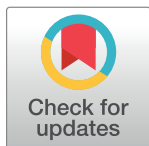# PLOS ONE

RESEARCH ARTICLE

In silico functional annotation of hypothetical proteins from the *Bacillus paralicheniformis* strain Bac84 reveals proteins with biotechnological potentials and adaptational functions to extreme environments

Md. Atikur Rahman[1], **Uzma Habiba Heme**[2], **Md. Anowar Khasru Parvez**[3]*

1 Institute of Microbiology, Friedrich Schiller University Jena, Thuringia, Germany, **2** Faculty of Biological Sciences, Friedrich Schiller University Jena, Thuringia, Germany, **3** Department of Microbiology, Jahangirnagar University, Savar, Dhaka, Bangladesh

* khasru73@juniv.edu

Abstract

Members of the *Bacillus* genus are industrial cell factories due to their capacity to secrete significant quantities of biomolecules with industrial applications. The *Bacillus paralicheniformis* strain Bac84 was isolated from the Red Sea and it shares a close evolutionary relationship with *Bacillus licheniformis*. However, a significant number of proteins in its genome are annotated as functionally uncharacterized hypothetical proteins. Investigating these proteins' functions may help us better understand how bacteria survive extreme environmental conditions and to find novel targets for biotechnological applications. Therefore, the purpose of our research was to functionally annotate the hypothetical proteins from the genome of *B. paralicheniformis* strain Bac84. We employed a structured in-silico approach incorporating numerous bioinformatics tools and databases for functional annotation, physicochemical characterization, subcellular localization, protein-protein interactions, and three-dimensional structure determination. Sequences of 414 hypothetical proteins were evaluated and we were able to successfully attribute a function to 37 hypothetical proteins. Moreover, we performed receiver operating characteristic analysis to assess the performance of various tools used in this present study. We identified 12 proteins having significant adaptational roles to unfavorable environments such as sporulation, formation of biofilm, motility, regulation of transcription, etc. Additionally, 8 proteins were predicted with biotechnological potentials such as coenzyme A biosynthesis, phenylalanine biosynthesis, rare-sugars biosynthesis, antibiotic biosynthesis, bioremediation, and others. Evaluation of the performance of the tools showed an accuracy of 98% which represented the rationality of the tools used. This work shows that this annotation strategy will make the functional characterization of unknown proteins easier and can find the target for further investigation. The knowledge of these hypothetical proteins' potential functions aids *B. paralicheniformis* strain Bac84 in effectively creating a new biotechnological target. In addition, the results may also facilitate a better understanding of the survival mechanisms in harsh environmental conditions.

Introduction

Bacillus paralicheniformis is a newly discovered species in the *Bacillus* genus [1]. It is phylogenetically closely related to *B. licheniformis* [1, 2]. In the biotechnology sector, *B. licheniformis* has already been employed to produce biochemicals, enzymes, antibiotics, and other products [1, 3]. Several current investigations have indicated that *B. paralicheniformis* species have a strong potential for the biosynthesis of antimicrobial compounds [4, 5]. One of the strains can also inhibit plant pathogenic microbes [6]. In this way, *B. paralicheniformis* may be of biotechnological relevance but still, it has remained largely unexplored.

B. paralicheniformis is a gram-positive, facultatively anaerobic, rod-shaped, motile, and endospore-forming *Bacillus* species [1]. The *B. paralicheniformis* strains are found in a variety of habitats, including soil, freshwater, marine, and niches associated with food [1, 4, 6]. This strain is adapted to survive in extreme conditions such as high osmolarity which provides it with metabolic capabilities similar to industrial strains [4]. The *B. paralicheniformis* strain Bac84 was isolated from the Red Sea which is an ecosystem of harsh, extremely saline, and high temperature [4]. Hence, this strain may be a potential microbial cell factory to produce both thermo-tolerant and osmotolerant enzymes that may be more suitable for use in industry as well as able to survive frequent exposure to these extreme conditions [7]. This particular strain showed promising antibacterial activity against three-indicator pathogens: *Salmonella typhimurium*, *Staphylococcus aureus*, and *Pseudomonas syringae* [8]. Additionally, one very closely related strain (*B. paralicheniformis* Strain GSFE7- 95% genome sequence similarity) has been reported to be involved in the promotion of halotolerant plant growth [9]. Besides, another closely related strain (*B. paralicheniformis* Strain CCMM B940 which shares 98.94% identity with *B. paralicheniformis* strain Bac84) can break down complex polysaccharides [10].

The genome of *B. paralicheniformis* strain Bac84 has been fully sequenced and published [4]. According to the National Center for Biotechnology Information database—NCBI repository, it encodes 4,237 proteins (CP023665.1). However, 414 coding sequences have been anticipated to encode for proteins without any expression and function-associated data. These sequences have been assigned as "hypothetical". These hypothetical proteins (HPs) have constituted a considerable portion (9.8% of the total number of proteins) of the genome. Functional annotation is necessary for these HPs to find the possible roles in the cell which can lead to an understanding of new structures, and functions in this bacterium. Several studies have revealed the expression of HPs [11–13]. Homology-based gene annotation has been assigned previously to predict the unknown functions of numerous HPs in several organisms [14–18]. Additionally, numerous bioinformatics tools are available to determine the functions of the HPs such as Pfam, InterPro, CATH, SUPERFAMILY, SMART, CDD-BLAST SCANPROSITE, and many more [17–23]. Moreover, the STRING database is also an essential way of protein-protein interaction (PPI) determination to understand the protein functions in a biological network [24–26]. Hence, the PPI study of these HPs can lead to inferences about their biological functions [27]. Furthermore, the tertiary structure modeling through homology searches utilizing the SWISS-MODEL server is important to find the function of unknown proteins [28].

In this study, we aimed to determine the functional roles of the HPs from the *B. paralicheniformis* strain Bac84. We utilized an annotation-based workflow to determine the functions of the HPs for the identification of new biotechnologically important proteins as well as novel proteins contributing to the survival of this bacterium in extreme environments. We successfully identified potential target proteins in the *B. paralicheniformis* strain Bac84. It may eventually be possible to develop new biotechnological applications based on further experimental validation of these identified proteins.

Materials and methods

Sequence retrieval

The genome of *B. paralicheniformis* strain Bac84 was used (CP023665.1). It has 4,376,831 bp in length containing 4413 genes. It encodes 4,237 proteins and 414 are HPs among those (https://www.ncbi.nlm.nih.gov/genome/). The HPs' sequences were obtained in FASTA format for the analyses (S1 Table).

Functional annotation of hypothetical proteins

Functional annotation was applied to the HPs to reveal their functions (Fig 1). Firstly, several publicly available tools and databases (Pfam, InterPro, CATH, SUPERFAMILY, SMART, SCANPROSITE, and CDD-BLAST) are listed in the S2 Table were used. These bioinformatics tools and databases assist to find the conserved domains and afterward categorize the proteins. Pfam [29], InterPro [30], SUPERFAMILY [20], and SCANPROSITE [31] were employed to interpret the functional roles of the HPs based on similarity. Additionally, SMART and CATH were used to search for functions of our HPs based on the domain architecture and to categorize the domains within the structural hierarchy respectively [32, 33]. Conserved Domain Database (CDD) was utilized to search conserved domains [34]. All these analyses were performed in the default parameters and the results are given in detail in the S3 Table. These web tools showed distinctive results and to perform downstream analyses, 37 HPs were filtered as these HPs exhibited functional domains or motifs in at least three of the bioinformatic tools (S4 Table).

We also have predicted the gene ontology of all the HPs using Argot[2.5] (Annotation Retrieval of Genel Ontology Terms) [35] (S5 Table) and the findings are illustrated in Fig 2.

We further used the FASTA sequences of the selected 37 HPs for manual annotation utilizing the Basic Local Alignment Search Tool (BLAST) [36]. Here, the NCBI nonredundant database and hits with an identity ≥ 90% were employed (S6 Table).

In addition, we used BPROM (in the default settings) to perform the promoter analysis of the 37 proteins [37]. All the DNA sequences were downloaded from the NCBI database. The Shine Dalgarno (SD) sequence was manually assigned in this case.

The DEG database was utilized to detect the essential genes with the screened 37 HPs [38]. The search was performed against the available genomes of *Bacillus subtilis 168*, and *Bacillus thuringiensis BMB171* in the default parameters (S7 Table).

Prediction of physicochemical parameters and the Sub-cellular localization

The physicochemical parameters of the selected 37 HPs were theoretically measured using Expasy's Protparam server [39]. The predicted properties such as molecular mass, isoelectric point (pI), extinction coefficient, the total number of +/- residues, extinction coefficient, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) were determined.

Determination of the protein cellular localization helps to estimate its function. In this study, PSORTb [40] and CELLO [41] were used to identify the proteins' location in the cell. PSORTb includes both lab experimental data sets as well as in silico predictions. In contrast, CELLO employs a two-level support vector machine (SVM) based system.

Furthermore, SOSUI [42], HMMTOP [43], TMHMM [44], and SignalP [45] were utilized to predict the transmembrane helices as well as determine the presence of signal peptide cleavage sites. All the results of these characterization analyses were listed in the S8 Table.

