



MINING OF NOVEL DRUG/VACCINE TARGETS FROM THE PROTEOME OF NEISSERIA GONORRHOEAE USING COMPUTATIONAL TOOLS THROUGH REVERSE VACCINOLOGY APPROACH.

Vaishnavi Modupalli¹ and Abiramavalli Moorthy^{2*}

Article History: Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

Abstract:

Aim: The present study intends to explore the cluster of uncharacterized proteins of *Neisseria gonorrhoea* using a series of computational tools to identify novel drug/vaccine candidates. *Gonorrhoea* is a sexually transmitting infection caused by a gram-negative pathogenic bacteria *N. gonorrhoea*. **Materials and Methods:** A total of 50 uncharacterized protein sequences of *N. gonorrhoeae* were retrieved from NCBI and analyzed using computational tools for studying their localization, membrane helices, physio-chemical 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 circumvent autoimmune reactions. **Results:** Out of 50, 41 were found to be highly virulent with a score of more than 1 and among the 41, 13 proteins possess antigenic properties. Out of 13 antigens, 2 candidates comprise epitopes and they are not human homologs. **Conclusion:** Hence, these proteins could be novel 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, *Gonorrhoeae*, Proteome, Virulence.

¹Research Scholar, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, TamilNadu, India, Pincode:602105

²Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, TamilNadu, India, Pincode: 602105

1. Introduction

Neisseria gonorrhoeae, a gram-negative human-specific pathogenic bacteria, responsible for a sexually transmitted disease, gonorrhoea globally. The World Health Organization has recently predicted that 106 million new cases emerge in adults annually (Shanmugham & Pan, 2013). This antique infection in humans persists as a serious issue, with a high prevalence rate globally and having a major influence on reproductive and neonatal health (Sintchenko, 2009). Treatment failures are linked with the current development of gonorrhoeae with reduced susceptibility to the third-generation antibiotics. So far, no effective vaccine exists to prevent gonococcal infections (Kanguane, 2009; Unemo, 2015); (Kaur et al., 2021)). Hence, there is an urgent need for fast-tracked research towards the identification of molecular targets for development of drugs and preventive vaccines to fight against antibiotic-resistant gonorrhoea.

In this study, Reverse Vaccinology, a novel strategy to recognize molecular targets for drug and vaccine (Mota & Rodriguez, 2017; Ouattara et al., 2015; Sanchez-Trincado et al., 2017) development without culturing the organism was adopted to predict potential drug/vaccine candidates from the proteome of the highly pathogenic species *N. gonorrhoeae*. Total proteome of this species was analyzed to identify uncharacterized proteins, which were retrieved and further subjected to in-silico characterization to catch the novel drug/vaccine targets. More than 700 studies related to characterization of hypothetical proteins of various microbes towards identification of novel drug targets using computational tools were found in PubMed in the past 5 years (Shanmugham & Pan, 2013). The most recent study was (Kaur et al., 2021), where the hypothetical proteins of the uropathogenic *E. coli* strain were characterized by bioinformatics tools to find antimicrobial drug targets.

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Dinesh Kumar et al., 2022; Kumar et al., 2022; Mahesh et al., 2022; Mohanavel et al., 2022; Ram et al., 2022; Rinesh et al., 2022; Sathish et al., 2022; Sudhan et al., 2022; Sundararaman et al., 2022; Vijayalakshmi et al., 2022; Yaashikaa et al., 2022). However, the lacunae identified here is that, although the genome and proteome of this organism is well studied, nearly one fourth of its proteome is annotated as hypothetical proteins. Hence, the aim of this study was to find potential vaccine candidates with their epitopes from the uncharacterized protein pool of *Neisseria*

gonorrhoeae by applying RV and immune-informatics methods. These identified epitopes could be considered as promising candidates for effective protein-based vaccines against gonorrhoeae.

2. Materials and Methodology

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.

The FASTA sequences of fifty uncharacterized proteins of *Neisseria gonorrhoeae* were retrieved from NCBI. The physicochemical properties including the molecular weight, pI, instability index, aliphatic index, extinction coefficient and GRAVY of the hypothetical proteins were predicted using ProtParam (<https://web.expasy.org/protparam/>) (Gasteiger et al., 2005), shown in Table 2. VICM pred tool (<http://www.imtech.res.in/raghava/vicmpred/>) was used for functional classification of hypothetical proteins (Saha & Raghava, 2006), shown in Table 3. Subcellular localization was predicted for the 50 sequences using CELLO2GO (<http://cello.life.nctu.edu.tw/cello2go/>) (Yu et al., 2014). Virulence and antigenicity properties of the fifty uncharacterized proteins were identified using VaxiJen ver. 2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>).

