IDENTIFICATION OF NOVEL VACCINE TARGETS FROM THE PROTEOME OF KLEBSIELLA PNEUMONIAE TO FIGHT AGAINST THE DEADLY INFECTION PNEUMONIA USING REVERSE VACCINOLOGY APPROACH

Moka Jaswanth Kumar[1], Abiramavalli Moorthy[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: This study intends to delve into the uncharacterized protein pool of Klebsiella pneumoniae to identify novel vaccine targets using computational methods.
Materials and Methods: A total of 50 hypothetical proteins of this organism were selected and retrieved from the NCBI database. Physicochemical properties of the selected 50 proteins were predicted using the PROTPARAM tool, followed by categorizing them based on their predicted functional class using the VICMpred tool. Virulence of these proteins were analyzed using the virulentpred tool. Localization of these proteins were studied using PSORT. The antigenicity of the selected virulent proteins were predicted using VaxiJen v2.0 followed by detection of B-Cell epitopes found in the antigenic proteins using ABCpred. The scores of both the tools were compared and concluded.
Results: Out of 50, 28 proteins were found to be virulent and among the 28 virulent proteins, 14 were localized in the cytoplasm, 11 were found to be present in the inner membrane and 2 were from periplasm while only one was found to be from the outer membrane. Among the 11 inner membrane virulent proteins, 4 were found to be an antigen while the one extracellular membrane protein also had antigenic properties. Thus these 5 antigenic proteins were further subjected to epitope prediction which revealed that they possessed B-cell and T-cell epitopes. **Conclusion:** Based on their virulence scores, antigenicity properties, and presence of epitopes, we conclude that these 5 proteins can be considered as potential drug and vaccine targets with further validation through in-vitro and in-vivo immunology experiments.

Keywords: Pneumonia, Pathogen, Novel Vaccine Targets, Virulence, Antigenicity, Epitopes, Bioinformatics tools.

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

1. Introduction

Klebsiella pneumoniae is a gram-negative human specific opportunistic pathogen instigating nosocomial infections(Fok et al. 1998). Klebsiella pneumoniae causes a wide range of infection which includes the urinary tract infection, pneumonia and liver abscess(Paczosa and Peruchini 2013). The WHO has categorized Klebsiella pneumoniae as a serious threat as they are resistant towards most of the antibiotics by adapting themselves according to the environment and also due to the rise and spread of hypervirulent strains. Even the highest class of antibiotics; carbapenems have failed to combat this bacterial pathogen permitting them to multiply their antibiotic-resistant varieties(Tusnady and Simon 2001). Despite the importance of this pathogen, its virulence determinants remain largely understudied . Almost all the antibiotics prescribed for this infection target the candidate genes involved in the crucial cellular pathways of the bacteria (Petchiappan, Naik, and Chatterji 2020)nevertheless these molecules of the pathogen are no longer susceptible to the antibiotics, leading to high mortality rates among patients. This has further led to a demand for novel therapeutic interventions to treat such infection.

The number of research articles published since 2017 on identifying novel molecular targets using in-silico approaches are nearly 40,000 in Google scholar (Koski 2001). Many publications that focus on various strategies to identify new drug and vaccine targets for pneumonia infection (Petchiappan, Naik, and Chatterji 2020) . For instance, a research group has utilized an in-silico approach to identify drug and vaccine targets from hypothetical proteins of S. pneumoniae(Paczosa and Peruchini 2013). Computational subtractive genomics approach has been employed for predicting the targets concluded with 2 proteins that can be used as potent drug targets (Mbaeyi et al. 2018). Another study on functional annotation of conserved hypothetical proteins from Haemophilus influenza has resulted in functional classification of these unexplored proteins with the help of advanced bioinformatics tools and databases (Koski 2001; Petchiappan, Naik, and Chatterji 2020).

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). The lacunae in the existing research is, the complete genome of K. pneumoniae, is publicly available and most of its protein coding genes are well-studied, yet approximately 30% of the genes is not characterized and annotated as hypothetical proteins (HPs)(Sundaravaradan et al. 2010). The functions of HPs in pathogens, their characterization has got much attention. They perform crucial roles in resistance development, host adaptation, induction of chemotaxis and treatment in wound healing. HPs are also involved in pathogens triggered disease development. Thus, the present study aims at identifying such novel drug/vaccine targets from the hypothetical proteins. They were retrieved from the NCBI database with the help of the keyword “hypothetical protein of K.pneumoniae”. They were further characterized using bioinformatics tools for prediction of subcellular localization, prediction of transmembrane helices, virulence prediction, antigenicity prediction and epitope prediction were done.

