**GOVT. DEGREE COLLEGE FOR WOMEN, KUKATPALLY**

**Department of Biotechnology**

**Certificate course- Immunoinformatics based approach to identify Biomarkers for infectious diseases**

**2020-21**

**Name of the Certificate course Conducted : Immunoinformatics based approach to identify Biomarkers for infectious diseases**

**Name of the Course coordinator : K.Geethanjali**

**No. of Students enrolled : 10**

**Date of commencement of classes : 9/03/2021**

**Date of ending of the course : 8/05/2021**

**No. of classes conducted : 30**

**No. of students appeared for final exam : 10**

**No. of students passed final exam : 10**

**Over all participation of the students : Satisfactory**

# GOVT. DEGREE COLLECGE FOR WOMEN GAJWEL, SIDDIPET

## Department of Biotechnology
**Certificate course**

**Immunoinformatics based approach to identify Biomarkers for infectious diseases**

**2020-21**

**Total students enrolled-10** **No. of hours-30 hr**

**Course objectives:**

- To Impart the basic understanding of cells and organs of immune system
- To enable the students to design experiments to evaluate underlying immunological basis of diseases.
- To educate the students regarding Auto immunity, Hyper sensitivity reactions in Health and disease
- To make the students aware of recent advanced in Immunological diagnosis methods and strategies
- IEDB resource tools for epitope prediction and vaccine development
- Tools for fast track developing diagnostics.

**Syllabus**

**Immune system**:

Types of immunity – Innate and Acquired immune mechanism
Organs and cells of immune system Immunity,
Antibody structure and classes; opsonisation; Antibody diversity
Types of Antigens, Antigenicity (factors affecting antigenecity); haptens; adjuvants
.

**Types of Immunity:**
Humoral immunity-B cell mediated Immunity
Cell mediated immunity: TC mediated immunity, NK cell mediated immunity, ADCC
Brief description of cytokines and MHC (MHC types and diversity)

**Vaccination**:
Vaccination: Discovery, principles, significance, Types of Vaccines (Traditional – live attenuated, heat killed, toxoids, Modern vaccines – recombinant, DNA, peptide)

**Immunological Techniques**:

Antigen-antibody reactions: Precipitation, agglutination, complement fixation, ELISA, RIA.
Hybridoma technology: Monoclonal antibodies and their applications in immune diagnosis.

**Immunoinformatics**

Basic requirements for working on tools
B-Cell epitope prediction tools
T-Cell epitope prediction tools
Role of protein domains for synthesizing the truncated proteins with large continuous epitopes.
Use of recombinant proteins for developing lateral flow assays (LFA) and ELISA for rapid diagnostics.

**Course Outcomes**

- The basic understanding of microbial physiology, cellular and molecular components of immune system, genetic control of elements will help the students to design experiments to evaluate underlying causes of diseases.
- The basic understanding of microbial physiology, cultural characteristics will enable the students to identify micro organisms that cause diseases
- For any sudden onset of disease out breaks- the approach to look into available data and applying the online tools.
- Basic hands on training in Immunoinformatics techniques

## List of students Enrolled

| S.No | Regd.No | Name Of The Student | No. of classes attended |
|---|---|---|---|
| 1 | 203919572006 | ALAGONDA SAI KRISHNA | 30 |
| 2 | 203919572008 | AMEENA FATIMA | 30 |
| 3 | 203919572015 | ARSHIYA MUBEEN | 30 |
| 4 | 203919572036 | CHIPPALA SWATHI | 30 |
| 5 | 203919572039 | DHADIGA MOUNIKA | 30 |
| 6 | 203919572041 | DONTHIREDDY SAI CHARITHA | 30 |
| 7 | 203919572048 | GANTE SAIVARMA | 30 |

| | | | |
|---|---|---|---|
| 8 | 203919572059 | KAKARAPARTI KARTHEEK | 30 |
| 9 | 203919572061 | KANCHARI VIMALA | 30 |
| 10 | 203919572069 | KOTTA APARNA | 30 |

