**International Journal of Bioinformatics and Biomedical Engineering**
Vol. 2, No. 1, 2016, pp. 30-39
http://www.aiscience.org/journal/ijbbe
ISSN: 2381-7399 (Print); ISSN: 2381-7402 (Online)

# Insilico Validation of Babesia Bovis Merozoite Surface Antigen-1, Merozoite Surface Antigen-2b and Merozoite Surface Antigen-2c Proteins for Vaccine and Drug Development

**Sofia B. Mohamed[1, *], Mohamed M. Hassan[2]**

[1]Vector and Biomedical Studies, Tropical Medicine Research Institute, Khartoum, Sudan
[2]Faculty of Medical Laboratory Sciences, Al Zaiem Al Azhari University, Khartoum, Sudan

## Abstract

The Babesia bovis MSA-1, MSA-2b and MSA-2c are members of the variable merozoite surface antigen (VMSA) family, they are encoded surface proteins that are proposed to mediate the initial attachment of the merozoite to the host erythrocyte so this protein are targeted for vaccine and drug design. So the aim of this study is to give an outlook for MSA-1, MSA-2b and MSA-2c proteins using bioinformatics tools to help the developing drug and vaccine. In the present study, an in silico techniques were initiated to characterize the properties and structure of the MSA-1, MSA-2b and MSA-2c proteins. Firstly, the Physico-chemical characterization were computed by ExPasy's (ProtParam). Then the functional site prediction was done using ScanProsite. Subsequently, for functional characterization were computed the transmembrane regions and phosphorylation sites by SOSUI server and NetPhos server respectively. Thereafter, secondary structure prediction was explored using GOR IV. Finally, the 3D structure of proteins was built by sequence homology using CPH models 3.2 servers envision by Chimera 1.8 programming. The model was further surveyed by ERRAT, this confirmation of the quality of the model. Our results revealed that MSA-2c may be stable for a wide range of temperatures and the MSA-1 and MSA-2b classified as an unstable protein. While the MSA-2b less stable in test tube than MSA-2c and MSA-1. Further, all proteins are acidic and hydrophilic in nature, negatively charged, membrane and serine is the most phosphorylated amino acid in the Proteins. Also, we detected the sequences belonging to the following families: ASN_GLYCOSYLATION, CK2_PHOSPHO_SITE, PKC_PHOSPHO_SITE, MYRISTYL and TYR_PHOSPHO_. The secondary structure prediction of these proteins revealed that MSA-1, MSA-2b and MsA-2c have predominant mixed secondary structures. The 3 D structure for proteins were modeling and we found the quality of 3D structures less than 90%. Based on the findings, it could be concluded that further characterization of the Babesia bovis proteins is novel and will be important for drug and vaccine designing or understanding the interactions between proteins.

## Keywords

Babesia Bovis, MSA-1, MSA-2b, MSA-2c Proteins, Insilico Analysis, Physico-Chemical Character, 3-D Modeling and Functional Characterization

Received: November 17, 2015 / Accepted: January 7, 2016 / Published online: January 17, 2016



* Corresponding author
E-mail address: sofiaziada@gmail.com (S. B. Mohamed)

# 1. Introduction

Parasites in the genus *Babesia* are tick-borne, apicomplexan hemoparasites that cause severe hemolytic anemia, abortion, cerebral babesiosis, and death in susceptible animals. During the first steps of erythrocyte invasion, *Babesia* species use molecules located on the parasite surface coat to bind the host cell. In Babesia bovis, the merozoite surface bears at least five proteins that belong to the variable merozoite surface antigen (VMSA) family [1]. Babesia bovis MSA-1 and MSA-2 are part of the VMSA [2]. These proteins are exposed to the host immune system and have immunodominant CD4_ T lymphocyte epitopes [3]. Also, these proteins are exposed on the surfaces of both merozoites and sporozoites and induce invasion-blocking antibodies in immunized animals [4]. Unlike MSA-1, which is encoded by a single-copy gene, the four MSA-2 proteins, MSA-2a1, -2a2, -2b, and -2c, are encoded by tandem arranged genes within an 8.3-kB genomic locus, that includes a copy of MSA-2a1, MSA-2a2, MSA-2b, and MSA-2c genes, grouped in this order, apparently due to gene duplication [2]. MSA-2b and MSA-2c have 25% identical. These structural differences among MSA-2a, MSA- 2b, and MSA-2c are reflected by the presence of unique B-cell epitopes on each protein and the absence of cross-reactive antibody [5]. The region(s) of these proteins targeted by blocking antibodies is unknown. Vaccines directing the immune response against proteins involved in erythrocyte invasion, including merozoite surface proteins, provide a potential control point that targets the extracellular merozoite stage [6]. However, antigenic variation poses a challenge to the use of surface antigens in vaccines. A role has been attributed to MSA-2b as a participant during invasion of B. bovis to the erythrocyte, by co-expressing along with MSA-2c or MSA-2a1, MSA-2a2, and although less conserved, it has been proposed as a vaccine candidate [2]. Recently, several groups have focused on the development of vaccines against *B. bovis* based on recombinant antigens [7], the merozoite surface antigens (MSAs) were considered to be a good vaccine candidate, as the antisera raised against the recombinant forms of these antigens blocked the erythrocyte invasion of *B. bovis.* However, the extensive polymorphism shown by MSAs, which might result in altered immune responses, is one of the obstacles for the development of MSA-based sub-unit vaccines [8]. Computational analysis of biological sequences has become an extremely rich field of modern science and a highly interdisciplinary area, where statistical and algorithmic methods play a key role [9]. A large number of online tools and servers are available from different sources for making prediction regarding the identification and structure of proteins [10]. Applying bioinformatics algorithms to facilitate vaccine design is a very powerful approach that is changing many paradigms of vaccine discovery [11]. Babesia bovis proteins such as MSA-1, MSA-2b and MSA-2c are not fully studied both concerned with its structure and function. Our main objective of this study to perform sequence analysis, identify conserved motifs, structure analysis and homology modeling of MSA-1, MSA-2b and MSA-2c proteins by using proteomics tools and online prediction servers that helped to understand the proteins for drug design and vaccine.

