1 **Identification of potential candidates for drug designing against *Klebsiella Pneumoniae***
2 **among hypothetical proteins of DNA Adenine Methyltransferase through subtractive**
3 **genomics approach**

4
5 Umairah Natasya Mohd Omeershhfudin ¹, Suresh Kumar ^{1*}

6 ¹ Department of Diagnostic and Allied Health Science, Faculty of Health and Life Sciences,
7 Management and Science University, 40100, Shah Alam, Selangor Darul Ehsan, Malaysia

8 *Corresponding author

9
10 **Address for Correspondence**

11 Dr.Suresh Kumar, Tel: +60-142734893. E-mail: sureshkumar@msu.edu.my Address:
12 Department of Diagnostic and Allied Health Science, Faculty of Health and Life Sciences,
13 Management and Science University, 40100, Shah Alam, Selangor Darul Ehsan, Malaysia

14
15
16
17
18
19
20
21
22
23
24
25
26
27

1 ABSTRACT

2 *Klebsiella Pneumoniae* is a gram-negative bacterium that is known for causing infection
 3 in nosocomial settings. As reported by WHO, this bacterial pathogen is classified as an
 4 urgent threat our most concern is that these bacterial pathogens acquired genetic traits
 5 that make them resistant towards antibiotics. The last class of antibiotics; carbapenems
 6 are not able to combat these bacterial pathogens allowing them to clonally expand their
 7 antibiotic-resistant strain. Most antibiotics target the essential pathways of the bacteria
 8 cell however these targets are no longer susceptible to the antibiotic. Hence in our study,
 9 we focus on *Klebsiella Pneumoniae* bacterial strains that contain DNA Adenine
 10 Methyltransferase domain which suggests a new potential site for a drug target. DNA
 11 methylation is seen to regulate the attenuation of bacterial virulence. In this study, all
 12 hypothetical proteins of *Klebsiella Pneumoniae* containing N6 DNA Adenine
 13 Methyltransferase domain were analysed for a potential drug target. About 32
 14 hypothetical proteins were retrieved from Uniprot. 19 proteins were selected through a
 15 step-wise subtractive genomics approach like a selection of non-homologous proteins
 16 against the human host, selection of bacterial proteins contains an essential gene, broad-
 17 spectrum analysis, druggability analysis, Non-homology analysis against gut
 18 microbiota. Through drug target prioritization like sub-cellular analysis, drug property
 19 analysis, anti-target non-homology analysis, virulence factor analysis and protein-
 20 protein interaction analysis one drug target protein (Uniprot ID: A0A2U0NNR3) was
 21 prioritized. Identified drug target docked with potential inhibitors like are mahanine
 22 (PubChem ID: 375151), curcumin (PubChem ID: 969516), EGCG (PubChem ID:
 23 65064), nanaomycin A (PubChem ID: 40586), parthenolide (PubChem ID: 7251185),
 24 quercetin (PubChem ID: 5280343) and trimethylaurintricarboxylic acid. Based on the
 25 molecular docking analysis, mahanine has the highest binding affinity. In order to
 26 identify novel natural inhibitor based on mahanine fingerprint search is performed
 27 against NPASS (Natural Product Activity and Species Source databases) and
 28 Koenimbine was identified as a novel natural inhibitor based on virtual screening.

29

30 **Keywords:** *Klebsiella pneumoniae*, DNA Methylation, Genomics, Anti-Bacterial
 31 Agents

1 **1. Introduction**

2 *Klebsiella Pneumoniae* belongs to the genus of *Enterobacteriae* and classified as
3 one of the Carbapenems-Resistant-Enterobacteriaceae (CRE) organisms. This organism
4 causes infection in the nosocomial setting leading to a global threat as the ability of the
5 bacterial pathogens to acquire mobile genetic traits making them resistant towards
6 antibiotics. *Klebsiella Pneumoniae* causes a wide range of infection which includes the
7 urinary tract infection, pneumonia and liver abscess (Paczosa 2016). The current drug
8 targets the cellular process of the bacterial pathogens of the translation, transcription,
9 and replication. However, the bacterial pathogens are still able to develop resistant
10 towards the antibiotics (Petchiappan and Chatterji 2017).

11 The concerning emergence of this bacterial pathogen MDR has become a global
12 threat and this MDR and as mentioned by WHO (2012). These bacterial pathogens
13 clonally expand acquiring genetical traits allowing them to develop resistant causing an
14 increase prevalence affecting the human population by suppressing the mortality and
15 morbidity rate. The current antibiotics are not able to combat this bacterial pathogen
16 allowing them to acquire mobile genetic traits that make them resistant towards the
17 antibiotics. Although, the carbapenems and colistin are identified as the 2 most potent
18 antibiotics however this bacteria producing resistant towards these antibiotics and
19 identified as *Klebsiella Pneumoniae* Carbapenems(KPCs) producing (Giordano et al.
20 2017).

21 DNA Adenine Methyltransferase(Dam) is seen as a promising drug target as it
22 involves in the epigenetics regulatory machinery that helps in sustaining the viability of
23 the bacterial pathogens and regulates the bacterial pathogenicity (Natasya, Ismail, and
24 Kumar 2019). Evidently, Dam is seen to regulate the modulation of that bacterial
25 pathogen making them a promising drug target for future drug discovery
26 (Giacomodonato et al. 2014). Hence, this study aims to identify novel antibiotic drug
27 targets by targeting the Dam for *Klebsiella Pneumonia*.

28

29 **2.0 Materials and Methods**

30 This study involves several *in silico* methods that include subtractive genomic
31 analysis, molecular docking and ADME test. The detailed workflow is summarized in
32 Figure 1, Figure 2 and Figure 3.

2.1 Retrieval of bacterial proteome

All available strain *Klebsiella Pneumoniae* hypothetical protein sequence containing DNA adenine methyltransferase was obtained from UniProt (Universal Protein Database) by using keyword organisms; “*Klebsiella Pneumoniae*” and term; N6 DNA adenine methyltransferase and sequence containing unreviewed status are selected. The UniProt database is the largest protein database that is resourced with detailed annotation of protein (Bateman et al. 2017).

