Int. J. Pharm. Investigation, 2025; 15(4):1227-1234.
https://www.jpionline.org

Original Article

# Identification of Essential Proteins as Promising Therapeutic Targets against Drug Resistant Strain of *Staphylococcus aureus*: An *in silico* Approach

**Monish Mukul Das[1,*], Sayan Chakraborty[1], Avijit Mondal[2], Tapas Guha[3]**

[1]Department of CSE AIML and CST, JIS College of Engineering, Kalyani, West Bengal, INDIA.
[2]Department of CSBS, Techno Main Salt Lake, Kolkata, West Bengal, INDIA.
[3]Department of CSE, Techno International, Newtown, Kolkata, West Bengal, INDIA.

## ABSTRACT

**Background:** *Staphylococcus aureus* is a gram-positive pathogen which contributes community-associated and nosocomial serious diseases and it has become resistant to various antimicrobials. Owing to rapid advent and continuous dispersal of MRSA strains, there is an urgent need to identify therapeutic targets to deal with global threat of persistent infections caused by MRSA strains. **Materials and Methods:** Machine learning approach (DNN), subtractive genomics channel analysis, qualitative characterization for cellular location of proteins, structure-based modelling and eventually, analysis for stereochemical quality and physico-chemical properties of target protein (WP_000249839.1) were employed to identify potential drug target. **Results:** 105 non-homologous essential proteins identified as common drug targets, were analysed using BLASTp against 'VFDB' which resulted in 25 virulent proteins. Among them, 2 were identified as antibiotic-resistant proteins. One out of 2 antibiotic resistant proteins, was identified as cytoplasmic protein (DapD) which was selected as drug target. The stereochemical quality and physico chemical analysis confirmed 'DapD' protein as potential drug target. **Conclusion:** Unique strategy employed in this study would be useful in future in comprehending pathogenesis of bacterium as well as developing and designing potent therapeutic candidates.

**Keywords:** MRSA, DNN, VFDB, DapD, MSSA, DEG, WGS, GA, SVM, ARG-ANNOT.

**Correspondence:**
**Dr. Monish Mukul Das**
Department of CSE AIML and CST, JIS College of Engineering, Kalyani-741235, West Bengal, INDIA.
Email: monishmukul.das@jiscollege.ac.in
ORCID: 0000-0003-1860-8674

**Received:** 06-02-2025;
**Revised:** 18-04-2025;
**Accepted:** 27-06-2025.

## INTRODUCTION

*Staphylococcus aureus* under the family of Staphylococcaceae, is a gram-positive and versatile pathogen that mainly causes minor to deadly infections to human and animals. This bacterium could cause infection to wounds of living creatures which was ascertained first time by Sir Alexander Ogston in 1881 (Ogston, 1881). Though a lot of advances has been made in the control of this pathogen, yet it depicts deadly effects while it is defiant to antimicrobials. Generally, this pathogen can be obtained from healthy organisms (Taylor and Unakal, 2021) and it causes serious infection while it enters into the cells or blood circulation (Schito, 2006; Turner *et al*., 2019). *S aureus* strains defiant to methicillin compound, is considered as MRSA strains while this bacterium becomes sensitive to methicillin compound, it is regarded as MSSA strains (Ochoa *et al*., 2020). MRSA strains, observed first time in 1961 are resistant to numerous antimicrobials (Mickymaray *et al*., 2018). Further on, it was reported that a widespread disease caused by MRSA new clones, was occurred throughout the developed countries during 1980's and 1990's (Nandhini *et al*., 2022). In 2005, a report revealed that mortality of patients caused by MRSA pathogens was higher than fatalities brought about by AIDS/HIV (Nandhini *et al*., 2022). However, disease caused by MRSA strains generally occurs through settings associated with either community or hospitals. The hospital-borne MRSA infection with higher occurrence was delineated globally from the 60's (Dantes *et al*., 2013). Nowadays, MRSA bacterium is considered as epidemic in majority of the hospitals throughout the world and it has become a serious global threat (Nandhini *et al*., 2022). The defiance of *S. aureus* strains to methicillin compound is related to clonal variability in SCC (*Staphylococcus cassette* Chromosome) containing 'mec gene complex' that provides defiance to all β-lactam antibiotics. The 'mec A gene' produce which binds with penicillin but it has less binding attraction for wide-ranging antibiotics containing lactam (Lindsay, 2013). A number of virulence factors like lipase, serine protease, cysteine proteases, leukocidins and alpha toxin are regulated by Quorum sensing process in MRSA strains (Jenul and Horswill, 2019). The functions of various virulence molecules