# 2. Materials and Methods

## 2.1. Extraction of Protein Sequences

The protein sequences of Babesia bovis were extracted from NCBI (National Center for Biotechnology) (http://www.ncbi.nlm.nih.gov/). It is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The protein sequences were retrieved in FASTA format. The aim was to analyze and characterize the proteins by computational methods.

## 2.2. Identification of Amino Acid Percentage Composition and Physico-Chemical Properties

The primary structure was predicted using the Expasy ProtParam server http://expasy.org/cgibin/ protparam). The parameters computed by ProtParam include the molecular weight (M.Wt), isoelectric point pI, amino acid composition, atomic composition, extinction coefficient (EC), estimated half-life, instability index(II), aliphatic index(AI) and grand average of hydropathicity (GRAVY). The amino acid and atomic compositions are self-explanatory. All the other parameters will be explained below [12].

### 2.2.1. Isoelectric Point (pI)

The calculated isoelectric point (pi) is useful for at pi the solubility is lost and the mobility in an electric field is zero. Isoelectric point is the pH at which the surface of the protein is covered with the charge but a net charge of the protein is zero.

### 2.2.2. Extinction Coefficients (EC)

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for analyzing a protein with a spectrophotometer when purifying it. It has been shown [13], that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of Tyrosine, Tryptophan and Cystine (Tyrosine does not absorb appreciably at wavelengths >260 nm, while Cystine does) at a given wavelength

### 2.2.3. Instability Index (II)

The instability index provides an estimate of the stability of a protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed [14], that there are certain dipeptides, the occurrence of which is significantly different in the unstable proteins compared with those in the stable ones. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.

### 2.2.4. Aliphatic Index (AL)

The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins [15].

### 2.2.5. Grand Average of Hydropathy (GRAVY)

The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values [16] of all the amino acids, divided by the number of residues in the sequence.

### 2.2.6. Estimated Half-Life

The half-life is a prediction of the time required for half of a protein in a cell to degrade after its synthesis. ProtParam relies on the "N-end rule", which relates the half-life of a protein to the identity of its N-terminal residue; the prediction is given for three model organisms; human, yeast, and *E. coli*. The identity of the N terminal residue of a protein plays an important role in determining its stability *in vivo*. Proteins have strikingly different half-lives *in vivo*, from seconds to hours, depending on the nature of the amino acid at the N terminus and the different models.

## 2.3. Fingerprinting Analysis

ScanProsite use of fingerprinting analysis; (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html) it is a large collection of biologically meaningful signatures that are described as patterns (regular expressions), used for short motif detection, or generalized profiles (weight matrices) for sensitive detection of larger domains. Each signature is linked to detailed annotation that provides useful biological information on the protein family, domain, or functional sites identified by the signature [17].

## 2.4. Hydrophobicity Analysis

Percentages of hydrophobic and hydrophilic residues were calculated from the percentage of Amino Acid composition.

## 2.5. Transmembrane Sequence Analysis

Transmembranase domains were predicted by using SOSUI server [18]. (http://harrier.nagahama-i-bio.ac.jp/sosui/SOSUI), which distinguishes between membrane and soluble proteins from amino acid sequences, and predicts the transmembrane helices for the former.

