Molecular Docking and Simulation Approach to Study the Inhibitory Effect of Rhamnolipid on Biofilm Producing Proteins in *E. coli* K12

Rohit Pritam Das, Banishree Sahoo, Manoranjan Arakha, Arun Kumar Pradhan

Centre for Biotechnology, Siksha O Anusandhan (Deemed to be University), Bhubaneswar, India

Abstract- Microbes have a proclivity for binding to cell surfaces and forming biofilms. The act of creating biofilms is the microbe's social activity while they are under stress. In humans, this form of cell aggregation leads to biofilm, which often leads to an infection. Despite their ability to form adhesion to the cell surface, biofilm has also drawn attention due to its involvement in chronic disorders. Accumulation of biofilm leads to a serious health concern showing high resistance to antibiotics. In order to address this concern, there is a desperate need to find out natural bioproducts like biosurfactants which could be an alternative to synthetic compounds. In the current study, the inhibitory effect of rhamnolipid against *E. coli* k-12 proteins that are involved in biofilm formation was studied through various computational approaches. In the molecular docking approach, the interaction between rhamnolipid and targeted proteins has been recorded. Rhamnolipid interacts with pgaC with the total highest energy of -8.91 kcal/mol, indicating a tight ligand-protein interaction. Further, to validate the interaction, a 10-ns molecular dynamics simulation was performed for pgaC and with rhamnolipid bound complex. The stability of biosurfactant and biofilm-producing protein was investigated using the RMSD, RMSF, Rg, and SASA plots. As a comparison to only protein, a complex Binding with rhamnolipid shows a stable RMSD value with minimal RMSF and Rg values, which indicates the tight interaction between rhamnolipid and pgaC. This could be a leading novel in silico approach to studying the inhibitory effect of biosurfactants against biofilm formation proteins.
© 2022 Tehran University of Medical Sciences. All rights reserved.
Acta Med Iran 2022;60(12):731-741.

Keywords: Rhamnolipid; Biofilm; In silico; Glycolipid; Interaction; K-12 strain; Groningen machine for chemical simulations (GROMACS) 5.1; Auto dock vina

Introduction

Bacterial biofilms are a major global health concern because of their ability to withstand antibiotics, host defense systems, and other external pressures, contributing to chronic infections that persist (1). Chemical surfactants, with their xenobiotic attributes, play a pivotal part in various fields in this revolutionary scientific era, with the worst long-term ramifications for the biological environment. As a result, biosurfactants (BS), which are derived from a variety of bacteria, are non-toxic, biodegradable, and environmentally sustainable, providing a viable alternative. Biosurfactants are amphiphilic microbial compounds that outperform conventional drugs in terms of

antibiofilm, low toxicity, and improved surface and interfacial efficiency. In bioremediation, hydrocarbon degradation, and bio-emulsions, biosurfactants play a critical role (2). Biosurfactant is also valued for their antibiofilm properties, which help to inhibit pathogenic biofilm formation (3). Among various classes of BS, glycolipids are microbial surface-active molecules made up of a carbohydrate moiety linked to fatty acids that are produced by a wide range of bacteria. BS based on glycolipids with a wide range of abilities to diminish surface and interfacial tension at the surface and interface emerged as the most effective ligand molecule for targeting biofilm-forming proteins (4). The primary objectives in selecting glycolipid biosurfactants are due to their capacity to form pores and destabilize biological

Corresponding Author: A.K. Pradhan
Centre for Biotechnology, Siksha O Anusandhan (Deemed to be University), Bhubaneswar, India
Tel: + 917978171574, Fax: + 9040003290, E-mail address: arunpradhan@soa.ac.in

membranes and also allow them to be used as antibacterial, antifungal, and hemolytic agents in biomedicine. Their antiviral and antitumor properties enable them to be used as therapeutic agents in pharmaceuticals. Glycolipids can also inhibit pathogenic bacteria's bio adhesion, making them useful as anti-adhesive agents and for disrupting biofilm formation (5). The ability of pathogenic bacteria to form biofilms determines their virulence. According to Fux *et al.*, antimicrobial drugs often fail to inhibit the growth of certain pathogenic bacteria due to the prevalence of biofilms (6). With multiple pathogenic strains displaying a wide variety of symptoms and just 20% of the genome shared by all strains, Escherichia coli is one of the most diverse bacterial species (7). The K-12 strain of *E. coli* is used as a model organism because of its adaptability to the laboratory environment. The inhibitory effect of glycolipid-based biosurfactant (Rahmnoloipid) on certain biofilm-producing as well as adhesive proteins present in *E. coli* K-12 is demonstrated here. And their interaction with the ligand compound has also been noticed. To colonize new areas, bacterial biofilms rely on cell transmission or dispersal. Bacteria spread their pathogenicity from an environmental reservoir to a host by dispersal. Thereupon targeting these dispersal proteins may lead to the development of a therapeutic agent (8). bdcA, a crucial regulatory protein with a symbolic function in the dispersal of biofilm from K-12, is thus considered an important therapeutic target (9). ymgA and ycgZ of strain K-12 play an important role in biofilm maturation by stimulating the two-component mechanism, and they are thought to be RcsB/C connector proteins (10). Similarly, overexpression of TabA in strain K-12 has a significant impact on biofilm development by repressing the fimbria gene, which aids biofilm dispersal on different surfaces (11). The same strain's mcbR promotes biofilm formation by inhibiting the secretion of colanic acid, an exopolysaccharide (12). According to Tuckerman DosC controls and influences biofilm formation in an oxygen-dependent manner (13). Eventually, biofilm formation is also influenced by two other proteins, pgaB and pgaD, which belong to the same gene family (14). Overexpression of mqsA decreases bacteria's resistance to oxidative stress through catalytic action, which in turn represses mqsA, which activates rpoS and aids biofilm formation (15). According to Kim JS *et al.*, bdm (biofilm-dependent modulation gene) plays a role in biofilm development (16). In *E. coli*, deletion of bdm causes repression of the flagella-forming gene, resulting in defective motility. In *E. coli*, csgD functions as a central master regulator of biofilm development in two ways. CsgD inhibits the genes fliE and fliEFGH, which are involved in flagellum formation, and activates the adrA gene, which is involved in the synthesis of cyclic di-GMP, a bacterial second messenger, repressing flagellum development and rotation in *E. coli* (17). In *E. coli*, the adhesion protein PGA plays an important role in biofilm formation and structural stability. PgaC is a glycosyltransferase that is required for the production of PGA (18). In this current study, we have preferred to look at how glycolipid biosurfactant, Rhamnolipid interacts with *E. coli* K-12 biofilm-producing proteins through certain in silico approaches.