**DOI:** 10.5530/ijpi.20250261

that cause staphylococcal disease are incomprehensible. Globally, $CC_5$, $CC_8$, $CC_{22}$, $CC_{30}$, $CC_{45}$ as clonal complex and $ST_{239}$ as sequence type have been identified as main isolates. Therefore, it is still obstruction to combat resistance of MRSA strains. In 2008, a study on pathogenesis of MRSA disease showed that MRSA strains are more virulent that that of MSSA. But strains of MRSA do not always show added virulence than that of MSSA (Gordon and Lowy, 2008). MRSA virulent strains have developed unique mechanism to resist quickly against newly introduced drugs (Kaur and Chate, 2015). As the functions of virulent factors are required by bacterium to cause pathogen to hosts, aiming at virulence certainly is a delightful which target *in vitro* cell growth of bacterium not unlike traditional drugs (Clatworthy, Pierson, and Hung, 2007). Actually, aiming at virulence as well as '*in vivo*' essentials is required to be explored in respect of therapeutic strategies (Clatworthy, Pierson, and Hung, 2007). However, scanty information is available in respect of this destructive MRSA TCH-32767 strain which belongs to Clonal Complex 5 ($CC_5$). There is always a gap between rapid development of new class antimicrobials and actual requirement which has entailed the need to explore potential drug targets against MRSA TCH-32767 strain RS to Ceftaroline fosamil by executing computational methods while complete genome sequence of this strain are available (McNeil *et al.*, 2022). In a previous work (Sangeetha *et al.*, 2020), drug target proteins were modelled using homology modelling technique and effective drug compounds were identified from PubChem databank for MRSA treatment. Further on, herbal drug compounds duly screened and evaluated by computer aided design, were used against infection caused by MRSA strains (Skariyachan *et al.*, 2011). The present study has been aimed to predict most virulent and antibiotic resistant non-homologous protein as potential drug target against MRSA TCH-32767 strain using *in silico* methods which have been devised meticulously depending on empirical findings and science publications indicating its appropriateness with grand achievements (Hossain *et al.*, 2013). In recent years, methods based on genetic data were executed to recognize potential targets in different types of pathogens and these methods are considered as the bottom line of antimicrobial development (Barh *et al.*, 2011; Hossain *et al.*, 2013). Methods employed in present investigation includes use of machine learning classifier (DNN), extraction of critical genomic data, subtractive genomics channel analysis, structure-based approach for modelling and validation of target protein and its physico chemical analysis.

## MATERIALS AND METHODS

In the present study, newer approach of Deep Neural Networks (DNN) combined with subtractive genomics analysis was applied to predict non-homologous essential genes from whole proteome of MRSA TCH-32767 strain (Das and Sarkar, 2022; Shoukat, Rasheed, and Sajid, 2012). The strategy employed, has been presented here under:

### Stage I

In the 1st stage, essential gene locus tags having e-score value >0.6 and gene locus tags of WGS (Whole Genome Sequences) of 15 strains of *S. aureus* were retrieved from e-Path(Fu *et al.*, 2012) and NCBI Gen Bank, respectively. Gene locus tags of whole genome sequences matched with essential gene locus tags were marked as 'essential genes' and the rest gene locus tags of whole genome sequences were denoted as 'non-essential genes'. It was observed that the sum total number of essential and non-essential sequence stood 1394 and 38161, respectively (Table 1). The strategy for redundancy reduction in majority class (non-essential gene sequence) was executed using CD-HIT (Fu *et al.*, 2012) with protein sequence similarity cutoff of 60% threshold to eliminate paralogous protein sequences resulting in non-paralogous protein sequences. Thereafter, random undersampling process was repeated 10 times with a view to decrease class imbalance in the dataset (Plaimas, Eils, and König, 2010) and increase efficacy of DNN classifier on new imbalance diminished dataset. With reference to each gene present in imbalance reduced dataset, 14-genome sequence-based features viz. L_aa (Length of amino acid), Gravy (hydropathicity of protein), Fop (Frequency of optimal codons), $N_c$ (effective no. of codons), CBI (Codon Bias Index), CAI (Codon Adaptation Index), GC (GC content of gene), $GC_{3S}$ (G+C content at 3rd position of synonymous codons), $G_{3S,}$ $C_{3S,}$ $A_{3S,}$ $T_{3S}$ (base composition at silent sites), L_sym (Length of system amino acids) and Aromo (frequency of aromatic amino acid), were carefully elicited using CodonW software(Sharp *et al.*, 2005) keeping '0' and '1' for essential and non-essential sequence, respectively as target variable classes. Eventually, 11-critical genetic features like $C_{3S,}$ $G_{3S,}$ CAI, CBI, Fop, $N_c$, $GC_{3S,}$ GC, L_aa, Gravy and Aromo were screened from sequence-based dataset comprising of familiar features using GA in combination with SVM-Gaussian Kernel. The Z-score normalization was performed using standard scaler function of 'sk-learn' genetic-opt module of python keeping Standard Deviation (SD) and mean '1' and '0', respectively(Das and Sarkar, 2022). The missing values were removed using 'Pandas' module in Python. DNN with MLP approach was newly designed (Figure 2), which was trained and validated with 80% and 10% of sub dataset, respectively. The rest 10% data of subset was then utilized for testing the classifier which was utilized for 10-fold cross validation purpose based on critical features of sub dataset. DNN was trained for 1000 epochs with 'Adam' optimizer. Accuracy metrics have been computed using confusion matrix calculated for respective 10 equal divisions of sub dataset and ultimate accuracy was ascertained based on average of accuracy metrics at the end of 10-f-CV. Area under Receiver Operating Characteristics (AU-ROC) (Figure 3) and Precision Recall (PR) (Figure 4) curves were generated with newly designed DNN with MLP using sub dataset for each fold. The scikit-learn 1.1.2 (python library) was used to execute the classifier. The 'Leaky-ReLU' and 'Sigmoid' activation function were implemented in hidden layers and output layer, respectively.

In addition, 'Binary cross entropy' was also used as loss function in DNN.

### Stage II

The complete genome of MRSA-TCH 32767 (ACC-ID NZ_CP064772.1) was downloaded from NCBI GenBank for predicting unseen essential genes. DEG 2020, VFDB, UniProt database (Universal Protein Resource) and ARG-ANNOT AAV6 were also downloaded. The critical genomic features of selected pathogen were obtained using CodonW and G.A. with SVM. Trained DNN designed in Stage I, was executed on 11 critical features of each genome to predict essential protein sequences which were used as input of subtractive genomics channel analysis where in CD-HIT, UniProt-BLASTp, DEG-BLASTp, VFDB-BLASTp and BLASTp against ARG-ANNOT were employed to identify non-homologous virulent and antibiotic-resistant essential proteins for therapeutic targets. The resultant set of 25 essential virulent proteins was treated with BLASTp against PSORTb V 3.0.2 (Yu *et al.*, 2010) to find out locations of these proteins which were considered as common drug targets. However, the resultant set of 105 essential proteins was subjected to BLASTp against VFDB (Chen *et al.*, 2016) to identify most virulent target proteins. Thereafter, target proteins were again treated with BLASTp against ARG-ANNOT AA V. 6 July 2019 (Kaur, Kalia, and Taneja, 2021) for shortlisted antibiotic-resistant proteins. Eventually, the 'DapD', was considered as potential drug target (NCBI ID: IXH 55_RS 00600). Overall flow chart and filtering of protein of subtractive channel analysis have been presented in Figure 1 and Table 2, respectively.