## 2.6. Prediction of Hydrophobic Residues

The hydrophobic residues were predicted by using Pepwheel (http://www.hpa-bioinfotools.org.uk/pise/pepwheel.html). Pepwheel program draws a helical wheel diagram for a protein sequence. This displays the sequence in a helical representation as if looking down the axis of the helix. It is useful for highlighting amphipathicity and other properties of residues around a helix. By default, aliphatic residues are marked with squares; hydrophilic residues are marked with diamonds, and positively charged residues with octagons, although this can be changed [19].

## 2.7. Prediction of Phosphorylation Sites

Phosphorylatiuon sites were predicted by using NetPhos (http://www.cbs.dtu.dk/services/NetPhos/). The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins [20].

## 2.8. Secondary Structure Generation

GOR IV (http://npsa-pbil.ibcp.fr/cgibin/npsa_automat.pl?page=npsa_gor4.html): The GOR (Garnier, Osguthorpe, and Robson) method is an information-based method for the prediction of secondary structures of proteins. There is no defined constant decision. GOR IV uses all possible pair frequencies within the window of 17 amino acid residues. The program gives two outputs: one is eye friendly and gives the sequence and the predicted secondary structure in parallel rows, with symbols H= α helix, E=extended or β strand and C=coil; the second gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one with the highest probability-compatible structure with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues [21].

## 2.9. Homology Modeling

The 3-dimensional structure expectation we utilized CPH models 3.2 servers (http://www.cbs.dtu.dk/services/CPHmodels/) to predict the PDB of proteins. It is a protein homology modeling server, where the template recognition is based on profile-to-profile arrangement, guided by secondary structure and presentation forecasts. Visualization and characterization of the protein model were done by Chimera (version 1.8) software [22].

### 2.10. Validation of 3D Model

Structural validation of protein model was done by ERRAT is a program for verifying protein structures determined by crystallography. Error values are plotted as a function of the position of a sliding 9-residue window. The error function is based on the statistics of non-bonded atom-atom interactions in the reporting structure (compared to a database of reliable high-resolution structures) (http://services.mbi.ucla.edu/ERRAT/) [23].

## 3. Result

In the first step, we selected sequences for MSA-1, MSA-2b and MSA-2c proteins of Babesia bovis from cattle in Sri Lankan. The proteins considered in this study showed in (Table 1). These protein sequences were retrieved from NCBI, a public domain database, in FASTA format and used for further analysis.

**Table 1.** Proteins considered for the study.

| Accession number [NCBI] | Length [Amino Acid] | Description |
|---|---|---|
| BAN78727.1 | 289 | Merozoite surface antigen-1 (babesia bovis) MSA-1 |
| BAN78765.1 | 275 | Merozoite surface antigen-2b (babesia bovis) MSA-2b |
| BAN78765.1 | 260 | Merozoite surface antigen-2c (babesia bovis) MSA-2c |

### 3.1. Primary Structure Prediction

The parameters computed by ProtParam. The Atom composition of proteins showed in (Table 2). The percentage of hydrophobic and hydrophilic residue content of the proteins showed in (Table 3). The amino acid composition showed in (Table4).

**Table 2.** Atom composition and formula of MSA-1, MSA-2b and MSA-2c proteins.

| Atomic composition | MSA-1 | MSA-2b | MSA-2c |
|---|---|---|---|
| Carbon (C) | 1353 | 1367 | 1300 |
| Hydrogen (H) | 2107 | 2124 | 2037 |
| Nitrogen (N) | 359 | 338 | 321 |
| Oxygen (O) | 447 | 425 | 409 |
| Sulfur (S) | 10 | 13 | 8 |
| Formula | $C_{1353}H_{2107}N_{359}O_{447}S_{10}$ | $C_{1367}H_{2124}N_{338}O_{425}S_{13}$ | $C_{1300}H_{2037}N_{321}O_{409}S_{8}$ |
| Total number of atoms | 4276 | 4267 | 4075 |

**Table 3.** Hydrophilic and hydrophobic residue content.

| Proteins | Percentage of Hydrophobic Residues | Percentage of Hydrophilic Residues | Net Hydrophobic Residues Content |
|---|---|---|---|
| MSA-1 | 46.7 | 53.3 | Low |
| MSA-2b | 46.8 | 53.2 | Low |
| MSA-2c | 45.5 | 54.5 | Low |

