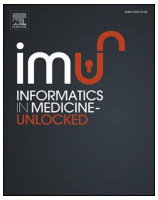
Informatics in Medicine Unlocked 31 (2022) 100998

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



An *in silico* approach towards identification of novel drug targets in *Klebsiella oxytoca*

Umme Hafsa [a], GS Chuwdhury [a], Md Kamrul Hasan [b], Tanveer Ahsan [a], Mohammad Ali Moni [c,d,*]

[a] Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
[b] Department of Theoretical and Computational Chemistry, University of Dhaka, Bangladesh
[c] Department of Computer Science and Engineering, Pabna University of Science and Technology, Bangladesh
[d] School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, Australia

ARTICLE INFO

Keywords:
Klebsiella oxytoca
Computer-aided drug design
Non-homologous protein
Metabolic pathways
In silico approach
Molecular docking

ABSTRACT

The prolonged use, incorrect diagnosis, unnecessary prescription, improper dosing over the year has transformed *klebsiella* organisms into resistant to antibiotics. The emergence of resistance to many antibiotics and drugs makes treatment options limited. Now, the antibiotic pipeline has become severely dry. So, there is a pressing necessity for a new drug against *Klebsiella oxytoca* infections as well as the identification of drug targets. In this study, a systematic proteome subtractive method is used to screen out the most potential drug targets of *Klebsiella oxytoca* that might facilitate the discovery of putative drug in the near future. The comparative proteome analysis of host and pathogen was made to identify the non-homologous proteins which showed no similarity with human host proteins. KEGG pathway analysis was made to identify common and unique metabolic pathways. A computational analysis was carried out to list out the indispensable non-homologous proteins of the pathogen. Essential proteins were predicted by the analysis of protein-protein interactions networks to reveal the proteins which are exigent for the survival of the pathogen, and these proteins also have a lethal effect when removed from the pathogen. In this study, 43 essential proteins were identified. To predict subcellular localization, CELLO (version 2.5) and PSORTb (version 3.00) tools were used. The druggability of proteins was predicted using the DrugBank database. Besides, the physiochemical properties of proteins were analyzed using the Protparam tool of ExPASy. After physiochemical properties analysis and the based on 3D structure availability in the Protein Data Bank, the homology model was built for only one influential drug target using MODELLER. In the end, molecular docking study was carried out to investigate the protein interactions with five different drugs.

1. Introduction

Klebsiella oxytoca is a non-motile, Gram-negative, rod-shaped bacillus in the *Enterobacteriaceae* family. It causes nosocomial infection in hospitalized patients as well as high morbidity and mortality [1,2], and it is the second most disease-causing agent of bacteremia after *Klebsiella pneomoniae* [3]. The most common infections caused by *Klebsiella oxytoca* are urinary tract infections, pneumonia, wound infections, and antibiotic-associated hemorrhagic colitis [4].

Generally, *Klebsiella oxytoca* is spread in healthcare environments. These environments might be nursing homes and ICU (intensive care units). While *Klebsiella oxytoca* (KO) resides inside intestines of a person, it is considered healthy and normal. If it spreads outside the intestinal tract, it can lead to severe infections. Increasingly, it has been present in the newborn baby's blood, and they are suffering from neonatal septicemia [5]. Neonatal septicemia symptoms are seizures, slow heart rate, temperature changes, jaundice, vomiting, diarrhea, low blood sugar, breaking difficulties, reduction in movements and sucking, and swollen abdomen. In the traditional drug discovery process, roughly ten to fifteen years are required to bring a new product to the market in the traditional drug discovery system. It is time-consuming and costly to determine drug targets in the wet lab. Therefore, drug target identification by computational based bioinformatics approach is the most preferable way because it reduces time and cost-effectiveness.

Antibiotic resistance is a global concern. Now, it is becoming increasingly threatening to public health. Like other enterobacteria

* Corresponding author. School of Health and Rehabilitation Sciences Faculty of Health and Behavioural SciencesThe University of Queensland St Lucia, QLD 4072, Australia.
E-mail address: m.moni@uq.edu.au (M.A. Moni).

https://doi.org/10.1016/j.imu.2022.100998
Received 31 January 2022; Received in revised form 17 June 2022; Accepted 19 June 2022
Available online 21 June 2022

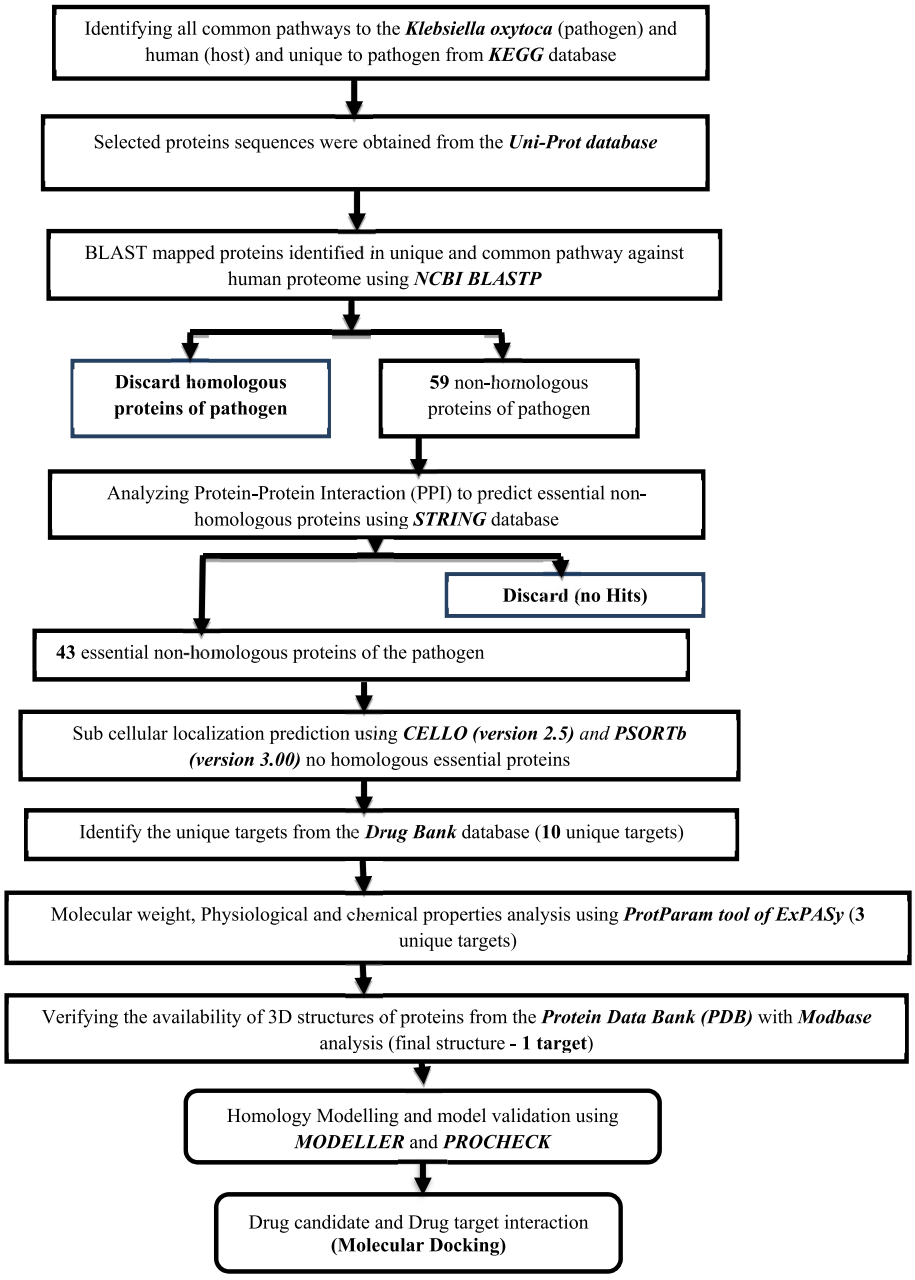


Fig. 1. The workflow of the overall analysis pipeline.

klebsiella oxytoca capable of acquiring antibiotic resistance. Evidently, the resistance of *Klebsiella* species to current antibiotics like penicillins, especially cephalosporinases, ampicillin and carbenicillin, carbapenemases and the oxyimino β-lactams suchas cefotaxime, ceftazidime and the monobactam, and aztreonam are increasing order [6,7].

