Ind. J. Pharm. Edu. Res., 2025; 59(1s):s256-s273.
https://www.ijper.org

Original Article

Identification of Novel Drug Targets and Potential Antibacterial Traditional Chinese Medicine Compounds for *Klebsiella pneumoniae* through *in silico* and Antibacterial Activity Evaluation

Yanping Li[1,2], Suresh Kumar[3,*], Hao Peng[4], Lihu Zhang[1,*]

[1]Department of Pharmacy, Jiangsu Medical College, Yancheng, Jiangsu, CHINA.
[2]Post Graduate Centre, Management and Science University, University Drive, Off Persiaran Olahraga, Selangor, MALAYSIA.
[3]Department of Diagnostic and Allied Health Science, Faculty of Health and Life Sciences, Management and Science University, Shah Alam, MALAYSIA.
[4]Jinling Hospital, Nanjing Medical University, Nanjing, Jiangsu Province, CHINA.

ABSTRACT

Background: *Klebsiella pneumoniae* is a ubiquitous opportunistic pathogen that poses a significant threat to hospitalized patients by causing a wide range of infections. The alarming increase in clinical resistance to all current antibiotics necessitates the urgent identification of novel therapeutic targets and development of effective antimicrobial agents. **Materials and Methods:** Using pan-genomic analysis of the core protein repertoire of *K. pneumoniae*, we applied a subtractive proteomics approach to uncover potential drug targets. This has led to the identification of CsgD as a promising candidate. Structural modelling and validation of CsgD were performed. A comprehensive virtual screening of 29,384 natural compounds sourced from Traditional Chinese Medicine (TCM) libraries was performed against CsgD, which yielded three candidates with low binding energies and desirable pharmacodynamic profiles, as validated by Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) prediction.100-nanosecond molecular dynamics simulations were performed to further substantiate their efficacy. **Results:** Through a rigorous virtual screening process using the TCM database, we identified three compounds with potential antibacterial activity against *K. pneumoniae*. In particular, mydriatin stands out as a potent CsgD inhibitor, demonstrating a significant inhibitory effect on antibiotic-resistant strains of *K. pneumoniae in vitro* antibacterial activity evaluation. **Conclusion:** This study identified mydriatin as a potential therapeutic agent targeting CsgD in *K. pneumoniae*, offering a promising strategy for the development of novel antimicrobials to combat drug-resistant infections caused by this pathogen. Our results highlight the importance of using natural product libraries and computational methods in the discovery and rational design of novel antibiotics to address the pressing challenges posed by multidrug resistance in *K. pneumoniae*.

Keywords: *In vitro* evaluation, *Klebsiella pneumoniae*, Molecular docking, Molecular dynamics simulation, Subtractive proteomics, Virtual screening.

Correspondence:
Dr. Suresh Kumar
Department of Diagnostic and Allied Health Science, Faculty of Health and Life Sciences, Management and Science University, Shah Alam, MALAYSIA.
Email: sureshkumar@msu.edu.my

Dr. Lihu Zhang
Department of Pharmacy, Jiangsu Vocational College of Medicine, Yancheng 224005, Jiangsu Province, CHINA.
Email: zlh800927@163.com

Received: 14-10-2024;
Revised: 28-11-2024;
Accepted: 13-12-2024.

INTRODUCTION

Klebsiella pneumoniae, a member of the *Enterobacteriaceae* family, is a prevalent cause of various opportunistic infections, encompassing urinary tract infections, pneumonia, and liver abscesses.[1,2] *K. pneumoniae* has emerged as a major threat to clinical and public health[3] owing to the emergence of Multidrug-Resistant (MDR) strains that produce Extended-Spectrum Beta-Lactamases (ESBL) or Carbapenemases (CRKp). Currently, antibiotics are ineffective in combating MDR pathogens. Although colistin and carbapenems are considered highly potent antibiotics, carbapenemase-producing strains of *K. pneumoniae* have developed resistance to these agents.[4,5] Therefore, there is a pressing demand for more effective treatments for multidrug-resistant infections. One way to achieve this is to identify bacterial proteins that could be targets of new classes of antibiotics. The identification of new drug targets is crucial for drug development. Compared to laboratory screening of macromolecules as drug targets, computational drug

DOI: 10.5530/ijper.20251069

Copyright Information :
Copyright Author (s) 2025 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia.[www.mstechnomedia.com]

development, including pan-genome analysis, structure-based drug design, and virtual screening, reduces time and money consumption.[6,7]

This study involved a thorough examination of both the core and pangenome, aiming to pinpoint therapeutic targets and uncover novel drug candidates among a diverse range of natural products, encompassing Traditional Chinese Medicine (TCM). The exploration of natural products to prevent and treat antibiotic-resistant bacterial infections has gained significant traction.[8,9] Our approach in this study involved a computational method based on the pangenome to analyze the core genome of *K. pneumoniae*. This analysis followed a structured sequence of steps that included nonhomology to human proteins, assessment of bacterial essentiality and virulence, druggability evaluation, broad-spectrum analysis, and nonhomology checks against human anti-targets. We examined host-pathogen interactions and performed nonhomology analysis against gut microbiota proteomes to identify human-specific, virulent, essential, and non-homologous targets. Subsequently, we generated the 3D structure of the identified drug targets. To further advance this investigation, we conducted molecular coupling studies to identify effective compounds from a library of natural products that could target the selected drug candidates. The protein-ligand complexes with the most favorable docking results were subjected to molecular dynamics simulations to assess the stability of the interactions.

MATERIALS AND METHODS

Materials

Data collection Core genome data of *K. pneumoniae* were retrieved from the Pan X (Pan-Genome Analysis and Exploration Database) database, and core protein sequences of 500 strains of *K. pneumoniae* were downloaded in FASTA format.

Source of strains

Thirty non-repetitive *K. pneumoniae* strains clinically isolated from Jiangsu Medical College Affiliated Hospital from June to December 2023 were collected. The specimens were obtained from sputum, bronchoalveolar lavage fluid, blood, pus, wound secretions, and urine. The strain was identified as *K. pneumoniae* using a Vitek-2 fully automatic microbiological identification drug sensitivity analyzer. *Escherichia coli* (ATCC 25922) was used as the quality control strain. Simultaneously, the ViteK 2 fully automatic drug susceptibility analyzer and paper diffusion (K-B) method were used for drug susceptibility testing to confirm whether it was a multi-resistant strain. The quality control strain was *K. pneumoniae* (ATCC 700603), and the results of the drug susceptibility test were interpreted according to the CLSI 2023 M100 standard,[10] The Minimum Inhibitory Concentration (MIC) of tigecycline was ≥2 μg/mL. Duplicate bacterial strains isolated from the same patient and site were eliminated. Two standard isolates of *Escherichia coli* (ATCC 25922) and *K. pneumoniae* (ATCC 700603) were obtained from the China Medical Microbial Strain Conservation Center. All cultures were stored in slant agar at 4ºC and used as stock cultures for 14 days.

Methods

Figure 1 presents a flowchart that outlines the process employed for discovering new drug targets and potential candidates for *K. pneumoniae*.

Retrieval of the core proteins of *K. pneumoniae*

The complete proteome sequences of 500 *K. pneumoniae* strains were downloaded from the Pan X: Pan-Genome Analysis and Exploration Database.[11] These 500 strains were previously isolated from all geographical regions of the world. Core proteins were obtained from pan-genome alignments. The core protein consisted of 666,500 amino acid sequences. Core proteins were also preprocessed to remove paralogs with 75% sequence identity using the CD-HIT suite.[12] Furthermore, sequences less than 100 amino acids in length were removed. After pre-processing, the sequences were reduced to 1,136 amino acid sequences.

Non-homology analysis against the human proteome

The BLAST algorithm against the human proteome (UP000005640; updated February 2022)[13] was performed with an E-value threshold >0.005 and restricted the sequence identity to <50%.[14] Subsequently, the homologous human proteins were removed.

Essentiality and virulence analysis

Essential genes are indispensable for pathogen survival.[13] We performed a BLAST search against the Database of Essential Genes (DEG) with a strict E-value threshold of <0.0001 to identify the essential genes.[15] To evaluate the virulence of protein targets, we used the Virulence Factor Database (VFDB),[16] which evaluates the pathogenic virulence associated with these target proteins.

Druggability analysis

Druggability refers to the ability of a protein to bind to a drug-like molecule. BLAST analysis of the hypothetical protein with an E-value<0.0001 against the DrugBank dataset[17] and the Therapeutic Target Database (TTD)[18] was performed to evaluate whether it was a drug-able protein.

Broad-spectrum analysis

A protein can be deemed a potential broad-spectrum protein target with homologues across multiple pathogenic organisms if it is found in at least 25 bacterial protein kingdoms. To identify such hypothetical proteins, the BLAST algorithm was utilized with an E-value threshold of less than 0.0001, and proteomic data

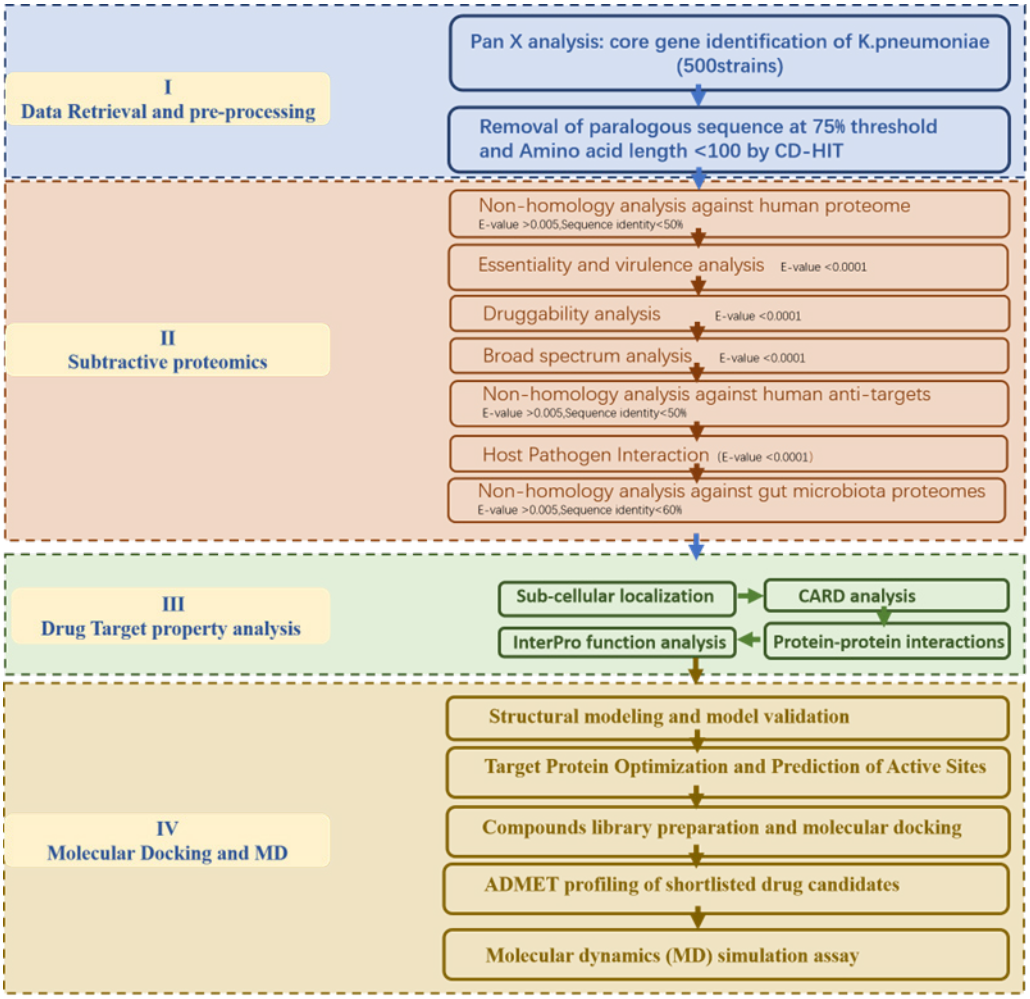

Figure 1: Flowchart illustrating the workflow of the current study, which integrates various approaches and tools to identify potential drug targets and candidates for treating *K. pneumoniae*.