**Table 4.** Amino acid composition (in %) MSA-1, MSA-2b and MSA-2c proteins using ProtParam tool.

| Amino acid | MSA-1 | MSA-2b | MSA-2c |
|---|---|---|---|
| Ala (A) | 7.6% | 6.9% | 6.9% |
| Arg (R) | 1.0% | 1.1% | 1.2% |
| Asn (N) | 5.2% | 4.4% | 4.6% |
| Asp (D) | 5.9% | 6.2% | 6.5% |
| Cys (C) | 1.4% | 2.2% | 0.8% |
| Gln (Q) | 3.1% | 3.3% | 3.8% |
| Glu (E) | 6.6% | 6.2% | 7.3% |
| Gly (G) | 7.6% | 3.6% | 4.2% |
| His (H) | 2.4% | 1.8% | 1.2% |
| Ile (I) | 2.4% | 4.0% | 2.7% |
| Leu (L) | 6.2% | 10.2% | 11.2% |
| Lys (K) | 8.0% | 8.4% | 9.2% |
| Met (M) | 2.1% | 2.5% | 2.3% |
| Phe (F) | 6.2% | 6.2% | 6.9% |
| Pro (P) | 5.2% | 6.9% | 3.5% |
| Ser (S) | 12.5% | 9.8% | 8.5% |
| Thr (T) | 6.9% | 8.4% | 9.2% |
| Trp (W) | 0.0% | 0.0% | 0.0% |
| Tyr (Y) | 1.7% | 3.6% | 3.1% |
| Val (V) | 8.0% | 4.4% | 6.9% |
| Pyl (O) | 0.0% | 0.0% | 0.0% |
| Sec (U) | 0.0% | 0.0% | 0.0% |
| (B) | 0.0% | 0.0% | 0.0% |
| (Z) | 0.0% | 0.0% | 0.0% |

### 3.2. Physicochemical Analysis

The Physio-chemical properties include the molecular weight, isoelectric point, total number of positive and negative residues, extinction coefficient, and grand average of hydropathicity were depicted in (Table 5).

**Table 5.** Physical and Chemical Characters of the Primary Structures of Predicted Proteins in Theory.

| Types of Proteins | M.wt | -R | +R | EC( M-1 cm-1) | GRAVY | pI |
|---|---|---|---|---|---|---|
| MSA-1 | 30875.3 | 36 | 26 | 7700 | -0.380 | 5.20 |
| MSA-2b | 30510.6 | 34 | 26 | 15275 | -0.292 | 5.20 |
| MSA-2c | 28963.8 | 36 | 27 | 12045 | -0.257 | 4.97 |

M.Wt. molecular weight; −R: number of negative residues (Arg + Lys); +R: number of positive residues (Asp + Glu); EC: extinction coefficient at 280 nm; GRAVY: grand average hydropathy, pI: isoelectric point.

### 3.3. Half Lifetime, Stability and Solubility

The estimated half-life, instability index(II) and aliphatic index(AI) of proteins showed in (Table 6).

**Table 6.** Estimated half-life of MSA-1, MSA-2b and MSA-2c proteins using ProtParam tool.

| Estimated have life | MSA-1 | MSA-2b | MSA-2c |
|---|---|---|---|
| mammalian reticulocytes, in vitro | 100 Hours | 30 Hours | 30 Hours |
| yeast, in vivo | >20 | >20 | >20 |
| Escherichia coli, in vivo). | >10 | >10 | >10 |
| N-terminal | V(val) | M(Met) | M(Met) |
| Instability index(II) | 39.89 | 43.29 | 31.33 |
| Aliphatic index(AL) | 64.43 | 74.87 | 81.00 |

## 3.4. Fingerprinting Analysis

The possible domains and characteristic motifs and patterns contained in MSa-1, MSA-2b and MSa-2c were investigated by ScanProsite. The result showed in (Table.7, 8 and 9).

**Table 7.** MSA-1 protein expression profiles using Scan Prosit.

| Motif side | Predicted feature | phosphorylation site | Position in protein sequence | Sequence |
|---|---|---|---|---|
| PKC_PHOSPHO_SITE | MOD_RES | Phosphoserine | 37-39 | SsR |
| N-glycosylation site | CARBOHYD | N-linked | 56-59 | NASL |
| | | N-linked | 148-151 | NKTK |
| CK2_PHOSPHO_SITE | MOD_RES | Phosphoserine | 108-111 | SaeE |
| | MOD_RES | Phosphoserine | 184-187 | TknD |
| | MOD_RES | Phosphothreonine | 210-213 | SngD |
| | MOD_RES | Phosphoserine | 217-220 | SsgD |
| | MOD_RES | Phosphoserine | 229-232 | SaaE |
| | MOD_RES | Phosphoserine | 249-252 | SptE |
| | MOD_RES | Phosphoserine | 251-154 | TepE |
| N-myristoylation site | _ | _ | 208 - 213 | GSsnGD |
| | _ | _ | 208 - 213 | GVpsAA |
| | _ | _ | 226 – 231 | GNlnGH |
| | _ | _ | 258 – 263 | GSsfTF |
| | _ | _ | 268 - 273 | GLtvAT |

MOD_RES= (Modified residue). CARBOHYD=(Glycosylation)