Many researchers have worked on different organisms, for example *Salmonella typhimurium LT2*, *Streptococcus pneumoniae, Klebsiella pneumoniae*, using the comparative genomic approach to identify drug and vaccine targets with regard to overcome the limitations of drug and vaccine [3,8–10]. A recent study has figured out novel drug targets in the pathogen *Leptospira* using *in silico* approach [11]. It has now been demonstrated that novel drug targets discovered by utilizing the subtractive genomic approaches are useful [12].

Antibiotic resistance has become more common now, and there is an urgent need for developing alternative drugs for better treatments. Furthermore, the drug pipeline is running very slowly and dry [13]. It has become the driving force behind developing a novel and effective drug. So, it is explored the possibility of identification of novel drug targets and designing of drugs against human pathogens. Proteomes of the pathogens are available to use for further study. Thus, in this present study, a strategy has been made for identifying the disease-causing agents in *Klebsiella oxytoca* with the help of the subtractive genomics approach for moving forward to the drug discovery process.

2. Material and methods

2.1. Identification of metabolic pathways

The overall workflow for identifying the putative drug targets in *Klebsiella oxytoca* is shown in Fig. 1. The KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database resource and knowledge base for genes function analysis with higher-order functional information by computerizing present knowledge on cellular processes and utilizing the biological system [14,15]. The metabolic pathways for both host and

pathogen were extracted from the KEGG database (https://www.kegg.jp/), and a manual comparison was made between the host and pathogen pathways. Then, the pathways that were presented in both human and *KO* metabolic pathways were considered as common pathways, and pathways that were not present in humans were selected as unique pathways. Furthermore, all proteins reside in both metabolic pathways that were accessed from the UniProt database (https://www.uniprot.org/), a repository of a comprehensive resource for protein sequences and annotation datasets [16,17].

2.2. Identification of non-homologous proteins

All retrieved proteins from common and unique pathways were imported to BLASTp tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi), and the similarity search of amino acid sequences was carried out against the human proteome [18]. The key aim of this search is to find proteins that show no similarity with human host proteins in the pathogenic bacteria *Klebsiella oxytoca*. Because the availability of non-homologous proteins might lead to cross-reactivity or side-effects of drug compounds with human proteins [19]. The homologous proteins were filtered out with the conditions of expectation value (e-value) ≤ 0.005 and minimum bit score 100. Proteins, that showed hit above the e-value 0.005 were selected as non-homologous proteins, and the remaining amino acid sequences of proteins were eliminated from the list [20–22]. Based on the previous study, the criteria for the e-value was selected [23].

2.3. Identification of essential proteins

To discover the essential proteins of *Klebsiella oxytoca*, resultant non-homologous proteins were further subjected to the STRING database (https://string-db.org/cgi/input.pl) [24], a search tool for recurring instances of neighbouring genes to build a protein-protein interaction network. Essential proteins are those which are crucial for the survival of pathogens and elimination of these proteins have a lethal effect on organisms [25]. In constructed protein-protein interaction network, the degree of nodes (i.e., proteins) is correlated with lethality [26,27]. Thus, essential proteins were identified by selecting hub nodes in the graph, i.e., nodes with three or higher degrees are selected as essential proteins in this study. The remaining proteins were omitted from the list, which shows the dispensability for their cellular process [28,29].

2.4. Subcellular localization prediction

The prediction of subcellular localization for essential proteins is one of the vital steps in search for drug targets. For identification of the biological significance of its function and subcellular localization, all predicted proteins that are non-homologous to humans and essential for pathogens were subjected to subcellular localization prediction. It was carried out by a precise and multicomponent approach using PSORTb (http://www.psort.org/psortb/) [30], a commonly used and powerful tool for predicting the localization of the Gram-negative bacterial protein. It offers unique advantages over other methods. If the confidential prediction is not possible, the result would be in an 'unknown' location [31]. It can identify the subcellular localization of the protein in five regions of a cell: (1) Cytoplasm, (2) Periplasm, (3) Outer membrane, (4) Inner membrane, and (5) Extracellular space [32]. The resultant datasets of PSORTb were further cross-checked using CELLO (http://cello.life.nctu.edu.tw/), which is another web-based tool for implementation of Support Vector Machine (SVM) classifiers to forecast the location of proteins in cells [32]. The cytoplasmic proteins were considered as a putative drug target. From the output of PSORTb and CELLO, proteins that did not show the same region in the cells were omitted from the list [33].

2.5. Druggability

Testing the druggability is another vital ground for drug target identification, i.e., the likelihood of being susceptible to regulate the function of the drug targets with a small drug compound [34,35]. Predicting the similar protein that binds to the drug is the most convenient way to assess the druggability of a protein [36,37]. Druggability of each selected non-homologous essential protein was evaluated by submitting them to the DrugBank Database (https://www.drugbank.ca/biodb/search/bonds/sequence) [38], an online resource that deals with dual fields bioinformatics and cheminformatics, which associates detailed information about drugs and drug targets. The resultant hits found with the DrugBank database were defined as druggable targets, whereas the non-hits were considered unique targets and needed to be evaluated experimentally.

2.6. Physicochemical properties analysis

Being non-homologous to human proteins and unavoidable for the survival of pathogens are not only specifications to select the most attractive drug targets, but some other physiochemical properties also play a significant role and could be useful as drug targets. The properties involve low molecular weight, grand average of hydropathicity (GRAVY), isoelectric point (pI), aliphatic index. All these requisite parameters were measured using ProtParam tools of ExPASy (https://web.expasy.org/protparam/) [39]. Further, Protein Data Bank (PDB) (https://www.rcsb.org/) [40] and Modbase (https://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi) [41] were assessed with the aim of searching for the availability of solved three-dimensional structures of targeted proteins.

2.7. Homology modelling

The homology modelling of possible drug targets was carried out by performing MODELLER v9.20 [42,43], a commonly used and precise computer program for predicting the three-dimensional structure of the 'target' protein. The basic steps include in comparative modelling are initial template selection, then best template selection having higher sequence identity with the 'target', query sequence alignment with the best template structure, model generation, and model assessment. For a three-dimensional structure building, template sequences were retrieved from the NCBI-Blast search against Protein Data Bank (PDB) [43]. Template structures were selected based on the sequence identity and query coverage. The most appropriate model was selected from the various models that were generated by MODELLER. The model was visualized by RasMol [44] software.

2.8. Model assessment

Homology modelling involves the model evaluation phase. The reliability of the simulated model was verified with the help of the Ramachandran plot, which was generated using PDBsum (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) [45]. The Ramachandran graph illustrates phi and psi dihedral angles for the residues of each amino acid in a protein. The graph is separated into favoured, allowed, and disallowed regions. If more than 90% of residues of amino acid lie in the favoured region, then the generated model is considered to be a good quality structure [46].