All hypothetical proteins retrieved from *Bacillus paralicheniformis strain Bac84* (CP023665.1): NCBI

414 HPs

The gene ontology of all 414 HPs: Argot2.5

Functional prediction: Pfam, InterPro, CATH, SUPERFAMILY, SMART, ScanProsite and CDD

HPs with function predicted by three or more tools

Excluded ← No

Yes → 37 HPs

Homologous search: BlastP

Promoter analysis: BPROM

Search for essential genes: DFG

Characterization: ProtParam, PSORTb, CELLO, SOSUI, TMMHM, HMMTOP, SignalP, STRING and SWISS-MODEL

Inference of the protein function

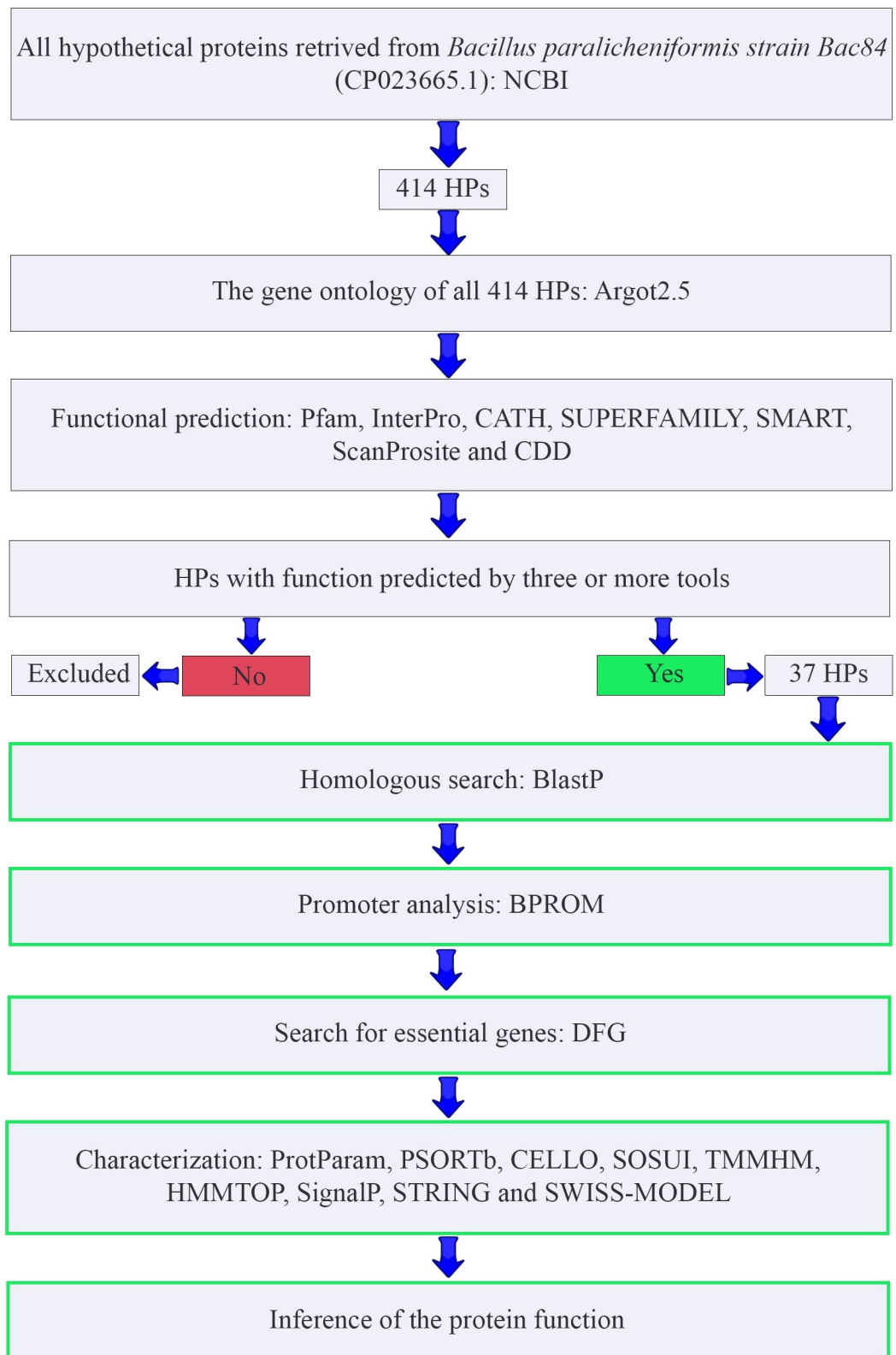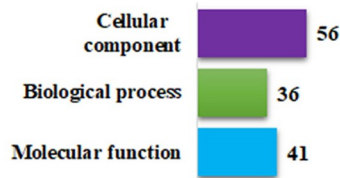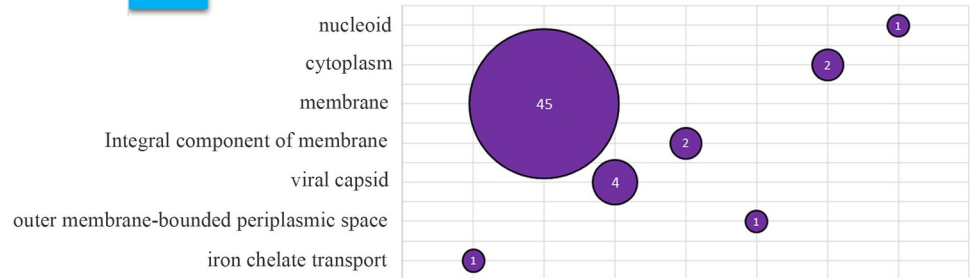**Fig 1. Workflow representing the overall design of the study.** The tasks listed in the green outlined boxes were applied only after the analyzed HPs showed the same function in at least three different bioinformatics tools.
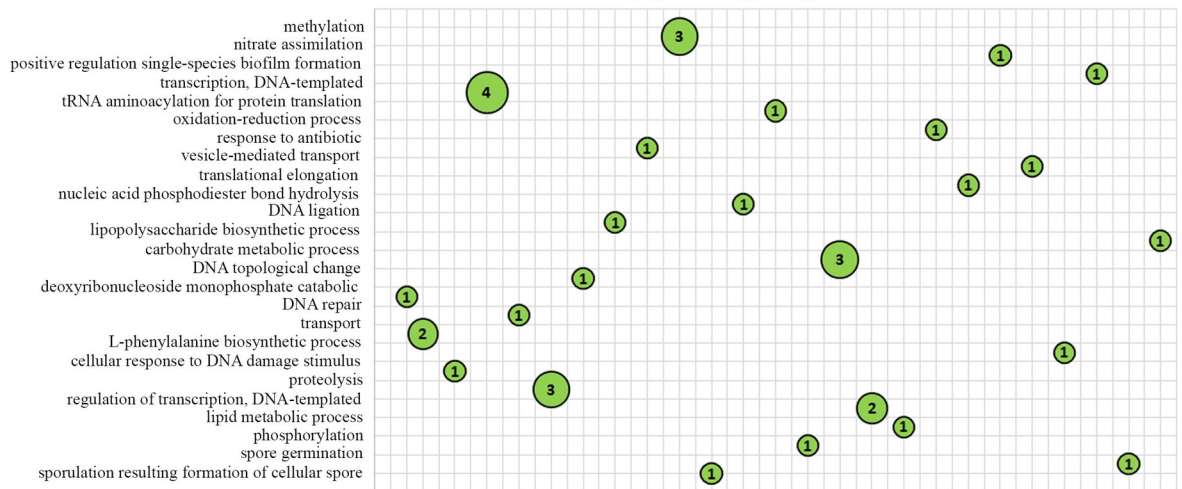
https://doi.org/10.1371/journal.pone.0276085.g001

(A) HPs distribution among the GO categories

(B) Cellular component

(C) Biological process
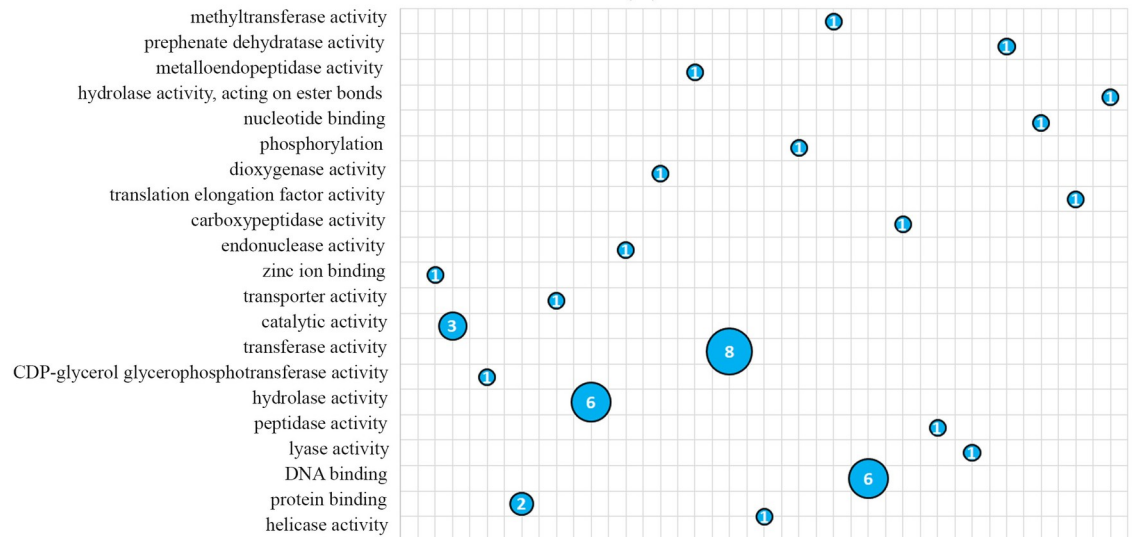

(D) Molecular function

Fig 2. The gene ontology of all the 414 HPs. (A) The distribution of the HPs among the three gene ontology categories. (B) Graph of the cellular components. (C) Graph of the biological processes. (D) Graph of the molecular functions. Here, the distribution of GO terms is presented on the Y axis and the area of the bubbles is relative to the number of proteins found in each category.

https://doi.org/10.1371/journal.pone.0276085.g002

Protein-protein interaction analysis

In this study, STRING software [24, 26] was used to predict interactive partners using a confidence score above 0.7 for ensuring the dependability of the predictions (S9 Table). We had to use the *Bacillus licheniformis* DSM 13 reference genome to generate the interaction networks as the dataset for any strain of *B. paralicheniformis* has not been available yet. Both the physical and functional associations were applied to compute the networks. The Cytoscape was used to visualize the interaction networks (S1 Fig).