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. The number of groups is 1 for each organism. The sample size is 50 proteins per group. Uncharacterized proteins were searched in the National Center for Biotechnology information (NCBI) database by typing the keywords “hypothetical proteins Klebsiella pneumoniae and a total of 50 protein sequences of hypothetical proteins of Klebsiella pneumoniae were selected for this study. The protein sequences were selected based on the criteria that their function was unknown hence hypothetical. The proteins were then characterized by several bioinformatics tools including ProtParam, PSORT, SignalP, TMHMM, tBLASTn, and VaxiJen. For immunoinformatics, ABCpred was used for B cell epitope prediction while CTLpred was employed for T cell epitope prediction.

About 50 hypothetical genes were retrieved and sequences with virulence potential were chosen by using the VOCM tool. Further, proteins without virulence and properties were eliminated. No statistical tools were used in this study. However, statistical values assigned by the tools called antigenicity score of > 0.5, and virulence score of > 1 determine the sequences as potential candidates.

3. Results

Table 1 provides the NCBI-retrieved putative protein sequences. The physico-chemical

parameters of the returned protein sequences predicted using PROTPARAM tools are shown in Table 2. The functional classification of hypothetical proteins discovered with the VICMpred tool is shown in Table 3. The subcellular localization of all proteins predicted using the CELLO2GO tool is shown in Table 4. The antigenicity features of the proteins discovered using the VaxiJen 2.0 programme are shown in Table 5. Using the Virulent Pred tool, Table 6 shows the pathogenicity score of the putative proteins. The presence of epitopes in the proteins investigated with the ABCpred and VaxiJen tools is shown in Table 7. Using the CTLpred tool, Table 8 demonstrates the existence of T-cell epitopes in the uncharacterized protein pool. The number of transmembrane helices discovered in the two putative antigenic proteins evaluated with the HMMTOP tool is shown in Table 9. The tBLASTn results of the two putative novel antigenic proteins are shown in Table 10, demonstrating that they are similar to any human proteins. These findings lead us to believe that these proteins could be used to develop drugs to combat Mycobacterium TB.

4. Discussion

The clinical pathogen Klebsiella pneumoniae is a causative agent of several nosocomial infections. It is a multidrug resistant pathogen, resulting in high mortality rates among patients which has further led to a demand for novel therapeutic intervention to treat such infections. Using a series of in-silico analyses, the present study has attempted to explore novel vaccine targets from the hypothetical proteins of K. pneumoniae (Fok et al. 1998).

About 195 studies related to in-silico characterization of hypothetical proteins of various pathogens were found in Pubmed in the last 5 years. The computational tools they have utilized for studying the physicochemical properties and functional classification were ProtParam, and CELLO2GO respectively(Al Sadah et al. 2021)(Fok et al. 1998; Mbaeyi et al. 2018). The present investigation mainly focused only on identifying virulence factors among these hypothetical proteins. From 50 hypothetical proteins only 5 proteins were found to be the most suitable candidates with virulence and antigenic potentials and these properties may help in the pathogen's survival. These observations were in concordance with previous reports on detection of virulence factors from uncharacterized protein pool using bioinformatics and immunoinformatics tools in several organisms (Huang and Lilley 2020) (Bhasin and Raghava 2004). Therefore these 5 antigens could be novel vaccine targets to combat pneumonia and other nosocomial infections.

The major limitation of this study is, this is a preliminary study focusing on identifying the potential candidate proteins with vaccine potential using a computational workflow whereas, if they are to be used for future vaccine development against pneumonia, they have to be validated through experimental methods in-vitro and in-vivo methods. Hence, the future scope of this study is structure prediction for these 5 proteins which would aid in structure-based inhibitors designing and experimental validation through immunological techniques could be done(Kanchi et al. 2021).

5. Conclusion

The strategies and workflow used in our study to classify these hypothetical proteins functionally and to identify potential vaccine candidates can be useful for designing experimental approaches geared towards the novel vaccine/drug development against the pathogen. Here, we have studied the protein properties of the 50 hypothetical proteins from K.pneumoniae and categorized them into different protein classes. Among these 50 HPs, 5 proteins were found to be antigenic and most virulent with epitopes. Hence we conclude that these candidate proteins could be considered as novel vaccine targets and be further validated.

Declarations

Conflict of Interest

The authors of this paper declare no conflict of interest.

