OriginalArticle

Mining of novel drug/vaccine targets from the proteome of Staphylococcus aureus using computational tools through reverse vaccinology approach

Vaishnavi Modupalli[1], John Arun[2*]

[1]Research Scholar,Department of Bioinformatics,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai,TamilNadu, India, Pincode:602105
[2]Project Guide, Corresponding author,Department of Bioinformatics,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, TamilNadu, India, Pincode: 602105

Abstract

Aim: The present study aims to investigate the cluster of unexplored proteins of *Staphylococcus aureus*using a series of bioinformatics tools to fetch novel drug/vaccine candidates. *Staphylococcus aureus*, a gram-positive multidrug resistant bacteria which is easily transmitting nosocomial infections. So far, no effective vaccine exists to prevent this pathogenic infection.**Materials and Methods:** A total of 50 uncharacterized protein sequences of *Staphylococcus aureus* were retrieved from NCBI and analyzed using computational tools for studying their localization, membrane helices, physicochemical properties, virulence factors, signal peptides, antigenicity, and epitopes. These proteins were then subjected to tBLASTn to compare against human proteome for confirming that they are not human homologs in order to bypass autoimmune reactions. **Results**: Two potential candidate proteins possessing virulence and antigenic properties, comprising epitopes and are not human homologs were found in this study. **Conclusion**: Hence, these proteins could be ideal drug/vaccine targets, however, further in-depth immuno-informatics and structural biology approaches are recommended with *in-vitro* and *in-vivo* experiments for validation.

Keywords: Immunoinformatics, Reverse Vaccinology, Novel Drug Targets, *Staphylococcus aureus*, Proteome, Virulence Factors

DOI: 10.47750/pnr.2022.13.S03.015

INTRODUCTION

Staphylococcus aureus, a gram-positive bacteria can cause inflammatory diseases, including skin infections, pneumonia, endocarditis, septic arthritis, osteomyelitis and abscesses (Sizar and Unakal 2021). It is both a commensal bacterium and a human pathogen. Approximately 30% of the human population is infected with *S. aureus*, it is one of the leading causes of bacteremia and infective endocarditis (Tong et al. 2015). The virulence factors released by bacteria enables it to colonize different environmental niches and start infecting. Skin and soft tissues are the primary sites of infection (Lung and Prince 2020). *S.aureus*is naturally resistant to every antibiotic, it acquired resistance by horizontal transfer to genes, chromosomal mutations. Infections caused by *S.aureus* have reached epidemic proportions globally.

Antibiotic resistance patterns were identified for 106 strains of methicillin-resistant staphylococcus aureus (MSRA). However, there is no vaccine discovered so far, new technologies leading to a better understanding of pathogenesis of staphylococcus is necessary for prevention and treatment of infection (Chambers and DeLeo 2009). This initiates a need for the development of alternative antimicrobial approaches to fight against multidrug-resistant *S.aureus(Maple, Hamilton-Miller, and Brumfitt 1989)*; (Mathers 2008); (Beeson et al. 2019); (Mota and Rodriguez 2017); (Böhme et al. 2019); (Tang et al. 2019). Genomic era brought new approaches, without the need of growing pathogens for designing the vaccines. One of those new approaches is Reverse Vaccinology, which relies on identifying the proteins with vaccine potentials (i.e; surface proteins or secreted proteins, prediction of T cell, B cell immune epitopes) using bioinformatic tools (Rappuoli et al. 2016). Therefore, finding novel drug targets with antigenic potential is a prerequisite for vaccine development. More than 610 research reports related to *in-silico* characterization of uncharacterized proteins of various microbes towards identification of novel drug targets were found in PubMed in the past 5 years. The most recent study was (Aguttu et al., 2021), where the hypothetical proteins of the *Plasmodium falciparum* were characterized by

computational tools to find vaccine targets.Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021)

However, the research gap identified here is that nearly one fourth of the proteome of *Staphylococcus aureus*is annotated as hypothetical proteins although the genome and proteome of this organism is well studied. Hence, this study aims to find potential vaccine candidates with their epitopes from the uncharacterized protein pool of *Staphylococcus aureus*by applying RV and *immune-informatics* methods. Those epitopes could be considered as promising candidates for effective protein-based vaccines against *S.aureus.*

Materials And Methods

The proposed work is done in the Bioinformatics lab, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, TamilNadu,India. There is no ethical approval as human samples are not involved. For each organism the number of groups is one. The sample size is 50 proteins per group.

