Protein docking and molecular dynamics simulation of predicted IcaA and IcaD protein structure from *Staphylococcus epidermidis* involved in PNAG biosynthesis and trafficking

Ramachandira Prabu[1]*, Mohanty Amaresh Kumar[2], Balaji Gupta V.L.N.T.[3], Yadav Rajesh Mudi[3] and Veeramani T.[1]

1. Department of Biotechnology, PRIST University, Vallam Campus, Thanjavur - 613 403, Tamil Nadu, INDIA
2. Department of Bioinformatics, Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry - 605 014, INDIA
3. Department of Basic Science, Vishnu Institute of Technology, Vishnupur, Bhimavaram - 534202, INDIA

*rcprabu86@gmail.com

Abstract

IcaA is poly-beta-1,6-N-acetyl-D-glucosamine (PNAG) synthase, an enzyme producing activated homopolymer chain of N-acetylglucosamine from Staphylococcus epidermidis. IcaA is an integral membrane-associated N-Acetylglucosamine transferase that requires accessory IcaD protein for maximal expression and synthesis. We identified icaA gene sequence carries reputed G-quadruplexes region and characterized 164 bp upstream promoter sequence of IcaADBC operon. We characterized transcription factor binding sequence derived from promoter region as PurR, DegU transcription factor and E-Box invariant sequence. We identified IcaD gene sequence carries E-Box sequence overlap with IcaB gene of the operon which may act as transcription start site (TSS) for IcaB gene expression. Due to lack of crystal structure in PDB database, protein threading methodology was used for construct IcaA and IcaD three-dimensional structure.

Our findings reveal that IcaA C-terminal helices region between Phe295-Arg394 connected by loops, may act as lid domain of the protein. Asp134, Ile224, Glu226, Asp227, Arg266, Trp267 are active site amino acids and binding free energy was calculated as ΔG score −5.13 KJ/mol using Autodock 4.2.6. Molecular dynamics simulation (MDS) was performed to evaluate stability of IcaA protein, IcaA-UDP-GluNAc binding complex and IcaD protein throughout trajectories captured within time scale 100 ns simulation period using GROMACS 4.5. Different binding energies were calculated for IcaA-UDP-GluNAc complex using GROMACS tools MM-PBSA.

Keywords: *Staphylococcus epidermidis*, Promoter region, Transcription factor, Protein three dimensional structures, MD simulation.

Introduction

Glycosyl transferases (GTs) are ubiquitous in nature and chemically diverse enzymes were reflecting on anomeric configuration. Epimers at each chiral central point, branching and linkage position, catalyze polysaccharide synthesis through glycosidic bonding between activated nucleotide sugars donors and acceptor.[1,4,5,13,25] Poly-β-1,6-N-acetyl-D-glucosamine (PNAG) homopolymer consists of N-acetyl-D-glucosamine molecules associated with glycosidic linkage between anomeric C1 and C6 carbons. UDP-GlcNAc is a vital key starting material for the biosynthesis of bacteria cell wall and act as nucleotide-sugar donar for complete activity of SseK glycosyltransferase.[2,9,16,35] Predominant role of bacterial GlcNAc signalling is associated with de nova synthesis during cell wall remodeling through catabolism, recycling of GlcNAc and provides enough UDP-GlcNAc for newly dividing cell growth.[7,10,15,28,30,32]

IcaA is an integral membrane protein comprising of four transmembrane (TM) helices and belongs to glycosyltransferase 2 (GT-2) family for synthesis of poly β-1,6-linked GlcNAc oligosaccharides from UDP-GlcNAc precursor. Generally inverting GT-2 enzymes proceed through direct displacement (S_N2 mechanism) accompanied by nucleophilic attack by anomeric carbon of the acceptor molecule via transfer of proton to catalytic bases and leaving group departure. GT-A and GT-B are two structural folds of GT-2 family of enzymes. GT-A is specifically dependent on nucleotide-sugar donor whereas other GT-B folds are dependent on lipid phosphosugar donar.

GT-A folds called Rossmann fold type co-ordinate divalent cations for catalysis. IcaA catalytic domain was present in cytosolic side of cell membrane for binding and cleavage of UDP (leaving phosphate group departure) and GlcNAc followed by GlcNAc shuttling to inner membrane region without aid of undecaprenyl phosphate lipid carrier. *icaA* gene expression is either directly regulated by promoter region or via IcaR/TcaR regulatory protein.[29] In other scenario, insertion element meditated *icaA* gene regulation through IS256 insertion or excision via on or off switch mechanism. IcaD called poly-beta-1,6-N-acetyl-D-glucosamine synthase, a short integral membrane protein consisting of two TM domain predicted for enhancing IcaA full function. *icaD* gene sequences were overlapped with *icaA* at 5′-end by 37 nt, results in *icaAD* genes were both co-expressed. *icaD* gene 3'-end overlapped with *icaB* gene at 5'-end by 4 nt.

IcaD function predicted to be involved in IcaA active folding confirmation and membrane insertion, may act as protein chaperone and in regulation of PNAG synthesis. Hydroxypropyltrimethyl ammonium chloride chitosan (HACC) may enhance antibacterial activity to inhibit microbial biofilm formation on the titanium surface and to prevent orthopedic implant-related infections by blocking *icaA* gene transcription. IcaA protein has been found to be predominant virulence gene identified among clinical isolates for biofilm formation than commensal *S. epidermidis* populations.[33]

Caenorhabditis elegans infection model reveals that *S. epidermidis* 9142, is a PIA producing strain showing no virulence in *icaA* insertion mutant than its isogenic lines. However, virulence was restored in the wild-type from complementation with *icaADBC* clone in biofilm-negative mutant.[26]

IcaAD homologs called PgaCD in *E. coli* found similar function with perceptive to structure, inner membrane localization, PNAG polymerization and regulation. PgaCD acts as c-di-GMP receptor, whereas c-di-GMP act as second messenger binding with two protein enhancing allosteric interaction and stabilizing complex for PNAG synthesis.[3,5,8,11,12,19,21,31] Bis-(3′-5′)-cyclic dimeric GMP (c-di-GMP) and guanosine-bis 3′, 5′(diphosphate) (ppGpp) are second messenger involved in signal transduction. Their concentration levels involved in *E.coli* PNAG production and biofilm development are directly proportional to PgaD and PgaA maximum expression level. Positive-biofilm forming ATCC 35984 strain and methicillin-resistant MRSE 287 strain were isolated from orthopaedic implant-related infected patient for understanding inhibitory nature of chinese herbal SanHuang decoction (SH) of *S. epidermidis* growth.

SH down-regulates *icaAD* gene expression and inhibition of dPNAG synthesis.[29] *icaA*, *icaD* and *aap* genes greatly improve biofilm-forming ability and clinical specimen has shown strongest attachment on silica gel surface from breast surgery were measured from confocal laser scanning microscope (CLSM). *icaD* and *mecA* gene were co-expressed, leading to increase antibiotics resistance. *icaD* gene were investigated that act as important virulence markers of slime formation and adhesive aptitude.

There is a stronger co-relation of *icaAD* gene expression altogether, which increases slime production rather than single *icaA* or *icaD* gene expression.[20] *icaA* or *icaD* gene function may be lost or mutated leads to decreased slime production. Isogenic biofilm-negative *icaA*-insertion mutant was less susceptible to polymorphonuclear neutrophils (PMNs) killing after opsonisation. Human β-defensin 3 (hBD-3) has been effective against methicillin-resistant *Staphylococcus epidermidis* (MRSE) shown to have decreased *icaA* and *icaD* transcription and increased *icaR* expression.[34] Protein threading methodology has been proposed for structural and functional elucidation of IcaAD protein. Predicted binding pattern of IcaA-UDP-GlcNAc represents active site amino acids and binding energy scores were manually examined using molecular visualization tools, following MD simulation with individual IcaD, IcaA, IcaA-UDP-GlcNAc complex in GROMACS 4.5. Trajectory period of molecular simulation was evaluated for complete structural stability of IcaD, IcaA, IcaA-UDP-GlcNAc complex protein. Different energies of IcaA-UDP-GlcNAc complexes were calculated using MM-PBSA.

