

# A Systematic *In Silico* Analysis of the *Legionellaceae* Family for Identification of Novel Drug Target Candidates

Seyyed Mohsen Sohrabi,<sup>1</sup> Mohsen Mohammadi,<sup>2</sup> Seyyede Nasim Tabatabaiepour,<sup>3</sup> Seyyede Zahra Tabatabaiepour,<sup>4</sup> Hossein Hosseini-Nave,<sup>5</sup> Mohammad Fazel Soltani,<sup>6</sup> Hosniyeh Alizadeh,<sup>7</sup> and Morteza Hadizadeh<sup>8</sup>

The family *Legionellaceae* consists of Gram-negative bacteria that are widely distributed in aquatic environments around the world. This family consists of a single genus, *Legionella*, that is recognized as an important cause of community-acquired pneumonia and hospital-acquired pneumonia. *Legionella* consists of intracellular pathogens, thus cellular pharmacokinetic and pharmacodynamic properties of an antibiotic against these bacteria as well as uptake and subcellular distribution into macrophages should be considered for a successful outcome of disease. Treatment strategies for *Legionella* infection require a combination of multiple antibiotics. Hence, because of the possible development of resistance to the drugs during therapy, a new alternative targeted therapy is yielding promising results. In this study, a comprehensive *in silico* target identification pipeline was performed on members of the family *Legionellaceae* to identify the best targets. Using a homology-based computational pipeline method, new drug targets were identified. Of 4,358 analyzed proteins, 18 proteins, including proteins involved in metabolism (amino acid, energy, and lipid metabolisms), cellular transport, cell division, and cell motility, were selected as the final putative drug targets. These proteins play an important role in the survival and propagation of *Legionella* infection. In conclusion, homology-based methods could improve the identification of novel drug targets and the drug discovery process, which can potentially be effective for the prevention and treatment of *Legionella* infections.

**Keywords:** homology-based method, *Legionellaceae*, drug targets, *in silico*

## Introduction

THE FAMILY *LEGIONELLACEAE* consists of the single genus *Legionella*. The common characteristic of these bacteria is that they are Gram-negative bacteria, obligate aerobes, and nonspore-forming rod-shaped cells.<sup>1</sup> The number of recognized species and serogroups of the genus *Legionella* continues to increase over recent years. Members of this genus were initially isolated from a wide range of natural and artificial aquatic environments, and also most species are motile and have maximum three polar flagella.<sup>1–3</sup> Several species have been recognized as pathogenic for humans and this can lead to Legionellosis infection varying

in severity from mild febrile disease of influenza-like (Pontiac fever) to a potentially fatal form of pneumonia (Legionnaires' disease).<sup>4,5</sup>

Pneumonia due to *Legionella* infection is an important public health problem that is frequently caused by *Legionella pneumophila* with either community- or hospital-acquired Legionellosis. Therefore, this organism has the potential to cause both outbreaks and sporadic cases.<sup>5,6</sup> The fatality rates associated with Legionnaires' disease among untreated patients are ~15–20%.<sup>7</sup> Owing to the small size of *Legionella* and adaptation to intracellular compartments, an appropriate treatment of the antimicrobial agents is crucial for a successful outcome of Legionnaires' disease; on this basis,

<sup>1</sup>Department of Biotechnology, Lorestan University, Khorramabad, Iran.

<sup>2</sup>Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, Lorestan University of Medical Sciences, Khorramabad, Iran.

<sup>3</sup>Department of Biotechnology, Shahrekord University, Shahrekord, Iran.

<sup>4</sup>Department of Plant Breeding, Faculty of Agriculture, Ilam University, Ilam, Iran.

<sup>5</sup>Department of Microbiology and Virology, School of Medicine, Kerman University of Medical Sciences, Kerman, Iran.

<sup>6</sup>Molecular Genetics and Genetic Engineering, Department of Crop Production and Plant Breeding, School of Agriculture, Razi University, Kermanshah, Iran.

<sup>7</sup>Endocrinology and Metabolism Research Center, Institute of Basic and Clinical Physiology Sciences, Kerman University of Medical Sciences, Kerman, Iran.

<sup>8</sup>Physiology Research Center, Institute of Basic and Clinical Physiology Sciences, Kerman University of Medical Sciences, Kerman, Iran.

intracellular penetration, accumulation, and distribution are important parameters to the activity of antibiotics against intracellular bacteria.<sup>8</sup> Thus, antimicrobial agents with adequate intracellular penetration into alveolar macrophages as well as subcellular compartments are more effective than antibiotics with poor intracellular penetration.

There are several major classes of antibiotics, including macrolides, fluoroquinolones, tetracyclines, rifamycins, and ketolides, which are used for the treatment of *Legionella* infections.<sup>5,7</sup> Because of the ability of *Legionella* spp. to survive and multiply in human macrophages, consequently, like other intracellular pathogens, antibiotics that penetrate cell membrane are drugs of choice. Aminoglycosides in general are too water soluble and penetration into biologic membranes is poor; therefore, they are taken up slowly by endocytosis, which results in an exclusively lysosomal localization. Whereas both macrolides and fluoroquinolones accumulate in phagocytes. In addition, some drugs are quickly cleared from phagocytes, consequently their concentration and intracellular distribution vary because of differences in their intracellular pH gradients of the cytosol and subcellular compartments.<sup>7,8</sup>

On the contrary, treatment of *Legionella* infection requires a combination of multiple antibiotics, thus this may pose a concern on the development of drug resistance during therapy. In recent years, several cases of drug resistance to routinely prescribed antimicrobials in *Legionella* isolates have been reported, among which *L. pneumophila* strains were more resistant than other *Legionella* species.<sup>7,9-14</sup> It is very important to note that hidden resistance cases have also been reported and have been suggested to be correlated with treatment failure and poor prognosis in Legionnaires patients.<sup>15</sup>

The current strategies to identify drug targets are relatively difficult, time consuming, and expensive, often resulting in very finite drug targets.<sup>16-18</sup> Furthermore, the stages of drug discovery and development using traditional methods are a linear multistep process and yield few drug targets with little feedback from intracellular information for guiding target selection. Therefore, with the arrival of genome research during the past decade, focus on drug development has shifted to computational comparative genomics to identify novel drug targets.<sup>16,19</sup>