2.2 Identification of non-homology sequence analysis

All of the hypothetical protein sequences retrieved were first screened to the only select non-homologous sequence. The aim of this is to restrict and identify the hypothetical protein that is not similar to the human proteome. By performing BLASTp (Basic Local Alignment for protein) against the National Centre Biotechnology Institute database (NCBI). The threshold was set to an e-value of <0.0005 (Anishetty, Pulimi, and Pennathur 2005).

2.3 Identification of essential genes

In order to identify a potential drug target, the non-homologous hypothetical protein that was first identified should consist of essential genes that are beneficial for the constituent cellular process of the cell. The essential genes are important to be identified as the major constituent to the survival of the pathogens. The hypothetical proteins are screened by using BLAST against the Database of Essential Genes (DEG) and the threshold values were set to default parameters of E-value <0.0001 (Zhang, Ou, and Zhang 2004).

2.3 Broad-spectrum analysis

The hypothetical protein that contains all the required essential genes were then identified to be presented in a broad spectrum of the bacterial kingdom. Bacterial proteins are classified as broad-spectrum if the proteome is presented in more than 25 bacterial protein kingdoms. In order to identify these hypothetical proteins, the proteins of essential genes were screened by BLASTp with an E-value of 0.005.

1 **2.4 Drugability analysis**

2 DrugBank is extensively used for a new drug target in *silico* computational aided
3 drug design (Hosen et al. 2015). The database consists of the data with more
4 comprehensive information on the drug target and action of the drug. The drug data
5 included in the database are approved by the Food and Drug Administration (FDA). In
6 order to obtain the drugable protein, the hypothetical protein is subjected to BLASTp
7 with an E-value of 0.001.

8 **2.5 Non-homology analysis against gut microbiota**

9 Gut microbiota plays an important role in human gastrointestinal. The
10 homologous protein giving similarities to the human gut may interact and bind with the
11 gut flora proteins leading to adverse pharmacokinetic side effects in the host. Hence, a
12 homologous protein that is similar to the human gun is removed by using BLASTp with
13 an e-value of 0.0001 (Altschul, Stephen F Gish, Warren Miller, Webb Myers 1990).

14 **2.6 Subcellular localization**

15 PSORTb 3.0 is an accurate predictor for bacterial protein subcellular
16 localization(SCL). Gram-negative bacterial protein is composed of 5 major localization
17 which includes cytoplasmic, inner membrane, periplasmic, outer membrane and
18 extracellular (Yu et al. 2010). In this study, proteins that are located at the membrane
19 channel and cytoplasmic are selected as they are more profound as a drug target (Barh
20 et al. 2011). The subcellular localization identification is performed by using PsortB
21 and the drug target located at cytoplasmic is selected.

22 **2.7 Drug target property**

23 Drugs are usually enzymes and they are involved in binding, signalling and
24 communication. According to Bakheet (2009), good drug target are summarized to 8
25 major key properties namely, hydrophobicity >-142.4, length of amino > 550, signal
26 motif present, no PEST motif, more than 2 N-glycosylated amino acids, not more than
27 1 O-glycosylated ser and mean pI of <7.2, presence of the transmembrane helix and
28 membrane location is cytoplasmic. Therefore, all of these aspects are first analyzed to
29 determine a good drug target.

To predict the length of amino acids, the ExPasy server is used to analyze the length of amino acids, hydrophobicity and pI. (Bendtsen et al. 2004). In order to identify the presence of transmembrane helix(TMHHM), the analysis is done via the TMHHM method (<http://www.cbs.dtu.dk/~krogh/TMHHM/>) (Krogh et al. 2001). PEST region is identified by via (<http://emboss.cbr.nrc.ca/cgi-bin/emboss/pepfind>). To analyze the O -glycosylation, The NetOglyc program via (<http://www.cbs.dtu.dk/services/NetOGlyc/>) is used while for N-glycosylation ser is accessed via (<http://www.cbs.dtu.dk/services/NetNGlyc/>) (L J Jensen et al. 2003).

2.8 Anti-target non-homology analysis

The anti-target non-homology analysis is performed to eliminate the anti-target receptor (Raman, Yeturu, and Chandra 2008). Anti-target proteins are eliminated with an E-value set to 0.005.

2.9 Drug data property

ChEMBL database provides the bioactivity, molecule, target and drug data from various data which includes the medicinal chemistry literature. The database provides an association for ligand and the specific biological target (Davies et al. 2015). The identified hypothetical protein was observed to show more matches from the ChEMBL as a good drug target.

2.10 Virulence factor analysis

Virulence Factor Database (VFDB) provides an extensive comprehension of the Virulence Factors that are characterized by 16 dominant bacterial pathogens (Chen et al. 2005). These virulence factors are one of the most crucial factors that cause bacterial pathogens to colonize the host and being detrimental to the host cell.

2.11 Protein-protein interaction

Protein functions are the main complex of the cellular phenotype and they are not independent. The networks of the interacting proteins elucidate the protein function. To obtain the protein-protein interaction (PPI) of *Klebsiella Pneumoniae*, STRING(Search Tool for the Retrieval of Interacting Genes/Protein) is used (Lars J Jensen et al. 2009). Neighboring protein with a high confidence score of greater than 0.7 is included.

1 **2.12 Binding site prediction**

2 Drugs bind to a specific site of the protein. The interaction between the binding
3 site helps to understand the physicochemical interaction between the drugs and the
4 protein and further interactions. The predictions are done by using DogSiteScorer, an
5 automated algorithm for pocket and drug ability prediction. The pocket is predicted by
6 mapping the protein to the grid and the Difference of Gaussian is used to filter and
7 identify the pocket regions on the protein surface (Volkamer et al. 2012).

8 **2.13 Metabolic Pathway Analysis**

9 Comparative metabolic pathway analysis is performed to identify the unique
10 interaction between the host and the identified protein by KEGG (Kyoto Encyclopedia
11 of Genes and Genomes). The output gives KO (KEGG Orthology) assignments that will
12 generate the KEGG pathway(Kanehisa et al. 2010). Metabolic pathway analysis is
13 essential to elucidate the predicted putative drugs.