from 181 pathogenic organisms was obtained from the European Bioinformatics Institute (EBI).[19,20]

Nonhomology analysis against human antitargets

Proteins that trigger dangerous side effects under the influence of a drug are called antitargets. We performed BLASTp analysis of protein sequences using a dataset of known human anti-targets gathered from the literature. We applied an E-value>0.005 and used a similarity threshold of < 50% to identify potential antitarget proteins. Proteins with similarity values<50% were retained.[21]

Host pathogen interaction

Homologous proteins were discarded to avoid an autoimmune response in the host. Non-homologous proteins were calculated using the BLAST algorithm in online databases such as HPIDB (version 2.0), PHIbase (version 4.2), and PHISTO (version 2.0). The value of E was set to <0.0001 and a 1% alignment cutoff of 1% was employed.[22-24]

Nonhomology analysis against gut microbiota proteomes

A protein homologous to the human intestinal microbiota can interact and form bonds with proteins in the intestinal microbiota, potentially resulting in unfavorable pharmacokinetic side effects in the host. As a precaution, any homologous protein similar to the human intestine was excluded using BLASTp with a stringent E-value of <0.0001, while the human gut microbiota database was assembled from relevant literature sources.[25,26]

Subcellular location

We conducted subcellular location analysis of the identified drug targets. Characterization of the subcellular location of a protein is essential for understanding its function. We employed PSORTb v 3.0[27] to predict the subcellular location of the selected proteins because it covers a wide range of cell morphologies found in archaeal and bacterial organisms.

Comprehensive antibiotic resistance

Automated BLAST alignment was performed for the protein sequence using the Comprehensive Antibiotic Resistance Database (CARD). The results were accompanied by additional information from CARD.[28]

InterPro function analysis

Predicting the functional families of hypothetical essential proteins is a crucial step in their characterization. To this end, we used InterProScan to predict the functional families of hypothetical individual protein sequences. Recognizing the functional family of a protein is a prerequisite for evaluating its druggability and allows rapid understanding of the role of the target protein during the drug development process.[29] InterProScan facilitates the analysis of a protein's function with annotations based on predictive signatures within the protein's primary sequence.[30]

Structural modeling and model validation

For structure-based drug design, a thorough understanding of the target protein's molecular function and three-dimensional structure is essential. Therefore, we employed a homology modeling technique to create a structural model of the target protein.An AlphaFold2 server was used for this purpose.[31,32] Furthermore, the quality of the modeled protein structure was evaluated for its stereochemical integrity using PROCHECK, VERIFY3D, and ERRAT on the SAVES server (https://saves.mbi.ucla.edu).

Molecular docking

Ligand library preparation

To prepare our ligand library, we performed a virtual screening against the TCMSP database (http://tcmspw.com/tcmsp.php), which contains 499 Chinese herbs and the composite components of each herb (approximately 29,384). Ligand optimization of antibacterial small molecules: We used ChemDraw Ultra 8.0 and Chem 3D 17.1 to construct 3D structures of antibacterial drug molecules, and then imported them into Maestro 11.2. The "Ligprep" tool to optimise the structures and transform the 3D structures with the "OPLS4" force field. The Ligprep tool with the "OPLS4" force field was used for structure optimization and 3D structure transformation.The molecules in the database were preprocessed for protonation, desalting, hydrogenation, generation of reciprocal isomers, generation of stereo conformations, and energy minimisation using Schrödinger's LigPrep module. Pre-processing resulted in 29,384 prescreened compounds as ligand libraries because of the reciprocal isomers. The Schrödinger 2023-1 QickProp module was used for the first screening round according to the pharmacophore rules "Lipinski Ro5" and "Verber Ro3," and 12, 227 eligible hit compounds were selected.[33]

Optimization of target proteins and prediction of active sites

Predictive atomic models of protein structures were obtained using AlphaFold 2 high-precision predictive protein tertiary structure software (https://colab.research.google.com/), which uses machine learning methods to aggregate existing knowledge of protein structures from protein databases with information from sequence comparisons, as well as physical and geometric constraints. Raw sequence PDB files were processed by Maestro 11.9 software to reduce them to 3D models, and the "Protein Preparation Wizard" suite was used to automatically assign bond sequences, add hydrogens, form zero-grade bonds with metals, convert selenomethionine to methionine, add missing side chains, and create possible disulfides. Side chains were added, possible disulfide bonds were created, heterogroups with distances greater than 50 nm from the coordination site were deleted, heteroprotonated states were created at pH 7.0, and restriction minimization was performed using the OPLS4 force field until the Root Mean Square error of the heavy atoms (RMSD) converged to 3 nm.

After structural optimization of the target proteins using Maestro 11.2 "Protein Preparation Wizard" tool for hydrogenation and dehydrogenation, the target proteins were subjected to the online predictive activity pocket software POCASA (https://g6altair.sci. hokudai.ac.jp) to predict the potential active sites of target proteins. Discovery Studio 3.5 software was used to obtain the coordinates of the active pockets, and then set the size and center of the docking box area by site inputting the coordinate values in the "Receptor Grid Generation" program. The size and central position of the docking box were set for the docking of small molecules of the drug ligands.

Molecular docking

Virtual screening coupling is a cost-effective and time-saving method for drug target development and discovery, which can help researchers reduce the number of compounds that require further experimental analysis. Docking simulations were performed using the Glide tool to assess the binding interactions between drug targets and ligands. Visualization of the docking poses was achieved using Discovery Studio version 3.5. A ligand-coupling tool was used to perform molecular coupling of the minimized target protein and its optimized original ligand. Docking occurs separately for different active sites in the target protein. The interaction between the natural compounds and the target protein was evaluated using the Glide score and Glide energy scoring functions, which consider hydrogen bonds, hydrophobic interactions, van der Waals forces, and other interactions. The larger the absolute value of the score, the more stable the docking complex and the better the binding interaction between the compounds and target protein. To analyze their interactions, comparisons were made between the docking scores

of the natural compounds (ligands) and the target protein, as well as between the scores of the original ligand and the target protein.

ADMET profile of selected drug candidates

Ensuring the safety of potential drugs is a primary concern for drug development. Early detection of serious drug toxicity and adverse effects is critical for streamlining drug development, both in terms of time and cost. Therefore, we subjected the shortlisted drug-like compounds to a comprehensive evaluation of their pharmacokinetic properties, including Absorption, Distribution, Metabolism, and Excretion (ADME). This analysis was performed using Discovery Studio 3.5. Subsequently, we assessed the toxicity profiles of the compounds, taking into account factors such as minimal human side effects, immunotoxicity, mutagenicity, teratogenicity, neurotoxicity, enhanced penetration, and carcinogenic potential.A comprehensive analysis was conducted using the pkCSM database. Carbapenems, an effective antibiotic for the treatment of *K. pneumoniae*, were selected as a reference to conduct a more comprehensive study on the effectiveness of the drug.

Molecular dynamics simulation studies

Using the Desmond software package with the OPLS4 force field, we conducted a molecular dynamics simulation test to further analyze the dynamics and stability of all the selected docked complexes. The solvent molecules were modeled using the TIP4P/EW solvent model within a cubic box of size 10, and counterions were introduced to balance the system's charge. To mimic the physiological saline concentration in humans, a 0.15 mol/L NaCl solution was incorporated. The simulation was run for 100 Nanoseconds (ns), with a trajectory captured every 4.8 Picoseconds (ps).The simulations were carried out on hardware with an nVidia GeForce GTX 4090 3200 MHz Graphics Processor (GPU) running on Linux Ubuntu. The 100 ns simulations lasted a maximum of 3-5 hr.

Prediction of biological activity based on the structure of small molecule compounds

The PASS system is an online service for activity prediction based on small-molecule active fragments, which allows the prediction of potential biological activities of drug-like compounds. The system includes more than 35,000 nonidentical compounds with different bioactivities and 500 different types of bioactivities, including pharmacological effects, mechanism of action, mutagenicity, carcinogenicity, teratogenicity, and embryotoxicity, and constitutes a large training set database of these compounds. The prediction of compound bioactivities is precisely based on the results obtained by analyzing the structure-bioactivity relationships of the training set, and the two-dimensional MOL-format structures of the constructed Chinese medicine monomer compounds were inputted into the PASS SAR training set database. After library search and structure matching, the portion with prediction accuracy Pa>0.9 was selected as the possible biological activity of the compounds studied and used as the basis for further research and analysis. Default system settings were used for all parameters in the calculation process.

Antibacterial *in vitro* activity test

Evaluation of antibacterial activity

The antibacterial activities of the compounds were determined using the LB liquid culture medium dilution method. First, an appropriate amount of LB liquid culture medium was sterilized using a high-pressure steam sterilizer and placed on a super-clean table for UV sterilization. LB culture medium was added to a sterile 96-well plate, and the target compound and control group solutions were added to the first well of each column of the 96-well plate. Second, the compound was diluted to the 10th well using the double-dilution method. Finally, the prepared bacterial solution was inoculated (except in the last column). After 24 hr of cultivation in a constant temperature incubator at 37°C, the OD600 value of each well was measured using an enzyme-linked immunosorbent assay (ELISA) reader. The data for each well were obtained using the following formula: Inhibition rate (%)=100% - (OD test well-OD background well)/(OD blank well-OD background well) × 100%. The antibacterial activity of each compound was determined by treatment.

Detection of Minimum Inhibitory Concentration (MIC)

The Minimum Inhibitory Concentration (MIC) was ascertained by employing the microdilution technique outlined in the Clinical and Laboratory Standards Institute (CLSI) guidelines. The determination of MIC helps differentiate between drugs that are effective against bacteria and those that are not. MIC was determined using the modified Mueller-Hinton (MH) broth microdilution method and incubated overnight in an incubator (Kavanagh *et al*., 2019). *K. pneumoniae* was grown to a McFarland standard optical density of 0.500, which corresponds to 1.5×10^8 colonies per milliliter (CFU/mL), which is necessary to determine whether the bacteria are susceptible or resistant to the antibiotic tested ('Recommendations for the Technique of *in vitro* Pharmacovigilance Tests,' 1996). Antibacterial activity tests were conducted *in vitro* on the screened compounds, utilizing tigecycline as a benchmark. Initially, compound solutions were prepared in DMSO at a concentration of 4 mg/mL and divided into three experimental groups (each represented by a column) and a control group. The first row received a mixture of 100 μL of compound solution and 100 μL of MH broth culture medium. Subsequent rows were prepared by serial dilution, where 100 μL from the previous row was transferred to the next. Additionally, 100 μL of culture medium-diluted bacterial solution were added to each well, ensuring a final volume of 200 μL per well.