Author contribution

Author Jaswanth was involved in data collection, data analysis, manuscript writing. Author Dr. Abiramavalli 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. Qbiogene Chennai
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

6. References

Al Sadah, Zhra M., Fajar A. Aldulijan, Dana A. Aljamea, and Maan A. Saleh. 2021. "The Impact of Women Violence on Depression among Female Medical Students at Imam Abdulrahman Bin Faisal University, Saudi Arabia." Journal of Family Medicine and Primary Care 10 (12): 4565–72.

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.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. "Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics." Bioresource Technology 350 (April): 126904.

Fok, Tai Fai, Cheuk Hon Lee, Eric M. C. Wong, Donald J. Lyon, William Wong, Pak Cheong Ng, Kam Lau Cheung, and Augustine F. B. Cheng. 1998. "Risk Factors for Enterobacter Septicemia in a Neonatal Unit: Case-Control Study." Clinical Infectious Diseases. https://doi.org/10.1086/514974.

Huang, L., and D. M. J. Lilley. 2020. "Crystal Structure of the SAM-SAH Riboswitch with SAH." https://doi.org/10.2210/pdb6yl5/pdb.

Kanchi, Suvardhan, Myalowenkosi I. Sabela, Mohd Shahbaaz, and Krishna Bisetty. 2021. "Sensitivity Enhancement of Pre-Capillary Chelation Method for the Separation of Metal Ions: Experimental and DFT Study." Current Analytical Chemistry. https://doi.org/10.2174/1573411016666200108145109.

Koski, T. 2001. Hidden Markov Models for Bioinformatics. Springer Science & Business Media.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. "A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies." Fuel. https://doi.org/10.1016/j.fuel.2022.123594.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. "A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations." Journal of Water Process Engineering. https://doi.org/10.1016/j.jwpe.2022.102712.

Mbaeyi, Chukwuma, Zubair Mufti Wadood, Thomas Moran, MJourn, Fazal Ather, Tasha Stehling-Ariza, Joanna Nikulin, et al. 2018. "Strategic Response to an Outbreak of Circulating Vaccine-Derived Poliovirus Type 2 - Syria, 2017-2018." MMWR. Morbidity and Mortality Weekly Report 67 (24): 690–94.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. "Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites." Bioinorganic Chemistry and Applications 2022 (January): 8559402.

Paczosa, Scott, and Chuck Peruchini. 2013. "Selling Professional and Financial Services Handbook." https://doi.org/10.1002/9781118728208.

Petchiappan, Anushya, Sujay Y. Naik, and Dipankar Chatterji. 2020. "RelZ-Mediated Stress Response in Mycobacterium Smegmatis : pGpp Synthesis and Its Regulation." Journal of Bacteriology. https://doi.org/10.1128/jb.00444-19.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." Journal of Healthcare Engineering 2022 (February): 2761847.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile." Fuel. https://doi.org/10.1016/j.fuel.2022.123798.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S.

Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model." Journal of Healthcare Engineering 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. "Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management." Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2022.153897.

Sundaravaradan, Naren, K. S. Tozammel Hossain, Vandana Sreedharan, Douglas J. Slotta, John Paul C. Vergara, Lenwood S. Heath, and Naren Ramakrishnan. 2010. "Extracting Temporal Signatures for Comprehending Systems Biology Models." Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '10. https://doi.org/10.1145/1835804.1835863.

Tusnady, G. E., and I. Simon. 2001. "The HMMTOP Transmembrane Topology Prediction Server." Bioinformatics. https://doi.org/10.1093/bioinformatics/17.9.849.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. "Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach." International Journal of Energy Research. https://doi.org/10.1002/er.7780.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. "A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants." Environmental Pollution 301 (May): 119035.

List of Tables

Table 1. List of hypothetical proteins retrieved from NCBI

S.No	Accession Number	Protein
1.	CEK42940.1	Hypothetical
2.	QGW 60714.1	Hypothetical
3.	CDL57047.1	Hypothetical
4.	ODM39876.1	Hypothetical
5.	ODM34760.1	Hypothetical
6.	PLP12834.1	Hypothetical
7.	ODM41053.1	Hypothetical
8.	ODM 41051.1	Hypothetical
9.	ODM41025.1	Hypothetical
10.	ODM 40980.1	Hypothetical
11.	ODM 40914.1	Hypothetical
12.	ODM 40663.1	Hypothetical
13.	ODM 40510.1	Hypothetical
14.	ODM40436.1	Hypothetical
15.	ODM 40356.1	Hypothetical
16.	ODM 40241.1	Hypothetical
17.	ODM 40153.1	Hypothetical
18.	ODM 40064.1	Hypothetical
19.	ODM39954.1	Hypothetical