14 **2.14 Gene ontology**

15 As the protein identified is uncharacterized protein, it is therefore essential to
16 identify the specific biological role of the protein. Understanding the biological role of
17 protein provides an insight into the protein specific function (Omeersshffudin, and
18 Suresh 2019). ThisThe biological role of the identified protein is accessed via Gene
19 Ontology (GO). GO consortium is used for biology unification for all shared
20 eukaryotes. The consortium constructs 3 ontologies which are the biological process,
21 molecular function and the cellular component (Gene et al. 2011). The ontologies
22 provide essential information to understand the biological role of the protein in specific
23 organisms.

24 **2.15 Homology modeling**

25 Uncharacterized protein structure is not available in the Protein Data Bank
26 (PDB) whilst the structural mechanism is vital to facilitate the understanding of the
27 ligand interaction and also channelling between the targeted protein (Kumar 2011a). The
28 targeted 3D protein structure is constructed by using the fully automated server SWISS-
29 MODEL. Firstly, homology modelling of targeted protein is first compared against a
30 similar protein structure template (Kumar 2011b). The template for the identified

protein was identified as the protein structure of bacteria *Escherichia Coli* containing the DNA adenine methyltransferase PDB ID: 4RTL(Horton et al. 2015).

The modeled protein structure is then verified to check the protein quality of the stereochemical by using PROCHECK (Laskowski, Molecular, and Thornton 1993), ProSA-web where errors in the 3D structure are recognized (Wiederstein and Sippl 2007), and ERRAT.

2.17 Ligand preparation

7 ligands are identified as DNA methyltransferase inhibitors (DNMTi) from the literature review (Saldívar-gonzález et al. 2018). The ligands are Mahanine (PubChem ID: 375151), Curcumin (PubChem ID: 969516), EGCG (PubChem ID: 65064), Nanaomycin A (PubChem ID: 40586), Parthenolide (PubChem ID: 7251185), Quercetin(PubChem ID: 5280343) and Trimethylaurintricarboxylic acid (PubChem ID: 263071). Ligands that can be developed for putative drug targets should not violate 5 Lipinski rules and therefore were first validify by the SWISS-ADME server. Ligands that results in any violation are not further included for analysis.

The 2D structure is obtained in SDF format and is retrieved from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>.) The SDF file is converted to PDB via OpenBABEL and SMILES (<http://cactus.nci.nih.gov/services/translate/>) (Weininger 1988). The converted PDB structures are minimized to PDBQT with AutoDock Vina tools. Non-polar hydrogen is added and merges to the ligands and gaisteiger charges are computed. The torsion of the ligand is defined and saved as a PDBQT extension.

2.18 Molecular docking

Molecular docking is performed to identify ligands that bind with the lowest affinity score to develop potential putative drug targets by using AutoDock Vina. The default exhaustiveness is set to 1.0 Å. Identified protein is configured by first adding all hydrogens to the protein and merging the non-polar hydrogen. The gaisteiger charges are computed and the protein is saved as pdbqt file.

The grid box is set based on the predicted binding site with the configuration values of the center grid box of x,y,z. The size of the dimension grid box is set to 30.0 Å. The binding affinity score is observed.

2.19 Identification of novel inhibitor through molecular fingerprint search of the prioritized ligand

To identify novel inhibitor, fingerprint search is performed by NPASS(Natural Product Activity and Species Source databases) to search compound that is similar to the prioritized ligand based on docking of DNA methyltransferase inhibitors (Zeng et al. 2018). The fingerprint search is done using by inputting SMILES of a prioritized ligand with setting fingerprint type (pubchem-881 fp) with threshold ≥ 0.90 search by structure function in NPASS database.

2.20 Virtual screening

Virtual screening is performed to dock against clusters of ligands by using AutoDock. The setting is set to default parameters of 1.0 Å with a dimension grid of 30.0 Å. The analysis is performed based on the binding energy score.

2.21 ADMET Test

Absorption, distribution, metabolism, excretion, and toxicity(ADMET) are the major aspects that are processed by the body as soon as the drug is administered into the body(Zhong 2017). These pharmacokinetic properties are accessed to indicate the site of action of the drugs. These aspects are analyzed by using the pkCSM database (Pires, Blundell, and Ascher 2015). PkCSM database provides optimization of these pharmacokinetics properties by using the graph-based signatures.

3.0 Result & Discussion

The aim of this study is to identify potential drug target among hypothetical proteins of *Klebsiella Pneumonia* containing DNA adenine methyltransferase and to identify novel DNMT inhibitors. All 32-hypothetical proteome of *Klebsiella Pneumoniae* containing DNA adenine Methyltransferase were retrieved from Uniprot to be analyzed as potential druggable protein. The druggable protein is first characterized by a subtractive genomics approach that characterizes protein of which suffices the following criteria; non-homologous to the human host, contains essential genes, presented in broad-spectrum bacterial kingdom, and non-homologous to the human gut microbiota.

The first consideration of the subtractive genomic approach is non-homology analysis. Homologous protein that is presented in the human host may react with the molecules as they may carry unwanted toxicity. To decrease the adverse side effects, non-homologous protein is selected as putative drug targets by subjecting the proteins sequence to BLAST with an E-value of 10^{-3} . Essential genes are known to be indispensable to maintain the cellular processes of the bacterial proteome to survive (Zhang, Ou, and Zhang 2004). To identify essential genes in the bacterial proteome, the sequences are subjected to BLASTp against the DEG database with an expected value of < 0.0001 .

Threatening multiple bacterial infections would require the target drug able protein to be common in the broad-spectrum bacterial kingdom(White and Kell 2004). Multiple target antibacterial is most preferred to be developed as drugs. To foretell whether these bacterial proteomes are broad-spectrum, all 32 protein is searched by performing BLAST with E-value of less than 0.0001 against NCBI bacterial pathogens database. This resulted in all 32 as non-homologous, contain essential genes and are presented in a broad spectrum of the bacterial kingdom.

3.1 Druggability analysis

Protein that is druggable is defined as being able to bind with drug molecules high strength of binding interaction (Wishart et al. 2008). These are known as high-affinity binding between the protein and ligand where it results in stronger attractions of intermolecular forces. One notable source containing comprehensive drug data is DrugBank. According to Hosen (2015), Drug Bank contains small molecule drugs, biotech drugs that are approved by FDA, nutraceuticals and experimental drugs entries. Targeted proteins are then BLASTp against the DrugBank database and out of 32 only 26 protein is characterized as being druggable.