Material and Methods

IcaA and IcaD protein characterization: *S. epidermidis* RP62A strain was selected and full length *IcaAD* gene sequence with promoter region was downloaded from GenBank (http://www.ncbi.nlm.nih.gov/BLAST). Quadruplex forming G-Rich sequences were performed with IcaA nucleotide sequences to detect QGRS (https://bioinformatics.ramapo.edu/QGRS/analyze.php). HD, DXD, TED, Q(Q/R)xRW, xED signature motif were identified from IcaA protein sequence. Scratch protein predictor was used for *in silico* overexpression of bacterial host, protein solubility, protein disorder region (http://scratch.proteomics.ics.uci.edu/). Musite deep tools were used for protein posttranslational modifications site prediction (https://www.musite.net/).

Other posttranslational modifications site prediction was done with Predictprotein (https://predictprotein.com). Database of transcriptional regulation in *Bacillus subtilis* (DBTBS) was used for identification of promoter cis-acting region and trans-acting elements (https://dbtbs.hgc.jp/). Metal ion-binding site prediction and docking server (MIB) were used for Mn2+ divalent ion binding site for IcaA protein (http://bioinfo.cmu.edu.tw/MIB/).

IcaA and IcaD protein 3D structure prediction: IcaA truncated region of 264 amino acids was used with deletion of N-terminal and C-terminal transmembrane region considered for three-dimensional protein structure prediction, built using threading methodology on I-tasser server (https://zhanggroup.org/ I-TASSER/). IcaD full length 101 amino acids sequences were derived for three-dimensional protein structure prediction. The predicted structure was screened for loop refinement using ModBase tool (http://modbase.compbio.ucsf.edu/modloop/).

The predicted model was evaluated with SAVES (http://nihserver.mbi.ucla.edu/SAVES/). ProFunc database was used for secondary structure prediction using ProMotif program and biochemical function of IcaA and IcaD protein (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/).
The transmembrane model quality score was analyzed from QMEANBrane.[27] PyMOL tools was used for 3D structure visualization.

IcaA protein-ligand docking: Largest area and volume were chosen specifically for active site prediction. IcaA-

UDP-GlcNAc blind and focused docking was performed using Schrodinger software. UDP-GlcNAc (UDP-N-acetyl-D-glucosamine) ligand structure and molecular formula ($C_{17}H_{27}N_3O_{17}P_2$) were retrieved from PubChem site [445675] in SDF file format (http://pubchem.ncbi.nlm.nih.gov). IcaA and UDP-GlcNAc conversion were prepared with protein preparation and LigPrep module of Schrodinger suite. Grid-box was generated to cover entire active site of IcaA protein structure. UCFC chimera molecular visualization tools were used to get insights into docking pattern and LIGPLOT was used to obtain 2D interaction plots (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/).

Molecular dynamics simulations: To evaluate protein structure stability of IcaA, IcaD and IcaA-UDP-GlcNAc ligand complex using molecular dynamics simulation was analyzed with Gromacs 5.1.4. Automated topology builder (ATB) was used to create topology files for UDP-GlcNAc. IcaA, IcaD, IcaA-UDP-GlcNAc complex were solvated with SPC water molecule in cubic shape box placed at the center. 1 nm distance was given between simulation box edge and IcaA-UDP-GlcNAc complex, to move freely in the immersed state completely. Electrostatic energy calculation was analyzed with Particle Mesh Ewald (PME) algorithms permitting Ewald summation and covalent bond constraints used to solve with linear constraint solver (LINCS) algorithms. Four chloride ions (Cl^-) were added to neutralize the system.

Energy minimization of IcaD, IcaA and IcaA-UDP-GlcNAc complex was done with steepest descent approach at 100 ps. Constant pressure, volume and temperature (300 K) were maintained to system equilibration at 100 ps with IcaD, IcaA, IcaA-UDP-GlcNAC complex. 100 ns time scale was set for final MD simulation run and trajectories were saved for further analysis. Root mean square deviation (RMSD), Root mean square fluctuation (RMSF), Radius of gyration, Hydrogen bond graph and Solvent accessible surface were generated using Xmgrace tools.[24] Energy calculation was performed using MM-PBSA (Molecular mechanics Poisson–Boltzmann surface area) from ligand complex file.

Results

Characterization of intercellular adhesion IcaA and IcaD protein: IcaA and IcaD gene accession no. U43366.1 and protein ID AAC06117.1 and AAC06120.1 were derived from NCBI. IcaA and IcaD gene sequence encoded 1239 and 306 nucleotides. Promoter upstream region -164 bp has been retrieved between *IcaR* gene termination site and *IcaADBC* transcription +1 start site (TSS). PurR transcription factor (TF) binding sites at - 92 to -105 nt binding site (5'-AATTACAAATATTA-3') and DegU at – 104 to -119 nt TF binding sites (5'-AAAAATTAAGTTAAAA-3') have been identified from cis-regulatory sequence.

E-Box with consensus sequence motif at position -36 to -41 nt (CATATG) is identified in *IcaA* promoter region. E-Box (CACGTG) is identified in *icaA* gene sequence position at 734 to 739 nt. E-Box consensus sequence CATATG was identified in *icaD* gene sequence position at 300 to 305 nt. Predicted one quadruplex forming G-Rich sequences were identified at start position 772 nt with 30 nucleotide length containing 4 GG repeats with significant G-score 20 (Figure 1).

IcaA and IcaD predicted solubility upon overexpression in bacteria host with range as 0.79132 and 0.786716. IcaA and IcaD predicted probability of antigenicity as 0.185069 and 0.081996. Predicted disorder amino acid residues as MVKPRQ (at 1st to 6th position) and RDI (at 98th -100th position) of IcaA protein have score ≥ 0.5. Predicted disorder amino acid residues as GNIQR (at 408th – 412th position) of IcaD protein have score ≥ 0.5.

DXD belonging to aspartate–any residue–aspartate motif has identified at different position as DAD (Asp134 - Asp136), DDD (Asp140–Asp142), DTD (Asp220–Asp222). Majority of GT-A fold represents invariant DXD motif for support nucleotide sugar and metal ion coordination at active site. xED belongs to inverting GT-A folds with a conserved Asp or Glu at different positon as ED (Glu149-Asp150), (Glu226-Asp227). Prediction of mn^{2+} divalent cation ion binding site other that HD signature motifs revealed two significant score as 1.849 amino acid position hit with Glu42E, His43 and 1.727 at amino acid position hits with Asp134, Asp136, Thr137 of IcaA protein.

N-glycosylation, sumoylation, acetylation, methylation, ubiquitination, hydroxylation are post-translational modification and phosphorylation predominately found at 5 different position in IcaA protein. Other deduced post-translational modification (PTM) sites are protein kinase C phosphorylation site, casein kinase II phosphorylation site, cAMP and cGMP-dependent protein kinase phosphorylation site and N-myristoylation site. 3 Pyrrolidone carboxylic acid and 1 methylation (PTM) site were identified in IcaD protein.

Structural and functional insights of IcaA and IcaD protein: IcaA and IcaD three-dimensional structure was retrieved from I-TASSER sever for threading based structure prediction, due to lack of specific PDB structure from PDB-PSI BLAST.

Position	Length	QGRS	G-Score
772	30	GGTGGTTTATGGAAACAAAGGGTTCGATGG	20

Figure 1: QGRS sequences of *icaA* DNA sequence.