Fifty hypothetical proteins of *Staphylococcus aureu*s were retrieved from NCBI. Protein parameters such as molecular weight, pI, Instability index, Aliphatic index, extinction coefficient and GRAVY of the hypothetical proteins were studied using ProtParam(https://web.expasy.org/protparam/)Table 2. Functional classification of those hypothetical proteins into various streams was done through the VICM pred tool (http://www.imtech.res.in/raghava/vicmpred/) was used Table 3. Subcellular localization was predicted for the 50 sequences using CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) (Horton et al. 2007). Virulence and antigenicty properties of the fifty uncharacterized proteins were identified using VaxiJen ver. 2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), shown in Table 5. The virulence nature of these proteins are listed in Table 6. The combination of ABCpred, and VaxiJen servers (http://ailab-projects1.ist.psu.edu:8080/bcpred/predict.html) allowed the prediction of overlapping antigenic B cell epitopes (Rammensee, Bachmann, and Stevanovic 2013; Rammensee, Bachmann, and Stevanović 1997)from the uncharacterized proteins. Out of the 35 antigenic B cell epitopes, 27 are non allergic and 8 are allergic. Antigenic B cell epitopes from the selected hypothetical proteins of *S.aureus*are presented in Table 7.

CTLPred server (http://crdd.osdd.net/raghava/ctlpred/) predicted a total of 6 cytotoxic T cell epitopes from the uncharacterized proteins studied (Bhasin and Raghava 2004; "Website," n.d.). A total of 2 out of 6 cytotoxic T cell epitope regions were predicted as antigens by VaxiJen server. Of these 1 antigenic epitope is found in OHS92109.1 and another one found in OHS88591.1. Both the antigenic epitopes are allergen. The results of the T cell epitopes are presented in Table 8. Characteristics of transmembrane helices in uncharacterized proteins was predicted by TMHMM based on the hidden Markov model and HMMTOP (http://www.enzim.hu/hmmtop/html/submit.html) (Koski 2001); (Tusnady and Simon 2001; "Website," n.d.). The predicted transmembrane helices are provided in Table 9. Avoiding interference against human immune mechanisms, these uncharacterized protein sequences are submitted to tBLASTn(https://blast.ncbi.nlm.nih.gov/) for non-human homologous (Heath and Ramakrishnan 2010). From the two sequences it is predicted that both the sequence had no significant similarity with humans. The Table 10 provides the results of BLAST.

Results

Table 1 shows the hypothetical protein sequences retrieved from NCBI. Table 2. Depicts the physio-chemical properties of the retrieved protein sequences predicted using PROTPARAM tools. Table 3 reveals the functional classification of hypothetical proteins found using the VICMpred tool. Table 4 shows the subcellular localization for all the proteins predicted using the CELLO2GO tool. Table 5 shows the antigenicity characteristics of the proteins revealed using VaxiJen ver. 2.0 tool. Table 6 demonstrates the virulence score of the hypothetical proteins using the Virulent Predtool.Table 7 depicts the presence of epitopes in the proteins studied using ABCpred and VaxiJen tools. Table 8 reveals the presence of T- cell epitopes in the uncharacterized protein pool using CTLpred tool. Table 9 shows the number of transmembrane helices found in the 2 putative antigenic proteins analyzed using the HMMTOP tool. Table 10 depicts the tBLASTn results of the 2 putative novel antigenic proteins, confirming if they are homologous to any human proteins. From these results, we can conclude that these proteins are potential candidates for drug designing against *S.aureus.*

Discussion

Two potential candidate proteins possessing virulence and antigenic properties, comprising epitopes and are not human homologs were found in this study. Hence, these proteins could be ideal drug/vaccine targets, however, further in-depth immuno-informatics and structural biology approaches are recommended with *in-vitro* and *in-vivo* experiments for validation.

Nosocomial infections due to *Staphylococcus aureus*are still a major cause of mortality particularly in the developing countries. Over the years, there has been rapid development of low-cost sequencing techniques which has led to generation of huge amounts of genomic and proteomic data; however, research on hypothetical proteins (HP) is yet to keep pace with. Currently, over 50% of the *S.aureus*proteins have no ascribed function. Characterization of HP may be useful in better understanding the organism's metabolic pathways, disease progression, drug development, and disease control strategies (Brüssow 2019). With a complete *S.aureus*genome sequence (Nene and Kole 2008) and advancement in bioinformatics, it is now possible to identify potential vaccine candidates using reverse vaccinology which reduces the time and cost of designing and identifying vaccine candidates (Rappuoli et al. 2016).