Tertiary structure prediction

Tertiary protein structures give significant insights into the molecular basis of protein function [46]. We used the SWISS-MODEL server [28] for homology modeling of the target proteins where only templates with an identity ≥ 30% were considered (S10 Table). The UCSF Chimera-1.16 was used to visualize the 3D structures as well as to perform the structural alignments (Fig 3A & 3B). Additionally, several predicted structures were also compared with the AlphaFold models to validate the structures.
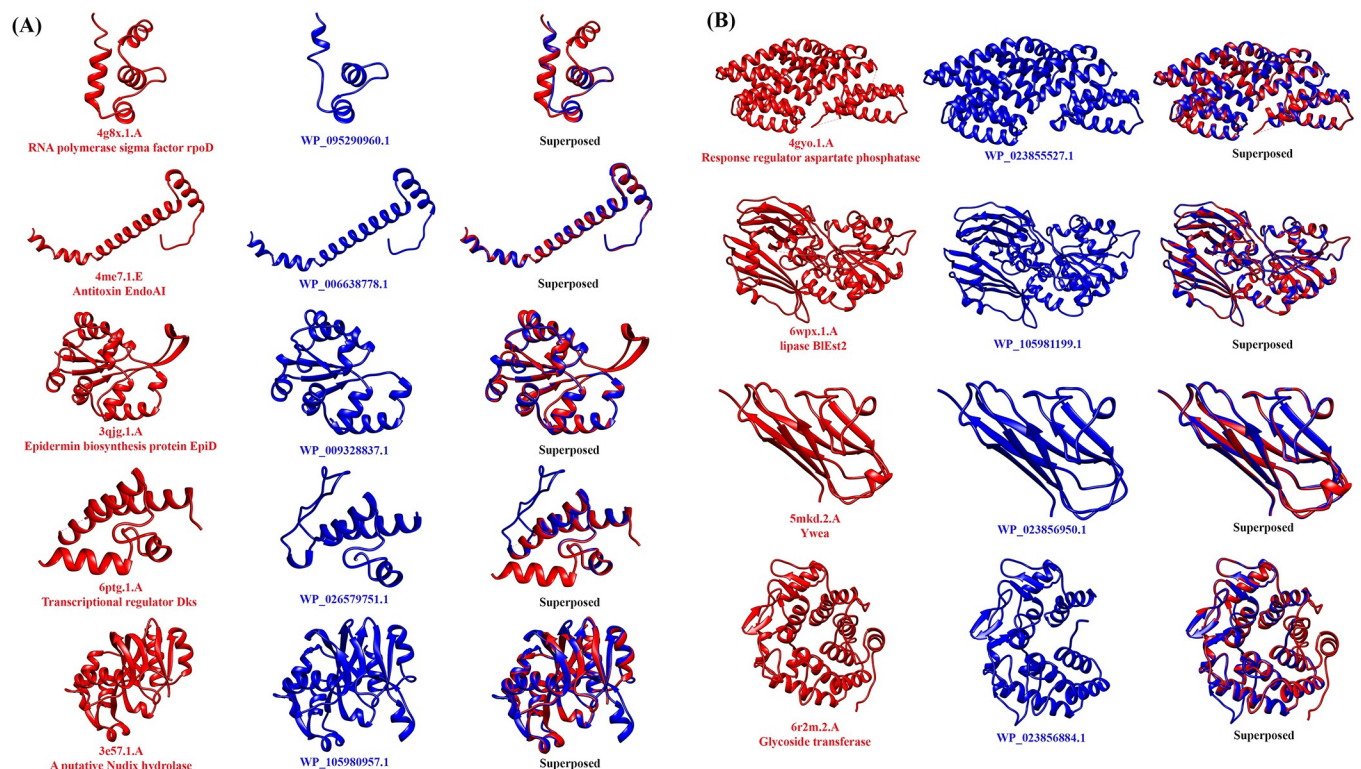
Fig 3. A & B. Tertiary structures analysis. Three-dimensional structures were modeled by the SWISS-MODEL server reliably using the templates with higher coverage, more than 30% of identity, and higher GMQE scores along with Ramachandran Favored percentages ≥90%. Only the templates determined by the X-ray crystallography with high resolution were used. The known proteins and the modeled structures are indicated in red and blue colors respectively. The proteins are orientated using the Chimera MatchMaker according to the optimal superposition of the matching residues.

https://doi.org/10.1371/journal.pone.0276085.g003

Performance assessment

We performed a ROC- receiver operating characteristic analysis with 100 functionally characterized proteins (S11 Table) from the genome of the *B. paralicheniformis* strain Bac84 to check the accuracy of the anticipated functions of our studied HPs [47]. These proteins were functionally checked using the seven databases used for our studied HPs.

For the interpretation, the binary numerals "1" and "0" were applied as the true positive and true negative respectively. The integers '2', '3', '4', and '5' were used to assess the prediction efficacy. After that, these datasets were submitted to the Web-based Calculator and calculated the specificity, sensitivity, accuracy, and the ROC area of each tool employed earlier for functional prediction of the HPs.

Results and discussion

Analysis of The hypothetical proteins from the *B. Paralicheniformis* strain Bac84 genome

DNA sequencing technologies are advancing, and high throughput sequencing technologies have allowed a significant number of bacterial genome sequencing. Sequence homology techniques are commonly used for the annotation of genes [48]. Nevertheless, these homology techniques alone are not always able to predict functions accurately and lead to false annotations [49]. Hence, multiple bioinformatic tools must be employed to assign functional annotations of HPs. In this study, we applied a number of effective tools and databases to do the annotation of HPs from the *B. paralicheniformis* strain Bac84.

We first identified the domains of the HPs which are structural, functional, and evolutionary parts of a protein, therefore providing the functional role of a protein [50]. We extensively analyzed all the 414 HPs sequences using Pfam, InterPro, CATH, SUPERFAMILY, SMART, SCANPROSITE, and CDD-BLAST (S3 Table). The results were evaluated aiming to assign functions to HPs and it revealed 37 HPs which demonstrated similar functions from three or more programs listed in Table 1. In this way, functional annotations were assigned with strong confidence to the HPs. For the rest HPs (n = 377), domains were recognized from less than three mentioned bioinformatic tools which are needed further assessments.

Further, the GO terms were determined using the ARGOT$^{2.5}$ server [35] that provides results based on the confidence scores. 133 HPs have GO term predictions among the 414 targets and the distribution among the GO categories was depicted in Fig 2. The rest of the HPs with no GO terms can be found in the S5 Table. Among the three categories, the largest cluster was cellular components followed by molecular functions and biological processes. We found seven different GO terminologies in the cellular component category including 45 having membrane function (Fig 2B). Although studying membrane proteins is difficult, it is well known that many membrane proteins play important roles in gram-positive bacteria's physiology [51, 52]. The membrane proteins come first in the interaction among cells and the environmental stresses [53]. These membrane HPs need to be analyzed as these may have considerable roles in the survival mechanism of the *B. paralicheniformis* strain Bac84 in extreme environments. For biological processes, twenty-five different GO terminologies were identified, mostly associated with transcription and DNA-related processes (Fig 2C). Transcriptional regulation is a crucial process for a living organism. The cell can respond to intracellular and external signals such as environmental cues or nutritional insufficiency through this transcription-controlling process. According to the GO annotation, the molecular function category showed twenty-one GO terminologies; mostly indicated to several enzymatic functions, and the others related to protein binding (Fig 2D). Here, the DNA and protein

Table 1. Hypothetical proteins functionally annotated from the *B. paralicheniformis* strain Bac84.

No.	HP ID	Inferred function
1	WP_158700706.1	Metal-dependent hydrolase
2	WP_230368348.1	Catalytic core DNA breaking-rejoining enzymes
3	WP_095290960.1	RNA polymerase sporulation sigma factor SigK
4	WP_026579962.1	YhzD-like protein
5	WP_224146215.1	Response regulator aspartate phosphatase
6	WP_095291534.1	The YqzH-like protein family
7	WP_003179940.1	The YgaB-like protein family
8	WP_020449960.1	Inner membrane protein YiaA-like
9	WP_105981192.1	YqaH-like protein
10	WP_020453622.1	Bacteriophage A118-like, holin
11	WP_006638778.1	Metal-responsive transcriptional regulator
12	WP_003180123.1	Sigma-M inhibitor protein YhdK
13	WP_025810847.1	Streptogramin lyase
14	WP_020450411.1	RlpA-like domain superfamily
15	WP_105980832.1	Phenylalanyl-tRNA synthetase
16	WP_009328837.1	Flavin-phosphopantothenoylcysteine decarboxylase/Flavin prenyltransferase
17	WP_003180732.1	Pathogenicity locus—Putative mitomycin resistance proteins
18	WP_199792123.1	YetA-like protein
19	WP_020451108.1	ESAT-6-like superfamily
20	WP_020451191.1	YkyB-like protein
21	WP_026579751.1	Transcription regulator DksA-related
22	WP_105980957.1	Nudix_Hydrolase super family
23	WP_023857538.1	YhzD-like protein
24	WP_020451915.1	Heat Shock protein (Hsp20 proteins)
25	WP_020452052.1	HesB-like domain superfamily
26	WP_026579290.1	YqfQ-like protein
27	WP_020452371.1	RmlC-like cupin superfamily
28	WP_234026546.1	Chromosome segregation protein SMC
29	WP_023855527.1	Response regulator aspartate phosphatase
30	WP_105981186.1	Putative phage metallopeptidase
31	WP_105981199.1	Alpha/Beta hydrolase fold
32	WP_003185659.1	Swarming motility protein SwrA
33	WP_023857076.1	Acyl-CoA N-acyltransferase
34	WP_023856950.1	BslA (Biofilm surface layer A)
35	WP_026580354.1	Immunity protein WapI-like/YxiJ super family
36	WP_023856884.1	Six-hairpin glycosidase superfamily
37	WP_020453535.1	Prephenate dehydratase

https://doi.org/10.1371/journal.pone.0276085.t001

interactions (sequence-specific and sequence non-specific binding) are involved in many biological processes including regulation of transcription, DNA repair, DNA modification, etc. [54]. Additionally, the proteins with enzymatic functions have potential biotechnological applications [55, 56].

Additionally, 15 HPs carried homologous sequences with described functions were found in BlastP analysis whereas the remaining HPs were matched to uncharacterized family proteins and/or hypothetical proteins (S6 Table). All the 15 HPs that matched with functional proteins in the BlastP analysis were functionally similar to the anticipated functions. We also analyzed the promoter regions of all 37 proteins. Promoter segments are required for the start

of transcription at a certain genomic site. Several conserved regions such as the Pribnow box and -35 box were determined along with the SD sequence (S2 Fig). These conserved sequences are vital for the binding of RNA polymerase and ribosome [57, 58]. The SD-sequence initiates the translation process and has a huge influence on protein expression levels [59, 60]. It was found that all 37 proteins have SD sequences. The findings from the promoter analysis of the 37 proteins indicate that further experimental validation is worth pursuing. We did not find any study regarding the experimental transcription data sets of the organism.

Furthermore, the DEG database was utilized to predict fundamental genes (S7 Table). This database adapts both in vitro and in vivo experiments to detect fundamental genes which are essential for cellular machinery [38]. Though different challenging lab experiments were used to detect the essential genes such as RNA interference, gene knockouts, and transposon mutagenesis [61], this DEG database offers an alternative for predicting essential genes. In our analysis, we did not find any essential genes among the targeted 37 HPs.