## Students attendence

| | Date | 9/3 | 10/3 | 11/3 | 12/3 | 13/3 | 16/3 | 6/4 | 7/4 | 8/4 | 9/4 | 10/4 | 4/5 | 5/5 | 6/5 | 8/5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S.NO | No. of Hours | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1. | ALAGONDA SAI KRISHNA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 2. | AMEENA FATIMA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3. | ARSHIYA MUBEEN | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 4. | CHIPPALA SWATHI | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 5. | DHADIGA MOUNIKA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 6. | DONTHIREDDY SAI CHARITHA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 7. | GANTE SAIVARMA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 8. | KAKARAPARTI KARTHEEK | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 9. | KANCHARI VIMALA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 10. | KOTTA APARNA | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

## Students Evaluation

II Year B.SC Biotechnology

| S.No | Regd.No | Name Of The Student | Max. Marks (20) |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 1 | 203919572006 | ALAGONDA SAI KRISHNA | 18 |
| 2 | 203919572008 | AMEENA FATIMA | 16 |
| 3 | 203919572015 | ARSHIYA MUBEEN | 18 |
| 4 | 203919572036 | CHIPPALA SWATHI | 19 |
| 5 | 203919572039 | DHADIGA MOUNIKA | 20 |
| 6 | 203919572041 | DONTHIREDDY SAI CHARITHA | 17 |
| 7 | 203919572048 | GANTE SAIVARMA | 20 |
| 8 | 203919572059 | KAKARAPARTI KARTHEEK | 19 |
| 9 | 203919572061 | KANCHARI VIMALA | 17 |
| 10 | 203919572069 | KOTTA APARNA | 19 |

GOVERNMENT COLLEGE FOR WOMEN, GAJWEL

Department of Biotechnology

CERTIFICATE COURSE

SUBJECT: Immunology & Immunoinformatics MAX.MARKS: 20
TIME: 30 MIN

I. Choose correct answers to the following – **(10 X 1 = 10 Marks)**

1. The T – cell receptor can bind to antigenic peptide ( )
   a) Only in the free form b) only when Presented on to MHC molecules
   c) Only when bound to Hapten d) Only when bound to antibody

2. CD3 + molecules are present on the surface of both these types of cells ( )
   a) B – Cells and T helper cells
   b) T – helper cells and NK cells
   c) T Cytotoxic cells cell and T Helper cells
   d) Tc cells and NK cells

3. Which one of the following is auto immune disorder ( )

a) Haemophilia - A
b) Sickle cell anemia
c) Type – 1 diabetes mellitus
d) Type -2 diabetis mellitus

4. Most potent antigens are ( )
   a) DNA b) Proteins c) Lipids d) Carbohydrates

5. Humoral immunity acquired passively by ( )
   a) Catching a virus from a friend by shaking hands
   b) Receiving a vaccine of influenza virus grown in eggs
   c) Receiving serum from some one who has recovered from an infection
   d) Receiving leucocytes from an immune family member

6. The antibody IgE and Mast cells are associated with ( )
   a) Contact hyper sensitivity
   b) Immuno floresence tests
   c) Electrophoresis
   d) Immediate hyper sensitivity
7. In primary Immune response the major antibody that appears in circulation is ( )

   a) IgG b) Ig A c) Ig D d) Ig E

8. Cells that release histamine and other vaso active substances in response to allergens are ( )
   a) neutrophills
   b) macrophages
   c) NK cells
   d) Mast cells

9. Of the MHC receptors, which of the following presents antigen to Cytotoxic T cells ( )
   a) Class I MHC b) Class II MHC c) Class III MHC d) all

10. What type of cell is fused with a myeloma cell when producing Monoclonal antibodies ( )
    a) B - lymphocyte b) T - lymphocyte c) Plasma cell d) Hepatocyte

II fill in the blanks – (10 X1 = 10 Marks)

1. The molecular weight of α chain of MHC molecule in class - I MHC

   _____

2. Expand HAT _____

3. The auto immune disorder to joints _____
4. Expand HLA _____
5. _____ organ is affected in insulin dependent diabetes mellitus.
6. Most abundant antibody in Circulation is _____
7. CD4-, CD8- cells are called _____
8. _____ cells activate both Tc cells and B - cells
9. Failure to respond to an antigen is a condition called _____
10. T cells mature in _____

**Material**

## Immunoinformatics to unveil the diagnostic markers of *Trichomonas foetus*

Bovine Trichomonosis is one of the underexplored diseases of cattle which resulting in reproductive failures. 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 resulted in Zoonotic transmission to humans. 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 for few structural features, Glycosylation patterns using Bioinformatics applications and antigenic potential using Immunoinformatics tools . Adhesin, Immuno-dominant variable surface antigen-like protein, and polymorphic membrane proteins were predicted as acidic enzymes, while GP-63-like (Clan MA, family M8) protein and Hypothetical protein (OHS95735.1) were predicted as alkaline enzymes.