3.2 Human gut microbiota analysis

The term gut microbiota is described as the large population of organisms that colonizes the intestinal tracts (Jandhyala et al. 2015). Gut microbiota is highly associated with human inflammatory diseases. The role of pathogens in human gut microbiota co-evolved by having a symbiotic relationship and promoting replication of pathogens (Joseph M. Pickard, Melody Y. Zeng, Roberta Caruso 2018). Homologous protein may lead to unintentional blockage of the proteins that are in the gut flora that

could cause adverse effects (Shanmugham and Pan 2013). To prevent this, homologous protein is removed by BLASTp against the NCBI database of gut flora with a threshold of less than 0.005. Out of 26 protein, 19 protein is found to be non-homologous to the human gut.

3.3 Subcellular localization analysis

One important determinant of developing a putative drug target is the characterization of the subcellular localization as it exhibits the main function of the protein (Uddin et al. 2017). Localization of the protein determines the protein function (Chordia, Lakhawat, and Kumar 2017). Proteins that are located at cytoplasmic regions are seen to be more favorable as drug targets as they contain an abundant enzyme making it more feasible for a drug target. The cellular localization is predicted by using PsortB. 9 protein is characterized by locating at the cytoplasmic membrane while 10 protein is unknown.

3.3 Drug target property analysis

To further understand the drug property of this 9 protein, the protein is analyzed based on 8 major key properties summarized by Bakheet (2009). The 8 key properties of a good drug target are hydrophobicity >-142.4 , length of amino > 550 , a signal motif present, no PEST motif, more than 2 *N*-glycosylated amino acids, not more than 1 O-glycosylated ser and mean pI of <7.2 , presence of the transmembrane helix and membrane location is cytoplasmic.

One of the driven forces of a good drug target is having high hydrophobicity. The balance of hydrophobicity of protein is significant to the folding of protein and aggregation (Zbilut et al. 2003). The higher the hydrophobicity of the protein, the better the resultant in the folding of the protein which indirectly aids in the functionality of the protein (Chordia, Lakhawat, and Kumar 2017). The stabilization of hydrophobicity of protein is also found to affect the interactions of the binding affinity between the protein-ligand (Snyder et al. 2011). The result shows that all protein has hydrophobicity of <-0.142 .

Another observed key aspect of drug targets is the isoelectric point (pI) of the protein which is the overall charges of the amino acids. The pI value determines the pH of the protein and the protein solubility. Higher pI values indicate that the protein is basic and vice versa. The observed mean value of a good drug target is that the pI value

1 should be <7.2. 3 proteins were identified as having a mean pI of <7.2: A0A3P4EC49,
2 A0A3P4UG76, and A0A2U0NNR3. Table 1 shows the screening of the 3 proteins.

3 The desirable amino acid length for drug target should be greater than 550 amino
4 acids in total length. The greater the length of amino acids the greater the surface area
5 of the protein interactions with drugs. Here, all the drug targets have less than 550
6 overall amino acids which may due to the type of the bacteria species. The signal peptide
7 cleavage aids in the transportation of the protein of the endoplasmic reticulum across
8 the membrane (Zimmermann et al. 2011). However, the result indicates that the
9 presence of signalP is less significant due to the localization of the protein in the
10 cytoplasmic.

11 PEST regions are regions of peptide that is rich with Proline(P), glutamic acid
12 (E), Serine (S) and Threonine(T). Proteins that have one or more PEST regions are
13 associated with having shorter intracellular half-lives as they are reported to cause
14 protein degradation (Rogers, Wells, and Rechsteiner 1860). All of the protein sequences
15 observed consist of at least 1 PEST region. The presence of transmembrane helix is
16 amino acids that portray flanking regions however results show there well absent
17 transmembrane helix in all observed protein.

18 Glycosylation is widely crucial and found abundant in polypeptide chain
19 modification (Nothaft and Szymanski 2010). Bacterial protein possessed two
20 glycosylation state which is the N-linked and O-linked glycosylation (Dell et al. 2010).
21 According to Bakheet (2009), most bacteria protein drug targets either have greater than
22 two N-glycosylated amino acids or one or no O-glycosylated ser. 4 proteins resulted in
23 having one and no O linked glycosylated ser and 3 protein has more than 2 N linked
24 glycosylated amino acids.

25 Based on the drug property 3 protein are selected to be further screened as they
26 possessed more drug target characteristic which is: A0A3P4EC49, A0A3P4UG76, and
27 A0A2U0NNR3.

28 **3.4 Identification of putative drug target**

29 It is necessary to prognosticate the probable drug candidate to fail. Therefore,
30 the protein needs to first be identified as either an anti target or target protein. The anti
31 target protein is a protein receptor that when it binds to a drug, it will cause adverse
32 pharmacokinetics side effects. Here, all 3 proteins are not anti-target proteins.

Out of these 3 proteins, protein A0A2U0NNR3 is seen to have more desirable drug property when analyzed by using chEMBL with 10 matches. Figure 4 greatly summarizes the screening process to identify the potential drug target. Also, this protein shows the highest pocket binding score of 0.82 when analyzed by using DogSiteScorer. The protein then further studies for its metabolic pathway. Based on the metabolic pathway analysis in the KEGG server, the protein is found to carry a unique pathway of DNA mismatch repair. The gene ontology to describe the specific biological role of the protein indicates that the protein carries a gene that functions as a site-specific DNA methyltransferase (adenine specific activity). The metabolic pathway and gene ontology are summarized in Table 2.

In order to understand the virulence mechanism, the protein is queried to Virulence Factor Databased. The finding resulted that the bacteria protein virulence factors are; VFG010749 (sdhB) Dot/Icm type IV secretion system effector, VFG001959 (hddC) capsular polysaccharide heptosyltransferase, VFG000166 (pchE) dihydroaeruginosic acid synthetase PchE [Py], VFG000272 (ureE) urease accessory protein (ureE), metalloch and VFG002139 (cdsD) Type III secretion system inner membrane r. The virulence factor is summarized in Table 3.