Physicochemical characterization and subcellular localization

To evaluate the physicochemical characteristics and their cellular distribution the sequences of the screened 37 HPs were used (S8 Table). Most of the studied proteins had molecular weight (MW) values over 10000 Da. Proteins with a lower MW ($<$ 10000 Da) need special modifications for analysis in the SDS-PAGE system [62]. Hence, the first few HPs with lower MW require special attention to perform further lab experiments. The pH value of a protein at which it carries no net electrical charge is known as isoelectric point pI. For our selected HPs, it ranged from 4.4 to 10.48 and 11 proteins have acidic nature (pI $<$ 7), whereas others were found to be basic. Along with the MW, the pI also helps in the laboratory analysis of proteins [63].

The aliphatic index (AI) is used to evaluate the protein thermostability and our HPs were in the range of 55.19–145.1. The range of temperatures at which a protein will be stable increases with increasing AI values [64]. Protein WP_003180123.1, associated with growth and survival after salt stress showed the highest value of 145.1. The instability index (II) was applied to get the idea regarding in vitro protein stability. 15 HPs were considered to be unstable, and 22 HPs were stable. The cut-off values $>$40 and $<$40 were used to categorize stable and unstable proteins, respectively [65]. The GRAVY indicates the interactive nature of a protein with water [66]. Among these 37 HPs, only four (WP_158700706.1; WP_003180123.1; WP_023857538.1 and WP_020453535.1) showed positive values which indicates that these might be hydrophobic.

Moreover, the cellular localization of proteins is vital for their biological functions in a specific environment [67–69]. Among the 37 HPs, most of the proteins were determined as cytoplasmic. Several cytoplasmic proteins are in the regulation of several functional processes including biosynthesis, regulatory activities, and transport which may help environmental bacteria to compete with the neighboring organisms in the same ecological niche [70]. Additionally, we only found 4 proteins to have signal peptides that are critically related to protein secretion [71].

Protein-protein interactions

To determine the interaction partners of the HPs, we performed a protein-protein interaction analysis [72]. In this study, protein WP_095290960.1, RNA polymerase sporulation sigma factor SigK showed a very strong interaction (score 0.930) with the sporulation stage IV protein A (SpoIVA) which is involved in sporulation [73]. WP_006638778.1 interacted with EndoA–a putative RNase (score 0.988) with functional endoribonuclease activity [74]. WP_009328837.1

was found to interact with the YacB (score 0.987) which catalyzes the phosphorylation of pantothenate [75]. The protein WP_023855527.1 showed interaction with the Raca protein which is required for the formation of axial filaments [76]. All these findings along with the other predictions (S9 Table and S2 Fig) strengthened our functional predictions.

Tertiary structure predictions

X-ray crystallography has become a robust approach to determining novel protein structures [77]. The functional annotation methods in combination with the protein structure analysis are evident to lead to the interpretation of uncharacterized proteins [78, 79]. In this study, we employed the protein structure homology-modeling server SWISS-MODEL to have the tertiary structures and used the UCSF Chimera software to visualize the models. Next, we compared the structures of known proteins with the modeled structures to check the degree of similarity (Fig 3A & 3B).

We successfully build the three-dimensional models for 9 HPs with identity above 30% and the details were listed in the S10 Table. We also checked the quality of the models with the Ramachandran plots and scores (S10 Table and S3 Fig). Structural comparisons were performed based on the Needleman-Wunsch algorithm [80]. We observed different percentages of structural similarities between the models and known proteins (S10 Table). The alignment results from the structural comparisons were shown in S4 Fig. The structural data collected for several HPs has validated the precise functional annotation. For instance, WP_105981199.1 and WP_023856950.1 showed high identities and resolutions which were functionally annotated as Alpha/Beta hydrolase and BslA (Biofilm surface layer A) respectively. The structures built for these two proteins were determined by X-ray crystallography from two *Bacillus sp.* and those two template proteins have similar functions as we predicted in this study. In this way, proteins with similar sequences usually exhibit similar functions. Proteins dissimilar to current PDB entries may correspond to novel functions. In addition, several final protein models were visualized using the Chimera 1.16 and compared to the predicted models suggested by AlphaFold (S5 Fig). We used the AlphaFold since Alpha-Fold has been demonstrated to be more accurate than Nuclear magnetic resonance spectroscopy (NMR) [81]. The findings showed similarities among the predicted models by Swiss-Model vs AlphaFold.

ROC performance measurement

The availability of genome sequences is increasing which is also allowing more scope to do the computational protein analysis. As these analysis methods are solely dependent on autonomic computing, the accuracy of these methods should be high. The ROC analysis is a broadly applied technique for evaluating the tool's accuracy. The employed pipeline had an average accuracy of 98 percent (Table 2), and the ROC analysis's findings supported the strong dependability of the tools used.

Table 2. ROC results of the tools used in this study.

Software	Accuracy (%)	Sensitivity (%)	Specificity (%)	ROC area
Pfam	99.0	98.0	100	0.99
InterPro	100.0	100.0	100.0	1
CATH	100.0	100.0	100.0	1
SUPERFAMILY	96.0	94.7	100.0	0.99
SCANPROSITE	97.0	93.8	100.0	0.99
SMART	98.0	97.0	100.0	1
CDD-BLAST	96.0	65.9	100.0	0.985

https://doi.org/10.1371/journal.pone.0276085.t002

Proteins with biotechnological potentials

We found several proteins that can be used for biotechnological applications.

WP_158700706.1 was predicted as a Metallo-dependent hydrolase (the amidohydrolase superfamily). This group includes numerous hydrolytic enzymes with a varied spectrum of substrates and reactions. The microbial obtained amidohydrolase possesses extensive biotechnological applications that include cosmetics, food, and therapeutics, especially as an anticancer/anti-proliferative agent [82, 83]. This hydrolase group also contains amylases and α-amylase derived from *B. licheniformis*, *B. amyloliquefaciens* and *B. stearothermophilus* which has been commercially used in fermentation, paper, and textiles industries [84, 85].

Protein WP_020453622.1 is a Bacteriophage A118-like, holin that involves the lysis of bacterial membrane [86]. These holins can be utilized for controlled pore formation and can promote the release of the desired products. Microorganisms are used and improved for the industrial manufacture of a wide range of substances, including pharmaceuticals and biofuels. These target compounds can be sequestered inside the cell causing toxic effects to the chassis without an efficient active efflux system. In this case, Holin-mediated cell lysis offers an efficient releasing mechanism [87]. One of the rate-limiting steps is releasing products from the microbial host for biotechnology-based chemical production on an industrial scale. Holins can provide an affordable and effective method of product release in many instances where the use of mechanical disruption or solvent extraction increases the cost of production [88]. Liu and Curtiss applied phage holin/endolysin cassettes containing a nickel-inducible signal transduction system into the chromosome of *Synechocystis sp*. strain PCC6803 which is being developed for biofuel production [89]. They successfully eliminated the chemical or mechanical removal step by just adding nickel to the culture medium resulting in cell lysis. Another group utilized a light-inducible lytic mechanism in the same cyanobacterium for similar purposes [90].

The protein WP_009328837.1 was predicted as Flavin-containing phosphopantothenoyl-cysteine decarboxylase which is involved in coenzyme A (CoA) biosynthesis [91]. CoA is a crucial cofactor involved in many metabolic processes including secondary metabolites production. These distinctive features make CoA an economically significant chemical compound in the cosmetic, and therapeutic industries [92]. Hence, the catalytic abilities of this enzyme make it of immense biotechnological significance.

The protein WP_020452371.1 is in the RmlC-like cupin superfamily and RmlC is a dTDP-sugar isomerase enzyme (dTDP—deoxythymidine diphosphates). This enzyme is involved in the L-rhamnose synthesis, commonly found in bacteria and plants [93, 94]. This sugar getting more interest due to its wide range of substrate specificity and its excellent potential for various unique sugars syntheses such as D-allose, D-cellulose, L-mannose, L rhamnulose, L-spotose, and L-talose [95]. Besides, rhamnose is combined with lipids to form rhamnolipids that can be used as potential biosurfactants [94].

The protein WP_105981199.1 contains an α/β-hydrolase fold that includes proteases, lipases, peroxidases, esterase, epoxide hydrolases, dehalogenases, and many others [96]. Therefore, this protein can be studied further to uncover its actual functionality as several hydrolases are being used in industrial processes [56]. Additionally, an α/β-hydrolase fold protein was also studied which is involved in the cyclic oligopeptide antibiotic 'thiostrepton' biosynthesis [97].

The protein WP_023857076.1 carries a structural domain found in numerous acyl-CoA acyltransferases including the N-acetyl transferase (NAT) [98]. Several NATs from *Bacillus sp*. Have shown the capability to metabolize xenobiotic compounds that are highly toxic contaminants of groundwater and soils [99]. This study showed that a class of industrial contaminants

or by-products of agrochemicals named "Arylamines" can be converted into less toxic states by *Bacillus* NATs. Hence, our WP_023857076.1 protein should be studied further to find out its bioremediation potential. Additionally, a synthetic N-acetyltransferase (MAT—methionine sulfone N-acetyltransferase) from a bacterial source was utilized to successfully design herbicide "Phosphinothricin" -resistant rice and Arabidopsis [100].

Different glycosyltransferases transfer sugar parts from donor molecules to acceptors to form glycosidic bonds and involve in disaccharides, oligosaccharides, and polysaccharides biosynthesis. Several microbial glycosyltransferases are frequently applied in food processes such as in the shelf-life improvement of bakeries, production of glucose, fructose, or dextrins, lactose hydrolysis, food pectins modification, and many others [101, 102]. In our study, protein WP_023856884.1 has the catalytic domain of the Six-hairpin glycosidase superfamily. To use this class of enzymes in different industrial conditions several enzymes functional in alkaline/ acidic pH and/or at high temperatures have been discovered from various microorganisms [103–105]. In several studies, bacterial glycosidases were characterized to improve human health and the treatment of different diseases [106, 107].

The WP_020453535.1 was anticipated to be a prephenate dehydratase that is involved in the biosynthesis of phenylalanine and phenylalanine is an essential amino acid for animals. Recently, the interest in microbial production of L- phenylalanine has increased [108]. It has been widely used in food and feeds as a taste and aroma enhancer, in pharmaceuticals as the drug's building block, as well as used in cosmetics as an ingredient [109, 110].

Proteins with adaptational functions to extreme environments

In this study, we identified 12 HPs that may have a significant role for *B. paralicheniformis* in the adaptation to extreme environments.

