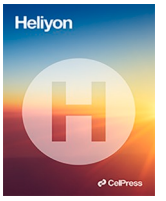
Heliyon 9 (2023) e21466

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Identification of repurposable drug targets in *Mycoplasma pneumoniae* using subtractive genomics, molecular docking and dynamics simulation

Zeshan Mahmud Chowdhury [a,1], Tabassum Binte Jamal [a,1], Ishtiaque Ahammad [a], Arittra Bhattacharjee [a], Anika Bushra Lamisa [a], Jannatul Maoa Jani [b], Md Fahim Israk [b], Mohammad Uzzal Hossain [a], Keshob Chandra Das [c], Chaman Ara Keya [d], Md Salimullah [c,*]

[a] *Bioinformatics Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka, 1349, Bangladesh*
[b] *Department of Biotechnology and Genetic Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, 8100, Bangladesh*
[c] *Molecular Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka, 1349, Bangladesh*
[d] *Department of Biochemistry and Microbiology, North South University, Bashundhara, Dhaka, 1229, Bangladesh*

ARTICLE INFO

Keywords:
Mycoplasma pneumoniae
Subtractive genomics
Essential proteins
Metabolic pathway
Antibiotic resistance

ABSTRACT

Mycoplasma pneumoniae is a significant causative agent of community-acquired pneumonia, causing acute inflammation in the upper and lower respiratory tract as well as extrapulmonary syndromes. In particular, the elderly and infants are at greater risk of developing severe, life-threatening pneumonia caused by *M. pneumoniae*. Yet, the global increase in antimicrobial resistance against antibiotics for the treatment of *M. pneumoniae* infection highlights the urgent need to explore novel drug targets. To this end, bioinformatics approaches, such as subtractive genomics, can be employed to identify specific metabolic pathways and essential proteins unique to the pathogen that could be potential targets for new drugs. In this study, we implemented a subtractive genomics approach to identify 61 metabolic pathways and 42 essential proteins that are unique to *M. pneumoniae*. A subsequent screening in the DrugBank database revealed three druggable proteins with similarity to FDA-approved small-molecule drugs, and finally, the compound CHEBI:97093 was identified as a promising novel putative drug target. These findings can provide crucial insights for the development of highly effective drugs that selectively inhibit the pathogen-specific metabolic pathways, leading to better management and treatment of *M. pneumoniae* infections.

1. Introduction

Mycoplasma pneumoniae is a small, slow-growing bacterium that causes respiratory tract infections. It is a common cause of atypical

* Corresponding author.Chief Scientific Officer and Director General Molecular Biotechnology Division National Institute of Biotechnology, Dhaka, Bangladesh.
E-mail address: dgnib.gov.bd@gmail.com (M. Salimullah).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2023.e21466
Received 11 May 2023; Received in revised form 5 October 2023; Accepted 21 October 2023
Available online 4 November 2023

pneumonia which is also known as "walking pneumonia" [1,2]. The bacterium lacks a cell wall and is classified as a member of the *Mycoplasmataceae* family [3]. *M. pneumoniae* infection is more prevalent in children and adults globally [1,2,4]. The infection accounts for 20–30 % of community-acquired pneumonia in adults and about 40 % in children >5 years old [5,6]. This gram-negative pleomorphic bacterium has the ability to penetrate the host cell membrane and cause direct damage to the host cells [7]. Clinical features of *M. pneumoniae* include sore throat, pharyngitis, hoarseness, fever, and a persistent cough that can yield non-bloody sputum. In severe cases, dyspnea may also be present [1]. The respiratory symptoms may last from several days to a few weeks. Moreover, it is recognized as a trigger for asthma exacerbations, with some patients having broncho-obstructive symptoms [2]. In addition to respiratory diseases, *M. pneumoniae* can generate complications that affect the skin, urogenital tract, certain sensory and digestive organs,

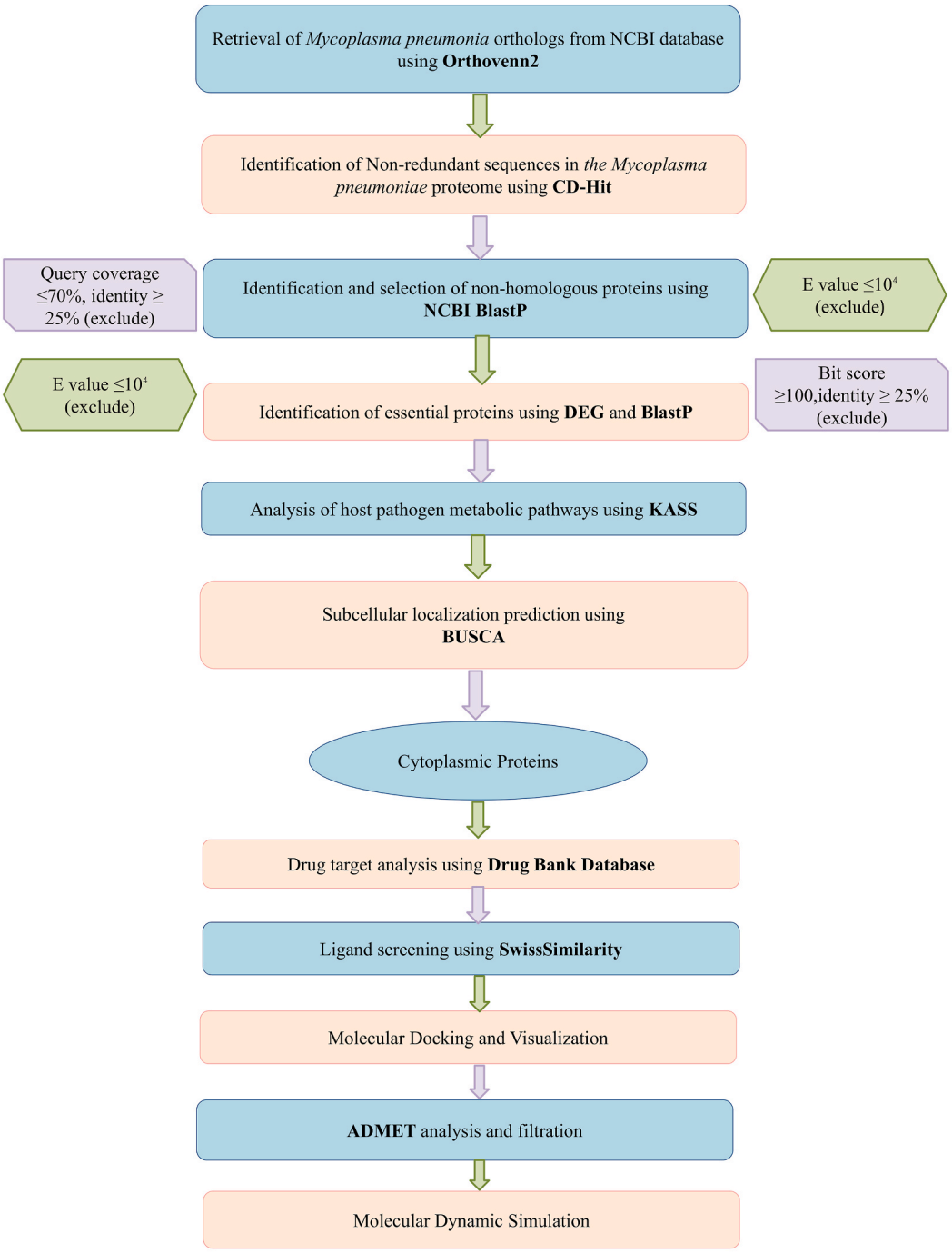


Fig. 1. Schematic representation of the flow chart for potential drug target identification.

cardiovascular, hematopoietic, and musculoskeletal systems, as well as the central nervous system [8,9].

Given the self-limiting nature of *M.pneumoniae* infections, clinicians routinely treat pneumonia with antibiotics like tetracycline, fluoroquinolones, and macrolide. Due to the lack of cell walls, antimicrobial agents like beta-lactam and glycopeptides are ineffective against this bacterium [10]. Tetracycline and macrolide families inhibit protein synthesis by acting on the bacterial ribosome. Fluoroquinolones, on the other hand, prevent DNA replication via binding to DNA gyrase. However, due to limited information on the safety of these antibiotics in children, macrolides have been the only preferred choice for treating *M. pneumoniae* infections in children [11,12]. Nevertheless, the emergence of macrolide-resistant *M. pneumoniae* (MRMP) strains highlights the importance of the prudent use of these drugs. The rates of MRMP vary among different regions and countries. Specifically, it has been observed that MRMP has experienced a notable increase from the early 2000s to the mid-2010s, particularly in East Asia [13,14]. In contrast, North America, Europe, and Australia have significantly lower rates of macrolide resistance, with rates generally remaining below 10 % [13]. Macrolide resistance occurs in *M. pneumoniae* with structural alterations in the 50S ribosomal subunit and prevention of antibiotic efflux into the cell [15]. Moreover, since 2004, tetracycline and quinolone antibiotics have been authorized for the treatment of macrolide-resistant or refractory *M. pneumoniae* infections in children over the age of eight [16]. However, uncontrolled use of these drugs in clinical practices, drug resistance has become an even more serious problem. The most frequent pneumonia-causing bacteria, *Haemophilus influenzae,* and *Streptococcus pneumoniae* are currently treated by vaccinations. Nevertheless, more research is necessary to find out if atypical respiratory disease-causing pathogens, such as *M. pneumoniae,* can make the same amount of advancement [17]. Considering the rise of multidrug-resistant *M. pneumoniae*, identification of the novel and potent therapeutic targets can be decisive to manage this pathogen effectively.

