Research Article

Atron, B., Identification and characterization of potential druggable targets among Essential Hypothetical Proteins of A. Baumannii, 2022. 5(2): p. 145-165. DOI: 10.38001/ijlsb.1009800

Identification and characterization of potential druggable targets among Essential Hypothetical Proteins of A. Baumannii

Bydaa Atron[1*]

ABSTRACT

Acinetobacter baumannii is a critical pathogen responsible for a wide range of infections. A. baumannii exhibits resistance to a variety of antibiotic classes, emphasizing that new therapeutic targets are urgently needed. In A. baumannii, ATCC 179778. Among 458 essential genes, 47 are uncharacterized and considered to be an essential hypothetical protein (EHPs). In this study, the functional characterization of EHPs was conducted utilizing variable computational tools. The physicochemical parameters, subcellular localization, domain identification, 3D structure, and virulence capabilities were predicted for the EHPs. According to our results, they were shown to be of a different functional category such as: transporters, enzymes, binding proteins, and virulence factors. Enzymes made up around 47% of the total and 17.6% were predicted as virulence factors. BLASTP analysis against human proteome was tested to identify proteins that is found exclusively in pathogen. The druggable property of the proteins was examined. Of 34, 27 essential pathogen-specific proteins could be considered as new pharmacological targets. In draggability analysis, one EHP turned out to be druggable while the others were novel. Our findings might assist in the innovation of new drugs for *Acinetobacter baumannii* infections.

ARTICLE HISTORY
Received
14 October 2021
Accepted
30 December 2021

KEYWORDS
Acinetobacter baumannii,
essential hypothetical proteins,
functional annotation,
drug targets

Introduction

Acinetobacter baumannii is an important bacterium that can induce a vast scope of diseases, including hospital-acquired infections. Antibiotic overuse and misuse have been increasing as a result, antibiotic-resistant strains of A. baumannii have arisen [1]. World Health Organization labeled the multidrug-resistant (MDR) A. baumannii (resistance to at least three separate classes of antimicrobial drugs) as one of three "critical priority pathogens" that in urgent need of novel therapeutics. [2]. A. baumannii is characterized by possessing intrinsic antimicrobial resistance mechanisms that can be expressed continuously or in response to antibiotic stress. The regulating of antibiotic transport via bacterial membranes, mutation of the antibiotic target site, and enzyme alterations that result in antibiotic neutralization are all types of antimicrobial resistance mechanisms.

[1] Sivas Cumhuriyet University, Faculty of Medicine, Department of Medical Microbiology, Sivas / Turkey
*Corresponding Autor: Bydaa Atron, e-mail: bidaahamad@gmail.com

[3]. More worryingly, A. baumannii can quickly develop resistance to all clinically available medicines, making illness caused by resistant strains exceedingly difficult to treat. Colistin-resistant isolates are spreading at an alarming rate throughout the world, making it critical to discover alternative therapeutics [4]. A proportion of genes is definitely vital for an organism. These genes are called essential genes. The proteins that they encode are called essential proteins (EPs). Essential genes and proteins are theoretically possible therapeutic targets since their deletion or inactivation causes the organism's death. As a result, predicting the essentiality of a gene in a pathogenic microbe might aid in the prospective therapeutics for the development of antimicrobial medicines. Database of Essential Genes (DEG) is a collection of essential elements from several organisms [5]. A large number of essential genes for A. baumannii have been identified using antisense RNA technique. However, some of them are hypothetical (uncharacterized) essential proteins (EHPs)[5]. Functional and structural characterization of these proteins will provide information on their structures, functionalities, and metabolic roles. [6]. There is evidence for the successful use of computational tools for the annotation of HPs of various pathogens, including *Vibrio parahaemolyticus* [7], *Candida dubliniensis* [8], *Chlamydia trachomatis* [9], *Leptospira interrogans* [10], *Mycobacterium tuberculosis* [11], *Haemophilus influenza* [12], *Neisseria meningitidis* MC58 [13], *Mycobacterium leprae* Br492 [14], and *Borrelia burgdorferi* [15].

Till now, no experimental information on EHPs of *A. baumannii*, the present attempt was made to annotate the function of these proteins using computational approach.

Material and Methods

Sequence retrieval and analysis

A.baumannii ATCC 179778 has 3587 coding genes, 458 (13%) of which are identified as essential genes in Database of Essential Genes. Out of the 458 essential genes, about 47 proteins with unknown function, and thus need characterization. Pseudo genes which are proteins having less than 100 amino acid residues were omitted in this study to reduce the misperceptions in functional annotation workflow (13 sequences were found with < 100 amino acids in this study), the remaining 34 were selected for the analysis.

Functional annotation and domain identification

Numerous publicly accessible bioinformatics tools and databases were utilized to assign functions to all A. baumannii EHPs. Tools involved: Pfam, NCBI-BLAST, Conserved domain database, and InterProScan [16].

BLASTp: Basic Local Alignment Search Tool

BLASP (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) is used to predict identical or similar proteins through the identification of the homologous sequences with known functions from the non-redundant NCBI protein database [16]. BLASP tool compares the amino acid chain of proteins and finds the best local alignment [17].

Pfam

Pfam (http://pfam.xfam.org/) is a comprehensive database of conserved protein families. This tool predicts protein families based on multiple sequence alignments and hidden Markov models [18]. The seed alignments are used to build profile hidden Markov models (HMMs) that can be used to search any sequence database for homologues in a sensitive and accurate fashion. Those homologues that score above the curated inclusion thresholds are aligned against the profile to make a full alignment [19].

InterProScan

The InterPro database (https://www.ebi.ac.uk/interpro/) provides an integrative families, domains and motif classification. InterProScan integrates various protein traditional authentication techniques from the InterPro project for motif identification [20]. InterPro also dispenses additional information such as descriptions, literature references and Gene Ontology (GO) terms, to produce a comprehensive resource for protein classification [21].

Conserved Domain Database (CDD)

The Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/) is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models

imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAMs) [22].

Physicochemical characterization

Physicochemical characterization of the EHPs was performed using ExPASy PROTPARAM proteomics tools (https://web.expasy.org/protparam/). The physicochemical characteristics of HP can contribute to the understanding of its products and the development of additional in vitro assays. ProtParam predicts the molecular weight (MW, given in Da), isoelectric point (IP), number of AAs (No of AA), instability index (II) and atomic composition of the targets. This tool also calculates other parameters, as the extinction coefficient (EC, given in $M^{-1}cm^{-1}$), useful in spectrophotometry experiments, aliphatic index (AI), related with thermostability and grand average of hydropathy (GRAVY), that indicates the interaction of proteins with water [23].