Vaxijen scores above 0.5 are shown in Table 5. The virulence nature of these proteins are listed in Table 6. Accurate identification of antigenic epitopes on a protein is important for the development of immunodiagnostics. 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 from the uncharacterized proteins. Out of the 50 proteins, only 9 were non-virulent while 41 were virulent proteins. Antigenicity scores were predicted for the 41 virulent proteins. Based on the antigenicity, 2 virulent antigens were selected for B cell epitopes prediction. 5 allergens were detected, shown 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 & Raghava, 2004; Website, n.d.-a). 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 TND82389.1 and another one found in TND59612.1. 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>) (Tusnady & Simon, 2001; Website, n.d.-b). 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/>) ((Altschul et al., 1990) for non-human homologs. From the two sequences it is predicted that both the sequence had no significant similarity with humans Table 10.

3. Results

Table 1 shows the hypothetical protein sequences retrieved from NCBI. Table 2. Depicts the physiochemical 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 Pred tool. 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 *Neisseria gonorrhoeae*.

4. Discussion

In this study, several bioinformatics and immunoinformatics tools were utilized for characterization of hypothetical proteins of *Neisseria gonorrhoeae* for identification of novel drug targets. Out of 50 hypothetical proteins studied, 2 (TND82389.1 and TND59612.1) were identified as potential vaccine target candidates against *Neisseria gonorrhoeae*. The BLAST(Heath & Ramakrishnan, 2010) results of the two selected

HPs suggest that they could be used for drug development without causing autoimmunity.

The current study has utilized immuno-informatics tools for characterization of hypothetical proteins of *N.gonorrhoeae* for vaccine development (Altschul et al., 1990). Out of the 50 proteins, only 9 were non-virulent while 41 were virulent proteins. Antigenicity scores were predicted for the 41 virulent proteins. Based on the antigenicity, 2 virulent antigens were selected for B cell epitopes prediction. 5 allergens were detected. Immunodominant epitopes that can induce specific immune responses could be a potential peptide vaccine candidate (Sanchez-Trincado et al., 2017). We have used a web server CELLO2GO for localization prediction with functional classification (Yu et al., 2014). A SVM based approach has been followed for prediction of virulence proteins (Garg & Gupta, 2008). Our results demonstrate a complete workflow for mining of vaccine candidates from unexplored protein pools of organisms using immunoinformatics.

Therefore, information generated herein states about the characteristics of the two novel drug targets which could shed insight into pathogenesis and can provide the basis for novel drug approaches. However it is essential that the selected drug candidates along with their epitopes be further validated for their immunogenicity and protective efficacy experimentally which is a limitation of this study. The future scope of this study is to validate the potential of these novel drug targets through wet lab experiments and to develop a drug against *Neisseria gonorrhoeae*.

5. 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 *N.gonorrhoeae* infection. In this study, two hypothetical proteins were selected through computational methods and verified as potential drug candidates against *gonorrhoeae*. 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 *gonorrhoeae*.

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 AM 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

6. References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. In *Journal of Molecular Biology* (Vol. 215, Issue 3, pp. 403–410). [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
- Bhasin, M., & Raghava, G. P. S. (2004). Prediction of CTL epitopes using QM, SVM and ANN techniques. In *Vaccine* (Vol. 22, Issues 23–24, pp. 3195–3204). <https://doi.org/10.1016/j.vaccine.2004.02.005>
- Dinesh Kumar, M., Godvin Sharmila, V., Kumar, G., Park, J.-H., Al-Qaradawi, S. Y., & Rajesh Banu, J. (2022). Surfactant induced microwave disintegration for enhanced biohydrogen production from macroalgae biomass: Thermodynamics and energetics. *Bioresource Technology*, 350, 126904.
- Garg, A., & Gupta, D. (2008). VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. In *BMC Bioinformatics* (Vol. 9, Issue 1). <https://doi.org/10.1186/1471-2105-9-62>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., 'everine, Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). <https://doi.org/10.1385/1-59259-890-0:571>
- Heath, L. S., & Ramakrishnan, N. (2010). Problem Solving Handbook in Computational Biology and Bioinformatics. Springer Science & Business Media.
- Kanguane, P. (2009). *Bioinformation Discovery: Data to Knowledge in Biology*. Springer Science & Business Media.
- Kaur, H., Singh, V., Kalia, M., Mohan, B., & Taneja, N. (2021). Identification and functional annotation of hypothetical proteins of uropathogenic strain CFT073 towards designing antimicrobial drug targets. *Journal of Biomolecular Structure & Dynamics*, 1–12.
- Kumar, J. A., Aravind Kumar, J., Sathish, S., Krithiga, T., Praveenkumar, T. R., Lokesh, S., Prabu, D., Annam Renita, A., Prakash, P., & Rajasimman, M. (2022). A comprehensive review on bio-hydrogen production from brewery industrial wastewater and its treatment methodologies. In *Fuel* (Vol. 319, p. 123594). <https://doi.org/10.1016/j.fuel.2022.123594>
- Mahesh, N., Balakumar, S., Danya, U., Shyamalagowri, S., Babu, P. S., Aravind, J., Kamaraj, M., & Govarthan, M. (2022). A review on mitigation of emerging contaminants in an aqueous environment using microbial bio-machines as sustainable tools: Progress and limitations. In *Journal of Water Process Engineering* (Vol. 47, p. 102712). <https://doi.org/10.1016/j.jwpe.2022.102712>
- Mohanavel, V., Kumar, K. R., Sathish, T., Velmurugan, P., Karthick, A., Ravichandran, M., Alfarraj, S., Almoallim, H. S., Sureshkumar, S., & JoshuaRamesh Lalvani, J. I. (2022). Investigation on Inorganic Salts K₂TiF₆ and KBF₄ to Develop Nanoparticles Based TiB₂ Reinforcement Aluminium Composites. *Bioinorganic Chemistry and Applications*, 2022, 8559402.
- Mota, M. M., & Rodriguez, A. (2017). *Malaria: Immune Response to Infection and Vaccination*. Springer.
- Ouattara, A., Barry, A. E., Dutta, S., Remarque, E. J., Beeson, J. G., & Plowe, C. V. (2015). Designing malaria vaccines to circumvent antigen variability. In *Vaccine* (Vol. 33, Issue 52, pp. 7506–7512). <https://doi.org/10.1016/j.vaccine.2015.09.110>
- Ram, G. D., Dinesh Ram, G., Praveen Kumar, S., Yuvaraj, T., Babu, T. S., & Balasubramanian, K. (2022). Simulation and investigation of MEMS bilayer solar energy harvester for smart wireless sensor applications. In *Sustainable Energy Technologies and Assessments* (Vol. 52, p. 102102). <https://doi.org/10.1016/j.seta.2022.102102>
- Rinesh, S., Maheswari, K., Arthi, B., Sherubha, P.,