In recent years, the field of computational biology has played an increasingly important role in the process of drug discovery, providing researchers with powerful tools and strategies [18]. However, designing a drug is a complex and time-consuming process [19]. One approach that has been particularly effective in identifying potential drug targets is comparative genomics [20]. This approach involves comparing the genomes of pathogenic microorganisms with those of their hosts to identify genes that are essential for the survival of the pathogen but are absent in the host. These non-homologous genes represent potential drug targets that can be exploited to develop novel antimicrobial agents [21]. Subtractive genomics is a specific type of comparative genomics that involves subtracting the genome of the host organism from that of the pathogen to identify potential drug targets that are unique to the pathogen. By combining subtractive genomics with other computational techniques, such as metabolic pathway analysis, researchers can further refine the search for druggable targets. Focusing on essential proteins involved in specific metabolic pathways unique to the pathogen, targets can be identified that are highly specific and less likely to interact with human proteins. This approach has been used successfully to identify essential, druggable, and non-homologous resistance proteins necessary for pathogen survival [22–24].

In this study, we have identified potential drug targets from the complete proteome of *M. pneumoniae* using a subtractive genomic approach. Rather than analyzing the whole genome, we particularly focused on the proteins which are non-homologous to the host and are also essential for the survival of the pathogens. If further experimentally proved, these findings might be effective against a wide range of harmful bacteria and aid in the development of drugs to treat *M. pneumoniae* infection.

2. Methodology

The study used a subtractive genomics approach to identify potential drug targets for *M. pneumoniae*, as displayed in Fig. 1.

2.1. Retrieval and identification of orthologs

The orthologs of *M. pneumoniae* were retrieved from OrthoVenn2 (https://orthovenn2.bioinfotoolkits.net/home) by comparing its proteome to the proteomes of *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Mycoplasma bovis*, *Mycoplasma gallisepticum,* and *Mycoplasma mycoides* [25].

2.2. Removal of duplicated sequences

The orthologs were run through the CD-HIT suite server (https://sites.google.com/view/cd-hit) to remove the paralogous sequences. CD-HIT is a webserver used for clustering biological sequences, mainly protein sequences, to decrease redundancy in protein sequence datasets. This valuable tool removes all the paralogous or duplicated sequences with a threshold value of 0.6, indicating 60 % identical sequences (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3516142/). Finally, the non-redundant sequences were selected for further analysis.

2.3. Screening of non-homologous protein

The non-redundant sequences were then subjected to Protein Basic Local Alignment Tool (Blastp)-based searches against humans [26]. Here, proteins with an expectation value (E value) of 0.001, query coverage $\leq$70 %, and identity $\leq$30 % were selected as non-homologous proteins. Non-homologous proteins are selected to prevent therapeutic drugs from binding to the active sites of the host proteins.

2.4. Identification of essential non-homologous proteins

Essential genes are essential for the survival of bacterial life. The Database of Essential Genes (DEG) (http://origin.tubic.org/deg/public/index.php) is a web server that predicts essential genes in bacteria based on a homologous sequence search. Protein sequences retrieved from the upstream analysis were subjected to Blastp against the database of essential genes to identify essential human non-homologous proteins [27]. Proteins that had a hit with expectation value ≤ 0.0000000001, identity $\geq$25 %, and bit score >100 were considered essential proteins.

2.5. Analysis of metabolic pathways

Unique metabolic pathways of *M. pneumoniae* from *Homo sapiens* (host) was retrieved using The Kyoto Encyclopedia of Genes and Genome (KEGG) pathway database [28]. The KEGG Automatic Annotation Server (KAAS) provided functional annotation of genes through comparisons against the manually curated KEGG GENES database. KAAS used BLAST KO (KEGG Orthology) assignments for generating KEGG pathways. *M. pneumoniae* proteins that shared a KO with *Homo sapiens* were excluded before proceeding to subsequent steps.

2.6. Prediction of subcellular localization

Gram-negative bacteria like *M. pneumoniae* have proteins that can be classified into different subcellular locations including 1) cytoplasm, 2) inner membrane, 3) periplasm, 4) outer membrane, and 5) extracellular space [24]. To identify the subcellular localization of metabolic proteins of our targeted organism the Bologna Unified Subcellular Component Annotator (BUSCA) webserver was used [29]. Afterward, the cytoplasmic proteins were subjected to druggability screening since they are considered potential drug targets [30]. In addition, conservation analysis of the cytoplasmic proteins was carried out using 80 *M. pneumoniae* strains.

2.7. Druggability analysis

The essential cytoplasmic proteins were then searched against the Drug bank (https://go.drugbank.com/) database using Blastp in order to evaluate their druggability [31,32]. The DrugBank is a unique resource of bioinformatics and cheminformatics that integrates extensive information about drugs with detailed data about their respective targets [31]. Proteins with a bit score higher than 100 and an E value less than 0.001 were considered potential drug targets or drugs. For proteins with corresponding drugs, ligand screening was conducted to assess their potentiality as drug candidates. Those without any hits were considered novel drug targets of *M. pneumoniae*.

2.8. Ligand screening

The druggable candidates were selected as model ligands so that compounds with similar properties could be identified. To accomplish this, the canonical simplified molecular-input line-entry system (SMILES) of the compounds were retrieved from NCBI PubChem (https://pubchem.nih.gov/) and then passed through the Swiss Similarity web tool for virtual similarity screening. SwissSimilarity (http://www.swisssimilarity.ch/) enables rapid ligand-based virtual screening [33] searching for the bioactive class of compounds in the ChEMBL and ChEBI databases. The ChEMBL database is used to rapidly calculate the molecular similarity based on chirality, shape, and electrostatics. The top 10 matches were selected as potential antimicrobial compounds against *M. pneumoniae.* Further, the SMILES of the matched compounds were converted to the pdb format using the Online SMILES Translator and Structure File Generator (https://cactus.nci.nih.gov/translate/).

Finally, molecular docking analysis was performed between the screened compounds and the targetted cytoplasmic proteins using PyRx (https://pyrx.sourceforge.io/) molecular docking software. PyRx is generally used to identify lead compounds with desired biological functions where small-molecule libraries are docked against macromolecules [34]). For a proper protein-ligand complex, the target macromolecule is expected to bind to the ligand with high binding affinity. The bond between a ligand and a macromolecule is stronger when the binding affinity is higher. The open-source tool PyRx makes use of AutoDock 4, OpenBabel, and AutoDock Vina in order to perform the docking [35–37]). In this study, the ligands and the receptors were converted into pdbqt format using OpenBabel. Each control drug and the screened ligands from the previous step were docked against their corresponding target proteins. The grid box for the docking was generated using the maximize option in order to find the most suitable binding pocket for the ligands. The docked complexes with the highest binding affinities were analyzed using PyMol [38]) and BIOVIA Discovery Studio Visualizer (https://discover.3ds.com) [39]).

2.9. Absorption, distribution, metabolism, excretion, and toxicity analysis

To develop new drugs, it is necessary to evaluate Absorption, Distribution, Metabolism, and Excretion (ADME) at an earlier stage of the drug discovery process, when the number of potential compounds is high but the access to physical samples is constrained [40]. Here, the ADME properties of the drug and the screened ligands with the highest binding affinity values were analyzed using the swissADME web tool [33]. In order to evaluate the toxicity properties, SMILES was provided to the pkCSM server (http://biosig.unimelb.edu.au/pkcsm/). Analogs with maximum intestinal absorption, negative hepatotoxicity, and negative results in the AMES test were selected for further analysis.

2.10. Molecular Dynamics Simulation

The apo receptor and ligand-receptor complex underwent a 100 ns Molecular Dynamics (MD) simulation using the GROningen MAchine for Chemical Simulations (GROMACS) (version 2020.6) [41]. The proteins were embedded in the TIP3 water model. Using the CHARMM36 m force-field, the whole system was energetically minimized [42]. Using K+ and Cl-ions, the systems were made neutral. Following the energy minimization, the system underwent isothermal-isochoric (NVT) and isobaric (NPT) equilibration. Then, a 100 ns production MD simulation was executed. Using MD simulation data, root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), and solvent accessible surface area (SASA) were calculated, as well as a hydrogen bond analysis. In RStudio, the ggplot2 package (https://ggplot2.tidyverse.org/) was used to create graphs for each of these analyses. All MD simulations were executed in high-performance simulation stations running on Ubuntu 20.04.4 LTS at the Bioinformatics Division, National Institute of Biotechnology, Bangladesh.

3. Result

The study employed a subtractive genomic approach to look for therapeutic target proteins that are indispensable for bacterial survival but are lacking in the host. The summary for the identification of novel drug targets in *M.pneumoniae* is portrayed in Fig. 2.

3.1. Identification of orthologous, non-paralogous, and non-homologous proteins

To compare the orthologs of *M. pneumoniae* with other members of the Mycoplasmatales family, Orthovenn2 was utilized which identified 227 orthologs. To ensure that the identified orthologs were non-redundant, a CD-HIT web server was utilized with 60 % identity. Here, all the orthologs were found to be non-paralogous. Next, the identified non-redundant orthologs were screened against NCBI Blastp to remove any proteins that exhibited significant homology to *Homo sapiens*. As a result, 71 non-homologous proteins were identified as suitable candidates for downstream analysis.