This study utilized several immuno-informatics tools for characterization of hypothetical proteins of *S.aureus*for vaccine development. Out of 50 hypothetical proteins studied, 2 (OHS92109.1, and OHS88591.1) were identified as potential drug targets against *S.aureus*. The BLAST(Heath and Ramakrishnan 2010) results of the two candidates suggest that they could be used for drug development without causing autoimmunity. VaxiJen server identified the 11 potential antigens out of 50 hypothetical proteins. These 11 proteins were assessed for virulence potential through virulentpred tool, which revealed that only 2 proteins possessed virulence (Nene and Kole 2008). Hence, these 2 were subjected to epitope prediction. The B cell epitopes were identified through ABCpred server while CTL **e**pitopes were predicted using CTLPred web server listed in Tables7 and 8. Immunodominant epitopes that can induce specific immune responses could be a potential peptide vaccine (Sanchez-Trincado, Gomez-Perosanz, and Reche 2017) . Our results demonstrate a complete workflow for mining of vaccine candidates from unexplored protein pools of organisms using immunoinformatics(Brüssow 2019).

The outcome of this study could provide insights into bacterial pathogenesis and can aid in drug development. However the main limitation of this study is that the selected drug candidates along with their epitopes should be further validated for their immunogenicity and protective efficacy experimentally if they are to be used for future drug development against *S.aureus.*

Conclusion

Reverse vaccinology is a promising strategy for the screening and identification of antigenic antigens with potential capacity to elicit cellular and humoral immune responses against *S.aureus*infection. In this study, two hypothetical proteins were selected through computational methods and verified as potential drug candidates against pneumonia. We therefore recommend further in-depth immunoinformatics and structural biology approaches together with *in-vitro* and *in-vivo* experiments to validate their immunogenicity and protective efficacy to completely decipher the vaccine targets against *S.aureus.*

Declarations

Conflict of Interest

The authors of this paper declare no conflict of interest.

Author Contribution

Author VM was involved in data collection, data analysis, manuscript writing. Author JA was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1.Qbiogen LLP, Chennai

2. Saveetha University

3.Saveetha Institute of Medical and Technical Sciences

4. Saveetha School of Engineering

References

1. Adhinarayanan, Rajesh, AravindhRamakrishnan, GopalKaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and DamodharanDillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
2. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite." *Biomass Conversion and Biorefinery*, August. https://doi.org/10.1007/s13399-020-00938-0.
3. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." *Biomass Conversion and Biorefinery*, May. https://doi.org/10.1007/s13399-021-01551-5.
4. Beeson, James G., Liriye Kurtovic, Carlota Dobaño, D. Herbert Opi, Jo-Anne Chan, Gaoqian Feng, Michael F. Good, Linda Reiling, and Michelle J. Boyle. 2019. "Challenges and Strategies for Developing Efficacious and Long-Lasting Malaria Vaccines." *Science Translational Medicine*. https://doi.org/10.1126/scitranslmed.aau1458.
5. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for PhotocatalyticTransesterification of Castor Oil." *Fuel* 304 (November): 121490.
6. Bhasin, Manoj, and G. P. S. Raghava. 2004. "Prediction of CTL Epitopes Using QM, SVM and ANN Techniques." *Vaccine*. https://doi.org/10.1016/j.vaccine.2004.02.005.
7. Böhme, Ulrike, Thomas D. Otto, Mandy Sanders, Chris I. Newbold, and Matthew Berriman. 2019. "Progression of the Canonical Reference Malaria Parasite Genome from 2002–2019." *Wellcome Open Research*. https://doi.org/10.12688/wellcomeopenres.15194.2.
8. Brüssow, Harald. 2019. *Hurdles for Phage Therapy (PT) to Become a Reality*. MDPI.
9. Chambers, Henry F., and Frank R. DeLeo. 2009. "Waves of Resistance: Staphylococcus Aureus in the Antibiotic Era." *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro2200.
10. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01312-4.
11. Heath, Lenwood S., and NarenRamakrishnan. 2010. *Problem Solving Handbook in Computational Biology and Bioinformatics*. Springer Science & Business Media.
12. Horton, P., K-J Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai. 2007. "WoLF PSORT: Protein Localization Predictor." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm259.
13. Jayanth, BellappuVenkat, Melvin Victor Depoures, GopalKaliyaperumal, DamodharanDillikannan, DilipsinghJawahar, KumaranPalani, and Ganesha Prasad MeravanigeeShivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
14. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and ShantanuPatil. 2020. "Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." *Process Biochemistry* 99 (December): 36–47.
15. Koski, T. 2001. *Hidden Markov Models for Bioinformatics*. Springer Science & Business Media.
16. Lung, Tania Wong Fok, and Alice Prince. 2020. "Consequences of Metabolic Interactions during Staphylococcus Aureus Infection." *Toxins*. https://doi.org/10.3390/toxins12090581.
17. Maple, P. A. C., J. M. T. Hamilton-Miller, and W. Brumfitt. 1989. "WORLD-WIDE ANTIBIOTIC RESISTANCE IN METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS." *The Lancet*. https://doi.org/10.1016/s0140-6736(89)90076-7.
18. Mathers, C. 2008. "Global Burden of Disease." *International Encyclopedia of Public Health*. https://doi.org/10.1016/b978-012373960-5.00335-x.
19. Mota, Maria M., and Ana Rodriguez. 2017. *Malaria: Immune Response to Infection and Vaccination*. Springer.
20. Nene, Vishvanath, and ChittaranjanKole. 2008. *Genome Mapping and Genomics in Animal-Associated Microbes*. Springer Science & Business Media.
21. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." *Fuel* 277 (October): 118166.
22. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." *Fuel* 278 (October): 118315.
23. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." *Biomass Conversion and Biorefinery*, March. https://doi.org/10.1007/s13399-021-01427-8.
24. Rammensee, Hans-Georg, Jutta Bachmann, and Stefan Stevanović. 1997. "The MHC Genes." *MHC Ligands and Peptide Motifs*. https://doi.org/10.1007/978-3-662-22162-4_2.
25. Rammensee, Hans-Georg, Jutta Bachmann, and Stefan Stevanovic. 2013. *MHC Ligands and Peptide Motifs*. Springer Science & Business Media.
26. Rappuoli, Rino, Matthew J. Bottomley, Ugo D'Oro, Oretta Finco, and Ennio De Gregorio. 2016. "Reverse Vaccinology 2.0: Human Immunology Instructs Vaccine Antigen Design." *Journal of Experimental Medicine*. https://doi.org/10.1084/jem.20151960.
27. Sanchez-Trincado, Jose L., Marta Gomez-Perosanz, and Pedro A. Reche. 2017. "Fundamentals and Methods for T- and B-Cell Epitope Prediction." *Journal of Immunology Research* 2017 (December): 2680160.
28. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.
29. Shanmugam, Rajasekaran, DamodharanDillikannan, GopalKaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 43 (23): 3064–81.
30. Sizar, Omeed, and Chandrashekhar G. Unakal. 2021. "Gram Positive Bacteria." In *StatPearls*. Treasure Island (FL): StatPearls Publishing.
31. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from