Sporulation aids bacterial survival in extreme environments by limiting active growth [111]. We found protein WP_095290960.1 as RNA polymerase sporulation sigma factor SigK which is involved in the gene expression controlling during sporulation [112]. Two HPs (WP_224146215.1 and WP_023855527.1) were identified to be the aspartate phosphatase, which regulates the phosphorelay for sporulation initiation by dephosphorylating Spo0F-P [113]. In this way, these HPs can be predicted to play crucial roles in adaption, and survival in extreme environments.

The protein WP_006638778.1 is a metal-responsive transcriptional regulator which can be engaged in the homeostasis and metabolism of any specific metal. These metal-responsive transcriptional regulators allow mechanisms for selective metal ion accumulation and utilization as well as tightly regulate intracellular metal trafficking mechanisms [114]. Metals can be limited in the environment or can be in high amounts that cause toxicity in extreme environments. Hence, a metal-responsive transcriptional regulator protein might be essential to the microorganism for the evolution and adaptation in that specific extreme environment [115]. Likewise, WP_026579751.1 is related to the transcription regulator DksA. It is an RNA polymerase-binding transcription factor and is involved in different stress conditions, including nitrosative stress, nutritional shortage, and other environmental stresses [116, 117]. So, this HP can be taken part in extreme environmental adaptations.

We detected a sigma-M inhibitor protein (WP_003180123.1). The sigma-M (YhdM) gene is essential for growth and survival in salt stress conditions [118]. Our predicted Sigma-M inhibitor WP_003180123.1 might play role in salt stress adaptation similarly to a previous study [119].

Protein WP_105980957.1 contains a Nudix hydrolase domain that hydrolyzes intracellular nucleotides, regulates their levels, and removes potentially toxic derivatives [120]. Some

superfamily members can degrade mutagenic, oxidized, and damaged nucleotides that may occur due to exposure to extreme environments [121].

As mentioned earlier, WP_023857076.1 carries a structural domain found in numerous acyl-CoA acyltransferases including- GCN5-related N-acetyltransferases (GNAT) and Glycine N-acyltransferase [122]. The proteins from these classes were studied and found to be involved in the adaptation to diverse environmental stress conditions including high salinity, pH tolerance, nutrient stress, etc. [123, 124].

Small Heat shock proteins are abundant molecular chaperones that counteract the aggregation of protein upon stress-induced unfolding [125]. We identified protein WP_020451915.1 as a heat shock protein (Hsp20). Several studies showed that Hsp20 responds to different environmental stresses including severe heat, hydrogen peroxide, desiccation, and osmotic shocks [126–129]. Therefore, WP_020451915.1 might have adaptational functions to extreme environments.

The HesB-like domain is observed in several microbial nitrogen fixation proteins that are associated with FeS-cluster assembly [130]. Previous studies found that proteins having a HesB-like domain are involved in different metal resistance and thermal stress conditions [131, 132]. HesB-like domain-containing protein WP_020452052.1 might also play role in survival in the extreme environment specifically in metal-rich or metal deficient conditions.

The WP_003185659.1 protein was identified as a swarming motility protein SwrA which is a transcription factor. It drives the *fla/che* operon, which encodes the components of the flagella, and causes swarming motility [133]. Another study showed that SwrA is involved in bacterial motility [134] and bacterial motility might be significant in extreme temperatures [135].

The WP_023856950.1 protein was predicted as a biofilm surface layer A (BslA) protein which acts as a hydrophobin and participates in biofilm assembly [136]. Certain microorganisms have great resistance to environmental challenges because of biofilm development [137–139]. Therefore, this protein might be crucial for adaptation to harsh environments.

Conclusions

Protein macromolecules are involved in numerous biological processes. Hence, functional annotation of proteins is crucial. An in silico approach was employed in this study to attribute functional annotation of HPs from the *B. paralicheniformis* strain Bac84 genome. We predicted the functions of 37 HPs from this bacterium. The determination of physicochemical parameters and subcellular localization were effective to understand the specific properties of the annotated proteins. The PPI and tertiary structures of these proteins were also explored which assisted to obtain more understanding of the annotated proteins. Several protein structures were also validated by the AlphaFold protein modeling. We identified several proteins with biotechnological potentials as well as proteins having the possibility to be involved in extreme environmental adaptation of the *B. paralicheniformis* strain Bac84. Moreover, the findings of this strategy suggested that it can be utilized to perform the predictive annotations of unknown proteins. The combination of such in-silico analysis with the proper lab experiments was successful to obtain functional annotations of HPs from different organisms [140–142]. Furthermore, the results also open prospects for further research of this bacterium for biotechnological applications.

Supporting information

S1 Fig. Protein-protein interaction networks obtained from STRING analysis. Networks are visualized using Cytoscape.
(PDF)

S2 Fig. Promoter analysis of the 37 proteins using BPROM.
(PDF)

S3 Fig. Ramachandran plots for the 3D models of the 9 proteins by the SWISS-MODEL serve.
(PDF)

S4 Fig. Alignment results from the superposition analysis.
(PDF)

S5 Fig. Comparison of the structures predicted by AlphaFold and Swiss-Model.
(PDF)

S1 Table. All the hypothetical proteins from the *B. paralicheniformis* strain Bac84.
(XLSX)

S2 Table. List of bioinformatics tools and databases used.
(XLSX)

S3 Table. Annotation dataset results for the 414 hypothetical proteins submitted to the workflow with Pfam, InterPro, CATH, SUPERFAMILY, SCANPROSITE, SMART, and CDD-Blast.
(XLSX)

S4 Table. List of selected HPs from the *B. paralicheniformis* strain Bac84.
(XLSX)

S5 Table. GO terms by Argot2.5 for all the HPs.
(XLSX)

S6 Table. Results of the BlastP search for similar sequences against the non-redundant (nr) database.
(XLSX)

S7 Table. Result of essential gene prediction using DEG database.
(XLSX)

S8 Table. List of predicted physicochemical parameters, sub-cellular localization, and prediction of transmembrane helices for the selected 37 HPs.
(XLSX)

S9 Table. Protein-protein interactions analyses of the 37 HPs.
(XLSX)

S10 Table. Tertiary structural information of HPs from *B. Paralicheniformis* strain Bac84.
(XLSX)

S11 Table. Dataset of functional annotation for 100 functionally known proteins from *B. paralicheniformis* strain Bac84 using the same pipeline used for the HP prediction.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Md. Atikur Rahman, Uzma Habiba Heme, Md. Anowar Khasru Parvez.

Data curation: Md. Atikur Rahman, Uzma Habiba Heme.

Formal analysis: Md. Atikur Rahman, Uzma Habiba Heme, Md. Anowar Khasru Parvez.

Investigation: Md. Atikur Rahman, Uzma Habiba Heme.

Methodology: Md. Atikur Rahman, Uzma Habiba Heme, Md. Anowar Khasru Parvez.

Project administration: Md. Atikur Rahman, Md. Anowar Khasru Parvez.

Resources: Md. Anowar Khasru Parvez.

Software: Md. Atikur Rahman, Md. Anowar Khasru Parvez.

Supervision: Md. Anowar Khasru Parvez.

Validation: Md. Atikur Rahman, Md. Anowar Khasru Parvez.

Visualization: Md. Atikur Rahman, Uzma Habiba Heme, Md. Anowar Khasru Parvez.

Writing – original draft: Md. Atikur Rahman.

Writing – review & editing: Md. Atikur Rahman, Uzma Habiba Heme, Md. Anowar Khasru Parvez.

References

1. Dunlap CA, Kwon S-W, Rooney AP, Kim S-J. Bacillus paralicheniformis sp. nov., isolated from fermented soybean paste. International journal of systematic and evolutionary microbiology. 2015; 65 (Pt_10):3487–92. https://doi.org/10.1099/ijsem.0.000441 PMID: 26296568

2. Du Y, Ma J, Yin Z, Liu K, Yao G, Xu W, et al. Comparative genomic analysis of Bacillus paralicheniformis MDJK30 with its closely related species reveals an evolutionary relationship between B. paralicheniformis and B. licheniformis. Bmc Genomics. 2019; 20(1):1–16.

3. Rey MW, Ramaiya P, Nelson BA, Brody-Karpin SD, Zaretsky EJ, Tang M, et al. Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacilluspecies. Genome biology. 2004; 5(10):1–12.

4. Othoum G, Bougouffa S, Razali R, Bokhari A, Alamoudi S, Antunes A, et al. In silico exploration of Red Sea Bacillus genomes for natural product biosynthetic gene clusters. BMC genomics. 2018; 19 (1):1–11.

5. Dhakal R, Chauhan K, Seale RB, Deeth HC, Pillidge CJ, Powell IB, et al. Genotyping of dairy Bacillus licheniformis isolates by high resolution melt analysis of multiple variable number tandem repeat loci. Food microbiology. 2013; 34(2):344–51. https://doi.org/10.1016/j.fm.2013.01.006 PMID: 23541201

6. Wang Y, Liu H, Liu K, Wang C, Ma H, Li Y, et al. Complete genome sequence of Bacillus paralicheniformis MDJK30, a plant growth-promoting rhizobacterium with antifungal activity. Genome Announcements. 2017; 5(25):e00577–17. https://doi.org/10.1128/genomeA.00577-17 PMID: 28642380

7. Nielsen J, Archer J, Essack M, Bajic VB, Gojobori T, Mijakovic I. Building a bio-based industry in the Middle East through harnessing the potential of the Red Sea biodiversity. Applied Microbiology and Biotechnology. 2017; 101(12):4837–51. https://doi.org/10.1007/s00253-017-8310-9 PMID: 28528426

8. Al-Amoudi S, Essack M, Simões MF, Bougouffa S, Soloviev I, Archer JA, et al. Bioprospecting Red Sea coastal ecosystems for culturable microorganisms and their antimicrobial potential. Marine drugs. 2016; 14(9):165. https://doi.org/10.3390/md14090165 PMID: 27626430

9. Albdaiwi R, Alhindi T, Hasan S. Draft Genome Sequence of Bacillus paralicheniformis Strain GSFE7, a Halotolerant Plant Growth-Promoting Bacterial Endophyte Isolated from Cultivated Saline Areas of the Dead Sea Region. Microbiology Resource Announcements. 2022:e00425–22. https://doi.org/10.1128/mra.00425-22 PMID: 35950866

10. Maski S, Ngom SI, Rached B, Chouati T, Benabdelkhalek M, El Fahime E, et al. Hemicellulosic biomass conversion by Moroccan hot spring Bacillus paralicheniformis CCMM B940 evidenced by glycoside hydrolase activities and whole genome sequencing. 3 Biotech. 2021; 11(8):1–13. https://doi.org/10.1007/s13205-021-02919-0 PMID: 34447652

11. Ijaq J, Bethi N, Jagannadham M. Mass spectrometry-based identification and characterization of human hypothetical proteins highlighting the inconsistency across the protein databases. Journal of Proteins and Proteomics. 2020; 11(1):17–25.