3.2. Identification of essential proteins

The screening of 71 non-homologous proteins using the DEG server resulted in 69 essential proteins for *M. pneumoniae* (Supplementary Table 1). These essential proteins play a vital role in the survival of the bacteria and must be non-homologous to prevent undesirable side effects.

3.3. Identification of pathogen-specific pathways

For the purpose of identifying potential new therapeutic targets, proteins participating only in pathogen-specific pathways were

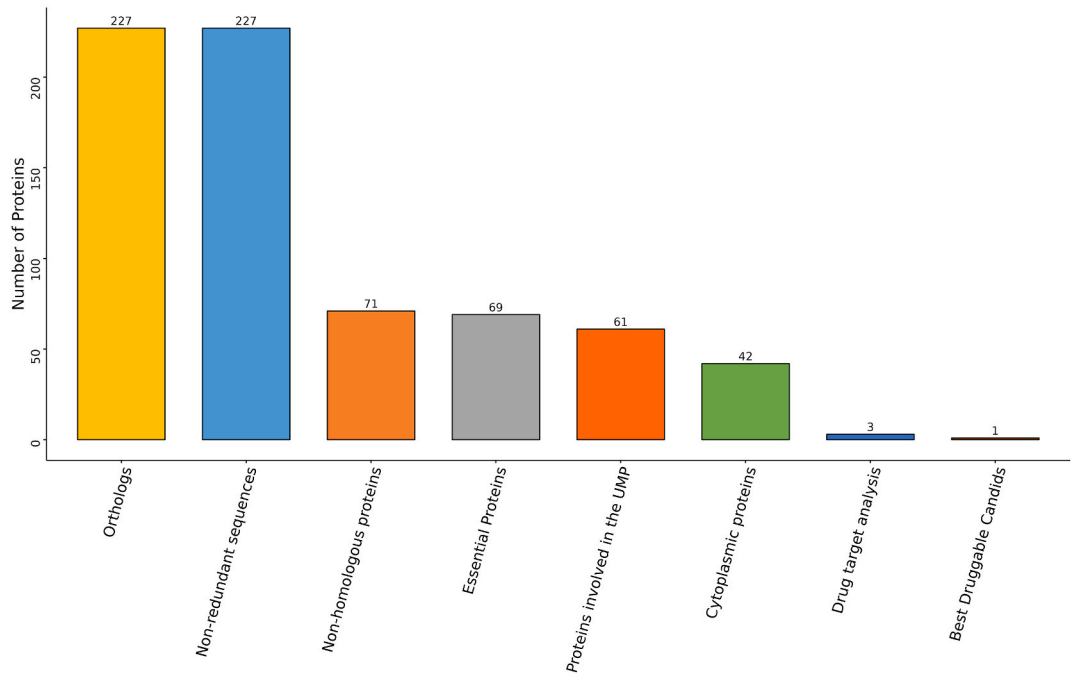


Fig. 2. Outline of potential therapeutic targets at different stages of screening.

taken into account. The KAAS server was used to screen all 69 of *M. pneumoniae*'s essential proteins, and the filtration revealed that 61 essential proteins were involved in metabolic pathways unique to *M. pneumoniae* (Supplementary Table 2, Fig. 3).

3.4. Subcellular localization analysis

Tracking the location of essential proteins is necessary to understand their functions in the relevant cell compartments as well as to facilitate the protein purification process in experimental setups. This is achieved by identifying whether the protein is cytoplasmic or membranous [34]. The analysis conducted via the BUSCA web server revealed that the retrieved proteins were mainly located in the cytoplasm and plasma membrane (Table 1, Fig. 4). Of the 61 essential proteins, 42 were observed to be present in the cytoplasm, while the remaining 19 proteins were identified as membrane proteins. For further analysis, only the cytoplasmic proteins were selected due to their high druggability compared to other subcellular locations. Additionally, conservation analysis of the cytoplasmic proteins revealed that all of the 42 cytoplasmic proteins were conserved within 80 *M. pneumoniae* strains (**Supplementary File 1**).

3.5. Identification of potential drug targets

The 42 proteins retrieved from the upstream analysis were filtered using Blastp against the DrugBank database. Two of the 42 proteins were identified to share similarities with three drug targets that have already been FDA-approved (Supplementary Table 3). These two proteins, DNA-directed RNA polymerase, and RNA polymerase sigma factor are enlisted in Table 2.

To expedite ligand-based virtual screening, the SwissSimilarity platform was employed. Here, three approved drugs were selected for further screening based on their electroshape scores. The top ten compounds most similar to their corresponding control drug were chosen for molecular docking. This screening protocol resulted in a total of 30 compounds as shown in Table 3.

3.6. Molecular docking and visualization

Following the identification of ligands using the SwissSimilarity platform, molecular docking investigations were conducted to evaluate their efficacy. The docking scores of the ligands against the intended protein targets were assessed, with the most efficacious ligand demonstrating the lowest docking score (Fig. 4, Supplementary Table 4). Here, Rifabutin, Rifamycin, and Fidaxomicin were selected as control. In the case of Rifabutin, nine druggable compounds were found to have higher binding affinity than the control whereas for the control Fidaxomicin, four compounds were found to have a higher binding affinity. Moreover, none of the novel targets of Rifamycin showed a satisfactory docking score to be identified as a therapeutic compound against M.pneumoniae infections. The molecular interactions between the ligand CHEBI97093 and the protein, DNA-directed RNA polymerase indicate a significant binding

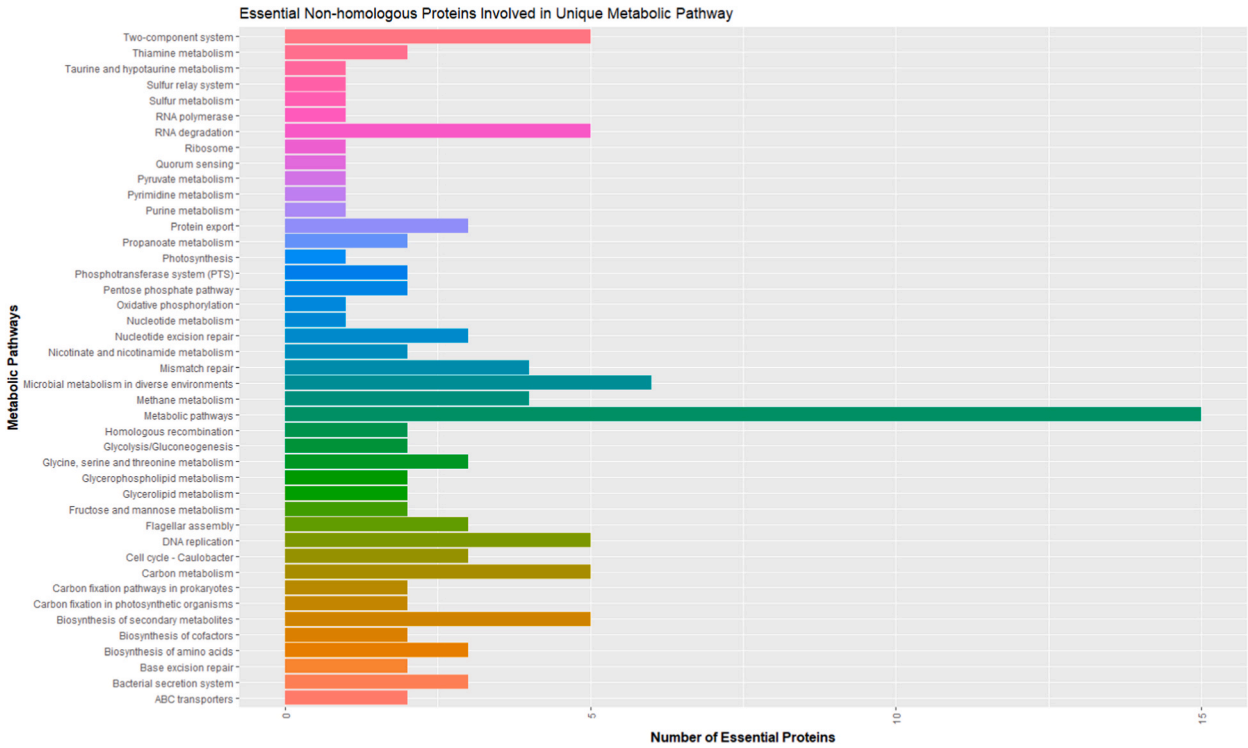


Fig. 3. Barplot analysis of non-homologous proteins involved in *M. pneumoniae's* unique metabolic p000athway.

Table 1
Subcellular localization prediction of essential proteins involved in unique metabolic pathway.

