Journal of Sustainability Science and Management
Volume 17 Number 2, February 2022: 18-26
eISSN: 2672-7226

IN SILICO IDENTIFICATION OF ANTIGENIC PROTEINS IN *Staphylococcus aureus*

NAWAL ZULKIPLY[1], MUHAMMAD EMIR RAMLI[1] AND MOHD FAKHARUL ZAMAN RAJA YAHYA[1,2]*

[1]Faculty of Applied Sciences, Universiti Teknologi MARA Shah Alam, 40450 Shah Alam, Malaysia. [2]Molecular Microbial Pathogenicity Research Group, Pharmaceutical and Life Sciences Community of Research, Universiti Teknologi MARA.

**Corresponding author: fakharulzaman@uitm.edu.my*
Submitted final draft: 10 June 2021 *Accepted: 17 June 2021* http://doi.org/10.46754/jssm.2022.02.002

Abstract: *Staphylococcus aureus,* a Gram-positive bacterium is recognized as an opportunistic pathogen in humans and livestock. Whole-cell proteome expression in *S. aureus* has previously been elucidated, however, antigenicity of *S. aureus* proteins has not been well investigated. The present work was performed to identify antigenic proteins expressed in *S. aureus* using *in silico* approach. The proteome information of *S. aureus* was retrieved from World-2DPAGE Repository. A total of 657 protein sequences of *S. aureus* were then downloaded from UniprotKB in FASTA format and were used as queries in VaxiJen, CELLO2GO, DEG, BLASTp, STRING and SWISS-MODEL programmes. Results demonstrated that 63% of *S. aureus* proteins were predicted as antigenic proteins. Majority of them were found to associate with catalytic activity and metabolic process. The antigenic *S. aureus* proteins, such as 50S ribosomal protein L21 and an uncharacterized protein, were identified as cytoplasmic proteins, essential for survival of *S. aureus,* non-host homologous and hub proteins in the protein interaction network. Homology modelling of 50S ribosomal protein L21 and uncharacterized protein yielded good models based on related structures from the Protein Data Bank. The findings of the present study suggest the potential use of the identified antigenic proteins in vaccine strategy against *S. aureus* infections.

Keywords: *Staphylococcus aureus,* antigenic, vaccine, *in silico.*

Introduction

Staphylococcus aureus is a Gram-positive, extracellular bacterium responsible for significant morbidity and mortality worldwide. It is a member of the normal human flora and often colonizes the skin and the upper respiratory tract. *S. aureus* expresses a broad range of virulence factors that include surface proteins covalently attached to the cell wall and secreted proteins expressed during infection. These exposed proteins are essential for the survival and proliferation of *S. aureus*, and the presence of many of them is conserved. Depending on the stage of infection and physiological conditions inside the host, virulence factor expression promotes binding to the extracellular matrix, colonization, invasion, and avoidance of the immune response (Misra *et al.*, 2018).

Structures of antigenic proteins of *S. aureus* enable them to be bound to either receptor of antigen-specific antibodies or β-cell antigen. Within the human body, the presence of these antigens activates immune responses. Longheu *et al.* (2020) identified secreted and cellular antigens of *S. aureus* which cause dairy sheep mastitis and their potential for production of vaccine. They found that a solid humoral host immune response was elicited by several *S. aureus* antigens including household proteins pyruvate kinase, elongation Factor Tu, dihydrolipoyl dehydrogenase, alpha-keto acid dehydrogenase and bifunctional autolysin. Implementing or incorporating these antigens into a vaccine product or program might help the host to monitor infections of *S. aureus* faster and more efficiently.

