenesis and biochemical characterization of a serine protease in Bacillus subtilis S1-4. *Genetics and Molecular Research*, 15(2), gmr7831.

Wang, H. Y., Liu, D. M., Liu, Y., Cheng, C. F., Ma, Q. Y., Huang, Q., & Zhang, Y. Z. (2007). Screening and mutagenesis of a novel Bacillus pumilus strain producing alkaline protease for dehairing. *Letters in Applied Microbiology*, 44(1), 1–6. https://doi.org/10.1111/j.1472-765X.2006.02039.x

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, 25(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Zolfaghari Emameh, R., Barker, H. R., Hytonen, V. P., & Parkkila, S. (2018). Involvement of beta-carbonic anhydrase genes in bacterial genomic islands and their horizontal transfer to protists. *Applied and Environmental Microbiology*, 84(15), e00771-18. https://doi.org/10.1128/AEM.00771-18

Zolfaghari Emameh, R., Barker, H., Tolvanen, M. E., Oruttay, C., & Parkkila, S. (2014). Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans. *Parasites & Vectors*, 7, 38. https://doi.org/10.1186/1756-3305-7-38

Zolfaghari Emameh, R., Kuuslahti, M., Nosrati, H., Lohi, H., & Parkkila, S. (2020). Assessment of databases to determine the validity of beta- and gamma-carbonic anhydrase sequences from vertebrates. *BMC Genomics*, 21(1), 352. https://doi.org/10.1186/s12864-020-6762-2

Zolfaghari, E. R., Barker, H. R., Tolvanen, M. E., Parkkila, S., & Hytonen, V. P. (2016). Horizontal transfer of beta-carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes. *Parasit Vectors*, 9, 152. https://doi.org/10.1186/s13071-016-1415-7