Materials and Methods

Target identification

Identifying target proteins are one of the crucial steps and to archive, this STITCH database (19) was used. It is useful for finding genes that may be involved in the formation of biofilms in *E. coli* (K12 MG1655). Fourteen proteins from the stitch database were identified and chosen for the intensive study due to their significance in biofilm development. The amino acid sequence information for the proteins was obtained through UniProtKB (20). The Basic Local Alignment Search Tool (BLAST) is a typical method for assessing local sequence similarity to a query sequence (21).

3D modelled structure prediction and optimization

Out of fourteen selected proteins, five were retrieved from ProteinDataBank (PDB) as their structure is already available (22). Structure resolution, domain completeness, sidechain completeness are some major criteria for the selection of protein from the PDB database. A template-based homology protein tertiary structure prediction server, Phyre2 (23), a web portal that allows to model, predict, and analyse proteins (24) was used to build 3d modelled structure. This server analysed the entire domain analysis as well as template information. The mod-Refiner server (25) used a knowledge-based simulation force field with high-resolution energy minimization to improve the physical properties of targeted proteins. With the aid of the PROCHECK server, the 3D structure content of the targeted proteins was analysed using the Ramachandran plot (26). PSI-BLAST was used to generate the evaluation results (Position-Specific Iterated - BLAST).

Ligand retrieval and preparation

Glycolipid biosurfactants derived from Pseudomonas

aeruginosa have been examined through NCBI. The chemical structure of Rhamnolipid was retrieved in SDF format with the assistance of the PubChem (27) database. All adequate data such as molecular weight, atomic structure, and smile structure has been collected, further required to construct the 3D structure of the ligand. Finally, with the assistance of ChemDoodle (28), 3D structure was predicted. Before proceeding with the molecular docking approach, the ligand was optimized once through proper channels with the help of Avogadro tools (29).

Virtual screening and interaction study by molecular docking analysis

All the proteins that were selected as the target have no active site that has previously been mentioned. Thus, blind docking has been performed by AutoDock to establish an interaction between ligands and proteins (30). Various docking tools and software have their own docking matrix score and hydrogen threshold value. The AutoDock tool is the most often recognized, open-source, and widely used tool for predicting binding interactions for MD simulations (31). To generate the ligand atom map, a grid box has been constructed around the protein surface in order to improve the likelihood of predicting the binding site. Before generating the grid box, all the proteins went through proper pre-processing channels. All hydrogens are attached to the nonpolar region of the protein, followed by the addition of Kollman charges and gasteiger charges to the ligand molecule. The ligand Docking parameter was increased by 100 times, significantly increasing the likelihood of ligands docking in the most favorable area. The best-docked complex was selected based on the highest negative docked energy score.

Protein-ligand binding interaction

UCSF CHIMERA analyses the post-docked complex to further understand the structural integrity and tightness (32). This is an interactive visualization tool that carries constructive data regarding molecular structures, along with density maps. Discovery Studio 4.1was used to study the interaction between targeted macromolecules and ligands (33). The ligand interaction parameter was restricted to 3.5 A^0. Within this distance, the H-bonds interaction was calculated. The interactive residues were labeled, and binding distances simultaneously were also measured. To get a good contrast visual clarity image, the background color has been set to white.

Molecular dynamics simulation

The internal and external dynamics of a protein are closely linked to its function. Since no experimental technique exists, the only way to study protein dynamics is to use a computational approach. The molecular dynamics study was performed by GROMACS 5.1.4 throughout 10 nanoseconds with CHARMM36 all-atom force field. The MD simulation has to be performed for both complex and unbound proteins. Structural level integrity and physical movement of atoms have been observed for the best-docked complex. The protein was placed in space, filling the dodecahedron with single-point charge water molecules. To neutralize, sodium ions from the system were introduced. In addition, the steepest descent methodologies were used to reduce the energy of the structures. To equilibrate the systems, the NPT and NVT canonical ensembles were utilized. The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) of the main chain backbone atoms were computed to analyze the conformational changes in the structural level integrity of atoms in docked complex (34). The radius of gyration has been carried out to measure the compactness of protein (35). SASA plot has been generated to develop an understanding of the folding and unfolding nature of the protein. A change in the SASA plot represents the rearrangement of the hydrogen bond.

Results and Discussion

Quality assessment and selection of 3D modeled targeted proteins

14 proteins (bdcA, ymgA, tabA, ycgZ, mcbR, dosC, pgaB, pgaD, mqsA,mcbA,bdm, csgD, pgaC, bsmA) were chosen for further study using the STITCH database because of their presence and positive regulation towards biofilm formation. The amino acid sequences of 14 *E. coli* K-12 target proteins were obtained in FASTA format from the UniProtKB database. The following is a list of 14 selected target proteins, along with their amino acid sequences (supplementary table 1). Out of the 14 chosen proteins, Five proteins, dosC (4ZVF), bdcA (5Z2L), mcbR (4P9F), pgaB (4P7O), and mqsA (3FMY), had their 3D crystal structure available in the protein database. As a result, the tertiary structures of the remaining 9 target proteins: bsmA, tabA, ycgZ, ygmA, mcbA, pgaD, bdm, csgD, and pgaC were predicted. The query coverage as well as the confidence score of the remaining targeted protein, was stated in supplementary table 2. A confidence score from 0 to 100 provides a structural