Subcellular localization analysis

The functions of proteins are mainly determined by their subcellular localizations in cells. SOSUI predicts subcellular localization of proteins together with the prediction of transmembrane helices [24]. SOSUI is a web-based tool with its Internet address (https://harrier.nagahama-i-bio.ac.jp/sosui/mobile/). The performance of this tool is 99% for the discrimination between two types of proteins (membrane against soluble protein) and 96% for the prediction of transmembrane helices [25].

Virulence factor prediction

The web server VICMpred (http://www.imtech.res.in/raghava/vicmpred/) has been used to predict virulence factors among EHPs. According to the sequences patterns and amino acid compositions VICMpred can categories gram-negative bacterial proteins into information molecules, proteins involved in the cellular process, and virulence factors [26]. Virulence factors allow the pathogens to establish themselves in the host. They include adhesions, toxins, and hemolytic molecules [27].

Host non-homology analysis

Here, proteins found exclusively in the pathogen were identified in this step. This analysis is crucial to identify proteins that are unique to bacteria and do not have any similarity with the host, in order to design specific antimicrobials. Protein BLAST search against

the human proteome with the e-value was adjusted to 0.0001 and cut-off to 100. Protein that showed no significant hits passed through this level [28,29].

Target Identification

A target is considered if proven to show the potentiality to bind with drug-like substances with high affinity. Following the prior step, the pathogen-specific proteins were subjected to a similarity search against the DrugBank targets repertoire. The existence of similar proteins with the same function in the DrugBank target list provides evidence for their draggability. The absence demonstrates the protein's novelty, and it is thus termed as a "novel target." [30].

Secondary structure prediction:

Determination of protein structure is essential for the understanding of protein function. The secondary structure of our models for HPs was estimated by SOPMA server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). SOPMA predicts a three-state description of the secondary structure: alpha-helix, beta-sheet and coil [31].

Homology Modeling and Validation

In order to recognize the protein functions of proteins at a molecular level, it is sometimes necessary to determine their 3D structure accurately and reliably. First, the 3D models of the EHPs (pdb format) were predicted by CPH model 3.2 server (https://services.healthtech.dtu.dk/service.php?CPHmodels-3.2). CPH model 3.2 is a protein homology modeling server. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions. This tool was previously accessible through (http://www.cbs.dtu.dk/) [32]. UCSF Chimera package used to visualize the 3D structure of predicted proteins. UCSF Chimera is a program for the interactive visualization and analysis of molecular structures and related data, including density maps, trajectories, and sequence alignments [33].

Model refinement

To test the correctness of 3D proteins mode, The structure results were cross checked by ERRAT and PROCHECK tools were used. [34]. ERRAT verifies protein structures determined by crystallography. Error values are plotted as a function of the position of a sliding 9-residue window. The error function is based on the statistics of non-bonded atom-atom interactions in the reported structure; compared to a database of reliable high-

resolution structures [34]. PROCHECK tool checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry [35].

Results and Discussion

Understanding the functions of essential hypothetical proteins is important since it facilitates the further comprehension of their role in physiological and pathological processes and the identification of novel classes of therapeutic targets and ultimately to combat with the infection. The present study utilized various available bioinformatics to identify drug and vaccine targets from essential hypothetical proteins of *A. baumannii*. In this paper, out of 458 *A. baumannii* essential genes, 47 EHPs were retrieved. After eliminating the pseudogenes, 34 EHPs were chosen for functional analysis. Bioinformatics methods were employed to assign functions, and they were found to be present in many functional groups (Table1).

Table 1 functionally annotated EHPs in A. baumannii ATCC 179778

EHP	DEG Accession Number	Function	Type	Pathogen specific
1	DEG10430024	Iron-sulfur cluster insertion protein ErpA	Metal binding	✕
2	DEG10430040	Uracil reductase RibD	Oxidoreductases	✔
3	DEG10430041	LPS export ABC transporter permease LptG	Permease	✔
4	DEG10430066	ABC transporter C-terminal domain	transporter	✔
5	DEG10430083	Ton b family C-terminal domain protein	transporter	✔
6	DEG10430087	Cytoskeletal protein RodZ	DNA binding	✔
7	DEG10430131	Biotin--[acetyl-CoA-carboxylase] ligase	ligase	✕
8	DEG10430134	3-oxoacyl-ACP reductase FabG	Oxidoreductases	✔
9	DEG10430136	hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	kinase	✔
10	DEG10430158	Solute carrier families 5 and 6-like	transporter	✕
11	DEG10430166	RNase H	endoribonucleases	✔
12	DEG10430174	LPS export ABC transporter periplasmic protein LptC	transporter	✔
13	DEG10430212	Putative integral membrane protein	transporter	✔
14	DEG10430249	Trna threonylcarbamoyl adenosine modification protein	hydrolase	✔
15	DEG10430253	Hypothetical protein	transporter	✔
16	DEG10430256	DUF3465 domain-containing protein	transporter	✔
17	DEG10430271	DUF3465 domain-containing protein	DUF	✔

Table 1 functionally annotated EHPs in A. baumannii ATCC 179778 (Continued)

18	DEG10430275	DIP1984 family protein	DUF	✔
19	DEG10430278	DUF3465 domain-containing protein	DUF	✔
20	DEG10430290	high frequency lysogenization protein HflD	hydrolase	✔
21	DEG10430321	DUF3465 domain-containing protein	DUF	✔
22	DEG10430344	Serine--trna ligase	ligase	✕
23	DEG10430346	Folate-binding protein	Transferases	✕
24	DEG10430349	Putative integral membrane protein	transporter	✔
25	DEG10430256	TamB, inner membrane protein subunit o	Secretory protein	✔
26	DEG10430361	membrane protein insertase	insertases	✕
27	DEG10430370	Glutathione-dependent formaldehyde-activating enzyme	Lyases	✕
28	DEG10430372	23S rRNA (adenine(2030)-N(6))-methyltransferase	transferases	✔
29	DEG10430374	rRNA maturation RNase YbeY	Ribosomal protein	✔
30	DEG10430415	p)ppGpp synthase/hydrolase	hydrolase	✕
31	DEG10430429	GNAT family N-acetyltransferas	transferases	✔
32	DEG10430446	DUF4175 domain-containing protein	DUF	✔
33	DEG10430451	Nucleotidyl transferase	transferases	✔
34	DEG10430452	GlmU superfamily N-acetylglucosamine-1-phosphate-uridyltransferase	transferases	✔

Around 16 (47%) of EHPs were enzymes,6 (17%) transporters,1(2%) binding protein, 5(14%) miscellaneous proteins, 4 (11%) domain of unknown function containing protein (DUF). Among EHPs 6 (17%) found to be virulence factors (Fig.1).