*In silico* drug target identification, which includes a series of algorithms to identify genes and proteins, is applied to the discovery and development of potential therapeutic targets. These approaches increase the efficiency of therapeutic targets, reduce timelines and costs of the drug discovery process, and also provide a closer look of the entire microorganism. Speed, low cost, and, even more importantly, providing a systematic view of the whole microorganism are the advantages of *in silico* techniques for identification and validation of drug targets. These features provide a great chance for asking questions that are often difficult to address experimentally. There has been an increase in the use of *in silico* identification methods to find drug targets in several studies.<sup>17,18,20,21</sup>

With the help of genome-sequencing projects of pathogens and humans, microbial drug target identification has been revolutionized in recent years. Among genomic strategies, subtractive genomics has been successfully used in the identification of microbial drug targets.<sup>22</sup> Various methods can be used to discover a proper therapeutic target such as comparative genomics models,<sup>23</sup> structure and sequence to func-

tion,<sup>24,25</sup> metabolic pathways,<sup>26</sup> and data mining.<sup>27</sup> The most used methods to identify new therapeutic targets are based on the sequence similarity using the Basic Local Alignment Search Tool (BLAST) algorithm.<sup>28</sup> In this study, a similarity-based method was used for genome-wide prediction of the drug targets in major pathogens of the family *Legionellaceae*, considering the *L. pneumophila* str. Paris as an important pathogen.

## Materials and Methods

### Step I

The complete proteome of *L. pneumophila* str. Paris that comprises 3,012 chromosomal proteins and 134 plasmid proteins was downloaded from the National Center for Biotechnology Information (NCBI) genome database. Virulence Factor Database (VFDB) is a practical database of virulence factors (VFs) for bacterial pathogens and also *Chlamydia* and *Mycoplasma*. All the major VFs of *L. pneumophila* str. Paris (127 proteins) were retrieved from VFDB. All literature reported that *L. pneumophila* str. Paris resistance proteins were downloaded from NCBI.<sup>21,29</sup> Using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) 10.0 tool and seven active prediction methods (neighborhood, gene fusion, co-occurrence, co-expression, experiments, databases, and text mining), interaction partners of *L. pneumophila* str. Paris resistance proteins with high confidence value (0.7) were predicted. The protein sequences of resistance genes and their partners were retrieved from NCBI and STRING database. Finally, a comparative metabolic pathway was performed to analyze between *L. pneumophila* str. Paris and human as the host. All the proteins in the different pathways of *L. pneumophila* str. Paris and unique proteins of *L. pneumophila* str. Paris were collected in common pathways (between *L. pneumophila* str. Paris and human). All proteins in the different pathways of human and unique proteins of human in common pathways were excluded. The amino acid sequences of the selected proteins were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and NCBI. Total proteins of step I (Greenlist) were transferred to step II.

### Step II

**Host homology search.** The main goal of this step is to find proteins in *L. pneumophila* str. Paris that have no similarity to the human proteome. This analysis minimizes the undesirable side effects of the drug.<sup>21,30</sup> The Greenlist proteins were subjected to BLASTp against nonredundant protein sequences of human (taxid: 9606) with an e-value of 0.005.<sup>21,30,31</sup> The Greenlist proteins that showed no hits for the 0.005 e-value were transferred to the next test (the test of essentiality).

**Test of essentiality.** The proteins in Greenlist were examined to recognize essential proteins using BLASTp search against DEG, a database of essential genes. The DEG BLASTp expected value cutoff was adjusted to  $10^{-5}$ . The hits with less than e-value of  $10^{-5}$ , identity  $\geq 25\%$ , and same annotated function of the query were selected as essential proteins.<sup>21,30,32</sup> The proteins in Greenlist that revealed hits with DEG BLASTp parameters were collected and transferred to the next step (gastrointestinal flora homology test).

The other proteins that revealed no hit in DEG BLASTp were excluded from the analysis.

Homology search for gastrointestinal flora. Unintentional targeting of vital molecules in the gut microorganisms may cause harmful effects on the host. To prevent this condition, a custom database was created from gut microflora reported in the literature and the proteins of Greenlist were subjected to BLASTp search against this custom database with an expected threshold of  $10^{-4}$ .<sup>17,20,21,32</sup> The Greenlist proteins that showed >10 hits were excluded and the rest were selected as Whitelist. The Whitelist comprises the primary potential drug targets of *L. pneumophila* str. Paris.

### Step III

Whole proteome comparisons. To find similarities, the proteome of important pathogens of the family *Legionellaceae* was compared. All proteomes belonging to these pathogens were compared in TaxPlot with a cutoff value of 200.

Homology search for *Legionellaceae*. To find homologues in important pathogens of the family *Legionellaceae*, primary potential drug targets of *L. pneumophila* str. Paris in Whitelist were subjected to BLASTp against a custom database. The BLASTp was done with an expected value of  $10^{-4}$ . Proteins with identity  $\geq 50\%$  were selected as final potential drug targets in the family *Legionellaceae* (Redlist) and were qualitatively characterized using various tools in step IV.

### Step IV

The biological function of an unknown protein in the Redlist (hypothetical proteins) was determined by means of Pfam tool. Subcellular localization sites of some proteins were recognized in the literature, and web-based tools, such as Prediction of Subcellular Localization of bacterial proteins. (PSLpred), Prediction of Protein SORTing (PSORT), and subcellular localization predictor (CELLO), were used to predict possible subcellular localization sites of Redlist proteins. PSLpred is a web server that uses Support Vector Machines (SVMs) to predict protein subcellular localization.<sup>33</sup> PSORT is a computer program for the prediction of protein localization sites in cells.<sup>34</sup> CELLO is a web server for protein subcellular localization prediction and functional gene ontology analysis.<sup>35</sup> In the broad-spectrum search, a custom database of bacterial pathogens reported in the literature was created.<sup>20</sup> Proteins in Redlist were evaluated using BLASTp search against a custom database with an e-value of 0.005 for the identification of broad-spectrum targets. In the DrugBank database,<sup>36</sup> Redlist proteins were subjected to BLASTp search with an e-value of  $10^{-5}$ . The presence of molecule with the same biological function in BLAST results shows the druggability of the target molecules, and absence shows the novelty of the target molecules.<sup>21,32</sup>

## Results

In this study, a subtractive homology-based method was used to identify a collection of potential therapeutic targets in some bacterial pathogens of the family *Legionellaceae*. A systematic pipeline that involved several computational tools and databases was developed (Fig. 1). Only these

proteins were selected as acceptable drug targets that successfully passed the pipeline.