2.9. Drug preparation and molecular docking calculations

Initial geometry of amikacin, aztreonam, ceftriaxone, tigecycline, and meropenem were taken from online chemical structure database named PubChem (https://pubchem.ncbi.nlm.nih.gov/) (PubChem CID: 37768, 5,459,211, 5479530, 54686904, and 441,130). Further geometry optimization was performed by utilizing Gaussian 09W Revision

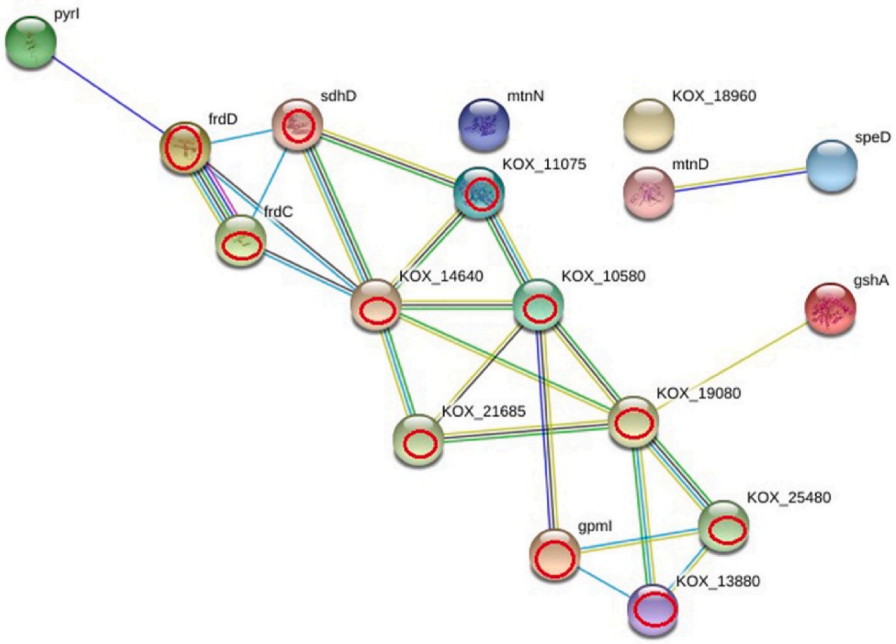


Fig. 2. STRING database image selected by confidence level 0.150 from non-homologous proteins of common pathways. Undirected colored edges between the proteins represent the different types of interaction evidence. Nodes with three or higher degree are selected as essential proteins that are indicated with red circles. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

D.01 [47]. Density functional theory (DFT), along with B3LYP [48], and 6-31G (d, p) basis set was used for optimization in the gas phase. Energy minimization of the protein structure was performed by using Swiss pdb viewer software (http://www.expasy.org/spdbv/) for the preparation of the molecular docking study.

Finally, optimized structures were subjected to molecular docking against the modelled protein, considering the protein as macromolecule, and drugs as ligand utilizing PyRx (version 0.8) software package. Flexible docking was performed considering the center grid box size 52.94 Å, 48.91 Å, 47.63 Å in x, y, and z directions respectively, where the whole protein was covered by the grid box.

3. Results and discussion

3.1. Identification of metabolic pathways

In this study, we used the computational strategy to identify the most attractive drug target for the pathogenic bacteria *Klebsiella oxytoca* through the protein subtractive approach. From the start, we collected information of metabolic pathways of *Klebsiella oxytoca* and human host from the KEGG online database. After performing a manual comparison between the metabolic pathways of the pathogen (KO) and host (human), 46 unique and 31 common metabolic pathways were identified in *Klebsiella oxytoca*. A total number of 132 proteins were found from 31 common pathways of host-pathogen (Table SI1), while 227 proteins were found from 46 unique pathogen pathways (Table SI2). Finally, we accessed all the protein sequences (amino acid sequence) for common and unique pathways from the UniProtKB online database (https://www.uniprot.org/help/uniprotkb), and the sequences were stored on our local machine as FASTA files.

3.2. Identification of non-homologous protein

With the aim of reducing the cross-reactivity or side effects, we excluded homologous proteins from the list using the NCBI-Blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) analysis against the human proteome. Total 59 proteins (18 proteins from common pathways, and 41 proteins from unique pathways) were remained those had no significant sequence similarity to human proteins as non-homologous.

Table 1
List of essential proteins selected by STRING database from non-homologous proteins of common pathways.

SL.	Protein Model Portal ID (3D Structure)	Gene	Protein Name
1.	A0A068H861	KOX_14,640	Putative fumarate hydratase
2.	A0A068H5T7	KOX_10,580	Putative hydratase
3.	A0A068HBS9	KOX_19,080	Putative oxidoreductase, Fe–S subunit
4.	A0A068HB58	sdhD	Succinate dehydrogenase cytochrome b556 small membrane subunit
5.	A0A068H5T5	frdC	Fumarate reductase subunit C
6.	A0A068H608	frdD	Fumarate reductase subunit D
7.	A0A068H7R8	KOX_13,880	Putative transketolase C-terminal subunit
8.	A0A068HD23	KOX_11,075	Bifunctional aconitate hydratase 2/ 2-methylisocitrate dehydratase
9.	A0A068HF19	KOX_21,685	Malate:quinone oxidoreductase
10.	A0A068HHD8	KOX_25,480	Fructose-bisphosphate aldolase
11.	A0A068HK09	gpmI	Phosphoglyceromutase

3.3. Essential proteins

To find the essentiality to the pathogen, the non-homologous proteins were further analyzed using the STRING online database tool. The STRING is a database of predicted and known protein interactions that work through functional and physical associations. In the protein-protein interaction (PPI) network in the STRING database, nodes represent proteins, and undirected edges that are connected to nodes represent the interaction between two or more proteins. The degree of nodes of proteins is correlated with lethality. For better prediction, nodes with lower seeds are excluded. We generated PPI networks with different confidence levels, and the network that had the highest number of connected nodes with a confidence score was selected. In this study,

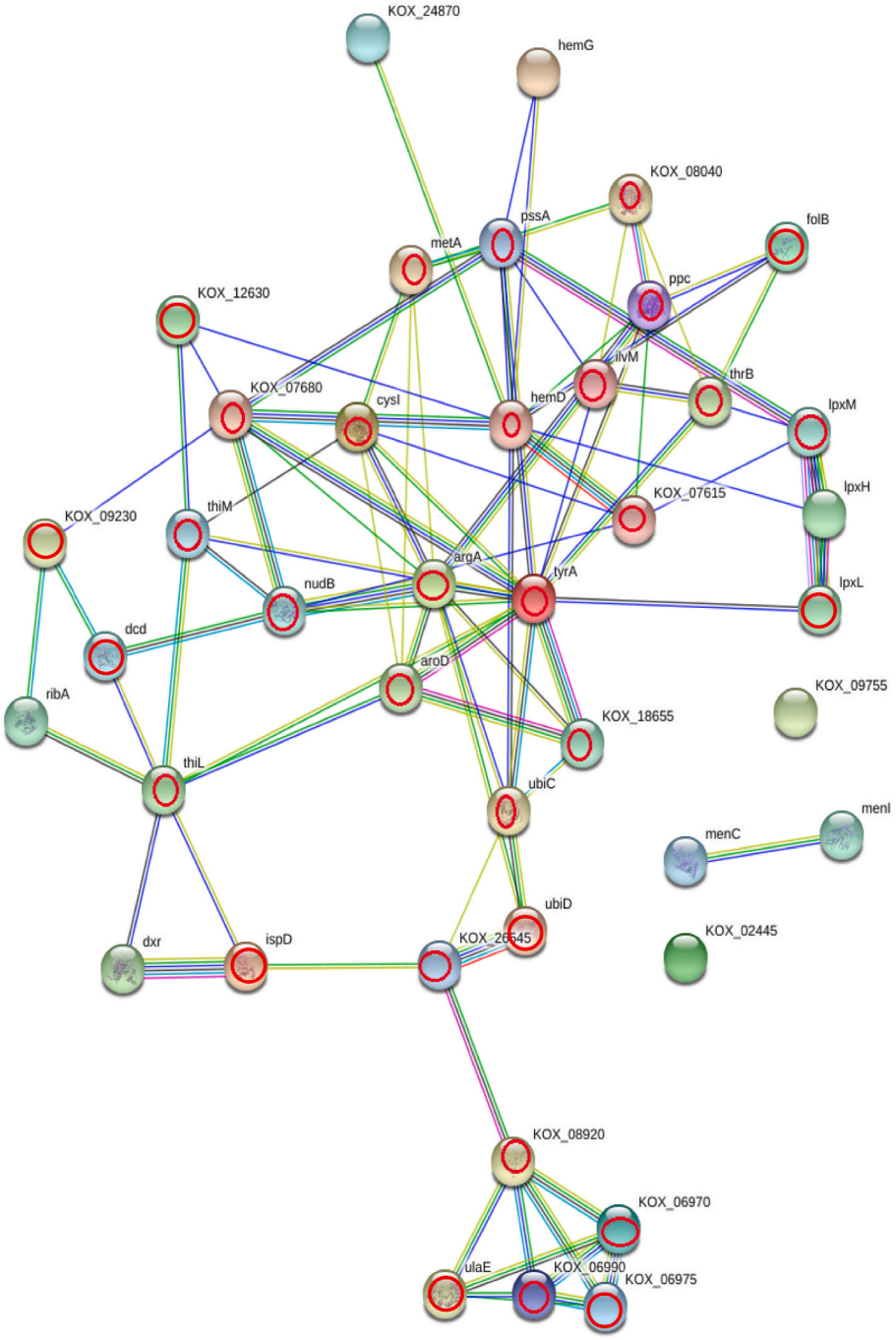


Fig. 3. STRING database image selected by confidence level 0.200 from non-homologous protein of unique pathways. Undirected colored edges between the proteins represent the different types of interaction evidence. Nodes with three or higher degree are selected as essential proteins that are indicated with red circles. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