3.4 Protein-protein interaction

Protein is not able to function alone, therefore the neighbouring protein is identified. Based on Figure 5, the interacting neighbouring protein helps to decipher the interactome mechanism in bacteria protein. Neighboring protein scored greater than 0.7 are considered as high confidence interacting protein which can be seen in Figure 6. The significant protein identified is tryptophanyl t-RNA synthetase, ribulose phosphate 3-epimerase, DNA mismatch repair endonuclease MutH, 3 dihydro quinate synthase, and DamX.

3.5 Molecular docking

Initially, the protein structure A0A2U0NNR3 is searched in Protein Data Bank. The search resulted that the protein structure is not available. The protein sequence is submitted to SWISS-MODEL for the construction of the protein structure. The modeled structure is imposed based on the retrieved template from the SWISS-MODEL template library of PDB ID: 4RTL. To ensure the structural quality of the model, the protein is first evaluated. The evaluation is done by using PROCHECK, ERRAT, and PROSA.

1 PROCHECK construct Ramachandran plot that gives 89.9% of residues in the favored
2 regions, 9.4% regions in the additional allowed regions and 0.4% for both residues in
3 generously and disallowed regions. According to Kumar (2015), good quality of protein
4 structure should give an overall of residues in the favored and allowed regions of over
5 90%. Here, the model gives a total of $89.9\% + 9.4\% = 99.3\%$.

6 The Z-score value from ProSA is -7.49 that indicates that the model is by the X-
7 ray crystallography method. ERRAT plot gives the overall factor quality score of
8 84.848 that gives an average quality score. The overall results can be referred to in
9 Figure 8 and based on the results, the protein model can be inferred for further study.
10 The ligands that are sourced from the literature review are first analyzed for any
11 violations of Lipinski Rules. Only ECGC is excluded as it gives 2 violations of Lipinski
12 Rules: N or O>10, NH or OH>5.

13 The inhibitors are docked and defined to which the predicted binding pocket by
14 using DogSiteScorer. Figure 7 shows the area in which the active region is identified.
15 The dimension grid is set to 30 Å. Based on the result tabulated in Table 4, Mahanine
16 gives the highest binding affinity score of -10.8 kcal/mol followed by Nanaomycin A
17 (-10.0 kcal/mol). Timethylaurintricarboxylic acid (-9.4 kcal/mol), Quercetin (-9.3
18 kcal/mol) and Curcumin give the lowest binding affinity score of -8.5 kcal/mol. The
19 molecular docking result is tabulated in Table 4.

20 **3.6 Virtual screening**

21 Based on the binding affinity score, Mahanine is further screened to identify a
22 compound with a similar composition as Mahanine. Mahanine is a carbazol alkaloid
23 extracted from the plant species *Murraya Koenigii*. Mahanine is previously reported as
24 one of the DNMT inhibitors. In order to identify new novel antibiotics derived from
25 plant, fingerprint search is performed. Based on the calculated fingerprint similarities,
26 22 compounds are found to be similar to Mahanine. 3 compound name is not available.
27 2 compounds: Koenimbine and Grinibine are reported as one of the carbazole alkaloids
28 of *Murraya Koenigii*.(Ramos et al. 2015).

29 The virtual screen result is tabulated in Figure 5. All the ligands are seen to no
30 violate Lipinski Rules. The virtual screening is performed by using AutoDock. The
31 virtual screening resulted that Koenimbine and Claurailas B have the highest binding
32 affinity score of -5.97 kcal/mol. Koenimbine is also one of the carbazole alkaloids of

the same plant species, can be explored as the novel antibiotics. Figure 9 shows the binding of Koenimbine and Table 5 shows the ligand-binding score and drug-likeness property.

3.8 ADMET test

The ADMET test was further evaluated for Koenimbine to understand the absorption, distribution, metabolism, excretion and toxicity of the compound. The analysis is analyzed and interpreted based on the guidelines of pkCSM (Pires, Blundell, and Ascher 2015). The compound absorption analysis resulted that the compound has high caco_2 permeability. The water solubility is $-4.618 \log/\text{mol/L}$. The compound absorbance in the human intestine is 93.479% which indicates a good absorption rate. It also results in being a P-glycoprotein II inhibitor which is significant in imposing pharmacokinetics effects.

In terms of distribution, koenimbine is seen to have slightly low distribution in the tissue plasma as the volume of distribution is 0.654 L/kg. Based on the toxicity measured, the *T.pyrifomis* toxicity indicates that the compound is highly toxic against bacteria with 0.948 ug/L. The compound is also seen to be mutagenic against bacteria which indicates that it can cause detrimental to the bacteria. The result is summarized in Table 6.

4.0 CONCLUSION

We identified and prioritized one DAM drug target from 32 proteins of *Klebsiella pneumonia* based on subtractive genomics, based on drug property, pocket analysis, pathway analysis, and structure analysis. Our proposed new computational pipeline approach helps to find out the drug target in a rapid way. We have also discovered that Koenimbine a natural bioactive chemical compound of *Murraya Koenigii* can potentially be a novel antibiotic We have also discovered that Mahanine can be used as a potential novel antibiotic to inhibit DAM containing bacterial pathogens.

Conflict of Interest: The authors declared that they have no conflict of interest.

Authors Contribution

Umairah Natasya Mohd Omeershffudin: Conceptualization, Investigation, Methodology, Formal analysis, Data curation, Writing - original draft.

1 Suresh Kumar : Conceptualization, Data curation, Validation, Visualization, Project
2 administration, Resources, Supervision, Roles/Writing - review & editing

4 ACKNOWLEDGMENT

5 The authors would like to thank the Faculty of Health Science, Management and
6 Science University for supporting facilities to carry out the research work.

8 **Funding:** No funding for this research.

