ORIGINAL ARTICLE

In Silico Comparative Structural and Residue Interaction Network Analysis of MATE Efflux Proteins in *P. aeruginosa* and *S. aureus*

Vandna Kumari[1] · **Shamsher S. Kanwar[1]**

Abstract

Bacterial multidrug resistance against front-line antibiotics has been widely observed by scientists and epidemiologists. Multidrug and toxic extrusion proteins are active transporters and play a vital role in the efflux system. Their main generalized function is the exclusion of various antibiotics and xenobiotics out of their host cells. Despite their wide presence in all living kingdoms, the structure–function relationship is not well understood. Protein characterization and structural analysis may help to figure out the important aspects of structure–function relationships. In the present study, multidrug and toxic extrusion protein sequences of *P. auruginosa* and *S. aureus* were characterized and comprehensive analysis of their structural levels was done. Residue interaction network algorithm further insight into amino acid residues has been designed. These proteins were also analyzed for hydrophilic and hydrophobic amino acid residues which indicated the predominance of hydrophobic amino acids in bacterial multidrug and toxic extrusion proteins primarily existing in the transmembrane region of the host cell. The residues involved in tertiary structure formation and the analysis of the network that emerges due to the interacting residues may reveal additional information about the structure–function relationship in a protein. The residue interaction networks were generated for these bacterial homologs of multidrug and toxic extrusion proteins has revealed the scale-free degree distribution and hydrophobic amino acids to be the major actors among these networks.

Keywords MATE · Gram-positive bacteria · Gram-negative bacteria · RIN · Antibiotics

1 Introduction

Pathogenic bacteria because of indiscriminate use of antibiotics have developed marked antibiotics, xenobiotics resistance and widely spread all over the world. Efflux pump systems are one of the major causes responsible for bacterial drug resistance, sustaining their survival [1]. Efflux pumps are transporter proteins localized in the cytoplasmic membrane and are active transporters requiring a source of chemical energy to perform their function [2]. These efflux pumps extrude a wide variety of structurally unrelated compounds and basically belong to five different families such as small multidrug resistance (SMR), resistance-nodulation-division (RND), ATP binding cassette (ABC), major facilitator superfamily (MFS) and multidrug and toxic compound extrusion (MATE).

MATE transporters facilitate the active efflux of a wide range of chemically and structurally diverse substances, including antimicrobials and chemotherapeutics, contributing to multidrug resistance in pathogenic bacteria and malignancies [3]. MATE proteins are present in all living kingdoms [4] and are secondary active transport systems which have a 12-membrane helix topology [5] driven by either Na^+ or H^+ [6, 7]. MATE efflux family proteins can range from ~400 to ~700 amino acids (aa)in length [8] whose toxic exclusion is important to maintain cell equilibrium [9]. The helices in MATE protein structure are arranged within two lobes which forms the cavities leading to the binding of substrates [10]. Multidrug efflux pumps work at the frontline to shield bacteria against antimicrobials by lowering drug intracellular concentrations [11].Their transporter function mostly energized by Na^+ confers multidrug resistance against bacterial pathogens and cancer cells [9]. Plants protect themselves against microbial pathogens by producing secondary metabolites through MATE transporters [4]. Whereas in human liver and kidney, they extrude xenobiotic cations [11]. MATE

✉ Shamsher S. Kanwar
kanwarss2000@yahoo.com

[1] Department of Biotechnology, Himachal Pradesh University, Summer Hill, Shimla 171005, India

family proteins pursue a mechanism as same as MFS transporter, i.e. rocker-switch mechanism [9]. MATE efflux proteins function as antiporters, causing the transporter to alter shape from outward to inward facing. A chemical substance (drug) attaches to the protein in the inward-facing conformation, causing a switch to the opposite shape and the drug to be extruded [12]. MATE has a substrate binding pocket at either the N-lobe or the C-lobe, and the substrate is extruded by bending the surrounding helix [13].

Protein characterization involves the depiction of a protein molecule's biological, chemical and physical properties. Protein computational characterization is the first necessary step to figure out the biological role of a protein [14] and is therefore important to determine the protein's current state. It also plays an important role in the study of the protein characteristics such as domains, oligomeric state, post-translational modifications, protein–protein interactions and protein–ligand interactions [15]. The early stage stability comparison of proteins should be based on computational followed by biophysical characterization [16]. A protein may also be visualized as a network of interacting residues and analysis of such a network that emerges due to interacting residues provides additional information about the structural and functional roles of the residues [17]. The protein interaction networks map depicts which proteins physically interact with one another [18]. The interface residue prediction can improve the understanding of molecular mechanisms of related process and functions [19]. Therefore, in the present study, an algorithm for residue interaction network (RIN) generation has been designed and comprehensive structural analysis has been used to reveal the important aa in MATE efflux transporter proteins.

2 Methodology

2.1 Dataset Retrieval and MATE Efflux Proteins Characterization

Dataset for MATE proteins of *P. aeruginosa* and *S. aureus* were retrieved from NCBI database [20]. These sequences were then exposed to the sorting and filtering in order to attain the non redundant sequences. The unique sequences were extracted on the basis of aspects such as removal of partial and uncultured sequences from the dataset, extraction out of the sequences having 400–700 aa length, grouping and sorting out of most similar sequences from the dataset were multiple alignment tool Multalin [21]. Multalin is a tool for multiple sequence alignment either proteins or nucleic acids and based on the dynamic programming approach.

2.2 Physicochemical Parameters Characterization

The physical parameters of MATE efflux proteins in *P. aeruginosa* and *S. aureus* were analyzed on the basis of number of aa and composition, molecular weight, instability index (II), theoretical pI, grand average hydrophobicity (GRAVY) and aliphatic index (AI)using ExPASy–ProtParam tool (http://web.expasy.org/protparam).

2.3 Primary Structure Analysis

The primary structure analysis of the polypeptide sequence for each MATE protein sequence was performed using ExPASy–ProtParam tool [22]. The number of hydrophilic aa, hydrophobic aa and their ratio was calculated for each MATE protein sequence using the Eqs. 1, 2 and 3 where $n_{hydrophob}$, $n_{hydrophil}$ and $Ratio_{hydrophob\_hydrophil}$ correspond to the number of hydrophilic aa, hydrophobic aa and their ratio. The average number of hydrophilic and hydrophobic aa in each protein sequence was also calculated.

$$n_{hydrophob} = \sum_{i=1}^{n}(n_G + n_A + n_V + n_L + n_I + n_M + n_P + n_F) \tag{1}$$

$$n_{hydrophil} = \sum_{i=1}^{n}(n_Q + n_N + n_S + n_H + n_T + n_Y + n_C + n_W) \tag{2}$$

*n is the total number of aain a protein sequence, n_G is the number of glycine aa, n_A is the number of alanine aa, n_V is the number of valine aa, n_L is the number of leucine aa n_I is the number of isoleucine aa, n_M is the number of methionine aa, n_P is the number of proline aa, n_F is the number of phenylalanine aa, n_Q is the number of glutamine aa, n_N is the number of asparagine aa, n_S is the number of serine aa, n_H is the number of histidine aa, n_T is the number of threonine aa, n_Y is the number of tyrosine aa, n_C is the number of cysteine aa, n_W is the number of tryptophan aa.

$$Ratio_{hydrophob\_hydrophil} = n_{hydrophob}/n_{hydrophil} \tag{3}$$

Furthermore, the helix topologies and number of transmembrane helices for MATE protein sequences were identified using web-based tools TMHM [23] and HMMTOP [24, 25]. TMHMM is a method of prediction of a number of transmembrane helices based on Hidden Markov model. HMMTOP is the web-based server for prediction of the topology of transmembrane membrane proteins and localization of helical transmembrane segments as well.