264 aa of IcaA truncated region (Ile28 to Ser289 aa) has taken for protein structure prediction, excluding N-terminal transmembrane region (Phe6 - Phe27 aa) and C-terminal transmembrane region (Leu290-Val315), (Tyr328-Ile354), (Ile367-Phe388) and lid domain region (Phe295-Arg394) (Figure 2). C-terminal transmembrane regions from Phe295 to Arg394 represented 5 helices at position (Phe295-Lys327), (Ser337-Ile346), (Thr349-Ile354) (Gly365-Leu369) (Asn380-Arg394). I-Tasser generated 6 out of 10 model hit with PDB-ID belonging to 4HG6A with normalized Z-score as -7.53.

Out of five model, fifth model was selected from c-score -3.36. Ramachandran plot showed that 96% amino acid residues were found in the favoured region and 3.5% amino acids in allowed regions and 0.4% in the outlier region (Figure 2). IcaD full length has been taken for protein structure prediction and I-Tasser generated 4th model was selected from c-score as – 4.07 (Figure 3). Another 2 out of 10 model hit with PDB-ID belonging to 6b87A of significant Z-score as -1.10 with respective to crystal structure of transmembrane protein TMHC2_E as synthetic construct expressed in *E. coli* expression system. Ramachandran plot shows that 97% amino acid residues were found in the favored region and 3.2% amino acids in allowed regions and 1% in the outlier region (Figure 3).

IcaA PDB secondary structure were predicted as 19 β-turns, 13 helices, 9 strands, 8 Helix-Helix interacs, 6 ϒ-turns, 2 sheets, 2 β-α-β units, 2 β-bulges, 1 β-hairpin, 1 psi loop. IcaD PDB secondary structure were predicted as 6 β-turns, 4 helices, 3 ϒ-turns, 2 Helix-Helix interacs. IcaA structure was visualized with QMembrane tools revealing that cylindrical like shaped membrane-spanning region sandwiched between large cytoplasmic domain for ligand binding and catalytic function, whereas membrane-spanning smaller transmembrane domain formed narrow channel confirmation for PNAG trafficking (Figure 2). IcaD protein structure resembles a rocket-shaped structure with cytoplasmic base and wide transmembrane helices rich in hydrophobic amino acids visualized with PyMOL (Figure 3).

IcaD three dimensional structure consists of 2 helices interface attached to the TM-region, connected by short loop at the periplasmic surface. Overall quality of IcaAD protein three dimensional model assessment from QMEANBane score predicted -100000 found to be within expected range and reliable (Figure 3). IcaA, IcaD cleft, tunnel, pore analysis reveal that interior surface of funnel was found to be higher, hydrophobic amino acids are alanine, valine, leucine, isoleucine, methionine.

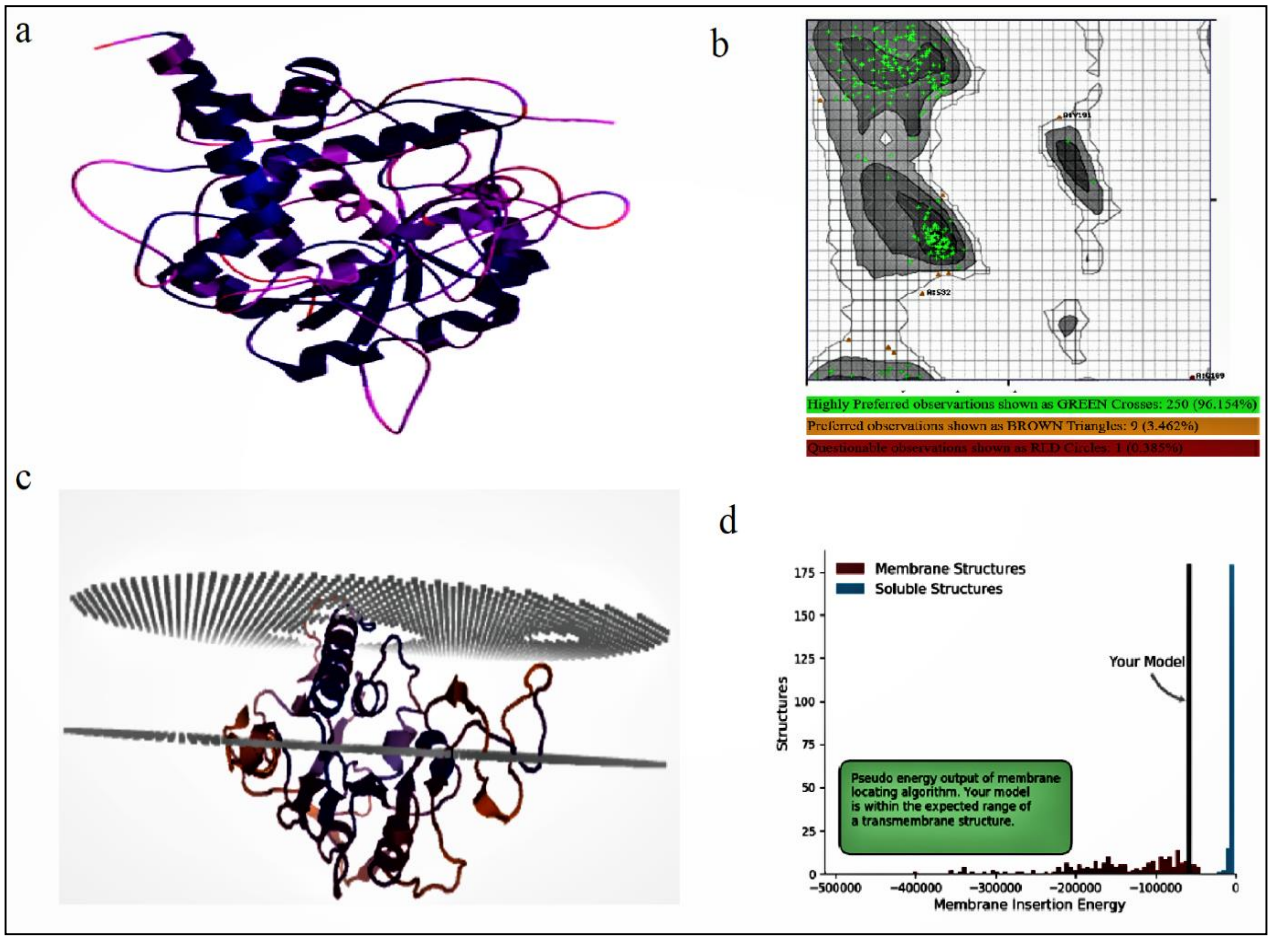

Figure 2: a. Predicted three-dimensional structure of IcaA protein, b. Ramachandran plot, c. Transmembrane model quality assessment with cytoplasmic membrane, d. QMEANBane model scoring