20.	ODM39942.1	Hypothetical
21.	ODM39924.1	Hypothetical
22.	ODM39923.1	Hypothetical
23.	ODM39920.1	Hypothetical
24.	ODM39862.1	Hypothetical
25.	ODM39714.1	Hypothetical
26.	ODM 39564.1	Hypothetical
27.	ODM39520.1	Hypothetical
28.	ODM39286.1	Hypothetical
29.	ODM39182.1	Hypothetical
30.	ODM39151.1	Hypothetical
31.	ODM38959.1	Hypothetical
32.	ODM 38375.1	Hypothetical
33.	ODM 38342.1	Hypothetical
34.	ODM 38315.1	Hypothetical
35.	ODM38280.1	Hypothetical
36.	ODM38268.1	Hypothetical
37.	ODM 38197.1	Hypothetical
38.	ODM 38190.1	Hypothetical
39.	ODM38155.1	Hypothetical
40.	ODM38087.1	Hypothetical
41.	ODM38064.1	Hypothetical
42.	ODM38057.1	Hypothetical
43.	ODM38056.1	Hypothetical
44.	ODM 37935.1	Hypothetical
45.	ODM 37919.1	Hypothetical
46.	ODM37834.1	Hypothetical
47.	ODM 37668.1	Hypothetical
48.	ODM 37640.1	Hypothetical
49.	ODM37515.1	Hypothetical
50.	ODM 37499.1	Hypothetical

Table 2. Physicochemical properties of the proteins predicted using PROTPARAM

S.No	Accession Number	Molecular Weight	PI	GRAVY	Instability index	Aliphatic index	Extinction coefficient
1	CEK42940.1	37938.43	9.03	0.012	31.34	106.01	30745
2	QGW 60714.1	18310.78	5.33	-0.147	50.58	83.55	25355

3	CDL57047.1	10395.51	9.4	0.532	43.49	127.37	12490
4	ODM39876.1	21460.91	4.26	-0.312	47.61	92.53	17085
5	ODM34760.1	33963.22	9.25	-0.415	29.51	73.61	33350
6	PLP12834.1	48632.56	5.18	0.017	28.7	95.65	10805
7	ODM41053.1	18804.24	9.14	-0.313	27.45	82.74	9065
8	ODM 41051.1	21076.02	6.19	-0.574	38.34	84.5	20190
9	ODM41025.1	24315.6	5.15	-0.346	41.65	88.94	9190
10	ODM 40980.1	8562.48	4.58	-0.603	44.98	75.6	17990
11	ODM 40914.1	34422.56	5.83	-0.267	52.67	92.59	82055
12	ODM 40663.1	9785.22	5.2	-0.163	45.79	98.39	5960
13	ODM 40510.1	11805.65	9.3	-0.072	42.17	81.55	12490
14	ODM40436.1	17100.52	5.52	-0.026	41.76	89.87	13075
15	ODM 40356.1	15586.22	9.52	0.027	26.29	105.66	11000
16	ODM 40241.1	19525.13	5.09	-0.121	37.72	102.32	8480
17	ODM 40153.1	9193.05	5.45	0.846	45.83	130.62	30605
18	ODM 40064.1	16924.33	9.42	-0.52	40.15	72.99	33585
19	ODM39954.1	82356.69	9.17	-0.515	40.52	74.77	104585
20	ODM39942.1	18637.24	6.97	0.616	37.27	122.99	23950
21	ODM39924.1	12818.49	4.86	-0.64	78.56	78.52	29115
22	ODM39923.1	26106.62	8.48	-0.431	38.05	89.83	17085
23	ODM39920.1	10294.05	9.25	0.097	38.77	111.77	1490
24	ODM39862.1	11742.14	4.67	-0.313	33.12	80.49	5960
25	ODM39714.1	14384.19	10.17	1.037	25.36	120.69	31970
26	ODM 39564.1	13623.43	6.9	-0.463	39.99	70	11920
27	ODM39520.1	48101.24	5.62	0.106	38.92	116.52	31400
28	ODM39286.1	60282.65	5.35	-0.157	51	96.73	83225
29	ODM39182.1	8033.91	4.97	-0.419	46.2	77.36	4470
30	ODM39151.1	25086.37	5.41	-0.376	24.85	78.12	57075
31	ODM38959.1	48005.78	5.45	0.126	31.01	95.91	45630
32	ODM 38375.1	8189.24	4.96	-0.147	59.48	101.47	1490