2.4 Secondary Structure Analysis

The secondary structure of characterized MATE efflux proteins in *P. aeruginosa* and *S. aureus* includes number of a-helices, b-turn, extended strand, b-sheet, coils. These secondary structure features were analyzed by ExPASy SIB Bioinformatics SOPMA tool. (https://npsaprabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html).

2.5 Tertiary Structure Analysis

The tertiary structures for MATE protein sequences were predicted using Modeller 9.21 [26]. Modeller provides a homology modeling-based approach for three-dimensional protein structure prediction. The protein identity was considered as > 25% for each protein. The observed PDB files for the tertiary structure of proteins were energy minimized with Swiss-Pdb viewer [27] in order to attain the protein model in an energetically favorable state. These tertiary modeled structures were verified as well as Ramachandran plots were constructed using PROCHECK tool [28] which helped in the visualization of energetically allowed and favorable regions for phi and psi backbone dihedral angles.

2.6 Molecular Dynamic Simulations and structure Optimization

Molecular dynamic simulations was performed using Gromacs 5.0 (http://www.gromacs.org/) to test the best model's stability. Structures of the predicted tertiary protein structures were optimized with the GROMACS 5.0 (GROningen MAchine for Chemical Simulations) package using the GROMOS96 53a6 force field for 1000 ns. The protein tertiary structures were subjected to a steepest descent energy minimization for 50,000 steps. The topology parameters for proteins were created using the GROMACS program.

2.7 RIN Generation

Residue interaction network has been generated for these energy minimized proteins. Cytoscape [29] was used to create a network. The C-αatoms of aa residue were considered as nodes and the distance between them we reconsidered as edges. The distances for all aa residues were calculated computationally. RIN provides the analysis of network centrality parameters such as degree centrality, closeness, and betweenness which indicate that the network characteristics of residues in quaternary interactions are differentiable from those of other residues. Centrality quantifies the topological importance of a node or edge in a network. The distance threshold for interacting residues was taken 4.0 to 8.5 Angstrom for each interaction of C_α atoms in aa residues.

Algorithm Design for RIN

The protein tertiary structure data constitute the three-dimensional coordinates for each atom throughout the protein structure. Assume the coordinates in protein structure data be x_i, y_i, z_i. A and B be considered as the first and second C-α atom.

Choose;

if x_i, y_i, $z_i \in$ C-α atoms: for each structural coordinates.

The input function includes the three dimensional coordinates (x_i, y_i, z_i) belonging t = to either A or B such that:

$$(x_1, x_2), (y_1, y_2), (z_1, z_2)$$

where {x_1, y_1, $z_1 \in$ A; x_2, y_2, $z_2 \in$ B.

Distances between any two C-α atom coordinates may provide the information about the bond formation among them at an optimal distance measure. The distances were calculated by the Euclidian distance calculation approach and a threshold for optimal distance corresponds to the bond-forming C-α atom.

Distance between two C-α atoms, $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$

In this equation x_1, y_1, z_1 corresponds to structural coordinates of first C-α of atom A, and x_2, y_2, z_2 corresponds to structural coordinates of the second C-α of atom B.

This algorithm is based on the RIN generation through the extraction of C-α atom coordinates from three-dimensional structure coordinates and Euclidian distance calculation. The distances calculated were then exposed to RIN generation and various centrality parameters analysis.

Degree centrality, $C_d(V)$

In a network graph, degree centrality is determined by the total number of direct linkages between nodes [30]. Nodes with a large number of neighbors (i.e. edges) have high degree centrality.

$$C_d(V) = \deg(V)$$

Closeness centrality, $CC(V)$

The term closeness centrality refers to the cumulative distances between two nodes [31]. It specifies which vertices have the shortest path to all others. The greater value of distance signifies lesser closeness.

$$C_C(V) = 1/\sum \text{dist}(u, v); u \in v$$

Eccentricity centrality, $C_e(V)$

The length of the largest shortest path beginning at a node in a graph is defined as its eccentricity i.e. maximum shortest path length from node u to all other nodes u in v [32]. The reciprocal of a node's eccentricity is defined as eccentricity centrality, $C_e(V)$.

$$E_{cc}(V) = \max \text{dist}(u, v) \, ; u \in v$$

The lower value of E_{cc} corresponds to higher C_e.

$$C_e(V) = 1/E_{cc}(V)$$

Betweenness centrality, $C_b(V)$

Betweenness centrality is used to quantify one node's position as a mediator in a network. If one node locates in the sole channel that other nodes must travel through, then this node is likely to be essential and has a high betweenness centrality [33]. It measures how often a vertex or edge is present in the set of all shortest paths.

$$C_b(V) = \sum (\sigma st(v)/\sigma st) \, ; s \neq t, \ s \neq v, \ v \neq t$$

σst represents the number of shortest paths from s to t which may or may not pass through the node v and $\sigma st(V)$ represents the number of shortest paths from s to t that pass through v.

3 Results and Discussion

3.1 Characterization and Sequence Analysis of MATE Proteins

Based on the various global features such as amino acid residue length, the number of transmembrane helices, a total of 16 MATE protein sequences were selected after sorting and filtering of data for *P. aeruginosa* and *S. aureus*.

3.2 Physicochemical Characterization

The physicochemical characterization of MATE proteins has provided the overview about the various characteristics and behavioral nature of MATE proteins of *P. aeruginosa* and *S. aureus* (Table 1). MATE proteins consists of 400 to 700 amino acids exhibiting 40% amino acid similarity in their protein sequences [34] with molecular mass ~ 54 kDa [35]. The selected dataset for MATE protein sequences were found to be comprised of 440 to 550 amino acid residues and of ~ 50,000 Da molecular weight and the amino acid residues were chiefly of hydrophobic nature. The instability index value in the range of less than 40 corresponded to the stability of these proteins as the proteins having value above 40 may be unstable. The higher value of aliphatic index corresponded to the higher thermal stability of MATE proteins. The isoelectric point for the MATE proteins was observed in the range of 8 to 10. Moreover, positive value of GRAVY confirmed the hydrophobic character of most amino acids in MATE protein sequences.

3.3 Primary Structure Analysis

The primary structure analysis of MATE proteins has revealed the amino acid residues Ala, Gly and Ile to be most occurring with the varied composition of these amino acids varied in each protein sequence (Fig. 1). However, the amino acid residues such as Lys, Cys and Trp were found to be lowest in the MATE protein sequences. The hydrophobic and hydrophilic amino acids analysis has shown ~ 60–70% amino acids to be of hydrophobic nature in each MATE protein sequence (Fig. 2). MATE transporters typically constitutes 12 transmembrane helices [36]. The predicted

Table 1 Physicochemical features of MATE protein sequences of *P. aeruginosa* and *S. aureus*

Accession	AA number	MW (Da)	pI	II	AI	GRAVY
>WP_061201177.1	477	50,858.25	9.75	29.66	114.23	0.61
>WP_078465331.1	466	49,764.01	9.66	27.55	115.04	0.64
>WP_034038516.1	453	48,652.41	10.69	32.13	127.44	0.67
>RAL80500.1	430	45,933.25	10.38	32.55	129.72	0.72
>PPB15020.1	445	47,771.53	10.07	33.13	129.51	0.75
>WP_050480324.1	467	50,641.22	9.46	42.69	118.82	0.46
>SQG59012.1	450	47,997.46	9.62	33	121.8	0.82
>WP_058160294.1	488	50,745.09	8.94	38.06	114.77	0.78
>WP_006191797.1	451	48,816.01	6.96	26.78	133.44	1.10
>SCS97867.1	423	45,556.08	7.68	24.57	132.58	1.06
>WP_075108090.1	451	48,921.8	8.67	22.25	133.81	1.05
>SUL88672.1	455	49,336.92	9.88	20.04	126.73	0.88
>WP_070046250.1	444	48,753.28	9.67	26.92	128.63	0.96
>WP_031905678.1	447	48,635.7	9.75	27.83	137.07	0.89
>BBD12010.1	489	54,582.54	9.87	42.11	131.17	0.88
>CXT95697.1	401	42,970.02	8.36	23.09	136.93	1.16

transmembrane helices were found to be ranging from 10 to 12 transmembrane helices which confirmed the topological similarity of these protein sequences to already existing MATE protein sequences. The presence of transmembrane helices and hydrophobic nature of these proteins has confirmed the occurrence of MATE proteins in transmembrane region of cell.