10 REFERENCES

- 11 [1]. Altschul, Stephen F Gish, Warren Miller, Webb Myers, Eugene W. 1990.
12 "Basic Local Alignment Tool.Pdf."
- 13 [2]. Anishetty, Sharmila, Mrudula Pulimi, and Gautam Pennathur. 2005. "Potential
14 Drug Targets in Mycobacterium Tuberculosis through Metabolic Pathway
15 Analysis." *Computational Biology and Chemistry* 29(5): 368–78.
- 16 [3]. Bakheet, Tala M, and Andrew J Doig. 2009. "Properties and Identification of
17 Human Protein Drug Targets." 25(4): 451–57.
- 18 [4]. Barh, Debmalya, Sandeep Tiwari, Neha Jain, and Amjad Ali. 2011. "In Silico
19 Subtractive Genomics for Target Identification in Human Bacterial
20 Pathogens." 72(November 2018): 162–77.
- 21 [5]. Bateman, Alex et al. 2017. "UniProt: The Universal Protein Knowledgebase."
22 *Nucleic Acids Research* 45(D1): D158–69.
- 23 [6]. Bendtsen, Jannick Dyrlov, Henrik Nielsen, Gunnar Von Heijne, and Søren
24 Brunak. 2004. "Improved Prediction of Signal Peptides : SignalP 3 . 0." : 783–
25 95.
- 26 [7]. Chen, Lihong et al. 2005. "VFDB: A Reference Database for Bacterial
27 Virulence Factors." *Nucleic Acids Research* 33(DATABASE ISS.): 325–28.
- 28 [8]. Chordia, Nikita, Kapil Lakhawat, and Anil Kumar. 2017. "Identification of
29 Drug Target Properties and Its Validation on Helicobacter Pylori." *Canadian
30 Journal of Biotechnology* 1(1): 44–49.
- 31 [9]. Davies, Mark et al. 2015. "ChEMBL Web Services : Streamlining Access to
32 Drug Discovery Data and Utilities." 43(April): 612–20.
- 33 [10]. Dell, Anne, Alaa Galadari, Federico Sastre, and Paul Hitchen. 2010.
34 "Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes
35 and Eukaryotes." *International Journal of Microbiology* 2010.
- 36 [11]. Gene, et al. 2011. "Gene Ontology : Tool for the Unification of
37 Biology." 25(1): 25–29.
- 38 [12]. Giacomodonato, N, Fernanda Buzzola, Mauro D Garc, and Marina
39 Gallo Calder. 2014. "Dam Methylation Regulates the Expression of SPI-5-
40 Encoded SopB Gene in Salmonella Enterica Serovar Typhimurium." 16: 615–
41 22.
- 42 [13]. Giordano, Author Cesira et al. 2017. "Expansion of KPC-Producing
43 Klebsiella Pneumoniae with Various MgrB - Mutations Giving Rise to

- Colistin-Resistance : The Role of ISL3 on Plasmids Department of Internal Medicine , Infectious Diseases Unit , University Hospital of Department of Medical Microbiology.” *International Journal of Antimicrobial Agents*.
<https://doi.org/10.1016/j.ijantimicag.2017.10.011>.
- [14]. Horton, John R et al. 2015. “Structures of Escherichia Coli DNA Adenine Methyltransferase (Dam) in Complex with a Non-GATC Sequence : Potential Implications for Methylation-Independent Transcriptional Repression GA.” 43(8): 4296–4308.
- [15]. Hosen, S M Zahid et al. 2015. “Drug Bank : An Update-Resource for in Silico Drug Discovery Drug Bank : An Update-Resource for in Silico Drug Discovery.” (June 2012).
- [16]. Jandhyala, Sai Manasa et al. 2015. “Role of the Normal Gut Microbiota.” 21(29): 8787–8803.
- [17]. Jensen, L J, R Gupta, H Stærfeldt, and S Brunak. 2003. “Prediction of Human Protein Function According to Gene Ontology Categories.” 19(5): 635–42.
- [18]. Jensen, Lars J et al. 2009. “STRING 8 — a Global View on Proteins and Their Functional Interactions in 630 Organisms.” 37(October 2008): 412–16.
- [19]. Joseph M. Pickard*, Melody Y. Zeng*, Roberta Caruso*, and Gabriel Núñez Department. 2018. “Gut Microbiota:Role in Pathogen Colonization, Immune Responses and Inflammatory Disease.” 279(1): 70–89.
- [20]. Kanehisa, Minoru, Susumu Goto, Miho Furumichi, and Mao Tanabe. 2010. “KEGG for Representation and Analysis of Molecular Networks Involving Diseases and Drugs.” 38(October 2009): 355–60.
- [21]. Krogh, Anders, È Larsson, Gunnar Von Heijne, and Erik L L Sonnhammer. 2001. “Predicting Transmembrane Protein Topology with a Hidden Markov Model : Application to Complete Genomes.”
- [22]. Kumar, Suresh. 2011a. “Comparative Modeling and Molecular Docking of Orphan Human CYP4V2 Protein with Fatty Acid Substrates: Insights into Substrate Specificity.” *Bioinformation* 7(7): 360–65.
- [23]. ———. 2011b. “Molecular Modeling and Identification of Substrate Binding Site of Orphan Human Cytochrome P450 4F22.” *Bioinformation* 7(4): 207–10.
- [24]. ———. 2015. “Computational Identification and Binding Analysis of Orphan Human Cytochrome P450 4X1 Enzyme with Substrates.” : 1–10.
- [25]. Laskowski, Roman Aleksander, European Molecular, and Janet Thornton. 1993. “PROCHECK : A Program to Check the Stereochemical Quality of Protein Structures.” (March 2014).
- [26]. Natasya, Umairah, Nurul Azira Ismail, and Suresh Kumar. 2019. *Computational Identification of Novel Inhibitors against Klebsiella Pneumoniae DNA Adenine Methyltransferase*.
- [27]. Nothaft, Harald, and Christine M. Szymanski. 2010. “Protein Glycosylation in Bacteria: Sweeter than Ever.” *Nature Reviews Microbiology* 8(11): 765–78.
- [28]. Paczosa, Michelle K. 2016. “Klebsiella Pneumoniae : Going on the Offense with a Strong Defense.” 80(3): 629–61.
- [29]. Petchiappan, Anushya, and Dipankar Chatterji. 2017. “Antibiotic