Spent Biomass of GanodermaLucidum Using Saccharomyces Cerevisiae." *Fuel* 306 (December): 121680.
32. Tang, Yong, Thomas R. Meister, Marta Walczak, Michael J. Pulkoski-Gross, Sanjay B. Hari, Robert T. Sauer, Katherine Amberg-Johnson, and Ellen Yeh. 2019. "A Mutagenesis Screen for Essential Plastid Biogenesis Genes in Human Malaria Parasites." *PLoS Biology* 17 (2): e3000136.
33. Tong, Steven Y. C., Joshua S. Davis, Emily Eichenberger, Thomas L. Holland, and Vance G. Fowler. 2015. "Staphylococcus Aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management." *Clinical Microbiology Reviews*. https://doi.org/10.1128/cmr.00134-14.
34. Tusnady, G. E., and I. Simon. 2001. "The HMMTOP Transmembrane Topology Prediction Server." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/17.9.849.
35. "Website." n.d. Bhasin, Manoj, and G. P. S. Raghava. 2004. "Prediction of CTL Epitopes Using QM, SVM and ANN Techniques." Vaccine. https://doi.org/10.1016/j.vaccine.2004.02.005.
36. ———. n.d. Tusnady, G. E., and I. Simon. 2001. "The HMMTOP Transmembrane Topology Prediction Server." Bioinformatics. https://doi.org/10.1093/bioinformatics/17.9.849.

List of Tables

Table 1. Retrieval of hypothetical protein sequences from NCBI

S.No	Accession Number	Protein
1.	BBA24007.1	Hypothetical
2.	OFO53938.1	Hypothetical
3.	OFO51842.1	Hypothetical
4.	KKI67441.1	Hypothetical
5.	OHS92154.1	Hypothetical
6.	OHS92109.1	Hypothetical
7.	OHS91934.1	Hypothetical
8.	OHS91929.1	Hypothetical
9.	OHS91773.1	Hypothetical
10.	OHS91612.1	Hypothetical
11.	OHS91604.1	Hypothetical
12.	OHS91589.1	Hypothetical
13.	OHS91480.1	Hypothetical
14.	OHS91477.1	Hypothetical
15.	OHS91476.1	Hypothetical
16.	OHS91447.1	Hypothetical
17.	OHS91403.1	Hypothetical
18.	OHS91070.1	Hypothetical
19.	OHS90801.1	Hypothetical
20.	OHS90699.1	Hypothetical
21.	OHS90493.1	Hypothetical
22.	OHS89870.1	Hypothetical
23.	OHS89598.1	Hypothetical