**Table 8.** MSA-2b protein expression profiles using scanprosit.

| Motif side | Predicted feature | phosphorylation site | Position in protein sequence | Sequence |
|---|---|---|---|---|
| CK2_PHOSPHO_SITE | MOD_RES | Phosphoserine | 20-23 | SasE |
| | MOD_RES | Phosphothreonine | 32-35 | TlhD |
| | MOD_RES | Phosphothreonine | 48-51 | TkeE |
| | MOD_RES | Phosphoserine | 93-96 | SlfD |
| | MOD_RES | Phosphoserine | 114-117 | SllE |
| | MOD_RES | Phosphoserine | 157-160 | SakD |
| | MOD_RES | Phosphoserine | 192-195 | TyeE |
| | MOD_RES | Phosphoserine | 125-128 | SqpD |
| | MOD_RES | Phosphothreonine | 244-247 | TsaD |
| | MOD_RES | Phosphoserine | 157-159 | SaK |
| TYR_PHOSPHO_SITE | - | - | 159-166 | Kds. DvkdY |
| CAMP_PHOSPHO_SITE | - | - | 199-202 | KKpS |
| N-myristoylation site | - | - | 213 - 218 | GTpgAQ |
| | | | 216 - 221 | GAqpAA |
| | | | 254 - 259 | GSsfTF |
| | | | 261 - 266 | GLtvAT |
| | | | 213 - 218 | GTpgAQ |
| N-glycosylation site | CARBOHYD | N-linked | 223-226 | NTSQ |

MOD_RES= (Modified residue). CARBOHYD=(Glycosylation)

**Table 9.** MSA-2c protein expression profiles using Scan Prosit.

| Motif side | Predicted feature | phosphorylation site | Position in protein sequence | Sequence |
|---|---|---|---|---|
| N-glycosylation site | CARBOHYD | N-linked | 17-20 | NATL |
| | CARBOHYD | N-linked | 193-196 | NFSA |
| | MOD_RES | Phosphoserine | 22-25 | SspE |
| | MOD_RES | Phosphoserine | 52-55 | SfgE |
| | MOD_RES | Phosphoserine | 60-63 | SfgE |
| | MOD_RES | Phosphoserine | 99-102 | SdfD |
| | MOD_RES | Phosphothreonine | 113-116 | TtkE |
| | MOD_RES | Phosphoserine | 148-151 | SgdD |
| | MOD_RES | Phosphoserine | 166-169 | SelD |
| | MOD_RES | Phosphothreonine | 205-208 | TgyE |
| PKC_PHOSPHO_SITE | MOD_RES | Phosphothreonine | 30-32 | TeR |
| | MOD_RES | Phosphothreonine | 113-115 | TtK |
| | MOD_RES | Phosphothreonine | 202-204 | TaK |
| | MOD_RES | Phosphoserine | 234-236 | SaK |
| N-myristoylation site | - | - | 70 - 75 | GMeaTS |
| | - | - | 127 - 132 | GMkeNI |
| | - | - | 239 - 244 | GSsfTF |
| | - | - | 246 - 251 | GLtvAT |

MOD_RES= (Modified residue). CARBOHYD=(Glycosylation)

## 3.5. Prediction of the Transmembarne Site

The SOSUI server performed the identification of transmembrane region. The transmembrane regions and their length were tabulated in (Table 10).

**Table 10.** Transmembrane sequence analysis of SOSUI server.

| Proteins | N terminal | Transmembrane region | C terminal | Type | Length | Average of hydrophobicity |
|---|---|---|---|---|---|---|
| MSA-1 | 1 | VLTFLMTAVCCAASFA | 16 | PRIMARY | 16 | -0.379931 |
| MSA-2b | 2 | IGKIFLLTACCCASLLSVSASE | 23 | PRIMARY | 22 | -0.292363 |
| MSA-2c | 1 | MVSFNIITVALCSTLFNATLAS | 22 | SECONDARY | 22 | -0.232576 |
| MSA-2c | 239 | GSSFTFGGLTVATLCYFVLSAF | 260 | PRIMARY | 22 | -0.232576 |

## 3.6. Helical Wheel Predicted by Pepwheel Program

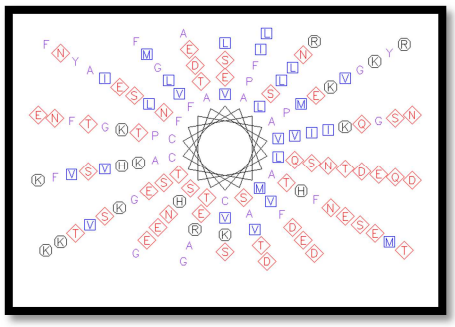

**Fig. 1.** Helical wheel predicted by Pepwheel for MSA-1 Protein. Aliphatic residues are marked with blue squire(V.I.L.M) Hydrophilic residues are marked with red octagons(T.E.Q.D.N.S). Hydrophobic residues are marked with free figure (G.A.C.P.F). Positively charged residues are marked with diamonds (R.K.H).