- Resistance : Current Perspectives.”
- [30]. Pires, Douglas E V, Tom L Blundell, and David B Ascher. 2015. “PkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures.”
- [31]. Raman, Karthik, Kalidas Yeturu, and Nagasuma Chandra. 2008. “TargetTB : A Target Identification Pipeline for Mycobacterium Tuberculosis through an Interactome , Reactome and Genome-Scale Structural Analysis.” 21: 1–21.
- [32]. Ramos, Lorena et al. 2015. “Bioorganic & Medicinal Chemistry Flavonoids as Noncompetitive Inhibitors of Dengue Virus NS2B-NS3 Protease : Inhibition Kinetics and Docking Studies.” *BIOORGANIC & MEDICINAL CHEMISTRY*: 2–6.
- [33]. Rogers, Scotr, Rodney Wells, and Martin Rechsteiner. 1860. “Acid Sequences Common to Rapidly Degraded Proteins : The PEST Hypothesis.” 234.
- [34]. Saldívar-gonzález, Fernanda I et al. 2018. “Inhibitors of DNA Methyltransferases From Natural Sources : A Computational Perspective.” 9(October): 1–10.
- [35]. Shanmugham, Buvaneswari, and Archana Pan. 2013. “Identification and Characterization of Potential Therapeutic Candidates in Emerging Human Pathogen Mycobacterium Abscessus: A Novel Hierarchical In Silico Approach.” *PLoS ONE* 8(3).
- [36]. Snyder, Phillip W et al. 2011. “Mechanism of the Hydrophobic Effect in the Biomolecular Recognition of Arylsulfonamides by Carbonic Anhydrase.”
- [37]. Uddin, Reaz et al. 2017. “Identification and Characterization of Potential Druggable Targets among Hypothetical Proteins of Extensively Drug Resistant Mycobacterium Tuberculosis (XDR KZN 605) through Subtractive Genomics Approach.” *European Journal of Pharmaceutical Sciences*.
- [38]. Umairah, Omeershffudin, and Kumar Suresh. 2019. “In Silico Approach for Mining of Potential Drug Targets from Hypothetical Proteins of Bacterial Proteome.” *International Journal of Molecular Biology* 4(4): 145–52.
- [39]. Volkamer, Andrea et al. 2012. “Combining Global and Local Measures for Structure-Based Druggability Predictions.”
- [40]. Weininger, David. 1988. “SMILES, a Chemical Language and Information System.” : 31–36.
- [41]. White, Thomas A., and Douglas B. Kell. 2004. “Comparative Genomic Assessment of Novel Broad-Spectrum Targets for Antibacterial Drugs.” *Comparative and Functional Genomics* 5(4): 304–27.
- [42]. WHO. 2012. “The Evolving Threat of Antimicrobial Resistance Options for Action.”
- [43]. Wiederstein, Markus, and Manfred J Sippl. 2007. “ProSA-Web : Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins.” 35: 407–10.
- [44]. Wishart, David S et al. 2008. “DrugBank : A Knowledgebase for Drugs , Drug Actions and Drug Targets.” 36(November 2007): 901–6.
- [45]. Yu, Nancy Y. et al. 2010. “PSORTb 3.0: Improved Protein Subcellular

- 1 Localization Prediction with Refined Localization Subcategories and
- 2 Predictive Capabilities for All Prokaryotes.” *Bioinformatics* 26(13): 1608–15.
- 3 [46]. Zbilut, Joseph P et al. 2003. “Protein Aggregation / Folding : The Role
- 4 of Deterministic Singularities of Sequence Hydrophobicity as Determined by
- 5 Nonlinear Signal Analysis of Acylphosphatase and A b (1 – 40).”
- 6 85(December): 3544–57.
- 7 [47]. Zeng, Xian et al. 2018. “NPASS: Natural Product Activity and Species
- 8 Source Database for Natural Product Research, Discovery and Tool
- 9 Development.” *Nucleic Acids Research* 46(D1): D1217–22.
- 10 [48]. Zhang, Ren, Hong-yu Ou, and Chun-ting Zhang. 2004. “DEG : A
- 11 Database of Essential Genes.” 32: 271–72.
- 12 [49]. Zhong, H A. 2017. “ADMET Properties: Overview and Current
- 13 Topics.” : 113–33.
- 14 [50]. Zimmermann, Richard, Susanne Eyrisch, Mazen Ahmad, and Volkhard
- 15 Helms. 2011. “Protein Translocation across the ER Membrane.” *Biochimica et*
- 16 *Biophysica Acta - Biomembranes* 1808(3): 912–24.

1 FIGURES LEGENDS

2 **Figure 1:** Overall workflow of identification of protein putative drug target

3 **Figure 2:** Extended workflow summarization

4 **Figure 3:** Figure shows the virtual screening method applied

5 **Figure 4:** Analysis summarization of the subtractive genomic analysis of the 32-
6 hypothetical protein of *Klebsiella Pneumoniae* containing DNA Adenine
7 Methyltransferase.

8 **Figure 5:** Shows the interacting protein between the main protein JG24_22730 of a
9 specific site of DNA Adenine Methyltransferase and the functional partner

10 **Figure 6:** The detailed predicted functional partner of the protein interaction with the
11 main protein of JG24_27300(Site-specific DNA-methyltransferase). The result
12 demonstrates scores and of which the function is either; neighborhood, gene fusion,
13 concurrence, and coexpression.

14 **Figure 7:** (A) Protein structure of A0A2U0NNR3 that is evaluated by using PyMOL.
15 Red: helix, yellow: sheets and green: loops. (B) The active site that is demonstrated by
16 the region colored peach is the binding site of the protein predicted by DogSiteScorer

17 **Figure 8:** Structure quality assessment (A) Ramachandran plot of protein
18 A0A2U0NNR3 by using SwissModel (B)Z-score of the residues (ProSA) (C) Local
19 model quality based on the sequence position

20
21 **Figure 9:** Binding of Koenimbine at the active site of protein A0A2U0NNR3

22

23

24

25

26

27

28

29

30

31

32

33

34

35

1 **TABLES**

2 **Table 1:** Table 1 shows the results when performing the ChEMBL approach for
3 identified targets and DogSiteScorer score pocket binding prediction of all 3
4 hypothetical proteins.