### Step I

In the complete proteome (chromosomal and plasmid) of *L. pneumophila* str. Paris, 1,589 proteins were produced by positive strand and 1,557 proteins were translated from negative strand. Distributions of the protein lengths in the *L. pneumophila* str. Paris proteome were in the range of 44–7,679 amino acids. All 3,146 *L. pneumophila* str. Paris proteins were transferred to the next step. The major VFs in *L. pneumophila* str. Paris have been divided into the following seven groups: adherence, intracellular survival, invasion, iron acquisition, regulation, secretion systems, and toxins. All VFs (127 proteins) were transferred to step II. Inhibition of VFs proteins could reduce the virulence of pathogenic bacteria. Nine proteins that confer resistance to antimicrobial drugs in *L. pneumophila* str. Paris were obtained from the literature and corresponding sequences were retrieved from NCBI. Using STRING 10.0 tool and seven active prediction methods (neighborhood, gene fusion, co-occurrence, co-expression, experiments, databases, and text mining) with a high confidence value (0.7), 342 interaction partners for resistance-causing proteins were predicted. Antimicrobial resistance proteins and their interaction partners are considered as potential drug targets and suppression of these proteins may block the antimicrobial resistance process.<sup>20</sup> All the antimicrobial resistance proteins and their interaction partners (351 proteins) were transferred to the next step. In the KEGG database, there are 299 metabolic pathways for human and 109 metabolic pathways for *L. pneumophila* str. Paris. Pathways of *L. pneumophila* str. Paris are classified as metabolism, genetic data processing, environmental data processing, and cellular processes. Of 109 metabolic pathways in *L. pneumophila* str. Paris, 30 pathways were unique and 79 pathways were common with humans. Sequences of 531 unique proteins from common pathways and 203 proteins from distinct pathways were selected and transferred to step II. Finally, 4,358 proteins, including chromosomal and plasmid proteome, VFs, resistance proteins and their interaction partners, and common and unique metabolic pathway proteins were collected as Greenlist and transferred to step II.

### Step II

A collection of 4,358 proteins that resulted from step I (Greenlist) were subjected to three sequential BLAST searches to find primary targets (Fig. 1). The first and third BLAST searches determined the proteins that were nonhomologous to human and gut microflora proteome. The essentiality of the proteins was assessed by second BLAST search against a known set of essential proteins in DEG database. Targeting of the proteins homologous to host by drug compounds could adversely alter the host metabolism. Therefore, filtering proteins homologous to human proteome is considered in several computational drug target discovery methods.<sup>17,21,30,32</sup> A total of 4,358 *L. pneumophila* str. Paris proteins, including chromosomal and plasmid proteome (3,146 proteins), pathogenicity islands and VFs (127 proteins), resistance proteins and their interaction partners (351 proteins), and common and unique metabolic pathway proteins (734 proteins), were subjected to host homology search

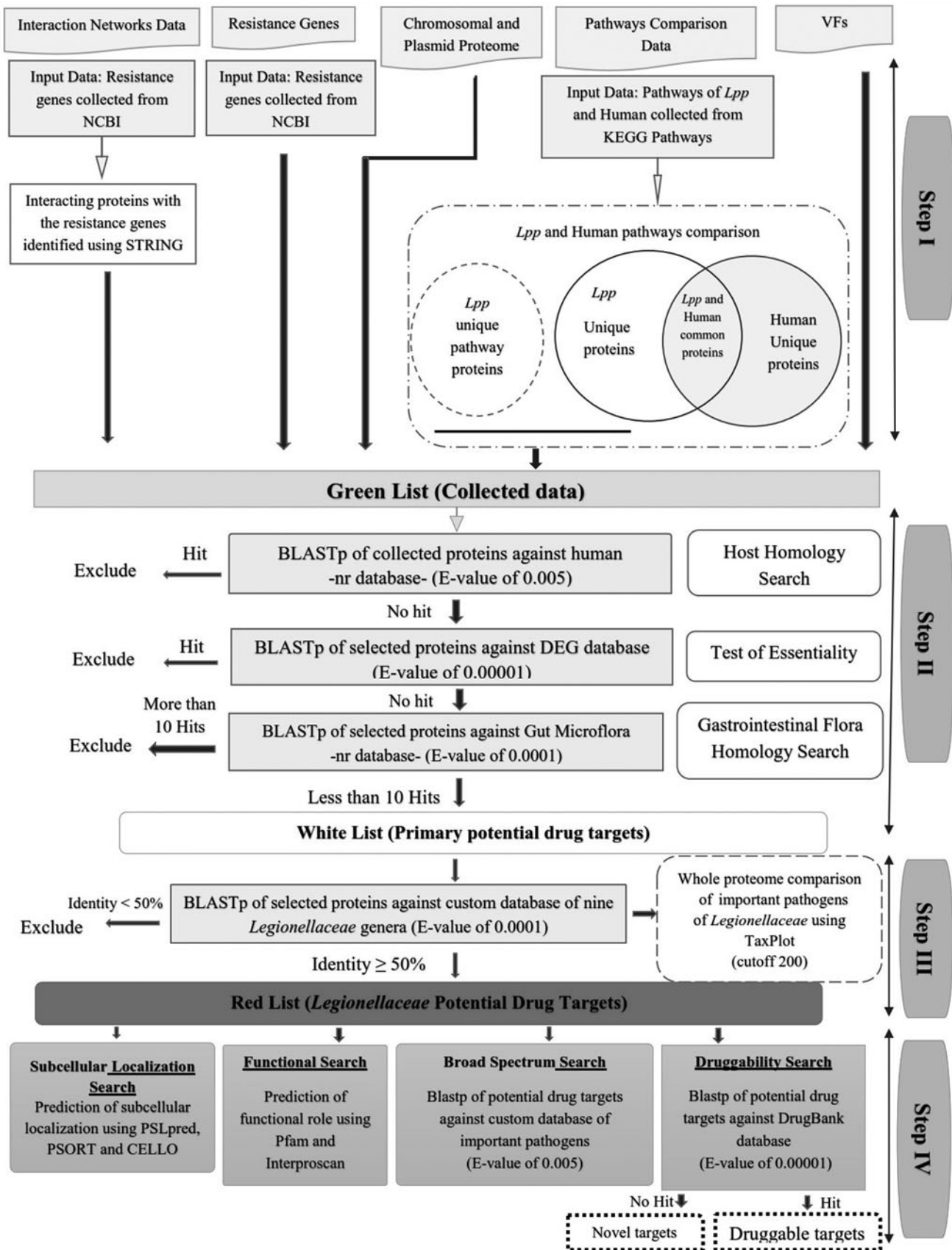


FIG. 1. The complete flowchart of the homology-based method.