33	ODM 38342.1	24267.2	9.74	-0.028	55.31	102.99	51575
34	ODM 38315.1	7745.6	3.94	-0.486	-0.486	88.79	19480
35	ODM38280.1	34552.79	5.96	-0.217	45.23	102.15	23380
36	ODM38268.1	8212.47	6.38	-0.731	19.46	95.82	27500
37	ODM 38197.1	10193.38	4.05	-0.031	36.92	75.74	1865
38	ODM 38190.1	28230.69	9.79	0.87	46.24	125.87	19940
39	ODM38155.1	17100.11	9.94	-0.006	53.88	97.02	33460
40	ODM38087.1	12545.25	4.85	-0.138	49.84	87.28	12740
41	ODM38064.1	43157.08	10.09	0.511	34.42	118.44	124120
42	ODM38057.1	24702.94	9.62	1.051	25.6	132.64	22710
43	ODM38056.1	14627.85	9.18	1.067	45.83	145.38	29575
44	ODM 37935.1	13947.53	4.33	-0.431	45	79.58	23950
45	ODM 37919.1	50936.86	6.74	-0.074	46.84	107.79	31440
46	ODM37834.1	20136.09	7.85	0.924	19.05	116.33	25440
47	ODM 37668.1	7235.18	4.45	0.041	31.28	109.55	7575
48	ODM 37640.1	21946.46	9.14	0.935	30.18	133.19	25105
49	ODM37515.1	31841	9.18	0.739	20.9	137.1	30480
50	ODM 37499.1	24720.69	4.25	-0.316	-0.316	89.37	47690

Table 3. Assigning proteins under different functional classes using CELLO2GO

S.No	**Accession Number**	**Functional class**	**Score**
1.	CEK42940.1	cellular process	-0.83870873
2.	QGW 60714.1	cellular process	1.4125244
3.	CDL57047.1	Metabolism molecule	2.3086116
4.	ODM39876.1	Metabolism molecule	0.95789324
5.	ODM34760.1	cellular process	0.85028507
6.	PLP12834.1	Metabolism molecule	-11.119463
7.	ODM41053.1	cellular process	5.726317
8.	ODM 41051.1	cellular process	1.3882346
9.	ODM41025.1	cellular process	1.7510281
10.	ODM 40980.1	Metabolism molecule	1.8785957
11.	ODM 40914.1	Cellular process	2.2908744
12.	ODM 40663.1	Metabolism molecule	1.2374383
13.	ODM 40510.1	Metabolism molecule	0.91812879

14.	ODM40436.1	Cellular process	0.88032908
15.	ODM 40356.1	Cellular process	1.092828
16.	ODM 40241.1	Cellular process	0.32520296
17.	ODM 40153.1	Cellular Process	0.4482564
18.	ODM 40064.1	Cellular process	1.4388512
19.	ODM39954.1	Metabolism molecule	0.97320013
20.	ODM39942.1	Cellular process	1.4503689
21	ODM39924.1	Metabolism molecule	0.27568006
22.	ODM39923.1	Metabolism molecule	0.45348925
23.	ODM39920.1	Cellular process	1.3536922
24.	ODM39862.1	Cellular process	0.65282994
25.	ODM39714.1	Cellular process	0.66527258
26.	ODM 39564.1	Cellular process	0.61654133
27.	ODM39520.1	Cellular process	2.0045757
28.	ODM39286.1	Metabolism molecule	2.1459901
29.	ODM39182.1	Cellular process	1.8019255
30.	ODM39151.1	Cellular process	1.987278
31.	ODM38959.1	Metabolism molecule	2.6727011
32.	ODM 38375.1	Cellular process	0.98884747
33.	ODM 38342.1	Cellular process	0.97634509
34.	ODM 38315.1	Metabolism molecule	1.6837046
35.	ODM38280.1	Metabolism molecule	0.32024358
36.	ODM38268.1	Cellular process	0.18494394
37.	ODM 38197.1	Metabolism molecule	-0.28833818
38.	ODM 38190.1	Cellular process	1.6972158
39.	ODM38155.1	Metabolism molecule	1.4216292
40.	ODM38087.1	Cellular process	0.93849715
41.	ODM38064.1	Metabolism molecule	7.1771962
42.	ODM38057.1	Metabolism molecule	3.4042632
43.	ODM38056.1	Metabolism molecule	0.7609872
44.	ODM 37935.1	Information molecule	0.91059675
45.	ODM 37919.1	Virulence Factors	1.0071275
46.	ODM37834.1	Metabolism molecule	1.3392627
47.	ODM 37668.1	Metabolism molecule	0.92415558
48.	ODM 37640.1	Metabolism molecule	-0.88505948
49.	ODM37515.1	Metabolism molecule	4.4557462
50.	ODM 37499.1	Virulence Factors	-5.209018