Accession	Subcellular Localization	*Homo sapiens*	Ko number
MPNA0250	cytoplasm	No	K01624
MPNA0340	plasma membrane	No	K03763
MPNA0530	cytoplasm	No	K02784
MPNA0670	cytoplasm	No	K02601
MPNA0740	cytoplasm	No	K03664
MPNA0800	plasma membrane	No	K02004
MPNA1240	cytoplasm	No	K03705
MPNA1250	cytoplasm	No	K03703
MPNA1400	cytoplasm	No	K06881
MPNA1540	cytoplasm	No	K02600
MPNA1800	cytoplasm	No	K02933
MPNA1840	plasma membrane	No	K03076
MPNA1870	cytoplasm	No	K02518
MPNA1910	cytoplasm	No	K03040
MPNA1950	plasma membrane	No	K16785
MPNA2100	cytoplasm	No	K03070
MPNA2220	plasma membrane	No	K04075
MPNA2230	cytoplasm	No	K06023
MPNA2240	plasma membrane	No	K13292
MPNA2320	cytoplasm	No	K02314
MPNA2490	cytoplasm	No	K06949
MPNA2590	plasma membrane	No	K23535
MPNA2600	plasma membrane	No	K23536
MPNA3000	plasma membrane	No	K05896
MPNA3010	cytoplasm	No	K06024
MPNA3140	cytoplasm	No	K03925
MPNA3170	cytoplasm	No	K03531
MPNA3280	cytoplasm	No	K01151
MPNA3310	cytoplasm	No	K03545
MPNA3360	plasma membrane	No	K00969
MPNA3410	plasma membrane	No	K03657
MPNA3490	cytoplasm	No	K09769
MPNA3500	plasma membrane	No	K08591
MPNA3520	cytoplasm	No	K03086
MPNA3530	cytoplasm	No	K02316
MPNA3570	cytoplasm	No	K01972
MPNA3780	plasma membrane	No	K02337
MPNA4010	cytoplasm	No	K03624
MPNA4150	cytoplasm	No	K02044
MPNA4170	plasma membrane	No	K02042
MPNA4280	cytoplasm	No	K00625
MPNA4460	cytoplasm	No	K02986
MPNA4600	plasma membrane	No	K03498
MPNA4760	cytoplasm	No	K00945
MPNA4820	plasma membrane	No	K09976
MPNA5330	cytoplasm	No	K00925
MPNA5390	cytoplasm	No	K02935
MPNA5460	cytoplasm	No	K03621
MPNA5470	cytoplasm	No	K07030
MPNA5500	cytoplasm	No	K03151
MPNA5950	cytoplasm	No	K01808
MPNA6040	plasma membrane	No	K02108
MPNA6210	cytoplasm	No	K12574
MPNA6240	cytoplasm	No	K02902
MPNA6270	cytoplasm	No	K08483
MPNA6280	cytoplasm	No	K15633
MPNA6320	cytoplasm	No	K09903
MPNA6590	cytoplasm	No	K00554
MPNA6600	cytoplasm	No	K02959
MPNA6800	plasma membrane	No	K03217
MPNA6860	plasma membrane	No	K02313

energy value of −11.2 kcal/mol. A 2D diagram of the receptor-ligand interaction generated using BIOVIA Discovery Studio Visualizer shows the interactions between the ligand and the target protein (Fig. 5, Table 4). Fig. 6 highlights the hydrophobic surface around the binding pocket of CHEBI97093 and the protein (see Fig. 7).

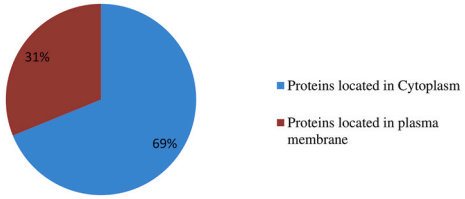


Fig. 4. A pie chart displaying the distribution of proteins in cytoplasm and plasma membrane.

Table 2
List of druggable target proteins identified via DrugBank.

Accession	KO number	Protein Name	Location	Approved Drugs	*Homo sapiens*
MPNA1910	K03040	DNA-directed RNA polymerase	Cytoplasm	Rifabutin, Rifamycin,	No
MPNA3520	K03086	RNA polymerase sigma factor	Cytoplasm	Fidaxomicin	No

Ligand screening.

Table 3
List of potential lead compounds and their analogs.

Protein	MPNA1910
Ligands (Drugs, CHEBL,CHEBI)	RIFABUTIN (control)
	CHEBI103560
	CHEBI111544
	CHEBI90884
	CHEBI97093
	CHEBI99015
	CHEMBL1834897
	CHEMBL1988195
	CHEMBL4173635
	CHEMBL4207303
	CHEMBL501238
Protein	MPNA1910
Ligands (Drugs, CHEBL,CHEBI)	RIFAMYCIN control
	CHEBI142726
	CHEBI29673
	CHEBI34349
	CHEBI34948
	CHEBI84571
	CHEMBL1373126
	CHEMBL1519702
	CHEMBL4300133
	CHEMBL4303200
	CHEMBL437765
Protein	MPNA3520
Ligands (Drugs, CHEBL,CHEBI)	FIDAXOMICIN control
	CHEMBL445546
	CHEMBL500998
	CHEMBL505136
	CHEMBL505392
	CHEMBL1255800
	CHEMBL1796211
	CHEMBL3357582
	CHEMBL4077022
	CHEMBL4540936
	CHEMBL4574960

3.7. ADMET analysis

The compound with the highest binding affinity was selected for ADMET screening and compared to currently available approved drugs based on various parameters such as Bioavailability Score, PAINS, Brenk, Lead Likeness, Synthetic Accessibility, and toxicity. SwissADME was used to analyze the toxicity of the compound, as shown in Supplementary Table 5, while PKCSM was deployed to evaluate AMES toxicity, Max. tolerated dose (human), hERG I inhibitor, hERG II inhibitor, Oral Rat Acute Toxicity (LD50), Oral Rat Chronic Toxicity (LOAEL), Hepatotoxicity, Skin Sensitisation, Skin Sensitisation, T.Pyriformis toxicity, and Minnow toxicity, as presented in Supplementary Table 6. The results of the PKCSM analysis indicated that the target is non-toxic, further supporting its

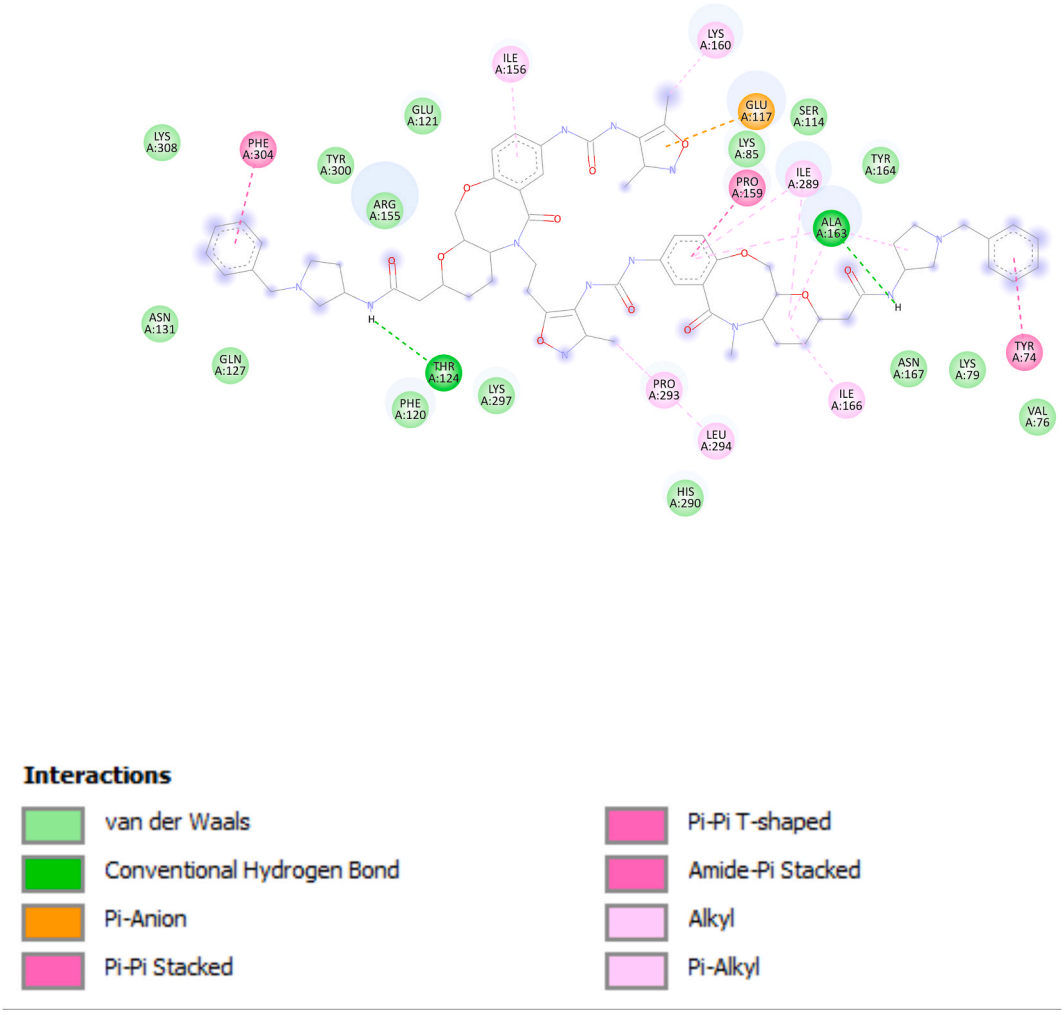


Fig. 5. Comparison of binding affinities between the control and the potential ligands.

potential as a drug candidate. Based on the outcomes of both SwissADME and PKCSM, it was concluded that CHEBI:97093 is the most promising drug candidate.

Molecular Dynamics Simulation of the apo-receptor and drug-receptor Complex.