24.	OHS89048.1	Hypothetical
25.	OHS89022.1	Hypothetical
26.	OHS88591.1	Hypothetical
27.	OHS83079.1	Hypothetical
28.	OHS87955.1	Hypothetical
29.	OHS87810.1	Hypothetical
30.	OHS87642.1	Hypothetical
31.	OHS83454.1	Hypothetical
32.	OHS87136.1	Hypothetical
33.	OHS87087.1	Hypothetical
34.	OHS86427.1	Hypothetical
35.	OHS86352.1	Hypothetical
36.	OHS85415.1	Hypothetical
37.	OHS85414.1	Hypothetical
38.	OHS85411.1	Hypothetical
39.	OHS85356.1	Hypothetical
40.	OHS85313.1	Hypothetical
41.	OHS85185.1	Hypothetical
42.	OHS85170.1	Hypothetical
43.	OHS85162.1	Hypothetical
44.	OHS85083.1	Hypothetical
45.	OHS84751.1	Hypothetical
46.	OHS84633.1	Hypothetical
47.	OHS84612.1	Hypothetical
48.	OHS84492.1	Hypothetical
49.	OHS84306.1	Hypothetical
50.	OHS84005.1	Hypothetical

Table 2. Physio-chemical properties of the retrieved protein sequences using PROTPARAM

S.No	Protein ID	Molecular weight	P_I	GRAVY	Instability index	Aliphatic index	Extinction coefficient
1.	BBA24007.1	900211.37	5.73	-0.776	23.47	76.03	216840
2.	OFO53938.1	125589.50	9.91	0.313	22.24	107.50	177380
3.	OFO51842.1	35632.59	5.97	-0.506	43.15	89.44	35425
4.	KKI67441.1	900652.90	5.74	-0.776	23.47	76.00	216840
5.	OHS92154.1	35209.59	5.73	0.128	46.25	111.25	9970
6.	OHS92109.1	13273.30	6.83	-0.125	35.82	108.00	4470
7.	OHS91934.1	25718.30	4.65	-0.296	33.99	96.31	18005
8.	OHS91929.1	11948.98	3.84	-0.814	68.75	79.22	9970
9.	OHS91773.1	33069.77	6.27	-0.494	34.03	87.23	19370
10.	OHS91612.1	8751.75	5.30	-0.822	42.44	77.08	5960
11.	OHS91604.1	10691.91	5.16	-0.798	36.12	54.62	15930
12.	OHS91589.1	10355.17	5.30	-0.124	29.50	112.53	1490
13.	OHS91480.1	8740.99	4.36	0.086	64.61	102.25	125
14.	OHS91477.1	8656.76	3.94	0.087	38.30	97.44	11710
15.	OHS91476.1	12486.13	4.35	-0.157	28.88	74.62	6210
16.	OHS91477.1	14484.94	10.46	1.022	49.25	141.32	17990
17.	OHS91403.1	19255.73	4.87	-0.324	35.86	89.94	12950
18.	OHS91070.1	29019.82	5.58	0.816	32.61	146.75	20400
19.	OHS90801.1	12037.39	4.33	-0.435	27.22	95.65	4470
20.	OHS90699.1	21703.58	5.60	-0.683	42.27	80.06	48485
21.	OHS90493.1	41918.33	9.34	-0.311	23.87	94.41	32110
22.	OHS89870.1	8533.50	9.99	1.282	21.84	150.48	17990
23.	OHS89598.1	33069.77	6.27	-0.494	34.03	87.23	19370
24.	OHS89048.1	36260.45	5.52	-0.121	27.73	102.45	10555
25.	OHS89022.1	20032.49	8.49	-0.832	51.68	81.10	21430
26.	OHS88591.1	13385.24	5.79	0.001	19.99	106.69	10430
27.	OHS83079.1	41890.27	9.34	-0.318	23.87	93.90	32110
28.	OHS87955.1	17717.04	9.19	-0.628	43.01	75.58	11460