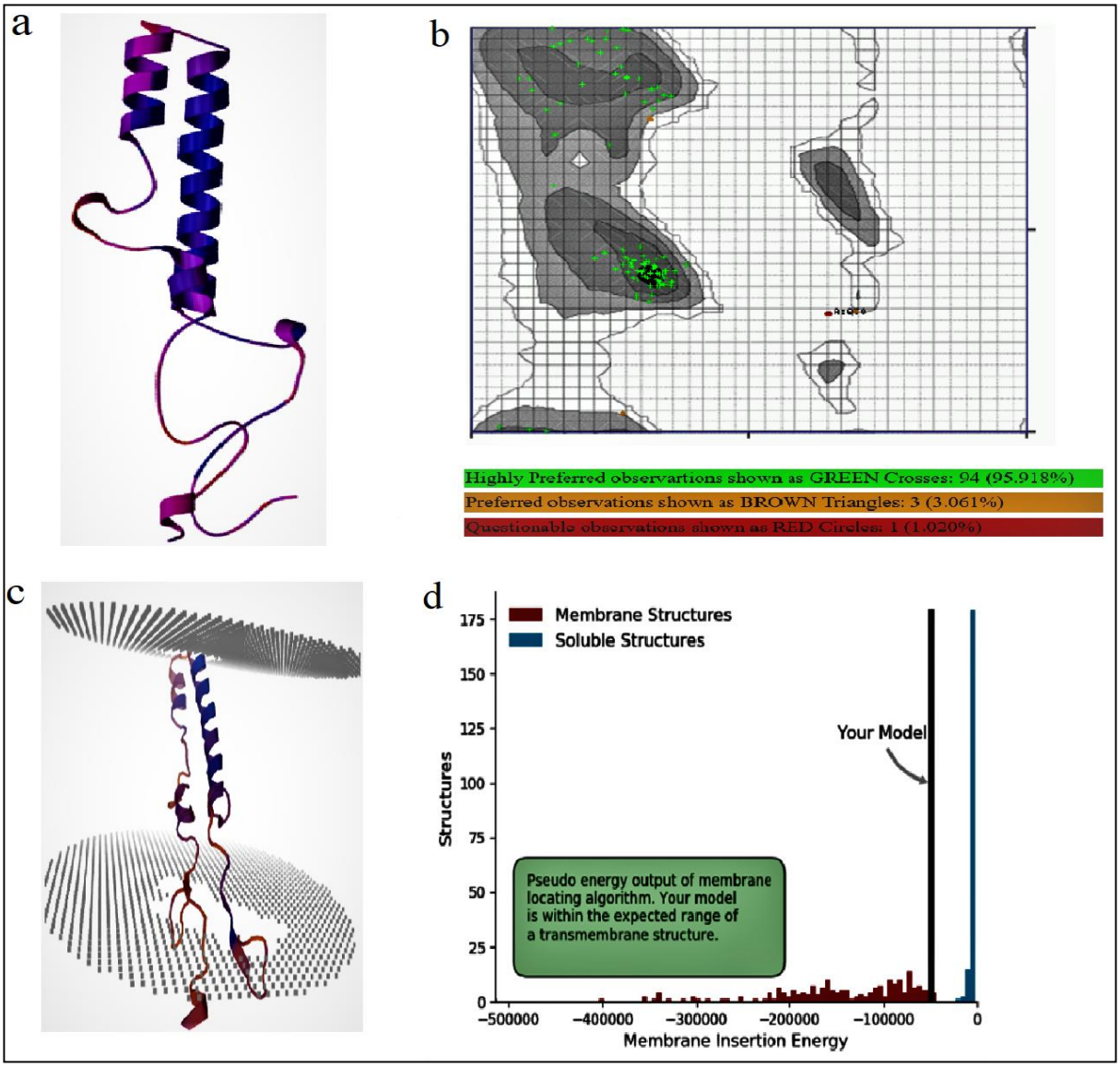

Figure 3: a. Predicted three-dimensional structure of IcaD protein, b. Ramachandran plot, c. Transmembrane model quality assessment with cytoplasmic membrane, d. QMEANBane model scoring

Active site prediction of IcaA protein: The predicted IcaA active site amino acids were Asp134, Ile224, Glu226, Asp227, Arg266, Trp267 hits with blind docking as coherent with predicted active site amino acid. IcaA-UDP-GlcNAc docking with significant binding energy was found as -5.13 kcal/mol. Docking pose analysis revealed ten hydrogen bonds (H_bonds) interactions between UDP-GlcNAc and active site amino acids residues of protein. IIe 201, Val 204, Thr 225, IIe 228, Gly 271 are amino acids from hydrophobic interaction (Figure 4). Here, we observed single H_bond with Ile224, Glu226, Arg 266, Trp267 residues of bond length 2.59 Å, 2.94 Å,3.02 Å, 2.93 Å, while two H_bond formation was observed with Asp134 of bond length 2.96 Å and 2.83 Å, whereas four H_bond formations were observed with Asp227 of bond length 2.92 Å, 2.91 Å, 2.67 Å, 2.74 Å.

Eventually, Trp267 pyrrole ring side chain of 6th nitrogen epsilon (E1) position forms hydrogen bond with α-phosphate group of oxygen of 5th carbon atom of UDP-GlcNAc. Ile224 functional carboxylic acid group bonds with 4th chiral carbon from hydroxyl group of pyronase ring of β-N-acetyl-D-glucosamine. Glu226 side chain (R group) of delta (δ) carbon 1 position forms hydrogen bond with 3rd chiral carbon hydroxyl group of pyronase ring of β-N-acetyl-D-glucosamine. Arg266 eta (η2) positon of positive charge from nitrogen atom of the amino group bonds with 3rd oxygen atom at β-phosphate group (Figure 4). Asp134 side chain (R group) of delta 2nd (δ2) position of oxygen atom forms hydrogen bond with α-phosphate group of 2nd position oxygen atom and β-phosphate group of 2nd oxygen atom.

Asp227 side chain (R group) of delta 1st position (δ1) of oxygen atom binds with uracil structure of uridine molecule at the 6th oxygen position of 4th carbon atom position. Next delta 1st positon (δ1) of oxygen atom binds with N-acetyl-D-glucosamine molecule at 4th chiral carbon from oxygen atom of hydroxyl group of pyronase ring. Asp227 side chain (R group) of delta 2nd (δ2) position of oxygen atom binds with

uracil structure of uridine molecule at 7th oxygen of 5th carbon position. Next Asp227 side chain (R group) of delta 2nd (δ2) position of oxygen atom binds with N-acetyl-D-glucosamine molecule of 2nd chiral carbon atom of 3rd nitrogen atom position. Hydrophobic interactions were observed with Isoleucine Ile201, Val204, Thr225, Ile228, Gly271 amino acids showing an arc with radiating spokes towards UDP-GlcNAc atom with contacts showing spokes radiating back (Figure 4).

Molecular dynamics simulations of IcaA, IcaA- UDP-GlcNAc and IcaD: Root mean square deviations (RMSD) was calculated for IcaA, IcaA-UDP-GlcNAc complex. The graphs were generated for protein backbone flexibility. Throughout simulation period, protein backbone stability was not disturbed and strong interaction with IcaA-UDP-GlcNAc complex at 10 ns and IcaA at 15 ns, with no significant fluctuations was seen. Root mean square fluctuation (RMSF) evaluated the residual mobility of IcaA and IcaA-UDP-GlcNAc molecule against residue number and graph was generated. High fluctuation was observed from 50 to 100 ns during trajectory period of MD simulation (Figure 5).

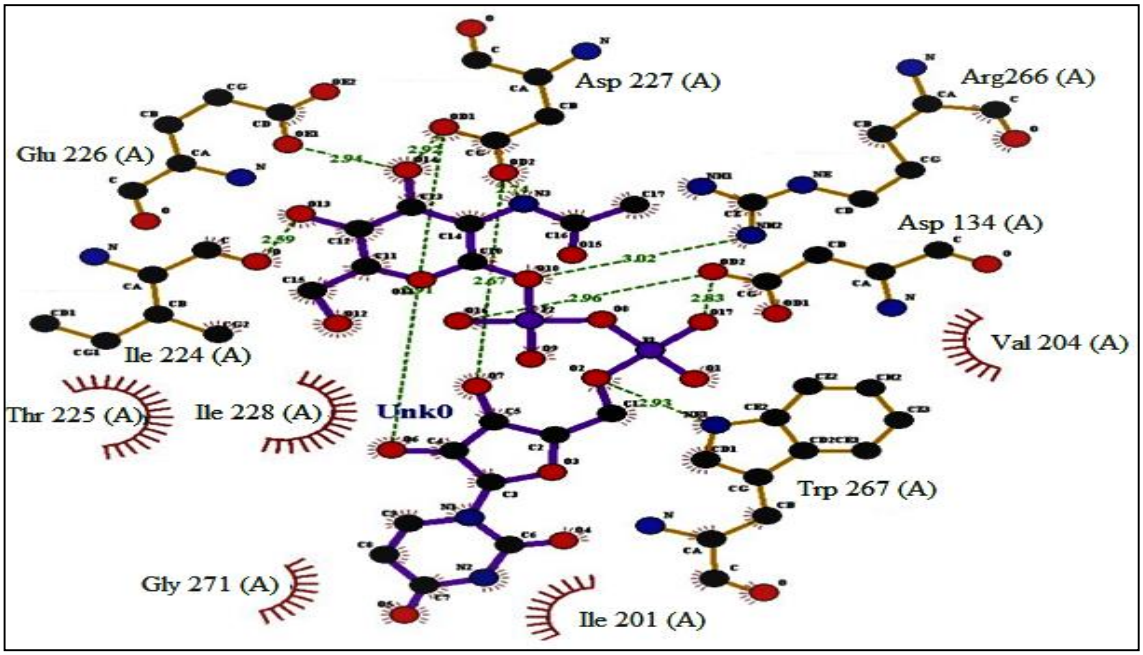

Figure 4: IcaA-UDP-GlcNAc docking.

