Received: 25 November 2023 | Revised: 28 February 2024 | Accepted: 15 March 2024 | Published online: 21 March 2024

RESEARCH ARTICLE

Medinformatics
2024, Vol. 1(2) 54–61
DOI: 10.47852/bonviewMEDIN42022148

BON VIEW PUBLISHING

In Silico Annotation and Immunoinformatics Guided Epitope Mapping of Potential Antigenic Proteins of *Trichomonas foetus*

Shravya Gorla[1], Shravani Akula[1], Deevya Awalaloo[1], Swathi Pathlavath[1], Neha Shireen[1], Alavala Renuka Devi[1] and Geethanjali Karli[1,*]

[1]*Department of Biotechnology, Indira Priyadarshini Govt. Degree College for Women, India*

Abstract: Bovine trichomonosis is one of the neglected tropical diseases of cattle that is resulting in severe reproductive failure. With present knowledge, disease diagnosis and maintaining the infected animals in the quarantine are the only available strategies. Several spillover incidences of *Trichomonas foetus* had also resulted in zoonotic transmission to humans. In spite of above circumstances, till date there are no point of care diagnostics developed for screening bovine trichomoniasis in cattle. In the light of above circumstances, there exists a demand for cost-effective diagnostic kits to be provided to farming community. This current study highlights evaluation of few surface proteins of *Trichomonas foetus* for the suitability as sero-diagnostic markers. Few target Proteins such as Adhesin, Immuno-dominant variable surface antigen-like protein, Polymorphic membrane protein - like protein, GP-63-like (Clan MA, family M8) protein and Hypothetical protein (OHS95735.1) were evaluated for suitable pH, Signal Peptide, protein glycosylation pattern using freely available Bioinformatics tools. Mapping of potential epitopes of all the target proteins was done using immunoinformatics tools. Among the above proteins, GP63-like protein, immuno-dominant variable antigenic domain-like protein, and polymorphic membrane protein-like proteins are most suitable as diagnostic targets, owing to their higher levels of glycosylation, large epitope domains, and showing structural similarities with the domains of known toxic proteins. On the other hand, adhesin protein has the potential to be exploited as a vaccine candidate. The above proteins are suitable to be expressed in suitable host system and validated the immunogenic potential by animal inoculation and by testing with the real samples.

Keywords: B-cell epitope prediction, signal peptide, point of care diagnostics, comparative genomics, 3D structure prediction

1. Introduction

The continual increase in the human population is creating a great demand for animal-based food products. However, livestock is under tremendous stress due to rise in the incidence of infectious diseases and rapid changes in environmental conditions [1, 2]. Reproduction rate of cattle is found to be as low as 35%, which is insufficient to meet global demand of milk and meat [3]. Artificial insemination with the sperm of high-quality bulls as one essential measure to improve breeding came into light in several parts of the world. However, due to a lack of semen quality analysis criteria, the incidence of sexually transmitted diseases such as brucellosis, bovine tuberculosis, para tuberculosis, infectious bovine rhinotracheitis, campylobacteriosis, foot and mouth disease, and bovine trichomonosis remains to be the bottleneck.

In light of the prevailing low breeding rates in cattle, minimum standard operating procedures for bovine breeding have been set up by World Organization of Animal Health and a few point-of-care diagnostic kits like iELISA and lateral flow assay (LFA) rapid kits were developed for diseases such as brucellosis [4, 5], bovine tuberculosis [6], para tuberculosis [7], and FMD [8]. However, there were no attempts made to develop screening tools for bovine trichomonosis in cattle. *Trichomonas foetus (T. foetus)* is a single-cellular microaerophilic parasite, usually found as a trophozoite and acquires a pseudocyst form at later stage [9]. *T. foetus* infestation can be suspected in the herd based on the presentation of symptoms like irregular oestrous cycles, the requirement of recurrent services/inseminations, and secondary infections due to hormonal imbalance (pyometra). Prolonged vaginitis and placentitis were found to cause fetal mortality in pregnant cows, which was eventually found to result in abortions as early as within 8 weeks of gestation [10]. During natural breeding, when the cows were bred by an infected bull, almost 30–90% of cows were found to get infected in the herd. On the other hand, an affected cow was found to transmit the parasitic infection to healthy bulls [11, 12]. The infection in the herd is characterized by frequent abortions and presence of large number of unbred cows [13]. Cows usually get infected through sexual transmission from infected bulls, artificial insemination with the infected semen. Except for few chemical methods of surface sterilization [14], till date there are no legal options proposed for elimination of parasite. *Trichomonas foetus* is reported to have zoonotic transmission to humans in a number of immune-compromised patients [15, 16] which

*Corresponding author: Geethanjali Karli, Department of Biotechnology, Indira Priyadarshini Govt. Degree College for Women, India. Email: mail2anjalee@yahoo.co.in

further extends the concern. Hence, there is a great demand to develop easily available, simple, and easy to use diagnostics.