29.	OHS87810.1	9724.03	4.59	-0.203	29.04	98.60	2980
30.	OHS87642.1	22191.00	4.65	-0.410	45.11	79.73	35410
31.	OHS83454.1	28983.66	5.59	0.742	34.34	143.29	21890
32.	OHS87136.1	14924.28	6.29	-0.250	44.88	98.17	21095
33.	OHS87087.1	10502.96	5.22	-0.320	24.70	98.54	15470
34.	OHS86427.1	19436.90	4.53	-0.712	27.97	60.68	41495
35.	OHS86352.1	8533.50	9.99	1.282	21.84	150.48	17990
36.	OHS85415.1	12485.19	4.52	-0.161	27.26	74.62	6210
37.	OHS85414.1	8656.76	3.94	0.087	38.30	97.44	11710
38.	OHS85411.1	8740.99	4.36	0.086	64.61	102.25	125
39.	OHS85356.1	19325.86	5.04	-0.314	31.47	90.00	12950
40.	OHS85313.1	14538.97	10.70	1.035	45.80	144.34	17990
41.	OHS85185.1	10355.17	5.30	-0.124	29.50	112.53	1490
42.	OHS85170.1	10691.91	5.16	-0.798	36.12	54.62	15930
43.	OHS85162.1	8751.75	5.30	-0.822	42.44	77.08	5960
44.	OHS85083.1	68621.30	8.08	0.001	31.51	106.74	46550
45.	OHS84751.1	14924.28	6.29	-0.250	44.88	98.17	21095
46.	OHS84633.1	8633.76	5.87	-0.947	26.65	69.32	4470
47.	OHS84612.1	11144.64	4.93	-0.393	32.68	87.70	5960
48.	OHS84492.1	10517.03	5.72	-0.324	22.93	98.54	15470
49.	OHS84306.1	21703.58	5.60	-0.683	42.27	80.06	48485
50.	OHS84005.1	8670.55	5.23	-1.089	32.18	57.40	7450

Table 3. Functional classification of hypothetical proteins using VICMpred.

S.No	Accession Number	Functional Class	Score
1.	BBA24007.1	Information and storage	-34.421617
2.	OFO53938.1	Metabolism molecule	38.80472
3.	OFO51842.1	Metabolism molecule	-0.13024919
4.	KKI67441.1	Information and storage	-34.421617
5.	OHS92154.1	Cellular process	1.7245893
6.	OHS92109.1	Cellular process	0.52953375
7.	OHS91934.1	Cellular process	0.25967325
8.	OHS91929.1	Information and storage	0.60415804
9.	OHS91773.1	Cellular process	0.23712935
10.	OHS91612.1	Cellular process	1.3259646
11.	OHS91604.1	Information and storage	0.063502072
12.	OHS91589.1	Cellular process	0.36029923
13.	OHS91480.1	Cellular process	1.0439275
14.	OHS91477.1	Information and storage	0.36443194
15.	OHS91476.1	Information and storage	0.8617311
16.	OHS91477.1	Metabolism molecule	-0.11811703
17.	OHS91403.1	Metabolism molecule	0.46164659
18.	OHS91070.1	Metabolism molecule	0.1525313
19.	OHS90801.1	Metabolism molecule	-0.256837
20.	OHS90699.1	Cellular process	0.93854621
21.	OHS90493.1	Cellular process	0.25569424
22.	OHS89870.1	Metabolism molecule	0.58552768
23.	OHS89598.1	Cellular process	0.23712935
24.	OHS89048.1	Cellular process	1.668612
25.	OHS89022.1	Metabolism molecule	0.65408653
26.	OHS88591.1	Cellular process	1.2514507
27.	OHS83079.1	Cellular process	0.2936825
28.	OHS87955.1	Metabolism molecule	-0.12944347
29.	OHS87810.1	Cellular process	-0.17933529

30.	OHS87642.1	Metabolism molecule	-0.046537022
31.	OHS83454.1	Metabolism molecule	0.73022272
32.	OHS87136.1	Information and storage	1.6308689
33.	OHS87087.1	Virulence factor	1.1866571
34.	OHS86427.1	Information and storage	0.17788394
35.	OHS86352.1	Metabolism molecule	0.58552768
36.	OHS85415.1	Information and storage	0.8245954
37.	OHS85414.1	Information and storage	0.36443194
38.	OHS85411.1	Cellular process	1.0439275
39.	OHS85356.1	Metabolism molecule	0.50182408
40.	OHS85313.1	Metabolism molecule	-0.34805698
41.	OHS85185.1	Cellular process	0.36029923
42.	OHS85170.1	Information and storage	0.063502072
43.	OHS85162.1	Cellular process	1.3259646
44.	OHS85083.1	Cellular process	4.1212716
45.	OHS84751.1	Information and storage	1.6308689
46.	OHS84633.1	Information and storage	0.71904284
47.	OHS84612.1	Information and storage	-0.56751745
48.	OHS84492.1	Virulence factors	1.8432142
49.	OHS84306.1	Cellular process	0.93854621
50.	OHS84005.1	Information and storage	0.28639416