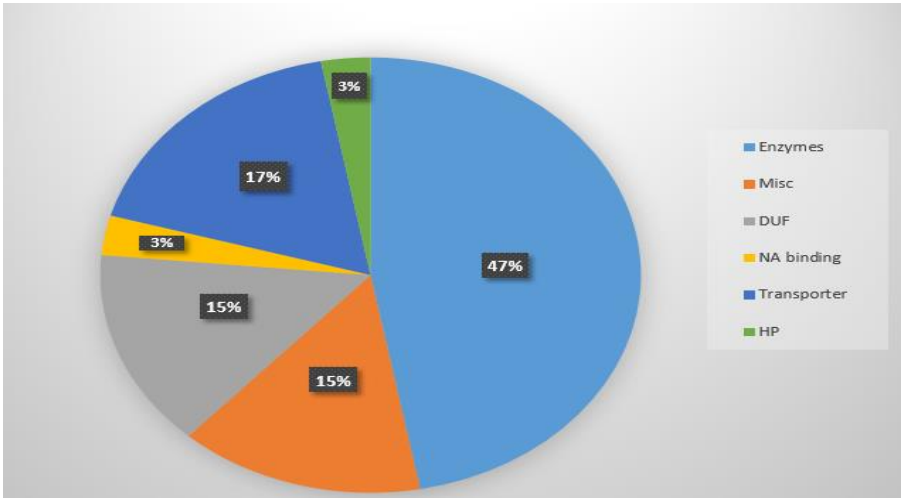

Fig 1 Predicted Essential Proteins

Exploring pathogenicity at the molecular level requires in-depth knowledge of these functional groups. The descriptions of the protein groups are shown below.

Functional annotation

Hydrolases

Catalyzes the hydrolysis of a bonds. Bacterial hydrolase enzymes tin the splitting of distinct peptidoglycan bonds. Hydrolases are also participated in a range of additional processes, notably peptidoglycan development, autolysis, and septal cleavage on cell division [36]. Two of the proteins in this study were identified to be hydrolases. Threonylcarbamoyl adenosine biosynthesis protein TsaE and ysogenization regulator HflD are hydrolyses. The TsaE family of enzymes is engaged in the creating of threonylcarbamoyl adenosine (t(6)A) [37].

High frequency lysogenization protein (HflD) is a membrane protein that suppresses lambda phage by binding to and degrading CII and, as a result, limiting its interactions with DNA [38].

Bifunctional ppGpp (guanosine 3'-diphosphate 5'-diphosphate), a rigorous response mediator that regulate a range of cellular processes in response to nutritional availability alterations. This enzyme catalyzes both the production of pppGpp, which is afterwards hydrolyzed to create ppGpp which has a synthase and hydrolase activity), and the hydrolysis of ppGpp [39].

Oxidoreductases

Oxidoreductases catalyze the transfer of electrons between electron donor and acceptor, generally with cofactors such as nicotinamide adenine dinucleotide phosphate (NADP) or nicotinamide adenine dinucleotide (NAD) [40]. Two EHPs were discovered to be members of the oxidoreductase family. Uracil reductase, RibD, and 3-oxoacyl-ACP reductase. Riboflavin synthesis is supported by uracil reductase [41]. Animals can obtain riboflavin from nutrition, whereas bacteria synthesize it from scratch. This vital part of bacterial metabolism is absent in humans, indicating that this biosynthetic route might be a source of antimicrobial therapeutic targets. Dawson et al. designed methods for the purification and crystallization of RibD. They were successful in obtaining two ordered crystal forms with probable binding sites that were ready for ligand discovery [42].

Ligases

In the current study, two EHPs were predicted to be ligases. Ligase enzyme catalyzes the creation of a new chemical link between two large molecules. Biotin-dependent acetyl-CoA carboxylase enzyme catalyzes ATP-dependent carboxylation of acetyl-CoA to malonyl-CoA. This process offers the required active substrate for fatty acid production [43]. This enzyme is well known to be a major target of antifungal and antibiotic medications [44,45]. The second anticipated ligase in this investigation is serine-tRNA ligase, which acts as a catalyst to the attachment of serine to tRNA (Ser). Which involved in in glycine, serine, and threonine metabolism and aminoacyl-trna biosynthesis.

Lyases

One EHP was predicted as lyase which is Glutathione-dependent formaldehyde-activating enzyme. In bacteria it serves in the pathway of formaldehyde detoxification [46].

Transferases

Six EHPs were predicted as transferases. Transferases catalyze the transfer of specific functional groups between molecules. They are involved in hundreds of different biochemical pathways throughout biology [47].

Aminomethyltransferase (T-protein) is part of the glycine cleavage system, along with P-protein, L-protein, and H-protein. This system is involved in glycine and serine catabolism pathways.

Ribosomal RNA large subunit methyltransferase methylates the adenine in 23S rRNA during ribosome biogenesis. The usage of extracellular DNA as neutrient is also mediated by this enzyme [48].

The N-acetyltransferases (NAT) is an enzyme that uses acetyl coenzyme A (CoA) to transfer an acetyl group to a substrate. This enzyme mediate pantothenate and CoA biosynthesis. Acetyl coenzyme A transfer activity is responsible for resistance to Aminoglycoside antibiotics [49].

UDP-N-acetylglucosamine diphosphorylase/glucosamine-1-phosphate N-acetyltransferase is part of the pathway for Amino sugar, nucleotide sugar, and O antigen nucleotide sugar biosynthesis [50]. The enzyme found in several bacteria such as *Escherichia coli*, *Bacillus subtilis*, and *Haemophilus influenza*, and has similar activity of glucosamine-1-phosphate N-acetyltransferase [51].

N-Acetylglucosamine-1-phosphate uridyltransferase abbreviated as GlmU, is a bifunctional enzyme exclusive to prokaryotes. GlmU is involved in the peptidoglycan and lipopolysaccharide biosynthesis [52]. Several studies were conducted to identify inhibitors specific to target this enzyme. [53,54].