### Stage III

The 'DapD' protein sequence was modelled (Figure S1) with high accuracy using Alpha Fold 3 model(Abramson *et al.*, 2024) and the modelled protein structure was validated employing 'PROCHECK' which generated 'Ramachandran plot' (Figure S2). The physicochemical analysis of 'DapD' protein sequence (WP_000249839.1) was executed (Table 3) using ProtParam tool (Gasteiger, 2005).

## RESULTS

The present study is a combination of efficient deep neural networks, unique subtractive genomics and other computational approaches to identify essential genes in *Staphylococcus aureus* strains and subsequent identification of potential therapeutic targets against MRSA TCH-32767 strain as exhibited in Table 1 and Figure 1, respectively. In this study newly designed DNN (Figure 2) was explored to analyse the accuracy of this classifier based on key genomic features and the result showed that DNN classifier achieved very good AUC score with higher precision-recall. It was revealed from Table 1 that essential gene number stood on lower side due to filtering of essential sequences having e-value <0.6. In addition, computational study displayed a total of 298 essential proteins in MRSA TCH-32767 strain and the steps for filtering protein data in subtractive channel analysis was delineated in Table 2. A complete workflow was presented in Figure 1. A sum total of the 292 proteins was obtained as non-paralogous after operation of CD-HIT program. The BLASTp program resulted in 145 proteins having strong resemblance with proteome of Homo sapiens. A total of 147 non-homologous essential proteomes was further subjected to BLASTp program against 'DEG' Database and 105 essential and non-homologous protein sequences were identified as homologs to essential proteomes which would be considered as common drug target proteins against MRSA TCH-32767. Further on, 105 proteins were against analysed using BLASTp program against 'VFDB' Database which resulted in 25 essential virulent proteins. However, only 2 out of 25 drug target proteins, were shortlisted as antibiotic-resistant proteins using BLASTp program against Database of ARG-ANNOT. Concurrently, in this research work, 'PSORTb v 3.0' defined 'DapD' (NCBI ID: IXH55_RS00600) as cytoplasmic protein and it was identified as most virulent and antibiotic resistant. This 'DapD' protein thus selected as drug target (Protein ID: WP_000249839.1) against MRSA TCH-32767 strain. Finally, 3D model of 'DapD' protein was performed using 'Alpha Fold 3' model (Figure S1) i.e. structure based method and quality of modelled, structure was validated using 'PROCHECK'

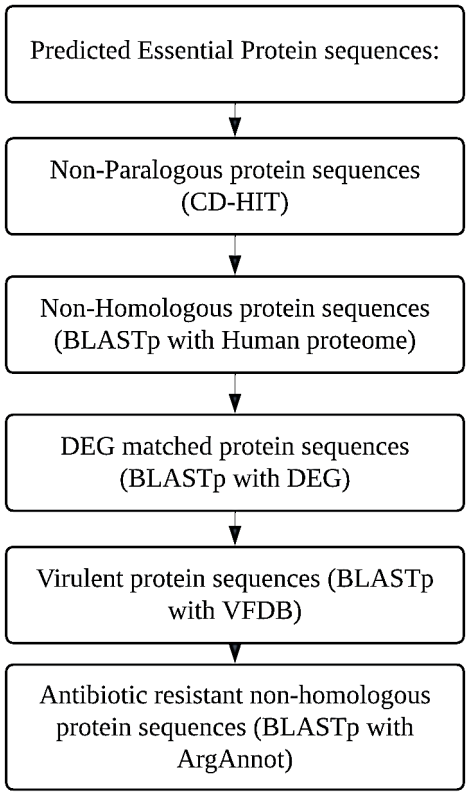


**Figure 1:** Overall flowchart showing subtractive genomic analysis.

**Table 1: Genomic sequences used in present study.**

| Sl. No. | Organism name | NCBI Accession ID | Essential sequence No. | Non-essential sequence No. | Total encoded sequence number |
|---|---|---|---|---|---|
| 1. | *Staphylococcus aureus* subsp. 11819-97 | CP003194.1 | 93 | 2540 | 2663 |
| 2. | *Staphylococcus aureus* subsp. 55/2053 | CP002388.1 | 90 | 2442 | 2532 |
| 3. | *Staphylococcus aureus* subsp. aureus USA300_FPR3757 | CP000255.1 | 93 | 2467 | 2560 |
| 4. | *Staphylococcus aureus* subsp. 71193 | CP003045.1 | 94 | 2529 | 2623 |
| 5. | *Staphylococcus aureus* subsp. CN1 | CP003979.1 | 94 | 2625 | 2719 |
| 6. | *Staphylococcus aureus* subsp. COL | CP000046.1 | 93 | 2580 | 2673 |
| 7. | *Staphylococcus aureus* subsp. ECT-R 2 | FR714927.1 | 93 | 2396 | 2489 |
| 8. | *Staphylococcus aureus* subsp. ED133 | CP001996.1 | 93 | 2560 | 2653 |
| 9. | *Staphylococcus aureus* subsp. ED98 | CP001781.1 | 93 | 2568 | 2661 |
| 10. | *Staphylococcus aureus* subsp. Newman | AP009351.1 | 92 | 2532 | 2624 |
| 11. | *Staphylococcus aureus* subsp. SA268 | CP006630.1 | 93 | 2563 | 2656 |
| 12. | *Staphylococcus aureus* subsp. TCH60 | CP002110.1 | 94 | 2579 | 2673 |
| 13. | *Staphylococcus aureus* subsp. VC40 | CP003033.1 | 93 | 2375 | 2468 |
| 14. | *Staphylococcus aureus* subsp. ST398 | AM990992.1 | 93 | 2606 | 2699 |
| 15. | *Staphylococcus aureus* subsp. NCTC8325 | CP000253.1 | 93 | 2799 | 2892 |