nodes with three or higher degrees ($K \geq 3$) were selected as essential proteins which represent the significant number of associations. We found 11 out of 18 non-homologous proteins in common pathways as essential proteins with the confidence level 0.150 is shown in Fig. 2 and listed in Table 1. The generated network was also tested by 0.9, 0.7, 0.4, and 0.2 confidence levels. Furthermore, out of 41 non-homologous proteins of unique pathways, we found 32 proteins as essential proteins with the confidence level of 0.200 (tested by 0.9, 0.7, 0.4, 0.300, 2.00, and 0.150 also) is demonstrated in Fig. 3 and listed in Table 2.

3.4. Prediction of subcellular localization

Predicting the location of a protein plays a vital role in the drug discovery and development process. Hence, all identified non-homologous essential proteins were subjected to both CELLO v2.5 and PSORTb tools separately, for subcellular locations prediction. The resultant outputs from the tools were cross-checked for getting out better drug targets (Table SI3). So, based on the prediction of CELLO v2.5 and PSORTb, we selected the proteins those were in the same location in cells. The proteins that reside in Cytoplasmic might be an attractive drug target. In this significant step, 24 non-homologous essential proteins were found from both common and unique

Table 2
List of essential proteins selected by STRING database from non-homologous proteins of unique pathways.

SL.	Protein Model Portal ID (3D Structure)	Gene	Protein Name
1.	A0A068H3W9	ppc	Phosphoenolpyruvate carboxylase
2.	A0A068H4B8	metA	Homoserine O-succinyltransferase
3.	A0A068H556	ilvM	Acetolactate synthase 2 regulatory subunit
4.	A0A068H581	hemD	uroporphyrinogen-III synthase
5.	A0A068H5D6	KOX_07615	Putative uroporphyrinogen III C-methyltransferase
6.	A0A068H5K2	KOX_08040	Malate synthase
7.	A0A068H5P1	ubiC	Chorismate pyruvate lyase
8.	A0A068H6L7	KOX_06970	PTS system lactose/cellobiose-specific transporter subunit IIB
9.	A0A068H726	aroD	3-dehydroquinate dehydratase
10.	A0A068H7A3	KOX_07680	Flavin mononucleotide phosphatase
11.	A0A068H8W3	hrB	Homoserine kinase
12.	A0A068HAP6	KOX_06990	PTS transporter subunit IIA-like nitrogen-regulatory protein PtsN
13.	A0A068HBX2	KOX_08920	Putative L-ascorbate 6-phosphate lactonase
14.	A0A068HDG5	nudB	Dihydroneopterin triphosphate pyrophosphatase
15.	A0A068HE50	thiL	Thiamine monophosphate kinase
16.	A0A068HEI2	lpxM	Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
17.	A0A068HFD6	thiM	Hydroxyethylthiazole kinase
18.	A0A068HFS3	KOX_26,545	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
19.	A0A068HGC9	pssA	Phosphatidylserine synthase
20.	A0A068HH49	cysI	Sulfite reductase subunit beta
21.	A0A068HHD3	KOX_18,655	Chorismate mutase
22.	A0A068HHE7	argA	N-acetylglutamate synthase
23.	A0A068HMC4	tyrA	Bifunctional chorismate mutase/ prephenate dehydrogenase
24.	A0A068H3K0	KOX_06975	PTS system ascorbate-specific transporter subunit IIC
25.	A0A068H524	KOX_09230	Anaerobic ribonucleoside triphosphate reductase
26.	A0A068H5G0	ubiD	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
27.	A0A068H872	KOX_12,630	2-dehydropantoate 2-reductase
28.	A0A068HBX6	ulaE	L-xylulose 5-phosphate 3-epimerase
29.	A0A068HF58	dcd	Deoxycytidine triphosphate deaminase
30.	A0A068HGL3	lpxL	Lipid A biosynthesis lauroyl acyltransferase
31.	A0A068HIB0	folB	Bifunctional dihydroneopterin aldolase/dihydroneopterin triphosphate 2′-epimerase
32.	A0A068HN42	ispD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase

pathways. The dissimilarity in the subcellular localization of essential proteins is shown in Fig. 4 by using the CELLO v2.5 and PSORTb tools.

3.5. Druggability

In drug target identification, the druggability test of a protein is a crucial step that was assessed based on the assumption that druggable protein targets should interact with the drug-like compound. Therefore, the DrugBank database was used to identify the drug targets. For this reason, 24 essential non-homologous proteins that were found in the prediction of subcellular localization step were subjected to the DrugBank database against BLASTp search with default settings. Hits found with DrugBank were considered as the common targets or druggable targets while remained were treated as unique drug targets that are further recommended for experimental validation. As demonstrated in Table 3, 10 out of 24 non-homologous proteins were found as novel unique targets.

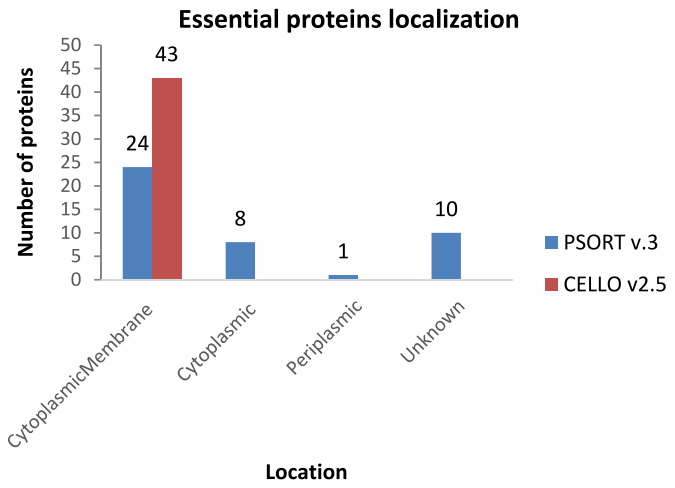


Fig. 4. The dissimilarity in subcellular localizations for identified essential proteins by using CELLO v2.5 and PSORTb tools.

Table 3
Ten unique drug targets were found in DrugBank.