Currently, there is only one commercially available kit, namely VetMAX™-Gold Trich Detection Kit, Thermo-Fischer Scientifics, USA, which is being used for trichomonosis detection in few specialized laboratories across the world. As early as 1990s, Yule et al. [17] developed an iELISA for the detection of *T. foetus*-specific antibodies in vaginal mucous samples and serum from naturally infected cows using the ethanol-fixed *T. foetus* parasites as antigen. Further, indirect ELISA (iELISA) was developed using TF1.17 surface antigen of *T. foetus* to demonstrate the presence of *T. foetus*-specific antibodies in the smegma of naturally infected cows [18] and in the smegma of bulls [19]. Further, Cobo et al. [20] examined the existence of *T. foetus*-specific antibodies in cows that were experimentally bred with the infected bulls using ethanol-fixed immobilized *T. foetus* parasites. Cobo et al. [21] have also developed an iELISA using *T. foetus* extracted antigen for detection of specific antibodies in serum. Due to the cost associated with bulk culturing of *T. foetus*, long incubation periods and the zoonotic potential (human Transmission) associated with *T. foetus* cells, the earlier assays exhibited limited scope for commercialization.

Till date, there are no commercially available simple detection kits for screening bovine trichomonosis in cattle. In our prior work, transcriptomics and proteomics-based expression profiles of *T. foetus* were evaluated and few surface proteins were proposed as suitable diagnostic targets [22]. The above analysis was also supported by analyzing the expression profiles of Trichomonas vaginalis, a closely related parasite to bovine parasite causing trichomonosis in humans. Due to its high incidence in human, extensive research work was carried out to reveal cell surface markers as virulence factors involved in pathogenesis [23, 24] that had further improved the confidence in choosing suitable markers. Since the time of whole genome sequencing of *T. foetus* available [25], several attempts were made to identify unique genes for developing molecular detection kits such as Polymerase Chain Reaction (PCR) and Real Time- Polymerase Chain Reaction (RT-PCR). However, the above assays were not suitable for point of care setting. For the first time, in this work we have chosen few highly expressed proteins from the published data and analyzed for immunogenic potential and suitability as a sero-diagnostic marker intended to develop antibody detection kits. As described by Tomar and De [26], immuno-informatics, often known as computational immunology, is the intersection of computer science and experimental immunology. It represents the use of computational approaches and resources to the analysis of immunological data. It not only aids in the processing of large amounts of data, but it also plays an important role in the development of novel diagnostics and therapeutics. It also investigates the possibility of combining immunoinformatics with systems biology for the development of personalized medicine. The current study focuses on identification of potential immunogenic domains of selected antigenic proteins suitable for developing antibody detection assays.

2. Materials and Methods

In this study, few candidate cell surface and surface associated proteins, such as GP63-like protein, adhesin, hypothetical protein (OHS95735.1), immuno-dominant variable antigenic domain-like protein, and polymorphic membrane protein, were analyzed for predicting various physico-chemical properties like the acidic or alkaline nature of proteins, presence of signal peptide, N and O-glycosylation sites, secondary and tertiary structural features as well as immunogenic epitope domain prediction. The overview of workflow is shown in Figure 1.

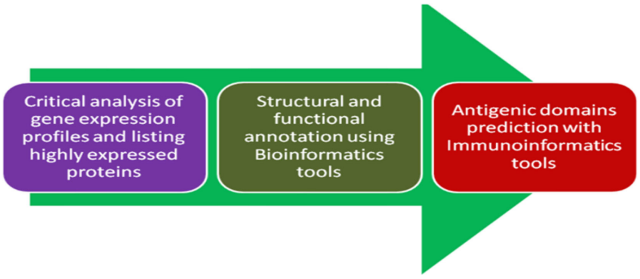

Figure 1. Work flow indicating the basic steps involved in identifying suitable proteins and epitope domains of target proteins suitable for developing diagnostics

2.1. Protein sequence retrieval

Amino acid sequence of target protein is quite important for prediction of various physico-chemical properties of the proteins using bioinformatics tools. The accession numbers of GP63-like protein (OHS97275.1), adhesin (OHT02241.1), hypothetical protein (OHS95735.1), immuno-dominant variable antigenic domain-like protein (OHT11175.1), polymorphic membrane protein (OHS93232.1) were submitted to "NCBI protein" (https://www.ncbi.nlm.nih.gov/protein/) server (NCBI accessed on Feb 2019). A freely searchable database of proteins maintained by "National Center for Biotechnological information" at "National Institute of Health", USA [27]. The amino acid sequences were obtained in the FASTA format. The sequence information will be used for prediction of various physico-chemical properties of the target proteins.

2.2. Prediction of suitable pH for the enzyme activity

Determining the optimal pH for the enzymatic activity of the proteins is very important for determining the pH for the media suitable for enhanced protein expression in heterologous systems. Amino acid sequences of test [28] proteins were submitted to "AcalPred" (http://lin-group.cn/server/AcalPred) with default parameters to predict a probability value that most likely determines maximum enzyme activity of protein at particular acidic or alkaline pH.

2.3. Signal peptide prediction

Signal peptide enables the protein to be a secretory. The "SignalP" 5.0 (http://www.cbs.dtu.dk/services/SignalP/) tool [29], a freely available tool, was used to predict the presence of signal peptide in the query proteins. Amino acid sequences of all these proteins were submitted to "SignalP" tool to predict the presence of any signal peptide. However, the prediction is not perfect and there are possible false negatives.