The Hydrophilic residues for protein sequences were predicted by using Pepwheel program. By default, aliphatic residues are marked with squares; hydrophilic residues are marked with diamonds, and positively charged residues with octagons.

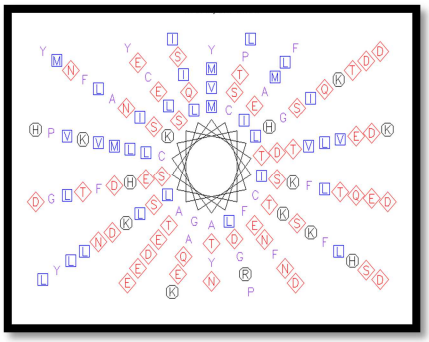

**Fig. 2.** Helical wheel predicted by Pepwheel for MSA-2b Protein. Aliphatic residues are marked with blue squire (V.I.L.M) Hydrophilic residues are marked with red octagons (T.E.Q.D.N.S). Hydrophobic residues are marked with free figure (G.A.C.P.F). Positively charged residues are marked with diamonds (R.K.H).

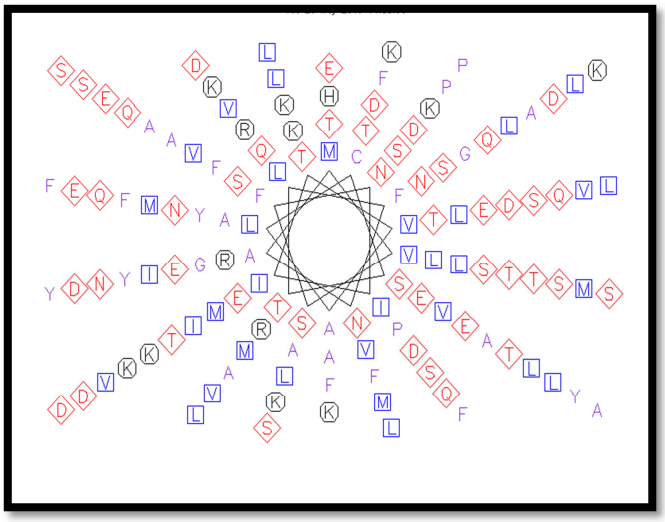

**Fig. 3.** Helical wheel predicted by Pepwheel for MSA-2c Protein. Aliphatic residues are marked with blue squire (V.I.L.M) Hydrophilic residues are marked with red octagons (T.E.Q.D.N.S). Hydrophobic residues are marked with free figure (G.A.C.P.F). Positively charged residues are marked with diamonds (R.K.H).

## 3.7. Prediction of Phosphorylation Sites

The NetPhos server predicted phosphorylation site of serine, threonine and tyrosine the result showed in (Table 11) and figure 4, 5 and 6.

**Table 11.** Phosphorylation sites predicted in MSa1, MSa2b and MSAc2 by using NetPhos.

| Proteins | Serine | Threonine | Tyrosine |
|---|---|---|---|
| MSA1 | 13 | 5 | 2 |
| MSA2b | 13 | 4 | 5 |
| MSA2c | 14 | 3 | 0 |

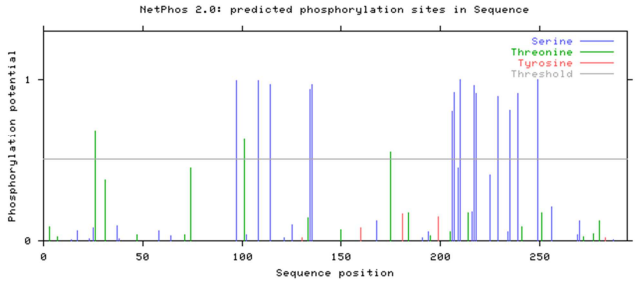


**Fig. 4.** Predicted Phosphorylation sites in MSA-1 protein of babesia bovis.

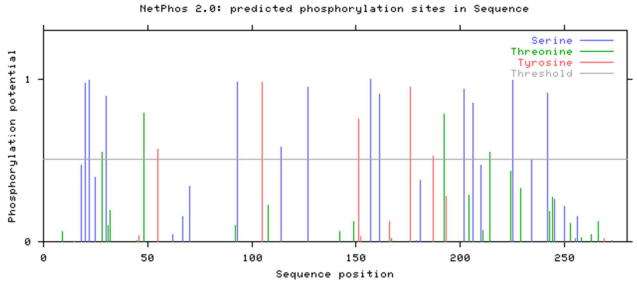


**Fig. 5.** Predicted Phosphorylation sites in MSA-2b protein of babesia bovis.

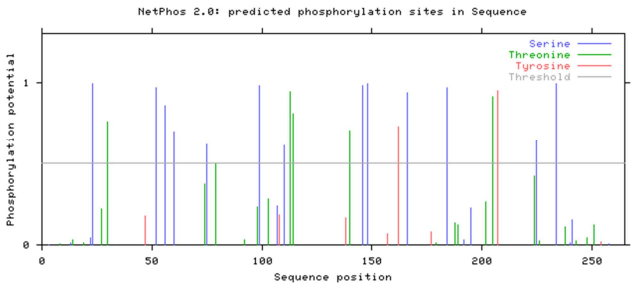


**Fig. 6.** Predicted Phosphorylation sites in MSA-2c protein of babesia bovis.

## 3.8. Protein Secondary Structure

The secondary structure predicted with the help of program GORIV. The result showed in (Table. 12) and figure 7, 8 and 9.