homology match. Query coverage below 60% was not considered to be further analyzed; therefore, out of 9 modeled proteins, 3 proteins, i.e., tabA, csgD, and pgaC have been considered to be a further screening process. Further structural validation of the remaining targeted proteins was conducted through energy minimization by the Mod Refiner tool. The root means squire deviation (RMSD) was calculated for each of the remaining proteins as compared to their native conformation. RMSD values of all 3 targeted proteins were less than 2 ^{0}A which was a good quality estimation (supplementary table 3) (36). Deviation has to measure in the form of H-bond, backbone topology, and side-chain conformation. Further Quality assessment of the predicted tertiary structure was obtained from PROCHECK through the "Ramachandran plot," wherethe most favored regions for our target proteins have been found (supplementary table 4). Structures that carry a disallowed region of less

than 2 were considered to be good structures. As per table 3, it was clear that the predicted structures matched the desired requirement. Ligands were docked against biofilm target proteins, and a docking phenomenon was carried out. The docked compounds were commissioned based on some parameters, such as maximum occupancy of binding pocket with free energy strength of hydrogen bonding and non-covalent interactions. After passing through these series of validation, a list of 8 proteins was considered to dock with our target ligand, i.e., Rhamnolipid. Out of these 8 proteins, 6 proteins with a low binding energy score are consideredas the best-interacted proteins towards the Rhamnolipid (supplementary table 5). A 100 times flexible running parameter was set to allow the ligand to search for 100 different sites on the protein surface for binding, thus creating a binding pocket residue for the selected 6 proteins mentioned in (figure1).
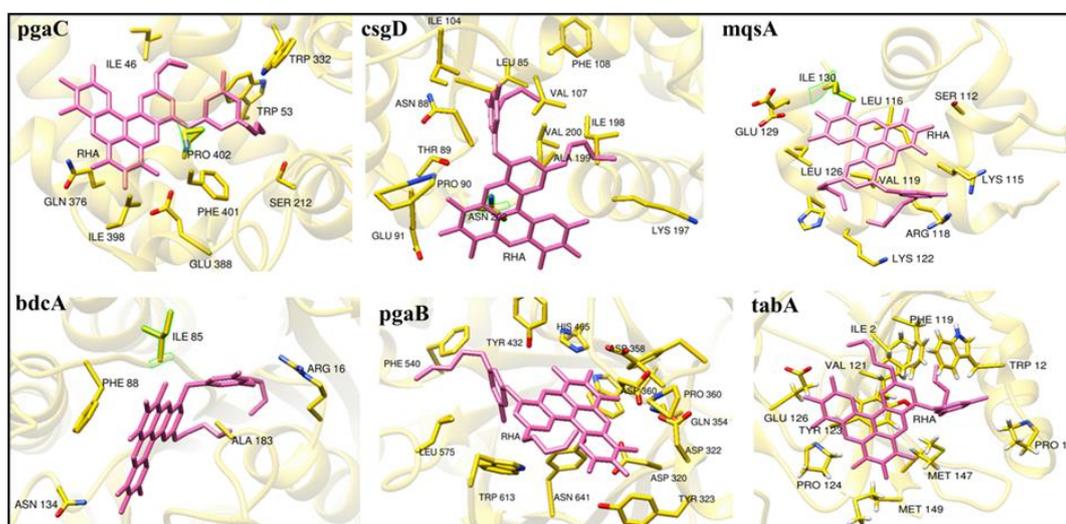

Figure 1. Interacting residues between Rhamnolipid with biofilm-producing proteins of E. coli K12.Protein molecules were highlighted in yellow and Rhamnolipid in pink

In this current process, binding energy was continuously noted for each protein until a favored region was found. The best confirmation with minimum binding energy-carrying high-affinity proteins was selected for further analysis. The same procedure was repeated for each of the targeted proteins. Among all the listed proteins, 6 proteins were found to show a strong binding affinity for rhamnolipid. Rhamnolipid was bound to mqsA, csgD, pgaB, pgaC, tabA, bdcA proteins with a score of less than - 6 kcal/mol binding energy. The region in the target protein where Rhamnolipid binds with high affinity was considered the binding site for the ligand illustrated in figure 1. In this way, the

Rhamnolipid binding sites in all the necessary six target proteins were determined (Figure 1).

Study of binding affinity between Rhamnolipid and biofilm-forming protein in *E. coli* (K12 MG1655) through the docking

After identifying the molecular interactions, Discovery Studio 4.1 was used to investigate the interaction between the selected targeted proteins and Rhamnolipid. The pocket regions of the top three docked proteins, pgaC, csgD, and mqsA, are depicted in the diagram (figure 2). An aromatic ring with highly strong hydrophobic amino acid residues is anticipated to

be involved in hydrophobic contact, as well as a cloud surface depiction of H-bond interaction between the same three targeted proteins and ligands, as indicated in (figure 3). The pgaC protein shows strong hydrophobicinteraction with Rhamnolipid, indicating towards tight interaction between rhamnolipid with the protein. A quick understanding of the H-Bond interaction residues of the target proteins with Rhamnolipid has been elucidated by looking at table 1. Within the pocket binding location, the pgaC protein surrounds the Rhamnolipid, establishing an H-bond with TRP 332, GLU 388, and GLN 376. GLU 129, LYS 115, and ARG 118 residues interact with the Rhamnolipid in close proximity at 3.5A0, according to mqsA. In pgaB, amino acids such as TYR 322, GLN 354, HIS 465, and