12. Jagannadham M, Abou-Eladab EF, Kulkarni HM. Identification of outer membrane proteins from an Antarctic bacterium Pseudomonas syringae Lz4W. Molecular & Cellular Proteomics. 2011; 10(6). https://doi.org/10.1074/mcp.M110.004549 PMID: 21447709

13. Jagannadham MV, Chowdhury C. Differential expression of membrane proteins helps Antarctic Pseudomonas syringae to acclimatize upon temperature variations. Journal of proteomics. 2012; 75 (8):2488–99. https://doi.org/10.1016/j.jprot.2012.02.033 PMID: 22418587

14. Doerks T, Von Mering C, Bork P. Functional clues for hypothetical proteins based on genomic context analysis in prokaryotes. Nucleic acids research. 2004; 32(21):6321–6. https://doi.org/10.1093/nar/gkh973 PMID: 15576358

15. Hawkins T, Kihara D. Function prediction of uncharacterized proteins. Journal of bioinformatics and computational biology. 2007; 5(01):1–30. https://doi.org/10.1142/s0219720007002503 PMID: 17477489

16. Vickers NJ. Animal communication: when i'm calling you, will you answer too? Current biology. 2017; 27(14):R713–R5. https://doi.org/10.1016/j.cub.2017.05.064 PMID: 28743020

17. Shahbaaz M, ImtaiyazHassan M, Ahmad F. Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. PloS one. 2013; 8(12):e84263. https://doi.org/10.1371/journal.pone.0084263 PMID: 24391926

18. Shahbaaz M, Hassan, Ahmad F. Correction: Functional Annotation of Conserved Hypothetical Proteins from Haemophilus influenzae Rd KW20. PLOS ONE. 2014; 9(1):10.1371/annotation/23d005b8-fe53-4b14-a31c-915be3e839b5. https://doi.org/10.1371/annotation/23d005b8-fe53-4b14-a31c-915be3e839b5

19. Ijaq J, Chandrasekharan M, Poddar R, Bethi N, Sundararajan VS. Annotation and curation of uncharacterized proteins-challenges. Frontiers in genetics. 2015; 6:119. https://doi.org/10.3389/fgene.2015.00119 PMID: 25873935

20. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of molecular biology. 2001; 313(4):903–19. https://doi.org/10.1006/jmbi.2001.5080 PMID: 11697912

21. Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome research. 2002; 12(10):1619–23. https://doi.org/10.1101/gr.278202 PMID: 12368255

22. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic acids research. 2012; 40(D1):D290–D301. https://doi.org/10.1093/nar/gkr1065 PMID: 22127870

23. Liu Z, Karmarkar V. Groucho/Tup1 family co-repressors in plant development. Trends in plant science. 2008; 13(3):137–44. https://doi.org/10.1016/j.tplants.2007.12.005 PMID: 18314376

24. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research. 2021; 49(D1):D605–D12. https://doi.org/10.1093/nar/gkaa1074 PMID: 33237311

25. Jeong H, Qian X, Yoon B-J, editors. Effective comparative analysis of protein-protein interaction networks by measuring the steady-state network flow using a Markov model. BMC bioinformatics; 2016: BioMed Central.

26. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. Correction to 'The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets'. Nucleic Acids Research. 2021; 49(18):10800-. https://doi.org/10.1093/nar/gkab835 PMID: 34530444

27. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I. Fundamentals of protein interaction network mapping. Molecular systems biology. 2015; 11(12):848. https://doi.org/10.15252/msb.20156351 PMID: 26681426

28. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research. 2018; 46(W1): W296–W303. https://doi.org/10.1093/nar/gky427 PMID: 29788355

29. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, et al. Pfam: The protein families database in 2021. Nucleic acids research. 2021; 49(D1):D412–D9. https://doi.org/10.1093/nar/gkaa913 PMID: 33125078

30. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic acids research. 2021; 49(D1):D344–D54. https://doi.org/10.1093/nar/gkaa977 PMID: 33156333

31. De Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural

residues in proteins. Nucleic acids research. 2006; 34(suppl_2):W362–W5. https://doi.org/10.1093/nar/gkl124 PMID: 16845026

32. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic acids research. 2021; 49(D1):D458–D60. https://doi.org/10.1093/nar/gkaa937 PMID: 33104802

33. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, et al. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic acids research. 2015; 43(D1):D376–D81. https://doi.org/10.1093/nar/gku947 PMID: 25348408

34. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic acids research. 2020; 48(D1):D265–D8. https://doi.org/10.1093/nar/gkz991 PMID: 31777944

35. Lavezzo E, Falda M, Fontana P, Bianco L, Toppo S. Enhancing protein function prediction with taxonomic constraints–The Argot2. 5 web server. Methods. 2016; 93:15–23. https://doi.org/10.1016/j.ymeth.2015.08.021 PMID: 26318087

36. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic acids research. 2008; 36(suppl_2):W5–W9. https://doi.org/10.1093/nar/gkn201 PMID: 18440982

37. Salamov VSA, Solovyevand A. Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture, biomedicine and environmental studies. 2011:61–78.

38. Luo H, Lin Y, Liu T, Lai F-L, Zhang C-T, Gao F, et al. DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. Nucleic acids research. 2021; 49(D1):D677–D86. https://doi.org/10.1093/nar/gkaa917 PMID: 33095861

39. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. 2005:571–607.

40. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010; 26(13):1608–15. https://doi.org/10.1093/bioinformatics/btq249 PMID: 20472543

41. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein science. 2004; 13(5):1402–6. https://doi.org/10.1110/ps.03479604 PMID: 15096640

42. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics (Oxford, England). 1998; 14(4):378–9. https://doi.org/10.1093/bioinformatics/14.4.378 PMID: 9632836

43. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001; 17(9):849–50. https://doi.org/10.1093/bioinformatics/17.9.849 PMID: 11590105

44. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001; 305 (3):567–80. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

45. Nielsen H, Tsirigos KD, Brunak S, von Heijne G. A brief history of protein sorting prediction. The protein journal. 2019; 38(3):200–16. https://doi.org/10.1007/s10930-019-09838-3 PMID: 31119599

46. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic acids research. 2003; 31(13):3381–5. https://doi.org/10.1093/nar/gkg520 PMID: 12824332

47. Swets JA, Dawes RM, Monahan J. Better decisions through science. Scientific American. 2000; 283 (4):82–7. https://doi.org/10.1038/scientificamerican1000-82 PMID: 11011389

48. Stormo GD. An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics. 2009; 27(1):3.1.–3.1. 7. https://doi.org/10.1002/0471250953.bi0301s27 PMID: 19728288

49. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS computational biology. 2009; 5(12):e1000605. https://doi.org/10.1371/journal.pcbi.1000605 PMID: 20011109

50. Rao VS, Srinivas K, Sujini G, Kumar G. Protein-protein interaction detection: methods and analysis. International journal of proteomics. 2014;2014. https://doi.org/10.1155/2014/147648 PMID: 24693427

51. Lee B-Y, Hefta S, Brennan P. Characterization of the major membrane protein of virulent Mycobacterium tuberculosis. Infection and immunity. 1992; 60(5):2066–74. https://doi.org/10.1128/iai.60.5.2066-2074.1992 PMID: 1563797

52. Desvaux M, Dumas E, Chafsey I, Hebraud M. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. FEMS microbiology letters. 2006; 256(1):1–15. https://doi.org/10.1111/j.1574-6968.2006.00122.x PMID: 16487313

53. Walian PJ, Allen S, Shatsky M, Zeng L, Szakal ED, Liu H, et al. High-throughput isolation and characterization of untagged membrane protein complexes: outer membrane complexes of Desulfovibrio vulgaris. Journal of proteome research. 2012; 11(12):5720–35. https://doi.org/10.1021/pr300548d PMID: 23098413

54. Karthik L, Kumar G, Keswani T, Bhattacharyya A, Chandar SS, Bhaskara Rao K. Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound. PloS one. 2014; 9(3): e90972. https://doi.org/10.1371/journal.pone.0090972 PMID: 24618707

55. Cabrera M, Blamey JM. Biotechnological applications of archaeal enzymes from extreme environments. Biological research. 2018; 51. https://doi.org/10.1186/s40659-018-0186-3 PMID: 30290805

56. Gurung N, Ray S, Bose S, Rai V. A broader view: microbial enzymes and their relevance in industries, medicine, and beyond. BioMed research international. 2013;2013.

57. Tripathi G. Cellular and Biochemical Science: IK International Pvt Ltd; 2010.

58. Pribnow D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proceedings of the National Academy of Sciences. 1975; 72(3):784–8. https://doi.org/10.1073/pnas.72.3.784 PMID: 1093168

59. Shine J, Dalgarno L. The 3′-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proceedings of the National Academy of Sciences. 1974; 71(4):1342–6. https://doi.org/10.1073/pnas.71.4.1342 PMID: 4598299

60. Ma J, Campbell A, Karlin S. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. Journal of bacteriology. 2002; 184(20):5733–45. https://doi.org/10.1128/JB.184.20.5733-5745.2002 PMID: 12270832

61. Wei W, Ning L-W, Ye Y-N, Guo F-B. Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. PloS one. 2013; 8(8):e72343. https://doi.org/10.1371/journal.pone.0072343 PMID: 23977285

62. Hashimoto F, Horigome T, Kanbayashi M, Yoshida K, Sugano H. An improved method for separation of low-molecular-weight polypeptides by electrophoresis in sodium dodecyl sulfate-polyacrylamide gel. Analytical Biochemistry. 1983; 129(1):192–9. https://doi.org/10.1016/0003-2697(83)90068-4 PMID: 6190419

63. da Costa WLO, Araújo CLdA, Dias LM, Pereira LCdS, Alves JTC, Araujo FA, et al. Functional annotation of hypothetical proteins from the Exiguobacterium antarcticum strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PloS one. 2018; 13(6): e0198965. https://doi.org/10.1371/journal.pone.0198965 PMID: 29940001

64. Ikai A. Thermostability and aliphatic index of globular proteins. The Journal of Biochemistry. 1980; 88 (6):1895–8. PMID: 7462208