In order to evaluate the stability of the systems, the Root Mean Square Deviation (RMSD) was measured. Change in RMSD value corresponds to conformational alterations caused by ligand binding. In Fig. 8, the green line represents the RMSD profile of the apo receptor, while the blue line represents the drug-receptor complex. The RMSD value of the drug-receptor complex was ~0.6 nm whereas the apo-receptor was ~0.3 nm. Hence, the ligand-bound conformation was different in dynamic conditions. Root Mean Square Fluctuation (RMSF) was calculated to determine the regional flexibility of the proteins. The higher the RMSF, the higher is the flexibility of a given amino acid position. Fig. 9 demonstrates the RMSF profile of the apo receptor and drug-receptor complex. The RMSF peaks were observed at ~150th residue, ~200th residue, ~250th residue,~300th residue, and the C-terminal part of the protein. In the 50th and 80th peaks, the apo receptor showed higher mobility. In ~150th and ~200th residues, the peaks of apo-receptor and drug-receptor complexes were different. For ~300th residue, binding of the drug reduced mobility. These differences of peaks indicated that the ligand binding altered the natural mobility of the protein. Solvent Accessible Surface Area (SASA) was implemented to predict the hydrophobic core stability of proteins. The higher the SASA value, the higher the chance of destabilization of the protein due to solvent accessibility. The SASA value of apo-receptor and drug-receptor was ~275 nm ^{2} finally (Fig. 10). The similarities between SASA values of the apo-receptor and drug-receptor complex suggested that the binding of the drug did not impact the hydrophobic core of the protein. The radius of gyration (Rg) was a measurement to determine the degree of compactness of the protein. A relatively steady value of the radius of gyration means stable folding of a protein. Fluctuation of the radius of gyration implies the unfolding of the protein. According to Fig. 11, the drug-receptor complex was less stable than the apo-receptor.

Table 4
Binding interactions between the ligand (CHEBI97093) and the target protein.

Bond formed	Interacting residues
Van der Waals	ARG:155
	ASN:131
	ASN:167
	GLN:127
	GLU:121
	HIS:290
	LYS:79
	LYS:85
	LYS:297
	LYS:308
	PHE:120
	SER:114
	TYR:164
	TYR:300
	VAL:76
Hydrogen Bond	ALA:163
	THR:124
Pi-Anion	GLU:117
Pi Bond	PHE:304
	PRO:159
	TYR:74
Alkyl Bond	ILE:156
	ILE:166
	ILE:289
	LEU:294
	LYS:160
	PRO:293

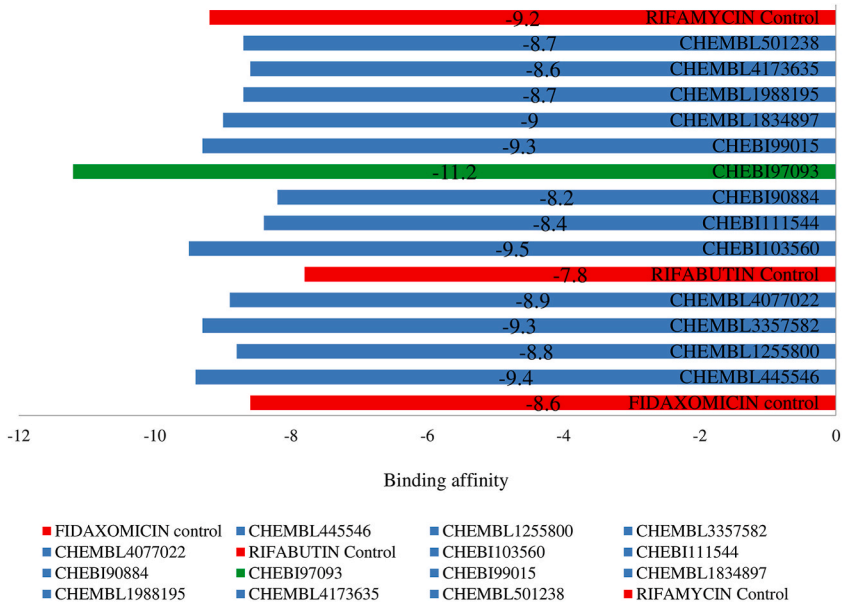


Fig. 6. Interactions between the ligand (CHEBI97093) and the target protein.

4. Discussion

Community-acquired pneumonia (CAP) is a prevalent respiratory infection that can be caused by a variety of atypical bacterial microorganisms [43]. *M. pneumoniae* is one of the leading causes of such infections, accounting for approximately 7–20% of CAP cases [5]. Even though *M. pneumoniae* infections are typically mild, they can cause severe and potentially fatal pneumonia that spreads within communities and families [44]. Patients may also develop complications involving multiple organs, including autoimmune and inflammatory responses [45]. Since mycoplasmas lack a cell wall, the number of antibiotics that can be used for treating mycoplasmal

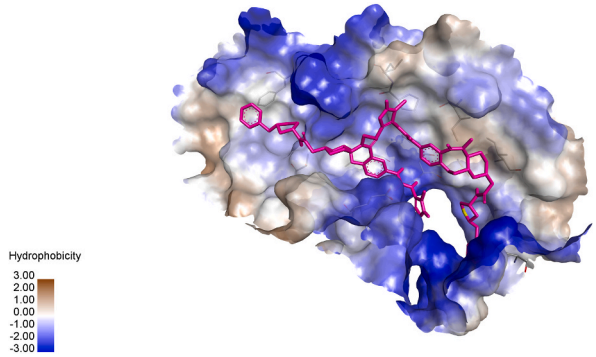


Fig. 7. Illustrates the receptor surface displayed by generating a hydrophobic surface. The degree of hydrophobicity is indicated by color range. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

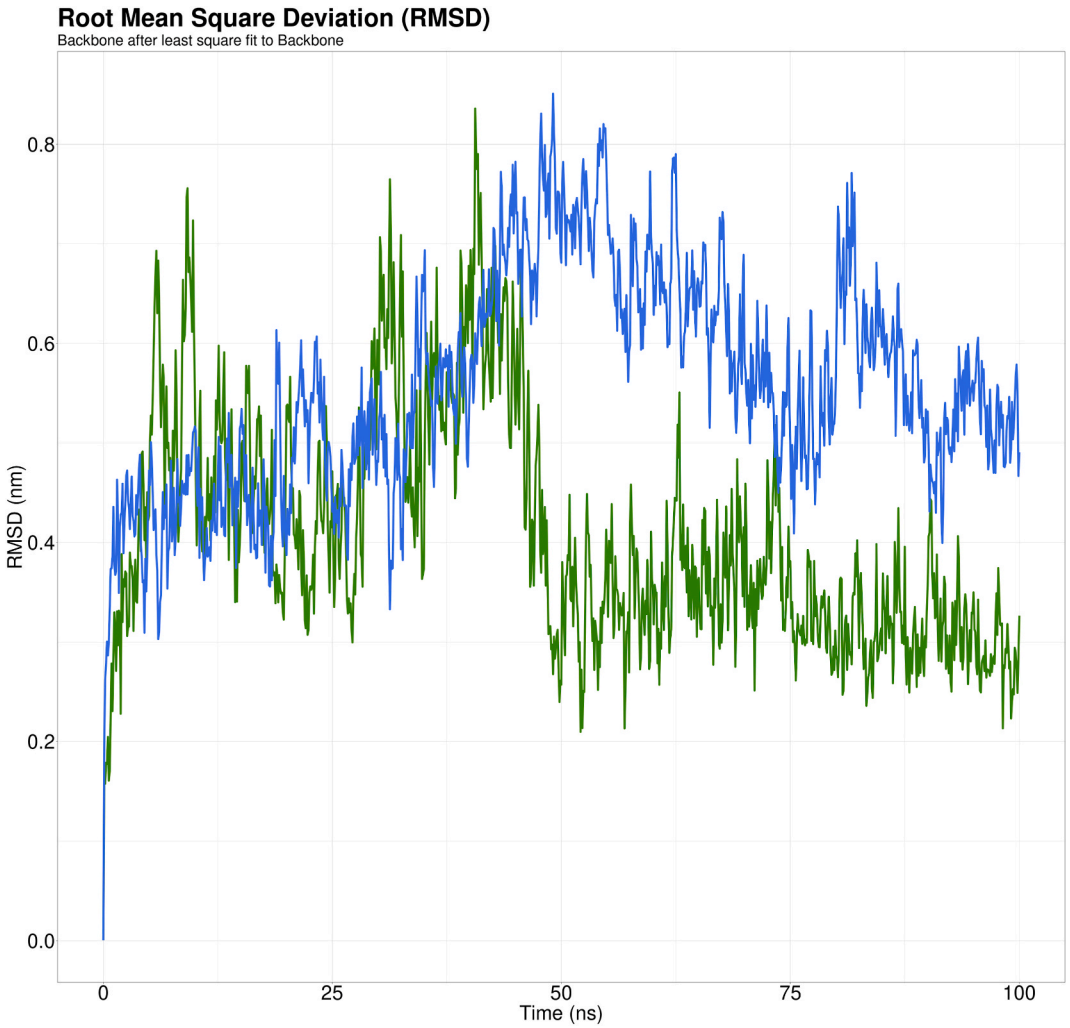


Fig. 8. Root Mean Square Deviation (RMSD) profile of apo-receptor DNA-directed RNA polymerase, alpha subunit (green) and drug-receptor complex (Blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

infections is limited [46]. The situation tends to be the worst due to the emergence of resistance to macrolides since the 2000s [47]. The search for new drugs to treat bacterial infections, including *M. pneumoniae*, is a critical public health concern, especially given the rise of antibiotic-resistant strains. Recent progress in the field of informatics and computational biology has significantly enhanced the