All of the test proteins, with the exception of the Immuno-dominant variable surface antigen-like protein, were devoid of signal peptide and all of the proteins listed above have moderate N- and O-glycosylation sites. Polymorphic membrane protein was strongly O-Glycosylated at 107 amino acid sites. Further analysis revealed that 2-Dimensional structural features were heterogeneously dispersed. Several proteins shared structural similarity with proteins domains of recognized toxins. Immuno-dominant variable surface antigen like protein shared a common domains with Immuno-modulating metalloprotease , whereas GP-63 like protein shared a domain similarity with Leishmanolysin. Largest predicted epitope domains in all of the test proteins were between 27 and 333 amino acids long.

Among the above list of proteins 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 can be exploited as a vaccine candidate. These proteins can be expressed in suitable host system and validate the immunogenic properties by animal inoculation as well as by testing with the real samples

**Key words**: B-Cell epitope prediction, Signal Peptide, Point of care diagnostics, Comparative genomics, 3-D structure prediction

## Introduction

Uncontrolled rise in population is associated with rapid increase in demand for food products including those of animal based. However, live stock is under tremendous stress due to rise in the risk of infectious diseases and rapid changes in environmental conditions (Compendium of Minimum standards of protocol 2014). Reproduction rate of Cattle are as low as 35%, which is insufficient to meet global demand of milk and meat (Document of The World Bank, (2012)), ( Annual report 2012-23). This is resulting in uncontrolled adulteration of animal based food products. Artificial insemination with the sperm of high-quality bulls as one essential measure to improve breeding that came into light in several parts of the world. However, due to a lack of semen quality analysis criteria, there is rapid surge of sexually transmitted infections viz., Brucellosis, Bovine Tuberculosis, Para Tuberculosis, Infectious Bovine Rhinotrecheitis, Camphylobacteriosis, Foot and Mouth disease and Trichomonosis. [1]

Trichomonosis has wide spread epidemiology on the globe (OIE Reference manual,2019). The causative agent, *Trichomonas foetus* is a single-cellular microaerophilic parasite, usually found as a Trophozoite (Warton 1979) and acquires a psuedocyst form at later stage (Mariante 2009). Early abortions as well as widespread infertility in the cows are the most obvious signs of Trichomonas infection. Identifying affected cattle and isolating them from the herd is the best and only feasible eradication strategy (Skirrow 1988). *Trichomonas foetus* is reported to have zoonotic transmission to humans in a number of immune-compromised patients (Yao C 2012), (Yao C. (2013) Which further extends the concern. Hence, there is a great demand to develop easily available, simple and easy to use diagnostics. To address the "Minimum Standard Operating Procedures for Bovine Breeding", little emphasis was paid in recent years to developing point-of-care diagnostics for screening Bovine Tuberculosis (Pucken 2017), Para Tuberculosis (Pucken 2017), and Brucellosis at breeding stations (Rudrama 2019),( Mallikarjuna 2017).

Till date, the whole world is lacking a commercially available simple detection kit for Bovine Trichomonosis. In our prior work, gene expression profiles of *T.foetus* were evaluated and few surface proteins were proposed as suitable diagnostic targets (Manasa 2019). *Trichomonas vaginalis*, is closely related parasite to Bovine parasite causing Trichomonosis in humans. Due to its high incidence, extensive

research work was carried out to reveal cell surface markers in the "surface proteome and virulence factors involved in pathogenesis (Miguel 2010) (Hirt 2013). These findings added to our understanding of *T.foetus* pathophysiology and provided a base for proposing suitable diagnostic markers in our earlier studies (Karli 2020).