Permeases

One permease was predicted. Permeases are membranous transporters that facilitate molecules diffusion around the membrane [55]. LPS export ABC transporter permease (LptG), an enzyme that carry out the function of lipopolysaccharide transport. This transport system uses energy from ATP hydrolysis in the cytoplasm to facilitate extraction of LPS from the outer face of the cytoplasmic membrane prior to transport to the cell surface. [56,57]. LptG system comprises seven different proteins (LptA, LptB, LptC, LptD, LptE, LptF, and LptG) that assemble into a complex spanning from the cytoplasm to the outer membrane [58].

Nucleic acid Binding Proteins

The bacterial nucleoid showes a multi-level hierarchical structural organization similar to eukaryotic chromatin [59]. This hierarchical structure maintains the global nucleoid organization and ensures the accessibility of particular chromosomal regions for DNA-dependent processes, such as replication, transcription, DNA repair, and recombination. The organization of the highly compacted yet dynamic nucleoid structure reflects the input of many different factors, including molecular crowding, depletion forces, DNA supercoiling, and nucleic acid binding proteins [60].

In this study one binding proteins was identified, which is cytoskeleton protein from RodZ family. It's a cytoskeletal protein that controls cell-shape by the manipulation of the length of the long axis [61].

Transporters

Six EHPs were identified as transporters. ABC transporter of C-terminal domain plays a role in ATP-binding cassette (ABC) transporter-dependent pathway. Polysaccharides O-antigenic (O-PSs) and Surface glycoconjugates are synthesized via this pathway [62]. Surface glycoconjugates are essential for cell envelope structure integrity, protection against host immunological responses, signaling cascades, and biofilm development [63]. ABC transporters act as antimicrobial targets [64]. TonB is a protein abundantly found in

many Gram-negative bacteria, and thought to regulate nutritional imports [65]. TonB system may have a function not just in protein import as well as in protein secretion [66]. Putative membrane protein Solute carrier 5 family and 6-like family; solute binding domain (SLC5). SLC5s co-transport sodium ions with other ions and sugers. [67].

Two EHPs were predicted to be membrane-associated proteins belong to EamA-like transporter family, which is the only drug/metabolite transporter that can be found in both prokaryote and eukaryotes [68].

LPS export ABC transporter periplasmic protein LptC, LptC is involved in lipopolysaccharide-assembly of gram-negative bacteria, which takes place in the outer membrane [69].

Endoribonucleases

It can split both single-stranded and double-stranded RNA. One protein was predicted to be Endoribonucleases. RNase H endonucleolytically cleaves RNA in RNA-DNA hybrid molecules. This activity is present in almost all organisms [70].

Virulence proteins

Pathogenic bacteria possess virulence factors that enable them to bring damage to the host. VICMpred predicted that among the EHP: 22 are related to cellular process, 1 in storage and information, 5 in metabolism, and 6 in virulence. (Table 2).

Antivirulence drugs are targeting virulence factors, renders the pathogen nonpathogenic (avirulent). In comparison to conventional antibiotics, it has been theorized that antivirulence agents have less resistance serendipity [71].

Table 2 VICMpred Result

Function Prediction	No
Cellular Process	22
Metabolism Molecule	5
Virulence factor	6
Information and storage	1

Physiochemical characterization

Physiochemical properties for the proteins are essential for understanding the protein's function and interaction with other proteins. The physicochemical characteristics involve the molecular weight, positive and negative residues, amino acids, the theoretical isoelectric point (pI), extinction coefficient, which is important in studying protein interactions, is calculated from amino acids composition and grand average of

hydropathicity (GRAVY). The GRAVY value is used to determine whether the molecules are hydrophobic or hydrophilic. The results are depicted in the supplementary file.

Cellular localization

Determining protein subcellular localization is vital for understanding the role of proteins in a cell, and it also aids in the process of drug discovery and delivery. Through which it facilitates the classification of them as drug and vaccine targets. Possible drug target is cytoplasmic protein. while the membranous one considered a vaccine target. 24 were found to be cytoplasmic proteins and 10 were membrane proteins. The membrane regions were classified as primary and secondary, and the length of the transmembrane regions is predicted as well. and if the protein is rich in hydrophobic amino acids (Table 3).

Potential Targets

An ideal drug target should not show any similarity to human proteome. Hence, a host non-homology analysis was performed. Functionally annotated 34 hypothetical proteins subjected to host non-homology analysis using a BLASTp search against the human database. Out of 35 proteins, 27 were found to be non-homologous (potential drug targets). A druggability analysis was performed. Out of 27, one protein was found to be previously reported as a drug target. 3-Oxo-acyl-acyl carrier protein (ACP) reductase (FabG) is significant in bacterial fatty acid synthesis and has been identified as a target [72,73].

Table 3 Prediction of cellular localization

EHP	Region	Transmembrane sequence	Type
1	Soluble		
2	Soluble		
3	10 --32	RIVGWVMRSALLLIVLSFALSEW	Primary
	184--206	KVASPFSLITLVLVACSFIFGPL	Primary
	209--231	QSMGFRLVIALFIGLGFYYLQDF	Secondary
	237--259	LVYAPSPAWFVLMPIILMFGAGS	Primary
4	Soluble		
5	Soluble		
6	87--105	ALLAILIIAVVSLIVMGVQ	Primary
7	Soluble		
8	Soluble		
9	Soluble		

Table 3 Prediction of cellular localization (Continued)