The control group contained a mixture of 100 µL of compound solution and 100 µL of MH broth. The positive control comprised 100 µL of tigecycline solution and 100 µL of diluted bacterial culture, while the negative control consisted of 100 µL of MHB culture medium and 100 µL of diluted bacterial culture. The culture medium control had 200 µL of culture medium per well, and the solvent-induced control tested the impact of DMSO on bacterial growth with 100 µL of DMSO solution and 100 µL of diluted bacterial culture.After incubation at 37ºC for a specified duration, the concentration corresponding to a visually clear solution was recorded as the minimum inhibitory concentration.

RESULTS

Pan-genome and core genome analysis

Information obtained from the Panx database revealed that *K. pneumoniae* harbored 3,554 core genes among its total of 27,196 genes. The amino acid sequences of the core proteins derived from these core gene alignments amounted to 666,500. Subsequent processing using the CD-HIT tool reduced this number to 1,284 sequences. After eliminating sequences with fewer than 100 amino acids, 1,136 core proteins remained (depicted in Figure 1). These pre-processed core proteins underwent a subtractive proteomics approach involving a series of analyses, as illustrated in Figure 1. A concise overview of the analysis is provided in Table 1.

Subtractive proteomics approach and drug target identification.

Non-homology analysis against the human proteome

The core proteins identified using pan-genome analysis served as the basis for the discovery of new therapeutic targets against *K. pneumoniae*. To this end, 1136 core proteins were thoroughly evaluated by BLAST testing and compared to the entire human proteome. The objective of this study was to ascertain the non-homologous characteristics of potential therapeutic targets by establishing a rigorous E-value threshold of greater than 0.005. Consequently, 1,041 proteins were identified as non-homologous to human expressed proteins, while the remaining proteins were eliminated through a similarity search conducted using BLAST against the human proteome.

Essentiality and virulence analysis

Analysis of essential proteins was carried out on the sequences derived from the non-homologous sequences to the human proteome. Essential proteins are indispensable for the survival of pathogens in any given scenario. By utilizing the BLAST algorithm against the Database of Essential Genes (DEG), critical proteins from *K. pneumoniae* were identified through a similarity search. Furthermore, Virulence Factors (VFs) within *K. pneumoniae* were found to be pivotal in processes like adhesion, invasion, colonization, and persistence within the host organism. They also serve as key regulators of infection. The identified VFs are promising targets for the pathogenicity of *K. pneumoniae*. To explore this potential in more detail, we performed BLASTp analysis of these proteins using the VFDB core dataset. In this study, a strict threshold of E values<0.0001 and an alignment limit of 1% were enforced. Based on the analysis of essentiality and virulence, 895 proteins were identified, which should be further screened for the identification of drug targets.

Druggability analysis

Proteins capable of forming robust bonds with drug molecules fall under the category of druggable proteins. These interactions are characterized by high-affinity connections between proteins and ligands, which rely on enhanced intermolecular forces. We relied on confirmed sources of comprehensive drug-related information, such as the Drugbank database[25] (version 5.0), which includes 4,159 non-redundant drug targets, and the Therapeutic Target Database (TTD) with 2,589 identified targets.[26]

Broad-spectrum analysis

An optimal potential drug candidate promises to address a wide range of future infections. A protein is a potential target for

Table 1: Subtractive genomic analysis steps for *K. pneumoniae*.

Sl. No.	Step	Identified protein
1	Core proteome of *K. pneumoniae*	666500
2	Removal of paralogous protein sequences at 60% threshold.	1284
3	Removal of protein sequences with < 100 amino acids.	1136
4	Non-homology analysis against human proteome.	1041
5	Essentiality and virulence analysis	895
6	Druggability analysis	469
7	Broad-spectrum analysis	469
8	Nonhomology analysis against human antitargets.	469
9	Host-pathogen interactions	319
10	Nonhomology analysis against gut microbiota.	1

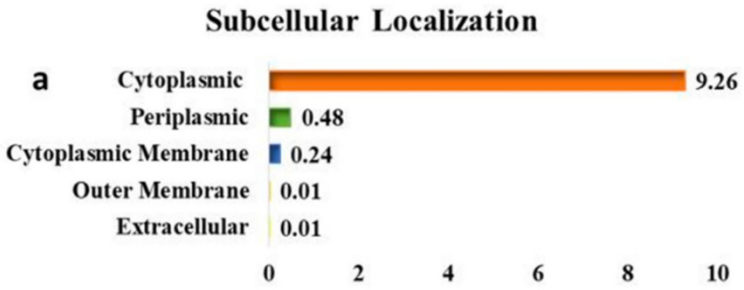

Figure 2: The subcellular localisation of the protein. The results indicate that the query protein belongs to the cytoplasmic region of the cell.

broad-spectrum drugs if it is not homologous and is present in over 25 different bacterial pathogens.[34]

Nonhomology analysis against human antitargets

The history of drug development is marked by the withdrawal of numerous candidates from the market owing to concerns about carcinogenicity. Therefore, evaluation of cross-reactivity and carcinogenic potential plays a crucial role in the development of effective pharmacological compounds.[19] Although we excluded non-homologous host proteins from non-paralogous sequences in our analysis, we also performed anti-target analyses. This additional step aims to prevent unintended interactions between drugs administered to treat pathogens and host antitargets, thereby reducing the risk of harmful side effects.These three crucial processes led to the identification of all 469 proteins, which were characterized as drug-like, nonhomologous antitarget proteins and occurred in a broad spectrum of the bacterial kingdom.

Host pathogen interaction

Homologous proteins were discarded to avoid an autoimmune response in the host. Regarding the specified query proteins, we identified a set of 319 proteins that showed characteristic host-pathogen interactions.

Nonhomology analysis against gut microbiota proteomes

The effectiveness of antibiotics is closely linked to the ecosystem of intestinal microbiota. Proteins similar to those in the human gut microbiota can potentially disrupt the balance of the gut microbiota during drug interactions. Through an in-depth analysis of nonhomologous proteins compared to the human gut microbiota, we found only one protein, the transcriptional regulator protein CsgD (KPHS_44520), with a percentage identity of less than 60%. This protein was identified as non-homologous and was further investigated.

Drug target property analysis

The DNA-binding transcriptional regulator protein CsgD (KPHS_44520) was selected as a new drug target for *K. pneumoniae* using subtractive genomic analysis. The proteins were characterized using the following criteria: subcellular location, CARD analysis, and InterPro functional analysis.

Subcellular location

Cellular localization was predicted using PSORTb. The transcriptional regulator protein CsgD (KPHS_44520) was identified within the cytoplasmic region, as shown in Figure 2, with a commendable localization score of 9.26. Proteins located in the cytoplasmic region offer distinct advantages as potential drug targets. This is primarily because of the abundance of enzymes in this cell compartment, which makes them more suitable for the action of targeted drugs. In fact, cytoplasmic proteins have been reported to be favorable therapeutic targets that can be effectively and easily targeted by drugs.[35,36]

CARD analysis

Proteins associated with resistance mechanisms and drug efflux pathways are promising therapeutic targets for controlling drug-resistant strains of *K. pneumoniae*. CARD includes tools for analyzing molecular sequences, including BLAST and Resistance Gene Identifier (RGI) software, to predict resistomes based on homology and SNP models. The amino acid sequence of the transcriptional regulator CsgD (KPHS_44520) was compared with the CARD database, and the annotation of the drug resistance gene (CARD) was analyzed. The sdiA gene had the highest identity (41%) and was related to the formation of bacterial biofilms. The results of the CARD analysis are displayed in the Supplementary Data Files (Table S1).

InterPro function analysis

Predicting the functional families of hypothetical essential proteins is a crucial step in their characterization. In this research, we employed InterProScan to predict the functional families of hypothetical protein sequences on an individual basis. Recognition of the functional family is crucial for determining the role of the target protein, particularly in the development of drugs for their specific function. As an online tool, InterProScan calculates the functional domains within a protein and predicts its superfamily, providing essential diagnostic signatures for protein classification. In our analysis, InterProScan provided predictions related to the functional transcription domain, which

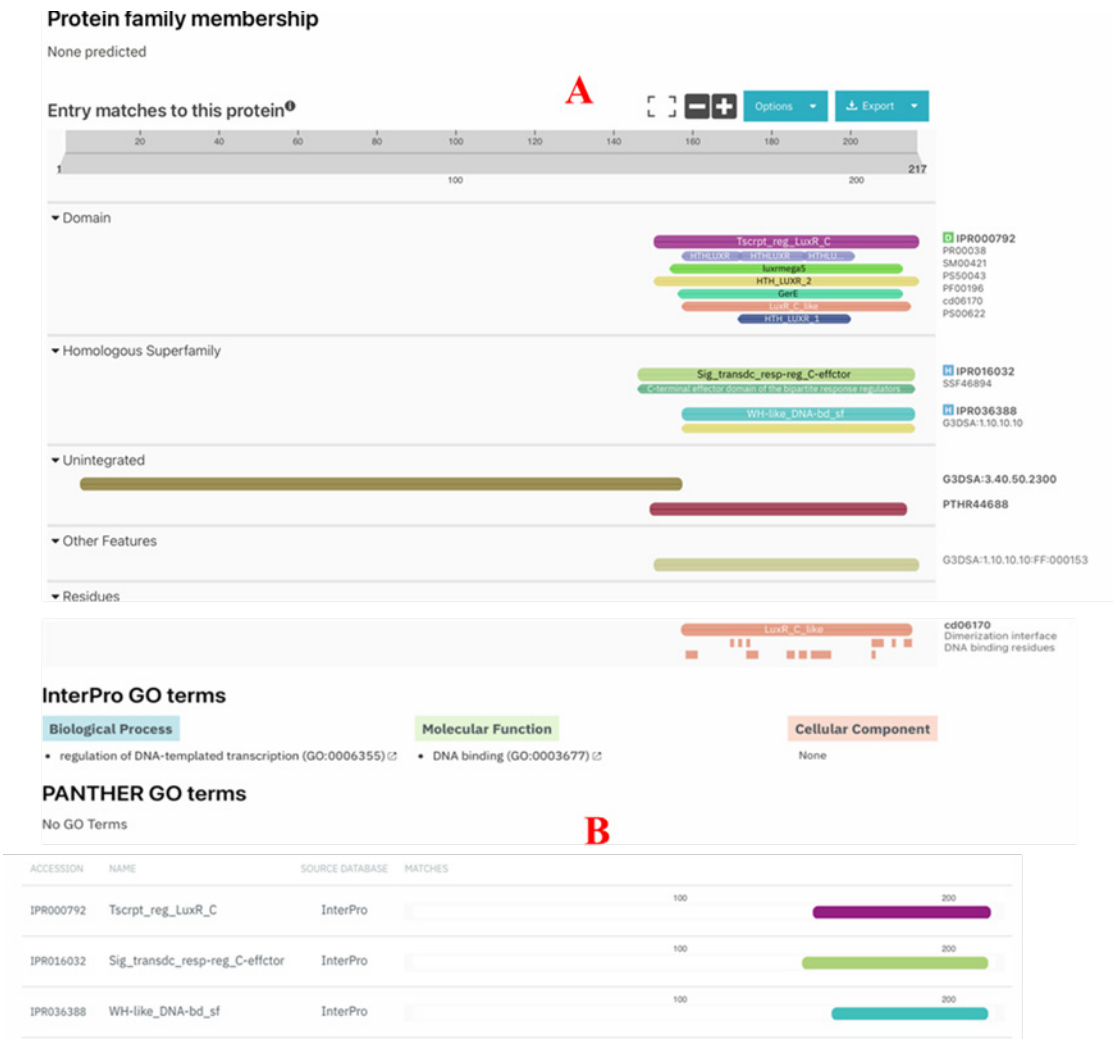

Figure 3: Presents the results obtained from InterProScan. Panel (A) displays the accession numbers associated with the functional domains, whereas Panel (B) organizes and lists the names of these domains.

serves as a diagnostic marker for the functional domain and types of compounds with which it interacts. This protein was found to be associated with the biological process of DNA-assisted transcription regulation, as shown in Figure 3.