Bioinformatics is a major advantage of scientific research, especially in the field of biomedicine. It links biological data with techniques for information storage, distribution, and analysis to support multiple areas of scientific research. The use of bioinformatic approach in the identification of drug and vaccine targets has recently gained a great deal of attention. In 2015, Delfani *et al.* performed *in silico* identification of potential vaccine candidates against *S. aureus*. They designed synthetic genes encoding the *clfA*, *isdB*, and *hlg* and successfully predicted their 3D structure. Recently, Shahid *et al.* (2020) studied potential drug targets in *Staphylococcus saprophyticus* using *in silico* proteomic approach and identified four cytoplasmic proteins as drug target candidates namely UDP-N-acetylenolpyruvoylglucosamine reductase, UDP-N-acetylmuramoylalanineD-glutamate ligase, D-alanine-D-alanine ligase and alanine racemase. In addition, whole-cell proteome expression in *S. aureus* has previously been elucidated using two -dimensional polyacrylamide gel electrophoresis. However, antigenic proteins from that proteome remains not well investigated. The present work was performed to identify antigenic proteins expressed in *S. aureus* using *in silico* approach.

Materials and Methods

Proteome Retrieval

The gel-based proteome information and list of *S. aureus* (strain Mu50/ATCC 700699, Taxon id: 158878) proteins were retrieved from World-2DPAGE Repository. A total of 657 FASTA sequences of *S. aureus* proteins were then downloaded from UniprotKB.

Analysis of the Protein Antigenicity

The protein sequences were used as queries in VaxiJen software for prediction of antigenic proteins. The threshold value was 0.4. Functional classification of identified antigenic proteins was performed based on SwissProt/TrEMBL database.

Analysis of Subcellular Localization

The protein sequences were used as queries in CELLO2GO for prediction of subcellular localization whether in periplasm, cytoplasm or membrane. The E-value of search was set to 1e-06.

Analysis of Protein Essentiality

The protein sequences were used as queries in DEG for the prediction of proteins essential for the survival of the organism. The E-value of search was set to 1e-06.

Sequence Similarity Search

The protein sequences were used as queries in BLASTp for identification of non-host homologous proteins. The parameters of search against *Homo sapiens* proteome were set to E-value > 1e-06 and sequence identity < 30%.

Prediction of B Cell Epitope

The protein sequences were used as queries in Immune Epitope Database for prediction of B cell epitope. The search used Bepipred Linear Epitope Prediction 2.0 method.

Analysis of Protein-Protein Interaction Network

The protein sequences were used as queries in the STRING database for prediction of protein interaction network. The confidence limit for the analysis of protein interaction networks was set to high confidence level (0.7).

Analysis of 3D Structure of Protein

Homology modelling was used to determine the 3D structure of protein. A BLASTp search with default parameters was performed against Protein Data Bank (PDB) to find suitable templates for homology modelling. SWISS-MODEL was used for homology model construction. Structural evaluation and stereochemical analyses were performed using PROCHECK.

Results and Discussion

Antigenic Proteins

Figure 1 shows the antigenic proteins expressed in *S. aureus*. A total of 414 (63%) *S. aureus* proteins were predicted to be antigenic proteins. Majority of them were found to be associated with catalytic activity (52%) and metabolic process (45%). Their antigenicity score ranged between 0.4 and 0.99. The *S. aureus* proteins with antigenicity score greater than 0.9 were selected for further analysis. The selected proteins were 50S ribosomal protein L21, uncharacterized protein and organic hydroperoxide resistance protein-like.

Subcellular Localization, Essentiality and Non-host Homology

Table 1 shows the selected antigenic proteins for further analyses. 50S ribosomal protein L21, uncharacterized protein and organic hydroperoxide resistance protein-like were analyzed for subcellular localization, essentiality and non-host homology. All the selected antigenic proteins were predicted to be localized in cytoplasm, essential to the survival of *S. aureus* and non-host homologous.

B-cell Epitopes

Table 2 shows B-cell epitopes of the selected antigenic proteins expressed in *S. aureus*. All the selected antigenic proteins expressed in *S.*