2.4. Prediction of glycosylation sites on antigenic proteins

Post-translational modifications like N-glycosylation and O-glycosylation are known to enhance the stability and immunogenicity of eukaryotic proteins. The N-glycosylated amino

acids were identified by uploading the FASTA sequences of all the proteins to the "NetNGly 1.0 Server" (http://www.cbs.dtu.dk/services/NetNGlyc/) [30]. O-Glycosylated amino acids were identified by submitting the FASTA sequence of the protein to "NetOglyc server 1.0" (http://www.cbs.dtu.dk/services/NetOGlyc/) [31].

2.5. Analysis of 2D structural details of the protein

Higher level of organization of protein structure correlates with the function of the proteins. To understand the various types of secondary structural features of a query protein, the FASTA sequences of the target proteins were uploaded to "SOPMA" tool (https://npsaprabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) – a "Secondary structure prediction" method [32]. The percentage of alpha helices, beta turns, and other extended turns assist with the prediction of function of the protein.

2.6. Prediction of 3D structure of the proteins

Understanding the 3D structure of the protein is fundamental for the prediction of function as well as target sites for drug designing. As these proteins of *T. foetus* were poorly annotated, computational tool of "EXPASY – SWISS MODEL" (https://swissmodel.expasy.org/) was used to predict the possible 3D structures of few proteins [33]. FASTA sequences of all test proteins were submitted to "SWISS MODEL" tool with default parameters. Few closely related templates from Protein Data Bank (PDB) and the percentage similarity along with description were downloaded and the model was built. Scanprosite tool from EXPASY (ScanProsite.expasy.org) [34] was used to predict the shared domains between the test and predicted proteins.

2.7. Retrieval of the amino acid sequences as suitable epitope domains from target proteins

Membrane proteins with continuous epitope domains are predicted to be involved in host pathogen interaction and inducing immune response. In this context, B-cell epitope prediction tools from Immune Epitope Database Tools (IEDB) Resource were used to map the extent of epitopes in the target proteins. Protein sequences of all target proteins were analyzed using the "BepiPred-2.0" algorithm (http://tools.iedb.org/bcell/) [35] at 0.5 threshold value. The extent of amino acids from largest epitope domain were noted.

3. Results and Discussion

3.1. Protein sequence retrieval

FASTA sequence of the proteins is the most acceptable amino acid sequence format for bioinformatics analysis. Protein sequences in the FASTA format for GP63-like protein, adhesin, hypothetical protein, immuno-dominant variable antigenic domain-like protein, and polymorphic membrane protein were retrieved from "NCBI protein" portal. The sequences for the above proteins were compiled and available in the Supplementary data 1.

3.2. Prediction of suitable pH for the enzyme activity

Proteins were analyzed using "AcalPred" server and probability score was obtained. The probability value determines a particular protein determining the likelihood maximum enzyme activity at acidic or alkaline pH. Predicted nature of the enzymatic activity for the above proteins is listed in Table 1.

In the above list, GP-63-like (Clan MA, family M8) protein and hypothetical protein were predicted to be alkaline enzymes with the probability score of above 0.9. If these proteins to be expressed in vitro in a heterologous host systems, the above predictions guide us to maintaining the pH of the culture medium in the acidic range, so that activity of the protein may not be lost prior to harvesting. Adhesin, immuno-dominant variable surface antigen-like protein, and polymorphic membrane proteins were predicted as acidic enzymes with the probability score of above 0.9.

The above predictions are quite useful during heterologous recombinant protein expression. For instance, GP-63-like (Clan MA, family M8) protein and hypothetical proteins need to be synthesized in the media optimized in the acidic range, whereas immuno-dominant variable surface antigen-like protein, adhesin, and polymorphic membrane protein-like protein are to be expressed in the medium optimized in the alkaline range, to minimize the enzymatic activity and optimal protein recovery.

3.3. Signal peptide prediction

The "SignalP 5.0" server predicted the following probability scores for each query protein. Except for the immuno-dominant variable antigenic domain-like protein, all the query proteins were shown to have very low likelihood score for having a signal peptide or being a secretary protein. The individual protein scores with analysis are given in Table 2.

As the above analysis is purely based on predictions tools, there are more chances of generating false positive results. Hence, all these findings further warrant wet lab analysis.