Structure prediction and validation

Because the Three-Dimensional (3D) structure of the identified drug target was missing from the PDB database, we used the homology modeling approach supported by the AlphaFold 2 server[32] to generate the 3D structure for the DNA-binding transcriptional regulator CsgD (KPHS_44520). The confidence values of the model structure for most residuals were remarkably high, reflecting the accuracy of the model (pLDDT > 85; Figure 4A). In this plot, lighter green indicates areas of higher error, whereas darker green indicates areas of lower error, representing the expected position error. AlphaFold provides a confidence value per residue (pLDDT) ranging from 0 to 100, where intercepts with a value below 50 may indicate a lack of structure or connectivity, as shown in Figure 4B.

The constructed model underwent validation using PROCHECK, VERIFY3D, and ERRAT tools.[37-39] According to the Ramachandran plot, 86.1% of the residues resided in the most favored regions, 7.0% were in the additionally allowed regions, 5.0% were in the generously allowed regions, and only four residues fell into the disallowed region, as depicted in Figure 4C. VERIFY3D assessed the structure and showed that 69.12% of the residues had an average 3D-1D score of ≥0.1 (Figure 4D). Furthermore, the ERRAT plot estimated a quality score of 88.835% for the developed model (Figure 4E).

Molecular docking

Initially, a High-Throughput Virtual Screening (HTVS) was conducted on the compound library previously screened for target proteins, and the top 10% compounds with the highest scores were selected using the Schrödinger Glide Grid module. This was followed by a Standard Precision (SP) screening, where compounds were ranked again based on their Glide Scores, and the top 10% were further identified. Lastly, an Extra Precision docking (XP) was carried out, and the top 10% compounds

were retained according to the Schrödinger Glide Grid module's criteria. In the third and final round of screening, which involved flexible docking, the compounds from the previous step were flexibly coupled, and the top 10 compounds with the highest coupling scores were identified as the final hits of this study. The virtual screening process and its results are illustrated in Figure 5.

In molecular coupling, the efficiency of a ligand is measured by its binding affinity to the target receptor protein, with the most efficient ligand having the lowest binding energy. To locate the active binding sites for target protein optimization, we used the online POCASA software. Based on the volume, 161 of the five active pockets were selected as the binding site (Figure 4F). Compounds from the TCM libraries were screened against the CsgD receptor, a DNA-binding transcriptional regulator, using the ligand-mounding tool and binding pocket parameters. The top ten compounds were selected based on their docking scores (Table 2).

These 10 selected compounds were manually visualized to verify their complex protein-compound interactions. The interactions of the docked hit compounds with CsgD were visualized using Discovery Studio 3.5 (Figure 6). The top 10 compounds in molecular coupling (Table 3) include hydrogen bonds between amino acids such as naringin-4-glucoside from the Chinese herb Wumei and GLU99, SER91, ALA92, ARG216, GLU159, ARG111, and GLN93 from the CsgD protein. Ingredients such as mydriatin, epiafzelechin, and ephedrine from the traditional Chinese medicine ephedra form hydrogen bonds with amino acids such as LYS94, GLU96, SER155, ASN70, LYS94, and SER155 in CsgD, respectively. Furthermore, mydriatin had a salt-bridging effect with GLU96 amino acids in CsgD. The astragalin, rengyoside C, and forsythide methyl ester components obtained from the Chinese herb Forsythia suspensa have intermolecular hydrogen bonds with amino acids such as GLU159, ASN70, GLU96, SER91, ALA92, SER65, GLN93, GLU99, SER155, LEU157, and LYS67, respectively. Furthermore, there is a salt bridge effect between astragalin and GLU96, and other compounds derived from the

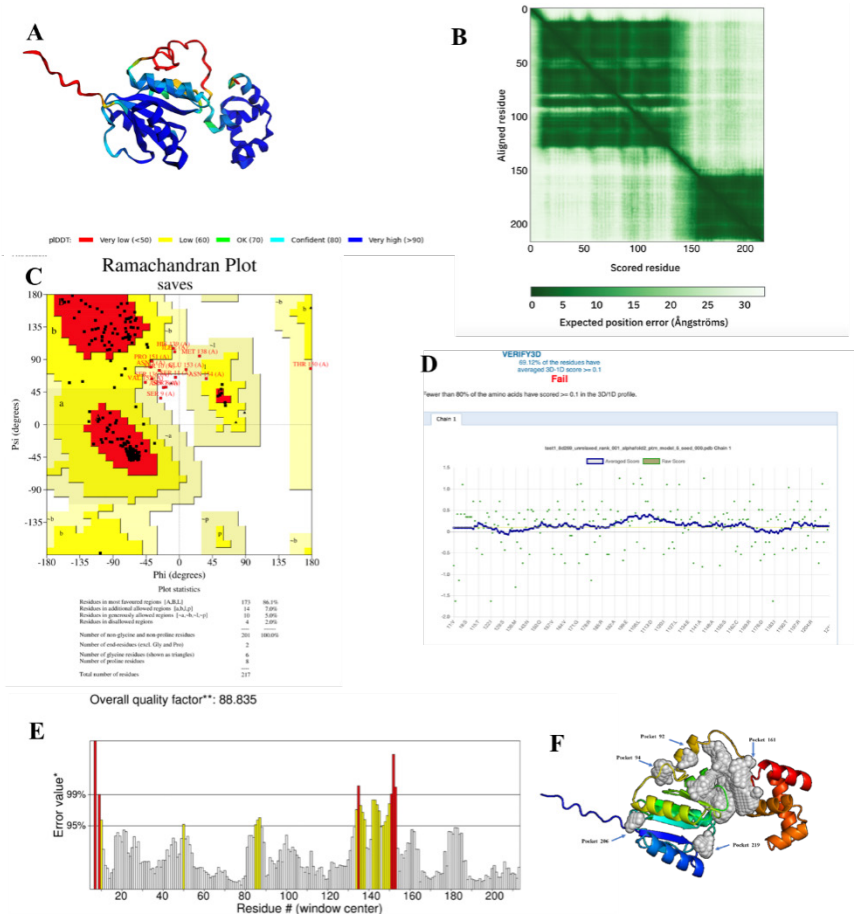

Figure 4: Presents the following: (A) The modeled structure of CsgD generated by AlphaFold2. (B) The anticipated alignment error of the model, where dark green signifies low error and light green denotes high error. (C) Validation of the model's structure through the Ramachandran plot using PROCHECK, revealing that 86.1% of the residues lie within the preferred region. (D) The VERIFY3D analysis confirms the protein's structure. (E) The ERRAT graph indicates an overall quality score of 88.835% for the model. (F) The active site, depicted as a white region, represents the binding site of the protein as predicted by POCASA.

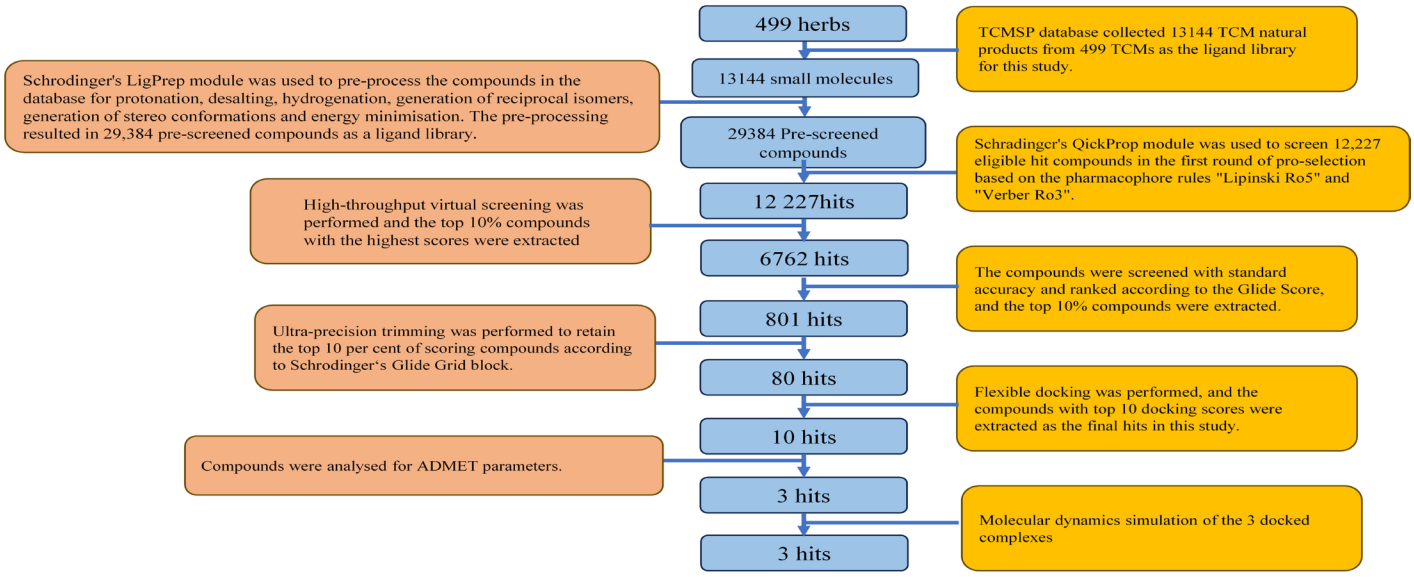

Figure 5: Virtual screening workflow and results.

Table 2: Top 10 docking studies for shortlisted TCM compounds.