3.4 Secondary Structure Analysis

The analysis using ExPASy tool has shown the occurrence of alpha helices, beta turns, extended strands and random coils in secondary structure of these proteins. The higher number of alpha helices has signified the thermal stability in MATE proteins of both *P. aeruginosa* and *S. aureus* (Fig. 3).

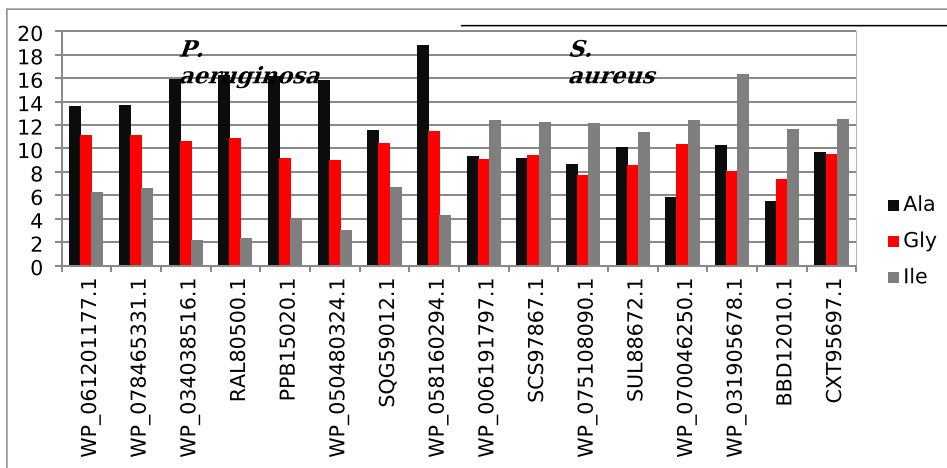**Fig. 1** Major contributing amino acids in MATE protein sequences

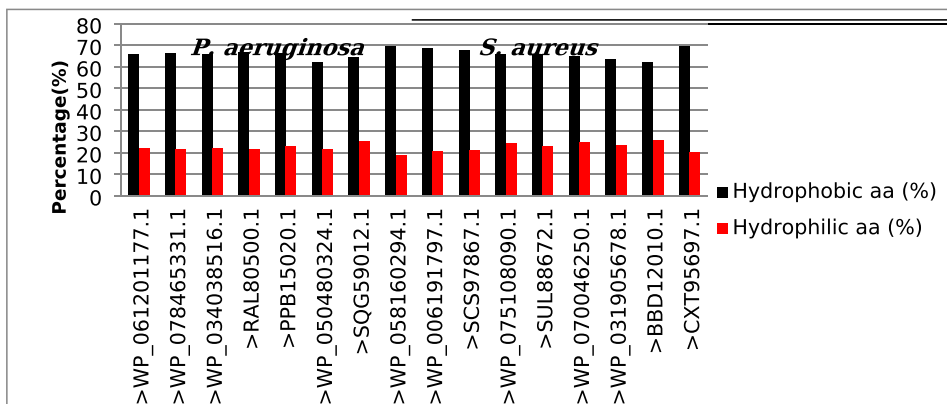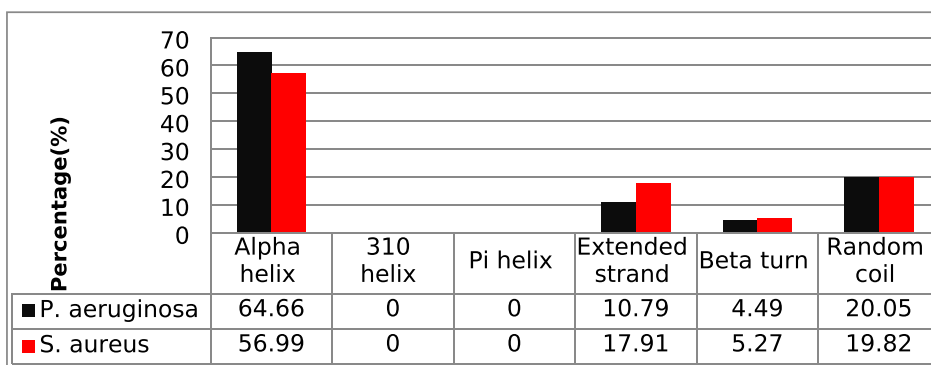**Fig. 2** Hydrophobic and hydrophilic composition in MATE protein sequences

Fig. 3 Secondary structure composition of MATE protein sequences

	Alpha helix	310 helix	Pi helix	Extended strand	Beta turn	Random coil
P. aeruginosa	64.66	0	0	10.79	4.49	20.05
S. aureus	56.99	0	0	17.91	5.27	19.82

3.5 Tertiary Structure Analysis

The predicted structures for bacterial homologs of MATE proteins have shown resemblance to the existing structures of MATE transporter proteins. The 12 transmembrane helices are arranged symmetrically with 6TMHs forming a bundle each named as N- and C-bundle [37].