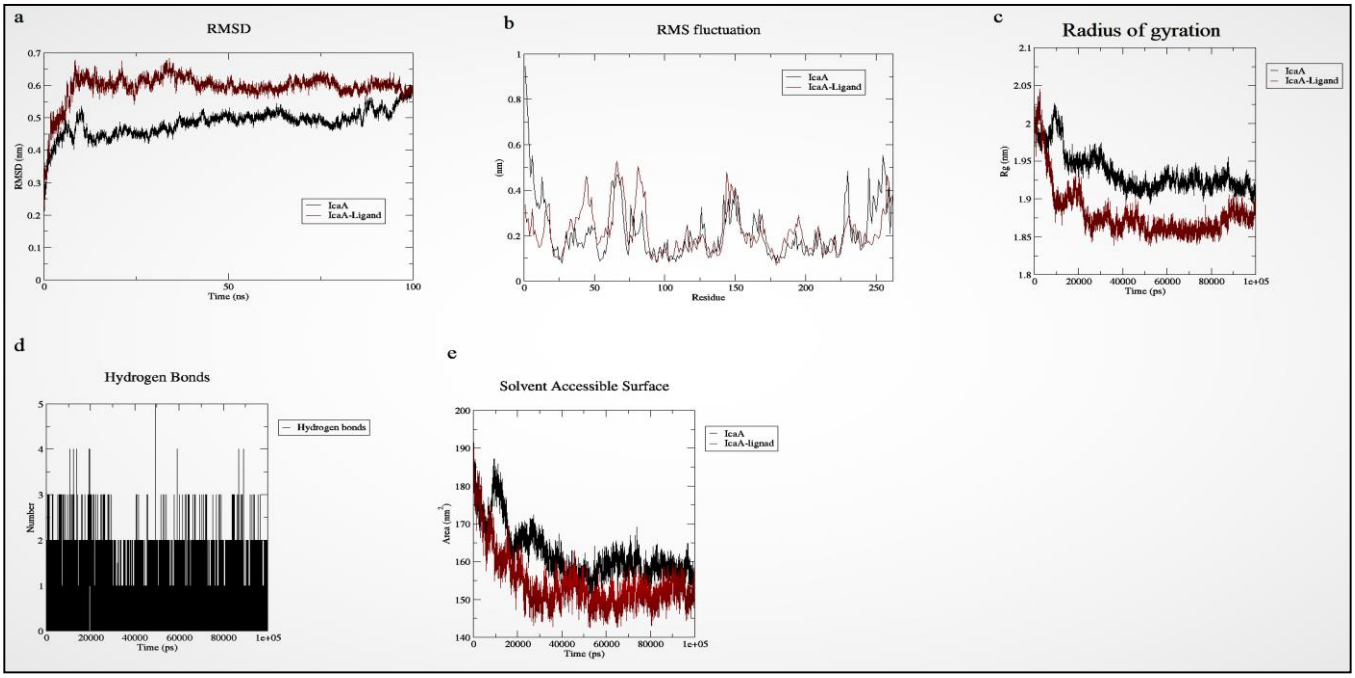

Figure 5: IcaA and IcaA-UDP-GlcNAc MD simulation a. RMSD, b. RMS fluctuation, c. Radius of gyration, d. Hydrogen bonds, e. Solvent accessible surface.

The radius of gyration (Rg) was calculated through high secondary structure compactness decreasing from 20,000 ps to 100,000 ps throughout 100 ns trajectory period of MD simulation. Rg graph starts at y-axis 1.98 nm of IcaA and IcaA-UDP-GlcNAc complex, eventually reaching to the lowest range to 1.9 nm (Figure 5). To determine hydrogen bond stability of IcaA-UPD-GlcNAc, binding sites were monitored throughout trajectory time scale 100 ns. The analysis revealed that H_bonds graph starts at Y-axis with the number of hydrogen bond from 0 to 5 and Y-axis with 0–100 ns period. Three hydrogen bond profiles were shown with 0–1, 1–2, 2–3, 3-4, 4-5. H_bond profile 0–1 range was calculated with no significant fluctuation and one breaking was observed, eventually maintaining stability throughout 100 ns.

Profile 1–2 range was calculated with no significant fluctuation, however least H_bond breakage was observed maintaining stability throughout 100 ns. Profile 2–3 range has a significant range of fluctuation and higher H_bond breakage was observed. Profile 3–4, 4-5 range has a significant range of fluctuation and higher H_bond breakage was observed (Figure 5).

Solvent accessibility surface area results observed that IcaA and IcaA-UDP-GlcNAc reside at 185 nm2 and 180 nm2 in the simulation graph throughout trajectory time scale 100 ns. IcaA and IcaA-UDP-GlcNAc complex showed a gradual decrease and reached to limit of 155 nm2 and 158 nm2 at 100 ns (Figure 5).

Different energies were calculated for icaA-UDP-GlcNAc complex with van der Waal energy as -294.947 +/- 11.700 kJ/mol, electrostattic energy as -54.453+/- 10.179 kJ/mol, polar solvation energy as 179.329 +/- 16.751 kJ/mol, SASA energy as -26.572+/- 1.223 kJ/mol, SAV energy as 0.000 +/- 0.000 kJ/mol, WCA energy as 0.000 +/- 0.000 kJ/mol and binding energy as -196.643 +/-13.473 kJ/mol.

Root mean square deviations (RMSD) was calculated for IcaD protein. The graphs were generated for protein backbone flexibility. Throughout simulation period, protein backbone stability was not disturbed and no significant fluctuations was seen. Root mean square fluctuation (RMSF) was evaluated for the residual mobility of IcaD against residue number and graph was generated (Figure 6). High fluctuation was observed from 90 ns during trajectory period of MD simulation. The radius of gyration (Rg) was calculated by high secondary structure compactness and shown decreasing from 60,000 ps to 100,000 ps throughout 100 ns trajectory period of MD simulation.

Rg graph starts at Y-axis 2.4 nm of IcaD, eventually reaching to the lowest range to 1.6 nm. Solvent accessibility surface area results observed IcaD at 108 nm2 in the simulation graph throughout trajectory time scale 100 ns and showed a gradual decrease and reached to limit of 75 nm2 at 100 ns (Figure 6).

Discussion

Glycosyltransferases (GTs) is one of the predominant enzymes involved in polysaccharides biosynthesis, secondary metabolites, glycoprotein, glycolipids biosynthesis and ubiquitous in nature. There is strong association with promoter regulatory sequence and level of RNA transcripts synthesis. PUR and Degu cis-regulatory region under proximal promoter range of sequence with perceptive strictly control downstream genes under spatial-temporal manner whereas E-box site under core promoter range of sequence is especially for common transcription machinery association and pre-initiation complex.