Table 4 Subcellular Localization of uncharacterized proteins using PSORT

S.No	**Accession Number**	**Localization**
1	CEK42940.1	Cytoplasmic
2	QGW 60714.1	Cytoplasmic
3	CDL57047.1	Inner membrane
4	ODM39876.1	Cytoplasmic
5	ODM34760.1	Periplasmic
6	PLP12834.1	Cytoplasmic
7	ODM41053.1	Outer membrane
8	ODM 41051.1	Cytoplasmic
9	ODM41025.1	Cytoplasmic

10	ODM 40980.1	Cytoplasmic
11	ODM 40914.1	Inner membrane
12	ODM 40663.1	Cytoplasmic
13	ODM 40510.1	Periplasmic
14	ODM40436.1	Cytoplasmic
15	ODM 40356.1	Inner membrane
16	ODM 40241.1	Periplasmic
17	ODM 40153.1	Inner membrane
18	ODM 40064.1	Cytoplasmic
19	ODM39954.1	Cytoplasmic
20	ODM39942.1	Inner membrane
21	ODM39923.1	Periplasmic
22	ODM39920.1	Cytoplasmic
23	ODM39862.1	Cytoplasmic
24	ODM39714.1	Inner membrane
25	ODM 39564.1	Cytoplasmic
26	ODM39520.1	Inner membrane
27	ODM39286.1	Cytoplasmic
28	ODM39182.1	Inner membrane
29	ODM39151.1	Cytoplasmic
30	ODM38959.1	Cytoplasmic
31	ODM 38375.1	Cytoplasmic
32	ODM 38342.1	Inner membrane
33	ODM 38315.1	Cytoplasmic
34	ODM38280.1	Cytoplasmic
35	ODM38268.1	Inner membrane
36	ODM 38197.1	Cytoplasmic
37	ODM 38190.1	Inner membrane
38	ODM38155.1	Inner membrane
39	ODM38087.1	Cytoplasmic
40	ODM38064.1	Inner membrane
41	ODM38057.1	Inner membrane
42	ODM38056.1	Inner membrane
43.	ODM38056.1	Inner membrane
44.	ODM 37935.1	Cytoplasmic
45.	ODM 37919.1	Cytoplasmic
46.	ODM37834.1	Inner membrane
47.	ODM 37668.1	Cytoplasmic
48.	ODM 37640.1	Inner membrane
49.	ODM37515.1	Inner membrane
50.	ODM 37499.1	Cytoplasmic

Table 5.Prediction of virulence potential of hypothetical proteins using VICM

S.NO	Accession Number	Virulent/Non Virulent	Score
1.	CEK42940.1	Virulent	0.536377
2.	QGW 60714.1	Virulent	0.63637
3.	CDL57047.1	Virulent	0.53637
4.	ODM39876.1	Non virulent	0.4363
5.	ODM34760.1	Virulent	0.4263
6.	PLP12834.1	Virulent	0.5364
7.	ODM41053.1	Virulent	0.3543
8.	ODM 41051.1	Virulent	-65474
9.	ODM41025.1	Non virulent	0.5364
10.	ODM 40980.1	Virulent	0.2537
11.	ODM 40914.1	Virulent	1.3674
12.	ODM 40663.1	Non virulent	0.3879
13.	ODM 40510.1	Non virulent	0.2626
14.	ODM40436.1	Virulent	0.3646
15.	ODM 40356.1	Virulent	1.3546
16.	ODM 40241.1	Non virulent	0.373
17.	ODM 40153.1	Virulent	0.4637
18.	ODM 40064.1	Non virulent	0.2637
20.	ODM39954.1	Non virulent	0.2537
21.	ODM39942.1	Virulent	0.263
22.	ODM39924.1	Virulent	0.276
23.	ODM39923.1	Non virulent	1.738
24.	ODM39920.1	Virulent	0.173
25.	ODM39862.1	Virulent	0.287
26.	ODM39714.1	Non virulent	0.2637
27.	ODM 39564.1	Virulent	0.265
28.	ODM39520.1	Non virulent	0.652
29.	ODM39286.1	Non virulent	0.364
30.	ODM39182.1	Non virulent	0.253
31.	ODM39151.1	Non virulent	0.5374
32.	ODM38959.1	Virulent	0.653
33.	ODM 38375.1	Non virulent	0.232
34.	ODM 38342.1	Virulent	0.637
35.	ODM 38315.1	Virulent	0.363
36.	ODM38280.1	Non virulent	0.5436
37.	ODM38268.1	Virulent	0.27378
38.	ODM 38197.1	Non virulent	0.3647
39.	ODM 38190.1	Virulent	0.377
40.	ODM38155.1	Non virulent	0.356
41.	ODM38087.1	Virulent	0.474
42.	ODM38064.1	Virulent	0.3675
43.	ODM38057.1	Non virulent	0.3534
44.	ODM38056.1	Virulent	0.374
45.	ODM 37935.1	Non virulent	0.364
46.	ODM 37919.1	Non virulent	0.2653
47.	ODM37834.1	Virulent	0.377
48.	ODM 37668.1	Non virulent	0.3278
49.	ODM 37640.1	Virulent	09.2536
50.	ODM 37499.1	Virulent	0.363764