against the whole proteome of the human. Proteins without any hits in BLAST search were considered as nonhomologous, whereas those showing hits were regarded as homologous. Of the 4,358 input proteins, 802 proteins that showed no hits against human proteome were selected and 3,556 proteins (homologous to human proteome) were excluded from the next step of the analysis. Essential genes are the genes of an organism that are vital for its survival. To preserve the vital process of an organism such as central metabolism, DNA replication, translation, cellular structure maintenance, and cellular transport, essential genes produce proteins. A drug target must be an essential protein for the survival of pathogenic microorganisms. Targeting of such proteins could cause growth inhibition and disrupt biological functions of microorganism. The Greenlist proteins were analyzed for determination of essentiality using DEG 11.3 database. Of 802 input proteins, 464 proteins were identified to be essential for the pathogen and selected for the next analyses. In addition, 338 proteins that showed no hits against DEG (nonessential) were excluded from the analysis. The gut microbiota plays an essential role in human metabolism, physiology, and nutrition, and it also plays an essential role in human health by preventing pathogens from colonization. Therefore, the decay of the gut microbiota population may cause problems in human health.<sup>37,38</sup> The essential resulting proteins from previous analyses were subjected to BLASTp against a custom database of the whole proteome of gut microbiota. Of 464 input proteins, 60 proteins were collected as Whitelist and 404 proteins showing >10 hits were excluded. Whitelist proteins were selected as primary potential drug target of *L. pneumophila* str. Paris and were used for the next step of the analysis.

### Step III

The reference proteome of *L. pneumophila* str. Paris was compared with the family *Legionellaceae* pathogens to find similarities. The Whitelist was subjected to BLASTp against a custom database of the family *Legionellaceae* pathogens from 60 proteins in the Whitelist. Eighteen proteins that showed identity  $\geq 50\%$  were selected as Redlist (Table 1). Redlist proteins were taken as the final potential drug targets in the family *Legionellaceae* (Table 2). The Redlist consists of proteins involved in metabolism, cellular transport, cell division, and cell motility. Proteins involved in metabolism and cellular transport provide a major contribution to the Redlist of final targets (13 of 18 proteins). Targets involved in multiple metabolic pathways are thought to be a more efficient drug target, and preventing the activity of such targets could increase lethal effects by blocking the activity of several metabolic pathways of the microorganism. Of the

18 Redlist targets, 8 targets (lpp0252, lpp0381, lpp1542, lpp1671, lpp1673, lpp2256, lpp2454, and lpp2758) were involved in >1 metabolic pathway (Table 2).

### Step IV

In this step, Redlist proteins were further explored using various analyses (functional search, subcellular localization search, broad-spectrum search, and druggability search) (Fig. 1). The biological function of unknown and hypothetical proteins in Redlist was determined by using Pfam tool. Eight proteins in Redlist, namely, lpp0375, lpp1542, lpp1720, lpp1729, lpp1975, lpp2256, lpp2683, and lpp2758 were unknown and hypothetical (Table 2). Hypothetical protein lpp0375 was determined to be an OstA-like protein. The OstA protein is known as an organic solvent tolerance protein.<sup>39</sup> Hypothetical protein lpp1542 was predicted to be a CDP-alcohol phosphatidyl transferase (CDP-AP). Phospholipids have key roles in the structure and function of all cell membranes. Members of CDP-alcohol phosphotransferase family are integral membrane proteins which are all involved in phospholipid biosynthesis. They catalyze the displacement of cytidine monophosphate (CMP) from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoridic anhydride bond.<sup>40</sup> Hypothetical protein lpp1720 was determined to be a flagellar FliJ protein, this family is found in bacterial motility proteins in flagellar system. Hypothetical protein lpp1729 was predicted to be an outer membrane (OM) lipoprotein carrier protein LolA that plays in the translocation of lipoproteins from the inner membrane (IM) to the OM.<sup>41</sup> Hypothetical protein lpp1975 was determined to be membrane-bound lytic transglycosylase (MltA)-specific insert domain. This beta barrel domain is found inserted into the MltA, a murein-degrading enzyme. This domain that contains three conserved aspartate residues may be involved in peptidoglycan binding.<sup>42</sup> Hypothetical protein lpp2256 was predicted to be a preprotein translocase subunit SecB. SecB is a cytoplasmic module common with membrane-bound multisubunit enzyme termed Sec protein translocase, which is the main component of the general secretory pathway (type II), involved in translocation of aborting polypeptides across the cytoplasmic membrane.<sup>43</sup> Hypothetical protein lpp2683 was determined to permease the YjgP/YjgQ family. Members of this family are integral membrane proteins of unknown function. Finally, hypothetical protein lpp2758 was predicted to be the cytochrome C1 family, which is the third complex in the electron transport chain. To identify cellular localization, Redlist proteins were analyzed using PSORT, PSLpred, and CELLO tools. In PSORTb and PSLpred, based on the localization score, reliability index, and expected accuracy, five targets were predicted to be cytoplasmic proteins, five targets were predicted to be cytoplasmic membrane proteins, two targets were predicted to be periplasmic, and two targets were predicted to be the OM protein. The cellular localization of four targets (lpp0252, lpp0922, lpp1673, and lpp2664) was predicted to be unknown. Furthermore, to obtain greater accuracy, the predicted subcellular localization of the targets was cross-evaluated using the CELLO tool. In CELLO, the location of six targets was determined as cytoplasmic, the location of two targets was determined

TABLE 1. NUMBER OF PROTEINS USED IN THE PERFORMED ANALYSES

Analysis	Input proteins	Output proteins	Failed proteins
Host homology search	4,358	802	3,556
Test of essentiality	802	464	338
Gastrointestinal flora homology search	464	60	404
<i>Legionellaceae</i> homology search	60	18	42