Table 4. Subcellular localization prediction for the proteins using CELLO2GO

S.No	**Accession Number**	**Localization**
1.	BBA24007.1	Extracellular
2.	OFO53938.1	Inner membrane
3.	OFO51842.1	Cytoplasmic
4.	KKI67441.1	Extracellular
5.	OHS92154.1	Inner membrane
6.	OHS92109.1	Cytoplasmic
7.	OHS91934.1	Cytoplasmic

8.	OHS91929.1	Cytoplasmic
9.	OHS91773.1	Cytoplasmic
10.	OHS91612.1	Cytoplasmic
11.	OHS91604.1	Cytoplasmic
12.	OHS91589.1	Cytoplasmic
13.	OHS91480.1	Cytoplasmic
14.	OHS91477.1	Cytoplasmic
15.	OHS91476.1	Cytoplasmic
16.	OHS91477.1	Inner membrane
17.	OHS91403.1	Cytoplasmic
18.	OHS91070.1	Inner membrane
19.	OHS90801.1	Cytoplasmic
20.	OHS90699.1	Cytoplasmic
21.	OHS90493.1	Cytoplasmic
22.	OHS89870.1	Inner membrane
23.	OHS89598.1	Cytoplasmic
24.	OHS89048.1	Cytoplasmic
25.	OHS89022.1	Inner membrane
26.	OHS88591.1	Cytoplasmic
27.	OHS83079.1	Cytoplasmic
28.	OHS87955.1	Inner membrane
29.	OHS87810.1	Cytoplasmic
30.	OHS87642.1	Cytoplasmic
31.	OHS83454.1	Inner membrane
32.	OHS87136.1	Cytoplasmic
33.	OHS87087.1	Cytoplasmic
34.	OHS86427.1	Cytoplasmic
35.	OHS86352.1	Inner membrane
36.	OHS85415.1	Cytoplasmic
37.	OHS85414.1	Cytoplasmic

38.	OHS85411.1	Cytoplasmic
39.	OHS85356.1	Cytoplasmic
40.	OHS85313.1	Inner membrane
41.	OHS85185.1	Cytoplasmic
42.	OHS85170.1	Cytoplasmic
43.	OHS85162.1	Cytoplasmic
44.	OHS85083.1	Inner membrane
45.	OHS84751.1	Cytoplasmic
46.	OHS84633.1	Cytoplasmic
47.	OHS84612.1	Cytoplasmic
48.	OHS84492.1	Cytoplasmic
49.	OHS84306.1	Cytoplasmic
50.	OHS84005.1	Cytoplasmic

Table 5. Evaluating the antigenicity of the proteins using VaxiJen ver. 2.0

S.No	**Accession Number**	**Antigenicity score**
1.	BBA24007.1	0.6113
2.	OFO53938.1	0.2748
3.	OFO51842.1	0.2580
4.	KKI67441.1	0.6120
5.	OHS92154.1	0.7373
6.	OHS92109.1	0.5365
7.	OHS91934.1	0.3239
8.	OHS91929.1	0.4391
9.	OHS91773.1	0.4664
10.	OHS91612.1	0.4202
11.	OHS91604.1	0.3570
12.	OHS91589.1	0.3460
13.	OHS91480.1	0.3317
14.	OHS91477.1	0.3396
15.	OHS91476.1	0.5909

16.	OHS91477.1	0.5381
17.	OHS91403.1	0.4915
18.	OHS91070.1	0.5296
19.	OHS90801.1	0.2441
20.	OHS90699.1	0.3145
21.	OHS90493.1	0.2898
22.	OHS89870.1	0.0492
23.	OHS89598.1	0.4664
24.	OHS89048.1	0.4360
25.	OHS89022.1	0.3483
26.	OHS88591.1	0.6375
27.	OHS83079.1	0.2945
28.	OHS87955.1	0.3812
29.	OHS87810.1	0.2208
30.	OHS87642.1	0.1108
31.	OHS83454.1	0.4999
32.	OHS87136.1	0.3032
33.	OHS87087.1	0.4140
34.	OHS86427.1	0.3222
35.	OHS86352.1	0.0492
36.	OHS85415.1	0.6012
37.	OHS85414.1	0.3396
38.	OHS85411.1	0.3317
39.	OHS85356.1	0.5044
40.	OHS85313.1	0.5216
41.	OHS85185.1	0.3460
42.	OHS85170.1	0.3570
43.	OHS85162.1	0.4202
44.	OHS85083.1	0.4535
45.	OHS84751.1	0.3032