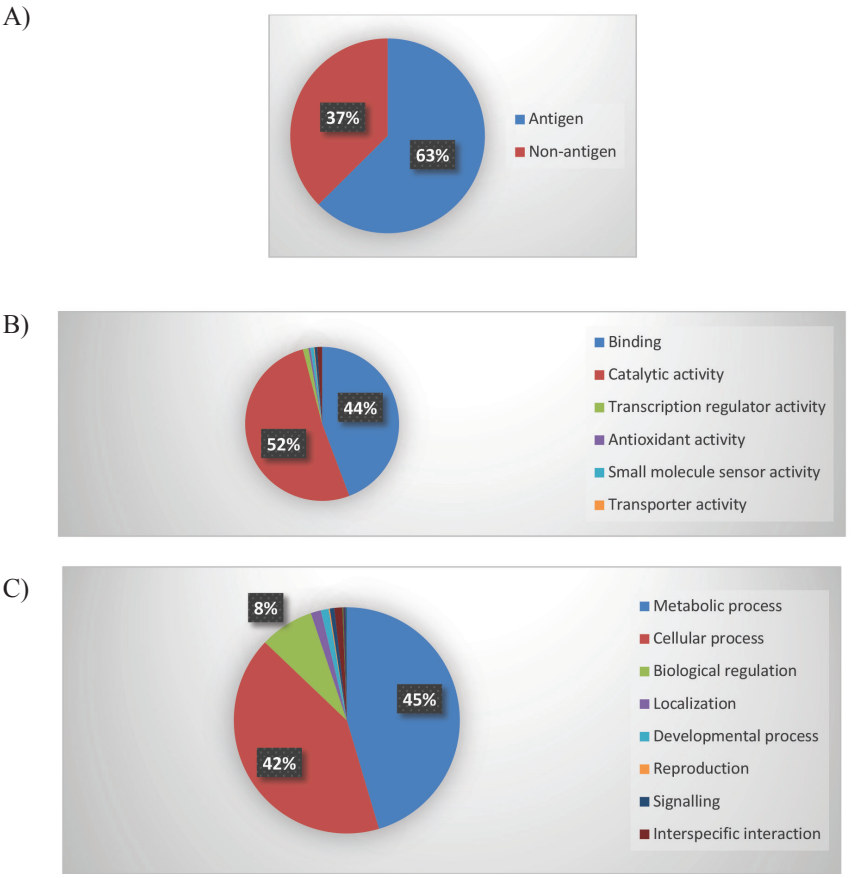

Figure 1: Distribution of antigenic proteins expressed in *S. aureus*. A) Percentage of antigenic proteins; B) Functional classification of antigenic proteins based on molecular function; C) Functional classification of antigenic proteins based on biological process

Table 1: Functional characteristics of the selected antigenic proteins expressed in *S. aureus*

ID	Proteins	Vaxijen	Cello	Deg	Blastp
Q99TK6	50S ribosomal protein L21	0.9318	Cytoplasmic	Essential	Non -host homologous
A0A0H3JT87	Uncharacterized protein	0.9372	Cytoplasmic	Essential	Non -host homologous
Q99VH8	Organic hydroperoxide resistance protein-like	0.9913	Cytoplasmic	Essential	Non -host homologous

Table 2: Prediction of B-cell epitopes by Immune Epitope Database

Proteins	Start	End	Epitopes
50S ribosomal protein L21	21	32	FVEKLDVNEGDT
	65	90	QGRGKKITVFTYKRRKNSKRKKGHRQ
Uncharacterized protein	22	57	DTPKDETKSTESNTNQDTNTTKDVIALKDVKTSPED
	105	138	KKVINKKTEKEDTVNENDNFKYSDAIDYKKAIKE
	143	155	FDGDIKEWSLEKD
Organic hydroperoxide resistance protein-like	24	41	RALDIDIVPPAQADGKAT
	102	115	NVISQEEAEKYLQM

aureus were predicted to contain several B-cell epitopes. Only epitope sequences with at least 10 amino residues are presented here.

Protein Interaction Networks

50S ribosomal protein L21 and uncharacterized proteins showed more than 10 functional linkages in the protein interaction network. They were classified as hub proteins in *S. aureus*. On the other hand, organic hydroperoxide resistance protein-like showed only five functional linkages in different sizes of protein interaction network. It was not classified as a hub protein in *S. aureus*.

3D Models

Templates 6ddg.1.D, 4exr.1.A and 6mjn.1.A were used to construct the 3D model of 50S ribosomal protein L21, uncharacterized protein and organic hydroperoxide resistance protein-like respectively. The sequence similarity between target and template for 50S ribosomal protein L21, uncharacterized protein and organic hydroperoxide resistance protein-like were 100%, 27.7% and 43.38% respectively. The 3D models of 50S ribosomal protein L21, uncharacterized protein and organic hydroperoxide resistance protein-like showed Ramachandran-favoured area of 94.9%, 97.24% and 97.42% respectively.