EHP	Region	Transmembrane sequence	Type
10	11--33	SPLMAFCLSFVIMATLAPTMGIQ	Primary
	40--62	FWLLWFGTMLLLALPVCYLEIAL	Primary
	80--110	RVVGWLAVVFIPFLAGNVLSTAS	Primary
	116--138	QFAPSISGQIIFAGLAVAALVLS	Primary
	143--165	QILILLMTLGVIASIVLANTMGS	Primary
	179--199	EWGNATVLALVASGLGLGLYW	Primary
	213--235	TKTVLPIWLAQLIAVVAFGFFSL	Primary
	239--261	LPVLTWIFTGVMTSALFVQLARE	Primary
	266--288	RQLMPVLQWVIIVVAIAVWAVPE	Primary
	294--316	TLILMLWGLLICLIYAVFAGWIM	Primary
	335--357	LWRIAVRIVLPLSIIVAMIAVIG	Primary
11	Soluble		
12	Soluble		
13	1--23	MQMIAWNYASASLLCYLWFKPDI	Secondary
	32--54	PWWLIVALGVILPSIFLCLAKSL	Primary
	59--81	IVKTEIAQRLSVVLSLLAAYFFF	Secondary
	87--109	SLKLLGIGLGIFAVLLILLGHF	Primary
	119--142	AIFALMSVWFGYAAVDILLKYTS	Primary
	147--169	FTLTLNLIFITAFVLSIIYLIFQ	Primary
14	Soluble		
15	Soluble		
16	Soluble		
17	Soluble		
18	Soluble		
19	Soluble		
20	Soluble		
21	Soluble		
22	Soluble		
23	Soluble		
24	1--23	MQMIAWNYASASLLCYLWFKPDI	Secondary
	32--54	PWWLIVALGVILPSIFLCLAKSL	Primary
	59--81	IVKTEIAQRLSVVLSLLAAYFFF	Secondary
	87--108	SLKLLGIGLGIFAVLLILLGHF	Primary
	119--141	AIFALMSVWFGYAAVDILLKYTS	Primary
	147--169	FTLTLNLIFITAFVLSIIYLIFQ	Primary

Table 3 Prediction of cellular localization (Continued)

EHP	Region	Transmembrane sequence	Type
25	Soluble		
26	39--61	SGCLPLLLQMPIFLALYWVLMES	Primary
	82--104	WFILPLIMGATMFAQQMLNPQPA	Secondary
	116--138	PIMFTVFMLFFPAGLVLYWIVNN	Primary
27	Soluble		
28	Soluble		
29	Soluble		
30	Soluble		
31	Soluble		
32	92--114	WLAILASLFLHVLLWLAYRAIAR	Primary
33	Soluble		
34	Soluble		

In drug and vaccine design the determination of the 3D structure is important. The most common prediction method is the Homology approach. CPH model as modeling was used in this study for EHPs. Visualization of these predictions is performed by Chimera software. Validation of all structures was performed using the ERRAT program (Figure 2)

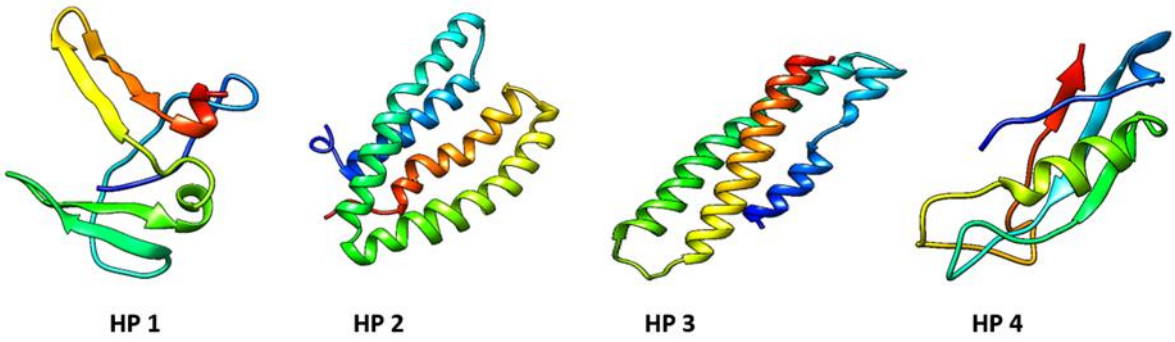

Fig 2 Predicted 3D Structure

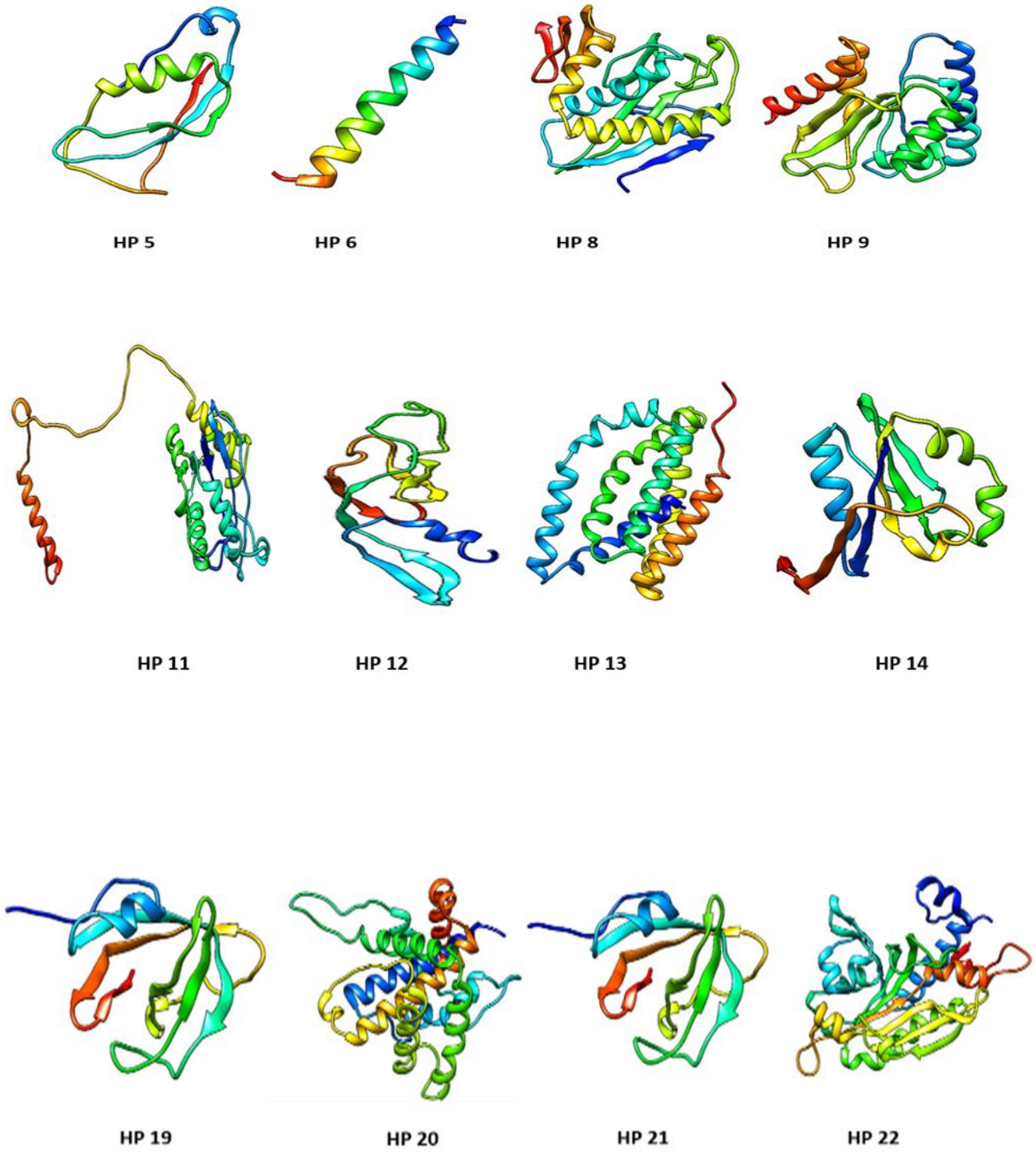

Fig 2 Predicted 3D Structure (Continued)