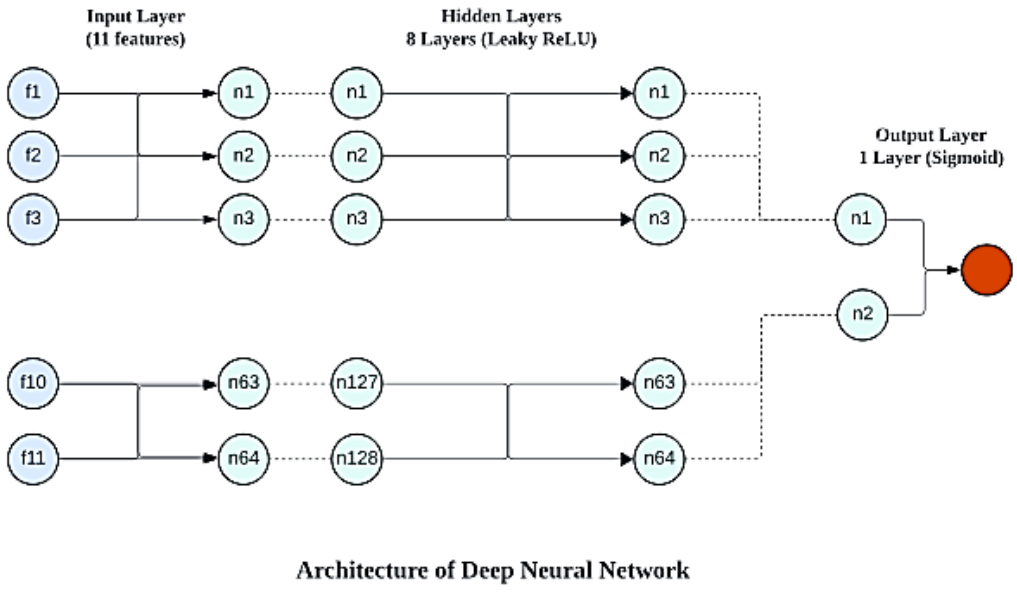


**Figure 2:** Framework of Deep Neural Network.

tool which resulted in 93% residues in the most favoured regions of Ramachandran plots (Figure S2). Further on, the outcome of Prot Param tool (Table 3) delineated the physico-chemical properties i.e. instability index value (35.72<40), isoelectric point (4.71<7), index value of GRAVY (0.108), positive charged residues (21), negative charged residues (32) and aliphatic index value (100) of 'DapD' protein sequence, which would play important functions for development of new antimicrobial against *S. aureus* strains.

## DISCUSSION

In the post-genomic period, treasure trove of completely genome sequence based empirical data for pathogens is being exploited for rapid bioinformatics identification of new potential drug targets (Uddin and Rafi, 2017) which in turn facilitates designing new drug compounds for therapeutic intervention (Miesel, Greene, and Black, 2003). Newly designed 'DNN' model displayed good performance suggesting that the training datasets

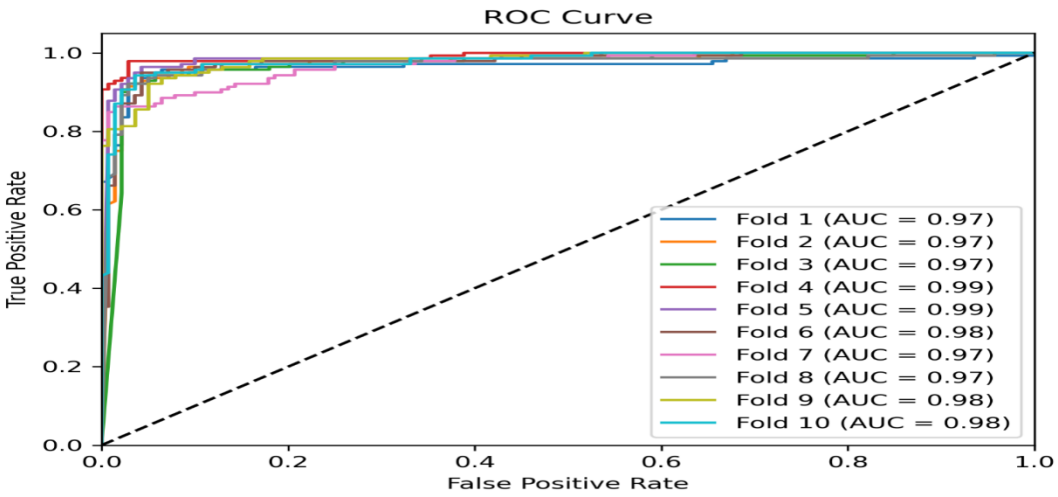


**Figure 3:** ROC curve for 10-fold cross validation.

of key genomic features were satisfactory to gain understanding of signal discrimination for prediction of unrevealed genome sequence in target pathogen. Further on, 10-f-cv displayed very high 'AUC' score of 0.98 for 'DNN' method with higher precision-recall which also supported its unique capability of using learning signals for prediction more accurately (Das and Sarkar, 2022; Hasan and Lonardi, 2020). However, the use of DNN model to identify essential gene sequences was consistent with previous work (Das and Sarkar, 2022; Hasan and Lonardi, 2020; Palaniappan and Mukherjee, 2011).