Vaccination is known to effectively control a wide range of diseases in human and veterinary health care. The immune system recognizes antigens as foreign substances, destroys them, and subsequently memorizes them. The ability to sequence the whole genome of a pathogenic microorganism has enabled *in silico screening* for the most probable protective antigens before performing confirmatory experiments. This approach which is known as reverse vaccinology offers several advantages such as high speed and low cost. VaxiJen represents the first *in silico* tool for alignment-independent prediction of protective antigens. It enables antigen classification solely based on the physicochemical properties of proteins without recourse to sequence alignment. On the other hand, Immune Epitope Database predicts B-cell epitope, the localized area of the antigen that binds to a specific antigen receptor

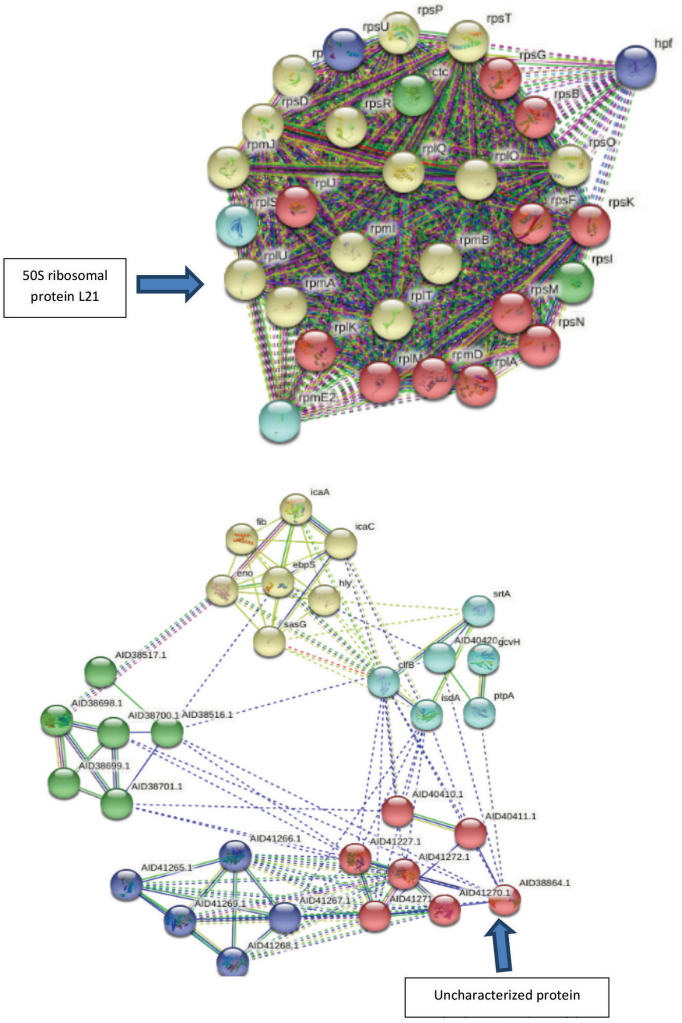

Figure 2**:** Identification of the selected antigenic proteins as hub proteins; Upper panel: functional linkages mediated by 50S ribosomal protein L21; Lower panel: functional linkages mediated by uncharacterized protein

on the surface of a B-cell, based on sequence characteristics of the antigen using amino acid scales and hidden markov models. The present study used *in silico* approach to predict the antigenicity and B cell epitopes of *S. aeureus* proteins. The similar approach was used to study the antigenicity of capsid protein of human bocavirus 1 (Kalyanaraman, 2018).

Antigenic composition differs among bacterial species. In experiment, the antigenic composition can be studied using isoelectrophoretic separation of protein mixtures followed by electrophoretic separation in a gel containing antibody. This combined approach offers high-resolution identification of various antigens, genetic variants of single antigens, and isoenzymes. In the present study, *in silico* approach successfully identified a high antigenic composition in *S. aureus* proteome. The use of multiple antigenic proteins in vaccine development provides greater immunogenic potential and protective immunity (Delfani *et al.,* 2015; Gatkowska *et al*., 2019).