**Table 12.** Percentage of residues forming alpha, Extended strand and coil structures computed by GOR IV server.

| | MSA-1 | MSA-2b | MSA-2c |
|---|---|---|---|
| Alpha helix (Hh) | 77 (26.64%) | 86 (31.27%) | 123 (47.31%) |
| Extended strand (Ee) | 46 (15.92%) | 41 (14.91%) | 29 (11.15%) |
| Random coil (Cc) | 166 (57.44%) | 148 (53.82%) | 108 (41.54%) |

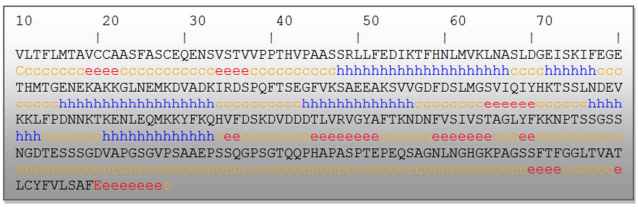


**Fig. 7.** MSA-1 Secondary protein structure. Alpha helix (h), Extended strand (Ee) and Random coil. Sequence length 289. (Mixed Structure).

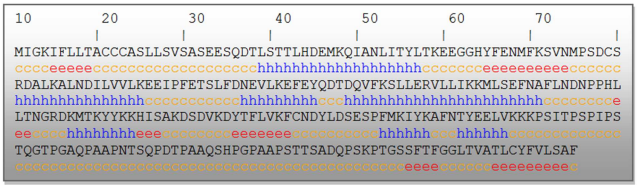


**Fig. 8.** MSA-2b Secondary protein structure. Alpha helix (h), Extended strand (Ee) and Random coil Sequence length 275. (Mixed Structure).

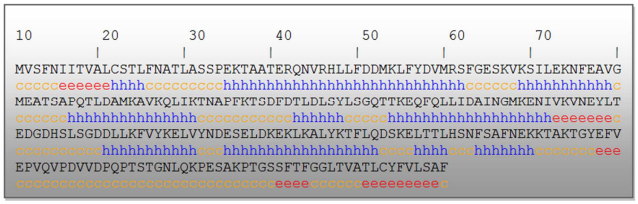


**Fig. 9.** MSA-2c Secondary protein structure. Alpha helix (h), Extended strand (Ee) and Random coil. Sequence length 260 (Mixed Structure).

## 3.9. 3D Structure of Protein

The 3-dimensional structure expectation we utilized CPH models 3.2 server. Visualization and characterization of the proteins model were done by Chimera (version 1.8) program. The result showed in figure 10, 11and 12.

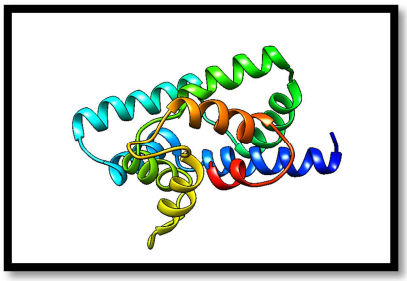

**Fig. 10.** Three dimensional structure of MSA-1 protein.

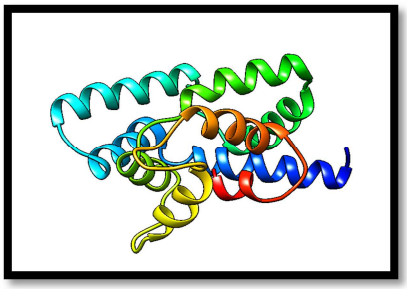

**Fig. 11.** Three dimensional structure of MSA-2b protein.

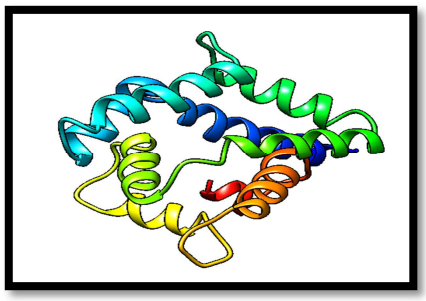

**Fig. 12.** Three dimensional structure of MSA-2c protein.

### 3.10. Validation of Proteins

The validation of the modeled structure was carried out using ERRAT. The result showed in figure 13, 14 and 15.