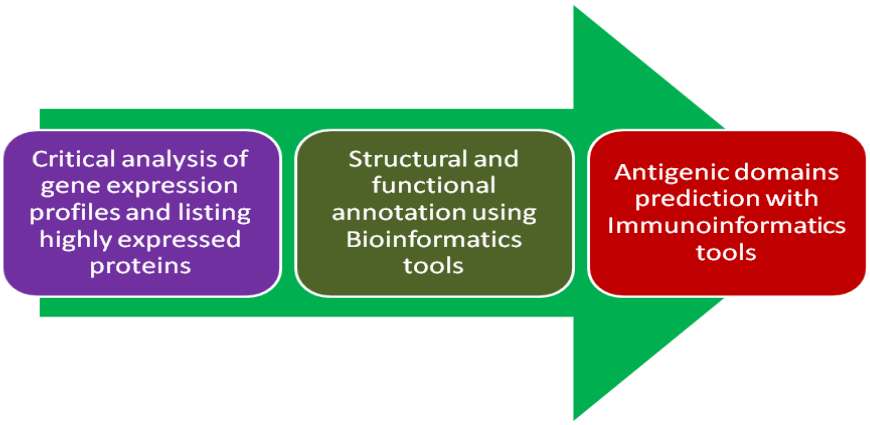


**Fig. 1. Graphical representation of the over all methodology involved in identifying suitable diagnostic markers**

In this current 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.

## Methods

***Protein sequence retrieval***

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. The amino acid sequences were obtained in the FASTA format.

***Prediction of suitable pH for the enzyme activity***

Amino acid sequences of test (Lin 2013) 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.

***Signal peptide prediction***

Signal peptide enables the protein to be a secretory. The "SignalP" 5.0 ( http://www.cbs. dtu.dk/services / SignalP/ ) tool (Armenteros 2019) a freely available tool was used to predict the of 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.

***Prediction of Glycosylation sites on antigenic proteins***

Post translational modification like N-Glycosylation and O-Glycosylation enhances 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/) (Gupta 2002). 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/) (Steentoft 2013).

***Analysis of 2-D structural details of the protein***

Higher level of organization of protein structure correlates with its function. 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/cgibin /npsa_automat.pl?page=/NPSA/npsa_sopma.html) – a "Secondary structure prediction" method (Geourjon 1995). The percentage of Alpha helices, Beta turns and other extended turns helpes in prediction of function of the protein.

***Prediction of 3-D structure of the Proteins***

Understanding the 3-D structure of the protein is vital to understand the 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 3-D structures of few proteins (Waterhouse 2018). FASTA sequences of all test proteins were submitted to "SWISS MODEL" tool with default parameters. Few closely related templates from PDB, percentage similarity and description were downloaded and allowed to build model. Scanprosite tool from EXPASY (ScanProsite.expasy.org) (Castro E 2006) was used to predict the shared domains between the test and predicted proteins.

***Retrieving the amino acid sequences as suitable antigenic domains***

Proteins with continuous epitopes are more preferred for developing diagnostic tool. "Immune epitope database" tools of "IEDB Resource" are freely available and maintained by "National Institute of Allergy and Infectious diseases" at NIH, USA has several tools to predict "B-Cell epitopes". Protein sequences in the amino acids of all target proteins were analyzed for epitope domains using the "BepiPred-2.0" algorithm (http://tools.iedb.org/bcell/) (Jespersen 2017) at 0.5 threshold value. The largest domain as well as the prediction of most suitable stretch of amino acids were noted.

**Result Analysis**

FASTA sequence of the proteins are 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 retrived from "NCBI protein" portal. The sequences for the above proteins were compiled and available in the Supplementary data.1.

Proteins were analyzed using "AcalPred" server and probability score was obtained. The probability value determines a particular protein determines 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.

**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 |
| **Immun dominant variable surface antigen like protein** | OHT11175.1 | acidic enzymes | 0.919904 |
| **Polymorphic membrane protein- like protein** | OHS93232.1 | acidic enzymes | 0.994233 |

GP-63-Glycoprotein 63 Kilo Daltons

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.

**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 peptode present. Non secretary protein |
| Adhesin | 0.0133 | No signal peptode present. Non secretary protein |
| hypothetical protein | 0.007 | No signal peptode present. Non secretary 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. Non secretary protein |

SEC-Secretary; SPI –Signal peptide

All the query proteins exhibited to have N-Glycosylation sites at a minimum of 1 to 8 amino acid residues. The predicted position of amino acids are presented in Table 3. Adhesin and Hypothetical proteins were lack of O-Glycosylation 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. Amino acid locations of glycosylation are listed in Table 3

**Table 3.Prediction of potential Glycosylation sites on the target proteins of *T.foetus***

| 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 |

N- Amino terminal; C-Corboxyl terminal

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 are tabulated in Table 4.