The observed protein sequences have shown two lobes, i.e. N-lobe and C-lobe in their topology and pseudo two fold symmetry. The PDB files for the energy minimized tertiary structures have been analyzed for their structural coordinates. The homology modeled structures followed by energy minimization were validated by the Ramachandran plot. The Ramachandran plot analysis for *P. aeruginosa* (> WP_078465331.1)has shown that 80.4% residues were observed in most favored regions, 15.2% residues
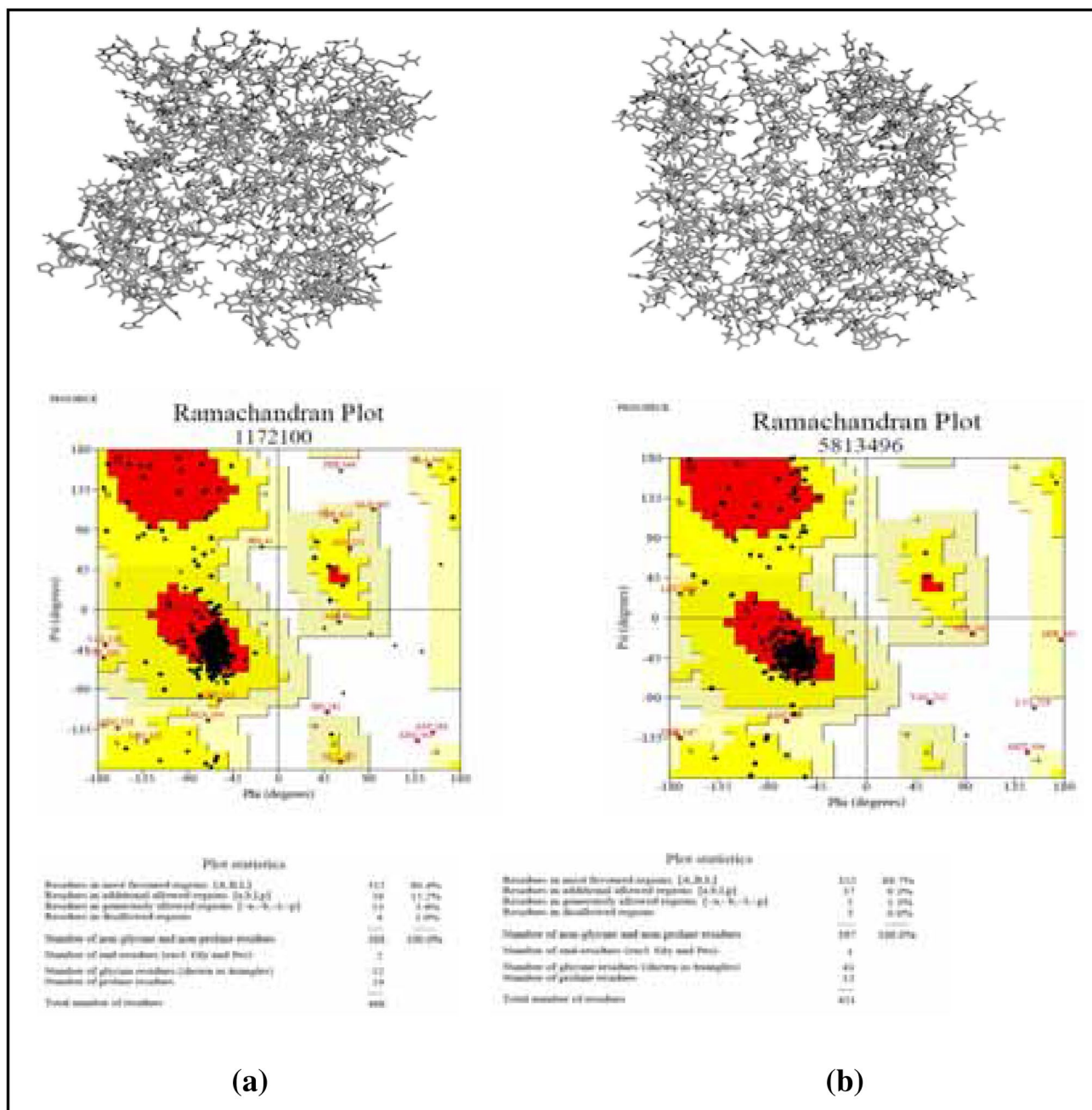

(a)

(b)

Fig. 4 Homology modeled structures, Ramachandran plot for MATE protein sequences **a** *P. aeruginosa* (> WP_078465331.1) and **b** *S. aureus*(> WP_006191797.1)

in additional allowed regions, 3.4% residues in generously allowed regions and 1% residues were observed in disallowed regions. However, the analysis results for *S. aureus* (> WP_006191797.1) has shown 88.7%, 9.3% and 1.3% residues in most favored regions, additional allowed regions, and generously allowed regions respectively. The residues in disallowed regions were observed to be 0.8% (Fig. 4).

3.6 Molecular Dynamic Simulations and Optimized Structure

The tertiary structures for *P. aeruginosa* (> WP_078465331.1) and *S. aureus*(> WP_006191797.1) MATE protein sequences were exposed to molecular dynamic simulations using GROMACS 5.0 using 53a6 force field. These optimized structures showed resemblance to already existing protein structures to their respective protein families in membrane and shown in Figs. 5, 6.In these figures; (a) part shows the side view of protein in membrane which shows their topological resemblance to their respective protein families and (b) part shows the top view which shows the cavity formation for substrate extrusion purposes in the respective protein.

The uniqueness of homology modeled structures was validated using RCSB-PDB pairwise structure alignment tool (https://www.rcsb.org/alignment).Structure alignment is the technique of comparing the molecular structures of two or more biomolecules to determine three-dimensional

form equivalences. The homology-modeled structures of *P. aeruginosa* (> WP_078465331.1) and *S. aureus* (> WP_006191797.1) were compared to the crystal structure of the *E. coli* cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter (PDB id: 3MKT) using rigid body alignment and superimposed structures (Fig. 7) were viewed using BIOVIA-Discovery studio 4.0 [38]. The relative orientations and locations of atoms inside each structure stay fixed during the alignment process in a rigid body alignment. Only the general forms of the structures are aligned in the resultant superposition. Rigid body alignments are ideal for identifying structural equivalences between proteins that are closely related evolutionarily and so have comparable forms.

The parameters root mean square deviation (RMSD), template modeling score (TM- Score), Score, sequence identity percentage (SI%), sequence similarity percentage (SS%) and length (Table 2) were used to describe the extent of overlap or similarity between the Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter of *E. coli* (PDB id: 3MKT) and with *P. aeruginosa* (> WP_078465331.1), *S. aureus* (> WP_006191797.1). In superposed structures, the RMSD is calculated between aligned pairs of backbone C-alpha atoms. The smaller the RMSD, the better the alignment of the structures. The topological similarity between the template and model structures is measured by the TM-score. The TM-score goes from 0 to 1, with 1 indicating a perfect match and 0 indicating no match between the two structures. Scores of
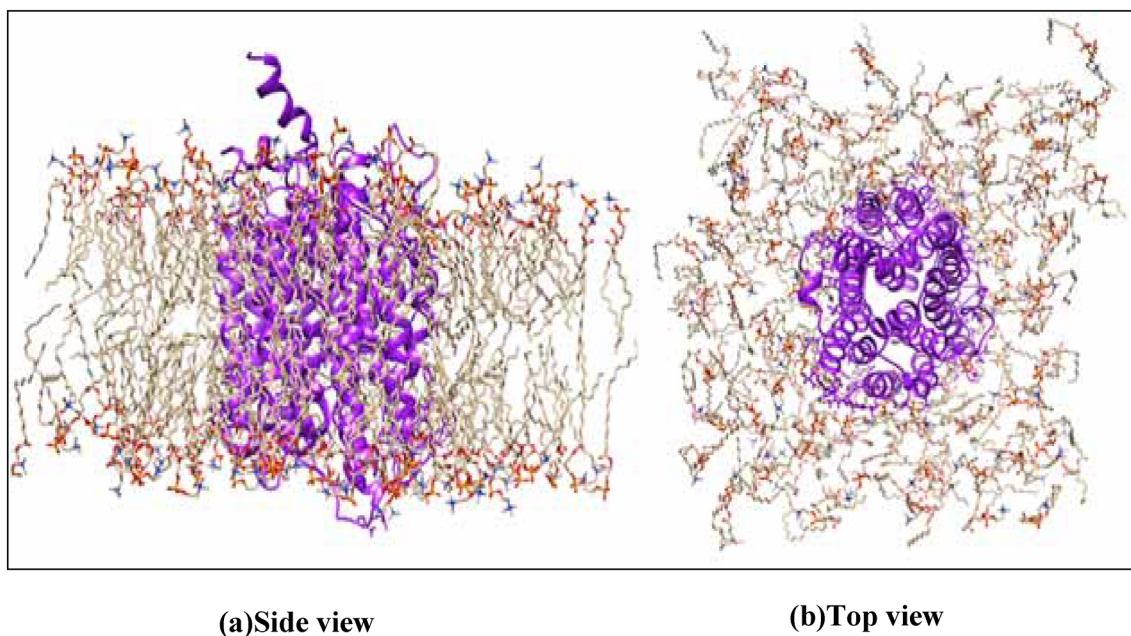