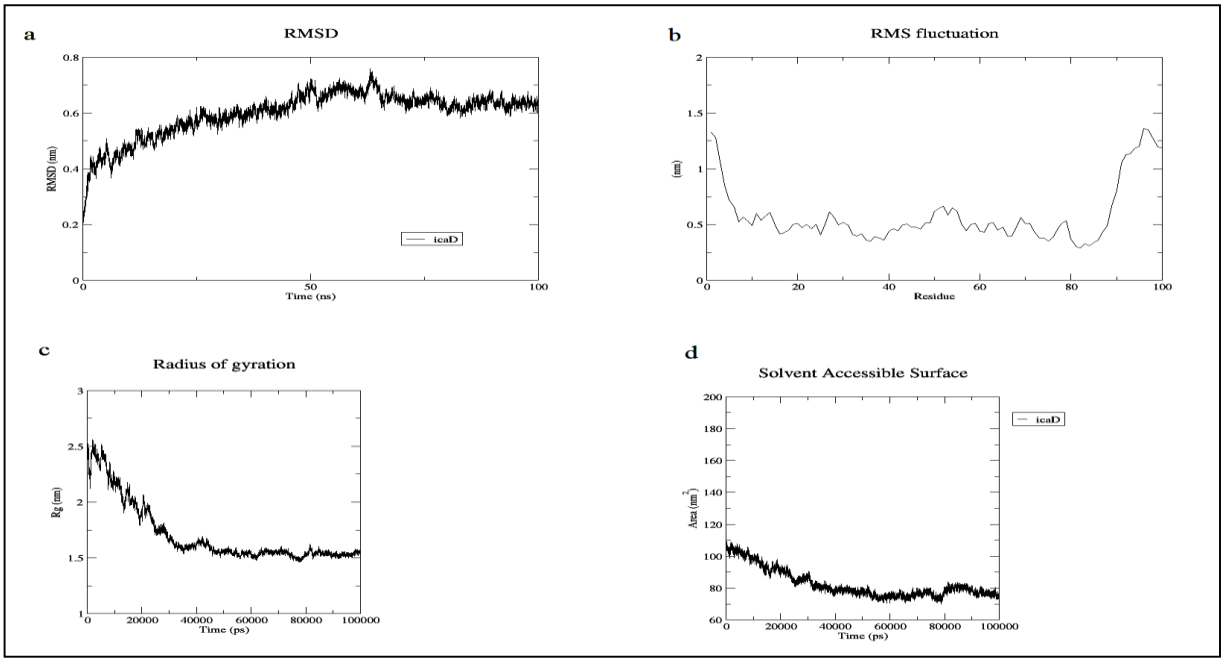

Figure 6: IcaD MD simulation a. RMSD, b. RMS fluctuation, c. Radius of gyration, d. Solvent accessible surface.

PurR transcriptional factor comes under LacI family of metabolite-responsive regulators and plays a critical role in purine biosynthesis regulation, metabolism and virulence in *S. aureus*.[31] PurR transcription factor binds directly to promoter genes acting as virulence responsive master regulators and pathogenic factors to control their gene expression and increased its tempers for pathogenesis. DegU is a pleiotropic and response regulator acting as two-component system with its cognate kinase DegS. DegU is involved in distinct response on post-exponential growth phase of bacteria, genetic competence development, exoprotease production, regulation of gene expression, biofilm formation, degradative enzymes regulation, capsule biosynthesis, swarming motility.[23]

E-Box called enhancer box is involved in gene regulation, where found in promoter or enhancer region acting as DNA response element with specific DNA signature motifs, CANNTG (N-strands for any nucleotide). IcaD protein is engaged on IcaA function with perceptive to enhancer and chaperonin, apart from that may act as promoter for *IcaBC* transcription due to E-box signature motif and may regulate gene expression or enhance level of gene expression.

E-box signature sequence plays a vital role in circadian genes, similar in function like CLOCK-related DNA elements for maintaining circadian rhythmicity acting as circadian oscillator for clock-controlled genes. Due to the presence of E-Box sequence, *icaA and icaD* gene may have other functions like circadian genes involved in adaptation to diverse range of environmental conditions.

G-quadruplexes are DNA and RNA highest-order confirmation systemically form G-rich sequences forming tetrads of hydrogen-bonded guanine bases. G-quartets term retrieved from highly stable G-quadruplex structures, eventually were arranged as stacked G-tetrads. Each guanine base is arranged in arrays of four square co-planar arrangements. This sequence motif of G-quadruplex has been identified in mammalian genome, telomere region, promoter region, protect DNA from nuclease, gene regulation. Quantitative analysis of bacterial membrane proteins is still challenging due to dynamical entities of protein structure modification. IcaA predicted protein solubility was found to be overexpressed in the *E. coli* system and insoluable in nature due to post-translational modification site, whereas *IcaD* is soluble in nature acting as molecular chaperone for IcaA protein and found to be less post-translational modification site.

DAD, DTD, DDD, TED, QRVRW motif of IcaA protein are responsible for ligand binding and enzymatic activity. Carboxylates group coordinate with ligand and act as catalytic base for close proximity of non-reducing end of UDP-GlcNAc, co-ordination of divalent cation binding activity. Aspartic acid of DAD domain at the position Asp134 and aspartic acid of TED domain at position Asp227 and glutamic acid at position Glu226 serve as base for catalytic domain requirements for mn2+ activity and interact with phosphate group primarily of UDP-GlcNAc.

Here, we predict that TED and DAD invariant motifs are essential for glycosyl transfer reaction and are independent of spatial arrangement of atoms of the reaction. xED motifs act as catalytic base with invariant Glu or Asp belonging to inverting GT-As folds functions to deprotonate incoming nucleophile of acceptor and start nucleophilic attack with direct displacement of phosphate leaving group.[14]

Invariant Q(Q/R)xRW) pentapeptide motif present in cellulose synthase, chitin synthase, hyaluronan synthase involved in structure maintenance and function of GT-A fold. Arginine and tryptophan of QRVRW motif at position Arg266 and Trp267 most likely coordinate diphosphate binding of UDP for purpose of phosphate group departure and behave as key catalytic pocket for support of glycosyl transfer for new glycosidic bond formation.[36] Hence, QRVRW motifs provide cytoplasmic entrance for UDP-GlcNAc and may act as glycosyl transfer tunnel region for PNAG trafficking. BcsA inner membrane-bounded glycosyltransferases called cellulose synthases invert anomeric carbon structure from α to β of newly binding UDP-activated glucose (UDP-Glc) in the production of nascent polysaccharide during reaction.

Crystal structure of BcsA subunit of *Rhodobacter* sphaeroides cellulose synthase with perceptive of Trp383 of Q(Q/R)xRW) region was involved in translocating polysaccharide. GT-A type mediated response recruits Mn^{2+} divalent cations for co-ordination and act as base for phosphate group living. This reaction mechanism is Sn^2 direct displacement mechanism with short-lived magnesium divalent cation ion and a catalytic nucleophile. Histidine–aspartate (HD) domain belongs to metalloprotein superfamily and predicted phosphohydrolase activity is involved in stress response, virulence, inflammation, nucleic acid metabolism, signal transduction. HD coordinates at least one metal ion in mono-, di- and trinuclear conformation catalyse range of hydrolysis to oxygenation reaction.[17]

Predicted Asn-X-Thr invariant sequence may necessitate hydrophobic interaction with perceptive to transmembrane protein fold to packing within phospholipid bilayer of IcaA protein. Asparagine residues is vital for interaction between oppositely charged and hydrophobic amino acids within α-helical topological sector. Deduced sumoylation PTM is member of small ubiquitin-like modifier (SUMO) at lysine position 40 and 242, predicted to be regulate IcaA covalent attachment for membrane trafficking, intracellular trafficking, signal transduction and membrane structure stabilization.[18] Acetylation PTM may be involved in cellular process, stress factor detrimental and pathogenesis.