65. Guruprasad K, Reddy BB, Pandit MW. 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 Engineering, Design and Selection. 1990; 4(2):155–61. https://doi.org/10.1093/protein/4.2.155 PMID: 2075190

66. Jaspard E, Macherel D, Hunault G. Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. PloS one. 2012; 7(5):e36968. https://doi.org/10.1371/journal.pone.0036968 PMID: 22615859

67. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics. 2006; 64(3):643–51. https://doi.org/10.1002/prot.21018 PMID: 16752418

68. Naqvi AAT, Shahbaaz M, Ahmad F, Hassan MI. Identification of functional candidates amongst hypothetical proteins of Treponema pallidum ssp. pallidum. PloS one. 2015; 10(4):e0124177. https://doi.org/10.1371/journal.pone.0124177 PMID: 25894582

69. Naqvi AAT, Shahbaaz M, Ahmad F, Hassan MI. Correction: Identification of Functional Candidates amongst Hypothetical Proteins of Treponema pallidum ssp. pallidum. PLOS ONE. 2018; 13(5): e0197452. https://doi.org/10.1371/journal.pone.0197452 PMID: 29758067

70. Nakashima H, Nishikawa K. Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. Journal of molecular biology. 1994; 238(1):54–61. https://doi.org/10.1006/jmbi.1994.1267 PMID: 8145256

71. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. A comprehensive review of signal peptides: Structure, roles, and applications. European journal of cell biology. 2018; 97(6):422–41. https://doi.org/10.1016/j.ejcb.2018.06.003 PMID: 29958716

72. Gazi M, Mahmud S, Fahim SM, Islam M, Das S, Mahfuz M, et al. Questing functions and structures of hypothetical proteins from Campylobacter jejuni: a computer-aided approach. Bioscience reports. 2020; 40(6). https://doi.org/10.1042/BSR20193939 PMID: 32458979

73. Roels S, Driks A, Losick R. Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in Bacillus subtilis. Journal of bacteriology. 1992; 174(2):575–85. https://doi.org/10.1128/jb.174.2.575-585.1992 PMID: 1729246

74. Pellegrini O, Mathy N, Gogos A, Shapiro L, Condon C. The Bacillus subtilis ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. Molecular microbiology. 2005; 56 (5):1139–48. https://doi.org/10.1111/j.1365-2958.2005.04606.x PMID: 15882409

75. Brand LA, Strauss E. Characterization of a new pantothenate kinase isoform from Helicobacter pylori. Journal of Biological Chemistry. 2005; 280(21):20185–8. https://doi.org/10.1074/jbc.C500044200 PMID: 15795230

76. Schumacher MA, Lee J, Zeng W. Molecular insights into DNA binding and anchoring by the Bacillus subtilis sporulation kinetochore-like RacA protein. Nucleic Acids Res. 2016; 44(11):5438–49. Epub 20160416. https://doi.org/10.1093/nar/gkw248 PMID: 27085804; PubMed Central PMCID: PMC4914108.

77. Chance MR, Bresnick AR, Burley SK, Jiang J-S, Lima CD, Sali A, et al. Structural genomics: a pipeline for providing structures for the biologist. Protein science: a publication of the Protein Society. 2002; 11 (4):723. https://doi.org/10.1110/ps.4570102 PMID: 11910018

78. Jez JM. Revisiting protein structure, function, and evolution in the genomic era. Journal of invertebrate pathology. 2017; 142:11–5. https://doi.org/10.1016/j.jip.2016.07.013 PMID: 27486121

79. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Deinhardt K, Darie CC. Protein–protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. Cellular and molecular life sciences. 2014; 71(2):205–28. https://doi.org/10.1007/s00018-013-1333-1 PMID: 23579629

80. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of molecular biology. 1970; 48(3):443–53. https://doi.org/10.1016/0022-2836(70)90057-4 PMID: 5420325

81. Fowler NJ, Williamson MP. The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure. 2022. https://doi.org/10.1016/j.str.2022.04.005 PMID: 35537451

82. Patel NY, Baria DM, Yagnik SM, Rajput KN, Panchal RR, Raval VH. Bio-prospecting the future in perspective of amidohydrolase L-glutaminase from marine habitats. Applied Microbiology and Biotechnology. 2021; 105(13):5325–40. https://doi.org/10.1007/s00253-021-11416-6 PMID: 34236482

83. Durthi CP, Pola M, Rajulapati SB, Kola AK, Kamal MA. Versatile and valuable utilization of amidohydrolase L-glutaminase in pharma and food industries: A review. Current Drug Metabolism. 2020; 21 (1):11–24. https://doi.org/10.2174/1574884715666200116110542 PMID: 31951174

84. Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D, Mohan R. Advances in microbial amylases. Biotechnology and applied biochemistry. 2000; 31(2):135–52. https://doi.org/10.1042/ba19990073 PMID: 10744959

85. Konsoula Z, Liakopoulou-Kyriakides M. Co-production of α-amylase and β-galactosidase by Bacillus subtilis in complex organic substrates. Bioresource Technology. 2007; 98(1):150–7.

86. Gründling A, Manson MD, Young R. Holins kill without warning. Proceedings of the National Academy of Sciences. 2001; 98(16):9348–52. https://doi.org/10.1073/pnas.151247598 PMID: 11459934

87. Saier MH Jr, Reddy BL. Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. Journal of bacteriology. 2015; 197(1):7–17. https://doi.org/10.1128/JB.02046-14 PMID: 25157079

88. Gao Y, Feng X, Xian M, Wang Q, Zhao G. Inducible cell lysis systems in microbial production of biobased chemicals. Applied microbiology and biotechnology. 2013; 97(16):7121–9. https://doi.org/10.1007/s00253-013-5100-x PMID: 23872961

89. Liu X, Curtiss R III. Nickel-inducible lysis system in Synechocystis sp. PCC 6803. Proceedings of the National Academy of Sciences. 2009; 106(51):21550–4. https://doi.org/10.1073/pnas.0911953106 PMID: 19995962

90. Miyake K, Abe K, Ferri S, Nakajima M, Nakamura M, Yoshida W, et al. A green-light inducible lytic system for cyanobacterial cells. Biotechnology for biofuels. 2014; 7(1):1–8.

91. Strauss E, Kinsland C, Ge Y, McLafferty FW, Begley TP. Phosphopantothenoylcysteine synthetase from Escherichia coli: identification and characterization of the last unidentified coenzyme A biosynthetic enzyme in bacteria. Journal of Biological Chemistry. 2001; 276(17):13513–6.

92. Suryatin Alim G, Iwatani T, Okano K, Kitani S, Honda K. In vitro production of coenzyme A using thermophilic enzymes. Applied and environmental microbiology. 2021; 87(14):e00541–21. https://doi.org/10.1128/AEM.00541-21 PMID: 33990309

93. Giraud M-F, Leonard GA, Field RA, Berlind C, Naismith JH. RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. Nature structural biology. 2000; 7(5):398–402. https://doi.org/10.1038/75178 PMID: 10802738

94. Kahraman H. The Importance of L-Rhamnose Sugar. Biomedical Journal of Scientific & Technical Research. 2019; 21:15906–8. https://doi.org/10.26717/BJSTR.2019.21.003606

95. Xu W, Zhang W, Zhang T, Jiang B, Mu W. L-Rhamnose isomerase and its use for biotechnological production of rare sugars. Applied microbiology and biotechnology. 2016; 100(7):2985–92. https://doi.org/10.1007/s00253-016-7369-z PMID: 26875877

96. Nardini M, Dijkstra BW. α/β Hydrolase fold enzymes: the family keeps growing. Current opinion in structural biology. 1999; 9(6):732–7.

97. Zheng Q, Wang S, Duan P, Liao R, Chen D, Liu W. An α/β-hydrolase fold protein in the biosynthesis of thiostrepton exhibits a dual activity for endopeptidyl hydrolysis and epoxide ring-opening/macrocyclization. Proceedings of the National Academy of Sciences. 2016; 113(50):14318–23. https://doi.org/10.1073/pnas.1612607113 PMID: 27911800

98. Burk DL. X-ray structure of the AAC(6')-Ii antibiotic resistance enzyme at 1.8 A resolution; examination of oligomeric arrangements in GNAT superfamily members. Protein Science. 2003; 12(3):426–37. https://doi.org/10.1110/ps.0233503 PMID: 12592013

99. Garefalaki V, Papavergi M-G, Savvidou O, Papanikolaou G, Felföldi T, Márialigeti K, et al. Comparative Investigation of 15 Xenobiotic-Metabolizing N-Acetyltransferase (NAT) Homologs from Bacteria. Applied and environmental microbiology. 2021; 87(19):e0081921–e. Epub 2021/09/10. https://doi.org/10.1128/AEM.00819-21 PMID: 34288706.

100. Yun C-S, Hasegawa H, Nanamiya H, Terakawa T, Tozawa Y. Novel BacterialN-Acetyltransferase Gene for Herbicide Detoxification in Land Plants and Selection Maker in Plant Transformation. Bioscience, Biotechnology, and Biochemistry. 2009; 73(5):1000–6. https://doi.org/10.1271/bbb.80777 PMID: 19420728

101. Bhatia Y, Mishra S, Bisaria VS. Microbial β-Glucosidases: Cloning, Properties, and Applications. Critical Reviews in Biotechnology. 2002; 22(4):375–407. https://doi.org/10.1080/07388550290789568 PMID: 12487426

102. Viikari L, Alapuranen M, Puranen T, Vehmaanperä J, Siika-Aho M. Biofuels. Advances in biochemical engineering/biotechnology. 2007;108.