Name	Structure	Rotatable bonds	Docking score	h-bond	Glide energy
Naringin 4'-glucoside		20	-7.368	0	-73.131
Mydriatin		3	-7.234	-0.656	-28.976
Astragalin		11	-7.19	-0.48	-59.644
Pinoresinol 4-O-glucoside		15	-7.005	-0.555	-61.673
Lopac-I-3766		6	-6.86	-0.712	-36.072
Rengyoside C		15	-6.857	-0.578	-53.372
Epiafzelechin		5	-6.838	-0.87	-36.686
Forsythidmethylester		10	-6.754	-0.178	-47.381
Ephedrine		4	-6.609	0.656	-30.392
Alpha-glucose		6	-6.565	0	-30.649

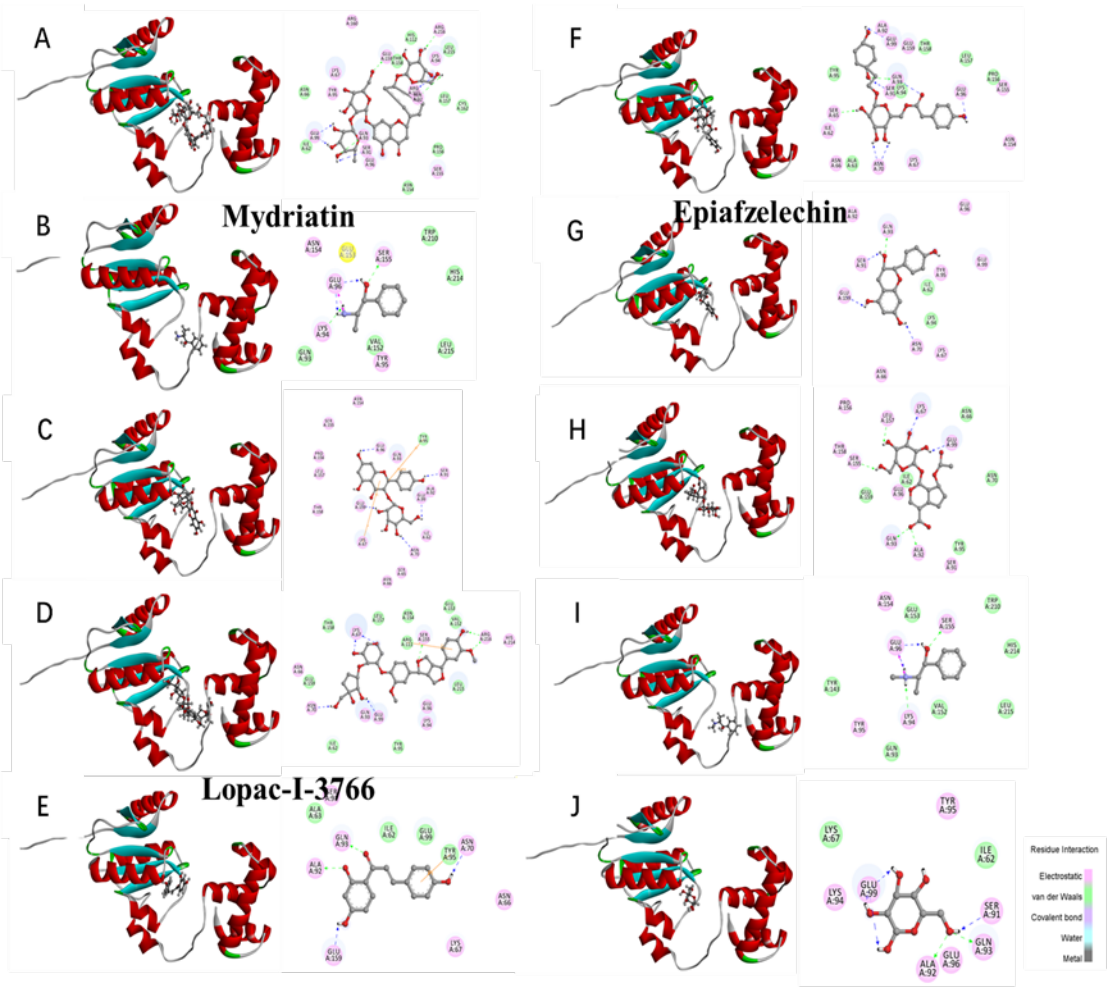

Figure 6: 2D and 3D images of the coupling between CsgD and the top 10 compounds.

Table 3: The top 10 interactions and their interacting residues.

Ligands Name	Herb name	H-Bonds	Salt Bridges	Pi-Pi
Naringin 4'-glucoside	wumei	GLU99 SER91 ALA92 ARG216 GLU159 ARG111 GLN93	--	--
Mydriatin	mahuang	LYS94 GLU96 SER155	GLU96	--
Astragalin	lainqiao	GLU159 ASN70 GLU99 GLU96 SER91 ALA92	--	TYR95
Pinoresinol 4-O-glucoside	banlangen	ASN70 GLU99 ARG216 SER155 LYS67	--	--
Lopac-I-3766	jixueteng	GLU159 GLN93 ALA92 ASN70	--	--
Rengyoside C	lianqiao	SER65 ASN70 GLU96 SER91 GLN93	--	--
Epiafzelechin	mahuang	ASN70	--	--
Forsythidmethylester	lianqiao	GLU99 SER155 LEU157 LYS67 GLN93	--	--
Ephedrine	mahuang	LYS94 GLU96 SER155	--	--
Alpha-glucose	banlangen	GLU99 SER91 GLN93	--	--

Table 4: ADME profiling of shortlisted TCM compounds.

Sl. No	Compounds	Water solubility	$Caco_2$ permeability	HIA	Skin permeability	BBB	Lipinski
0	Carbapenems	-0.346	1.12	87.418	-2.735	-0.312	Yes
1	Naringin 4'-glucoside	-2.814	-0.673	0	-2.735	-2.18	NO
2	Mydriatin	-1.213	1.167	88.229	-2.712	-0.245	Yes
3	Astragalin	-2.863	0.306	48.052	-2.735	-1.514	NO
4	Pinoresinol 4-O-glucoside	-3.374	0.653	71.027	-2.735	-1.7	Yes
5	Lopac-I-3766	-3.06	0.955	91.096	-2.735	-2.752	Yes
6	Rengyoside C	-3.111	0.372	41.448	-2.737	-1.357	Yes
7	Epiafzelechin	-3.254	1.077	91.482	-2.735	-0.818	Yes
8	Forsythidmethylester	-2.501	-0.531	11.514	-2.735	-1.506	Yes
9	Ephedrine	-1.389	1.559	91.464	-2.698	-0.307	Yes
10	Alpha-glucose	-1.377	-0.249	21.51	-3.041	-0.943	Yes

Table 5: Toxicity analysis of the potent compounds identified in TCM.

Sl. No	Compounds	MES toxicity	Max. tolerated dose(log mg/kg/ day/)	Oral rat acute toxicity (LD_{50})	Oral rat chronic toxicity (LOAEL)	Hepatoxicity	Skin sensitisation	*T. pyriformis* toxicity (ug/L)	Minnow toxicity (mM)	Spider web for ADME
1	Naringin 4'-glucoside	NO	0.035	2.434	5.833	NO	NO	0.285	9.019	
2	Carbapenems	NO	1.633	1.53	2.432	Yes	NO	0.272	3.236	
3	Mydriatin	NO	-0.318	2.896	1.38	NO	Yes	0.285	2.026	
4	Astragalin	NO	0.582	2.546	4.53	NO	NO	0.285	6.735	
5	Pinoresinol 4-O-glucoside	NO	-0.478	3.488	5.432	NO	NO	0.285	2.496	--
6	Lopac-I-3766	NO	0.118	2.427	.049	NO	NO	0.285	2.081	
7	Rengyoside C	NO	0.252	3.733	3.677	Yes	NO	0.285	4.926	--
8	Epiafzelechin	NO	0.136	2.365	2.215	NO	NO	0.519	2.75	
9	Forsythid-methylester	NO	0.999	2.36	3.882	NO	NO	0.285	5.768	
10	Ephedrine	NO	-0.359	2.942	1.369	Yes	Yes	-0.023	1.887	
11	Alpha-glucose	NO	1.896	1.214	3.897	NO	NO	0.285	5.083	

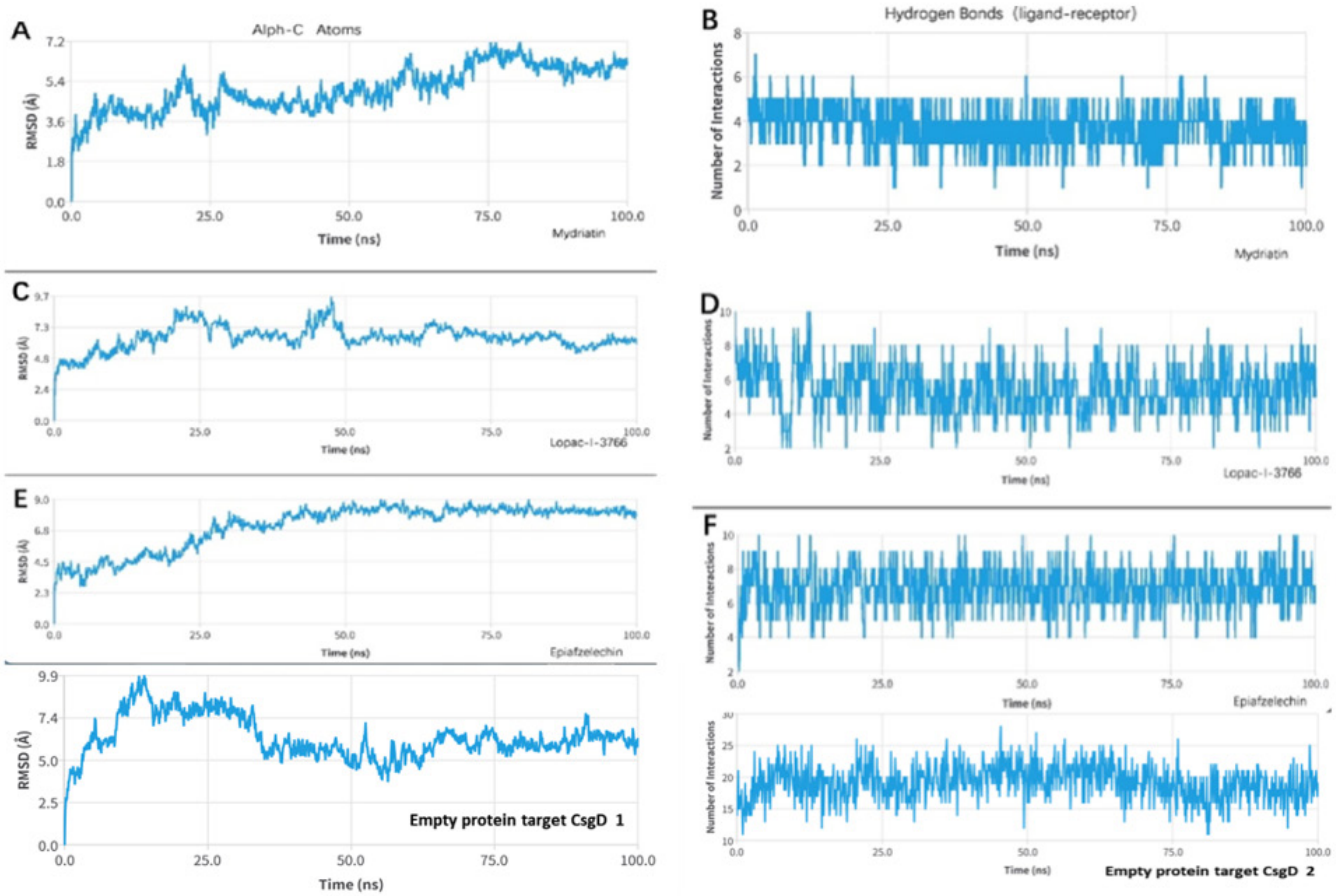

Figure 7: (1) displays the RMSD graphs of the protein and ligands throughout a 100 ns dynamic simulation. (2) Number of hydrogen bonds formed between ligands and the CsgD during dynamic simulation of 100 ns A, B: Mydriatin; C, D: Lopac-I-3766; E, F: Epiafzelechin.

two Chinese herbs, Caulis spatholobi and Radix Isatidis, also have many hydrogen bonds.