TABLE 2. LIST OF PUTATIVE DRUG TARGETS IN THE *LEGIONELLACEAE* FAMILY

No.	Description	NCBI accession	KEGG ID	KEGG pathway ID	Length, aa
1	Catalase peroxidase	CAH11399	lpp0252	lpp01100 lpp01110 lpp00360 lpp00380	749
2	Hypothetical protein lpp0375	CAH11523	lpp0375	No hits!	839
3	Preprotein translocase secE	CAH11529	lpp0381 lpp0381	lpp0381	123
4	Heme exporter protein CcmC	CAH12071	lpp0920	lpp02010	251
5	Cytochrome c-type biogenesis protein CcmE	CAH12073	lpp0922	No hits!	143
6	Hypothetical protein lpp1542	CAH12693	lpp1542	lpp00564 lpp01110	255
7	Arginine <i>N</i> -succinyltransferase, beta chain	CAH12823	lpp1671	lpp00330 lpp01100	346
8	Succinylarginine dihydrolase	CAH12825	lpp1673	lpp00330 lpp01100	448
9	Septum formation topological specificity factor	CAH12842	lpp1690	No hits!	89
10	Hypothetical protein lpp1720	CAH12872	lpp1720	lpp02040	151
11	Hypothetical protein lpp1729	CAH12881	lpp1729	No hits!	202
12	Hypothetical protein lpp1975	CAH13127	lpp1975	No hits!	398
13	Hypothetical protein lpp2256	CAH13409	lpp2256	lpp03060 lpp03070	162
14	Catalase-peroxidase KatB	CAH13607	lpp2454	lpp01100 lpp01110 lpp00360 lpp00380	721
15	Cell division protein FtsQ	CAH13817	lpp2664	No hits!	239
16	Hypothetical protein lpp2683	CAH13836	lpp2683	lpp02010	359
17	Hypothetical protein lpp2758	CAH13911	lpp2758	lpp01100 lpp00190 lpp02020	248
18	Putative TatA protein (twin arginine translocation)	CAH14126	lpp2973	lpp03060 lpp03070	61

aa, amino acid; NCBI, National Center for Biotechnology Information; KEGG, Kyoto Encyclopedia of Genes and Genomes.

as periplasmic, the location of nine targets was determined as IM, and the location of one target was determined as OM. Of the four targets characterized as unknown by PSORTb and PSLpred, one target was predicted as cytoplasmic, one target was predicted as periplasmic, and two targets were predicted as IM (Table 3). Six proteins with cytoplasmic localization and nine proteins with IM localization and one protein with OM localization could, respectively, possibly serve as drug and vaccine targets (Table 3). In a broad-spectrum search, BLASTp against a custom database of the whole proteome of bacterial pathogens resulted in the recognition of broad-spectrum targets. Results of the broad-spectrum search showed that 7, 8, and 3 targets were present in >100 pathogens, between 50 and 100 pathogens, and <50 pathogens, respectively. Seven proteins of Redlist that were present in >100 pathogens were considered as broad-spectrum targets (Table 3). Drug targets involved in vital metabolic process seem to be a broader spectrum than the other targets such as VFs. VFs are not often broad-spectrum drug targets because of high specificity. Targeting of broad-spectrum proteins by drug molecules may facilitate the destruction of a wide range of pathogenic bacteria. One target that is present in <25 pathogens was considered as the family *Legionellaceae*-specific targets. Such specific target proteins may decrease the threat of antimicrobial resistance development in a wide range of pathogenic bacteria.<sup>20</sup>

Druggability search was the final analysis in this approach. Druggability of Redlist targets was assessed using BLASTp against DrugBank database. Similarity search showed that four targets (lpp0252, lpp1671, lpp1673, and lpp2454) are homologous to each other or more known targets in DrugBank database with an expected value of  $10^{-5}$ . BLASTp against DrugBank also showed that the lpp0252 target has three experimental inhibitors [2-amino-3-(1-hydroperoxy-1 h-indol-3-yl) propan-1-ol, ethionamide, and isoniazid], lpp1671 has one experimental inhibitor (formic acid), lpp1673 has two experimental inhibitors (N ~ 2 ~ -succinylarginine and N ~ 2 ~ -succinylornithine), and lpp2454 has three experimental inhibitors [2-amino-3-(1-hydroperoxy-1 h-indol-3-yl) propan-1-ol, ethionamide, and isoniazid]. All these four druggable targets are involved in vital pathways of amino acid metabolism of *L. pneumophila* str. Paris. Fourteen targets, showing no homology with DrugBank database, were considered as novel drug targets, which should be further evaluated experimentally (Table 3).

## Discussion

Recent advances in whole genome sequencing and computational methods have facilitated the discovery of novel antimicrobial agents and could enable us to combat antibiotic resistance in bacteria. In addition, they have provided

TABLE 3. QUALITATIVE CHARACTERIZATION OF PUTATIVE DRUG TARGETS

No.	NCBI ID	KEGG ID	Function prediction	Subcellular localization	Broad-spectrum condition	Druggability
1	CAH11399	lpp0252	Not required	Periplasmic	Yes	Druggable
2	CAH11523	lpp0375	OstA-like protein/ organic solvent tolerance protein	Outer membrane	Yes	Novel
3	CAH11529	lpp0381	Not required	Inner membrane	No	Novel
4	CAH12071	lpp0920	Not required	Inner membrane	No	Novel
5	CAH12073	lpp0922	Not required	Inner membrane	No	Novel
6	CAH12693	lpp1542	CDP-alcohol phosphatidyltransferase	Inner membrane	No	Novel
7	CAH12823	lpp1671	Not required	Cytoplasmic	Yes	Druggable
8	CAH12825	lpp1673	Not required	Cytoplasmic	No	Druggable
9	CAH12842	lpp1690	Not required	Cytoplasmic	No	Novel
10	CAH12872	lpp1720	Flagellar FliJ protein	Cytoplasmic	No	Novel
11	CAH12881	lpp1729	Outer membrane lipoprotein carrier protein LolA	Cytoplasmic	Yes	Novel
12	CAH13127	lpp1975	MltA-specific insert domain/ 3D domain	Inner membrane	Yes	Novel
13	CAH13409	lpp2256	Preprotein translocase subunit SecB	Cytoplasmic	No	Novel
14	CAH13607	lpp2454	Not required	Periplasmic	Yes	Druggable
15	CAH13817	lpp2664	Not required	Inner membrane	No	Novel
16	CAH13836	lpp2683	Predicted permease YjgP/ YjgQ family	Inner membrane	Yes	Novel
17	CAH13911	lpp2758	Cytochrome C1 family	Inner membrane	No	Novel
18	CAH14126	lpp2973	Not required	Inner membrane	No	Novel