3.4. Prediction of glycosylation sites on antigenic proteins

All the test proteins of the current study were found to contain either N/O-glycosylated amino acids or both as presented in Table 3. Adhesin and hypothetical proteins were predicted to have only N-glycosylated amino acids and lacking of O-glycosylation

Table 1. Prediction of alkaline/acidic nature of the antigenic proteins of *T. foetus*

Name of the protein	NCBI accession number	Predicted nature of the enzyme	Probability score
GP-63-like (Clan MA, family M8) protein	OHS97275.1	Alkaline enzyme	0.937129
Adhesin	OHT02241.1	Acidic enzymes	0.931316
Hypothetical protein	OHS95735.1	Alkaline enzyme	0.957141
Immuno-dominant variable surface antigen-like protein	OHT11175.1	Acidic enzymes	0.919904
Polymorphic membrane protein-like protein	OHS93232.1	Acidic enzymes	0.994233

Note: GP-63, glycoprotein 63 kilodaltons

Table 2. Prediction of signal peptide of the antigenic proteins of *T. foetus*

Name of the protein	Likelihood score (SEC/SPI)	Impression
GP-63-like (Clan MA, family M8) protein	0.023	No signal peptide present. Nonsecretary protein
Adhesin	0.0133	No signal peptide present. Nonsecretary protein
Hypothetical protein	0.007	No signal peptide present. Nonsecretary protein
Immuno-dominant variable surface antigen-like protein	0.8849	Has a signal peptide It is a secretary protein
Polymorphic membrane protein-like protein	0.004	No signal peptide present. Nonsecretary protein

Note: SEC, secretary; SPI, signal peptide

Table 3. Potential glycosylation sites of the target proteins of *T. foetus* predicted using NetNGly 1.0 and NetOGly 1.0 server

Name of the protein	Sites of N-glycosylation	Sites of O-glycosylation
GP-63-like (Clan MA, family M8) protein	99,112,124,189,409,489,560	74,345,348,349.353,445,540,544,549
Adhesin	59,65,104,159,184,244,278,293	–
Hypothetical protein (OHS95735.1)	310	–
Immuno-dominant variable surface antigen-like protein	79,89,102, 239	63
Polymorphic membrane protein-like protein	211,276,681	377–638,685

Note: N, amino terminal; C, carboxyl terminal

sites. GP-63-like protein was having 9 amino acid sites and one site for immuno-dominant variable surface antigen-like protein. Polymorphic membrane protein was found to have hundreds of O-glycosylation sites and was identified as heavily O-glycosylated protein.

As mentioned by Brooks [36], protein glycosylation plays a crucial role in several biological activities, such as cell adhesion and signaling, stability, function, and immunogenicity of the proteins. Similarly, Moremen et al. [37] revealed that glycosylation of cellular proteins has a considerable influence on interactions with the external environment by acting as ligands for cell adhesion, macromolecule interactions, and pathogen invasion. As described by Breloy and Hanisch [38], O-glycosylation of amino acids was found to impact protein structure, cellular protein sorting, resistance to proteolysis, and cellular communication. Further, de Haas et al. [39] have established that N-glycosylation of proteins was found to influence pathogen recognition, cell–cell interaction, and migration in the host tissue.

3.5. Analysis of 2D structural details of the protein

Amino acid sequences of target proteins were analyzed through "SOPMA" tool analyzed to unveil the secondary structural features with varied fractions. The percentage of alpha helices, beta turns, and random coils is tabulated in Table 4.

The 2D patterns are heterogeneously distributed among all the target proteins. GP-63-like (Clan MA, family M8) protein has predominant random coiling of 51.61%. Adhesin has equal proportions of extended strands and random coiling at 42.02% and 41.22%, respectively. Hypothetical protein (OHS95735.1) has alpha helices at 40.11%. Immuno-dominant variable surface antigen-like protein has 47.04% of random coiling as predominant 2D structure. Polymorphic membrane protein-like protein has 21.90% of extended strand as major form.

3.6. Prediction of 3D structure of the proteins

"EXPASY – SWISS MODEL" tool compared the target proteins to the available PDB structures. The results displayed are few predominant templates in building the 3D model of the antigenic proteins of *T. foetus* from PDB server. The predicted 3D structures of the target proteins modeled by "SWISS MODEL" tool that were build using few important templates from PBD data for all individual target protein are available in Figure 2(A)–(E).

The corresponding percentage similarly and the description of individual templates are available in Table 5.

All the target proteins exhibited variable degree of structural similarity with known toxins. GP-63-like (Clan MA, family M8) protein showed 24.55% with the Leishmanolysin. There were evidences that major glycoprotein GP-63 (Leishmanolysin) in Leishmania was having key role in establishing parasitism in the host [40]. Adhesin protein exhibited structural 9.38% similarity with "Bordetella Pertussis Virulence Factor P.69 Pertactin" [41]

Table 4. The 2-dimentional structural features of the antigenic proteins of *T. foetus*

Name of the protein	Alpha helix	Extended strand	Beta turn	Random coil
GP-63-like (Clan MA, family M8) protein	29.03%	15.97%	3.39%	51.61%
Adhesin	5.59%	42.02%	11.17%	41.22%
Hypothetical protein	40.11%	23.56%	9.43%	26.90%
Immuno-dominant variable surface antigen-like protein	24.81	25.19	2.96	47.04
Polymorphic membrane protein-like protein	15.56	21.90	5.51	17.02

Table 5. Few predominant templates in building the 3D model of the antigenic proteins of *T. foetus*