(a)Side view   **(b)Top view**

Fig. 5 Energy minimized and membrane embedded structure of *P. aeruginosa* (> WP_078465331.1) MATE protein; Template PDB ID: 3MKT (Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter of *E. coli*)
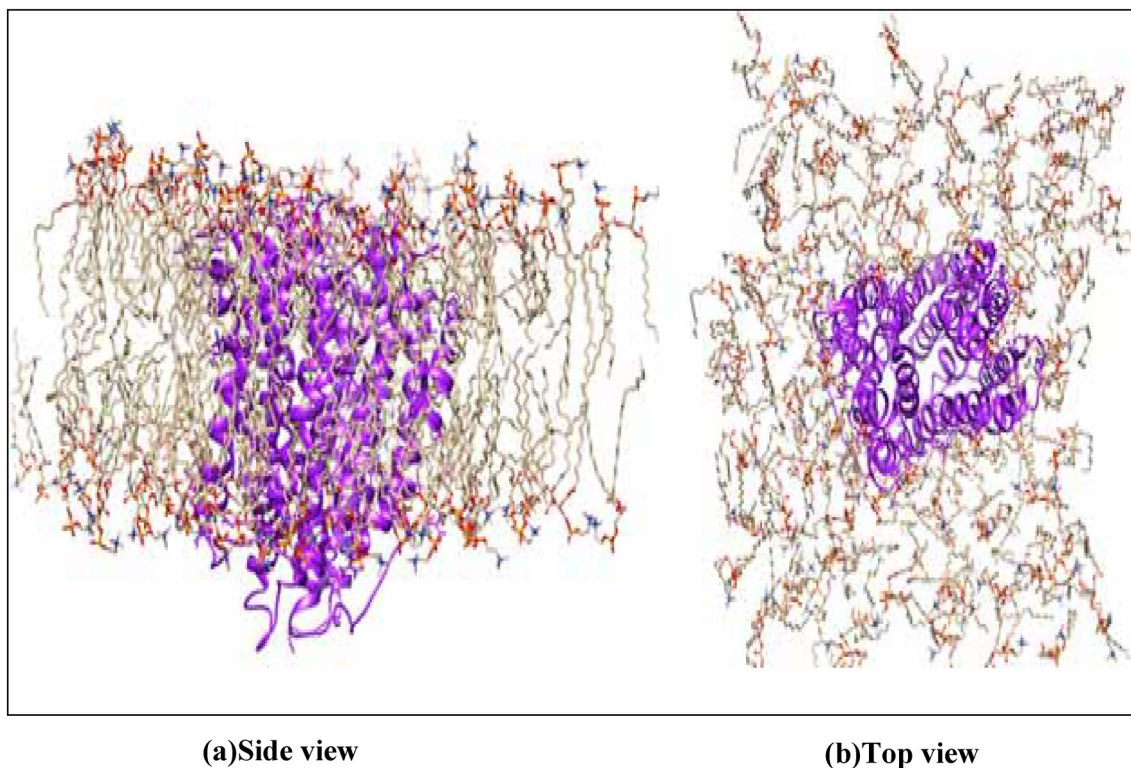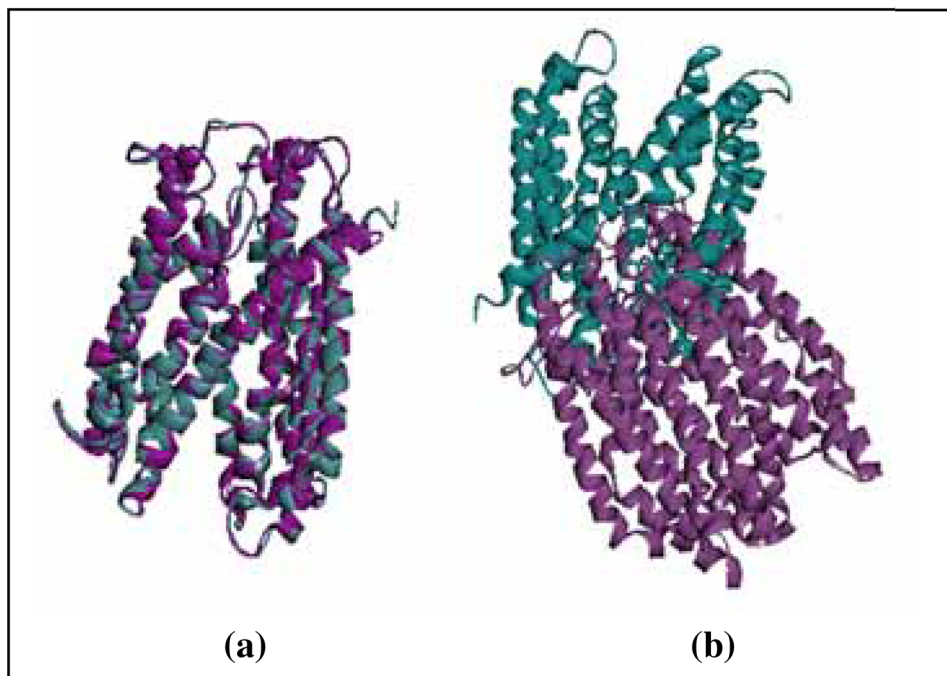
(a)Side view   **(b)Top view**

Fig. 6 Energy minimized structure of *S. aureus* (> WP_006191797.1) MATE protein; Template PDB ID: 7SP5 (Eukaryotic phosphate transporter protein)

Fig. 7 Superimposed structure of Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter of *E. coli*, PDB id: 3MKT(cyan) with **a** *P. aeruginosa* (> WP_078465331.1), **b** *S. aureus* (> WP_006191797.1)

(a)   **(b)**

0.2 normally suggest that the proteins are unrelated, but scores of > 0.5 usually indicate that the proteins share the same protein fold. Score is a structural similarity metric

that is unique to the alignment method utilized. SI % is the percentage of sequence-identical paired residues in an alignment. SS% is the percentage of paired residues

Table 2 Pairwise structure alignment characterization parameters

Parameters	*P. aeruginosa*	*S. aureus*
RMSD	0.35	3.09
TM-Score	0.97	0.80
Score	1236.52	868.23
SI%	42	16
SS%	62	35
SS%	455	418

in the alignment that are sequencely similar. Length is the number of structurally equivalent residue pairs in the alignment.

The parameter values revealed a strong three-dimensional structural similarity between the *E. coli* Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter, PDB id: 3MKT, and *P. aeruginosa* (> WP 078,465,331.1). When comparing two structures, the most usually reported statistic is RMSD, however it is susceptible to local structural variation. Even if the remainder of the structure is perfectly aligned, the RMSD value is high if a few residues in a loop are poorly aligned. The alignment of the *S. aureus* (> WP 006,191,797.1) with the *E.coli* Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter, PDB id: 3MKT, reveals that both proteins are structurally comparable with RMSD 3.09, TM-Score 0.8 despite the low sequence identity (16%).

3.7 RIN Analysis and Actor Residues Identification

The observed residue interaction networks has revealed the scale-free degree distribution which correspond MATE proteins to be a scale-free ideal network. The residue interaction network and various centrality parameters for *P. aeruginosa* (> WP_078465331.1) and *S. aureus* (> WP_006191797.1)were computed (Fig. 8). The majorly contributing nodes has been presented. The results for centrality values (Table 3) have shown the maximum betweenness centrality for LEU282, Ile244, MET27, ALA293 and GLN21 in *P. aeruginosa* and GLN395, LEU253, MET25, ALA247 and ILE273 in *S. aureus*, which suggested that these residues have control to pass over the information to other residues and their removal may disrupt the connection between the residues in the protein structure. The nodes GLY183, PHE182, ALA351, ASP42, LEU115 and LEU409, ASN219, ALA408, PHE190, and GLY411 were found to have the maximum degree to which network tends to cluster together. The residues LEU65, ILE244, ALA368, GLY191 and VAL296 in *P. aeruginosa* and ALA302, LEU293, ILE80, VAL74, and GLY75 were found to have maximum degree centrality values which measure the maximum number of nodes connected to these residues (nodes).