MltA, membrane-bound lytic transglycosylase; CDP, cytidine diphosphate.

important insights into the molecular mechanisms of underlying diseases. The development of new bioinformatic approaches and modulation and optimization of existing strategies can increase the accuracy of drug target identification.<sup>44</sup> This study has shown how bioinformatics and *in silico* approaches can be used in the drug discovery processes. As biological systems are complex networks of many metabolic and nonmetabolic pathways, the explanation of such systems may be obtained by considering large-scale studies. Focusing on a single drug target at a time in conventional drug discovery methods may not always produce acceptable or expected results. Comprehensive study of biological pathways and whole cell systems provides wider insights into the fitness of a potential drug target. Some proteins in an organism that may be predicted as good drug targets, when viewed in the context of a biological system, may not actually be vital. Analyzing drug targets in the context of a biological system can help in evaluating the criticality of the individual targets in the cell pathways. Of 4,358 proteins analyzed, 18 (Redlist) were selected as final putative drug targets in the family *Legionellaceae* (Table 2). Putative targets were filtered to exclude nonviable candidates based mostly on importance of survival, lack of homology to the human host, well-known biological function, and conserved between *Legionellaceae* species. The Redlist consists of proteins involved in metabolism (amino acid, energy, and lipid metabolisms), cellular transport, cell division, and cell motility. These essential genes that play a great role in cell survival encode the proteins to control a central metabolism, replicate DNA, translate genes into proteins, maintain a basic cellular structure, and mediate transport processes within or out of the cell.<sup>45</sup> The essential proteins are involved in many biological pathways that

represent molecular interaction networks between the pathogen and the host. Essentiality search shows some vital proteins that are required by the pathogen to perform important roles for their survival, growth, and replication. Therefore, a broader search of bioinformatics is a critical stage to predict potential drug target in the pathogens. Proteins involved in metabolism and cellular transport provide a major contribution to the Redlist of final targets (14 of 18 proteins). Of the 18 Redlist protein targets, 8 targets (lpp0252, lpp1542, lpp1671, lpp1673, lpp2256, lpp2454, lpp2758, and lpp2973) are involved in >1 metabolic pathway (Table 2). Restricted targeting of some specific pathways may cause development of multidrug resistance among pathogenic bacteria.<sup>21</sup> In general, approaches that consider all essential pathways of organisms can be more successful to identify efficient drug targets. Targets involved in multiple metabolic pathways are thought to be a more efficient drug target, and preventing the activity of such targets could increase lethal effects by blocking the activity of several metabolic pathways of the microorganism. Targeting of metabolic enzymes for cancer therapy is a hot topic for drug discovery.<sup>46</sup> Molecular structures of membrane transporter proteins are significant for drug discovery. These proteins are important coplayers in cellular systems and are known molecular components of many disease processes. The membrane transporter proteins are targeted by many currently used drugs and have a major potential as targets for new drug development. A large number (60–70%) of the current known drug targets are proteins set in a cellular membrane, and membrane proteins are among the most interesting macromolecules to study using structural biology techniques. High-resolution structural knowledge about proteins placed in a cellular membrane is

not only of pivotal importance for developing new drugs with therapeutic potential, but is also necessary for the knowledge of the molecular mechanisms of cellular signaling and function.<sup>47</sup> The lpp2758 target is involved in the bacterial two-component system. By using the two-component system, the pathogens perceive the changes in the environment and respond to it. Moreover, involvement in the pathogenicity of the organisms and the absence of these systems in humans make them attractive drug targets.<sup>48</sup> Therefore, inhibition of these proteins can reduce the growth rate and virulence of the pathogen. Targeting of proteins involved in DNA replication, DNA repair, DNA recombination, and cell division (lpp2664 and lpp1690) may disrupt the pathways essential for pathogen survival, growth, and reproduction. Seven broad-spectrum targets are involved in crucial processes such as cellular transport, environmental information processing, and metabolism (Tables 2 and 3). Drug targets involved in the vital metabolic process seem to have a broader spectrum than the other targets such as VFs. Because of high specificity, VFs are not often broad-spectrum drug targets. Targeting of broad-spectrum proteins by drug molecules may facilitate the destruction of a wide range of pathogenic bacteria. Eleven *Legionellaceae*-specific targets are suitable for the development of narrow-spectrum antibiotic. Such specific target proteins may decrease the threat of development of antimicrobial resistance in a wide range of pathogenic bacteria. Based on the localization analysis, most targeting proteins were located in the membrane (Table 3). Inhibition of target proteins located on the membrane and extracellular proteins is important because of their crucial role as VFs assisting pathogens to spread and proliferate within the host. Knowledge of protein localization is valuable for learning the function and the interaction of different proteins. When other information is not available, the subcellular localization will also be effective in the annotation for new proteins. In drug discovery process, subcellular localization data can help understand therapeutic intervention points. For example, because of their localization, secreted proteins and membrane proteins are easily available by drug molecules.<sup>49</sup> The Gram-negative bacteria are surrounded by two membranous structures, the IM and the OM. The IM that is called the plasma membrane has a trilamellar structure surrounded with bacterial protoplasm and composed of a phospholipids bilayer. Most of the membrane proteins that have role in energy production, lipid biosynthesis, protein secretion, and transport system are conserved in bacteria, but their cellular location is different. In bacteria, these proteins are located in the IM.<sup>50,51</sup> The OM also presents a trilamellar structure (with couple electron dense leaflets, outer and inner) in the electron micrograph and consists of proteins, containing porins, receptors, and an asymmetric distribution of lipids.<sup>50</sup> The OM of Gram-negative bacteria provides a difficult barrier that must be overcome. There are essentially two pathways that antibiotics can take within the OM: a lipid-mediated pathway for hydrophobic antibiotics and general dissemination porins for hydrophilic antibiotics, which consequently may lead to antibiotic resistance.<sup>52</sup> Druggability refers to the ability of a target molecule to bind with high affinity to the drug molecules. Druggability is one of the most important characteristics of a target molecule. Four druggable targets (lpp0252, lpp1671, lpp1673, and lpp2454) are involved in vital process such as oxidant-antioxidant system, metabolism, and bio-