In addition, the evaluation of 'DNN' model based on critical genomic features yielded good result which also indicated that GA combined with 'SVM' screened critical genetic features more effectively and these features could be utilized for essential genes prediction with higher accuracy. The subtractive genomic analysis which is one of the broadly used methods to identify therapeutic targets was employed in this work to filter a large number of essential genes lacking in host but existing in pathogen (Hema *et al.*, 2015; Hosen *et al.*, 2014; Kaur, Kalia, and Taneja, 2021). The specific subtractive channel proteome analysis identified 2 among 105 homologs to essential proteomes as virulent and antibiotic-resistant. Therefore, these essential proteomes were regarded as usual drug targets. Among 25 virulent proteins, 'DapD' cytoplasmic protein identified as most virulent and antibiotic resistant non-homologous protein was selected as potential drug target against MRSA TCH-32767 strain, which would no more hinder human enzymatic functions. Since meso-DAP (Diaminopimelic acid) and l-lysine are biosynthetic products of 'DapD' protein and biologically significant for peptidoglycan production of cell wall (Van Heijenoort, 2001) as well as cellular viability under stressed condition (Jia *et al.*, 2023), blocking biological activities of 'DapD' by potent inhibitor would certainly destroy the pathogenic bacteria. It was reported that 'DapD' protein i.e. tetrahydro dipicolinate N-succinyl transferase

**Table 2: Subtractive filtering of proteins.**

| Sl. No. | Procedure followed | No. of sequences |
|---|---|---|
| 1. | Sum total number of essential protein sequence. | 298 |
| 2. | Non-paralogous proteins (CD-HIT at 60% threshold word length=3). | 292 |
| 3. | Non-homologous protein sequences (e-value=1e-5). | 147 |
| 4. | BLASTp with DEG database (e-value=1e-5). | 105 |
| 5. | BLASTp with VFDB (e-value=1e-5). | 25 |
| 6. | BLASTp with ARG-ANNOT database (e-value=1e-5). | 2 |

reduces bactericidal activities (Jia *et al.*, 2023). The 'DapD' was also reported as potential drug target in pathogens viz. Klebsiella pneumoniae, Mycobacterium tuberculosis and Neisseria gonorrhoeae (Garg, Tewari, and Raghava, 2010; Jamil, Khan, and Uddin, 2024; Tanwer *et al.*, 2020). Therefore, potent inhibitor i.e. drug compound would only affect pathogen's biological system without inhibiting host's physiological functions (Hema *et al.*, 2015). In addition, most favoured region of Ramachandran plot's displayed 93% amino acid residues which indicated good quality and reliability of modelled 'DapD' protein (Laskowski *et al.*, 1993).

The physicochemical analysis has defined 'DapD' non-homologous protein as stable, electronegative, hydrophilic and soluble (Uddin and Rafi, 2017). Owing to rapid increase in antimicrobial resistance and pathogenesis, Machine Learning (ML), especially deep learning Neural Networks (DNN) and *in silico* subtractive genomics approaches have been extensively employed to identify potential vaccine and therapeutic targets

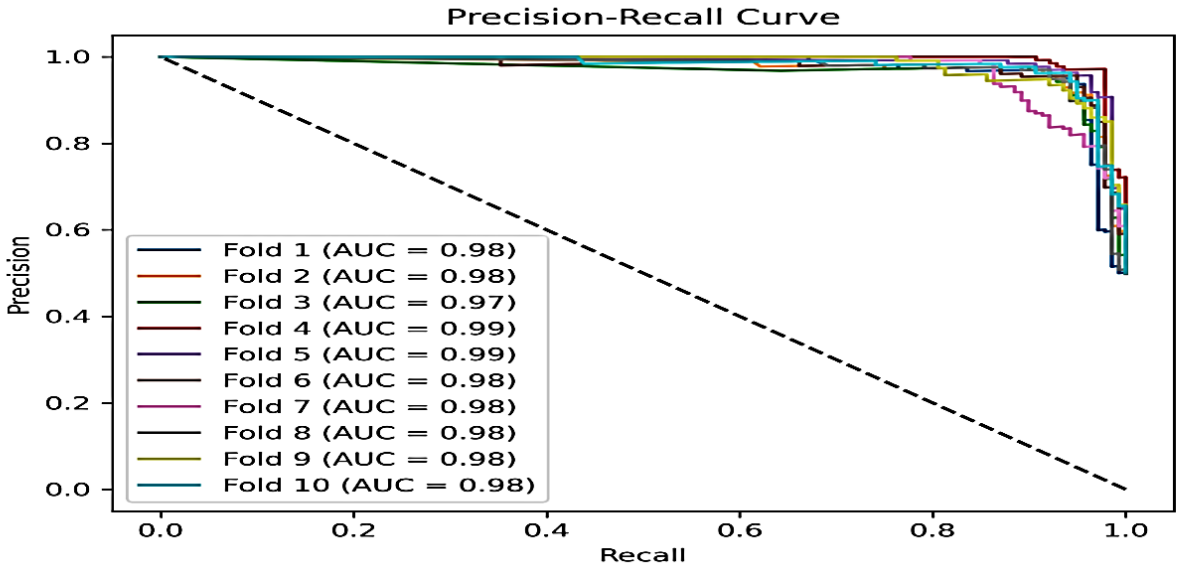


**Figure 4:** Precision Recall curve for 10-fold cross validation.

**Table 3: Physico-chemical properties of target protein sequence (WP_000249839.1) by ProtParam tool.**

| Sl. No. | Properties of target protein | Values |
|---|---|---|
| 1. | Molecular weight | 25257.87 |
| 2. | No. of positively charged residues (Ala+Glu) | 21 |
| 3. | No. of negatively charged residues (Arg+Lys) | 32 |
| 4. | Instability index | 35.72 |
| 5. | Aliphatic index | 100 |
| 6. | Grand average of hydropathicity (GRAVY) | 0.108 |
| 7. | Isoelectric point | 4.71 |