TYR 432 form H-bonds at a distance of 3.5^0A. csgD interacts with Rhamnolipid through two closely interacting residues, VAL 107 and GLU 91. bdcA interacts with Rhamnolipid by three near-interacting residues, ARG 16, GLY 86, and ASN 134. Finally, Rhamnolipid shared close residues with tabA and pgaC in near interaction proximity. TRP 332, GLN 376, and GLU 388 are the interaction residues in tabA and pgaC, respectively. Finally, a molecular simulation dynamics approach has been imposed to clarify the stability of molecular docking results. The above findings reveal that pgaC shows the highest binding energy with rhamnolipid has been consideredto be further validated through MD simulation.
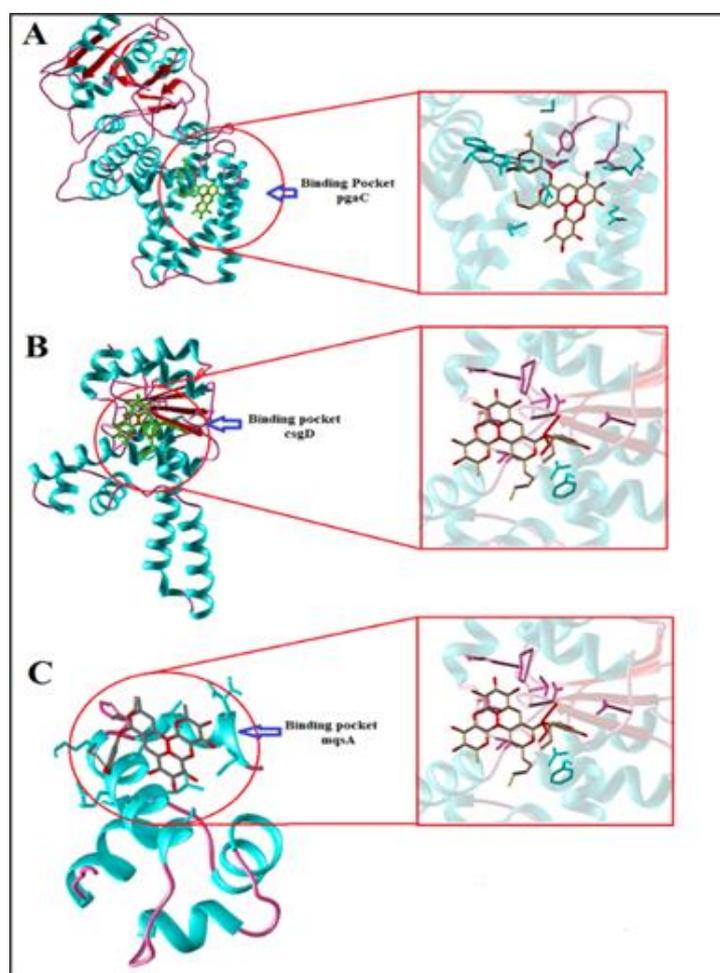

Figure 2. Binding site analyses of shortlisted target proteins with a close view: (A) Binding pocket of pgaC (B) Binding pocket of csgD (C) Binding pocket of mqsA
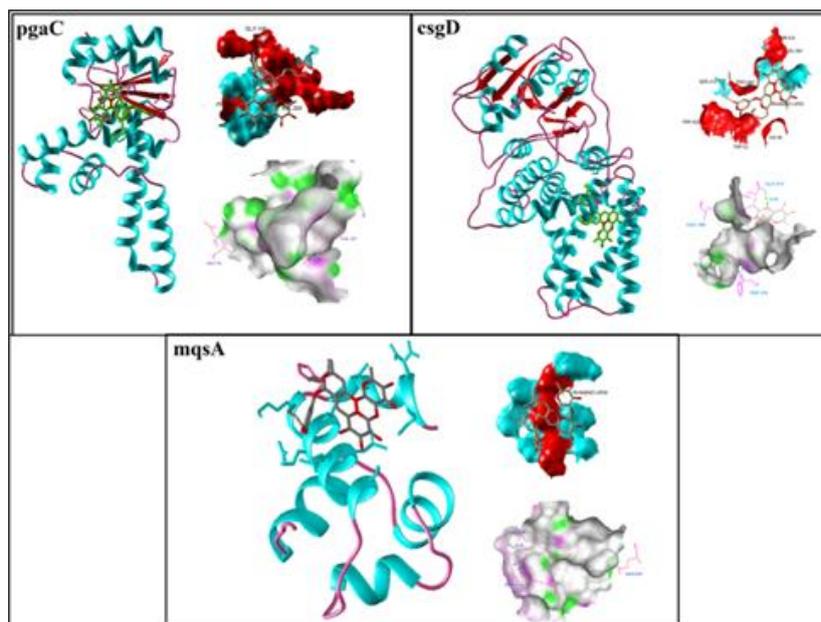
Figure 3. A complete analysis of hydrogen bonds involved in the interaction between rhamnolipid and final shortlisted target proteins. Red surface indicates the hydrophobic residues present in the binding site. Green and pink color represents the hydrogen acceptor and hydrogen donor residues, respectively

Table 1. Showing close molecular interaction between protein atoms and ligands, along with their bond name with a distance within 3.50A

Protein name	Residues Name	Bond name	Distance (A^0)
mqsA	GLU 129	O---O	3.04
	LYS 115	N---O	3.35
	ARG 118	N---O	3.29
pgaB	ASP332	O---O	3.14
	GLN 354	N---O	2.65
	TYR 432	O---O	2.96
csgD	HIS 465	N---O	2.79
	VAL 107	N---O	2.71
	GLU 91	N---O	3.12
bdcA	ARG 16	N---O	3.21
	GLY 86	H---O	2.52
	ASN 134	O---O	3.39
		N---O	3.18
tabA	GLY 125	H---O	2.02
	TRP 332	N---O	3.06
pgaC	GLN 376	N---O	3.29
	GLU 388	O---O	2.81