These nodes carrying the maximum degree centrality values may be responsible for steric bulk which corresponds to the steric hindrance in the protein structure. The RIN analysis for MATE protein sequences have shown the hydrophobic amino acids to have maximum centrality values, corresponding the hydrophobic amino acids as the major acting amino acids within networks which may be because of the predominance of hydrophobic amino acids throughout the networks for MATE protein sequences.

Protein network representations offer a systems approach to topological study of complicated three-dimensional structures and may contribute to a better understanding on structure–function links. There are a variety of online and stand-alone tools for constructing RINs from protein structures, such as NAPS [39], WebPSN [40], RINalyzer [41–43], NeEMO [44], Ring 2.0 [45], RIP-MD [46], PSN Ensemble [47], pyInteraph, [48], INTAA [49] and gRINN [50, 51]. These tools/software have relatively sophisticated user requirements, and some of them deliver protein network information in a more complicated manner. PSN-Ensemble, pyInteraph and INTAA, online web servers which require MATLAB installation and process results from Interaction Energy matrices only. gRINN is a standalone tool which requires topology/ trajectory files and also provide the results in the form of energy networks. RINalyzer provides the simultaneous, interactive 2D visualization and exploration of a RIN together with the corresponding molecular 3D structure having compatibility with Cytoscape only and as well as require JAVA plugin. However, NAPS, RIP-MD, WebPSN and RING 2.0 are web servers providing the results based on user specified residues but not available in standalone versions. The present approach may be used as a standalone method for RIN based on user specified residues. In this specific case, the centrality parameters were calculated based distance topology from the amino acid residue co-ordinates of PDB files observed after MD simulation. The Euclidian distances between the C_α atoms of amino acid residues were calculated followed by a distance threshold. The residue interaction network visualization and centrality calculations can be done directly by any residue interaction viewing programme, in the present specific case Cytoscape was used. The technique is broad in scope, and it can inspire and suggest strategies for developing standalone RIN analyzers for user specific residues.

The major findings of this study has provided the generalization of key features and computational physicochemical characteristics for MATE efflux proteins in *P. aeruginosa* and *S. aureus*. The primary, secondary and tertiary structure analysis provided a better understanding and comparative analysis of features, aa residues and protein folds for both the bacterial species. RIN analysis showed the major actor aa residues which are necessary for the specific folds with the protein structures based on their centrality parameters. The
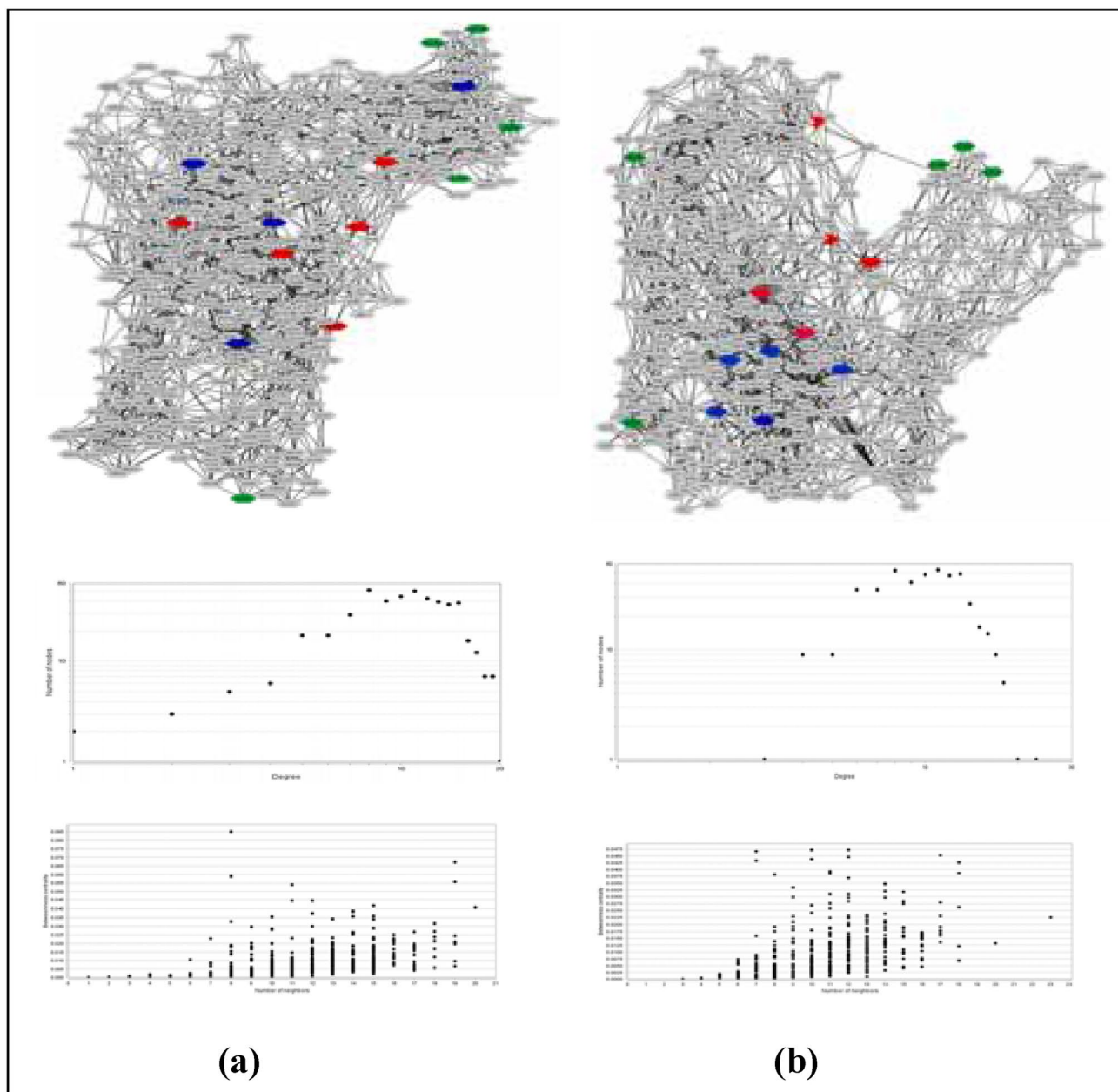
Fig. 8 RIN, Degree centrality graph and betweenness centrality graph (degree centrality: blue, betweenness centrality: red, clustering coefficient: green) **a** *P. aeruginosa* (> WP_078465331.1) and **b** *S. aureus* (> WP_006191797.1)

algorithm design provided a standalone approach for RIN analysis for the user specified residues. The detailed findings showed that after sorting and filtering data for *P. aeruginosa* and *S. aureus*, a total of 16 MATE protein sequences were chosen. MATE protein physicochemical characterization has provided an overview of the diverse properties and behavioural nature of MATE proteins. The chosen dataset for MATE protein sequences was found to be made up of 440 to 550 amino acid residues with a molecular weight of 50,000 Da, and the amino acid residues were mostly hydrophobic. Proteins with an instability index value less than 40 were stable, but proteins with a value more than 40 were potentially unstable. The higher the value of the aliphatic index, the greater the thermal stability of MATE proteins. The MATE proteins isoelectric point was found to be between 8 and 10. GRAVY result validated the hydrophobic nature of most amino acids in MATE protein sequences. The amino acid residues Ala, Gly, and Ile were shown to be the