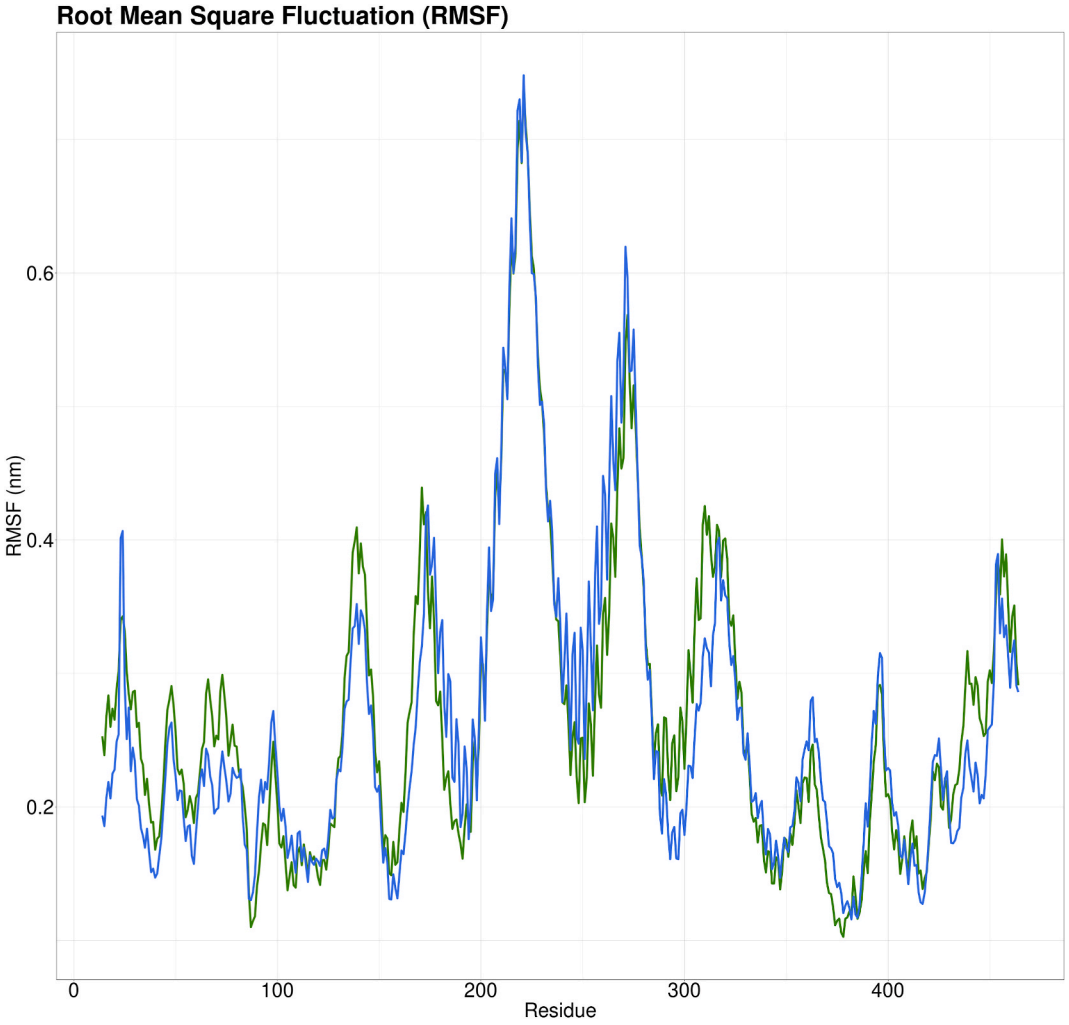


Fig. 9. Root Mean Square Fluctuation (RMSF) profile of apo-receptor DNA-directed RNA polymerase, alpha subunit (green) and drug-receptor complex (Blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

efficiency of the drug discovery process, providing a promising opportunity to identify novel treatments for bacterial infections like *M. pneumoniae*.

In this study, a subtractive genomics method was implemented to discover potential drugs that could be used against *M. pneumoniae*. The entire proteome of *M. pneumoniae*, comprising 227 proteins, was screened using a comparative analysis of orthologs shared across the *Mycoplasmatales* family ensuring the safety and efficacy of the drugs. To prevent certain genes from being overrepresented in the analysis and to avoid computational complexities, paralogous sequences were eliminated. This ensured that the resulting data was accurate, representative, and could be effectively handled. Bacteria rely on a set of genes, referred to as essential genes, for their survival. As such, these genes are ideal targets for the development of antibacterial drugs [48]. Thus essential genes were screened from non-paralogous sequences. However, these essential genes may also be present in humans, leading to significant fatal side effects. Therefore, to prevent cross-reactivity and minimize the risk of harmful events, non-homologous essential proteins were selected as the optimal targets for developing inhibitors to produce new drugs. This strategy of targeting non-homologous sequences could be the most effective approach to developing drugs with a higher level of specificity, resulting in fewer off-target effects and less toxicity.

Comparative analysis of the metabolic pathways of the host and pathogen revealed the presence of pathogen-specific pathways, which are defined as pathways present only in the bacterium. These pathways represent potential targets for drug development as they are essential to bacterial survival but not present in human cells. Furthermore, these pathways contain proteins that are involved in key metabolic processes, including energy production, nucleotide synthesis, and lipid metabolism, highlighting the potential of targeting these pathways to disrupt bacterial growth and survival. Nonetheless, it is essential to consider the subcellular localization of these proteins since they can have multiple locations. Membrane-bound proteins are challenging to purify and assay [49]. Hence, cytoplasmic proteins are preferred as drug targets as they are more accessible to small molecules and drugs. The analysis showed that 69 %

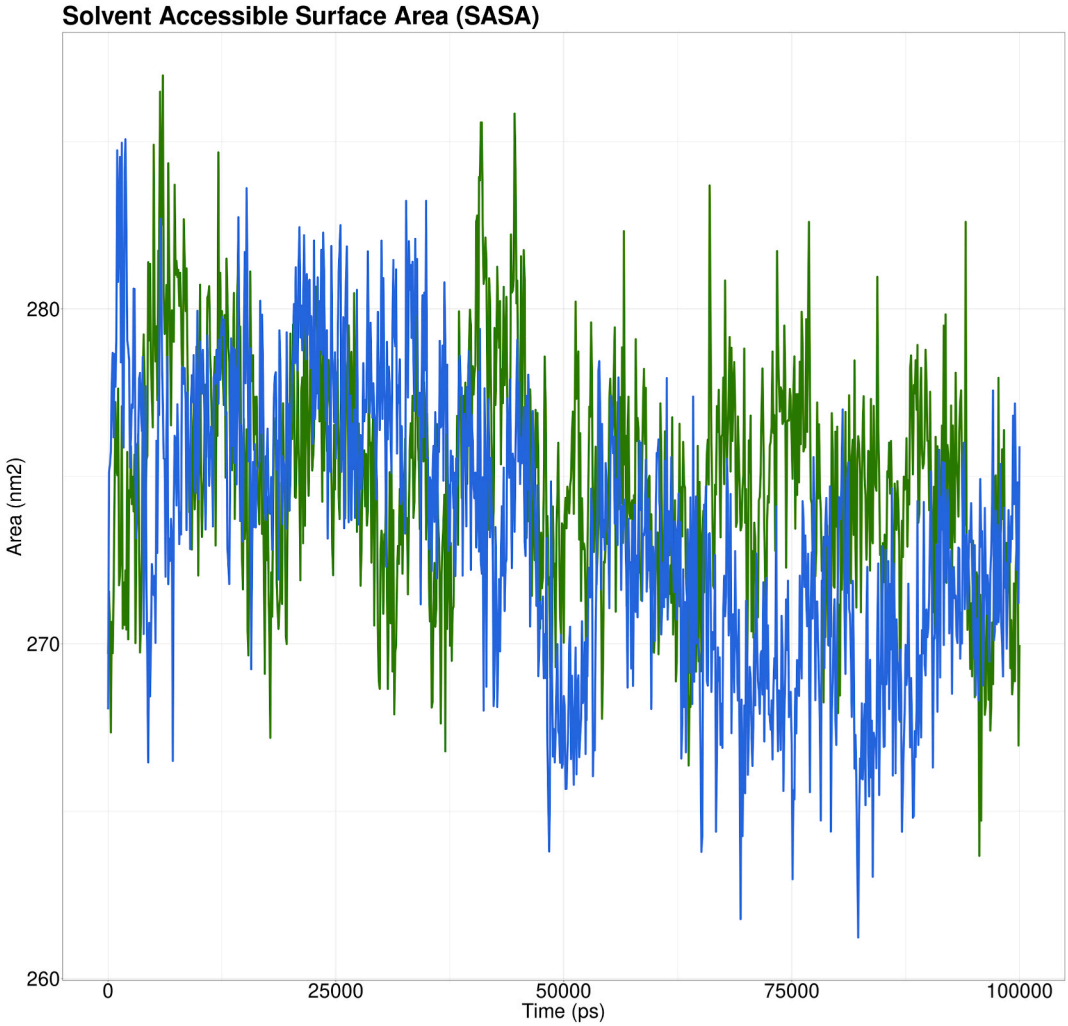


Fig. 10. Solvent Accessible Surface Area (SASA) profile of apo-receptor DNA-directed RNA polymerase, alpha subunit (green) and drug-receptor complex (Blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