synthesis of macromolecule. Either of these proteins is a target for conventional antibiotics. Fourteen novel drug targets are suitable for development of new antimicrobials and should be further evaluated experimentally. Currently, several computational methods such as comparative genomics, data mining, structure and sequence to function, and metabolic pathways are used for the identification and characterization of the drug targets. Most of these target discovery methods enable us to identify potential drug target candidates based on the main criteria of specificity and essentiality. Besides these properties, a suitable drug target must be specific to the pathogen, avoid harmful side effect, and should be a very important protein for survival of the pathogen. Suppression of such drug targets can result in effective control of the pathogen without any harmful effects on the host. Our homology-based method considers the essentiality, specificity, druggability, subcellular localization, function, and broad-spectrum condition of drug targets. Although a sequence similarity of the protein does not ensure the same structures or binding properties, using such homology-based methods could ease the optimization and production of new drugs and vaccines. In conclusion, the homology-based method<sup>17,20,21,32</sup> was applied to identify novel putative drug targets in the family *Legionellaceae*. The results of this study identified several proteins in the genomes of the family *Legionellaceae* pathogens that can be targeted for effective drug design and development. Thus, many of these putative drug targets that are involved in several vital metabolic pathways, such as energy metabolism, amino acid metabolism, and lipid metabolism, designing drug molecules against these targets could be very effective for the treatment of *Legionella* infections. Development of new drug against such targets will be precise to the pathogen and considerably decrease the harmful side effects to the host. The efficiency of already available antimicrobial drugs can be tested using this method. Targeting of proteins involving several essential metabolic pathways may facilitate the efficient treatment of infections. The findings of such studies facilitate the design and development of novel antimicrobial drugs against the family *Legionellaceae* and other pathogens.

### Acknowledgments

The authors would like to thank all staff of the Bioinformatics Lab, School of Medicine, Kerman University of Medical Sciences, for their kind cooperation.

### Disclosure Statement

No competing financial interests exist.

### References

1. Palusińska-Szys, M., and M. Cendrowska-Pinkosz. 2009. Pathogenicity of the family *Legionellaceae*. Arch. Immunol. Ther. Ex. 57:279–290.
2. Gomez-Valero, L., C. Rusniok, and C. Buchrieser. 2009. *Legionella pneumophila*: population genetics, phylogeny and genomics. Infect. Genet. Evol. 9:727–739.
3. Zhan, X.Y., Hu, C.H., and Q.Y. Zhu. 2010. Research advances of *Legionella* and Legionnaires' disease. Front. Med. China. 4:166–176.



4. Bartram, J. 2007. *Legionella* and the Prevention of Legionellosis. World Health Organization.
5. Diederer, B. 2008. *Legionella spp.* and Legionnaires' disease. *J. Infect.* 56:1–12.
6. Cordevant, C., J.S. Tang, D. Cleland, and M. Lange. 2003. Characterization of members of the *Legionellaceae* family by automated ribotyping. *J. Clin. Microbiol.* 41:34–43.
7. Edelstein, P.H. 1995. Antimicrobial chemotherapy for legionnaires' disease: a review. *Clin. Infect. Dis.* 21:S265–S276.
8. Tulkens, P.M. 1991. Intracellular distribution and activity of antibiotics. *Eur. J. Clin. Microbiol. Infect. Dis.* 10:100–106.
9. Barker, J., H. Scaife, and M. Brown. 1995. Intrapathogenic growth induces an antibiotic-resistant phenotype of *Legionella pneumophila*. *Antimicrob. Agents Chemother.* 39:2684–2688.
10. Nielsen, K., P. Hindersson, N. Høiby, and J.M. Bangsborg. 2000. Sequencing of the *rpoB* gene in *Legionella pneumophila* and characterization of mutations associated with rifampin resistance in the *Legionellaceae*. *Antimicrob. Agents Chemother.* 44:2679–2683.
11. Nielsen, K., J.M. Bangsborg, and N. Hoiby. 2000. Susceptibility of *Legionella* species to five antibiotics and development of resistance by exposure to erythromycin, ciprofloxacin, and rifampicin. *Diagn. Microbiol. Infect. Dis.* 36:43–48.
12. De Giglio, O., C. Napoli, G. Lovero, G. Diella, S. Rutigliano, G. Caggiano, and M.T. Montagna. 2015. Antibiotic susceptibility of *Legionella pneumophila* strains isolated from hospital water systems in Southern Italy. *Environ. Res.* 142:586–590.
13. Bruin, J.P., B.M. Diederer, E.P. Ijzerman, J.W. Den Boer, and J.W. Mouton. 2013. Correlation of MIC value and disk inhibition zone diameters in clinical *Legionella pneumophila* serogroup 1 isolates. *Diagn. Microbiol. Infect. Dis.* 76:339–342.
14. Mallegol, J., P. Fernandes, R.G. Melano, and C. Guyard. 2014. Antimicrobial activity of solithromycin against clinical isolates of *Legionella pneumophila* serogroup 1. *Antimicrob. Agents Chemother.* 58:909–915.
15. Shadoud, L., I. Almahmoud, S. Jarraud, J. Etienne, S. Larrat, C. Schwebel, J.F. Timsit, D. Schneider, and M. Maurin. 2015. Hidden selection of bacterial resistance to fluorquinolones in vivo: the case of *Legionella pneumophila* and humans. *EBioMedicine* 2:1179–1185.
16. Ginsburg, G.S., and J.J. McCarthy. 2001. Personalized medicine: revolutionizing drug discovery and patient care. *Trends Biotechnol.* 19:491–496.
17. Hadizadeh, M., S.N. Tabatabaiepour, S.Z. Tabatabaiepour, Hosseini Nave, H., M. Mohammadi, and S.M. Sohrabi. 2018. Genome-wide identification of potential drug target in *Enterobacteriaceae* family: a homology-based method. *Microb. Drug Resist.* 24:8–17.
18. Lewis, K. 2013. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* 12:371–387.
19. Amineni, U., D. Pradhan, and H. Marisetty. 2010. In silico identification of common putative drug targets in *Leptospira* interrogans. *J. Chem. Biol.* 3:165–173.
20. Raman, K., K. Yeturu, and N. Chandra. 2008. Target TB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst. Biol.* 2:109.
21. Shanmugham, B., and A. Pan. 2013. Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical in silico approach. *PLoS One* 8:e59126.
22. Vetrivel, U., G. Subramanian, and S. Dorairaj. 2011. A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. *HUGO J.* 5:25–34.
23. Abadio A.K.R., E.S. Kioshima, M.M. Teixeira, N.F. Martins, B. Maigret, and M.S.S. Felipe. 2011. Comparative genomics allowed the identification of drug targets against human fungal pathogens. *BMC Genomics* 12:75.
24. Darapaneni, V., V.K. Prabhaker, and A. Kukol. 2009. Large-scale analysis of influenza A virus sequences reveals potential drug target sites of non-structural proteins. *J. Gen. Virol.* 90:2124–2133.
25. Geyer, J.A., S.T. Prigge, and N.C. Waters. 2005. Targeting malaria with specific CDK inhibitors. *Biochim. Biophys. Acta* 1754:160–170.
26. Huthmacher, C., A. Hoppe, S. Bulik, and H.-G. Holzhutter. 2010. Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis. *BMC Syst. Biol.* 4:120.
27. Bakheet, T.M., and A.J. Doig. 2010. Properties and identification of antibiotic drug targets. *BMC Bioinformatics* 11:195.
28. Haupt, V.J., and M. Schroeder. 2011. Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief. Bioinform.* 12:312–326.
29. Jacob, J.T., G.E. Klein, R. Laxminarayan, Z. Beldavs, R. Lynfield, A.J. Kallen, P. Ricks, J. Edwards, A. Srinivasan, and S. Fridkin. 2013. Vital signs: carbapenem-resistant *Enterobacteriaceae*. *MMWR* 62:165–169.
30. Sarkar, M., L. Maganti, N. Ghoshal, and C. Dutta. 2012. In silico quest for putative drug targets in *Helicobacter pylori* HPAG1: molecular modeling of candidate enzymes from lipopolysaccharide biosynthesis pathway. *J. Mol. Model.* 18:1855–1866.
31. Anishetty, S., M. Pulimi, and G. Pennathur. 2005. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comp. Biol. Chem.* 29:368–378.
32. Butt, A.M., I. Nasrullah, S. Tahir, and Y. Tong. 2012. Comparative genomics analysis of *Mycobacterium ulcerans* for the identification of putative essential genes and therapeutic candidates. *PLoS One* 7:e43080.
33. Bhasin, M., A. Garg, and G. Raghava. 2005. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* 21:2522–2524.
34. Nakai, K., and P. Horton. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *TIBS* 24:34–35.
35. Yu, C.-S., C.-W. Cheng, W.-C. Su, K.-C. Chang, S.-W. Huang, J.-K. Hwang, and C.-H. Lu. 2014. CELLO2GO: a web server for protein subCELLular LOCALization prediction with functional gene ontology annotation. *PLoS One* 9:e99368.
36. Law, V., C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, and V. Neveu. 2013. Drug Bank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42:D1091–D1097.
37. Koch, L. 2015. Shaping the gut microbiome. *Nat. Rev. Microbiol.* 13:4. Available at <https://www.nature.com/articles/nrmicro3401>
38. Walter, J., and R. Ley. 2011. The human gut microbiome: ecology and recent evolutionary changes. *Annu. Rev. Microbiol.* 65:411–429.
39. Aono, R., T. Negishi, and H. Nakajima. 1994. Cloning of organic solvent tolerance gene *ostA* that determines