SL.	Uniprot ID	Protein name
1.	A0A068H4B8	Homoserine O-succinyltransferase
2.	A0A068H5K2	Malate synthase
3.	A0A068H7A3	Flavin mononucleotide phosphatase
4.	A0A068H861	Putative fumarate hydratase
5.	A0A068HDG5	Dihydroneopterin triphosphate pyrophosphatase
6.	A0A068HE50	Thiamine monophosphate kinase
7.	A0A068HGC9	Phosphatidylserine synthase
8.	A0A068HMC4	Bifunctional chorismate mutase/prephenate dehydrogenase
9.	A0A068H5G0	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
10.	A0A068HBX6	L-xylulose 5-phosphate 3-epimerase

3.6. Physiochemical properties analysis

In the prioritization strategy, we explored a variety of criteria to classify the more specific drug target for better results. These parameters including lower molecular weight, lower isoelectric point, and higher hydrophobicity (show lower polarity) are playing a very important role in drug targets identification. The Isoelectric Point (pI) value of proteins was calculated as less than 10. For most of the proteins, the Instability Index was above 40, which indicates that they would remain unstable. The Instability Index of less than 40 shows the stability. The proteins had negative Grand Average Hydropathicity (GRAVY) scores on average, which meant that they are hydrophilic (having a tendency to mix with, dissolve in, or be wetted by water) in nature. The Aliphatic index (Ai) value of proteins was above 84. Based on these calculations, the Ai values were calculated quite high, which indicates that the proteins would remain stable over an array of temperatures. These criterion included several cutoff values, i.e., Molecular weight $<$100K Da; PI $<$ 7.2; Instability index $<$40; Ai in between 85 and 104; Hydrophobicity $>$ -0.240. The physiochemical properties of the proteins are illustrated in Table SI4. However, by the above comparison, only 3 out of 10 were selected for the next experiment which are unmarked, on the other hand, 7 were rejected, which are marked by gray color in Table SI4.

3.7. Homology modelling

The three-dimensional (3D) structure was available only for the protein **A0A068HBX6** after checking the availability of 3D structures of 3 proteins from the Protein Data Bank (PDB) with the Modbase tool. Homology modelling was carried out to generate a 3D structure of **A0A068HBX6** protein using MODELLER v9.20. The query sequence of the protein was subjected to the BLASTp search against PDB that

Table 4
DOPE and GA341 score for the generated model of A0A068HBX6 protein.

Filename	molpdf	DOPE score	GA341 score
qseq1.B99990001.pdb	1370.14709	−36044.33203	1.00000
qseq1.B99990002.pdb	1418.36584	−35956.98438	1.00000
qseq1.B99990003.pdb	1345.41772	−36031.97266	1.00000
qseq1.B99990004.pdb	1344.84961	−35872.92969	1.00000
qseq1.B99990005.pdb	1392.72681	−36290.39844	1.00000

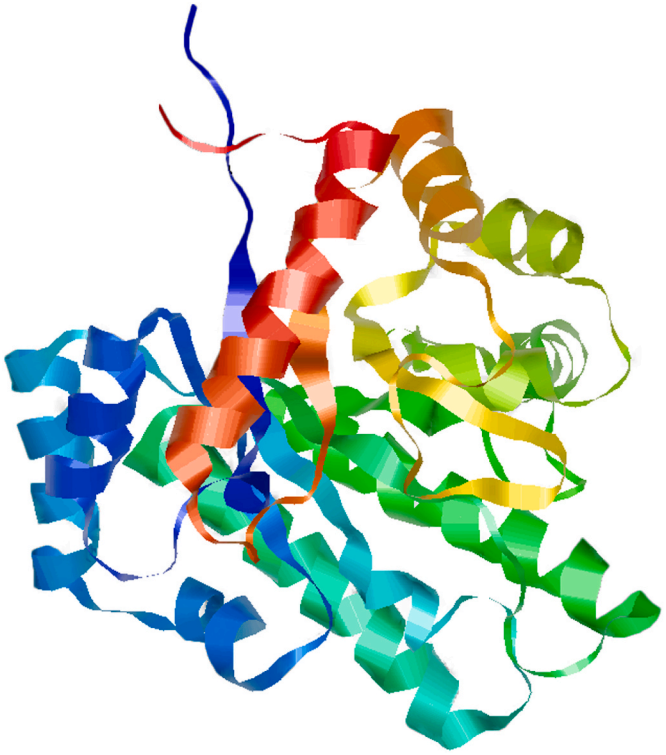


Fig. 5. 3D structure of the selected A0A068HBX6 protein for the targeted sequence in RasMol software.

revealed four template sequences, namely 3cqi, 3cqh, 6btm and 4pgl with higher sequence identity. The sequence identity of 3cqi, 3cqh, 6btm, and 4pgl were 94%, 90%, 28%, and 22%, respectively. These four template sequences were used for homology modelling. After performing modelling using the MODELLER, five 3D models were built for the query sequence of the protein **A0A068HBX6**. The primary principles for selecting the best one among the several models are the lowest DOPE (discrete optimized protein energy) score [49] with the highest GA341 score [50]. According to the summary of built models (Table 4), the 5th model qseq1.B99990005.pdb was selected as it had the lowest DOPE score of −36290.39844 with the highest GA341 score of 1.00000. Therefore, the 3D structure for the selected **A0A068HBX6** protein model "qseq1.B99990005.pdb" was visualized in RasMol (http://www.openrasmol.org/), and it is shown in Fig. 5.

3.8. Model assessment

The final selected **A0A068HBX6** protein's model "qseq1.B99990005.pdb" was evaluated and analyzed with the help of Ramachandran plot using the PROCHECK database [46], which provides a pictorial summary of 3D structure is demonstrated in Fig. 6. In the summary of the Ramachandran plot is shown in Table 5 with appropriate statistics, it is found that out of 284 amino acid residues, 234 residues (93.2%) laid in the core or most favoured region, and 17 residues (6.8%) exist in the allowed region. By these statistics, it was

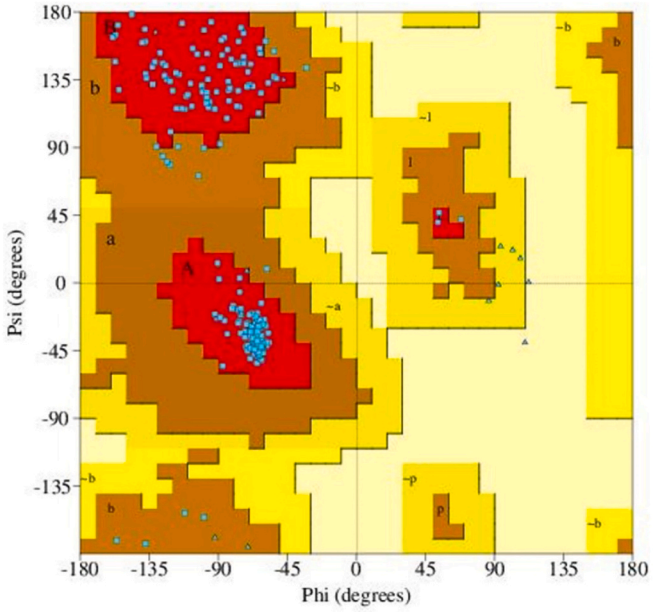


Fig. 6. The Ramachandran plot shows the stereochemical quality of 3D model generated by PROCHECK. The colouring on the plot signifies the phi-psi backbone conformational regions, where the red regions indicate the most favoured regions. The additional allowed regions and generously allowed regions are shown in brown and yellow field, respectively. The light-yellow areas correspond to the disallowed regions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 5
The Ramachandran plot statistics for the 'qseq1.B99990005.pdb' 3D structure model.

	Number of Residues	Percentage of Residues
Most favoured regions [A,B,L]	234	93.2%
Additional allowed regions [a,b,l,p]	17	6.8%
Generously allowed regions [~a,~b,~l,~p]	0	0.0%
Disallowed regions [XX]	0	0.0%
Non-glycine and non-proline residues	(234 + 17+0 + 0) = 251	(93.2 + 6.8+0 + 0)% = 100%
End-residues (excl. Gly and Pro)	2	
Glycine residues	19	
Proline residues	12	
Total number of residues	(251 + 2+19 + 12) = 284	

considered that the generated model "qseq1.B99990005.pdb" was good and reliable as above 90% residues placed in the most favour region [51,52].