of the identified proteins were located in the cytoplasm, while the rest were found in plasma membranes. All of these cytoplasmic proteins were found to be conserved among 80 *M. pneumoniae* strains. Subsequently, druggability assessment of cytoplasmic proteins revealed that two of them share similarities with three FDA-approved drug targets, Rifabutin, Rifamycin, and Fidaxomicin. To accelerate ligand-based virtual screening, compounds were chosen for molecular docking from each drug based on their electroshape scores. From the 30 molecules that were docked, 2-[(2R,4aS,12aS)-8-[[[(3,5-dimethyl-4-isoxazolyl)amino]-oxomethyl]amino]-5-methyl-6-oxo-2,3,4,4a,12,12a-hexahydropyrano[2,3-c][1,5]benzoxazocin-2-yl]-N-[(3S)-1-(phenylmethyl)-3-pyrrolidinyl]acetamide (CHEBI97093) had the highest binding affinity of −11.2 kcal/mol, which was higher than the control Rifabutin. Additionally, the analysis of the ligand-receptor complex revealed that the analog formed more conventional hydrogen bonds and did not have any unfavorable interactions with the receptors indicating its potential for further exploration.

The availability of a biomolecule in a bioactive form at the intended target site is crucial for its efficacy as a medication. Evaluation of the ADMET properties of a drug candidate can help determine its behavior, toxicity, and fate in the human body. This includes assessing its ability to be absorbed in the intestine, cross the blood-brain barrier, subcellular localization, and potential harm to the body [50]. ADMET analysis using pkCSM-based toxicity tests revealed no hepatotoxic properties and similar bioavailability to the approved drug. The drug-likeness and interactions with cytochrome P450 were also within acceptable limits. Molecular Dynamic (MD) simulation was used to examine the stability of the apo-receptor and drug-receptor complex. The drug-receptor complex had a higher RMSD value than the apo-receptor, indicating that the binding of the drug caused conformational changes in the receptor structure. Peaks were observed at the 150th residue, 200th residue, 250th residue, 300th residue, and C-terminal region of the protein, as determined by the RMSF analysis. The apo-receptor was more mobile at the 50 to 80 amino acids and the drug-receptor complex was less mobile at the 300th residue. The SASA analysis suggested that the hydrophobic core stability of apo-receptors and drug receptors was comparable. The drug-receptor complex was less stable than the apo-receptor, as demonstrated by the Rg analysis. Overall, these

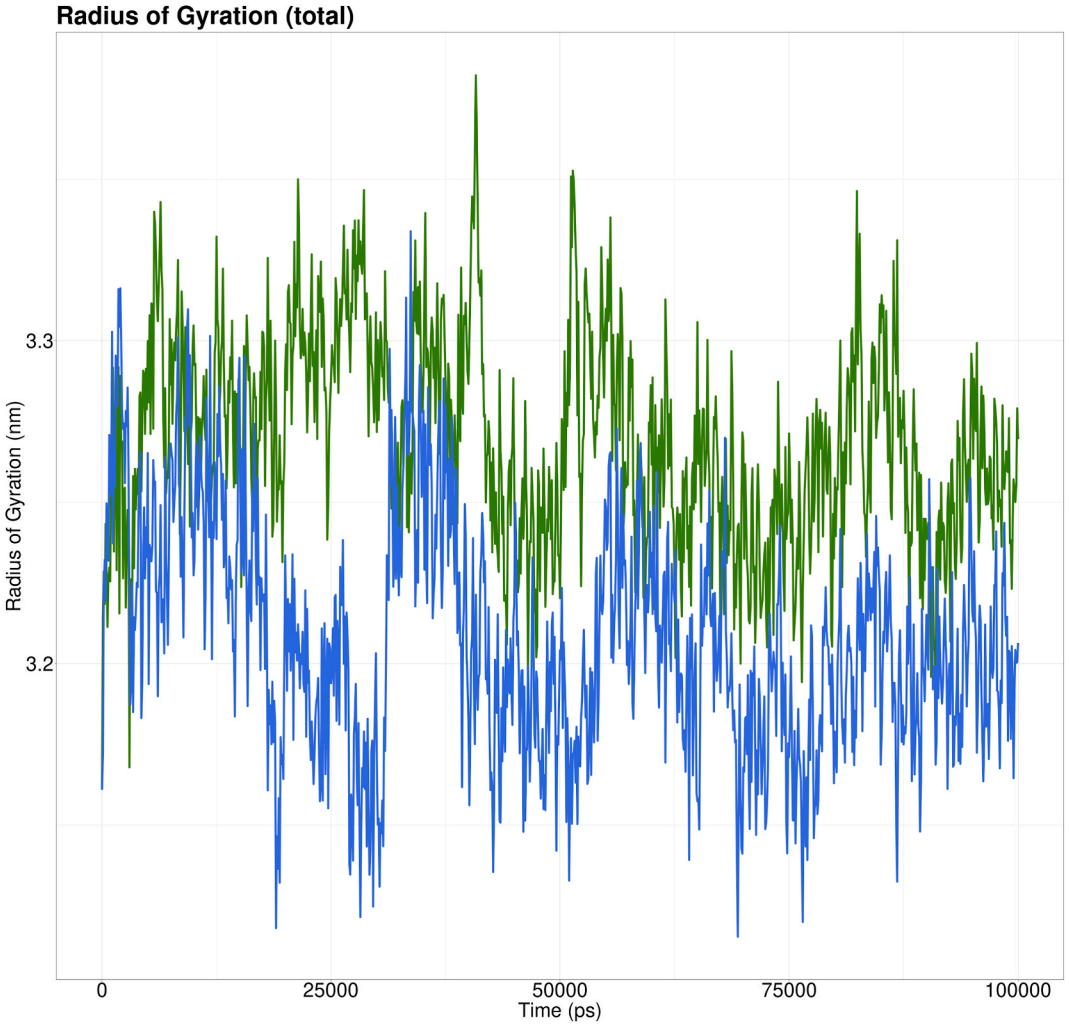


Fig. 11. Radius of Gyration (Rg) of apo-receptor DNA-directed RNA polymerase, alpha subunit (green) and drug-receptor complex (Blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

results suggest that the binding of the drug affects the structural stability of the receptor and induces conformational changes that modify the mobility of particular protein regions without affecting the hydrophobic core. These results suggest that CHEBI: 97093 (an aromatic amine) has the potential to disrupt the function of *M. pneumoniae* DNA-directed RNA polymerase. Since this polymerase is essential for the survival of the pathogen, disruption of this protein with CHEBI: 97093 will inhibit the bacterial pathogenesis at transcriptional level. Further study is required to assess the effectiveness of CHEBI: 97093 and their rationally designed analogs using *in vitro* and *in vivo* models.

5. Conclusion

The discovery of therapeutic targets against pathogens has undergone a radical change as a result of the investigation of the genomes and proteomes of pathogens. This study utilized a subtractive genomic strategy to identify essential druggable non-homologous proteins against *M. pneumoniae*, resulting in the identification of CHEBI:97093 as a promising drug candidate with better druggable properties and high binding affinity to the receptor. This candidate has the potential to treat *M. pneumoniae* infection, thus the results may advance the evaluation of already approved drugs. However, to fully ascertain the therapeutic potential of this candidate, additional *in vitro* and *in vivo* investigations are necessary to validate its efficacy.

Data availability

Data included in article/supp. material/referenced in article.

CRediT authorship contribution statement

Zeshan Mahmud Chowdhury: Writing – review & editing, Methodology, Investigation. **Tabassum Binte Jamal:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Ishtiaque Ahammad:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Arittra Bhattacharjee:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Anika Bushra Lamisa:** Formal analysis, Writing – original draft, Visualization. **Jannatul Maoa Jani:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Md Fahim Israk:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Mohammad Uzzal Hossain:** Supervision. **Keshob Chandra Das:** Supervision. **Chaman Ara Keya:** Writing – review & editing. **Md Salimullah:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21466.