Table 3 Centrality parameters for RIN in *P. aeruginosa* (> WP_078465331.1) and *S. aureus*(> WP_006191797.1)

	P. aeruginosa (> WP_078465331.1)		*S. aureus* (> WP_006191797.1)	
Degree centrality	LEU65	20	ALA302	20
	ILE244	19	LEU293	18
	ALA368	19	ILE80	18
	GLY191	19	VAL74	17
	VAL296	17	GLY75	17
Betweenness centrality	LEU282	0.084	GLN395	0.047
	ILE244	0.067	LEU253	0.044
	MET27	0.034	MET25	0.042
	ALA293	0.031	ALA247	0.031
	GLN21	0.03	ILE273	0.029
Clustering coefficient	GLY183	0.70	LEU409	0.70
	PHE182	0.70	ASN219	0.70
	ALA351	0.60	ALA408	0.60
	ASP42	0.57	PHE190	0.57
	LEU115	0.50	GLY411	0.52

most often present in MATE proteins, with the proportion of these amino acids varying in each protein sequence. The amino acid residues Lys, Cys, and Trp were discovered to be the least abundant in the MATE protein sequences. The hydrophobic and hydrophilic amino acid study revealed that 60–70% of the amino acids in each MATE protein sequence are hydrophobic. The predicted transmembrane helices ranged from 10 to 12 transmembrane helices, confirming the topological closeness of these protein sequences to previously identified MATE protein sequences. The presence of transmembrane helices and the hydrophobic nature of these proteins confirms the existence of MATE proteins in the cell's transmembrane region. ExPASy analysis revealed the presence of alpha helices, beta twists, extended strands, and random coils in the secondary structure of these proteins. The greater the number of alpha helices, the greater the heat stability in *P. aeruginosa* and *S. aureus* MATE proteins. The anticipated structures of bacterial homologs of MATE proteins resemble the known structures of MATE transporter proteins. The reported protein sequences feature two lobes, namely N-lobe and C-lobe, and pseudo two fold symmetry. The structural coordinates of the energy-minimized tertiary structures were determined using PDB data. The Ramachandran plot validation for *P. aeruginosa* (> WP 078,465,331.1) revealed that 80.4% of residues were found in the most preferred areas, 15.2% in further permitted regions, 3.4% in generously allowed regions, and 1% in banned regions. The analysis results for *S. aureus* (> WP_006191797.1) has shown 88.7%, 9.3% and 1.3% residues in most favored regions, additional allowed regions,

and generously allowed regions respectively. The residues in disallowed regions were observed to be 0.8% (Fig. 4). The optimized structures for *P. aeruginosa* (> WP_078465331.1) and *S. aureus* (> WP_006191797.1) showed resemblance to already existing protein structures to their respective protein families in membrane and the cavity formation for substrate extrusion purposes in the respective protein. The parameter values revealed a strong three-dimensional structural similarity between the *E. coli* Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter, PDB id: 3MKT, and *P. aeruginosa* (> WP 078,465,331.1). The alignment of the *S. aureus* (> WP 006,191,797.1) with the *E.coli* Cation-bound Multidrug and Toxin Compound Extrusion (MATE) transporter, PDB id: 3MKT, reveals that both proteins are structurally comparable with RMSD 3.09, TM- Score 0.8 despite the low sequence identity (16%). The observed residue interaction networks revealed the scale-free degree distribution, indicating that MATE proteins are a scale-free perfect network. The results for centrality values showed that LEU282, Ile244, MET27, ALA293 and GLN21 have the highest betweenness centrality in *P. aeruginosa* and GLN395, LEU253, MET25, ALA247, and ILE273 in *S. aureus*, implying that these residues have control over passing information to other residues and that their removal may disrupt the connection between the residues in the protein structure. The nodes GLY183, PHE182, ALA351, ASP42, LEU115 and LEU409, ASN219, ALA408, PHE190, and GLY411 were found to have the highest degree of network clustering. The *P. aeruginosa* residues LEU65, ILE244, ALA368, GLY191 and VAL296 and ALA302, LEU293, ILE80, VAL74, and GLY75 were discovered to have the highest degree centrality values, which quantify the number of nodes associated to these residues (nodes). These nodes with the highest degree centrality values might be in charge of steric bulk, which translates to steric hindrance in the protein structure. The RIN analysis for MATE protein sequences revealed that hydrophobic amino acids had the highest centrality values, corresponding to hydrophobic amino acids as the major acting amino acids within networks, which could be attributed to the predominance of hydrophobic amino acids throughout the networks for MATE protein sequences. NAPS, WebPSN, RINalyzer, NeEMO, Ring 2.0, RIP-MD, PSN Ensemble, pyInteraph, INTAA, and gRINN are some online and standalone tools for building RINs from protein structures. These tools/software have relatively advanced user requirements, and some give protein network information in a more intricate way. Some of them, however, are web servers that provide results based on user-specified residues but are not available in standalone versions. The current approach may be utilised as a stand-alone RIN method based on user-specified residues.

4 Conclusion

MATE antiporters in bacteria reduce the effective cellular concentration of a drug inside the cell(s) rendering them resistant to the toxic effect(s) of an antibiotic. A sound understanding of the structure, molecular mechanism and the residues involved in the functioning of these efflux transporter proteins is lacking in the literature. In this study, a comparative structural and RIN analysis of MATE protein sequences of *P. aeruginosa* and *S. aureus* has been observed. The physicochemical properties analysis have confirmed the major contributing global features of MATE protein sequences. The analysis of various structural levels of MATE protein sequences have shown the hydrophobic aa to be majorly contributing throughout MATE protein sequences and Ala, Gly and Ile to be majorly contributing aa. The secondary structure analysis have shown the predominating alpha helices and tertiary structure modeling and validation have shown the resemblance to the already existing MATE protein sequences. An algorithm is developed for RIN generation which has been followed by the analysis of important residues in MATE protein tertiary structure formation. The important residues in the MATE protein's structure formation have been looked into and analyzed on the basis of different centrality parameters through RIN analysis. The structural and RIN analysis of MATE protein sequences have shown the important aa in MATE proteins which might be helpful in new drug target site identification. The algorithm designed for RIN generation uses the bond formation within the C-α atoms in MATE protein structure and may be used for bond formation analysis for other atoms in MATE protein structure. Despite of online web servers and tools, this algorithm provides a standalone approach for RIN analysis. This algorithm may be applied for RIN generation in efflux transporter proteins and for other studies including protein tertiary structure analysis as well. The actor residues observed from RIN provided the major structural insight towards the MATE protein structure analysis.

Declarations

Conflict of interest Authors declare that they have no conflict of interest with the parent institute or among themselves.