Table 4. 2-D 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 ( OHS95735.1) | 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 |

"EXPASY –SWISS MODEL" tool compared the target proteins to the available PDB structures. The results displayed are few predominant templates in building the 3-D model of the antigenic proteins of *T.foetus* from PDB server. The corresponding percentage similarly and the description of individual templates are available in Table 5.

**Table 5. Few predominant templates in building the 3-D 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 brochiseptica 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% |
| **Immun 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<br>The crystal structure of a functionally unknown conserved protein mutant from Bacillus anthracis str. Ames | 16.39% |
| | 7jtv.1.A | Immuno modulating metalloprotease<br>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<br>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<br>Crystal Structure of the periplasmic alginate epimerase AlgG T265N mutant | 11.72% |
| | 1bhe.1.A | Polygalacturonase<br>Polygalacturonase From Erwinia Carotovora Ssp. Carotovora | 12.28% |
| | 4xm3.1.A | Tail spike protein<br>Tailspike protein mutant E372A of E. coli bacteriophage HK620 in complex with pentasaccharide | 9.68% |

NADH-Nicotimaide Adenine Dinucleotide; IMP-Inositol Monophospahate

"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 were identified and presented in Table.6

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

| Name of the 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 |
| Immun dominant variable surface antigen -like protein | 53 | 41-130 |
| Polymorphic membrane protein- like protein | 333 | 304-636 |

GP-63 –Glycoprotein 63KDa

The most likely 3-D structures modeled by "SWISS MODEL" Tool for individual protein are available from Fig.2(a)-(e)

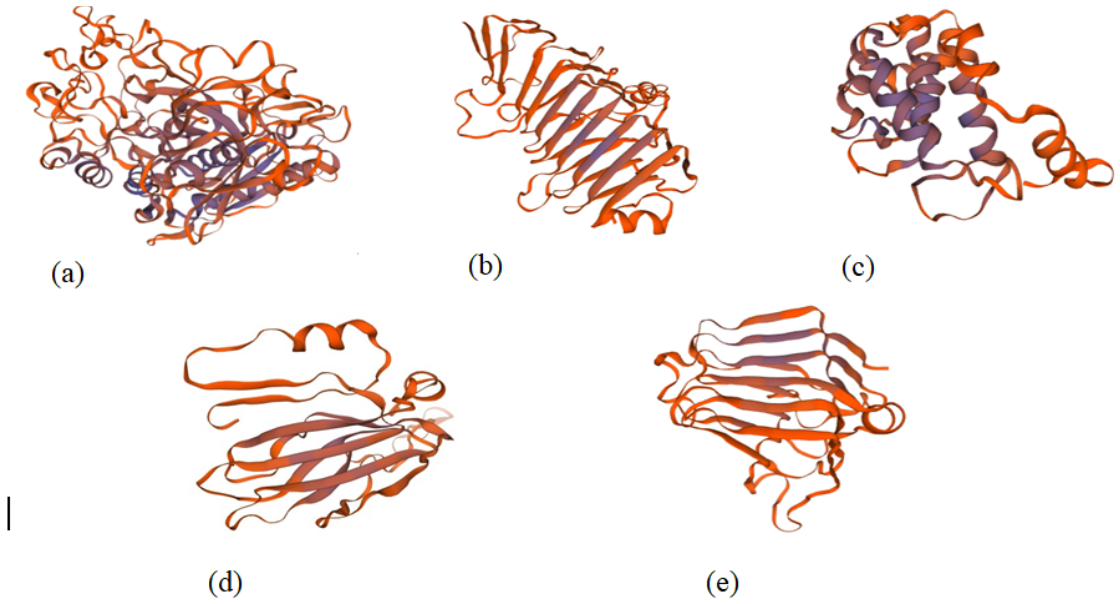


Fig. 2 SWISS MODEL 3-D structures of various antigenic proteins of *T.foetus*

(a) GP-63 like (Clan MA, family M8)protein (b) Adhesin (c) Hypothetical protein (OHS95735.1) (d) Immuno-dominant variable surface antigen like protein (e) Polymorphic membrane protein like protein

**Discussion**

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, Polymorphic membrane proteins were predicted as acidic enzymes with the Probability score above 0.9. Hence these proteins need to be preferably expressed in the culture media with optimized pH in the alkaline range, so as to retain the enzymatic activity.