Name of the protein	PDB ID of the template	Description	Sequence identity
GP-63-like (Clan MA, family M8) protein	6nbx.1.G	NADH-quinone oxidoreductase subunit J T. elongatus NDH	23.53%
	1lml.1.A	Leishmanolysin	24.55%
Adhesin	1dab.1.A	P.69 Pertactin The structure of Bordetella pertussis virulence factor P.69 pertactin	9.38%
	6bea.1.A	Autotransporter protein UpaB Crystal structure of the autotransporter UpaB from *E. coli* strain CFT073	15.79%
	2iou.1.G	Pertactin extracellular domain Major tropism determinant P1 (Mtd-P1) variant complexed with Bordetella bronchiseptica virulence factor pertactin extracellular domain (Prn-E)	14.45%
Hypothetical protein (OHS95735.1)	3c72.1.B	Geranylgeranyl transferase type-2 subunit beta engineered RabGGTase in complex with a peptidomimetic inhibitor	20.13%
	4l9p.1.B	CaaX farnesyltransferase beta subunit Ram1 Crystal structure of *Aspergillus fumigatus* protein farnesyltransferase complexed with the FII analog, FPT-II, and the KCVVM peptide	21.71%
	2wy8.1.A	Complement c3d fragment Staphylococcus aureus complement subversion protein Sbi-IV in complex with complement fragment C3d	15.23%
Immuno-dominant variable surface antigen-like protein	6bbo.1.B	MCherry fluorescent protein Crystal structure of human APOBEC3H/RNA complex	23.40%
	6bbo.1.B	MCherry fluorescent protein Crystal structure of human APOBEC3H/RNA complex	23.40%
	5ev7.1.A	Conserved domain protein The crystal structure of a functionally unknown conserved protein mutant from Bacillus anthracis str. Ames	16.39%
	7jtv.1.A	Immuno modulating metalloprotease structure of IMPa from Pseudomonas aeruginosa in complex with an O-glycopeptide	16.11%
Polymorphic membrane protein-like protein	6qps.1.A	Polysaccharide lyase family 6 Structural characterization of a mannuronic acid-specific polysaccharide family 6 lyase enzyme from human gut microbiota	13.51%
	4ozy.1.A	Poly(beta-D-mannuronate) C5 epimerase Crystal structure of the periplasmic alginate epimerase AlgG T265N mutant	11.72%
	1bhe.1.A	Polygalacturonase Polygalacturonase from Erwinia Carotovora Ssp. Carotovora	12.28%
	4xm3.1.A	Tail spike protein Tail spike protein mutant E372A of *E. coli* bacteriophage HK620 in complex with pentasaccharide	9.68%

Note: NADH, nicotinamide adenine dinucleotide; IMP, inositol monophosphate

which clearly states that adhesin could have possible role in causing virulence. Immuno-dominant variable surface antigen-like protein had 16.11% structural similarity with "Immunomodulating metalloprotease from Pseudomonas aeruginosa," whereas the polymorphic membrane protein-like protein had 13.51% structural similarity with "Polysaccharide Lyase Family 6". Since the total percentage similarity with the participating templates in structure building was >40%, the predicted models in Figure 2 are quite significant [42]. These findings clearly indicate the possible toxicity of the test proteins and their significant role in host parasite interactions.

3.7. Retrieving the amino acid sequences as suitable epitope domains

"BepiPred-2.0" tool analyzed the protein sequences and the most possible amino acids were displayed as epitopes in the sequence form as well as graphics. The largest domains as well as the most suitable stretch of amino acids suitable for cloning are identified and presented in Table 6.

All the above proteins are predicted to have largest domains ranging from 27 to 333 residues. According to Buus et al. [43], ultrahigh-density peptide microarrays analysis of linear epitopes of protein expressed sequence tags (PrESTs) against rabbit polyclonal antibodies found that the majority of the epitopes were present in the 6–12 amino acid residue range. Further, based on epitope and paratope contact residue analysis, Stave and Lindpaintner [44] reported that amino acid lengths >20 amino acids long were detected in 111 X-ray crystallographic structures of antigen–antibody.

Immunogenic domains of about minimum of 90 amino acids long in the extracellular domain of the protein were selected, which is expected to project at least 10–15 epitopes of the particular protein. The predicted epitope domains can be cloned and expressed as recombinant antigen in eukaryotic host systems like yeast. These proteins can either expressed individually or as