Uniprot Id	ChEMBL - targets identified	DoGSiteScorer
A0A3P4EC49	3	0.81
A0A3P4UG76	5	0.81
A0A2U0NNR3	10	0.82

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

1 **Table 3:** Virulence factor of the hypothetical protein of A0A2U0NNR3.

UniProt ID	Virulence factors
A0A3P5S6Q8	<u>VFG010749</u> (sdhB) Dot/Icm type IV secretion system effector
	<u>VFG001959</u> (hddC) capsular polysaccharide heptosyltransferase
	<u>VFG000166</u> (pchE) dihydroaeruginosic acid synthetase PchE [Py]
	<u>VFG000272</u> (ureE) urease accessory protein (ureE), metalloch
	<u>VFG002139</u> (cdsD) Type III secretion system inner membrane r

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

1 **Table 4:** Binding affinity between the protein A0A2U0NNR3 against the 6 ligands

Bioactive compound	Binding affinity(kcal/mol)
Curcumin	-8.5
Mahanine	-10.8
Nanaomycin A	-10.0
Parthenolide	-7.6
Quercetin	-9.3
Trimethylaurintricarboxylic acid	-9.4

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

Table 5: Binding affinity score between the protein A0A2U0NNR3 against 22 ligands from similarity findings and drug-likeness property

Natural product name	Product name	Pubchem ID	Binding energy Kcal/mol	Rotatable bonds	ADME TEST (Lipinski)
NPC162484	Koenimbine	97487	-5.97	2	Yes
NPC190007	Mahanimbilol	5353739	-5.95	2	Yes; 1 violations
NPC193777	Glybomine B	11208660	-5.96	2	Yes
NPC201697	Clausenawalline C	57409124	-5.96	2	Yes
NPC205934	Mahanimbine	167963	-5.95	2	Yes; 1 violation: MLOGP>4.15
NPC209769	Claurailas B	51039826	-5.97	2	Yes
NPC212535	3-Methyl-9H-Carbazol-2-ol	3459141	-5.94	2	Yes
NPC243834	Sid506287	375146	-5.95	2	Yes
NPC267423	n.a.	375144	-5.96	2	Yes; 1 violation: MLOGP>4.15
NPC291535	Sid506290	375149	-5.94	2	Yes
NPC43787	Grinimibine	96943	-5.92	2	Yes
NPC470931	n.a.	71717470	-5.89	2	Yes; 1 violation: MLOGP>4.15
NPC475085	Sid506292	375151	-5.91	2	Yes
NPC475112	Sid506289	375148	-5.96	2	Yes
NPC476044	n.a.	44583761	-5.89	2	Yes; 1 violation: MLOGP>4.15
NPC476106	5-(3,5-Dimethyl-3,11-dihydropyrano[3,2-a]carbazol-3-yl)-2-methylpent-1-en-3-ol	375155	-5.96	2	Yes
NPC477532	3-[(2E)-3,7-dimethylocta-2,6-dienyl]-8-[3-[(2E)-	71725436	-5.95	2	No; 2 violations: MW>500,

	3,7-dimethylocta-2,6-dienyl]-9-methoxy-3,5-dimethyl-11H-pyrano[3,2-a]carbazol-10-yl]-3,5-dimethyl-11H-pyrano[3,2-a]carbazol-7-ol				MLOGP>4.15
NPC48353	Glycoborinine	10446329	-5.95	2	Yes
NPC70259	(+)-Mahanimbicine	4072580	-5.94	2	Yes; 1 violation: MLOGP>4.15
NPC70956	Euchrestine B	15060943	-5.96	2	Yes; 1 violation: MLOGP>4.15
NPC72211	(-)-O-Methylmahanine	71716235	-5.96	2	Yes; 1 violation: MLOGP>4.15
NPC94943	Siamenol	477436	-5.95	2	Yes

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

1 **Table 6 : ADMET test for Koenimbine**

Absorption						
Water solubility Numeric (log mol/L)	Caco2 permeability Numeric (log Papp in 10 ⁻⁶ cm/s)	Intestinal absorption (human) (% Absorbed)	Skin permeability Numeric (log Kp)	P-glycoprotein substrate	P-glycoprotein I inhibitor	P-glycoprotein II inhibitor
-4.618	1.411	93.479	-2.742	Yes	No	Yes
Distribution						
VDss(human)		Fraction unbound (human)		BBB permeability (log BB)		CNS permeability (log PS)
0.654		0.024		0.421		-1.789
Metabolism						
CYP2D6 substrate	CYP3A4 substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor
No	Yes	Yes	Yes	yes	No	No
Excretion						
Total clearance (log ml/min/kg)			Renal OCT2 substrate			
0.477			No			
Toxicity						
AMES toxicity		Max tolerated dose (human) (log mg/kg/day)	Hepatotoxicity	Skin sensitization		Minnow toxicity (log mM)
Yes		-0.312	Yes	No		-0.436

2

3

1

2 **Figure 1:** Overall workflow of identification of protein putative drug target

3

4

1
2 **Figure 2:** Extended workflow summarization
3
4

1

2 **Figure 3:** Figure shows the virtual screening method applied

3

4

5

6

7

8

9

Figure 4: Analysis summarization of the subtractive genomic analysis of the 32 hypothetical protein of *Klebsiella Pneumoniae* containing DNA Adenine Methyltransferase.

1
2 **Figure 5:** Shows the interacting protein between the main protein JG24_22730 of a
3 specific site of DNA Adenine Methyltransferase and the functional partner

Figure 6: The detailed predicted functional partner of the protein interaction with the main protein of JG24_27300(Site-specific DNA-methyltransferase). The result demonstrates scores and of which the function is either; neighborhood, gene fusion, concurrence, and coexpression.

1
2 **Figure 7:** (A) Protein structure of A0A2U0NNR3 that is evaluated by using PyMOL.
3 Red: helix, yellow: sheets and green: loops. (B) The active site that is demonstrated by
4 the region colored peach is the binding site of the protein predicted by DogSiteScorer.
5

Figure 8: Structure quality assessment (A) Ramachandran plot of protein A0A2U0NNR3 by using SwissModel (B)Z-score of the residues (ProSA) (C) Local model quality based on the sequence position

Figure 9: Binding of Koenimbine at the active site of protein A0A2U0NNR3