- Vijay, A., Sridhar, S., Rajendran, T., & Waji, Y. A. (2022). Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms. *Journal of Healthcare Engineering*, 2022, 2761847.
- Saha, S., & Raghava, G. P. S. (2006). VICMpred: An SVM-based Method for the Prediction of Functional Proteins of Gram-negative Bacteria Using Amino Acid Patterns and Composition. In *Genomics, Proteomics & Bioinformatics* (Vol. 4, Issue 1, pp. 42–47). [https://doi.org/10.1016/s1672-0229\(06\)60015-6](https://doi.org/10.1016/s1672-0229(06)60015-6)
- Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of Immunology Research*, 2017, 2680160.
- Sathish, T., Mohanavel, V., Arunkumar, M., Rajan, K., Soudagar, M. E. M., Mujtaba, M. A., Salmen, S. H., Al Obaid, S., Fayaz, H., & Sivakumar, S. (2022). Utilization of Azadirachta indica biodiesel, ethanol and diesel blends for diesel engine applications with engine emission profile. In *Fuel* (Vol. 319, p. 123798). <https://doi.org/10.1016/j.fuel.2022.123798>
- Shanmugham, B., & Pan, A. (2013). Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach. *PloS One*, 8(3), e59126.
- Sintchenko, V. (2009). *Infectious Disease Informatics*. Springer Science & Business Media.
- Sudhan, M. B., Sinthuja, M., Pravinth Raja, S., Amutharaj, J., Charlyn Pushpa Latha, G., Sheeba Rachel, S., Anitha, T., Rajendran, T., & Waji, Y. A. (2022). Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model. *Journal of Healthcare Engineering*, 2022, 1601354.
- Sundararaman, S., Aravind Kumar, J., Devisigamani, P., & Devarajan, Y. (2022). Emerging pharma residue contaminants: Occurrence, monitoring, risk and fate assessment – A challenge to water resource management. In *Science of The Total Environment* (Vol. 825, p. 153897). <https://doi.org/10.1016/j.scitotenv.2022.153897>
- Tusnady, G. E., & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. In *Bioinformatics* (Vol. 17, Issue 9, pp. 849–850). <https://doi.org/10.1093/bioinformatics/17.9.849>
- Unemo, M. (2015). Current and future antimicrobial treatment of gonorrhoea - the rapidly evolving Neisseria gonorrhoeae continues to challenge. *BMC Infectious Diseases*, 15, 364.
- Vijayalakshmi, V. J., Arumugam, P., Ananthi Christy, A., & Brindha, R. (2022). Simultaneous allocation of EV charging stations and renewable energy sources: An Elite RERNN-m2MPA approach. In *International Journal of Energy Research* (Vol. 46, Issue 7, pp. 9020–9040). <https://doi.org/10.1002/er.7780>
- Website. (n.d.-a). 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>
- Website. (n.d.-b). Tusnady, G. E., and I. Simon. 2001. “The HMMTOP Transmembrane Topology Prediction Server.” *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/17.9.849>
- Yaashikaa, P. R., Kumar, P. S., Jeevanantham, S., & Saravanan, R. (2022). A review on bioremediation approach for heavy metal detoxification and accumulation in plants. *Environmental Pollution*, 301, 119035.
- Yu, C.-S., Cheng, C.-W., Su, W.-C., Chang, K.-C., Huang, S.-W., Hwang, J.-K., & Lu, C.-H. (2014). CELLO2GO: a web server for protein subCELLular LOCALization prediction with functional gene ontology annotation. *PloS One*, 9(6), e99368.
- Preethi, P. S., Hariharan, N. M., Vickram, S., Manian, R., Manikandan, S., Subbaiya, R., ... & Awasthi, M. K. (2022). Advances in bioremediation of emerging contaminants from industrial wastewater by oxidoreductase enzymes. *Bioresource Technology*, 127444.