Molecular dynamics simulations

A 10 ns Molecular Dynamics Simulation was used to determine the stability of both the protein and the docked complex, as well as the influence of the interaction on biofilm inhibition. The RMSD values for both protein and complex escalated dramatically in the first 2 ns of the simulation, but the protein-ligand combination was stabilized at about 5 ns. When comparing the stability of bound and unbound protein, it is obvious that the ligand-bound complex has archived its stability much before the protein has. During the simulation, the rhamnolipid bound complex was stabilized at about 5 ns and maintained its stability and compactness throughout the simulation time period. While the free protein underwent alterations. The trajectory demonstrates that the complex got more compact and stable with time, indicating strong bonding between rhamnolipid and pgaC. The radius of gyration (Rg) plot was used to further investigate the compactness of the complex. The protein Rg value drops sharply up to 7 ns before stabilizing (figure 4B), whereas the ligand-bound complex displays a stable Rg value of about 5 ns. As a result, it may be deduced that the rhamnolipid-bound complex is more compact,

implying a tight link between rhamnolipid and pgaC. SASA (Solvent Accessible Surface Area) is a parameter used to investigate changes in the hydrogen bond network. It also reflects the re-engagement of Amino acid side chains with the surrounding water molecule. The SASA plot provides detailed information about protein folding. As can be seen in (figure 4C), the complex protein degrades rapidly up to 1.5ns before stabilizing completely. The nature of complex proteins has shrunk as a result of this decline. These findings further support the information that rhamnolipid and pgaC are compact in nature and tightly interacted.

RMSF stands for protein residue information. By combining the Residue information from the RMSF plot (figure 4D), it is obvious from the graph that the majority of the fluctuations in ligand-free protein occur near the terminal and loop regions. Due to the existence of loop sections, the area 55-75 in unbound protein shows large variations. Similar huge variances can also be found in the (115-145) and (345-380) areas. Rhamnolipid molecule tends to stabilize these fluctuations by either forming H-bonds between these residues or by implementing some conformational changes.
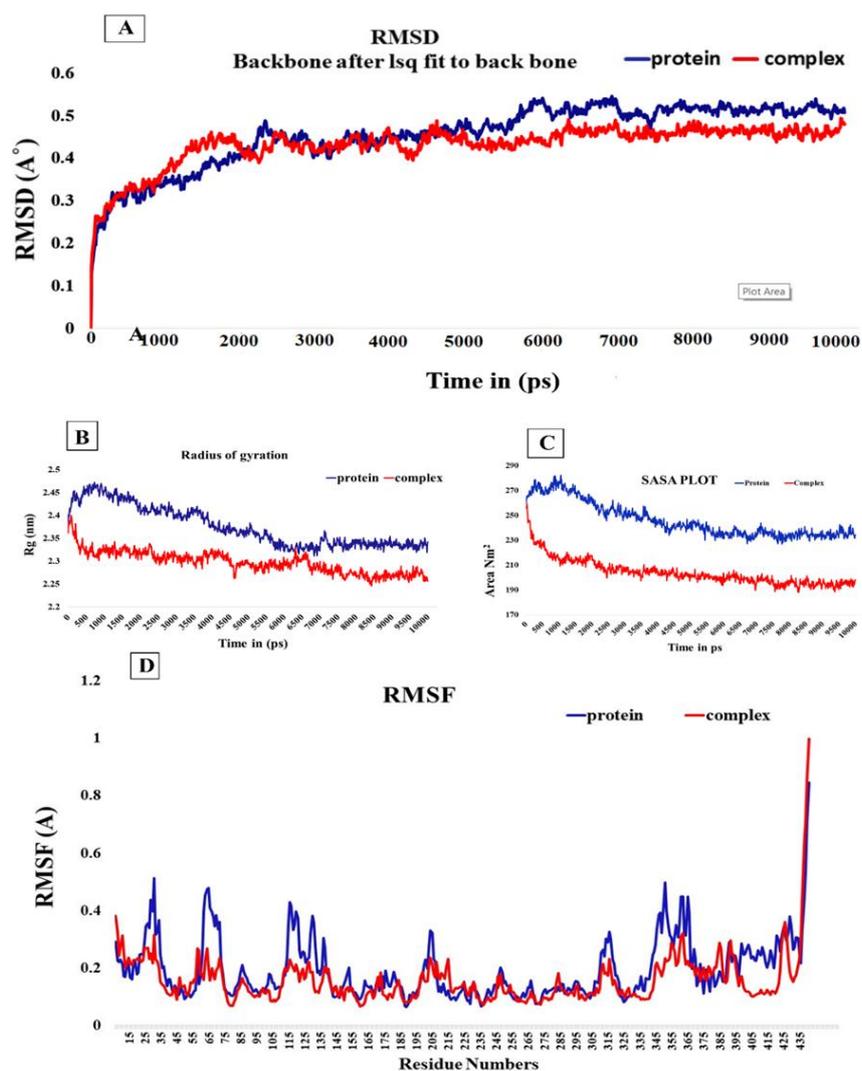

Figure 4. Protein-ligand complex simulation. A, Root mean square deviation plot of backbone, (B) Radius of gyration, (C) SASA plot, (D) Root mean square Fluctuation plot, while complex with Rhamnolipid and lead compound during 10 ns of molecular dynamics simulation are shown

Due to its disease-causing qualities and critical role in increasing intrinsic resistance to antibiotics, biofilm research is becoming a growing subject in microbial

research. To counter antibiotic resistance properties, biosurfactants are treated as an alternate. Hence understanding the structural properties of biofilm-

producing proteins is important for targeting these proteins. Biosurfactants are the natural outcomes of certain bacteria. Thus, understanding its molecular behavior is crucial. Prediction of the tertiary structure of the protein based on NMR and X-ray crystallography is very challenging. As a result, various in silico approaches have been used to anticipate the 3D structure; then, after understanding the ligand binding properties, molecular docking was performed. Docking score revealed the highest ligand affinity towards pgaC out of various other proteins that also have some role in biofilm formation. To understand the structural integrity, MD simulations have been performed. It is clear from the RMSD and RMSF that the complex acquires structural integrity after interacting with the ligand. From the RMSD plot, it is clear that the BS was able to stabilize the structure. RMSF plot also describes the flexibility of the protein as it is clearly understood from the plot that after MD simulation, the structure acquires a compact nature further validated through the radius of the gyration plot. After ligand binding, the structural integrity became more compact and stable. This is indicating towards a tight interaction between ligands and proteins. Hence from these findings, it can be concluded that Rhamnolipid should consider as a lead compound to target biofilm protein.