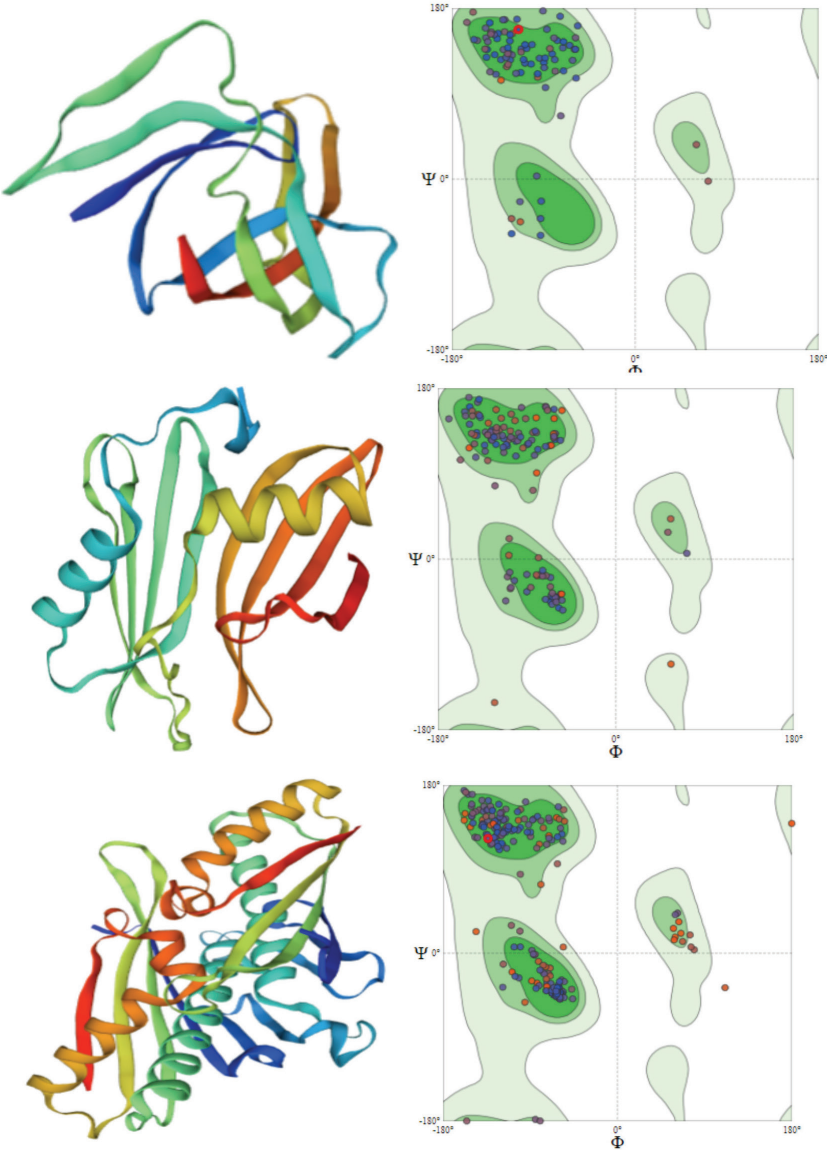

Figure 3: 3D models and their Ramachandran plots. Top: 50S ribosomal protein L21; middle: Uncharacterized protein; bottom: Organic hydroperoxide resistance protein-like

Herein, the majority of antigenic proteins of *S. aureus* were associated with catalytic activities and metabolic processes. For example, UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase plays an important role in the peptidoglycan synthesis in *S. aureus* (Zaveri & Patnala, 2016a). This result is also in agreement with an immunoproteomic study showing that the immunogenic proteins of *Streptococcus pneumoniae* was involved in glycolytic pathway (Ling *et al.,* 2004). In *Strongyloides stercoralis*, the antigenic proteins are known to play roles in metabolic process, energy generation, and oxidation-reduction, cell division, cell signaling and transportation, and regulation of muscular contraction (Rodpai *et al.,* 2017).