ADMET profiling

This phase plays a crucial role in the initial assessment of drug safety during the drug discovery process, aiming to minimize drug toxicity and side effects. Using carbapenems, the last class of antibiotics, as a benchmark, the study found that none of the 10 selected natural products exhibited inhibitory effects on CYP1A2, CYP2C19, CYP2C9, CYP2D6, or CYP3A4 enzymes. These compounds also demonstrated low permeability across the blood-brain barrier but significant permeability through the skin. Evaluation using the Caco2 model showed inhibition of P-glycoprotein. Additionally, these compounds adhered to Lipinski's five rules, suggesting favorable lead- and drug-like properties, as summarized in Table 4. Potential toxic effects of these drugs include toxicity, neurotoxicity, immunotoxicity, mutagenicity, and carcinogenicity. To guarantee safety, the toxicity of the selected molecules was evaluated using the pkSCM online server. The results indicated that, except for rengyoside C and ephedrine, which showed mutagenicity and hepatotoxicity in the AMES test, no other compounds exhibited these effects. Skin sensitization tests were negative for all compounds except mydriatin and ephedrine, as detailed in Table 5.

Mydriatin, Lopac-I-3766, and Epiafzelechin were the three compounds that best met the ADMET and toxicity analysis criteria. Mydriatin shows strong ADMET performance with good water solubility, high $Caco_2$ permeability, good HIA, BBB penetration, and Lipinski rule compliance. It also has low toxicity, with no AMES toxicity, relatively low acute oral toxicity, and no skin sensitization. Lopac-I-3766 performed well in ADMET tests, showing good water solubility, $Caco_2$ permeability, HIA, BBB penetration, and Lipinski rule compliance. It also has low toxicity, with no AMES toxicity, moderate acute oral acute toxicity, and no hepatotoxicity. Epiafzelechin exhibits favorable ADMET properties with good water solubility, $CaCo_2$ permeability, HIA, BBB penetration, and Lipinski rule compliance. It also has low toxicity, with no AMES toxicity, moderate acute oral toxicity, and no skin sensitization. These three compounds appear to be the best candidates based on the balance of ADMET profiles.

Molecular dynamics simulation studies

To verify the stability of ligand binding to the protein, we calculated the Root Mean Square Deviation (RMSD) of the

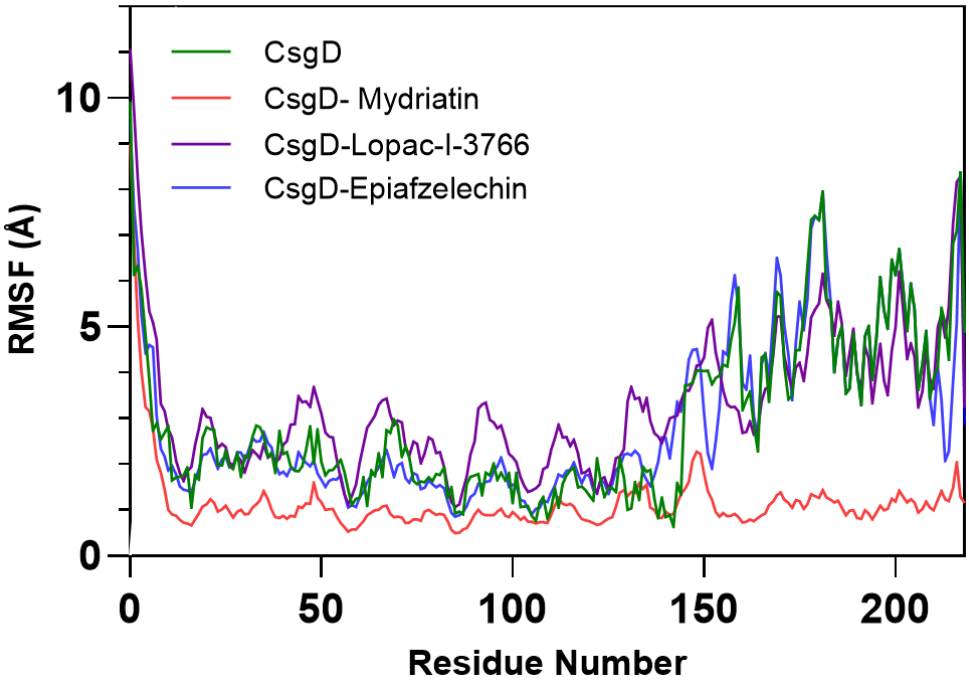

Figure 8: Fluctuations in protein residues during 100 ns dynamic simulations determined by RMSF values.

protein backbone and the small ligand molecule conformation based on the superposition of the protein backbone in 100 ns trajectories of the three ligand-protein complexes (Figure 7 (1)). The RMSD of the isoprotein backbone of the Lopac-I-3766 ligand and epiafzelechin in the system was stable below 2 after 50 ns, the simulation trajectory reached equilibrium, and the protein and ligand binding was stable. It is important to note that small molecules such as mydriatin exhibit large fluctuations in RMSD in the ligand conformation.

Hydrogen bonds play an important role in biological macromolecules and are one of the main driving forces of protein-ligand interactions. They play an indispensable role in the stabilization of protein-ligand complexes.[40] The number of hydrogen bonds between the ligand and the protein was relatively stable after 25 ns, and the number of Lopac-I-3766 and epiafzelechin ligands was greater than that of mydriatin. The more hydrogen bonds there are, the more stable the bond between the ligand and the protein. As the protein-ligand binding in Figure 7 (2) shows, the structural fragments that play a key role are concentrated at one end of the ligand, while the other end is exposed outside the binding pocket. When the molecular weight of the ligand is large and the molecular skeleton is long, there are more parts outside the binding region, and this part moves more freely, which may be the reason for the relatively large RMSD fluctuations of some small molecules.

To investigate the dynamic behavior of the target protein when interacting with three ligands, we calculated the

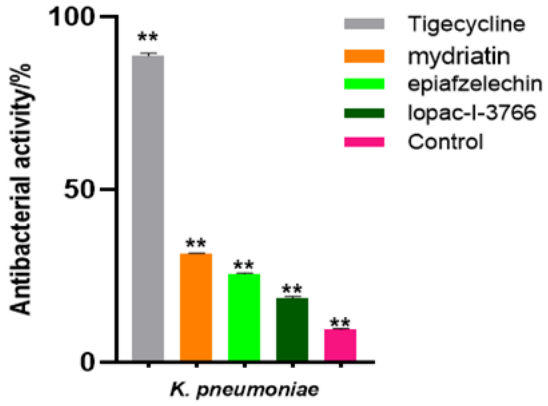

Figure 9: The antibacterial activity of compounds against *K. pneumoniae*.

Root-Mean-Square Fluctuations (RMSFs). This analysis offers insights into the flexibility of individual residues within a protein. The detailed findings are illustrated in Figure 8. An RMSF analysis of the protein CsgD with 217 inhibitors showed that the three protein-inhibitor complexes exhibited similar patterns in the numerical distribution of RMSFs. Specifically, the residues at the binding site underwent significant fluctuations within the range of 150-185. Compared to mydriatin, epiafzelechin and Lopac-I-3766 induced greater variations in the binding site residues. The overall fluctuations indicated that the protein-ligand complex was stable. Generally, during the 100-ns dynamic simulation, the transcriptional regulator protein CsgD was found to be stable when combined with these three compounds.

Table 6: MIC determination results of the 3 compounds (μg/mL).

Compounds	Strains No.	Concentration (μg/mL)					MIC range
		62.500	125.000	250.000	500.000	1000.000	
Mydriatin	30	0	12	12	6	0	125～500
Epiafzelechin	30	0	0	12	12	6	250～1000
Lopac-I-3766	30	0	0	4	18	8	250～1000

Utilizing the PASS tool to predict the biological activity of the three compounds

The identified hit compounds were searched for structural similarity, and their biological activities were estimated using the Prediction of Activity Spectrum of Substances (PASS) tool. The results of biological activity prediction of the selected compounds based on the ligand structural fragments are shown in the Supplementary Data Files Figure S1, where Pa is the probability of having the corresponding biological activity and Pi is the probability of non-biological activity. From the above results, it can be seen that the probabilities of being Active (Pa) were compared with the structures to determine the most active molecules for the predicted biological activity. The highest biological activity was predicted to be the membrane integrity agonist activity of the three compounds, with likelihood >0.8.

Antibacterial *in vitro* activity test

Evaluation of antibacterial activity

After conducting molecular docking and molecular dynamics analyses, mydriatin, epiafzelechin, and lopac-I-3766 were chosen for further evaluation of their antibacterial activity. The broth microdilution method was employed to test their activity against *K. pneumoniae* (ATCC 700603), following the guidelines set by the Clinical and Laboratory Standards Institute (CLSI). The results, presented in Figure 9, indicated that at a concentration of 200 μg/mL, there was a notable difference in the inhibitory activity of *K. pneumoniae* among the three compounds. Mydriatin exhibited the most promising effect, with an inhibition rate of approximately 25%, while the other two compounds showed inhibition rates of around 10%.

(The concentration of each compound treatment is 200 μg/mL and the error bars represent the mean±SD; the treatment group was subjected to one-way ANOVA with the corresponding control group, ** $p<0.01$, * $p<0.05$.)

Minimum Inhibitory Concentration (MIC) for 3 compounds

The three compounds, mydriatin, epiafzelechin, and lopac-I-3766, were detected by the broth microdilution method against 30 clinically isolated multidrug-resistant *K. pneumoniae* strains. The Minimum Inhibitory Concentrations (MIC) of the *K. pneumoniae* strains were determined, and the results showed that mydriatin, epiafzelechin, and lopac-I-3766 had significant antibacterial effects on the *K. pneumoniae* strains (Table 6). Mydriatin showed the best antibacterial effect, with an MIC of 125-500 μg/mL, and 80% of the strains had an MIC of 125-250 μg/mL. Although more *in vitro* experiments need to be carried out to confirm the inhibitory effects of mydriatin on *K. pneumoniae*, the antimicrobial activities of the compounds are consistent with virtual screening and molecular modelling, which also indicates that mydriatin is a promising compound. The antibacterial activities of the other two compounds were similar, with MICs of 250～1000 μg/mL. In the positive control group, the solution was clarified, which indicated that the growth was sterile and tigecycline was qualified. In the negative control group, the growth of *K. pneumoniae* was good, indicating that the bacterial solution was qualified, and the solution in the medium control group was clarified, which indicated that the presence of *K. pneumoniae* made the solution turbid, and the inoculation of *K. pneumoniae* in the diluted medium containing DMSO showed good growth, indicating that DMSO did not inhibit the growth of *K. pneumoniae*.

DISCUSSION

There is currently a growing interest in the application of computational methods and approaches for the discovery and development of effective drug targets.[41] However, high-throughput experimental sequencing data for most infectious bacteria are still unavailable. Consequently, the elucidation and identification of key drug targets are mainly based on bioinformatic predictions. Given the increasing prevalence of drug resistance in pathogens, *in silico* subtractive genomic analysis has emerged as a widely used strategy for precisely identifying strain-specific drug targets.[42,43] Contemporary studies have adopted subtractive proteomics to reveal and predict novel drug candidates. The scope of this research focused on one of the most important clinical pathogens, *K. pneumoniae*. Widely recognized for its effectiveness, this strategy has been crucial in the search for new drug targets against various dangerous pathogens.