of individual strain (Ashraf *et al.*, 2022; Das and Sarkar, 2022; Uddin *et al.*, 2019). Bioinformatics methods including deep neural networks and subtractive genomics analysis employed in this study, were previously used for identification of potential therapeutic targets against diverse life-threatening pathogens (Goodswen *et al.*, 2021; Hasan and Lonardi, 2020; Hema *et al.*, 2015; Tanwer *et al.*, 2020). The aims of this investigation was to recognize new potential therapeutic targets against a versatile and drug resistant pathogen i.e. MRSA-TCH 32767 strain for development of another new drug candidate. Therefore, further experimental validation is required to be carried out to evaluate this drug target protein. Therefore, presently used *in silico* combined with *in vitro* study as well as preclinical trial is required to be carried out to evaluate this potential drug target protein.

## CONCLUSION

There are enough opportunities for future study on role of critical genomic features and fine-tuning of discriminatory features for upgradation of deep neural network's applicability. Essential genes identified might be useful to gain insight in to pathogenesis of MRSA bacteria. The present study could provide a momentum for future research in finding lead compound as potent drug candidate against MRSA pathogenic bacteria.

## ACKNOWLEDGEMENT

The authors are grateful to JIS College of Engineering, Kalyani, West Bengal, for providing the facilities for this research work.

## AUTHOR CONTRIBUTIONS

Conceptualization, Experimentation, Material preparation, Data collection and analysis was performed by [Monish Mukul Das]. The first draft of the manuscript was written by [Monish Mukul Das] and all authors commented on previous versions of the manuscript. Manuscript review was done by [Sayan Chakraborty], Editing was done by [Avijit Mondal] and Visualization was done by [Tapas Guha]. All authors read and approved the final manuscript.

## ETHICAL APPROVAL

This portion of research is not applicable in present *in silico* work as this research is neither related to clinical data designing nor human experimentation. Moreover, the entire data has been obtained from online databank.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

## ABBREVIATIONS

**MRSA:** Methicillin Resistant *Staphylococcus aureus*; **DNN:** Deep Neural Network; **VFDB:** Virulent Factor Database; **DapD:** Tetrahydrodipicolinate N-Succinyltransferase; **DEG:** Database of Essential Genes; **MSSA:** Methicillin Sensitive *Staphylococcus aureus*; **WGS:** Whole Genome Sequence; **GA:** Genetic Algorithm; **SVM:** Support Vector Machine; **ARG:** Antibiotic Resistance Gene; **ANNOT:** Annotation.

## REFERENCES


Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., & Bambrick, J. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold. Nature, 3, 1–3.

Ashraf, B., Atiq, N., Khan, K., Wadood, A., & Uddin, R. (2022). Subtractive genomics profiling for potential drug targets identification against Moraxella catarrhalis. PLOS One, 17(8), Article e0273252. https://doi.org/10.1371/journal.pone.0273252

Barh, D., Tiwari, S., Jain, N., Ali, A., Santos, A. R., Misra, A. N., Azevedo, V., & Kumar, A. (2011). *In silico* subtractive genomics for target identification in human bacterial pathogens. Drug Development Research, 72(2), 162–177. https://doi.org/10.1002/ddr.20413

Chen, L., Zheng, D., Liu, B., Yang, J., & Jin, Q. (2016). VFDB 2016: Hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Research, 44(D1), D694–D697. https://doi.org/10.1093/nar/gkv1239

Clatworthy, A. E., Pierson, E., & Hung, D. T. (2007). Targeting virulence: A new paradigm for antimicrobial therapy. Nature Chemical Biology, 3(9), 541–548. https://doi.org/10.1038/nchembio.2007.24

Dantes, R., Mu, Y., Belflower, R., Aragon, D., Dumyati, G., Harrison, L. H., Lessa, F. C., Lynfield, R., Nadle, J., Petit, S., Ray, S. M., Schaffner, W., Townes, J., Fridkin, S., & Emerging Infections Program–Active Bacterial Core Surveillance MRSA Surveillance Investigators. (2013). National burden of invasive methicillin-resistant Staphylococcus aureus infections, United States, 2011. JAMA Internal Medicine, 173(21), 1970–1978. https://doi.org/10.1001/jamainternmed.2013.10423

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Garg, A., Tewari, R., & Raghava, G. P. S. (2010). Virtual Screening of potential drug-like inhibitors against lysine/DAP pathway of Mycobacterium tuberculosis. BMC Bioinformatics, 11(Suppl. 1), S53. https://doi.org/10.1186/1471-2105-11-S1-S53

Gasteiger, E. (2005). Protein identification and analysis tools on the Expasy server. In The proteomics protocols handbook. Human Press, Incorporated.