List of Tables

Table 1. Retrieval of hypothetical protein sequences from NCBI

| S.No | Accession Number | Protein |
|-------------|-------------------------|----------------|
| 1. | TND31022.1 | Hypothetical |
| 2. | TND72443.1 | Hypothetical |
| 3. | TNE11935.1 | Hypothetical |
| 4. | TND85735.1 | Hypothetical |
| 5. | TND83855.1 | Hypothetical |
| 6. | TND83585.1 | Hypothetical |
| 7. | TND82887.1 | Hypothetical |
| 8. | TND 82552.1 | Hypothetical |
| 9. | TND82389.1 | Hypothetical |
| 10. | TND81549.1 | Hypothetical |
| 11. | TND81364.1 | Hypothetical |
| 12. | TND 81197.1 | Hypothetical |
| 13. | TND80996.1 | Hypothetical |
| 14. | TND80889.1 | Hypothetical |
| 15. | TND78182.1 | Hypothetical |
| 16. | TND 77224.1 | Hypothetical |
| 17. | TND 76516.1 | Hypothetical |
| 18. | TND76318.1 | Hypothetical |
| 19. | TND76046.1 | Hypothetical |

| | | |
|-----|------------|--------------|
| 20. | TND75975.1 | Hypothetical |
| 21. | TND74566.1 | Hypothetical |
| 22. | TND74456.1 | Hypothetical |
| 23. | TND74320.1 | Hypothetical |
| 24. | TND74264.1 | Hypothetical |
| 25. | TND74186.1 | Hypothetical |
| 26. | TND72406.1 | Hypothetical |
| 27. | TND70760.1 | Hypothetical |
| 28. | TND70162.1 | Hypothetical |
| 29. | TND70113.1 | Hypothetical |
| 30. | TND69367.1 | Hypothetical |
| 31. | TND67982.1 | Hypothetical |
| 32. | TND67902.1 | Hypothetical |
| 33. | TND67835.1 | Hypothetical |
| 34. | TND67641.1 | Hypothetical |
| 35. | TND67622.1 | Hypothetical |
| 36. | TND67320.1 | Hypothetical |
| 37. | TND65560.1 | Hypothetical |
| 38. | TND64513.1 | Hypothetical |
| 39. | TND64292.1 | Hypothetical |

| | | |
|-----|------------|--------------|
| 40. | TND63482.1 | Hypothetical |
| 41. | TND61983.1 | Hypothetical |
| 42. | TND61836.1 | Hypothetical |
| 43. | TND61815.1 | Hypothetical |
| 44. | TND59612.1 | Hypothetical |
| 45. | TND58874.1 | Hypothetical |
| 46. | TND58694.1 | Hypothetical |
| 47. | TND50235.1 | Hypothetical |
| 48. | TND48128.1 | Hypothetical |
| 49. | TND48028.1 | Hypothetical |
| 50. | TND47662.1 | Hypothetical |

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

| S.No | Protein ID | Molecular weight | PI | GRAVY | Instability index | Aliphatic index | Extinction coefficient |
|------|------------|------------------|------|--------|-------------------|-----------------|------------------------|
| 1. | TND31022.1 | 12191.25 | 9.60 | -0.636 | 37.24 | 59.82 | 11375 |
| 2. | TND72443.1 | 10734.49 | 8.66 | -0.303 | 48.59 | 88.19 | 15595 |
| 3. | TNE11935.1 | 10445.06 | 8.24 | -0.286 | 61.97 | 89.89 | 20970 |
| 4. | TND85735.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 5. | TND83855.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 6. | TND83585.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 7. | TND82887.1 | 14484.82 | 9.72 | -0.196 | 21.26 | 76.59 | 18115 |
| 8. | TND82552.1 | 22182.18 | 9.69 | -0.299 | 23.34 | 66.58 | 39420 |
| 9. | TND82389.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 10. | TND81549.1 | 22154.17 | 9.78 | -0.285 | 23.09 | 66.58 | 39420 |
| 11. | TND81364.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |

| | | | | | | | |
|-----|------------|----------|------|--------|-------|--------|-------|
| 12. | TND81197.1 | 14484.82 | 9.72 | -0.916 | 21.26 | 76.59 | 18115 |
| 13. | TND80996.1 | 22182.18 | 9.69 | -0.299 | 23.34 | 66.58 | 39420 |
| 14. | TND80889.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 15. | TND78182.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 16. | TND77224.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 17. | TND76516.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 18. | TND76318.1 | 14542.86 | 9.55 | -0.219 | 20.34 | 76.59 | 18115 |
| 19. | TND76046.1 | 22182.18 | 9.69 | -0.299 | 23.34 | 66.58 | 39420 |
| 20. | TND75975.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 21. | TND74566.1 | 14484.82 | 9.72 | -0.196 | 21.26 | 76.59 | 18115 |
| 22. | TND74456.1 | 22154.17 | 9.78 | -0.416 | 23.09 | 66.58 | 39420 |
| 23. | TND74320.1 | 16615.07 | 9.30 | -0.196 | 68.92 | 74.97 | 10220 |
| 24. | TND74264.1 | 22182.18 | 9.69 | -0.285 | 23.34 | 66.58 | 39420 |
| 25. | TND74186.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 26. | TND72406.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 27. | TND70760.1 | 22301.35 | 9.78 | -0.270 | 23.02 | 66.25 | 39420 |
| 28. | TND70162.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 29. | TND70113.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 30. | TND69367.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 31. | TND67982.1 | 22182.18 | 9.69 | -0.299 | 23.34 | 66.58 | 39420 |
| 32. | TND67902.1 | 14484.82 | 9.72 | -0.196 | 21.26 | 76.59 | 18115 |
| 33. | TND67835.1 | 22154.17 | 9.78 | -0.285 | 23.09 | 66.58 | 39420 |
| 34. | TND67641.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 35. | TND67622.1 | 14466.79 | 9.72 | -0.177 | 20.63 | 79.48 | 18115 |

| | | | | | | | |
|-----|------------|----------|-------|--------|-------|--------|-------|
| 36. | TND67320.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 37. | TND65560.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 38. | TND64513.1 | 22154.17 | 9.78 | -0.285 | 23.09 | 66.58 | 39420 |
| 39. | TND64292.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 40. | TND63482.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 41. | TND61983.1 | 14426.78 | 9.88 | -0.157 | 21.26 | 77.33 | 18115 |
| 42. | TND61836.1 | 22182.18 | 9.69 | -0.299 | 23.34 | 66.58 | 39420 |
| 43. | TND61815.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 44. | TND59612.1 | 9603.18 | 10.93 | -0.699 | 57.87 | 74.57 | 19730 |
| 45. | TND58874.1 | 22182.18 | 9.69 | -0.299 | 23.34 | 66.58 | 39420 |
| 46. | TND58694.1 | 16615.07 | 9.30 | -0.416 | 68.92 | 75.97 | 10220 |
| 47. | TND50235.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 48. | TND48128.1 | 8599.77 | 4.62 | -0.359 | 20.50 | 101.35 | 11460 |
| 49. | TND48028.1 | 14542.86 | 9.55 | -0.219 | 20.34 | 76.59 | 18115 |
| 50. | TND47662.1 | 22182.18 | 9.69 | -0.299 | 20.34 | 66.58 | 39420 |

Table 3. Functional classification of hypothetical proteins using VICMpred.

| S.No | Accession Number | Functional Class | Score |
|------|------------------|------------------|-------------|
| 1. | TND31022.1 | Cellular Process | -0.44186837 |
| 2. | TND72443.1 | Cellular Process | 1.3151377 |
| 3. | TNE11935.1 | Cellular Process | 1.440621 |
| 4. | TND85735.1 | Cellular Process | 1.3820883 |

| | | | |
|-----|------------|---------------------|-------------|
| 5. | TND83855.1 | Cellular Process | 1.3820883 |
| 6. | TND83585.1 | Cellular Process | 1.3820883 |
| 7. | TND82887.1 | Metabolism Molecule | 1.2941522 |
| 8. | TND82552.1 | Cellular Process | 1.5539649 |
| 9. | TND82389.1 | Cellular Process | 0.4397392 |
| 10. | TND81549.1 | Cellular Process | 0.20524353 |
| 11. | TND81364.1 | Cellular Process | 0.4397392 |
| 12. | TND81197.1 | Metabolism Molecule | 1.2941522 |
| 13. | TND80996.1 | Cellular Process | 0.15539649 |
| 14. | TND80889.1 | Cellular Process | 0.4397392 |
| 15. | TND78182.1 | Cellular Process | 1.3820883 |
| 16. | TND77224.1 | Cellular Process | 1.3820883 |
| 17. | TND76516.1 | Cellular Process | 1.3820883 |
| 18. | TND76318.1 | Metabolism Molecule | -0.82011318 |
| 19. | TND76046.1 | Cellular Process | 0.15539649 |
| 20. | TND75975.1 | Cellular Process | 0.4397392 |
| 21. | TND74566.1 | Metabolism Molecule | 1.2941522 |
| 22. | TND74456.1 | Cellular Process | 0.20524353 |
| 23. | TND74320.1 | Cellular Process | 0.4397392 |
| 24. | TND74264.1 | Cellular Process | 0.15539649 |