To decipher the core proteome and uncover novel, potent therapeutic targets, a subtractive genome analysis was conducted on the core genome. This study identified approximately 1041 proteins that do not share homology with the human proteome,

among which 895 were deemed essential for the survival of the pathogen. Among these, the transcriptional regulator CsgD emerged as a promising therapeutic target. As a pivotal transcriptional regulator involved in biofilm formation, CsgD plays a vital role in the survival and proliferation of bacteria.[44-46] Previous studies have recognized CsgD as a potential drug target for various pathogens, such as Escherichia coli and Salmonella.[47] Furthermore, CsgD plays a crucial role in stress resistance and represents a potential target for treating and controlling biofilm formation.[48] Therefore, interventions targeting CsgD may provide important targets for the development of novel antibacterial therapies, particularly with potential application value in controlling drug-resistant bacterial biofilms.The protein structure of CsgD was modeled using AlphaFold2 and subsequently validated using PROCHECK, ERRAT, and VERIFY3D tools.

Previous studies have recognized CsgD as a potential drug target for various pathogens, such as *Escherichia coli* and *Salmonella*.[47] Furthermore, CsgD plays a crucial role in stress resistance and represents a potential target for treating and controlling biofilm formation.[48] The protein structure of CsgD was modeled using AlphaFold2 and subsequently validated using PROCHECK, ERRAT, and VERIFY3D tools.

In recent years, research on the antibacterial effects of TCM has increased gradually. As early as the 1980s, studies found that TCM could eliminate bacterial resistance to antibiotics. Currently, heat purification and detoxification medicines are the most effective TCMs for eliminating bacterial resistance, mainly honeysuckle, Scutellaria baicalensis, isatis root, Forsythia suspensa, bupleurum, shegan, dandelion, and Houttuynia cordata.[42,43] Given the vast number of Traditional Chinese Medicines (TCMs) and the intricacy of their constituents, pinpointing those with antibacterial properties through conventional methods is extremely challenging. High-throughput screening offers a substantial enhancement in the efficiency of active compound identification.[49] Additionally, virtual screening can quickly narrow the selection of target molecules from a large number of component libraries.The main components extracted from the Chinese herbs, as reported in TCMSP,[50] were then subjected to virtual screening. First, we followed the Lipinski Ro5 and Verber Ro3 Principles to filter out TCM compounds that violate the principle of drug likeness. Glide software was used to perform molecular docking calculations with three different accuracies, namely high-throughput HTVS, standard precision SP, and ultra-precision XP, gradually reducing the size of the candidate compound library. The top 10 compounds characterized by docking scores ranging from -6.6 to -7.4 were identified as promising candidates.

In the initial phases of drug discovery, computational profiling of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) is pivotal for monitoring the hit-to-lead discovery process and subsequent lead optimization. Schrodinger's QikDrop program was utilized to predict the ADMET properties of the hit molecule. A comprehensive drug-like evaluation was conducted by incorporating predictions based on both Lipinski's Rule of Five (Ro5) and Verber's Rule of Three (Ro3). The ADMET profile of these compounds assessed their ADME properties, non-toxicity, and drug-likeness, ensuring their safe use in both *in vitro* and *in vivo* studies. Among the evaluated compounds, mydriatin, Lopac-I-3766, and epiafzelechin emerged as the top three that met the ADMET and toxicity analysis standards. While 100-nanosecond molecular dynamics simulations of these protein-ligand complexes revealed their stability after approximately 25 nanoseconds, complementing our docking analysis and yielding acceptable ADMET parameters, it's important to note that the predicted results may still contain errors due to potential differences in compound behavior in natural environments. Therefore, laboratory validation and *in vivo* expression of these compounds against *K. pneumoniae*, as well as studies on their effects on human immune responses, are necessary to complement and validate the computational analysis. However, the application of advanced virtual screening technology and Molecular Dynamics (MD) simulations to demonstrate the effect of Traditional Chinese Medicine (TCM) monomers on *K. pneumoniae* has enhanced drug screening efficiency, reduced experimental costs, accelerated the modernization of TCM, and promoted its internationalization.

Mydriatin and Epiafzelechin are ephedrine and flavanol compounds of the Chinese medicinal herb Mahuang. Mahuang has antipyretic, antibacterial, and antiviral effects, while epiafzelechin has antioxidant and antibacterial effects.[51] Lopac-I-3766 is a bioactive component of the Chinese medicinal herb, Jixueteng. Jixueteng has anti-inflammatory, disinfectant, and antitumor effects, which were further tested for MIC using a broth microdilution test. The results showed that Mydriatin, Epiafzelechin and Lopac-I-3766 had significant antibacterial effects against clinical isolates of *K. pneumoniae* with MIC values ranging from 250 to 1000 μg/mL. Mydriatin had the best antibacterial effect, with an MIC of 125-500 μg/mL, against clinical isolates of *K. pneumoniae. In vitro* descriptors also supported the virtual screening approach. Therefore, Mydriatin can be used as a potential drug candidate against the pathogenic bacterium *K. pneumoniae*.

At present, the specific mechanism of the direct antibacterial effect of TCM is not clear, and current research results fall mainly into four categories:[52,53] changing the permeability of cell membranes; inhibiting the activity of enzymes in bacteria, thereby interfering with their metabolism; inhibiting the synthesis of proteins and nucleic acids; oxidising the active functional groups in bacteria. The confirmed structure of the 3 compounds was subjected to the computer program PASS for biological activity. The Probabilities of being active (Pa) were compared with those of the structures to determine the most active molecules for predicted biological

activity. The highest biological activity was predicted to be the membrane integrity agonist activity of the three compounds. In fact, it has been proposed that the antibacterial effects of some TCMs could be explained mainly by their ability to cause direct damage to the bacterial membrane[54] or at least as a consequence of destabilizing it to allow them access to their intracellular targets.[55] Further research is needed to elucidate the specific antibacterial mechanisms of these TCMs.

CONCLUSION

Given the increasing resistance of *K. pneumoniae t*o existing antibiotics, the search for novel drug candidates is ongoing. In the present study, a novel anti-*K. pneumoniae* drug target CsgD was predicted using pan-genomics and subtractive proteomics. The molecular docking technique was used to virtually screen and a series of TCM compounds against the CsgD. Three highly potent drug candidates were identified as the most promising lead. *In silico* ADMET analysis of the selected ligands showed positive pharmacokinetic and toxicity profiles. MD simulations of the most stable complexes showed that ligand compound 1 (mydriatin) bound well to the target protein CsgD in a dynamic manner. Subsequently, *in vitro* antibacterial experiments showed that mydriatin significantly inhibited *K. pneumoniae*. Therefore, mydriatin can be used as a potential drug candidate against the pathogenic bacteria *K. pneumoniae.* Our findings represent a valuable asset in the discovery and rational design of novel antibiotics to combat the challenges posed by multidrug resistance in *K. pneumoniae*.

ACKNOWLEDGEMENT

The authors thank anonymous reviewers for their valuable suggestions.

CONFLICT OF INTEREST

The authors declare that there is no confict of interest.

AUTHOR CONTRIBUTIONS

Kumar conceptualized and designed this study. Yanping Li performed bioinformatics analysis and *in vitro* identification and wrote the original manuscript. Hao Peng participated in the antibacterial in vitro activity test. Lihu Zhang supervised and edited the manuscript and provided funding. All authors contributed to the editorial changes in the manuscript. All the authors have read and agreed to the published version of the manuscript. All authors participated sufficiently in the work and agreed to be accountable for all aspects of the work.

SUMMARY

Klebsiella pneumoniae, a prevalent opportunistic pathogen, threatens hospitalized patients with various infections and growing antibiotic resistance. Through pan-genomic analysis, we identified CsgD as a drug target and virtually screened 29,384 natural compounds from Traditional Chinese Medicine libraries. Among them, mydriatin emerged as a potent CsgD inhibitor, demonstrating significant inhibitory effects on antibiotic-resistant *K. pneumoniae* strains *in vitro*. This study suggests mydriatin as a promising therapeutic agent and underscores the importance of natural product libraries and computational methods in antibiotic discovery against multidrug-resistant *K. pneumoniae*.

REFERENCES

1. Paczosa MK, Mecsas J. Klebsiella pneumoniae: Going on the Offense with a Strong Defense. Microbiol Mol Biol Rev. 2016;80(3):629-61. doi: 10.1128/MMBR.00078-15, PMID 27307579.
2. YL, SK, Id O, LZ, HW, H W. Characteristics of antibiotic resistance mechanisms and genes of klebsiella. D - 101672167. 2023(- 2391-5463 (Print)): 20230707.
3. Shrivastava SR, Shrivastava PS, Ramasamy J. World Health Organization releases global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. J Med Soc. 2018;32(1):76-. doi: 10.4103/jms.jms_25_17.
4. Giordano C, Barnini S, Tsioutis C, Chlebowicz MA, Scoulica EV, Gikas A, *et al.* Expansion of KPC-producing Klebsiella pneumoniae with various mgrB mutations giving rise to colistin resistance: the role of ISL3 on plasmids. Int J Antimicrob Agents. 2018;51(2):260-5. doi: 10.1016/j.ijantimicag.2017.10.011, PMID 29097338.
5. Li Y, Kumar S, Zhang L, Wu H. *Klebsiella pneumonia* and Its antibiotic Resistance: A bibliometric Analysis. BioMed Res Int. 2022; 2022:1668789. doi: 10.1155/2022/1668789, PMID 35707374.
6. Basharat Z, Jahanzaib M, Yasmin A, Khan IA. Pan-genomics, drug candidate mining and ADMET profiling of natural product inhibitors screened against Yersinia pseudotuberculosis. Genomics. 2021; 113(1 Pt. 1):238-44. doi: 10.1016/j.ygeno.2020.12.015, PMID 33321204.
7. Kumar S. Computational identification and binding analysis of orphan human cytochrome P450 4X1 enzyme with substrates. BMC Res Notes. 2015;8(1):9. doi: 10.1186/s13104-015-0976-4, PMID 25595103.
8. Barbosa F, Pinto E, Kijjoa A, Pinto M, Sousa E. Targeting antimicrobial drug resistance with marine natural products. Int J Antimicrob Agents. 2020;56(1):106005. doi: 10.1016/j.ijantimicag.2020.106005, PMID 32387480.
9. Rossiter SE, Fletcher MH, Wuest WM. Natural products as platforms to overcome antibiotic resistance. Chem Rev. 2017;117(19):12415-74. doi: 10.1021/acs.chemrev.7b00283, PMID 28953368.
10. Wayne PA, CLINICAL AND LABORATORY STANDARDS INSTITUTE. Performance standards for antimicrobial susceptibility testing; 2011.
11. Ding W, Baumdicker F, Neher RA. panX: pan-genome analysis and exploration. Nucleic Acids Res. 2018;46(1):e5. doi: 10.1093/nar/gkx977, PMID 29077859.
12. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics -OXFORD-; 2010.
13. Rolf A, Amos B, Wu CH, *et al.* UniProt: the Universal Protein knowledge base. Nucleic Acids Res. 2004;32((Database issue)):D115-119.
14. 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(1):19842. doi: 10.1038/srep19842, PMID 26806463.
15. Ren Z, DEG YL. 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res. 2009;37((Database issue)):D455-458.
16. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2022;50(D1):D912-7. doi: 10.1093/nar/gkab1107, PMID 34850947.
17. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. Oxford University Press; 2018.
18. Ying Z, Yintao Z, Xichen L, *et al.* Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. Nucleic Acids Res. 2022.
19. Shanmugham B, Pan A. Identification and Characterization of Potential Therapeutic Candidates in Emerging Human Pathogen Mycobacterium abscessus: A Novel Hierarchical in silico Approach. PLOS One. 2013;8(3):e59126. doi: 10.1371/journal.pone.0059126, PMID 23527108.
20. Nagasuma C, Kalidas Y, Karthik R. targetTB: A target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis. BMC Syst Biol. 2008;2(1):1-21.
21. Khan MT, Mahmud A, Iqbal A, Hoque SF, Hasan M. Subtractive genomics approach towards the identification of novel therapeutic targets against human *Bartonella bacilliformis*. Inform Med Unlocked. 2020;20:100385. doi: 10.1016/j.imu.2020.100385.