Table 6: Prediction of antigenicity property of hypothetical proteins using VaxiJen v2.0

S.NO	Accession Number	Antigenicity score

1.	CEK42940.1	0.4256
2.	QGW 60714.1	0.5144
3.	CDL57047.1	0.3740
4.	ODM39876.1	0.3740
5.	ODM34760.1	0.3131
6.	PLP12834.1	0.5680
7.	ODM41053.1	0.5633
8.	ODM 41051.1	0.2727
9.	ODM41025.1	0.4790
10.	ODM 40980.1	0.2595
11.	ODM 40914.1	0.3002
12.	ODM 40663.1	0.3531
13.	ODM 40510.1	0.6249
14.	ODM40436.1	0.5780
15.	ODM 40356.1	0.6222
16.	ODM 40241.1	0.4
17.	ODM 40153.1	0.6561
18.	ODM 40064.1	0.5802
19.	ODM39954.1	0.3994
20.	ODM39942.1	0.4018
21.	ODM39924.1	0.3820
22.	ODM39923.1	0.4744
23.	ODM39920.1	0.5181
24.	ODM39862.1	0.3832
25.	ODM39714.1	0.3313
26.	ODM 39564.1	0.4284
27.	ODM39520.1	0.2861
28.	ODM39286.1	0.3462
29.	ODM39182.1	0.3091
30.	ODM39151.1	0.3154
31.	ODM39151.1	0.3045
32.	ODM 38375.1	0.1200
33.	ODM 38342.1	0.3100
34.	ODM 38315.1	0.4093
35.	ODM38280.1	0.3589
36.	ODM38268.1	0.4057
37.	ODM 38197.1	0.5506
38.	ODM 38190.1	0.4104
39.	ODM38155.1	0.3559
40.	ODM38087.1	0.3351
41.	ODM38064.1	0.3715
42.	ODM38064.1	0.5438
43.	ODM38056.1	0.5168
44.	ODM 37935.1	0.3067
45.	ODM 37919.1	0.4811
46.	ODM37834.1	0.1733
47.	ODM 37668.1	0.2930
48.	ODM 37640.1	0.4385
49.	ODM37515.1	0.2565
50.	ODM 37499.1	0.4069

Table 7. Prediction of B-cell epitopes from the 5 candidate antigenic virulent proteins using ABCpred.