| | | | |
|-----|------------|---------------------|-------------|
| 25. | TND74186.1 | Cellular Process | 0.4397392 |
| 26. | TND72406.1 | Cellular Process | 1.3820883 |
| 27. | TND70760.1 | Cellular Process | 0.20524353 |
| 28. | TND70162.1 | Cellular Process | 0.4397392 |
| 29. | TND70113.1 | Cellular Process | 1.3820883 |
| 30. | TND69367.1 | Cellular Process | 1.3820883 |
| 31. | TND67982.1 | Cellular Process | 0.15539649 |
| 32. | TND67902.1 | Metabolism Molecule | 1.2941522 |
| 33. | TND67835.1 | Cellular Process | 0.20524353 |
| 34. | TND67641.1 | Cellular Process | 0.4397392 |
| 35. | TND67622.1 | Metabolism Molecule | -0.80065533 |
| 36. | TND67320.1 | Cellular Process | 0.4397392 |
| 37. | TND65560.1 | Cellular Process | 1.3820883 |
| 38. | TND64513.1 | Cellular Process | 0.20524353 |
| 39. | TND64292.1 | Cellular Process | 0.4397392 |
| 40. | TND63482.1 | Cellular Process | 1.3820883 |
| 41. | TND61983.1 | Metabolism Molecule | 1.6554162 |
| 42. | TND61836.1 | Cellular Process | 0.15539649 |
| 43. | TND61815.1 | Cellular Process | 0.4397392 |
| 44. | TND59612.1 | Cellular Process | 1.2447075 |

| | | | |
|-----|------------|---------------------|-------------|
| 45. | TND58874.1 | Cellular Process | 0.15539649 |
| 46. | TND58694.1 | Cellular Process | 0.4397392 |
| 47. | TND50235.1 | Cellular Process | 1.3820883 |
| 48. | TND48128.1 | Cellular Process | 1.3820883 |
| 49. | TND48028.1 | Metabolism Molecule | -0.82011318 |
| 50. | TND47662.1 | Cellular Process | 0.15539649 |

Table 4. Subcellular localization prediction for the proteins using CELLO2GO

| S.No | Accession Number | Localization |
|------|------------------|----------------------------|
| 1. | TND31022.1 | Cytoplasmic |
| 2. | TND72443.1 | Inner Membrane,Cytoplasmic |
| 3. | TNE11935.1 | Inner Membrane |
| 4. | TND85735.1 | Inner Membrane,Cytoplasmic |
| 5. | TND83855.1 | Inner Membrane,Cytoplasmic |
| 6. | TND83585.1 | Inner Membrane,Cytoplasmic |
| 7. | TND82887.1 | Periplasmic |
| 8. | TND82552.1 | Outer Membrane |
| 9. | TND82389.1 | Cytoplasmic |
| 10. | TND81549.1 | Outer Membrane |
| 11. | TND81364.1 | Cytoplasmic |
| 12. | TND81197.1 | Periplasmic |
| 13. | TND80996.1 | Outer Membrane |
| 14. | TND80889.1 | Cytoplasmic |
| 15. | TND78182.1 | Inner Membrane,Cytoplasmic |
| 16. | TND77224.1 | Inner Membrane,Cytoplasmic |
| 17. | TND76516.1 | Inner Membrane,Cytoplasmic |
| 18. | TND76318.1 | Periplasmic |
| 19. | TND76046.1 | Outer Membrane |

| | | |
|-----|------------|----------------------------|
| 20 | TND75975.1 | Cytoplasmic |
| 21. | TND74566.1 | Periplasmic |
| 22. | TND74456.1 | Outer Membrane |
| 23. | TND74320.1 | Cytoplasmic |
| 24. | TND74264.1 | Outer Membrane |
| 25. | TND74186.1 | Cytoplasmic |
| 26. | TND72406.1 | Inner Membrane,Cytoplasmic |
| 27. | TND70760.1 | Outer Membrane |
| 28. | TND70162.1 | Cytoplasmic |
| 29. | TND70113.1 | Inner Membrane,Cytoplasmic |
| 30. | TND69367.1 | Inner Membrane,Cytoplasmic |
| 31. | TND67982.1 | Outer Membrane |
| 32. | TND67902.1 | Periplasmic |
| 33. | TND67835.1 | Outer Membrane |
| 34. | TND67641.1 | Cytoplasmic |
| 35. | TND67622.1 | Periplasmic |
| 36. | TND67320.1 | Cytoplasmic |
| 37. | TND65560.1 | Inner Membrane,Cytoplasmic |
| 38. | TND64513.1 | Outer Membrane |
| 39. | TND64292.1 | Cytoplasmic |
| 40. | TND63482.1 | Inner Membrane,Cytoplasmic |
| 41. | TND61983.1 | Periplasmic |
| 42. | TND61836.1 | Outer Membrane |
| 43. | TND61815.1 | Cytoplasmic |
| 44. | TND59612.1 | Cytoplasmic |
| 45. | TND58874.1 | Outer Membrane |
| 46. | TND58694.1 | Cytoplasmic |
| 47. | TND50235.1 | Inner Membrane,Cytoplasmic |
| 48. | TND48128.1 | Inner Membrane,Cytoplasmic |
| 49. | TND48028.1 | Periplasmic |

| | | |
|-----|------------|----------------|
| 50. | TND47662.1 | Outer Membrane |
|-----|------------|----------------|