46.	OHS84633.1	0.3291
47.	OHS84612.1	0.3394
48.	OHS84492.1	0.3896
49.	OHS84306.1	0.3145
50.	OHS84005.1	0.4036

Table 6 . Virulence prediction for the proteins using Virulentpred tool

S.No	**Accession Number**	**Virulent\Non Virulent**	**Score**
1.	BBA244007.1	Virulent	1.0073
2.	KKI67441.1	Virulent	1.0074
3.	OHS92154.1	Non-Virulent	-0.990
4.	OHS92109.1	Virulent	1.1701
5.	OHS91476.1	Non-Virulent	-0.770
6.	OHS91447.1	Virulent	1.0386
7.	OHS91070.1	Virulent	0.0207
8.	OHS88591.1	Virulent	1.1346
9.	OHS85415.1	Non-Virulent	-0.577
10.	OHS85356.1	Non-Virulent	-0.990
11.	OHS85313.1	Virulent	0.9485

Table 7. Epitope prediction using ABCpred and VaxiJen

Accession Number	B cell epitope	ABC Pred	Vaxijen	Allergen FP
OHS92109.1	YLCCCSARESMVKVTD	0.92	0.7691	Non- allergen
	GIYIDVYCALKHGVNI	0.88	1.0023	Non- allergen
	HMPREFSTAPHYLCCC	0.87	0.1756	Non- allergen
	HETICALPRTEINHMP	0.82	0.0498	Non- allergen
	ESMVKVTDYSNSKLGK	0.82	0.4861	Non- allergen
	NMTAIEPKQINIHITQ	0.80	1.2731	Non- allergen
	SEVEGITGHFAELKET	0.79	0.7258	Non- allergen
	PRTEINHMPREFSTAP	0.79	0.0463	Non- allergen
	IESKEDGIYIDVYCAL	0.79	1.1354	Non- allergen
	AELKETNLEKVSRKNL	0.77	0.9468	Allergen
	HSHYPTHETICALPRT	0.75	-0.3716	Non- allergen
	KHGVNISKTANKIQTS	0.74	0.9253	Allergen
	QTSIFNSISNMTAIEP	0.70	-0.1430	Allergen
	GKVEIAPEVLSVIASI	0.69	0.2924	Non- allergen
	SKTANKIQTSIFNSIS	0.62	0.4935	Allergen
	EKVSRKNLSRDLKIES	0.60	1.0265	Allergen
	TDYSNSKLGKVEIAPE	0.59	0.7200	Non- allergen
OHS88591.1	YGKIDISNEVIASVVG	0.89	0.0759	Non- allergen
	HMPREFSTAPHYLCCC	0.87	0.1756	Non- allergen
	SNEVIASVVGGKAVEC	0.84	0.7685	Non- allergen
	VVDIDMYIIVSYGVKI	0.84	1.2271	Non- allergen
	KAVECYGIVGMASRQQ	0.83	0.4589	Non- allergen
	SVNSINIYVQGVRVNN	0.83	1.3259	Allergen
	HETICALPRTEINHMP	0.83	0.0498	Non- allergen
	SMTLEISNDYGKIDIS	0.82	0.6723	Non- allergen
	LCCCSARESMTLEISN	0.81	0.8756	Non- allergen
	PRTEINHMPREFSTAP	0.79	0.0463	Non- allergen
	NYAKGIKVTENNGVVD	0.79	0.7800	Non- allergen
	IVSYGVKISEVANNVQ	0.77	0.5229	Non- allergen

	HSHYPTHETICALPRT	0.76	-0.3716	Non- allergen
	QSTVKYTLEKSLNVSV	0.75	0.9549	Non- allergen
	VRDGIAEILGHENYAK	0.74	0.0320	Allergen
	VGMASRQQVRDGIAEI	0.70	0.3612	Non- allergen
	KISEVANNVQSTVKYT	0.58	0.4712	Non- allergen
	KVTENNGVVDIDMYII	0.55	1.2397	Allergen

Table 8. T- cell epitope prediction of uncharacterized proteins using CTLpred

Accession Number	**Peptide**	**Start**	**TC pred score**	**Antigenicity score**	**Allergencity**
OHS92109.1	SHYPTHETI	2	0.990	-0.2840	Non-allegen
	IASIATSEV	65	0.990	0.1467	Allergen
	CALKHGVNI	115	0.990	0.5998	Allergen
OHS88591.1	SHYPTHETI	2	0.990	-0.2840	Allergen
	ISNEVIASV	54	0.980	0.2517	Allergen
	SINIYVQGV	147	0.980	1.3315	Allergen

Table 9. Transmembrane helices prediction for the 2 putative antigenic proteins using HMMTOP

Accession Number	**TMHMM Score**	**HMMTOP Score**
OHS92109.1	0	0
OHS88591.1	0	0

Table 10. Screening of the 2 putative antigenic proteins using tBLASTn for off targets

Accession Number	**Non-human Homologous**
OHS92109.1	No significant similarity found
OHS88591.1	No significant similarity found