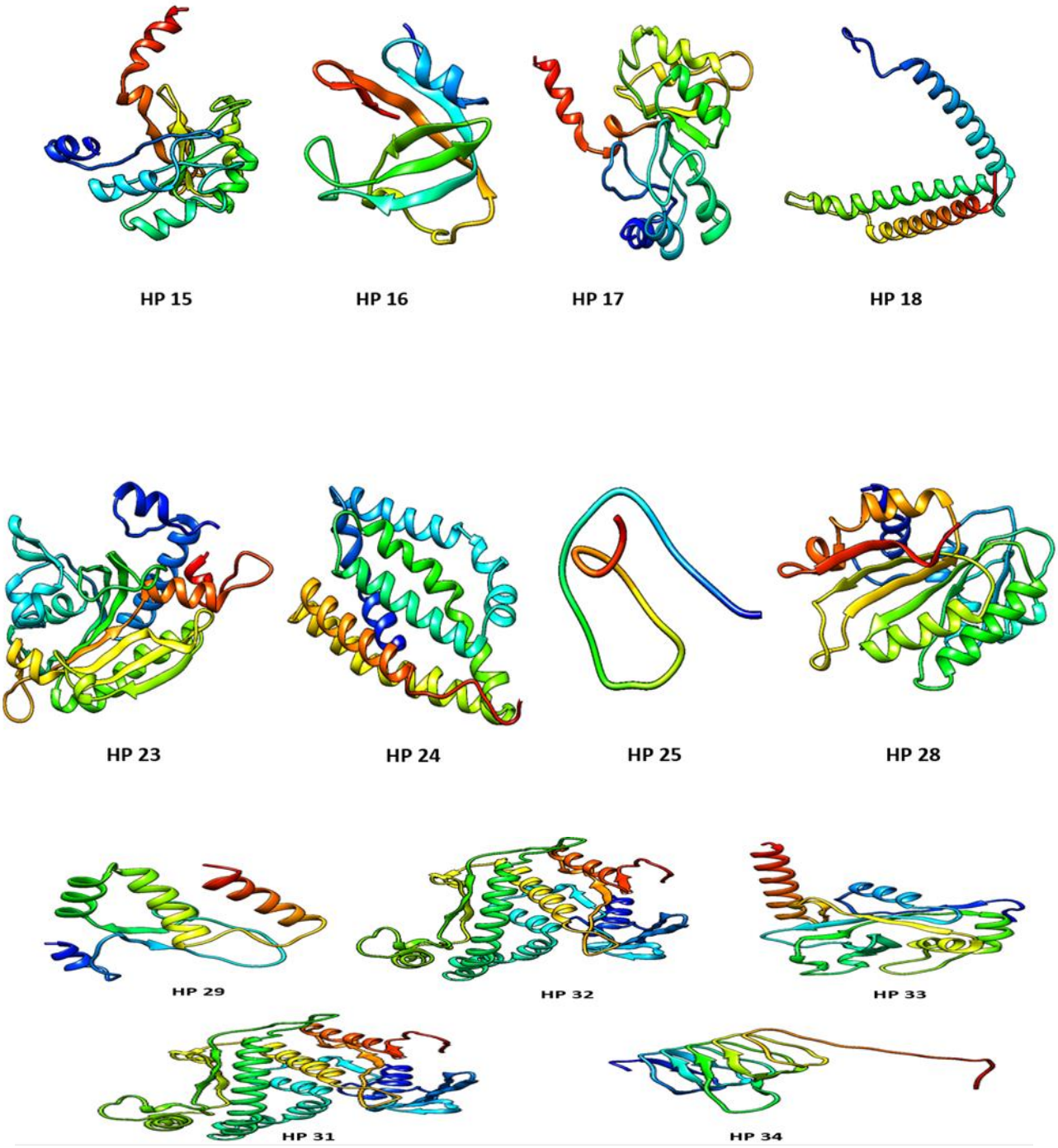

Fig 2 Predicted 3D Structure (Continued)

Conclusion

Exploring pathogenic microorganism's essential gene's function is crucial in medical research. The use of robust Insilco approaches, have been highlighted to prove the immense utilities of the knowledge databases and tools, toward characterization of

essential hypothetical proteins from *A. baumannii* ATCC 179778, whose genome wide information is elusive due to the presence of vast number of uncharacterized sequences. Characterizing these proteins will enhance our understanding of their virulence capacity and pathogenicity.

In this study, a variety of bioinformatic tools was employed to functionally characterize essential hypothetical proteins from *A. baumannii* ATCC 179778. The strategy applied predicted the function of more than 85% of hypothetical proteins pertaining to several key functional domains. Yet, some of EHPs predictions were not achievable due to scarcity of data. The cellular localization of these proteins was predicted using subcellular localization analysis. Subcellular localization helps in differentiating drug targets from vaccine targets. Furthermore, six of them were virulence factors. Host non-homology analysis shown that 27 proteins were specific to pathogen, so they are potential to be drug target candidates. 3-oxoacyl-ACP reductase is a known target [72]. The other proteins were identified as 'novel targets,' which require requires additional experimental validation. Hence, this study may facilitate future studies on the predicted EHPs as novel therapeutic targets for the drug and vaccine development.

Abbreviations

A. baumannii: *Acinetobacter baumannii*, CDD: Conserved domain database, BLAST: Basic Local Alignment Search Tool

Funding

No funding was received.

Availability of data and material

Please contact the corresponding author for any data request.