- n-hexane tolerance level in *Escherichia coli*. *Appl. Environ. Microbiol.* 60:4624–4626.
40. Nogly, P., I. Gushchin, A. Remeeva, A.M. Esteves, N. Borges, P. Ma, A. Ishchenko, S. Grudin, E. Round, I. Moraes, V. Borshchevskiy, H. Santos, V. Gordeliy, and M. Archer. 2014. X-ray structure of a CDP-alcohol phosphatidyltransferase membrane enzyme and insights into its catalytic mechanism. *Nat. Commun.* 5:4169. DOI: 10.1038/ncomms5169. Available at <https://www.nature.com/articles/ncomms5169>
  41. Matsuyama, S.I., N. Yokota, and H. Tokuda. 1997. A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of *Escherichia coli*. *EMBO J.* 16:6947–6955.
  42. Van Straaten, K.E., B.W. Dijkstra, W. Vollmer, and A.-M.W. Thunnissen. 2005. Crystal structure of MltA from *Escherichia coli* reveals a unique lytic transglycosylase fold. *J. Mol. Biol.* 352:1068–1080.
  43. Kumamoto, C.A., and A.K. Nault. 1989. Characterization of the *Escherichia coli* protein-export gene *secB*. *Gene* 75: 167–175.
  44. Read, T.D., S.R. Gill, H. Tettelin, and B.A. Dougherty. 2001. Finding drug targets in microbial genomes. *Drug Discov. Today* 6:887–892.
  45. Zhang, R., and Y. Lin. 2008. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.* 37:D455–D458.
  46. Vander Heiden, M.G. 2011. Targeting cancer metabolism: a therapeutic window opens. *Nat. Rev. Drug Discov.* 10: 671.
  47. Ravna, A.W., G. Sager, S.G. Dahl, and I. Sylte. 2008. Membrane Transporters: Structure, Function and Targets for Drug Design. *Transporters as Targets for Drugs*. Springer, Berlin, Heidelberg, pp. 15–51.
  48. Capra, E.J., and M.T. Laub. 2012. Evolution of two-component signal transduction systems. *Annu. Rev. Microbiol.* 66:325–347.
  49. Peng, C., and F. Gao. 2014. Protein localization analysis of essential genes in prokaryotes. *Sci. Rep.* 4:6001.
  50. Chatterjee, S.N., and K. Chaudhuri. 2012. Gram-Negative Bacteria: The Cell Membranes. In *Outer Membrane Vesicles of Bacteria*. Springer, Springer-Verlag Berlin Heidelberg, pp. 15–34.
  51. Silhavy, T.J., D. Kahne, and S. Walker. 2010. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2:a000414.
  52. Delcour, A.H. 2009. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta* 1794:808–816.

Address correspondence to:  
*Morteza Hadizadeh, MS*  
*Physiology Research Center*  
*Institute of Basic and Clinical Physiology Sciences*  
*Kerman University of Medical Sciences*  
*Kerman, 76198-13159*  
*Iran*

*E-mail:* [morteza.hadizade@gmail.com](mailto:morteza.hadizade@gmail.com)