Methylated PTM sites in IcaA and IcaD may be responsible for constitutive and irreversible of bacteria chemical stimuli to sensory adaptation, signal transduction. Lysine

ubiquitination PTM may be involved in protein stability and membrane trafficking. Protein phosphorylation by addition of phosphate group may be involved in transition from apolar state to polar state and may promote association with other molecule via kinase activity for activation.[22] Hydroxyproline from hydroxylation PTM may be involved in the triple helix structure stabilization. Proline hydroxylation is crucial thing for hypoxia promoting factor during stimulus of hypoxia state.

IcaD protein of pyrrolidone carboxylic acid PTM may be involved in the association with multi-pass membrane and extracellular protein, however pyrrolidone carboxylic acid PTM level is associated with decreased state reported in atopic dermatitis and lesional skin. IcaA predicted N-myristoylation post-translational modification (PTMs) is responsible for oncogenesis, cellular localization, signal transduction and acts against viral and bacterial infection.

Lack of crystal structure for intercellular adhesion protein has been major limitation of understanding bacterial biofilm formation and quorum sensing. IcaA consists of nucleic acid binding motif, catalytic domain and only lid domain present at C-terminal position.[9] *Thermomyces lanuginosus* lipase enzymes of mobile lid domain act as specifically covers. This inactive state of lipase enzymes with respect to substrate is inaccessible whereas open conformation displays catalytic activity. Low catalytic activity of lipase enzymes was reported in aqueous to closed confirmation and lid domains in the closed state.

Higher catalytic activity was reported in hydrophobic environment leading to open confirmation and lid domain is opened. Full length IcaA protein with more than two helices are present at C-terminal position confirming the presence of lid domain conformation found in close proximity to the active site, due to the reason that UDP-GlcNAc ligand docking failed with full length IcaA. Removal of transmembrane region particularly lid domain region leads to UDP-GlcNAc substrate docking with IcaA protein.

Overall IcaA structure close resembles to bacterial cellulose synthase of *Cereibacter sphaeroides* and IcaD structure resembles close to transmembrane protein TMHC2_E is rich in hydrophobic amino acids effective for protein stabilization within phospholipid bilayer. IcaA was characterized by repeated alpha-helices and beta-sheets in its structure resemble like sandwich for nucleotide binding. Hydrophobic interaction of IcaA protein is stabilizing tertiary structure and is the principal force behind stability and folding determination in phospholipid bilayer.

Isoleucine, valine, alanine, leucine and methionine constitute 33.8% and 39.7% predominant hydrophobic amino acid residues, which are thermodynamically favorable with perceptive to IcaA and IcaD folding and opposed by desolvation of charged and polar group. IcaA metabolic pathways recruit intercellular adhesion operon IcaDBC undergoing temporal expression during extreme environmental niches for bacterial biofilm formation, whereas extensive lineage of co-expression partners as bhp, atlE, eno predicted similar cellular function.

Cellulose synthesis and translocation proposed model for BcsA Rhodobacter sphaeroides with perceptive to Cys318 and Trp383 during glycosyl transfer with newly added sugar moiety via acetal linkage should rotate during synthesis in the polymer plane. This phenomenon of direction of rotation determined by β-1,4 glucan formation and steric interaction and second repeated UDP-GlcNAc was involved in the same process in opposite direction results of steric constraints. Stability of IcaA, IcaA-UDP-GlcNAc complex and IcaD was compactly folded and conformation in the native state was derived from solvent accessible surface area.

IcaA is stably folded throughout the point of time with the relative lowest Rg value considered with tight packing of IcaA protein in terms of Helix-helix interaction. Exposed hydrophobic amino acids and buried hydrophilic amino acids are found in IcaA, IcaD protein, hence there is possibility of predicting protein structure and folding characteristics from estimation of solvent accessibility surface area.

Conclusion

We reported cis-acting DNA sequence of PurR (metabolite-responsive) and DegU (Response regulator) present in the proximal promoter region of *icaADBC* operon model. Our novel findings of E-Box DNA signature motifs in *icaA* core promoter region and particularly *IcaD* gene sequence overlap with *IcaB* gene, may act as promoter region for enhancing *IcaB, IcaC* gene expression. We identified *icaA* DNA sequence with reputed G-quadruplexes region, predicted to be involved against unambiguous nuclease activity and promotes gene regulation. We highlighted three dimensional structure of IcaA *Staphylococcus epidermidis*, which gain insights into structure and function of GT-A folds of GT-2 type of family protein.

Our key finding on active sites, protein signature sequence, PTMs characterization necessitates UDP-GlcNAc binding and trafficking. MD simulation and different energy calculation will be major interface with perceptive to IcaA, IcaD and IcaA-ligand complex stability for potential drug targets and major interface of quorum sensing biology.

Acknowledgement

Authors are thankful to the Pondicherry University for providing bioinformatics facilities to carry out the work.

References

1. Alamery S., Al Ajmi A., Wani T.A. and Zargar S.S., *In silico* and *in vitro* exploration of poziotinib and olmutinib synergy in lung cancer: role of hsa-miR-7-5p in regulating apoptotic pathway marker genes, *Medicina (B Aires)*, **59**, 1923, 10.3390/medicina59111923 **(2023)**

2. Alam M., Abser M.N., Kumer A., Bhuiyan M.M.H., Akter P., Hossain M.E. and Chakma U., Synthesis, characterization, antibacterial activity of thiosemicarbazones derivatives and their computational approaches: quantum calculation, molecular docking, molecular dynamic, ADMET, QSAR, *Heliyon*, **24**, 16222, 10.1016/j.heliyon.2023.e16222 **(2023)**

3. Assefa M. and Amare A., Biofilm-associated multi-drug resistance in hospital acquired infections: a review, *Infect Drug Resist*, **15**, 5061-8, doi: 10.2147/IDR.S379502 **(2022)**

4. Bilal M.S., Ejaz S.A., Zargar S., Akhtar N., Wani T.A., Riaz N., Aborode A.T., Siddique F., Altwaijry N., Alkahtani H.M. and Umar H.I., Computational investigation of 1, 3, 4 oxadiazole derivatives as lead inhibitors of VEGFR 2 in comparison with EGFR: density functional theory, molecular docking and molecular dynamics simulation studies, *Biomolecules*, **12**, 1612, 10.3390/biom12111612 **(2022)**

5. Cheng X., Zhu Y., Huang J., Li Y., Jiang X. and Yang Q., A neutral polysaccharide from Persicaria hydropiper (L.) Spach ameliorates lipopolysaccharide-induced intestinal barrier injury via regulating the gut microbiota and modulating AKT/PI3K/mTOR and MAPK signaling pathways, *J. Ethnopharmacol*, **320**, 10.1016/j.jep.2023.117403 **(2024)**

6. Choi E.H., Jung M., Choi J. and Hwang J., Pyrrolidone carboxylic acid levels or caspase-14 expression in the corneocytes of lesional skin correlates with clinical severity, skin barrier function and lesional inflammation in atopic dermatitis, *J. Dermatol Sci.*, **76(3)**, 231-9, 10.1016/j.jdermsci.2014.09.004 **(2014)**

7. Ding D., Wang B., Zhang X., Zhang J., Zhang H., Liu X., Gao Z. and Yu Z., The spread of antibiotic resistance to humans and potential protection strategies, *Ecotoxicol. Environ. Saf.*, **254**, 114734, 10.1016/j.ecoenv.2023.114734 **(2023)**

8. Donmig J.C., Patricia C.N.G., Matthew L.C.G., Charlotte D.V.F., Acelah E.C., Emmanuel P.D., Chloe R.M.Y.G., Trisha N.R.F. and Mary R.F.L., Molecular Characterization and Homology Modeling of Intercellular Adhesion Regulatory (IcaR) Proteins in Biofilm-Producing Staphylococcus Species, *Asian Journal of Biological and Life Sciences*, **12**, 301-309, 10.5530/ajbls.2023.12.41 **(2023)**