Acknowledgments

References

1. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: a common cause of persistent infections. Science 1999;284:1318-22.

2. Rahman PK,Gakpe E. Production, characterisation and applications of biosurfactants-review. Biotechnology2008;7:360-70.

3. Pradhan AK, Pradhan N, Mall G, Panda HT, Sukla LB, Panda PK, et al. Application of lipopeptide biosurfactant isolated from a halophile: Bacillus tequilensisch for inhibition of biofilm. Appl BiochemBiotechnol2013;171:1362-75.

4. Mnif I, Ellouz-Chaabouni S, Ghribi D. Glycolipid biosurfactants, main classes, functional properties and related potential applications in environmental biotechnology. J Polym Environ2018;26:2192-206.

5. Inès M, Dhouha G. Glycolipid biosurfactants: Potential related biomedical and biotechnological applications. Carbohydr Res2015;416:59-69.

6. Fux CA, Stoodley P, Hall-Stoodley L, Costerton JW. Bacterial biofilms: a diagnostic and therapeutic challenge. Expert Rev Anti Infect Ther2003;1:667-83.

7. Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced Escherichia coli genomes. MicrobEcol2010;60:708-20.

8. Kaplan JÁ. Biofilm dispersal: mechanisms, clinical implications, and potential therapeutic uses. J Dent Res2010;89:205-18.

9. Lord DM, Baran AU, Wood TK, Peti, Page R.Bdca, a protein important for escherichia coli biofilm dispersal, is a short-chain dehydrogenase/reductase that binds specifically to nadph. PloS one 2014;9: e105751.

10. Lee J, Page R, García-Contreras R, Palermino JM, Zhang XS, Doshi O, et al. Structure and function of the Escherichia coli protein YmgB: a protein critical for biofilm formation and acid-resistance. J Mol Biol2007;373:11-26.

11. Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, et al, Escherichia coli k-12: A cooperatively developed annotation snapshot—2005. Nucleic Acids Res2006;34:1-9.

12. Zhang XS, García-Contreras R, Wood TK. Escherichia coli transcription factor yncc (mcbr) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein ybim (mcba). ISME J2008;2:615-31.

13. Tuckerman JR, Gonzalez G, Sousa EH, Wan X, Saito JA, AlamM,et al. An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-gmp control. Biochemistry 2009;48:9764-74.

14. Wang X, Preston III and T. Romeo. The pgaabcd locus of escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation. J Bacteriol2004;186:2724-34.

15. Wang X, Kim Y, Hong SH, Ma Q, Brown BL, Pu M, et al. 2011. Antitoxin mqsa helps mediate the bacterial general stress response. Nat Chem Biol2011;7: 359-66.

16. Kim JS, Kim YJ, Seo S, SeongMJ, Lee K. Functional role of bdm during flagella biogenesis in escherichia coli. CurrMicrobiol2015;70:369-73.

17. Ogasawara H, Yamamoto K, Ishihama A. Role of the biofilm master regulator csgd in cross-regulation between biofilm formation and flagellar synthesis. J Bacteriol2011;193:2587-97.

18. Itoh Y, Rice JD, Goller C, Pannuri A, Taylor J, Meisner J, et al. Roles of pgaabcd genes in synthesis, modification, and export of the escherichia coli biofilm adhesin poly-β-1, 6-n-acetyl-d-glucosamine. J Bacteriol2008;190:3670-

80.

19. SzklarczykD, Santos A, Von Mering C, Jensen LJ, Bork P, Kuhn M. Stitch 5: Augmenting protein–chemical interaction networks with tissue and affinity data. Nucleic Acids Res2016;44:D380-4.

20. Consortium U.Uniprot: A worldwide hub of protein knowledge. Nucleic Acids Res2019;47:D506-15.

21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol1990;215:403-10.

22. Berman HM, Battistuz TN, Bhat WF, Bluhm PE,Bourne K, Burkhardt Z,et al. The protein data bank. Acta Crystallogr D Biol Crystallogr2002;58:899-907.

23. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc2015;10:845-58.

24. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, et al. Eplant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. Plant Cell 2017;29:1806-21.

25. Xu, D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J2011;101:2525-34.

26. Laskowski RA, Rullmann JA, MacArthur MW, KapteinR, Thornton JM. Aqua and procheck-nmr: Programs for checking the quality of protein structures solved by nmr. J Biomol NMR1996;8: 477-86.

27. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al, Pubchem 2019 update: Improved access to chemical data. Nucleic Acids Res2019;47:D1102-9.

28. Todsen WL. Chemdoodle 6.0. J Chem Inf Model2014;54:2391-93.

29. Balaji GL, Rajesh K, Priya R, Iniyavan P, Siva R, Vijayakumar V. Ultrasound-promoted synthesis, biological evaluation and molecular docking of novel 7-(2-chloroquinolin-4-yloxy)-4-methyl-2H-chromen-2-one derivatives. Med Chem Res 2013;22:3185-92.

30. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform2012;4:1-7.

31. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, GoodsellDS. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. J Comput Chem 2004;30:2785-91.

32. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ucsf chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004;25: 1605-12.

33. Hemalatha CN,Muthkumar VA. Application of 3d qsar and docking studies in optimization of perylene diimides as anti cancer agent. Indian J Pharm Educ Res 2018;52:666-75.

34. Desai VH, Kumar SP, Pandya HA, Solanki HA. Receptor-guided de novodesign of dengue envelope protein inhibitors. Appl BiochemBiotechnol2015;177:861-78.