3.9. Molecular docking interactions analysis

Molecular docking is an important tool to investigate the binding affinity of drugs with the protein [53]. We studied the binding interactions of five selected drugs with the modelled protein (qseq1.B99990005.pdb), among them the docking interactions of amikacin and aztreonam are shown in Fig. 7. The greater the negative score of binding affinity, reveals the stronger binding interactions of ligands with the protein. Here, Aztreonam, meropenem, and ceftriaxone have relatively higher bonding scores than the others, that indicates they have interaction with the protein more strongly [54]. Aztreonam, meropenem, and ceftriaxone has binding affinity value of −7.8 kcalmol^{-1}, -7.7 kcalmol^{-1}, and -7.6 kcalmol^{-1}, respectively, which are tabulated in Table 6.

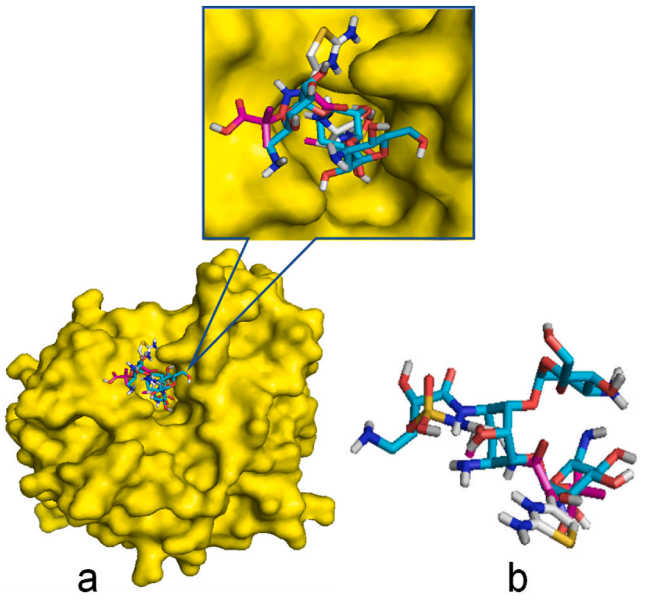


Fig. 7. Molecular docking. (a): docked conformation, and (b) superimposed view of amikacin and aztreonam with the modelled protein (qseq1. B99990005.pdb).

The other two ligands (Amikacin and tigecycline) have relatively lower binding affinity scores, i.e., −6.7 kcalmol^{-1} and -7.2 kcalmol^{-1}.

There are some other parameters such as hydrogen bonding, halogen bonding, and hydrophobic interactions that are also related to the bonding interactions of the protein-ligand complexes [55]. Hydrogen bond distances of less than 2.3 Å increases the binding affinity score between ligand and protein [54,56]. So, the non-covalent bond interactions in the drug-protein complex were examined (Table 6 and Fig. 8). In this study, all the drugs have mostly conventional hydrogen bonds with the amino acid residue, and among them, amikacin has lowest number of interactions but less bond distance. Ceftriaxone and tigecycline also have carbon-hydrogen bonds. Alkyl, Pi-alkyl, Anion, Pi-Anion and Pi-Donor bond interactions also been shown by most of the drugs.

Table 6
Binding affinity scores of the five selected drugs with the modelled protein.

Drugs	Binding affinity score (kcalmol^{-1})	Residue in contact	Interaction type	Bond distance (Å)
Amikacin	−6.7	LYS221	H	1.94548
		GLU42	H	2.31864
		GLU251	H	2.0036
		ASP158	H	1.99692
Aztreonam	−7.8	SER77	H	3.0882
		ARG80	H	3.09451
		LYS213	H	3.15086
		LYS221	H	3.215
		TRP253	H	3.2583
		ASP41	H	2.64982
		GLU42	PA	3.7884
		TYR122	Pd	3.09164
Ceftriaxone	−7.6	PHE225	H	3.10876
		GLU227	H	3.30537
		SER254	H	2.92154
		TRP270	H	3.03617
		THR256	H	1.9047
		TRP253	H	2.20893
		GLU227	H	2.97987
		TRP270	C	3.27973
		GLU263	C	3.29478
		VAL223	C	3.70647
		LYS266	A	4.80072
Meropenem	−7.7	ARG80	H	3.15353
		ARG80	H	2.93757
		PHE220	H	3.18974
		LYS221	H	3.38334
		GLU155	H	2.61056
		LYS221	A	3.79296
		VAL219	A	3.70771
		TRP253	Pi-A	4.49596
Tigecycline	−7.2	CYS20	H	3.26155
		TRP21	H	2.87957
		TRP52	H	3.0432
		TRP52	H	2.90554
		ASP45	H	2.88132
		ALA59	H	1.98036
		GLU19	C	3.34659
		ASP45	C	3.57015
		ALA59	Pi-A	4.62021

H= Conventional Hydrogen Bond, C= Carbon Hydrogen Bond, A = Alkyl, Pi-A = Pi-Alkyl, Pd= Pi-Donor Hydrogen Bond, PA= Pi-Anion.

4. Discussion

The identification of novel drugs and vaccines is now becoming increasingly crucial due to the growing resistance to existing drugs. To overcome this limitation, there arises the necessity of exploring the possibility of identifying new drug targets in the death-causing bacteria *Klebsiella oxytoca*. Nowadays, computational based techniques have become the most effective choice for identifying novel drug targets [10,33,57–59]. Researchers have found out potential drug targets in several pathogenic bacteria using comparative and subtractive genomic approaches [3,8–10,60–62]. However, in this study, we applied an *in silico* comprehensive subtractive genomics approach for discovering prospective drug targets that are involved in life-threatening infection to *Klebsiella oxytoca*; non-existent in the human host organisms.

The initial stage of the analysis revealed that the pathogenic organism *Klebsiella oxytoca* is made up of a wide range of unique metabolic pathways that are essential for survival and have no human homologs. Primarily, 227 and 132 proteins were identified in 46 unique and 31 common metabolic pathways, respectively. The homologous proteins were excluded from the list as the existence of these proteins can cause cross-reactivity that is detrimental to the host [19,63]. The protein-protein interaction (PPI) networks analysis was carried out to identify essential proteins based on selecting the hub nodes with three or higher interacting partners. At this stage, 43 essential proteins were discovered using PPI network analysis in STRING database. Several studies have shown that PPI network analysis is effective in term of finding out the hub proteins which might be lethal for pathogen's survival [64–66].

On the other hand, the study of the subcellular localization of proteins can significantly increase the accuracy regarding target identification during the drug discovery process. It is important for unravelling protein functions that are engaged in various cellular processes. Hence, in this step, the 43 selected essential proteins were further analyzed to predict the location of proteins in cells using PSORTb. Furthermore, to increase the accuracy of prediction, the proteins were further cross-checked with CELLO. After subcellular location prediction in cells of 43 essential proteins, 24 out of 43 proteins were subjected to the DrugBank database for druggability testing. 10 out of 24 proteins were determined based on their druggability analysis.

In the prioritization step, physiochemical properties of unique targets that are resident of the cytoplasm were estimated for getting better drug targets. By considering several parameters, such as lower molecular weight, lower isoelectric point, higher hydrophobicity and instability index, 3 out of 10 drug targets were prioritized. Among them, protein Thiamine monophosphate kinase encoded by the gene *thiL* is responsible for the *Thiamine biosynthesis* pathway and 3-octaprenyl-4-hydroxybenzoate carboxy-lyase encoded by the gene *ubiD* is associated with Ubiquinone biosynthesis pathway. Accordingly, protein L-xylulose 5-phosphate 3-epimerase coded by the gene *UlaE* belongs to the *Ascorbate degradation* pathway. However, all these three targets are