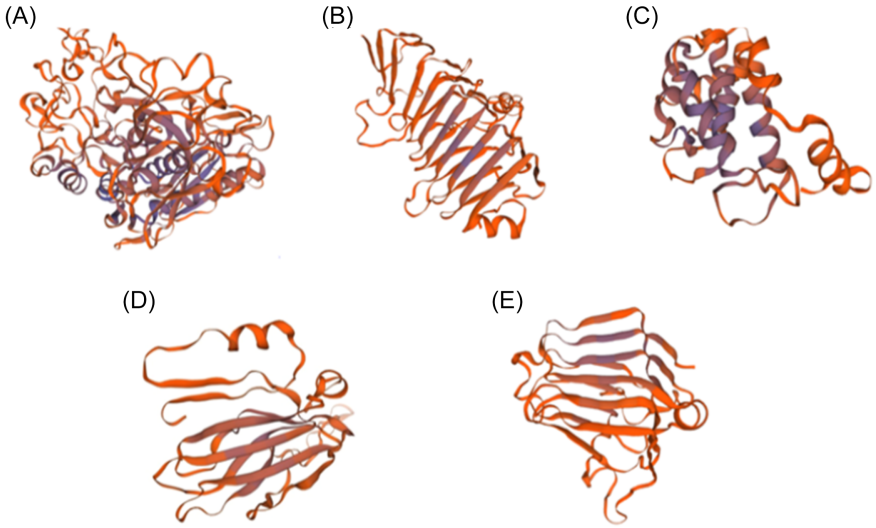

Figure 2. 3D structures of various antigenic proteins of *T. foetus* predicted using SWISS MODEL tool that were build based on few predominant templates from PDB as shown in Table 5. (A) GP-63-like (Clan MA, family M8) protein, (B) adhesin, (C) hypothetical protein (OHS95735.1), (D) immuno-dominant variable surface antigen-like protein, and (E) polymorphic membrane protein-like protein.

Table 6. Prediction of antigenic domains of target proteins of *T. foetus*

Name of the target protein	Largest epitope (no. of amino acids)	Domain suitable for cloning (AA-domain)
GP-63-like (Clan MA, family M8) protein	52	320–480
Adhesin	27	200–340
Hypothetical protein (OHS95735.1)	33	200–340
Immuno-dominant variable surface antigen-like protein	53	41–130
Polymorphic membrane protein-like protein	333	304–636

Note: GP-63, glycoprotein 63KDa

chimera protein using multiple antigens. Such recombinant proteins would serve as source of antigen for designing indirect ELISA for the large-scale screening as well as for development of rapid kit LFA that are suitable for onsite screening for *T. foetus*-specific antibodies in the infected animals by the farmers.

4. Conclusions

GP63-like protein, adhesin, hypothetical protein, immuno-dominant variable antigenic domain-like protein, and polymorphic membrane protein-like protein of *T. foetus* were chosen for current investigation, owing to their many folds expression in gene expression profiles and consistently expressed transcripts and proteins in our previous work. Such proteins were known to have a significant role in host–pathogen interactions. As *T. foetus* genome was sequenced recently [25], the gene function and protein structural and functional annotations were not made available to the public use. Hence, we investigated structural characteristics, structure similarity characteristics, and immunogenic potential using a variety of freely available bioinformatics tools.

The above analysis indicates that GP63-like protein, immuno-dominant variable antigenic domain-like protein, and polymorphic membrane protein-like proteins are most suitable proteins as diagnostic targets, owing to their higher levels of glycosylation, large epitope domains, and showing structural similarities with the domains of known toxic proteins. Adhesin protein might be exploited as a vaccine candidate. However, these findings warrant further studies on expression of recombinant proteins in suitable eukaryotic host systems, analysis of expressed protein for antigenicity in lab animals, and capturing wide range of antibodies in the infected host.

Acknowledgement

The authors acknowledge the team members of Genomix Molecular Diagnostics, Kukatpally, Hyderabad, India for their constant support for this current study.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Supplementary Information

The supplementary file is available at https://doi.org/10.47852/bonviewMEDIN42022148

References

[1] Independent Evaluation Group. (2012). *National dairy support project*. Retrieved from: https://documents1.worldbank.org/curated/en/766871613087488778/pdf/India-National-Dairy-Support-Project.pdf

[2] Government of India. (2014). *Compendium of minimum standards of protocol & standard operating procedures for bovine breeding*. Retrieved from: https://agriportal.cg.nic.in/ahd/PDF_common/SOP_for_Artificial_Insemination.pdf

[3] Government of India. (2022). *Annual report 2021–22*. Retrieved from: https://dof.gov.in/sites/default/files/2022-04/Annual_Report_2021_22_English.pdf

[4] Chinthaparthi, M. R., Anumolu, V. K., Mallepaddi, P. C., Revathi, P., Manasa, M., Poonati, R., . . . , & Vijayaraghavan, R. (2017). Sero prevalence and validation of in-house igm ELISA kit for the detection of Brucella antibody in human serum samples. *Asian Journal of Microbiology, Biotechnology and Environmental Sciences*, *19*(4), 975–980.

[5] Manasa, M., Revathi, P., Chand, M. P., Maroudam, V., Navaneetha, P., Raj, G. D., . . . , & Rathnagiri, P. (2019). Protein-G-based lateral flow assay for rapid serodiagnosis of brucellosis in domesticated animals. *Journal of Immunoassay and Immunochemistry*, *40*(2), 149–158. https://doi.org/10.1080/15321819.2018.1541803

[6] Pucken, V. B., Knubben-Schweizer, G., Döpfer, D., Groll, A., Hafner-Marx, A., Hörmansdorfer, S., . . . , & Hartnack, S. (2017). Evaluating diagnostic tests for bovine tuberculosis in the southern part of Germany: A latent class analysis. *PLoS ONE*, *12*(6), e0179847. https://doi.org/10.1371/journal.pone.0179847

[7] Punati, R. D., Mallepaddi, P. C., Poonati, R., Jain, M., Maity, S. N., Sohal, J. S., . . . , & Polavarapu, R. (2019). Development and validation of rapid, sensitive and inexpensive protein G-based point of care diagnostic assay for serodiagnosis of paratuberculosis at resource-limited areas. *Current Trends in Biotechnology and Pharmacy*, *13*(3), 232–242.