22. Ammari MG, Gresham CR, Mccarthy FM, Bindu N, HPIDB. HPIDB 2.0: a curated database for host-pathogen interactions [database]. Database (Oxford). 2016; 2016:baw103. doi: 10.1093/database/baw103, PMID 27374121.
23. Durmuş Tekir S, Çakır T, Ardiç E, Sayılırbaş AS, Konuk G, Konuk M, *et al*. ÇT. PHISTO: pathogen-host interaction search tool. Bioinformatics. 2013;29(10):1357-8. doi: 10.1093/bioinformatics/btt137, PMID 23515528.
24. Martin U, Rashmi P, Arathi R, Irvine AG, Helder P, Hammond-Kosack KE. The pathogen-host Interactions database (PHI-base): additions and future developments. Nucleic Acids Res. 2015d1:645-60.
25. Gomaa EZ. Human gut microbiota/microbiome in health and diseases: a review. Antonie Leeuwenhoek J Microbiol Serology. 2020;113(12):2019-40. doi: 10.1007/s10482-020-01474-7, PMID 33136284.
26. Sha S, Ni L, Stefil M, Dixon M, Mouraviev V. The human gastrointestinal microbiota and prostate cancer development and treatment. Investig Clin Urol. 2020; 61 Suppl 1:S43-50. doi: 10.4111/icu.2020.61.S1.S43, PMID 32055753.
27. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608-15. doi: 10.1093/bioinformatics/btq249, PMID 20472543.
28. Alcock BP, Raphenya AR, Lau TT, Tsang KK, Mcarthur AG. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2019;48(D1).
29. Jelle S, Rieza A, Jan-Willem V. Deep genome annotation of the opportunistic human pathogen Streptococcus pneumoniae D39. Nuclc Acids Res. 2018;19:19.
30. Nuka G, Potter S, Yong SY, *et al.* InterProScan 5: Large scale protein function classification; 2016.
31. David A, Islam S, Tankhilevich E, Sternberg MJ. The AlphaFold database of protein structures: a biologist's guide. J Mol Biol. 2022;434(2):167336. doi: 10.1016/j.jmb.2021.167336, PMID 34757056.
32. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al.* Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. doi: 10.1038/s41586-021-03819-2, PMID 34265844.
33. Pink. Shaw: RK, *et al.* Fresh fruit and vegetables as vehicles for the transmission of human pathogens.
34. Hossain T, Kamruzzaman M, Choudhury TZ, Mahmood HN, Nabi AH, Hosen MI. Application of the subtractive genomics and molecular docking analysis for the identification of novel putative drug targets against Salmonella enterica subsp. enterica serovar Poona. BioMed Res Int. 2017; 2017:3783714. doi: 10.1155/2017/3783714, PMID 28904956.
35. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, et al. *In silico* subtractive genomics for target identification in human bacterial pathogens. Drug Dev Res. 2011;72(2):162-77. doi: 10.1002/ddr.20413.
36. Masomian M, Ahmad Z, Gew LT, Poh CL. Development of next generation *Streptococcus pneumoniae* vaccines conferring broad protection. Vaccines. 2020;8(1):132. doi: 10.3390/vaccines8010132, PMID 32192117.
37. Dym O, Eisenberg D, Yeates TO. Detection of errors in protein models: international Tables for crystallography; 2012.
38. Eisenberg D, Lüthy R, Bowie J, Luthy R, Eisenberg D, Dr VERIFY E. 3D: assessment of protein models with three-dimensional profiles. A method to identify protein sequences that fold into a known three-dimensional structure; 1997.
39. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26(2):283-91. doi: 10.1107/S0021889892009944.
40. Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, *et al.* Insights into protein-ligand interactions: mechanisms, models, and methods. Int J Mol Sci. 2016;17(2):144. doi: 10.3390/ijms17020144, PMID 26821017.
41. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E *et al.* The Swiss-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003;31(1):365-70. doi: 10.1093/nar/gkg095, PMID 12520024.
42. Cai W, Fu Y, Zhang W, Chen X, Zhao J, Song W, *et al.* Synergistic effects of baicalein with cefotaxime against *Klebsiella pneumoniae* through inhibiting CTX-M-1 gene expression. BMC Microbiol. 2016;16(1):181. doi: 10.1186/s12866-016-0797-1, PMID 27502110.
43. Yun-Ning L, Xiao-Feng LI, Xu-Xia B. Pharmacy DO. The review on active antibacterial ingredients of Chinese medicine and the antibacterial mechanism. Glob Trad Chin Med. 2015.
44. Cristina, Solano, Begoña, *et al.* Genetic analysis of *Salmonella enteritidis* bio film formation: critical role of cellulose. Mol Microbiol. 2002.
45. Gerstel U, Römling U. Oxygen tension and nutrient starvation are major signals that regulate AGFD promoter activity and expression of the multicellular morphotype in *Salmonella typhimurium*. Environ Microbiol. 2001;3(10):638-48. doi: 10.1046/j.1462-2920.2001.00235.x, PMID 11722544.
46. Liu Z, Niu H, Wu S, Huang R. CsgD regulatory network in a bacterial trait-altering biofilm formation. Emerg Microbes Infect. 2014;3(1):e1. doi: 10.1038/emi.2014.1, PMID 26038492.
47. Uddin RR, Rafi S. Structural and functional characterization of a unique hypothetical protein (WP_003901628.1) of Mycobacterium tuberculosis: a computational approach. Med Chem Res. 2017;26(5):1029-41. doi: 10.1007/s00044-017-1822-0.
48. Yan CH, Chen FH, Yang YL, Zhan YF, Herman RA, Gong LC, *et al.* The transcription factor CSGD contributes to engineered Escherichia coli resistance by regulating biofilm formation and stress responses. Int J Mol Sci. 2023;24(18):13681. doi: 10.3390/ijms241813681, PMID 37761984.
49. Henrich CJ, Beutler JA. Matching the power of high throughput screening to the chemical diversity of natural products. Nat Prod Rep. 2013;30(10):1284-98. doi: 10.1039/c3np70052f, PMID 23925671.
50. Ru J, Li P, Wang J, Zhou W, Li B, Huang C, *et al.* TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Cheminform. 2014;6:13. doi: 10.1186/1758-2946-6-13, PMID 24735618.
51. Iftikhar A, Falodun A, Josephs GC, Hussain H, Langer P. Characterization and Antimicrobial Evaluation of Epiafzelechin from the Stem Bark of *Calliandra surinamensis* Benth. 2013.
52. Zhengrong W. Mechanism study on the effect of Shuanghuanglian powder injection on the drug resistance of multidrug-resistant *Escherichia coli* Beijing University of Traditional Chinese Medicine; 2013.
53. Chen Guoping SL, Qingfeng F, Lv S, Yonglin S, Jin Z, *et al.* Research progress on the influencing factors and control measures of severe hand foot mouth disease Public Health and Preventive Medicine. 2019;30(4):5.
54. Ming, Liu, Wei, *et al.* The direct anti-MRSA effect of emodin via damaging cell membrane. Applied Microbiology and Biotechnology; 2015.
55. Alves DS, Perez FL, Estepa A, Micol V. Membrane-related effects underlying the biological activity of the anthraquinones emodin and barbaloin. Biochem Pharmacol; 2004(3):68.

Cite this article: Li y, Kumar s, Zhang l. Identification of Novel Drug Targets and Potential Antibacterial Traditional Chinese Medicine Compounds for Klebsiella pneumoniae through in silico and Antibacterial Activity Evaluation. Indian J of Pharmaceutical Education and Research. 2025;59(1s):s256-s273.

Table S1: The amino acid sequence of the transcriptional regulator CsgD (KPHS_44520) was compared with the CARD database and the annotation of the drug resistance gene (CARD) was analyzed. The results of the CARD analysis are displayed in the supplementary data files.

card-result

Score	Acccession	Name	Evalue	Identity	Species
67	3003792	Enterococcus faecalis liaR mutant conferring daptomycin resistance	6.62058E-15	39	Enterococcus faecalis V583
63	3003078	Enterococcus faecium liaR mutant conferring daptomycin resistance	1.58532E-13	38	Enterococcus faecium DO
38	3000826	sdiA	0.00017804	41	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2
35	3003893	Escherichia coli uhpA with mutation conferring resistance to fosfomycin	0.00239421	37	Escherichia coli str. K-12 substr. MC4100
35	3003893	Escherichia coli uhpA with mutation conferring resistance to fosfomycin	0.00239421	37	Escherichia coli str. K-12 substr. MC4100
32	3000832	evgA	0.0207827	30	Escherichia coli O157:H7 str. Sakai

Mydriatin: C[C@@H]([C@@H](C1=CC=CC=C1)O)N(smiles)

Pa	Pi	Activity
0,919	0,003	Arylacetonitrilase inhibitor
0,907	0,003	NADPH peroxidase inhibitor
0,879	0,016	Membrane integrity agonist
0,866	0,004	Omptin inhibitor
0,861	0,004	Arginine 2-monooxygenase inhibitor
0,856	0,004	Cardiovascular analeptic
0,855	0,005	Nicotinic alpha6beta3beta4alpha5 receptor antagonist
0,862	0,013	Ubiquinol-cytochrome-c reductase inhibitor
0,852	0,005	Glutamyl endopeptidase II inhibitor
0,851	0,004	NADPH-cytochrome-c2 reductase inhibitor

Lopac-I-3766 C1 = CC (= CC = C1 / C = C C (= O) C2 = C (C = C (C = C 2) O) O (smiles)

Pa	Pi	Activity
0,947	0,003	Mucomembranous protector
0,929	0,005	Membrane integrity agonist
0,917	0,003	Feruloyl esterase inhibitor
0,897	0,002	Monophenol monooxygenase inhibitor
0,876	0,008	Chlordecone reductase inhibitor
0,865	0,004	JAK2 expression inhibitor
0,867	0,015	Aspulvinone dimethylallyltransferase inhibitor
0,845	0,018	Ubiquinol-cytochrome-c reductase inhibitor
0,828	0,003	1-Acylglycerol-3-phosphate O-acyltransferase inhibitor
0,826	0,003	APOA1 expression enhancer

Epiafzelechin C1[C@H]([C@H](OC2=CC(=CC(=C21)O)O)C3=CC=C(C=C3)O)O (smiles)

All Pa>Pi Pa>0,3 Pa>0,7 ok

Pa	Pi	Activity
0,984	0,001	Membrane integrity agonist
0,958	0,001	Pectate lyase inhibitor
0,959	0,003	Mucomembranous protector
0,958	0,002	HMOX1 expression enhancer
0,952	0,002	Fibrinolytic
0,952	0,003	TP53 expression enhancer
0,946	0,001	SULT1A3 substrate
0,942	0,002	Reductant
0,936	0,002	UGT1A6 substrate
0,934	0,002	Antimutagenic

Figure S1: Experimental prediction of biological activity of the 3 compounds based on the PASS tool.