References

1. Ayoub Moubareck, C. and D. Hammoudi, Insights into Acinetobacter baumannii: A Review of Microbiological, Virulence, and Resistance Traits in a Threatening Nosocomial Pathogen. Antibiotics, 2020. 9(3): p. 119.
2. Nasr, P., Genetics, epidemiology, and clinical manifestations of multidrug-resistant Acinetobacter baumannii. The Journal of hospital infection, 2020.104(1): p. 4–11.
3. Kyriakidis, I., et al., Acinetobacter baumannii Antibiotic Resistance Mechanisms. Pathogens, 2021. 0(3): p.373.
4. Butler, D., et al., Multidrug Resistant Acinetobacter baumannii: Resistance by Any Other Name Would Still be Hard to Treat. Current infectious disease reports, 2020. 21(12): p. 46.
5. Luo, H., et al., DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. Nucleic acids research, 2021. 49(1): p. 677–686.
6. Ijaq, J., et al., A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. BMC Bioinformatics, 2019. 20(1): p. 14
7. Kodama, T., Identification of Two Gatekeeper Proteins or Vibrio parahaemolyticus T3SS2. International Journal of Infectious Diseases, 2019. 79(S1): p. 79-75.

8. Karkowska-Kuleta, J., et al., Characteristics of Extracellular Vesicles Released by the Pathogenic Yeast-Like Fungi Candida glabrata, Candida parapsilosis and Candida tropicalis. Cells, 2020. 9(7): p. 1722.
9. Chen, C., et al., The hypothetical protein CT813 is localized in the Chlamydia trachomatis inclusion membrane and is immunogenic in women urogenitally infected with C. trachomatis. Infection and immunity, 2006. 74(8): p. 4826–4840.
10. Paratsaphan, S., et al., Characterization of a Novel Peptide from Pathogenic Leptospira and Its Cytotoxic Effect. Pathogens (Basel, Switzerland), 2020. 9(11): p. 906.
11. Yang, Z., X. Zeng, and K, Tsui, Investigating function roles of hypothetical proteins encoded by the Mycobacterium tuberculosis H37Rv genome. BMC Genomics, 2019. 20: p.394
12. Shahbaaz, M., M. Hassan, and F, Ahmad, Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. PloS one, 2013. 8(12): P. e84263.
13. Sharma, N., Genomic annotation of the hypothetical proteins of Neisseria Meningitidis K1207. Journal of advance scientific research, 2019. 10 (4): P. 76-82.
14. Tkáčová, Z., et al., Identification of the proteins of Borrelia garinii interacting with human brain microvascular endothelial cells. Ticks and tick-borne diseases, 2020. 11(4): P. 101451.
15. Ali, A., Ahmad, S., et al., Modeling Novel Putative Drugs and Vaccine Candidates against Tick-Borne Pathogens: A Subtractive Proteomics Approach. Veterinary Sciences, 2020.7(3): P. 129.
16. Attwood, TK. And F. Goblet, A Critical Guide to BLAST. F1000Research, 2018. 18(7): p.1435
17. de Oliveira, B., H. Pedrini, and Z. Dias, Ensemble of Template-Free and Template-Based Classifiers for Protein Secondary Structure Prediction. International journal of molecular sciences, 2021. 22(21): p. 11449.
18. Greener, G., M. Kandathil, and D. Jones, Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. Nature communications, 2019. 10(D1): P. 3977.
19. Mistry, J., Chuguransky., et al., Pfam: The protein families database in 2021. Nucleic acids research, 2021.49(1):p.412–419.
20. Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., et al., The InterPro protein families and domains database: 20 years on. Nucleic acids research, 2021. 49(D1): p. 344–354.
21. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. Nucleic acids research, 2019. 47(D1), 330–338.
22. Yang, M., et al., NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. Current protocols in bioinformatics, 2020. 69(1): p. 90.
23. Gasteiger, et al., ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research, 2005. 31(1): p. 3784–3788
24. Saha, S., VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition, Genomics Proteomics Bioinformatics. 2017. 45: p. 42–47.
25. Mitaku, S., T. Hirokawa, and T. Tsuji, Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces, Bioinformatics, 2018. 18(4): p. 608-616
26. Jadhav, A., B. Shanmugham, A. Rajendiran, Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2014. 27: p.300–308.
27. Saha, S. and G. Raghava, VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. Genomics, proteomics & bioinformatics, 2006. 4(1): p. 42–47.

28. Jadhav, A., et al., Clostridium-DT(DB): a comprehensive database for potential drug targets of Clostridium difficile. Computers in biology and medicine, 2013. 43(4): p. 362–367.
29. Brochhausen, M., Schneider, J., Malone, D., Empey, P. E., Hogan, W. R., & Boyce, R. D. (2014). Towards a foundational representation of potential drug-drug interaction knowledge. CEUR workshop proceedings, 1309, 16–31.
30. Shanmugham, B. and A. Pan, Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach. PloS one, 2013. 8(3): p. e59126.
31. Friedrich, A., et al., SM2PH-db: an interactive system for the integrated analysis of phenotypic consequences of missense mutations in proteins involved in human genetic diseases. Human mutation, 2010. 31(2),: p. 127–135.
32. Nielsen, M., Lundegaard, C., Lund, O., & Petersen, T. N. (2010). CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. Nucleic acids research, 38(Web Server issue), W576–W581. https://doi.org/10.1093/nar/gkq535
33. Rahbar, M,. S. Rasooli, M. Gargari, A potential in silico antibody-antigen based diagnostic test for precise identification of Acinetobacter baumannii. Journal of Theoretical Biology, 2012. 294: p. 29–39.
34. Colovos, C. and O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions. Protein science, 1993. 2(9): p.1511-1519.
35. Laskowski, R., PROCHECK: a program to check the stereochemical quality of protein structures. Journal Of Applied Crystallography, 193. 26(2): p. 283-291.
36. Busto, E., V. Gotor-Fernandez, and V. Gotor, ChemInform Abstract: Hydrolases: Catalytically Promiscuous Enzymes for Non-Conventional Reactions in Organic Synthesis. Cheminform, 2011. 42(8): p. 4504-4523
37. Deutsch, C., et al ., Biosynthesis of threonylcarbamoyl adenosine (t6A), a universal tRNA nucleoside. The Journal of biological chemistry, 2012. 287(17):p. 13666–13673.
38. Kihara, A., Y. Akiyama, and K. Ito, Revisiting the lysogenization control of bacteriophage lambda. Identification and characterization of a new host component, HflD. The Journal of biological chemistry, 2001. 276(17): p. 13695–13700.
39. Sato, M., et al., Bacterial alarmone, guanosine 5'-diphosphate 3'-diphosphate (ppGpp), predominantly binds the beta' subunit of plastid-encoded plastid RNA polymerase in chloroplasts. Chembiochem : a European journal of chemical biology, 2009. 10(7): p. 1227–1233.
40. Lu, W., et al., Extraction and Quantitation of Nicotinamide Adenine Dinucleotide Redox Cofactors. Antioxidants & redox signaling, 2018. 28(3): p. 167–179.
41. Macheroux, P., B. Kappes, and S. Ealick, Flavogenomics--a genomic and structural view of flavin-dependent proteins. The FEBS journal, 2011. 278(15): p. 2625–2634.
42. Dawson, A., Structure of diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil reductase from Acinetobacter baumannii. Acta crystallographica. Section F, Structural biology and crystallization communications, 2013. 69(6): P. 611–617.
43. Tong, L., Structure and function of biotin-dependent carboxylases. Cellular and molecular life sciences : CMLS, 2013. 70(5): p. 863–891.
44. Shen, Y., A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product. Molecular and cell, 2004.16(6): p. 881–891.
45. Campbell, J. and J. Cronan, Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. Annual review of microbiology, 2001. 55: p. 305–332.
46. Goenrich, M., et al., SA glutathione-dependent formaldehyde-activating enzyme (Gfa) from Paracoccus denitrificans detected and purified via two-dimensional proton exchange NMR spectroscopy. The Journal of biological chemistry, 2002. 277(5): p. 3069–3072.