9. Hardwidge P.R., Qaidi S.E., Scott N.E. and Hays M.P., Arginine glycosylation regulates UDP-GlcNAc biosynthesis in Salmonella enterica, *Sci Rep.*, **12**, 5293 **(2022)**

10. Hlashwayo D.F., Noormahomed E.V., Bahule L., Benson C.A., Schooley R.T., Sigaúque B., Barrett K.E. and Bila C.G., Susceptibility antibiotic screening reveals high rates of multidrug resistance of Salmonella, Shigella and Campylobacter in HIV infected and uninfected patients from Mozambique, *BMC Infect. Dis.*, **23**, 255, 10.1186/s12879-023-08219-7 **(2023)**

11. Huang Y., Wang Y., Pannuri AA., Ni D., Zhou H., Cao X., Lu X. and Romeo T., Structural Basis for Translocation of a Biofilm-supporting Exopolysaccharide across the Bacterial Outer Membrane, *J. Biol Chem.*, **291(19)**, 10046–10057, 10.1074/jbc.M115.711762 **(2016)**

12. Jenal U., Steiner S., Lori C. and Boehm A., Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein-protein interaction, *EMBO J.*, **32(3)**, 354-68 **(2013)**

13. Kang X., Ma Q., Wang G., Li N., Mao Y., Wang X., Wang Y. and Wang G., Potential mechanisms of quercetin influence the ClfB protein during biofilm formation of *staphylococcus aureus*, *Front. Pharmacol*, **13**, 10.3389/fphar.2022.825489 **(2022)**

14. Kannan N., Taujale R., Li S., Edison A.S. and Moremen K.W., Deep evolutionary analysis reveals the design principles of fold A glycosyl transferases, *Elife*, **9**, e54532, 10.7554/eLife.54532 **(2020)**

15. Kim Y.K. et al, Antibiofilm effects of quercetin against Salmonella enterica biofilm formation and virulence, stress response and quorum-sensing gene expression, *Food Control*, **137(8)**, 108964 **(2022)**

16. Konopka J.B., N-Acetylglucosamine Functions in Cell Signaling, *Scientifica*, **15**, 10.6064/2012/489208 **(2012)**

17. Langton M., Sun S., Ueda C., Markey M., Chen J., Paddy I., Jiang P., Chin N., Milne A. and Pandelia M.E., The HD-Domain Metalloprotein Superfamily: An Apparent Common Protein Scaffold with Diverse Chemistries, *Catalysts*, **10(10)**, 1191, 10.3390/catal10101191 **(2020)**

18. Leithe E., Sirnes S., Bruun J., Yohannes Z. and Rivedal E., The gap junction channel protein connexin 43 is covalently modified and regulated by SUMOylation, *J. Biol Chem*, **287(19)**, 15851-61, 10.1074/jbc.M111.281832 **(2012)**

19. Liang X., Tu C., Li Y., Sun J., Zhao R., Ran J., Jiao L., Huang J. and Li J., Inhibitory mechanism of quercetin on Alicyclobacillus acidoterrestris, *Front. Microbiol*, **14**, 10.3389/ fmicb.2023.1286187 **(2023).**

20. Liu B., Zhou S., Chao X., Fei M. and Dai Y., Analysis of S. Epidermidis icaA and icaD genes by polymerase chain reaction and slime production: a case control study, *BMC Infect Dis*, **13**, 242 **(2013)**

21. Majumdar G. and Mandal S., Exploring the inhibitory role of persicaria hydropiper bioactive compounds against 2KID protein associated with Staphylococcus aureus biofilm formation: molecular docking and pharmacological property analysis, *Res. J. Pharm. Tech.*, **16**, 3189-3194, 10.52711/0974-360x.2023.00524 **(2023)**

22. Muzio L.L., Ardito F., Giuliani M., Perrone D. and Troiano G., The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review), *Int J Mol Med.*, **40(2)**, 271-280, 10.3892/ijmm.2017.3036 **(2017)**

23. Ogura M., Yoshikawa H. and Chibazakura T., Regulation of the Response Regulator Gene degU through the Binding of SinR/SlrR and Exclusion of SinR/SlrR by DegU in *Bacillus subtilis*, *J Bacteriol*, **196(4)**, 873–881, 10.1128/JB.01321-13 **(2014)**

24. Pronk S., Páll S., Schulz R., Larsson P., Bjelkmar P. and Apostolov R., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics*, **29(7)**, 845–854, 10.1093/bioinformatics/btt055 **(2013)**

25. Rehman U.M., Ali A., Ansar R., Arafah A., Imtiyaz Z., Wani T.A., Zargar S. and Ganie S.A., *In Silico* molecular docking and dynamic analysis of natural compounds against major non-structural proteins of SARS-COV-2, *J. Biomolecul. Struct. Dyn*, **41**, 9072-9088, 10.1080/07391102.2022.2139766 **(2023)**

26. Rohde H., Büttner H. and Mack D., Structural basis of Staphylococcus epidermidis biofilm formation: mechanisms and molecular interactions, *Front Cell Infect Microbiol*, **5**, 10.3389/fcimb.2015.00014 **(2015)**

27. Schwede T., Studer G. and Biasini M., Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane), *Bioinformatics*, **30(17)**, i505–i511, 10.1093/bioinformatics/btu457 **(2014)**

28. Shamsudin N.F. et al, Antibacterial effects of flavonoids and their structure-activity relationship study: a comparative interpretation, *Molecules*, 10.3390/molecules27041149 **(2022)**

29. Tan H., Zhang S., Wang P. and Shi X., Inhibitory properties of Chinese Herbal Formula SanHuang decoction on biofilm formation by antibiotic-resistant Staphylococcal strains, *Sci Rep.*, **11**, 7134, 10.1038/s41598-021-86647-8 **(2021)**

30. Thomas G.H., Atkin K.E., MacDonald S.J. and Potts J.R., A different path: Revealing the function of staphylococcal proteins in biofilm formation, *FEBS Letters*, **588(10)**, 1869-1872 **(2014)**

31. Torres V.J., Sause W.E., Balasubramanian D. and Irnov I., The purine biosynthesis regulator PurR moonlights as a virulence regulator in *Staphylococcus aureus*, *PNAS*, **116**, 13563-13572, 10.1073/pnas.1904280116 **(2019)**

32. Wang G., Wang Y., Yao L., Gu W., Zhao S., Shen Z., Lin Z., Liu W. and Yan T., Pharmacological activity of quercetin: an updated review, *Evid.-Base. Complement. Alternat. Med.*, 10.1155/2022/3997190 **(2022)**

33. Yu F., Guo Y., Ding Y., Liu L., Shen X., Hao Z., Duan J., Jin Y. and Chen Z., Antimicrobial susceptibility, virulence determinants profiles and molecular characteristics of Staphylococcus epidermidis isolates in Wenzhou, eastern China, *BMC Microbiology*, **19**, 157 **(2019)**

34. Zhang X., Zhu C., Tan H., Cheng T., Shen H., Shao J., Guo Y. and Shi S., Human β-defensin 3 inhibits antibiotic-resistant Staphylococcus biofilm formation, *J Surg Res.*, **183(1)**, 204-13, 10.1016/j.jss.2012.11.048 **(2013)**

35. Zheng L., Meng J., Jiang K., Lan H., Wang Z., Lin M., Li W., Guo H., Wei Y. and Mu Y., Improving protein-ligand docking and screening accuracies byincorporating a scoring function correction term, *Brief. Bioinform*, **23**, 10.1093/bib/bbac051 **(2022)**

36. Zimmer J., Morgan J.L.W. and Strumillo J., Crystallographic snapshot of cellulose synthesis and membrane translocation, 10.1038/nature11744 **(2013)**.

(Received 16th February 2025, accepted 23rd April 2025)