35. Soreghan B, KosmoskiJ, Glabe C. Surfactant properties of alzheimer's a beta peptides and the mechanism of amyloid aggregation. J Biol Chem1994;269:28551-4.

36. Reva BA, Finkelstein AV, Skolnick J. What is the probability of a chance prediction of a protein structure with anrmsd of 6 å? Fold Des1998;3:141-7.

Supplementary table 1. Identification of biofilm target proteins along with their amino acid sequence which retrieved from UniProtKB

Protein name	Aminoacid sequence
bdcA	>sp|P39333|BDCA_ECOLI Cyclic-di-GMP-binding biofilm dispersal mediator protein OS=Escherichia coli (strain K12) OX=83333 GN=bdcA PE=1 SV=2 MGAFTGKTVLILGGSRGIGAAIVRRFVTDGANVRFTYAGSKDAAKRLAQETGATAVFTDS ADRDAVIDVVRKSGALDILVVNAGIGVFGEALELNADDIDRLFKINIHAPYHASVEAARQ MPEGGRILIIGSVNGDRMPVAGMAAYAASKSALQGMARGLARDFGPRGITINVVQPGPID TDANPANGPMRDMLHSLMAIKRHGQPEEVAGMVAWLAGPEASFVTGAMHTIDGAFGA
ygmA	>sp|P75992|YMGA_ECOLI Probable two-component-system connector protein YmgA OS=Escherichia coli (strain K12) OX=83333 GN=ymgA PE=2 SV=1 MKTSDNERIKYEITGQAVLQILRMKINFSLQTLIKQLLVMKSAEEDAFRRDLIDSIIRDF SNSDSGGPNRRTATADNKSMFNGKKINRIH
tabA	>sp|P0AF96|TABA_ECOLI Toxin-antitoxin biofilm protein TabA OS=Escherichia coli (strain K12) OX=83333 GN=tabA PE=1 SV=1 MIIGNIHNLQPWLPQELRQAIEHIKAHVTAETPKGKHDIEGNRLFYLISEDMTEPYEARR AEYHARYLDIQIVLKGQEGMTFSTQPAGAPDTDWLADKDIAFLPEGVDEKTVILNEGDFV VFYPGEVHKPLCAVGAPAQVRKAVVKMLMA
ycgZ	>sp|P75991|YCGZ_ECOLI Probable two-component-system connector protein YcgZ OS=Escherichia coli (strain K12) OX=83333 GN=ycgZ PE=2 SV=1 MHQNSVTLDSAGAITRYFAKANLHTQQETLGEIVTEILKDGRNLSRKSLCAKLLCRLEHA TGEEEQKHYNALIGLLFE