Table 5. Virulence prediction for the proteins using Virulentpred tool

| S.No. | Accession Number | Virulent/Non virulent | Score |
|-------|------------------|-----------------------|--------|
| 1. | TND31022.1 | Virulent | 0.9633 |
| 2. | TND72443.1 | Non-Virulent | -0.656 |
| 3. | TNE11935.1 | Virulent | 1.0240 |
| 4. | TND85735.1 | Virulent | 1.0981 |
| 5. | TND83855.1 | Virulent | 1.0981 |
| 6. | TND83585.1 | Virulent | 1.0981 |
| 7. | TND82887.1 | Non-Virulent | -0.284 |
| 8. | TND82552.1 | Virulent | 0.7117 |
| 9. | TND82389.1 | Virulent | 1.0443 |
| 10. | TND81549.1 | Virulent | 0.6502 |
| 11. | TND81364.1 | Virulent | 1.0443 |
| 12. | TND81197.1 | Non-Virulent | -0.284 |
| 13. | TND80996.1 | Virulent | 0.7117 |
| 14. | TND80889.1 | Virulent | 1.0443 |
| 15. | TND78182.1 | Virulent | 1.0981 |
| 16. | TND77224.1 | Virulent | 1.0981 |
| 17. | TND76516.1 | Virulent | 1.0981 |
| 18. | TND76318.1 | Non-Virulent | -0.644 |
| 19. | TND76046.1 | Virulent | 0.7117 |
| 20. | TND75975.1 | Virulent | 1.0443 |
| 21. | TND74566.1 | Non-Virulent | -0.284 |
| 22. | TND74456.1 | Virulent | 0.6502 |

| | | | |
|-----|------------|--------------|--------|
| 23. | TND74320.1 | Virulent | 1.0443 |
| 24. | TND74264.1 | Virulent | 0.7117 |
| 25. | TND74186.1 | Virulent | 1.0443 |
| 26. | TND72406.1 | Virulent | 1.0981 |
| 27. | TND70760.1 | Virulent | 0.6502 |
| 28. | TND70162.1 | Virulent | 1.0443 |
| 29. | TND70113.1 | Virulent | 1.0981 |
| 30. | TND69367.1 | Virulent | 1.0981 |
| 31. | TND67982.1 | Virulent | 0.7117 |
| 32. | TND67902.1 | Non-Virulent | -0.284 |
| 33. | TND67835.1 | Virulent | 0.6502 |
| 34. | TND67641.1 | Virulent | 1.0443 |
| 35. | TND67622.1 | Non-Virulent | -0.541 |
| 36. | TND67320.1 | Virulent | 1.0443 |
| 37. | TND65560.1 | Virulent | 1.0981 |
| 38. | TND64513.1 | Virulent | 0.6502 |
| 39. | TND64292.1 | Virulent | 1.0443 |
| 40. | TND63482.1 | Virulent | 1.0981 |
| 41. | TND61983.1 | Non-Virulent | -0.220 |
| 42. | TND61836.1 | Virulent | 0.7117 |
| 43. | TND61815.1 | Virulent | 1.0443 |
| 44. | TND59612.1 | Virulent | 1.0552 |
| 45. | TND58874.1 | Virulent | 0.7117 |
| 46. | TND58694.1 | Virulent | 1.0443 |

| | | | |
|-----|------------|--------------|--------|
| 47. | TND50235.1 | Virulent | 1.0981 |
| 48. | TND48128.1 | Virulent | 1.0981 |
| 49. | TND48028.1 | Non-Virulent | -0.644 |
| 50. | TND47662.1 | Virulent | 0.7117 |

Table 6. Evaluating the antigenicity of the proteins using VaxiJen ver. 2.0

| S.No | Accession Number | Antigenicity Score |
|------|------------------|--------------------|
| 1. | TND31022.1 | 0.4710 |
| 2. | TND72443.1 | 0.2110 |
| 3. | TNE11935.1 | 0.2728 |
| 4. | TND85735.1 | 0.3409 |
| 5. | TND83855.1 | 0.3409 |
| 6. | TND83585.1 | 0.3409 |
| 7. | TND82887.1 | 0.4759 |
| 8. | TND82552.1 | 0.3877 |
| 9. | TND82389.1 | 0.5063 |
| 10. | TND81549.1 | 0.3795 |
| 11. | TND81364.1 | 0.5063 |
| 12. | TND81197.1 | 0.4759 |
| 13. | TND80996.1 | 0.3828 |
| 14. | TND80889.1 | 0.5063 |
| 15. | TND78182.1 | 0.3409 |
| 16. | TND77224.1 | 0.3409 |
| 17. | TND76516.1 | 0.3409 |
| 18. | TND76318.1 | 0.4775 |