Protein subcellular localization is a key functional attribute. Prediction of subcellular localization of unknown proteins is useful to obtain information about their functions, enhance understanding of the disease mechanisms and assist the drug development. The basis of prediction strategy for bacterial proteins is based on N-terminal signal peptides and transmembrane segments. In the present study, 50S ribosomal protein L21 and an uncharacterized protein were predicted to associate with cytoplasm. This result contradicts Misra *et al*. (2018) demonstrating antigenic membrane proteins of *S. aureus* expressed in bovine milk during mastitis. However, intracellular and cytoplasmic proteins, such as elongation factors and ribosomal proteins may also function as potential vaccine targets due to their high abundance (Foulston *et al.,* 2014). The essential, non-host homologous, antigenic and virulent proteins of *S. aureus* that are associated with the cytoplasm have previously been reported (Zaveri & Patnala, 2016b).

Essential proteins are indispensable proteins for microorganisms to grow and survive under certain environment. Nowadays, many works on computational prediction of essential proteins involve sequence similarity search and analysis of protein interaction network. In the present study, 50S ribosomal protein L21 and an uncharacterized protein were predicted to be essential to the survival of *S. aureus* and also non-homologous to human host. According to Vogeley *et al.* (2016), essential and non-host homologous bacterial proteins represent the potential vaccine targets. The essential and non-host homologous proteins have also been widely investigated in the biofilm-forming bacteria (Othman & Yahya, 2019).

Prediction of protein interaction using STRING database is based on neighbourhood, fusion-fission events, occurrence, co-expression, text mining and data imported from public databases of physical interactions. The large-scale analysis of protein interaction network is important for understanding the organizational and functional properties of individual proteins. The number of interactions that a protein has in the network has been linked to its indispensability. Essential proteins generally have more interactions than the nonessential ones. Herein, 50S ribosomal protein L21 and an uncharacterized protein were predicted to be hub proteins in *S. aureus* because they showed more than 10 functional linkages in the protein interaction network. Targets with high number of functional linkages are considered as hub proteins and are important in drug and vaccine development (Kitano, 2002; Yahya *et al.,* 2017).

Homology modeling predicts the 3D structure of a query protein through the sequence alignment of template proteins. In general, the process of homology modeling involves four steps namely target identification, sequence alignment, model building and model refinement. In the present study, the 3D model of uncharacterized proteins showed high quality and was probably a lipoprotein. The role of lipoprotein as a vaccine target has previously been reported (Chong *et al.,* 2015). Meanwhile, the high-quality 3D model of antigenic proteins may be useful for designing immunotherapy against *S. aureus* (Tarek *et al.,* 2018).

Conclusion

The present study predicts a high composition of antigenic proteins expressed in *S. aureus.* The selected antigenic proteins fulfilled the common criteria of vaccine targets namely the presence of B-cell epitope, essentiality and non-host homology. Interestingly, two of them showed high number of functional linkages in the protein interaction network, making them valuable vaccine targets. The findings presented herein provide important information for the development of potential vaccine against *S. aureus* pathogen. The vaccines developed this way are specific to *S. aureus*, the occurrence of resistance and toxicity for the host could be minimized.

Acknowledgements

This research was funded by the Malaysian Ministry of Higher Education under Fundamental Research Grant Scheme, FRGS/1/2019/STG04/UITM/02/2.

References

Chong, P., Huang, J. H., Leng, C. H., Liu, S. J., & Chen, H. W. (2015). Recombinant lipoproteins as novel vaccines with intrinsic adjuvant. *Advances in Protein Chemistry and Structural Biology, 99,* 55-74.

Delfani, S., Imani Fooladi, A. A., Mobarez, A. M., Emaneini, M., Amani, J., & Sedighian, H. (2015). *In silico* analysis for identifying potential vaccine candidates against *Staphylococcus aureus. Clinical and Experimental Vaccine Research, 4*(1), 99-106.

Foulston, L., Elsholz, A. K., DeFrancesco, A. S., & Losick, R. (2014). The extracellular matrix of *Staphylococcus aureus* biofilms comprises cytoplasmic proteins that associate with the cell surface in response to decreasing pH. *MBio*, *5*(5), e01667-14.