All the test proteins except Immuno-dominant variable surface antigen like protein, all the other target proteins lack signal peptide. Hence all other proteins are non secretary proteins. Immuno-4dominat variable surface antigen-like protein are secretory proteins. With regard to glycosylation pattern, GP-63 like (Clan MA, family M8) protein is moderately glycosylated with 7 N-Glycosylated sites and 9 O-Glycosylated sites. Adhesin protein has only N-Glycosylation at 8 sites. Hypothetical protein is

N-glycosylated only at 310 site and no O-Glycosylation sites predicted. 4 N-Glycosylation sites and a single O-Glycosylation stes were predited for Immuno-dominant variable surface antigen like protein. Polymorphic membrane protein has three N-Glycosylation sites, but surprisingly got heavily O-Glycosylated at 107 amino acid residues.

The 2-D 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 2-D structure. Polymorphic membrane protein like protein has 21.90% of extended strand as major form.

Several 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( Button 1988) Adhesin protein exhibited structural 9.38% similarity with “Bordetella Pertussis Virulence Factor P.69 Pertactin”, (Hijnen 2004) 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”, where as 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 the Fig.2 are quite significant (Pearson 2013) These evidences lead us to understand the possible toxicity of the test proteins and their significant role in host parasite interactions.

All the above proteins are predicted to have largest domains ranging from 27-333 residues. For cloning these antigenic proteins, we had predicted amino acid stretches of more than 100 amino acids as suitable long continuous epitopes.

## Conclusions

GP63 like protein, Adhesin, Hypothetical protein, Immuno-dominant variable antigenic domain like protein, Polymorphic membrane protein like protein of *T.foetus* were chosen for current investigation, owing to their many fold 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 in recently (Benchimol 2017), 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 qualities using a variety of freely available bioinformatics tools.

## Limitations and future Prospects

From the above list of proteins 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 can be exploited as a vaccine candidate. All these prediction warrants further validation by expressing the protein in suitable host system and testing with real samples

### List of Abbreviations

*T.foetus (Trichomonas foetus*); *T.vaginalis (Trichomonas vaginalis*); PDB-Protein Data Bank

### Additional Information

Supplementary file 1. FASTA Sequences of Target antigenic proteins of Trichomonas foetus

Supplementary file 2. Graphical representations and antigenic domains of Target proteins of Trichomonas foetus. The supplementary information can be obtained from Corresponding author on request.


### Acknowledgments

We acknowledge the Team members of Genomix Molecular Diagnostics, Kukatpally, Hyderabad, INDIA for their constant support for this current study. We also acknowledge the DST-CURIE lab, Dept. of Biotechnology, SPMVV, Tirupati, INDIA for providing the computational facility.


## References


[1] Compendium of Minimum standards of protocol & Standard operating procedures For Bovine breeding Government of India Ministry of Agriculture Department of Animal Husbandry, Dairying & Fisheries, INDIA (2014)
[2] Document of The World Bank, National Dairy Support Project.(2012)
[3] Annual report Department Of Animal Husbandry, Dairying & Fisheries Ministry Of Agriculture Government Of India New Delhi (2012-13)
[4] Trichomoniasis. OIE Reference manual - Chapter 3.04.15- Accessed on May 2019.
[5] Warton, A., Honigberg, B,M.: Structure of Trichomonads as revealed by scanning electron microscopy. *J. Protozoot.*, 26, 56–6 (1979).