47. Martínez, S., et al ., The Classification and Evolution of Enzyme Function. Biophysical journal, 2015. 109(6): p. 1082–1086.
48. Punekar, A. S., et al., Structural and functional insights into the molecular mechanism of rRNA m6A methyltransferase RlmJ. Nucleic acids research, 2013. 41(20): p. 9537–9548.
49. Favrot, L., S. Blanchard, and O. Vergnolle, Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. Biochemistry, 2016. 55(7): p. 989–1002.
50. Wakil, S., K. Stoops, and C. Joshi, Fatty acid synthesis and its regulation. Annual review of biochemistry, 1983. 52: p. 537–579.
51. Gehring, A. M., Acetyltransfer precedes uridylyltransfer in the formation of UDP-N-acetylglucosamine in separable active sites of the bifunctional GlmU protein of Escherichia coli. Biochemistry, 1996. 35(2): p. 579–585.
52. Gehring, A. M., Acetyltransfer precedes uridylyltransfer in the formation of UDP-N-acetylglucosamine in separable active sites of the bifunctional GlmU protein of Escherichia coli. Biochemistry, 1996.35(2): p. 579–585.
53. Sharma, R., Escherichia coli N-Acetylglucosamine-1-Phosphate-Uridyltransferase/Glucosamine-1-Phosphate-Acetyltransferase (GlmU) Inhibitory Activity of Terreic Acid Isolated from Aspergillus terreus. Journal of biomolecular screening, 2016. 21(4): p. 342–353.
54. Pereira, M. P., High-throughput screening identifies novel inhibitors of the acetyltransferase activity of Escherichia coli GlmU. Antimicrobial agents and chemotherapy, 2009. 53(6): p. 2306–2311.
55. Fath, M. J. and R. Kolter, ABC transporters: bacterial exporters. Microbiological reviews, 1993. 57(4): p. 995–1017.
56. Raetz, C. R., and C. Whitfield, Lipopolysaccharide endotoxins. Annual review of biochemistry, 2002. 71: p. 635–700.
57. Simpson, B. W., et al., Identification of Residues in the Lipopolysaccharide ABC Transporter That Coordinate ATPase Activity with Extractor Function. mBio, 2016. 7(5): p. 1716-1726.
58. Moura, E., et al., Thanatin Impairs Lipopolysaccharide Transport Complex Assembly by Targeting LptC–LptA Interaction and Decreasing LptA Stability. Frontiers In Microbiology, 2020. 11(99).
59. Dillon, S. C. and C. Dorman, Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nature reviews. Microbiology, 2010. 8(3): p. 185–
60. Hołówka, J. and J. Zakrzewska-Czerwińska, Nucleoid Associated Proteins: The Small Organizers That Help to Cope With Stress. Frontiers in microbiology, 2020. 11: p. 590.
61. Shiomi, D., M. Sakai, and B. Niki, Determination of bacterial rod shape by a novel cytoskeletal membrane protein. The EMBO journal, 2008. 27(23): p. 3081–3091.
62. Linton, K. J. and C. Higgins, The Escherichia coli ATP-binding cassette (ABC) proteins. Molecular microbiology, 1998. 28(1): p. 5–13.
63. Messner, P., Schäffer, C., & Kosma, P. (2013). Bacterial cell-envelope glycoconjugates. Advances in carbohydrate chemistry and biochemistry, 69, 209–272. https://doi.org/10.1016/B978-0-12-408093-5.00006-X
64. Sperandeo, P., Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. Journal of bacteriology, 2008. 190(13): p. 4460–4469.
65. Gómez-Santos, N., A TonB-dependent transporter is required for secretion of protease PopC across the bacterial outer membrane. Nature Communications, 2019.10(1).
66. Bröer, S. and U. Gether, The solute carrier 6 family of transporters. British journal of pharmacology, 2012.167(2): p. 256–278.
67. Jack, D., N. Yang, and M. Saier, The drug/metabolite transporter superfamily. European journal of biochemistry, 2001. 268(13): p. 3620–3639.

68. Västermark, Å., et al., Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6) before the radiation of Viridiplantae. BMC evolutionary biology, 2011. 11(123).
69. Sperandeo, P., M. Martorana, and A. Polissi, The lipopolysaccharide transport (Lpt) machinery: A nonconventional transporter for lipopolysaccharide assembly at the outer membrane of Gram-negative bacteria. The Journal of biological chemistry, 2017. 292(44): p. 17981–17990.
70. Allen, R. C., Targeting virulence: can we make evolution-proof drugs?. Nature reviews. Microbiology, 2014. 12(4): p. 300–308.
71. Chan, D. and J. Vogel, Current understanding of fatty acid biosynthesis and the acyl carrier protein. The Biochemical journal, 2010. 430(1): p. 1–19.
72. Moynié, L., A Substrate Mimic Allows High-Throughput Assay of the FabA Protein and Consequently the Identification of a Novel Inhibitor of Pseudomonas aeruginosa FabA. Journal of molecular biology, 2016. 428(1): p.108–120.