Gatkowska, J., Dzitko, K., Ferra, B. T., Holec-Gąsior, L., Kawka, M., & Dziadek, B. (2019). The impact of the antigenic composition of chimeric proteins on their immunoprotective activity against chronic toxoplasmosis in mice. *Vaccines*, *7*(4), 154.

Kalyanaraman, N. (2018). *In silico* prediction of potential vaccine candidates on capsid protein of human bocavirus 1. *Molecular Immunology*, *93*, 193-205.

Kitano, H. (2002). Systems biology: A brief overview. *Science*, *295*(5560), 1662-1664.

Ling E., Feldman, G., Portnoi M., Dagan, R., Overweg, K., Mulholland, F., Chalifa-Caspi, V., Wells, J., & Mizrachi-Nebenzahl, Y. (2004). Glycolytic enzymes associated with the cell surface of *Streptococcus pneumoniae* are antigenic in humans and elicit protective immune responses in the mouse. *Clinical and Experimental Immunology*, *138*(2), 290-298.

Longheu, C. M., Azara, E., Marogna, G., Addis, M. F., & Tola, S. (2020). Identification of secreted and cellular antigens of *Staphylococcus aureus* causing dairy sheep mastitis and their potential for vaccine development. *Veterinary Immunology and Immunopathology*, *230*, 110149.

Misra, N., Pu, X., Holt, D. N., McGuire, M. A., & Tinker, J. K. (2018). Immunoproteomics to identify *Staphylococcus aureus* antigens expressed in bovine milk during mastitis. *Journal of Dairy Science*, *101*(7), 6296-6309.

Othman, N. A., & Yahya, M. F. Z. R. (2019). *In silico* analysis of essential and non-homologous proteins in *Salmonella typhimurium* biofilm. *Journal of Physics: Conference Series,* 1349 012133.

Rodpai, R., Intapan, P. M., Thanchomnang, T., Sanpool, O., Janwan, P., Laummaunwai, P., Wongkham, C., Insawang, T., & Maleewong, W. (2017). Identification of antigenic proteins in *Strongyloides stercoralis* by proteomic analysis. *Parasitology Research*, *116*(6), 1687-1693.

Shahid, F., Ashfaq, U. A., Saeed, S., Munir, S., Almatroudi, A., & Khurshid, M. (2020). *In silico s*ubtractive proteomics approach for identification of potential drug targets in *Staphylococcus saprophyticus*. *International Journal of Environmental Research and Public Health*, *17*(10), 3644.

Tarek, M. M., Shafei, A. E., Ali, M. A., & Mansour, M. M. (2018). Computational prediction of vaccine potential epitopes and 3-dimensional structure of XAGE-1b for non-small cell lung cancer immunotherapy. *Biomedical Journal*, *41*(2), 118-128.

Vogeley, L., El Arnaout, T., Bailey, J., Stansfeld, P. J., Boland, C., & Caffrey, M. (2016). Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin. *Science*, *351*(6275), 876-880.

Yahya, M. F. Z. R., Hamid, U. M. A., Norfatimah, M. Y., & Kambol, R. (2014). *In silico* analysis of essential tricarboxylic cycle enzymes from biofilm-forming bacteria. *Trends in Bioinformatics, 7*(1), 19-26.

Yahya, M. F. Z. R., Alias, Z., & Karsani, S. A. (2017). Subtractive protein profiling of *Salmonella typhimurium* biofilm treated with DMSO, *Protein J, 36,* 286-298.

Zaveri, K., & Patnala, K. (2016a). Prediction of promiscuous epitope studies of spa antigen in *Staphylococcus aureus*: An insight on the peptide-based vaccine. *International Journal of Pharmacy and Pharmaceutical Sciences*, *8*(7), 386-391.

Zaveri, K., & Patnala, K. (2016b). Screening of putative therapeutics candidates in superbug (*Staphylococcus aureus*): A systematic *in silico* approach. *Asian Journal of Pharmaceutical and Clinical Research, 9,* 1-9.