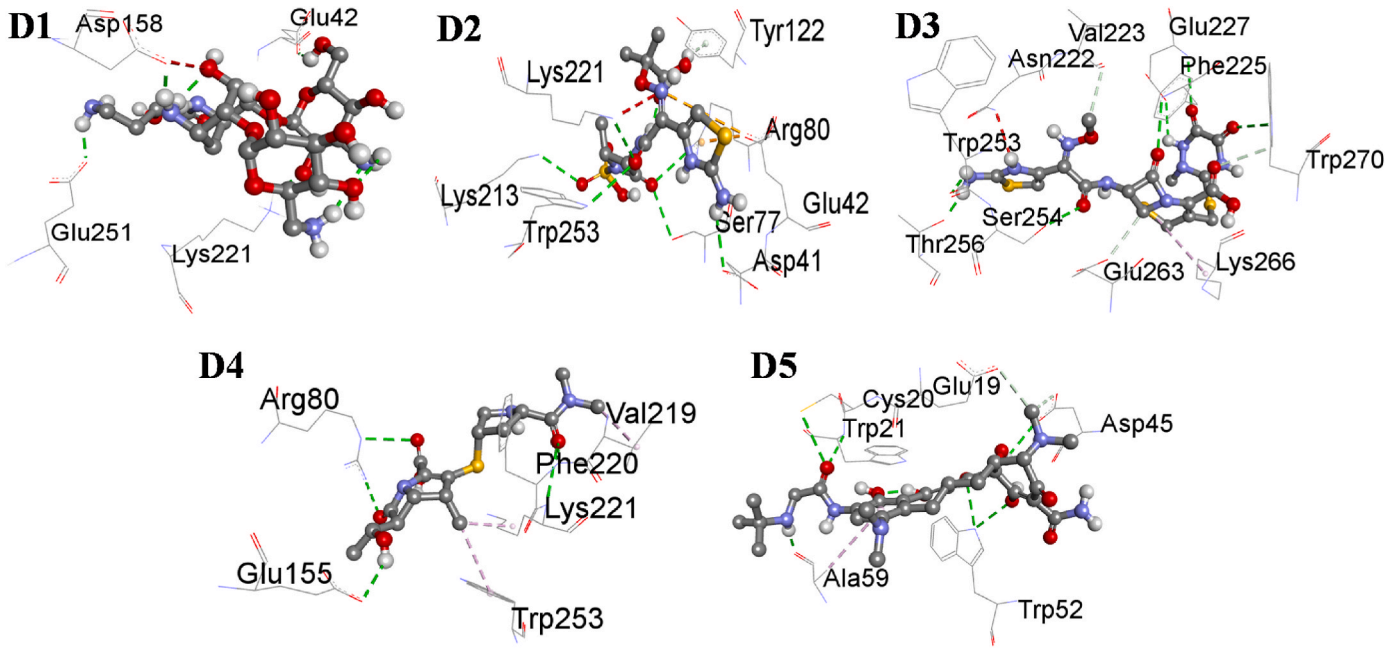


Fig. 8. Nonbonding interactions of five selected drugs with the modelled protein.

indispensable for pathogen survival.

Consequently, based on 3D structure availability, homology modelling was performed on protein **A0A068HBX6** (*ulaE*) to analyze structural characteristics. The model structures of **A0A068HBX6** were revealed using the four templates with a higher sequence identity score. Based on the lowest DOPE score of −36290.39844 and the highest GA341 score of 1.00000, the model structure qseq1.B99990005.pdb was selected as the best model. Next, the 3D protein model was evaluated by the Ramachandran plot that was calculated to 93.2% residues and located in the most favoured region, which indicates the model was highly valid and most reliable.

Virtual screening of protein-ligand complex was completed to investigate the interaction between protein and drugs by molecular docking analysis. Binding affinity score is the key point in this analysis, where the better the binding score, reveals better interactions. The binding affinity score for the three selected drugs Aztreonam, meropenem, and ceftriaxone is of −7.8 kcalmol^{-1}, -7.7 kcalmol^{-1}, and -7.6 kcalmol^{-1}, respectively. Their relatively better binding scores suggest their better interactions with the modelled protein. Non-covalent bond suggests the stability and efficacy of the bond. Among the selected drugs, most of them have shown better non-covalent interactions.

Herein, we performed the subtractive genomic approach (Figure SI1) to identify significant drug targets in *Klebsiella oxytoca*, and the suggested drugs would be progressed to the drug design and discovery process.

5. Conclusion

In recent years, *Klebsiella* organisms have become important pathogens in nosocomial infections. Due to the emergence of resistance to existing drugs, our research work is focused on drug target identification to improve the treatment of bacterial infectious diseases caused by *Klebsiella oxytoca*. In this *in silico* based approach, we identified a series of proteins that could be used as drug targets. In this study, by performing a comparative metabolic pathway analysis, a set of proteins were identified as drug targets. The targets that were identified are indispensable for the growth of the organisms. The identification of 10 drug targets provides the foundation for the computer-aided drug design process against *Klebsiella oxytoca*. On the basis of the physicochemical properties analysis of these 10 proteins, three of them were selected for further analysis. These three proteins are related to three distinct pathways, i.e., *ulaE* from the Ascorbate degradation pathway, *ubiD* from the Ubiquinone biosynthesis pathway, and *thiL* from the Thiamine biosynthesis pathway. Five models were predicted using the homology modelling analysis, and the best one was selected from them. Molecular docking reveals the biological interactions with drugs where most of them show good binding interactions with the modelled protein and nonbonding interactions suggest stability of the bonds.

Computer-aided drug design with the help of bioinformatics tools is more useful. Moreover, the *in silico* based approach reduces the time and complications, which could make it faster in clinical trials.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2022.100998.

References

[1] Moradigaravand D, Martin V, Peacock SJ, Parkhill J. Population structure of multidrug-resistant Klebsiella oxytoca within hospitals across the United Kingdom and Ireland identifies sharing of virulence and resistance genes with K. pneumoniae. Genome biology and evolution 2017;9:574–84.
[2] Sabio JM, López-Gómez M, Jiménez-Alonso J. Spontaneous spondylodiscitis caused by Klebsiella oxytoca [5]. 2002.
[3] Talukdar S, Bayan U, Saikia KK. In silico identification of vaccine candidates against Klebsiella oxytoca. Comput Biol Chem 2017;69:48–54.
[4] Högenauer C, Langner C, Beubler E, Lippe IT, Schicho R, Gorkiewicz G, et al. Klebsiella oxytoca as a causative organism of antibiotic-associated hemorrhagic colitis. N Engl J Med 2006;355:2418–26.
[5] Sahly H, Podschun R. Clinical, bacteriological, and serological aspects of Klebsiella infections and their spondylarthropathic sequelae. Clin Diagn Lab Immunol 1997;4:393.
[6] Avcıoğlu NH, Bilkay IS. Comparative assessment of five clinical Klebsiella isolates in terms of antibiotic resistance and plasmid profiles/Beş farklı klinik Klebsiella türünün antibiyotik dirençliliği ile plazmid profillerinin karşılaştırmalı olarak değerlendirilmesi. Turk J Biochem 2015;40:448–55.
[7] Diagbouga S, Salah F, Sadji A, Dabire A, Nadembega C, Kere A, et al. Detection of high prevalence of TEM. SHV/CTX-M Genes in ESBL Producing and Multidrug Resistant Klebsiella; 2016.
[8] Sharma OP, Kumar MS. Essential proteins and possible therapeutic targets of Wolbachia endosymbiont and development of FiloBase-a comprehensive drug target database for lymphatic filariasis. Sci Rep 2016;6:19842.
[9] Mondal SI, Ferdous S, Jewel NA, Akter A, Mahmud Z, Islam MM, et al. Identification of potential drug targets by subtractive genome analysis of