References

[1] S. Kashyap, M. Sarkar, Mycoplasma pneumonia: clinical features and management, Lung. India Off Organ. Indian Chest Soc. 27 (2) (2010) 75–85.
[2] R. D'Alonzo, E. Mencaroni, L. Di Genova, D. Laino, N. Principi, S. Esposito, Pathogenesis and treatment of neurologic diseases associated with mycoplasma pneumoniae infection, Front. Microbiol. 9 (2018), 2751.
[3] K.B. Waites, D.F. Talkington, Mycoplasma pneumoniae and its role as a human pathogen, Clin. Microbiol. Rev. 17 (4) (2004) 697–728 (table of contents).
[4] T.P. Atkinson, M.F. Balish, K.B. Waites, Epidemiology, clinical manifestations, pathogenesis and laboratory detection of Mycoplasma pneumoniae infections, FEMS Microbiol. Rev. 32 (6) (2008), 956–73.
[5] T. Kishaba, Community-Acquired Pneumonia Caused by Mycoplasma pneumoniae: How Physical and Radiological Examination Contribute to Successful Diagnosis, Front. Med. 3 (28) (2016) 00028.
[6] S. Kumar, Mycoplasma pneumoniae: a significant but underrated pathogen in paediatric community-acquired lower respiratory tract infections, Indian J. Med. Res. 147 (1) (2018 Jan) 23–31.
[7] F.M. Sánchez-Vargas, O.G. Gómez-Duarte, Mycoplasma pneumoniae—an emerging extra-pulmonary pathogen, 105–15, Clin. Microbiol. Infect. 14 (2) (2008 Feb 1).
[8] M. Narita, Classification of extrapulmonary manifestations due to mycoplasma pneumoniae infection on the basis of possible pathogenesis, Front. Microbiol. 23 (7) (2016).
[9] S. Yimenicioğlu, A. Yakut, A. Ekici, K. Bora Carman, E. Cagrı Dinleyici, Mycoplasma pneumoniae infection with neurologic complications, 647–51, Iran. J. Pediatr. (Engl. Ed.) 24 (5) (2014 Oct).
[10] O.A. Chernova, E.S. Medvedeva, A.A. Mouzykantov, N.B. Baranova, V.M. Chernov, Mycoplasmas and their antibiotic resistance: the problems and prospects in controlling infections, Acta Naturae 8 (2) (2016) 24–34.
[11] K. Kim, S. Jung, M. Kim, S. Park, H.J. Yang, E. Lee, Global trends in the proportion of macrolide-resistant mycoplasma pneumoniae infections: a systematic review and meta-analysis, JAMA Netw. Open 5 (7) (2022 Jul 11), e2220949.
[12] Mycoplasma pneumoniae, Antibiotics and Resistance | CDC [Internet], 2023 [cited 2023 Apr 27]. Available from: https://www.cdc.gov/pneumonia/atypical/mycoplasma/hcp/antibiotic-treatment-resistance.html.
[13] S. Pereyre, J. Goret, C. Bébéar, Mycoplasma pneumoniae: current knowledge on macrolide resistance and treatment, Front. Microbiol. 7 (2016) 974.
[14] F. Zhao, G. Liu, J. Wu, B. Cao, X. Tao, L. He, et al., Surveillance of macrolide-resistant Mycoplasma pneumoniae in Beijing, China, from 2008 to 2012, Antimicrob. Agents Chemother. 57 (3) (2013 Mar), 1521–3.
[15] G.P. Dinos, The macrolide antibiotic renaissance, Br. J. Pharmacol. 174 (18) (2017 Sep) 2967–2983.
[16] D.X. Guo, W.J. Hu, R. Wei, H. Wang, B.P. Xu, W. Zhou, et al., Epidemiology and mechanism of drug resistance of Mycoplasma pneumoniae in Beijing, China: a multicenter study, Bosn. J. Basic Med. Sci. 19 (3) (2019 Aug 20), 288–96.
[17] T.C. Vilela Rodrigues, A.K. Jaiswal, A. de Sarom, L. de Castro Oliveira, C.J. Freire Oliveira, P. Ghosh, et al., Reverse vaccinology and subtractive genomics reveal new therapeutic targets against Mycoplasma pneumoniae: a causative agent of pneumonia, R. Soc. Open Sci. 6 (7) (2019 Jul 31), 190907.
[18] X. Lin, X. Li, X. Lin, A review on applications of computational methods in drug screening and design, Mol Basel Switz 25 (6) (2020 Mar 18) 1375.
[19] N. Dhingra, Computer-aided drug design and development: an integrated approach [Internet]. IntechOpen, in: J. Akhtar, Badruddeen, M. Ahmad, M. Irfan Khan (Eds.), Drug Development Life Cycle, 2022 [cited 2023 Apr 27]. Available from: https://www.intechopen.com/chapters/82398.
[20] M.Y. Galperin, E.V. Koonin, Searching for drug targets in microbial genomes, Curr. Opin. Biotechnol. 10 (6) (1999 Dec) 571–578.
[21] A.M. Butt, S. Tahir, I. Nasrullah, M. Idrees, J. Lu, Y. Tong, Mycoplasma genitalium: a comparative genomics study of metabolic pathways for the identification of drug and vaccine targets, Infect. Genet. Evol. 12 (1) (2012 Jan 1) 53–62.
[22] B. Ashraf, N. Atiq, K. Khan, A. Wadood, R. Uddin, Subtractive genomics profiling for potential drug targets identification against Moraxella catarrhalis, PLoS One 17 (8) (2022 Aug 25), e0273252.
[23] D. Barh, S. Tiwari, N. Jain, A. Ali, A.R. Santos, A.N. Misra, et al., In silico subtractive genomics for target identification in human bacterial pathogens, Drug Dev. Res. 72 (2) (2011 Mar) 162–177.
[24] MdT. Khan, A. Mahmud, A. Iqbal, S.F. Hoque, M. Hasan, Subtractive genomics approach towards the identification of novel therapeutic targets against human Bartonella bacilliformis, Inform. Med. Unlocked 20 (2020 Jan 1), 100385.
[25] L. Xu, Z. Dong, L. Fang, Y. Luo, Z. Wei, H. Guo, et al., OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species, Nucleic Acids Res. 47 (W1) (2019 Jul 2) W52–W58.
[26] S. McGinnis, T.L. Madden, BLAST: at the core of a powerful and diverse set of sequence analysis tools, Web Server issue, Nucleic Acids Res. 32 (2004 Jul 1) W20–W25.
[27] R. Zhang, H.Y. Ou, C.T. Zhang, DEG: a database of essential genes, Nucleic Acids Res. 32 (Database issue) (2004 Jan 1) D271–D272.
[28] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007 Jul) W182–W185. Web Server issue).

[29] C. Savojardo, P.L. Martelli, P. Fariselli, G. Profiti, R. Casadio, BUSCA: an integrative web server to predict subcellular localization of proteins, Nucleic Acids Res. 46 (W1) (2018 Jul 2) W459–W466.
[30] M. Karim, N. Islam, G. Jewel, Silico identification of potential drug targets by subtractive genome analysis of Enterococcus faecium do, bioRxiv 2 (14) (2020) 948232.
[31] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (D1) (2018 Jan 4) D1074–D1082.
[32] D.S. Wishart, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. 34 (90001) (2006 Jan 1) D668–D672.
[33] V. Zoete, A. Daina, C. Bovigny, O. Michielin, SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening, J. Chem. Inf. Model. 56 (8) (2016 Aug 22) 1399–1404.
[34] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx [cited 2023 Mar 15]. pp. 243–50, in: J.E. Hempel, C.H. Williams, C.C. Hong (Eds.), Chemical Biology [Internet], Methods in Molecular Biology, vol. 1263, Springer New York, New York, NY, 2015. Available from: http://link.springer.com/10.1007/978-1-4939-2269-7_19.
[35] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, et al., AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (16) (2009 Dec) 2785–2791.
[36] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminf. 3 (1) (2011 Oct 7) 33.
[37] O. Trott, A. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J Comput Chem 31 (2) (2010) 455–461.
[38] S. Yuan, H.C.S. Chan, Z. Hu, Using PyMOL as a platform for computational drug design, WIREs Comput. Mol. Sci. 7 (2) (2017) e1298.
[39] S.S. Pawar, S.H. Rohane, Review on discovery Studio: an important tool for molecular docking, Asian J. Res. Chem. 14 (1) (2021) 1–3.
[40] M.P. Doogue, T.M. Polasek, The ABCD of clinical pharmacokinetics, Ther. Adv. Drug Saf. 4 (1) (2013 Feb) 5–7.
[41] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (16) (2005 Dec) 1701–1718.
[42] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, et al., CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods 14 (1) (2017 Jan) 71–73.
[43] J.S. Brown, Community-acquired pneumonia, Clin Med Lond Engl 12 (6) (2012 Dec) 538–543.
[44] M. Tsokos, Fatal respiratory tract infections with mycoplasma pneumoniae [Internet], in: M. Tsokos (Ed.), Forensic Pathology Reviews, Humana Press, Totowa, NJ, 2004, https://doi.org/10.1007/978-1-59259-786-4_8.
[45] J. Hu, Y. Ye, X. Chen, L. Xiong, W. Xie, P. Liu, Insight into the pathogenic mechanism of mycoplasma pneumoniae, Curr. Microbiol. 80 (1) (2022 Dec 2) 14.
[46] C. Bébéar, S. Pereyre, O. Peuchant, Mycoplasma pneumoniae: susceptibility and resistance to antibiotics, 423–31, Future Microbiol. (4) (2011 Apr;6).
[47] K.B. Waites, A. Ratliff, D.M. Crabb, L. Xiao, X. Qin, R. Selvarangan, et al., Macrolide-resistant mycoplasma pneumoniae in the United States as determined from a national surveillance program, J. Clin. Microbiol. 57 (11) (2019 Nov).
[48] F.R. Fields, S.W. Lee, M.J. McConnell, Using bacterial genomes and essential genes for the development of new antibiotics, Biochem. Pharmacol. 134 (2017 Jun 15) 74–86.
[49] E.P. Carpenter, K. Beis, A.D. Cameron, S. Iwata, Overcoming the challenges of membrane protein crystallography, Curr. Opin. Struct. Biol. 18 (5) (2008 Oct) 581–586.
[50] L. Guan, H. Yang, Y. Cai, L. Sun, P. Di, W. Li, et al., ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness, MedChemComm 10 (1) (2019 Jan 1) 148–157.