Goodswen, S. J., Barratt, J. L. N., Kennedy, P. J., Kaufer, A., Calarco, L., & Ellis, J. T. (2021). Machine learning and applications in microbiology. FEMS Microbiology Reviews, 45(5), Article fuab015. https://doi.org/10.1093/femsre/fuab015

Gordon, R. J., & Lowy, F. D. (2008). Pathogenesis of methicillin-resistant Staphylococcus aureus infection. Clinical Infectious Diseases, 46(Suppl. 5), S350–S359. https://doi.org/10.1086/533591

Hasan, M. A., & Lonardi, S. (2020). DeeplyEssential: A deep neural network for predicting essential genes in microbes. BMC Bioinformatics, 21(Suppl. 14), 367. https ://doi.org/10.1186/s12859-020-03688-y

Hema, K., Priyadarshini, V. I., Pradhan, D., Munikumar, M., Sandeep, S., Pradeep, N., Suchitra, M., & Umamaheswari, A. (2015). Identification of putative drug targets and vaccine candidates for pathogens causing atherosclerosis. Biochemistry and Analytical Biochemistry, 4(2), 1. https://doi.org/10.4172/2161-1009.1000175

Hosen, M. I., Tanmoy, A. M., Mahbuba, D.-A., Salma, U., Nazim, M., Islam, M. T., & Akhteruzzaman, S. (2014). Application of a subtractive genomics approach for *in silico* identification and characterization of novel drug targets in Mycobacterium tuberculosis F11. Interdisciplinary Sciences, Computational Life Sciences, 6(1), 48–56. https://doi.org/10.1007/s12539-014-0188-y

Hossain, M., Chowdhury, D. U. S., Farhana, J., Akbar, M. T., Chakraborty, A., Islam, S., & Mannan, A. (2013). Identification of potential targets in Staphylococcus aureus N315 using computer aided protein data analysis. Bioinformation, 9(4), 187–192. https://doi.org/10.6026/97320630009187

Jamil, F., Khan, K., Zahra, N. A., & Uddin, R. (2024). Lysine metabolism pathway as a target for drug repurposing: *In silico* approach against carbapenem-resistant Klebsiella pneumoniae. Gene Reports, 37, Article 102028. https://doi.org/10.1016/j.genrep.2024.102028

Jenul, C., & Horswill, A. R. (2019). Regulation of Staphylococcus aureus virulence. Microbiology Spectrum, 7(2). https://doi.org/10.1128/microbiolspec. gpp1123-0031-2018.

Jia, J., Zheng, M., Zhang, C., Li, B., Lu, C., Bai, Y., Tong, Q., Hang, X., Ge, Y., Zeng, L., Zhao, M., Song, F., Zhang, H., Zhang, L., Hong, K., & Bi, H. (2023). Killing of Staphylococcus aureus persisters by a multitarget natural product chrysomycin A. Science Advances, 9(31), Article eadg5995. https://doi.org/10.1126/sciadv.adg5995

Kaur, D. C., & Chate, S. S. (2015). Study of antibiotic resistance pattern in methicillin resistant Staphylococcus aureus with special reference to newer antibiotic. Journal of Global Infectious Diseases, 7(2), 78–84. https://doi.org/10.4103/0974-777X.157245

Kaur, H., Kalia, M., & Taneja, N. (2021). Identification of novel non-homologous drug targets against Acinetobacter baumannii using subtractive genomics and comparative metabolic pathway analysis. Microbial Pathogenesis, 152, Article 104608. https://doi.org/10.1016/j.micpath.2020.104608

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26(2), 283–291. https://doi.org/10.1107/S0021889892009944

Lindsay, J. A. (2013). Hospital-associated MRSA and antibiotic resistance-what have we learned from genomics? International Journal of Medical Microbiology, 303 (6–7), 318–323. https://doi.org/10.1016/j.ijmm.2013.02.005

McNeil, J. C., Sommer, L. M., Vallejo, J. G., Hulten, K. G., Kaplan, S. L., & Flores, A. R. (2022). Reduced ceftaroline susceptibility among invasive MRSA infections in children: A clinical and genomic investigation. Antimicrobial Agents and Chemotherapy, 66(10), Article e0074522. https://doi.org/10.1128/aac.00745-22

Mickymaray, S., Alturaiki, W., Al-Aboody, M. S., Mariappan, P., Rajenderan, V., Alsagaby, S. A., Kalyanasundram, U., & Alarfajj, A. A. (2018). Antibacterial efficacy of bacteriocin produced by marine Bacillus subtilis against clinically important extended spectrum beta-lactamase strains and methicillin-resistant Staphylococcus aureus. International Journal of Medical Research and Health Sciences, 7(2), 75–83.

Miesel, L., Greene, J., & Black, T. A. (2003). Genetic strategies for antibacterial drug discovery. Nature Reviews. Genetics, 4(6), 442–456. https://doi.org/10.1038/nrg1086

Mukul Das, M. M., & Sarkar, K. (2022). Evaluation of machine learning classifiers for predicting essential genes in Mycobacterium tuberculosis strains. Bioinformation, 18(12), 1126–1130. https://doi.org/10.6026/973206300181126

Nandhini, P., Kumar, P., Mickymaray, S., Alothaim, A. S., Somasundaram, J., & Rajan, M. (2022). Recent developments in methicillin-resistant Staphylococcus aureus (MRSA) treatment: A review. Antibiotics, 11(5), 606. https://doi.org/10.3390/antibiotics11050606

Ochoa, S. A., Cruz-Córdova, A., Mancilla-Rojano, J., Escalona-Venegas, G., Esteban-Kenel, V., Franco-Hernández, I., Parra-Ortega, I., Arellano-Galindo, J., Hernández-Castro, R., Perez-López, C. F., De la Rosa-Zamboni, D., & Xicohtencatl-Cortes, J. (2020). Control of methicillin-resistant Staphylococcus aureus strains associated with a hospital outbreak involving contamination from anesthesia equipment using UV-C. Frontiers in Microbiology, 11, Article 600093. https://doi.org/10.3389/fmicb.2020.600093

Ogston, A. (1881). Report upon micro-organisms in surgical diseases. British Medical Journal, 1(1054), 369. b362. https://doi.org/10.1136/bmj.1.1054.369

Palaniappan, K., & Mukherjee, S. (2011). Predicting essential" genes across microbial genomes: A machine learning approach. 2011 10th International Conference on Machine Learning and Applications and Workshops.