References

1. Garcia IR, de Oliveira Garcia FA, Pereira PS, Coutinho HD, Siyadatpanah A, Norouzi R, Wilairatana P, de Lourdes PM, Nissapatorn V, Tintino SR, Rodrigues FF (2022) Microbial resistance: the role of efflux pump superfamilies and their respective substrates. Life Sci 295:120391
2. Kumar S, Floyd JT, He G, Varela MF (2013) Bacterial antimicrobial efflux pumps of the MFS and MATE transporter families: a review. Dev Antimicrob Agents Chemother. 7:1–21
3. Claxton DP, Jagessar KL, Mchaourab HS (2021) Principles of alternating access in multidrug and toxin extrusion (MATE) transporters. J Mol Biol 433(16):166959
4. Omote H, Hiasa M, Matsumoto T, Otsuka M, Moriyama Y (2006) The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. Trends Pharmacol Sci 27:587–593
5. Mishra MN, Daniels L (2013) Characterization of the MSMEG_2631 Gene Encoding a MATE Family Protein in Mycobacterium smegmatis and Exploration of Its Polyspecific Nature Using Biolog Phenotype MicroArray. J Bacteriol 195:1610–1621
6. Ficici E, Zhou W, Castellano S, Faraldo-Gomez JD (2018) Broadly conserved Na+-binding site in the N-lobe of prokaryotic multidrug MATE transporters. Proc Natl Acad Sci 115:E6172–E6181
7. Kusakizako T, Claxton DP, Tanaka Y, Maturana AD, Kuroda T, Ishitani R, Mchaourab HS, Nureki O (2019) Structural basis of H⁺-dependent conformational change in a bacterial MATE transporter. Structure 27(2):293–301
8. Huang L, Wu C, Gao H, Xu C, Dai M, Huang L, Hao H, Wang X, Cheng G (2022) Bacterial multidrug efflux pumps at the frontline of antimicrobial resistance: an overview. Antibiotics 11(4):520
9. Dijun Du, van Veen HW, Murakami S, Pos KM, Luisi BF (2015) Structure, mechanism and cooperation of bacterial multidrug transporters. Curr Opin Struct Biol 33:76–91
10. Miyauchi H, Moriyama S, Kusakizako T, Kumazaki K, Nakane T, Yamashita K, Hirata K, Dohmae N, Nishizawa T, Ito K, Miyaji T, Moriyama Y, Ishitani R, Nureki O (2017) Structural basis for xenobiotic extrusion by eukaryotic MATE transporter. Nat Commun 8:1633
11. Becker ML, Visser LE, Van Schaik RH, Hofman A, Uitterlinden AG, Stricker BH (2009) Genetic variation in the multidrug and toxin extrusion transporter protein influences the glucose-lowering effect of metformin in patients with diabetes: a preliminary study. Diabetes 58:745–749
12. Krah A, Huber RG, Zachariae U, Bond PJ (2020) On the ion coupling mechanism of the MATE transporter ClbM. Biochimica et Biophysica Acta (BBA)-Biomembranes. 1862(2):183137
13. Kusakizako T, Miyauchi H, Ishitani R, Nureki O (2020) Structural biology of the multidrug and toxic compound extrusion superfamily transporters. Biochimica et biophysica Acta (BBA)-biomembranes. 1862(12):183154
14. Bienkowska J (2014) Computational characterization of proteins. Expert Rev Proteomics 2:129–138
15. Geerlof A, Brown J, Coutard B, Egloff MP, Enguita FJ, Fogg MJ, Gilbert RJ, Groves MR, Haouz A, Nettleship JE, Nordlund P, Owens RJ, Ruff M, Sainsbury S, Svergun DI, Wilmanns M (2006) The impact of protein characterization in structural proteomics. Acta Crystallogr D BiolCrystallogr 62:1125–1136
16. Gentiluomo L, Svilenov HL, Augustijn D, El Bialy I, Greco ML, Kulakova A, Indrakumar S, Mahapatra S, Morales MM, Pohl C, Roche A (2019) Advancing therapeutic protein discovery and

development through comprehensive computational and biophysical characterization. Mol Pharm 17(2):426–440

17. Nadezhda TD et al (2011) Analyzing and visualizing residue networks of protein structures. Trends Biochem Sci 36(4):179–182

18. Braberg H, Echeverria I, Kaake RM, Sali A, Krogan NJ (2022). From systems to structure- using genetic data to model protein structures. Nature Reviews Genetics. 1–3.

19. Jiao X, Ranganathan S (2017) Prediction interface residue based on the features of residue interaction network. J Theor biol 432:49–54

20. National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information.

21. Mitchell C (1993) MultAlin–multiple sequence alignment. Bioinformatics 9:614

22. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server

23. Walker JM (ed) (2005) The proteomics protocols handbook. Humana Press, Totowa, pp 571–607

24. Krogh A, Larsson B, Von HG, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

25. Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17:849–850

26. Combet C, Blanchet C, Geourjon C, Deleage G (2000) Network protein sequence analysis. Trends Biochem Sci 25:147–150

27. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A (2006) Comparative protein structure modeling using modeller. Curr Protoc Bioinform 15:561–5630

28. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723

29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. J App Cryst 26:283–291

30. Shannon P, Markiel A, Ozier O, Baliga SN, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

31. Bolland JM (1988) Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks. Social Netw 10(3):233–253

32. Beauchamp MA (1965) An improved index of centrality. Behav Sci 10(2):161–163

33. Hage P, Harary F (1995) Eccentricity and centrality in networks. Social Networks 17:57–63

34. Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41

35. Handzlik J, Matys A, Kieć-Kononowicz K (2013) Recent advances in multi-drug resistance (MDR) efflux pump inhibitors of Gram-positive bacteria *S. aureus*. Antibiotics 2:28–45

36. Chen L, Liu Y, Liu H, Kang L, Geng J, Gai Y, Ding Y, Sun H, Li Y (2015) Identification and expression analysis of MATE genes involved in flavonoid transport in blueberry plants. PLoS ONE 10(3):e0118578

37. Radchenko M, Symersky J, Nie R, Lu M (2015) Structural basis for the blockade of MATE multidrug efflux pumps. Nat Commun 6:7995

38. Eisinger ML, Nie L, Dorrbaum AR, Langer JD, Michel H (2018) The xenobiotic extrusion mechanism of the MATE transporter NorM_PS from *Pseudomonas stutzeri*. J Mol Biol 430:1311–1323

39. Dassault Systèmes (2016) Discovery studio 4.0. BIOVIA. Dassault Systèmes, San Diego

40. Chakrabarty B, Parekh N (2016) NAPS: network analysis of protein structures. Nucleic Acids Res 44:W375–W382

41. Seeber M, Felline A, Raimondi F, Mariani S, Fanelli F (2015) WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. Bioinformatics 31:779–781

42. Doncheva NT, Assenov Y, Domingues FS, Albrecht M (2012) Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc 7:670–685

43. Doncheva NT, Klein K, Domingues FS, Albrecht M (2011) Analyzing and visualizing residue networks of protein structures. Trends Biochem Sci 36:179–182

44. Doncheva NT, Klein K, Morris JH, Wybrow M, Domingues FS, Albrecht M (2014) Integrative visual analysis of protein sequence mutations. BMC Proc 8:S2

45. Giollo M, Martin JMA, Walsh J, Ferrari C, Tosatto CES (2014) NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation. BMC Genomics 15(Suppl 4):S7

46. Piovesan D, Minervini G, Tosatto CES (2016) The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res 44:W367–W374 (**Web Server issue**)

47. Riquelme SC, Garate JA, Acle TP, Martin JMA (2018) RIP-MD: a tool to study residue interaction networks in protein molecular dynamics. PEER J 6:e5998

48. Bhattacharyya M, Bhat CR, Vishveshwara S (2013) An automated approach to network features of protein structure ensembles. Protein Sci 22:1399–1416

49. Tiberti M, Invernizzi G, Lambrughi M, Inbar Y, Schreiber G, Papaleo E (2014) PyInteraph: a framework for the analysis of interaction networks in structural ensembles of proteins. J Chem Inf Model 54:1537–1551

50. Galgonek J, Vymetal J, Jakubec D, Vondrasek J (2017) Amino Acid Interaction (INTAA) web server. Nucleic Acids Res 45:W388–W392

51. Sercinoglu O, Ozbek P (2018) gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations. Nucleic Acids Res 46:W554–W562 (**Web Server issue**)