[8] Sharma, G. K., Mahajan, S., Matura, R., Subramaniam, S., Ranjan, R., Biswal, J., . . . , & Pattnaik, B. (2015). Diagnostic assays developed for the control of foot-and-mouth disease in India. *World Journal of Virology*, *4*(3), 295–302. https://doi.org/10.5501/wjv.v4.i3.295

[9] Meyer Mariante, R., Coutinho Lopes, L., & Benchimol, M. (2004). *Tritrichomonas foetus* pseudocysts adhere to vaginal epithelial cells in a contact-dependent manner. *Parasitology Research*, *92*, 303–312. https://doi.org/10.1007/s00436-003-1026-z

[10] World Organization for Animal Health. (2018). *Chapter 3.4.15. Trichomonosis of OIE terrestrial manual*. Retrieved from: https://www.woah.org/fileadmin/Home/fr/Health_standards/tahm/3.04.15_TRICHOMONOSIS.pdf

[11] BonDurant, R. H. (2005). Venereal diseases of cattle: Natural history, diagnosis, and the role of vaccines in their control. *Veterinary Clinics of North America: Food Animal Practice*, *21*(2), 383–408. https://doi.org/10.1016/j.cvfa.2005.03.002

[12] Jin, Y., Schumaker, B., Logan, J., & Yao, C. (2014). Risk factors associated with bovine trichomoniasis in beef cattle identified by a questionnaire. *Journal of Medical Microbiology*, *63*(6), 896–902. https://doi.org/10.1099/jmm.0.074971-0

[13] Collántes-Fernández, E., Fort, M. C., Ortega-Mora, L. M., & Schares, G. (2018). *Trichomonas*. In M. Florin-Christensen & L. Schnittger (Eds.), *Parasitic protozoa of farm animals and pets* (pp. 313–388). Springer. https://doi.org/10.1007/978-3-319-70132-5_14

[14] Martin, K. A., Kovach, K., Moscoso, E., Carreiro, E., Chelladurai, J. R. J., & Brewer, M. T. (2023). Assessment of *in vitro* efficacy for common surface disinfectants and antiseptics against *Tritrichomonas foetus* trophozoites. *Frontiers in Veterinary Science*, *10*, 1282274. https://doi.org/10.3389/fvets.2023.1282274

[15] Yao, C. (2012). Opportunistic human infections caused by *Tritrichomonas* species: A mini-review. *Clinical Microbiology Newsletter*, *34*(16), 127–131. https://doi.org/10.1016/j.clinmicnews.2012.07.004

[16] Yao, C. (2013). Diagnosis of *Tritrichomonas foetus*-infected bulls, an ultimate approach to eradicate bovine trichomoniasis in US cattle? *Journal of Medical Microbiology*, *62*(1), 1–9. https://doi.org/10.1099/jmm.0.047365-0

[17] Yule, A., Skirrow, S. Z., & Bonduran, R. H. (1989). Bovine trichomoniasis. *Parasitology Today*, *5*(12), 373–377. https://doi.org/10.1016/0169-4758(89)90298-6

[18] Ikeda, J. S., BonDurant, R. H., & Corbeil, L. B. (1995). Bovine vaginal antibody responses to immunoaffinity-purified surface antigen of *Tritrichomonas foetus*. *Journal of Clinical Microbiology*, *33*(5), 1158–1163. https://doi.org/10.1128/jcm.33.5.1158-1163.1995

[19] Rhyan, J. C., Wilson, K. L., Wagner, B., Anderson, M. L., Bondurant, R. H., Burgess, D. E., . . . , & Corbeil, L. B. (1999). Demonstration of *Tritrichomonas foetus* in the external genitalia and of specific antibodies in preputial secretions of naturally infected bulls. *Veterinary Pathology*, *36*(5), 406–411. https://doi.org/10.1354/vp.36-5-406

[20] Cobo, E. R., Cano, D., Rossetti, O., & Campero, C. M. (2002). Heifers immunized with whole-cell and membrane vaccines against *Tritrichomonas foetus* and naturally challenged with an infected bull. *Veterinary Parasitology*, *109*(3–4), 169–184. https://doi.org/10.1016/S0304-4017(02)00301-1

[21] Cobo, E. R., Corbeil, L. B., Gershwin, L. J., & BonDurant, R. H. (2009). Preputial cellular and antibody responses of bulls vaccinated and/or challenged with *Tritrichomonas foetus*. *Vaccine*, *28*(2), 361–370. https://doi.org/10.1016/j.vaccine.2009.10.039

[22] Karli, G., Polava, R., & Varada, K. (2020). Comparative omics based approach to identify putative immunogenic proteins of *Trichomonas foetus*. In *Advances in Computational and Bio-Engineering: Proceeding of the International Conference on Computational and Bio Engineering*, 583–592. https://doi.org/10.1007/978-3-030-46939-9_51