| | | |
|-----|------------|--------|
| 19. | TND76046.1 | 0.3877 |
| 20. | TND75975.1 | 0.5063 |
| 21. | TND74566.1 | 0.4759 |
| 22. | TND74456.1 | 0.3795 |
| 23. | TND74320.1 | 0.5063 |
| 24. | TND74264.1 | 0.3877 |
| 25. | TND74186.1 | 0.5063 |
| 26. | TND72406.1 | 0.3409 |
| 27. | TND70760.1 | 0.3795 |
| 28. | TND70162.1 | 0.5063 |
| 29. | TND70113.1 | 0.3409 |
| 30. | TND69367.1 | 0.3409 |
| 31. | TND67982.1 | 0.3877 |
| 32. | TND67902.1 | 0.4759 |
| 33. | TND67835.1 | 0.3795 |
| 34. | TND67641.1 | 0.5063 |
| 35. | TND67622.1 | 0.4631 |
| 36. | TND67320.1 | 0.5063 |
| 37. | TND65560.1 | 0.3409 |
| 38. | TND64513.1 | 0.3795 |
| 39. | TND64292.1 | 0.5063 |
| 40. | TND63482.1 | 0.3409 |
| 41. | TND61983.1 | 0.4792 |
| 42. | TND61836.1 | 0.3877 |

| | | |
|-----|------------|--------|
| 43. | TND61815.1 | 0.5063 |
| 44. | TND59612.1 | 0.5832 |
| 45. | TND58874.1 | 0.3877 |
| 46. | TND58694.1 | 0.5063 |
| 47. | TND50235.1 | 0.3409 |
| 48. | TND48128.1 | 0.3409 |
| 49. | TND48028.1 | 0.4775 |
| 50. | TND47662.1 | 0.3877 |

Table 7. Epitope prediction using ABCpred and VaxiJen

| Accession Number | B cell epitope | ABC Pred | Vaxijen | AllergenFP |
|------------------|------------------|----------|--------------|--------------|
| TND82389.1 | ATAYIPPNDFQPNCDI | 0.94 | 0.6862 | Non allergen |
| | YVESRYHSSMDFAVDE | 0.93 | 0.7389 | Allergen |
| | VVEIISDVFNREAR | 0.91 | 0.2618 | Non allergen |
| | SERIAGNRRHEAEMPL | 0.87 | 0.7465 | Non allergen |
| | EMPLPAPCRFAKPAAS | 0.86 | 0.3267 | Non allergen |
| | ALPRTEINEPHNEISS | 0.84 | 0.3236 | Non allergen |
| | CQLSHAATAYIPPNDF | 0.83 | 0.3481 | Non allergen |
| | RNEARDYVESRYHSSM | 0.82 | 0.4032 | Allergen |
| | ARKVMHSEHSRRRSV | 0.81 | 0.9817 | Non allergen |
| | PTHETICALPRTEINE | 0.80 | -0.0904 | Non allergen |
| | RRLGLTQGQHNELRKI | 0.78 | 0.5530 | Non allergen |
| | FKMAGDRARLKMVHSE | 0.74 | 0.9619 | Allergen |
| | VDELEIQHRFFHILTP | 0.72 | 0.9929 | Non allergen |
| | FFHILTPQQQMWLSS | 0.66 | 0.0307 | Allergen |
| | CRFAKPAASFLSMALL | 0.65 | 0.0751 | Non allergen |
| HSSMDFAVDELEIQHR | 0.56 | 0.9804 | Non allergen | |
| TND59612.1 | CSVWQSRRLMRQNVWR | 0.96 | -0.3362 | Non allergen |
| | HQCVLKPRFLDNKCRL | 0.94 | 0.4770 | Non allergen |
| | HEAEMSTKRTVYQLNP | 0.89 | 0.8923 | Non allergen |
| | SERIAGNRRHEAEMST | 0.89 | 1.0007 | Non allergen |
| | ALPRTEINEPHNEISS | 0.84 | 0.3236 | Non allergen |
| | AYCHWASRDKSPNLQN | 0.82 | 0.3177 | Allergen |
| | PTHETICALPRTEINE | 0.80 | -0.0904 | Non allergen |
| | NKCRLKPSDGICSVWQ | 0.72 | 0.5774 | Non allergen |
| | KRTVYQLNPFLSSGQA | 0.63 | 0.2719 | Non allergen |

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

| Accession Number | Peptide | Start | TC pred score | Antigenicity score | Allergenicity |
|------------------|-----------|-------|---------------|--------------------|---------------|
| TND82389.1 | ELRKIRAAF | 97 | 0.990 | 0.5473 | Allergen |
| | CDIRRLGLT | 83 | 0.960 | 0.1565 | Allergen |
| | AASFLSMAL | 53 | 0.950 | 0.2372 | Non-allergen |
| TND59612.1 | RLKPSDGIC | 94 | 0.990 | 1.7805 | Non-allergen |
| | NLSIRLHQC | 74 | 0.970 | 1.7753 | Allergen |
| | KPSDGICSV | 96 | 0.970 | 1.1334 | Non-allergen |

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

| Accession Number | TMHMM Score | HMMTOP Score |
|------------------|-------------|--------------|
| TND2389.1 | 0 | 0 |
| TND59612.1 | 0 | 0 |

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

| Accession Number | Non-human Homologous |
|------------------|---------------------------------|
| TND2389.1 | No significant similarity found |
| TND59612.1 | No significant similarity found |