Inhibitory effect of rhamnolipid on biofilm producing protein

mcbR	>sp|P76114|MCBR_ECOLI HTH-type transcriptional regulator McbR OS=Escherichia coli (strain K12) OX=83333 GN=mcbR PE=1 SV=2 MPGTGKMKHVSLTLQVENDLKHQLSIGALKPGARLITKNLAEQLGMSITPVREALLRLVS VNALSVAPAQAFTVPEVGKRQLDEINRIRYELELMAVALAVENLTPQDLAELQELLEKLQ QAQEKGDMEQIINVNRLFRLAIYHRSNMPILCEMIEQLWVRMGPGLHYLYEAINPAELRE HIENYHLLLAALKAKDKEGCRHCLAEIMQQNIAILYQQYNR
dosC	>sp|P0AA89|DOSC_ECOLI DiguanylatecyclaseDosC OS=Escherichia coli (strain K12) OX=83333 GN=dosC PE=1 SV=1 MEMYFKRMKDEWTGLVEQADPPIRAKAAEIAVAHAHYLSIEFYRIVRIDPHAEEFLSNEQ VERQLKSAMERWIINVLSAQVDDVERLIQIQHTVAEVHARIGIPVEIVEMGFRVLKKILY PVIFSSDYSAAEKLQVYHFSINSIDIAMEVMTRAFTFSDSSASKEDENYRIFSLLENAEE EKERQIASILSWEIDIIYKILLDSDLGSSLPLSQADFGLWFNHKGRHYFSGIAEVGHISR LIQDFDGIFNQTMRNTRNLNNRSLRVKFLLQIRNTVSQIITLLRELFEEVSRHEVGMDVL TKLLNRRFLPTIFKREIAHANRTGTPLSVLIIDVDKFKEINDTWGHNTGDEILRKVSQAF YDNVRSSDYVFRYGGDEFIIVLTEASENETLRTAERIRSRVEKTKLKAANGEDIALSLSI GAAMFNGHPDYERLIQIADEALYIAKRRGRNRVELWKASL
pgaB	>sp|P75906|PGAB_ECOLI Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase OS=Escherichia coli (strain K12) OX=83333 GN=pgaB PE=1 SV=1 MLRNGNKYLLMLVSIIMLTACISQSRTSFIPPQDRESLLAEQPWPHNGFVAISWHNVEDE AADQRFMSVRTSALREQFAWLRENGYQPVSIAQIREAHRGGKPLPEKAVVLTFDDGYQSF YTRVFPILQAFQWPAVWAPVGSWVDTPADKQVKFGDELVDREYFATWQQVREVARSRLVE LASHTWNSHYGIQANATGSLLPVYVNRAYFTDHARYETAAEYRERIRLDAVKMTEYLRTK VEVNPHVFVWPYGEANGIAIEELKKLGYDMFFTLESGLANASQLDSIPRVLIANNPSLKE FAQQIITVQEKSPQRIMHIDLDYVYDENLQQMDRNIDVLIQRVKDMQISTVYLQAFADPD GDGLVKEVWFPNRLLPMKADIFSRVAWQLRTRSGVNIYAWMPVLSWDLDPTLTRVKYLPT GEKKAQIHPEQYHRLSPFDDRVRAQVGMLYEDLAGHAAFDGILFHDDALLSDYEDASAPA ITAYQQAGFSGSLSEIRQNPEQFKQWARFKSRALTDFTLELSARVKAIRGPHIKTARNIF ALPVIQPESEAWFAQNYADFLKSYDWTAIMAMPYLEGVAEKSADQWLIQLTNQIKNIPQA KDKSILELQAQNWQKNGQHQAISSQQLAHWMSLLQLNGVKNYGYYPDNFLHNQPEIDLIR PEFSTAWYPKND
pgaD	>sp|P69432|PGAD_ECOLI Biofilm PGA synthesis protein PgaD OS=Escherichia coli (strain K12) OX=83333 GN=pgaD PE=1 SV=1 MNNLIITTRQSPVRLLVDYVATTILWTLFALFIFLFAMDLLTGYYWQSEARSRLQFYFLL AVANAVVLIVWALYNKLRFQKQQHHAAYQYTPQEYAESLAIPDELYQQLQKSHRMSVHFT SQGQIKMVVSEKALVRA
mqsA	>sp|Q46864|MQSA_ECOLI Antitoxin MqsA OS=Escherichia coli (strain K12) OX=83333 GN=mqsA PE=1 SV=1 MKCPVCHQGEMVSGIKDIPYTFRGRKTVLKGIHGLYCVHCEESIMNKEESDAFMAQVKAF RASVNAETVAPEFIVKVRKKLSLTQKEASEIFGGGVNAFSRYEKGNAQPHPSTIKLLRVL DKHPELLNEIR
mcbA	>sp|P0AAX6|MCBA_ECOLI Uncharacterized protein McbA OS=Escherichia coli (strain K12) OX=83333 GN=mcbA PE=2 SV=2 MKKCLTLLIATVLSGISLTAYAAQPMSNLDSGQLRPAGTVSATGASNLSDLEDKLAEKAR EQGAKGYVINSAGGNDQMFGTATIYK
bdm	>sp|P76127|BDM_ECOLI Protein bdm OS=Escherichia coli (strain K12) OX=83333 GN=bdm PE=2 SV=2 MFTYYQAENSTAEPALVNAIEQGLRAQHGVVTEDDILMELTKWVEASDNDILSDIYQQTI NYVVSGQHPTL
csgD	>sp|P52106|CSGD_ECOLI CsgBAC operon transcriptional regulatory protein OS=Escherichia coli (strain K12) OX=83333 GN=csgD PE=1 SV=1 MFNEVHSIHGHTLLLITKSSLQATALLQHLKQSLAITGKLHNIQRSLDDISSGSIILLDM MEADKKLIHYWQDTLSRKNNNIKILLLNTPEDYPYRDIENWPHINGVFYSMEDQERVVNG LQGVLRGECYFTQKLASYLITHSGNYRYNSTESALLTHREKEILNKLRIGASNNEIARSL FISENTVKTHLYNLFKKIAVKNRTQAVSWANDNLRR
pgaC	>sp|P75905|PGAC_ECOLI Poly-beta-1,6-N-acetyl-D-glucosamine synthase OS=Escherichia coli (strain K12) OX=83333 GN=pgaCPE=1 SV=1 MINRIVSFFILCLVLCIPLCVAYFHSGELMMRFVFFWPFFMSIMWIVGGVYFWVYRERHW PWGENAPAPQLKDNPSISIIIPCFNEEKNVEETIHAALAQRYENIEVIAVNDGSTDKTRA ILDRMAAQIPHLRVIHLAQNQGKAIALKTGAAAAKSEYLVCIDGDALLDRDAAAYIVEPM LYNPRVGAVTGNPRIRTRSTLVGKIQVGEYSSIIGLIKRTQRIYGNVFTVSGVIAAFRRS ALAEVGYWSDDMITEDIDISWKLQLNQWTIFYEPRALCWILMPETLKGLWKQRLRWAQGG AEVFLKNMTRLWRKENFRMWPLFFEYCLTTIWAFTCLVGFIIYAVQLAGVPLNIELTHIA ATHTAGILLCTLCLLQFIVSLMIENRYEHNLTSSLFWIIWFPVIFWMLSLATTLVSFTRV MLMPKKQRARWVSPDRGILRG
bsmA	>sp|P39297|BSMA_ECOLI Lipoprotein BsmA OS=Escherichia coli (strain K12) OX=83333 GN=bsmA PE=2 SV=2 MVSRKRNSVIYRFASLLLVLMLSACSALQGTPQPAPPVTDHPQEIRRDQTQGLQRIGSVS TMVRGSPDDALAEIKAKAVAAKADYYVVVMVDETIVTGQWYSQAILYRK

Supplementary table 2.Identifying query coverage as well as confidence score of homology modelling

Protein name	Query coverage %	Confidence score %
bsmA	63	99.99
ycgZ	63	97.40
ygmA	54	42.60
mcbA	63	99.90
bdm	63	97.40
pgaC	95	100.00
tabA	99	99.99
csgD	93	100.00
pgaD	39	72.80

Supplementary table 3. Root Mean Square Deviation (RMSD) and TM score of target proteins after energy minimization with respect to their native conformation

Protein name	RMSD (A^0)
pgaC	1.010
tabA	1.302
csgD	0.526

Supplementary table 4. Ramachandran plot analysis of predicted proteinsanalysis of ligand-Receptor interactions

Protein name	% Present in Disallowed region
pgaC	0.5
tabA	1.6
csgD	0.0

Supplementary table 5. Molecular docking energy (kcal/mol) of Rhamnolipid against selected Target proteins. Efficient binding energy lower than -6 (kcal/mol) withcorresponding target proteins were highlighted as bold, and the name mention in italic design designated as top 3 docked protein as per the binding energy

Protein name	Energy of docked proteins in (kcal/mol)
pgaC	-8.91
tabA	-6.08
csgD	-7.75
dosC	-5.08
bdcA	-7.04
mcbR	-5.72
pgaB	-7.30
mqsA	-7.54