Escherichia coli O157: H7: an in silico approach. Comput Biol Chem Adv Appl: AABC 2015;8:49.
[10] Hadizadeh M, Tabatabaiepour SN, Tabatabaiepour SZ, Hosseini Nave H, Mohammadi M, Sohrabi SM. Genome-wide identification of potential drug target in enterobacteriaceae family: a homology-based method. Microb Drug Resist 2018; 24:8–17.
[11] Gupta R, Verma R, Pradhan D, Jain AK, Umamaheswari A, Rai CS. An in silico approach towards identification of novel drug targets in pathogenic species of Leptospira. PLoS One 2019;14.
[12] Bhoi P, Sahu TK, Patel A. Identification of novel drug targets in Streptococcus pneumoniae using subtractive genomic approach. CSVTU International Journal of Biotechnology, Bioinformatics and Biomedical 2019;4:79–86.
[13] Cressey D. Traditional drug-discovery model ripe for reform. Nature 2011;471: 17–8.
[14] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
[15] Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 2010;38:D355–60.
[16] Magrane M. UniProt Knowledgebase: a hub of integrated protein data. Database 2011. 2011.
[17] Pundir S, Magrane M, Martin MJ, O'Donovan C, Consortium U. Searching and navigating UniProt databases. Current protocols in bioinformatics 2015;50:1.27. 21–21.27. 10.
[18] Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res 2008;36. W5-W9.
[19] Sarkar M, Maganti L, Ghoshal N, Dutta C. In silico quest for putative drug targets in Helicobacter pylori HPAG1: molecular modeling of candidate enzymes from lipopolysaccharide biosynthesis pathway. J Mol Model 2012;18:1855–66.
[20] Collins J, Coulson A, Lyall A. The significance of protein sequence similarities. Bioinformatics 1988;4:67–71.
[21] Pearson WR. [15] Effective protein sequence comparison. Methods in enzymology. Elsevier; 1996. p. 227–58.
[22] Pearson WR. Comparison of methods for searching protein sequence databases. Protein Sci 1995;4:1145–60.
[23] Chawley P, Samal HB, Prava J, Suar M, Mahapatra RK. Comparative genomics study for identification of drug and vaccine targets in Vibrio cholerae: MurA ligase as a case study. Genomics 2014;103:83–93.
[24] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2012;41:D808–15.
[25] Folador EL, de Carvalho PVSD, Silva WM, Ferreira RS, Silva A, Gromiha M, et al. In silico identification of essential proteins in Corynebacterium pseudotuberculosis based on protein-protein interaction networks. BMC Syst Biol 2016;10:103.
[26] Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–2.
[27] Estrada E. Virtual identification of essential proteins within the protein interaction network of yeast. Proteomics 2006;6:35–40.
[28] Kaçar B, Gaucher EA. Experimental evolution of protein–protein interaction networks. Biochem J 2013;453:311–9.
[29] Khuri S, Wuchty S. Essentiality and centrality in protein interaction networks revisited. BMC Bioinf 2015;16:109.
[30] Gardy JL, Spencer C, Wang K, Ester M, Tusnady GE, Simon I, et al. PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic Acids Res 2003;31:3613–7.
[31] Gardy JL, Brinkman FS. Methods for predicting bacterial protein subcellular localization. Nat Rev Microbiol 2006;4:741–51.
[32] Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci 2004;13:1402–6.
[33] Oany AR, Mia M, Pervin T, Hasan MN, Hirashima A. Identification of potential drug targets and inhibitor of the pathogenic bacteria Shigella flexneri 2a through the subtractive genomic approach. In: Silico pharmacology. vol. 6; 2018. p. 11.
[34] Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. Nat Rev Drug Discov 2008;7:900–7.
[35] Butt AM, Nasrullah I, Tahir S, Tong Y. Comparative genomics analysis of Mycobacterium ulcerans for the identification of putative essential genes and therapeutic candidates. PLoS One 2012;7.
[36] Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. Drug Discov Today 2005;10:1675–82.
[37] Keller TH, Pichota A, Yin Z. A practical view of 'druggability. Curr Opin Chem Biol 2006;10:357–61.
[38] Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res 2010;39: D1035–41.
[39] Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. Totowa: Humana Press; 2005. p. 571–607.
[40] Deshpande N, Addess KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, et al. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. Nucleic Acids Res 2005;33: D233–7.
[41] Pieper U, Webb BM, Dong GQ, Schneidman-Duhovny D, Fan H, Kim SJ, et al. ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 2014;42:D336–46.
[42] Chandrasekaran G, Hwang EC, Kang TW, Kwon DD, Park K, Lee J-J, et al. Computational Modeling of complete HOXB13 protein for predicting the functional effect of SNPs and the associated role in hereditary prostate cancer. Sci Rep 2017;7: 1–18.
[43] Webb B, Sali A. Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics 2016;54:5.6. 1–5.6. 37.
[44] Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci 1995;20:374–6.
[45] Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 2005;33:D266–8.
[46] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26: 283–91.
[47] Frisch A. Gaussian 09W reference. 2009. p. 25. Wallingford, USA.
[48] Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 1988;37:785.
[49] Shen M-y, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2507–24.
[50] Doong S-H. Protein homology modeling with heuristic search for sequence alignment. In: 2007 40th annual Hawaii international conference on system sciences (HICSS'07). IEEE; 2007. 128-128.
[51] Sheik S, Sundararajan P, Hussain A, Sekar K. Ramachandran plot on the web. Bioinformatics 2002;18:1548–9.
[52] Carrascoza F, Zaric S, Silaghi-Dumitrescu R. Computational study of protein secondary structure elements: Ramachandran plots revisited. J Mol Graph Model 2014;50:125–33.
[53] Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 2011;7:146–57.
[54] Uzzaman M, Hasan MK, Mahmud S, Fatema K, Matin MM. Structure-based design of new diclofenac: physicochemical, spectral, molecular docking, dynamics simulation and ADMET studies. Inform Med Unlocked 2021;25:100677.
[55] Patil R, Das S, Stanley A, Yadav L, Sudhakar A, Varma AK. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. PLoS One 2010;5:e12029.
[56] Matin MM, Bhuiyan MM, Kabir E, Sanaullah AF, Rahman MA, Hossain ME, et al. Synthesis, characterization, ADMET, PASS predication, and antimicrobial study of 6-O-lauroyl mannopyranosides. J Mol Struct 2019;1195:189–97.
[57] Agamah FE, Mazandu GK, Hassan R, Bope CD, Thomford NE, Ghansah A, et al. Computational/in silico methods in drug target and lead prediction. Briefings Bioinf 2020;21(5):1663–75.
[58] Gupta R, Rai CS. Identification of novel drug targets in pathogenic Aspergillus fumigatus: an in silico approach. In: International conference on recent developments in science, engineering and Technology. Springer; 2019. p. 151–60.
[59] Wadood A, Jamal A, Riaz M, Khan A, Uddin R, Jelani M, et al. Subtractive genome analysis for in silico identification and characterization of novel drug targets in Streptococcus pneumonia strain JJA. Microb Pathog 2018;115:194–8.
[60] Pasala C, Chilamakuri CSR, Katari SK, Nalamolu RM, Bitla AR, Umamaheswari A. An in silico study: novel targets for potential drug and vaccine design against drug resistant H. pylori. Microb Pathog 2018;122:156–61.
[61] Shahid F, Ashfaq UA, Saeed S, Munir S, Almatroudi A, Khurshid M. In silico subtractive proteomics approach for identification of potential drug targets in Staphylococcus saprophyticus. Int J Environ Res Publ Health 2020;17:3644.
[62] Uddin R, Khalil W. A comparative proteomic approach using metabolic pathways for the identification of potential drug targets against Helicobacter pylori. Genes & Genomics 2020:1–23.
[63] Sah PP, Bhattacharya S, Banerjee A, Ray S. Identification of novel therapeutic target and epitopes through proteome mining from essential hypothetical proteins in Salmonella strains: an in silico approach towards antivirulence therapy and vaccine development. Infect Genet Evol 2020:104315.
[64] Folador EL, de Carvalho PVSD, Silva WM, Ferreira RS, Silva A, Gromiha M, et al. In silico identification of essential proteins in Corynebacterium pseudotuberculosis based on protein-protein interaction networks. BMC Syst Biol 2016;10:1–9.
[65] Ghosh AK, Saini S, Das S, Mandal A, Sardar AH, Ansari MY, et al. Glucose-6-phosphate dehydrogenase and Trypanothione reductase interaction protects Leishmania donovani from metalloid mediated oxidative stress. Free Radic Biol Med 2017;106:10–23.
[66] Gupta V, Haider S, Sood U, Gilbert JA, Ramjee M, Forbes K, et al. Comparative genomic analysis of novel Acinetobacter symbionts: a combined systems biology and genomics approach. Sci Rep 2016;6:1–12.