Accession Number	B-Cell epitope	ABC pred	Vaxijen	AllergenFP
CEK42940.1	RLMHDAAAWMSAKGTP	0.87	-0.1243	Allergen
	MSAKGTPAWDVARIDR	0.84	0.2408	NonAllergen
	PTHETICALPRTEINP	0.84	0.0106	NonAllergen
	IVGCCTLSAEDPEFWP	0.83	0.7231	NonAllergen
	AVRRTHAGRGVSSALI	0.82	0.9949	NonAllergen
	TRRSLEEPTWALUFKOING	0.82	0.5605	Allergen
	LASMIDKLESIELLAP	0.81	-0.1219	NonAllergen
	VASCSD GVGCC TULSA	0.81	0.8421	NonAllergen
	TFN PGW DPT FIERLE	0.80	1.0049	NonAllergen
	RLGF VDTF PGWP	0.80	0.9243	Allergen
	AC SGDI DEV VRL MHDA	0.79	-0.0086	NonAllergen
	NIAEM DEEP NVVA	0.79	0.8658	NonAllergen
	PRTEINPLASMIDKLE	0.79	0.0986	NonAllergen
	SSALIEACRHAARTQG	0.72	0.2732	Allergen
	VARIDRTFAETFVLRS	0.67	0.9296	NonAllergen
	RTQGCAKLRLDCHPNL	0.66	0.19	NonAllergen
	ACRHAARTQGCAKLRL	0.64	1.3405	NonAllergen
	ERWORLDUOCOLLYER	0.57	1.2373	NonAllergen
QGW 60714.1	EGCUGCUGUVGVGGV	0.45	0.5678	Allergen
	DERRY DSDT TTTT	0.56	0.6578	NonAllergen
	DGHDJYFUGHHGGLH	0.67	-0.6775	Non Allergen
	FHDHDFYGFUGHIHJ	0.66	0.5566	Non Allergen
	ERYN JD RIKY FREE	0.67	0.5423	Non Allergen
CDL57047.1	DBC HDVCHDVHVDHV	0.66	0.6775	Non Allergen
	GGGG FC GEIGERT	0.7	0.4466	Non Allergen
	DVE HD FEY FYFE DYE	0.55	0.5567	Non Allergen
	FDU WFD WFDIF	0.56	0.5677	Non Allergen
	EDEDVEDFEYDFYDD	0.34	0.4456	Non Allergen
	YWFDYFEDYEFDYFF	0.45	0.6677	Non Allergen
	DFYUFWYFEFYEWIU	0.67	0.5544	Non Allergen
	HDFC CHEVY	0.36	0.4349	Non Allergen
	BDKJ DUGDUGI	0.67	0.7648	Non Allergen
	DFDFDFDF	0.78	0.5464	Non Allergen
	GDVGDVCDGCDYEFY	0.45	0.6575	Non Allergen
	SCX ECFY WE DEFY	0.67	0.4252	Non Allergen
	GCDFDYWFQDYFDFD	0.64	0.8750	Non Allergen
ODM39876.1	SCVDYFCYEDFCYFV	0.60	0.7647	Non Allergen
	FWEF FRFV CGEC	0.45	0.6454	Allergen
	DVREVREVREGVGHH	0.34	0.5636	Allergen
	TYT HYTHE HYTHE YHT	0.12	0.6547	Allergen
	RGEGEHGTHTHTRHT	0.45	0.6467	Non Allergen
	GSRTC FYFU GGUG	0.40	0.7373	Allergen
ODM34760.1	KJFC WFL FGHJ	0.52	0.6364	Allergen
	FEVG FB TRHTR BBG	0.35	0.6367	Allergen
	YYTJUHBWFEVRTBN	0.45	0.6367	Non Allergen
	TYJNYTHYGREFVRE	0.34	0.7648	Non Allergen
	FGHTRHTYHYHJYVG	0.56	0.5859	Non Allergen
	YHTRGRTTRHYRHYT	0.45	-0.7847	Non Allergen

Table 8. Prediction of T-cell epitopes from the 5 candidate antigenic virulent proteins using CTLpred

Accession Number	Peptide	TCPred Score	Antigenicity score	Allergenicity
>QGW60714.1	JFKKDKJFD	0.678	0.5498	Allergen
>PLP12834.1	FHGDEFFJD	0.345	0.3789	Non Allergen
>ODM41053.1	YEGGO CHECK	0.567	0.1237	Non Allergen
>ODM40436.1	GHETTO JGTHF	0.236	0.6548	Allergen
>ODM40153.1	DFDFFFF	0.567	0.2723	Non Allergen
>ODM38197.1	VFVFVF LEVY	0.536	0.3744	Allergen
>ODM38057.1	CVDCHVCVV	0.548	0.3344	Non Allergen

Table 9. Prediction of transmembrane helices for the 5 candidate antigenic virulent proteins using TMHMM and HMMTOP

Accession Number	TMHMM Score	HMMTOP Score
>QGW60714.1	0	0
>PLP12834.1	0	0
>ODM41053.1	0	0
>ODM40436.1	0	0
>ODM40153.1	0	0
>ODM38197.1	0	0
>ODM38057.1	0	0

Table 10. BLAST results of the 5 candidates compared against human proteome

Accession Number	Non- human Homologous
>QGW60714.1	No significant similarities found
>PLP12834.1	No significant similarities found
>ODM41053.1	No significant similarities found
>ODM40436.1	No significant similarities found
>ODM40153.1	No significant similarities found
>ODM38197.1	No significant similarities found