[6] Mariante, R.M., Lopes, L.C., Benchimol, M.: Tritrichomonas foetus pseudocysts adhere to vaginal epithelial cells in a contact-dependent manner. *Parasitology Research, 92*, 303-312 (2003) DOI: 10.1007/s00436-003-1026-z
[7] Skirrow, S,Z.,Bondurrant, R,H.: Bovine trichomoniasis. *Vet. Bull*., 58, 591–603. (1988)
[8] Yao C (2012). Opportunistic human infections caused by Tritrichomonas species: a mini-review. Clin. Microbiol. News, 34, 127–131. https://doi.org/ 10.1016/j. clinmicnews. 2012 .07.004
[9] Yao C. (2013) Diagnosis of Tritrichomonas foetus-infected bulls, an ultimate approach to eradicate bovine trichomoniasis in US cattle? *J. Med. Microbiol*., 62, 1–9.
[10] Pucken V-B, Knubben-Schweizer G, Döpfer D, Groll A, Hafner-Marx A, Hörmansdorfer 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
[11] Rudrama, D. P., Prudhvi, C. M., Revathi, P., Mukta, J., Soumendra, N. M., Jagdip, S. S., Sudhakar, P., Kavi, K. PB., and Rathnagiri, P. (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
[12] Mallikarjuna, R. C., Anumolu, V.K., Prudhvi, C. M., Revathi, P., Manasa, M., Rathnagiri, P., Vijayaraghavan, R. (2017) Sero Prevalence And Validation Of In-House IgM Elisa Kit For The Detection Of Brucella Antibody In Human Serum Samples *Asian Jr. of Microbiol. Biotech*. Env. Sc. 19 (4), 975-980
[13] Manasa, M. , Revathi, P., Prudhvi, C. M., Maroudam, V., Navaneetha, P., Dhinakar, Raj,G., Kavi, K. PB., Rathnagiri, P. (2019) Protein-G-based lateral flow assay for rapid serodiagnosis of brucellosis in domesticated animals, *Journal of Immunoassay and Immunochemistry*, 40:2, 232-242,
[14] 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*: MCP. 9(7), 1554–1566. doi:10.1074/mcp.M000022-MCP201
[15] Hirt R. P.( 2013) Trichomonas vaginalis virulence factors: an integrative overview. *Sexually transmitted infections*. 89(6), 439–443. doi:10.1136/sextrans-2013-051105
[16] Karli G., Polava R., Varada K.(2020) Comparative Omics Based Approach to Identify Putative Immunogenic Proteins of Trichomonas Foetus. *Learning and Analytics in Intelligent Systems*, 2020. 15. Springer, Cham. https://doi.org/10.1007/978-3-030-46939-9_51
[17] NCBI Resource Coordinators (2018). Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 46(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095
[18] Hao Lin, Wei Chen, Hui Ding. (2013) AcalPred: A sequence-based tool for discriminating between acidic and alkaline enzymes. *PloS ONE*, 8(10): e75726.
[19] José Juan Almagro Armenteros, Konstantinos D. Tsirigos, Casper Kaae Sønderby, Thomas Nordahl Petersen, Ole Winther, Søren Brunak, Gunnar von Heijne and Henrik Nielsen (2019) *Nature Biotechnology*, 37, 420-423, doi:10.1038/s41587-019-0036-z
[20] Gupta R, Brunak S.( 2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*.310-22.

[21] Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H.( 2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J*, 32(10):1478-88, May 15, . (doi: 10.1038/emboj.2013.79.

[22] Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* Dec;11(6):681-4. doi: 10.1093/bioinformatics/11.6.681

[23] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 46(W1), W296-W303

[24] De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006).ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. 1;34(Web Server issue):W362-5.

[25] Jespersen MC, Peters B, Nielsen M, Marcatili P. (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 3;45(W1):W24-W29. doi: 10.1093/nar/gkx346

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

[27] Hijnen M, Mooi FR, van Gageldonk PG, Hoogerhout P, King AJ, Berbers GA.( 2004) Epitope structure of the Bordetella pertussis protein P.69 pertactin, a major vaccine component and protective antigen. *Infect Immun*. 72(7):3716-23. doi: 10.1128/IAI.72.7.3716-3723.2004

[28] Pearson W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, *Chapter 3*, Unit3.1. https://doi.org/10.1002/0471250953.bi0301s42

[29] Benchimol, M., de Almeida, L., 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

# GOVT. DEGREE COLLEGE, KUKATPALLY.MEDCHAL Dt.)
# TELANGANA, INDIA

## *Certificate of Appreciation*

This is to certify that Mr.K.Kartheek, B,Sc BT.Z.C, II Yr has successfully completed the certificate course on "*Immunoinformatics based approach to identify Biomarkers for infectious diseases*" offered by Dept of Biotechnology for the academic year 2020-21.

Mrs.K.Geethanjali
Lecturer in Biotechnology

Dr.N.Alivelu Mangamma
Principal