Plaimas, K., Eils, R., & König, R. (2010). Identifying essential genes in bacterial metabolic networks with machine learning methods. BMC Systems Biology, 4, 56. https://doi.org/10.1186/1752-0509-4-56

Sangeetha, M., Saranya, T. S., Sathianarayanan, S., Hima Vyshnavi, A. M., & Krishnan Namboori, P. K. (2020). Design and development of potential flavonoid moiety for PBP2a inhibition for MRSA therapy-a computational technique. Biomedical and Pharmacology Journal, 13(2), 687–692. https://doi.org/10.13005/bpj/1933

Schito, G. C. (2006). The importance of the development of antibiotic resistance in Staphylococcus aureus. Clinical Microbiology and Infection, 12(Suppl. 1), 3–8. https://doi.org/10.1111/j.1469-0691.2006.01343.x

Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F., & Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Research, 33(4), 1141–1153. https://doi.org/10.1093/nar/gki242

Shoukat, K., Rasheed, N., & Sajid, M. (2012). Subtractive genome analysis for *in silico* identification and characterization of novel drug targets in C. trachomatis strain D/UW-3/Cx. International Journal of Current Research, 4(6).

Skariyachan, S., Krishnan, R. S., Siddapa, S. B., Salian, C., Bora, P., & Sebastian, D. (2011). Computer aided screening and evaluation of herbal therapeutics against MRSA infections. Bioinformation, 7(5), 222–233. https://doi.org/10.6026/97320630007222

Tanwer, P., Kolora, S. R. R., Babbar, A., Saluja, D., & Chaudhry, U. (2020). Identification of potential therapeutic targets in Neisseria gonorrhoeae by an *in silico* approach. Journal of Theoretical Biology, 490, Article 110172. https://doi.org/10.1016/j.jtbi.2020.110172

Taylor, T. A., & Unakal, C. G. (2021). Staphylococcus aureus. StatPearls [Internet].

Turner, N. A., Sharma-Kuinkel, B. K., Maskarinec, S. A., Eichenberger, E. M., Shah, P. P., Carugati, M., Holland, T. L., & Fowler, Jr., V. G. (2019). Methicillin-resistant Staphylococcus aureus: An overview of basic and clinical research. Nature Reviews. Microbiology, 17(4), 203–218. https://doi.org/10.1038/s41579-018-0147-4

Uddin, R., & Rafi, S. (2017). Structural and functional characterization of a unique hypothetical protein (WP_003901628.1) of Mycobacterium tuberculosis: A

computational approach. Medicinal Chemistry Research, 26(5), 1029–1041. https://doi.org/10.1007/s00044-017-1822-0

Uddin, R., Siddiqui, Q. N., Sufian, M., Azam, S. S., & Wadood, A. (2019). Proteome-wide subtractive approach to prioritize a hypothetical protein of XDR-Mycobacterium tuberculosis as potential drug target. Genes and Genomics, 41(11), 1281–1292. https://doi.org/10.1007/s13258-019-00857-z

Van Heijenoort, J. (2001). Recent advances in the formation of the bacterial peptidoglycan monomer unit. Natural Product Reports, 18(5), 503–519. https://doi.org/10.1039/a804532a

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 26(13), 1608–1615. https://doi.org/10.1093/bioinformatics/btq249.

**Cite this article:** Das MM, Chakraborty S, Mondal A, Guha T. Identification of Essential Proteins as Promising Therapeutic Targets against Drug Resistant Strain of *Staphylococcus aureus*: An *in silico* Approach. Int. J. Pharm. Investigation. 2025;15(4):1227-34.

**Figure S1:** Predicted protein structure using Alphafold 3.

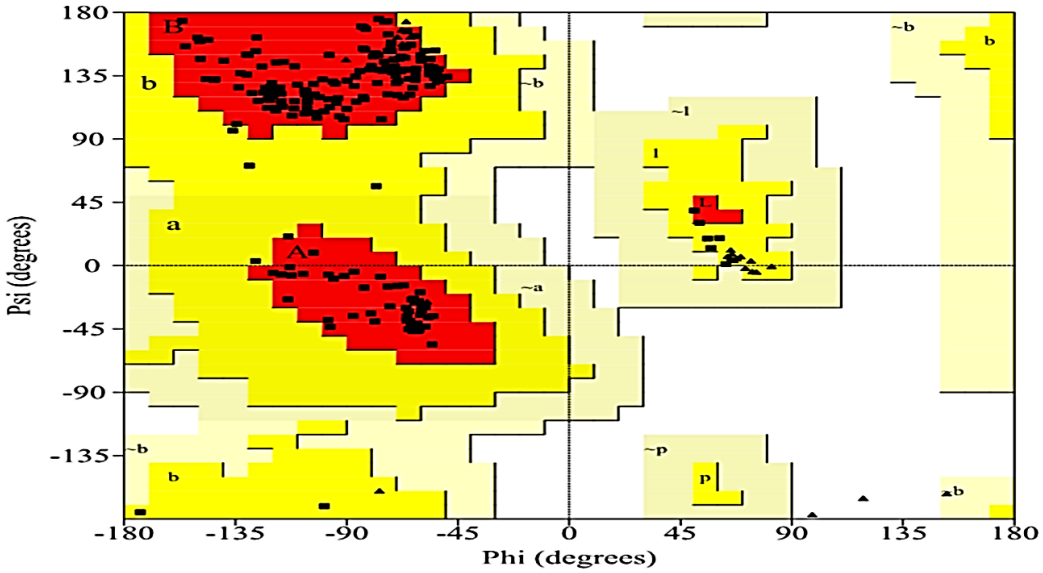


**Figure S2:** PROCHECK generated, Ramachandran plot showing evaluation of 'Dap D' modelled protein.