103. Amin K, Tranchimand S, Benvegnu T, Abdel-Razzak Z, Chamieh H. Glycoside hydrolases and glycosyltransferases from hyperthermophilic archaea: Insights on their characteristics and applications in biotechnology. Biomolecules. 2021; 11(11):1557. https://doi.org/10.3390/biom11111557 PMID: 34827555

104. Schröder C, Blank S, Antranikian G. First glycoside hydrolase family 2 enzymes from Thermus antranikianii and Thermus brockianus with β-glucosidase activity. Frontiers in bioengineering and biotechnology. 2015; 3:76. https://doi.org/10.3389/fbioe.2015.00076 PMID: 26090361

105. Thuan NH, Sohng JK. Recent biotechnological progress in enzymatic synthesis of glycosides. Journal of Industrial Microbiology and Biotechnology. 2013; 40(12):1329–56. https://doi.org/10.1007/s10295-013-1332-0 PMID: 24005992

106. Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, et al. Bacterial glycosidases for the production of universal red blood cells. Nature biotechnology. 2007; 25(4):454–64. https://doi.org/10.1038/nbt1298 PMID: 17401360

107. Tiels P, Baranova E, Piens K, De Visscher C, Pynaert G, Nerinckx W, et al. A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes. Nature biotechnology. 2012; 30(12):1225–31. https://doi.org/10.1038/nbt.2427 PMID: 23159880

108. Gerigk M, Bujnicki R, Ganpo-Nkwenkwa E, Bongaerts J, Sprenger G, Takors R. Process control for enhanced L-phenylalanine production using different recombinant Escherichia coli strains. Biotechnology and bioengineering. 2002; 80(7):746–54. https://doi.org/10.1002/bit.10428 PMID: 12402320

109. Sprenger GA. From scratch to value: engineering Escherichia coli wild type cells to the production of L-phenylalanine and other fine chemicals derived from chorismate. Applied microbiology and biotechnology. 2007; 75(4):739–49. https://doi.org/10.1007/s00253-007-0931-y PMID: 17435995

110. Zhou H, Liao X, Wang T, Du G, Chen J. Enhanced L-phenylalanine biosynthesis by co-expression of pheAfbr and aroFwt. Bioresource technology. 2010; 101(11):4151–6. https://doi.org/10.1016/j.biortech.2010.01.043 PMID: 20137911

111. Huang M, Hull CM. Sporulation: how to survive on planet Earth (and beyond). Current genetics. 2017; 63(5):831–8. https://doi.org/10.1007/s00294-017-0694-7 PMID: 28421279

112. Zheng L, Halberg R, Roels S, Ichikawa H, Kroos L, Losick R. Sporulation regulatory protein GerE from Bacillus subtilis binds to and can activate or repress transcription from promoters for mother-cell-specific genes. Journal of molecular biology. 1992; 226(4):1037–50. https://doi.org/10.1016/0022-2836 (92)91051-p PMID: 1518043

113. Parashar V, Mirouze N, Dubnau DA, Neiditch MB. Structural basis of response regulator dephosphorylation by Rap phosphatases. PLoS biology. 2011; 9(2):e1000589. https://doi.org/10.1371/journal. pbio.1000589 PMID: 21346797

114. Finney LA O'Halloran TV. Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. Science. 2003; 300(5621):931–6.

115. Musiani F, Zambelli B, Bazzani M, Mazzei L, Ciurli S. Nickel-responsive transcriptional regulators. Metallomics. 2015; 7(9):1305–18. https://doi.org/10.1039/c5mt00072f PMID: 26099858

116. Crawford MA, Henard CA, Tapscott T, Porwollik S, McClelland M, Vázquez-Torres A. DksA-dependent transcriptional regulation in Salmonella experiencing nitrosative stress. Frontiers in microbiology. 2016; 7:444. https://doi.org/10.3389/fmicb.2016.00444 PMID: 27065993

117. Łyżeń R, Maitra A, Milewska K, Kochanowska-Łyżeń M, Hernandez VJ, Szalewska-Pałasz A. The dual role of DksA protein in the regulation of Escherichia coli pArgX promoter. Nucleic Acids Research. 2016; 44(21):10316–25. https://doi.org/10.1093/nar/gkw912 PMID: 27915292

118. Horsburgh MJ, Moir A. σM, an ECF RNA polymerase sigma factor of Bacillus subtilis 168, is essential for growth and survival in high concentrations of salt. Molecular microbiology. 1999; 32(1):41–50.

119. Yoshimura M, Asai K, Sadaie Y, Yoshikawa H. Interaction of Bacillus subtilis extracytoplasmic function (ECF) sigma factors with the N-terminal regions of their potential anti-sigma factors. Microbiology. 2004; 150(3):591–9. https://doi.org/10.1099/mic.0.26712-0 PMID: 14993308

120. Bessman MJ, Frick DN, O'Handley SF. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. Journal of Biological Chemistry. 1996; 271(41):25059–62. https://doi.org/10.1074/jbc.271.41.25059 PMID: 8810257

121. Fisher DI, Cartwright JL, Harashima H, Kamiya H, McLennan AG. Characterization of a Nudix hydrolase from Deinococcus radiodurans with a marked specificity for (deoxy) ribonucleoside 5'-diphosphates. BMC biochemistry. 2004; 5(1):1–8. https://doi.org/10.1186/1471-2091-5-7 PMID: 15147580

122. Trievel RC, Rojas JR, Sterner DE, Venkataramani RN, Wang L, Zhou J, et al. Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. Proceedings of the National Academy of Sciences. 1999; 96(16):8931–6. https://doi.org/10.1073/pnas.96.16.8931 PMID: 10430873

123. Dash A, Modak R. Protein acetyltransferases mediate bacterial adaptation to a diverse environment. Journal of Bacteriology. 2021; 203(19):e00231–21. https://doi.org/10.1128/JB.00231-21 PMID: 34251868

124. Favrot L, Blanchard JS, Vergnolle O. Bacterial GCN5-related N-acetyltransferases: from resistance to regulation. Biochemistry. 2016; 55(7):989–1002. https://doi.org/10.1021/acs.biochem.5b01269 PMID: 26818562

125. Bepperling A, Alte F, Kriehuber T, Braun N, Weinkauf S, Groll M, et al. Alternative bacterial two-component small heat shock protein systems. Proceedings of the National Academy of Sciences. 2012; 109(50):20407–12. https://doi.org/10.1073/pnas.1209565109 PMID: 23184973

126. Cocotl-Yanez M, Moreno S, Encarnacion S, Lopez-Pliego L, Castaneda M, Espín G. A small heat-shock protein (Hsp20) regulated by RpoS is essential for cyst desiccation resistance in Azotobacter vinelandii. Microbiology. 2014; 160(3):479–87. https://doi.org/10.1099/mic.0.073353-0 PMID: 24385478

127. Khaskheli GB, Zuo F, Yu R, Chen S. Overexpression of small heat shock protein enhances heat-and salt-stress tolerance of Bifidobacterium longum NCC2705. Current Microbiology. 2015; 71(1):8–15. https://doi.org/10.1007/s00284-015-0811-0 PMID: 25842174

128. Singh H, Appukuttan D, Lim S. Hsp20, a small heat shock protein of Deinococcus radiodurans, confers tolerance to hydrogen peroxide in Escherichia coli. Journal of Microbiology and Biotechnology. 2014; 24(8):1118–22. https://doi.org/10.4014/jmb.1403.03006 PMID: 24743570

129. Ventura M, Canchaya C, Zhang Z, Fitzgerald GF, van Sinderen D. Molecular characterization of hsp20, encoding a small heat shock protein of Bifidobacterium breve UCC2003. Applied and environmental microbiology. 2007; 73(14):4695–703. https://doi.org/10.1128/AEM.02496-06 PMID: 17513584

130. Zheng L, Cash VL, Flint DH, Dean DR. Assembly of iron-sulfur clusters: identification of an iscSUA-hscBA-fdx gene cluster from Azotobacter vinelandii. Journal of Biological Chemistry. 1998; 273 (21):13264–72. https://doi.org/10.1074/jbc.273.21.13264 PMID: 9582371

131. Braz VS, Marques MV. Genes involved in cadmium resistance in Caulobacter crescentus. FEMS Microbiology Letters. 2005; 251(2):289–95. https://doi.org/10.1016/j.femsle.2005.08.013 PMID: 16168577

132. Crapoulet N, Barbry P, Raoult D, Renesto P. Global transcriptome analysis of Tropheryma whipplei in response to temperature stresses. Journal of bacteriology. 2006; 188(14):5228–39. https://doi.org/10.1128/JB.00507-06 PMID: 16816195

133. Ogura M, Tsukahara K. SwrA regulates assembly of Bacillus subtilis DegU via its interaction with N-terminal domain of DegU. The journal of biochemistry. 2012; 151(6):643–55. https://doi.org/10.1093/jb/mvs036 PMID: 22496484

134. Ghelardi E, Salvetti S, Ceragioli M, Gueye SA, Celandroni F, Senesi S. Contribution of surfactin and SwrA to flagellin expression, swimming, and surface motility in Bacillus subtilis. Applied and environmental microbiology. 2012; 78(18):6540–4. https://doi.org/10.1128/AEM.01341-12 PMID: 22773650

135. Dall'Agnol HP, Baraúna RA, de Sá PH, Ramos RT, Nóbrega F, Nunes CI, et al. Omics profiles used to evaluate the gene expression of Exiguobacterium antarcticum B7 during cold adaptation. BMC genomics. 2014; 15(1):1–12. https://doi.org/10.1186/1471-2164-15-986 PMID: 25407400

136. Kobayashi K, Iwano M. BslA (YuaB) forms a hydrophobic layer on the surface of Bacillus subtilis biofilms. Molecular microbiology. 2012; 85(1):51–66. https://doi.org/10.1111/j.1365-2958.2012.08094.x PMID: 22571672

137. De Carvalho CC. Marine biofilms: a successful microbial strategy with economic implications. Frontiers in marine science. 2018; 5:126.

138. Souza-Egipsy V, Vega JF, González-Toril E, Aguilera Á. Biofilm mechanics in an extremely acidic environment: microbiological significance. Soft Matter. 2021; 17(13):3672–80. https://doi.org/10.1039/d0sm01975e PMID: 33683248

139. Yin W, Wang Y, Liu L, He J. Biofilms: the microbial "protective clothing" in extreme environments. International journal of molecular sciences. 2019; 20(14):3423.

140. Barta ML, Thomas K, Yuan H, Lovell S, Battaile KP, Schramm VL, et al. Structural and biochemical characterization of Chlamydia trachomatis hypothetical protein CT263 supports that menaquinone synthesis occurs through the futalosine pathway. Journal of Biological Chemistry. 2014; 289 (46):32214–29. https://doi.org/10.1074/jbc.M114.594325 PMID: 25253688

141. Choi H-P, Juarez S, Ciordia S, Fernandez M, Bargiela R, Albar JP, et al. Biochemical characterization of hypothetical proteins from Helicobacter pylori. PLoS One. 2013; 8(6):e66605. https://doi.org/10.1371/journal.pone.0066605 PMID: 23825549

142. Zhang W, Culley DE, Gritsenko MA, Moore RJ, Nie L, Scholten JC, et al. LC–MS/MS based proteomic analysis and functional inference of hypothetical proteins in Desulfovibrio vulgaris. Biochemical and biophysical research communications. 2006; 349(4):1412–9. https://doi.org/10.1016/j.bbrc.2006.09.019 PMID: 16982031