[23] de Miguel, N., Lustig, G., Twu, O., Chattopadhyay, A., Wohlschlegel, J. A., & Johnson, P. J. (2010). Proteome analysis of the surface of *Trichomonas vaginalis* reveals novel proteins and strain-dependent differential expression. *Molecular & Cellular Proteomics*, *9*(7), 1554–1566. https://doi.org/10.1074/mcp.M000022-MCP201

[24] Hirt, R. P. (2013). *Trichomonas vaginalis* virulence factors: An integrative overview. *Sexually Transmitted Infections*, *89*(6), 439–443. https://doi.org/10.1136/sextrans-2013-051105

[25] Benchimol, M., de Almeida, L. G. P., Vasconcelos, A. T., de Andrade Rosa, I., Reis Bogo, M., Kist, L. W., & de Souza, W. (2017). Draft genome sequence of *Tritrichomonas foetus* strain K. *Genome Announcements*, *5*(16), e00195–17. https://doi.org/10.1128/genomea.00195-17

[26] Tomar, N., & De, R. K. (2014). Immunoinformatics: A brief review. In R. K. De & N. Tomar (Eds.), *Immunoinformatics* (pp. 23–55). Humana Press. https://doi.org/10.1007/978-1-4939-1115-8_3

[27] Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., . . . , & Zbicz, K. (2017). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

[28] Lin, H., Chen, W., & Ding, H. (2013). AcalPred: A sequence-based tool for discriminating between acidic and alkaline enzymes. *PLoS ONE*, *8*(10), e75726. https://doi.org/10.1371/journal.pone.0075726

[29] Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., . . . , & Nielsen, H. (2019). Signal P 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology, 37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

[30] Gupta, R., & Brunak, S. (2001). Prediction of glycosylation across the human proteome and the correlation to protein function. In *Biocomputing 2002: Proceedings of the Pacific Symposium*, 310–322. https://doi.org/10.1142/9789812799623_0029

[31] Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., . . . , & Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, *32*(10), 1478–1488. https://doi.org/10.1038/emboj.2013.79

[32] Geourjon, C., & Deléage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, *11*(6), 681–684. https://doi.org/10.1093/bioinformatics/11.6.681

[33] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . , & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

[34] de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., . . . , & Hulo, N. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, *34*, W362–W365. https://doi.org/10.1093/nar/gkl124

[35] Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*(W1), W24–W29. https://doi.org/10.1093/nar/gkx346

[36] Brooks, S. A. (2009). Strategies for analysis of the glycosylation of proteins: Current status and future perspectives. *Molecular Biotechnology*, *43*, 76–88. https://doi.org/10.1007/s12033-009-9184-6

[37] Moremen, K. W., Tiemeyer, M., & Nairn, A. V. (2012). Vertebrate protein glycosylation: Diversity, synthesis and function. *Nature Reviews Molecular Cell Biology*, *13*(7), 448–462. https://doi.org/10.1038/nrm3383

[38] Breloy, I., & Hanisch, F. G. (2018). Functional roles of O-glycosylation. *Molecules*, *23*(12), 3063. https://doi.org/10.3390/molecules23123063

[39] de Haas, P., Hendriks, W. J. A. J., Lefeber, D. J., & Cambi, A. (2020). Biological and technical challenges in unraveling the role of N-glycans in immune receptor regulation. *Frontiers in Chemistry*, *8*, 55. https://doi.org/10.3389/fchem.2020.00055

[40] Button, L. L., & McMaster, W. R. (1988). Molecular cloning of the major surface antigen of leishmania. *Journal of Experimental Medicine*, *167*(2), 724–729. https://doi.org/10.1084/jem.167.2.724

[41] Hijnen, M., Mooi, F. R., van Gageldonk, P. G., Hoogerhout, P., King, A. J., & Berbers, G. A. M. (2004). Epitope structure of the *Bordetella pertussis* protein P.69 pertactin, a major vaccine component and protective antigen. *Infection and Immunity*, *72*(7), 3716–3723. https://doi.org/10.1128/iai.72.7.3716-3723.2004

[42] Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, *42*(1), 3.1.1–3.1.8. https://doi.org/10.1002/0471250953.bi0301s42

[43] Buus, S., Rockberg, J., Forsström, B., Nilsson, P., Uhlen, M., & Schafer-Nielsen, C. (2012). High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Molecular & Cellular Proteomics*, *11*(12), 1790–1800. https://doi.org/10.1074/mcp.M112.020800

[44] Stave, J. W., & Lindpaintner, K. (2013). Antibody and antigen contact residues define epitope and paratope size and structure. *The Journal of Immunology*, *191*(3), 1428–1435. https://doi.org/10.4049/jimmunol.1203198

How to Cite: Gorla, S., Akula, S., Awalaloo, D., Pathlavath, S., Shireen, N., Devi, A. R., & Karli, G. (2024). In Silico Annotation and Immunoinformatics Guided Epitope Mapping of Potential Antigenic Proteins of *Trichomonas foetus*. *Medinformatics*, *1*(2), 54–61. https://doi.org/10.47852/bonviewMEDIN42022148