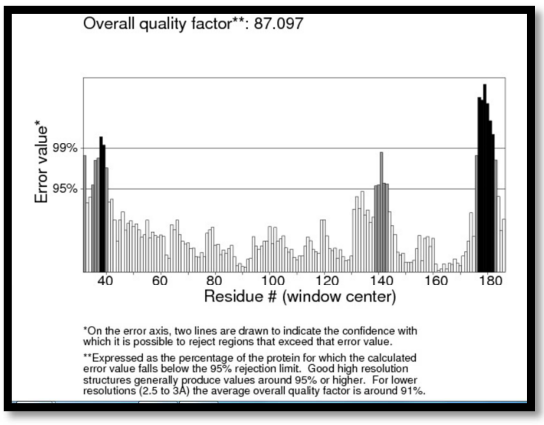


**Fig. 13.** The plot statistics value of response receiver regulator protein of MSA-1 obtained using (ERRAT) modeling tool. ERRAT results of CPH model for MSA-1. Black bars represent mis folded regions. On the error axis two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value CPH model.

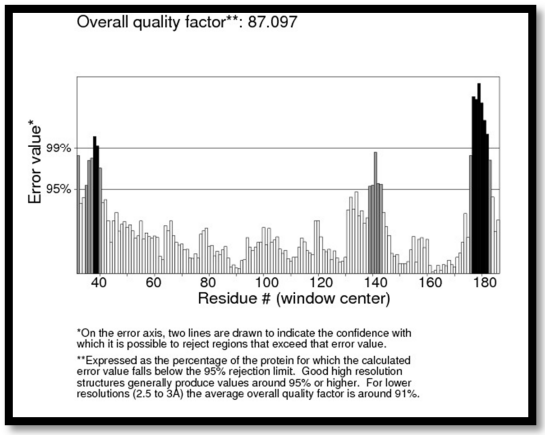


**Fig. 14.** The plot statistics value of response receiver regulator protein of MSA-2b obtained using (ERRAT) modeling tool. ERRAT results of CPH model for MSA-2b. Black bars represent mis folded regions. On the error axis two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value CPH model.

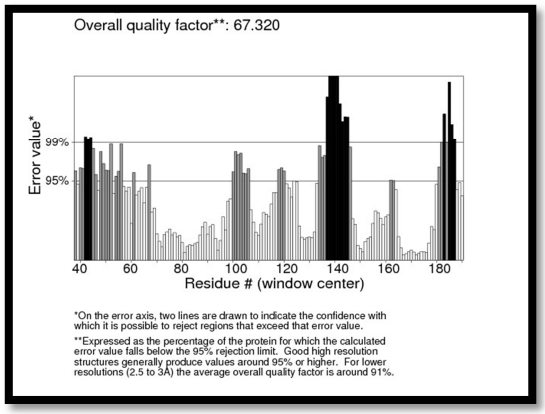


**Fig. 15.** The plot statistics value of response receiver regulator protein of MSA-2c obtained using (ERRAT) modeling tool. ERRAT results of CPH model for MSA-2c. Black bars represent mis folded regions. On the error axis two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value CPH model.

## 4. Discussion

In silico analysis is a suitable option for finding drug targets and choosing strategies to treat and control diseases. MSA-1, MSA-2b and MSA-2c proteins of babesia bovis have been chosen mainly to study their physicochemical properties, primary and secondary structures by biocomputation techniques. Physicochemical characterization studies give a good idea about the properties such as pI, EC, AI, GRAVY and II that is essential and vital in providing data about the proteins and their properties. The result of amino acid composition showed all proteins abundant from serine, while Pyrrolysine and Selenocysteine are absent. Also all proteins lack of cysteine, that is indicates of the low stability of proteins due to lack of disulphide bonds. The phisico-chemical properties showed that all proteins are hydrophilic in nature due to a low of Hydrophobic residues and all proteins are negatively

charged due to aspartic acid and glutamic acid these outcome significance with acidic nature of proteins. The computed pI value of proteins was less than 7 (pI<7) demonstrated that target proteins were considered as acidic. This additionally corresponds to the number of negatively charged residues presented in the proteins. The computed isoelectric point will be useful for developing buffer system for purification by isoelectric focusing method. Furthermore, we found the EC of MSA-1, MSA-2b and MSA-2c proteins at 280nm is low, this result indicates of low concentration of Cys, Trp and Tyr. The computed protein concentration and EC help in the quantitative study of protein–protein and protein–ligand interactions in solution. The high aliphatic index (AL) in MSA-2b and MSA-2c indicates that these proteins are stable for a wide range of temperature range, while the low (AL)of MSA-1 is indicative of instability and it is structure more flexible when compared to other Proteins .The Instability index ( II) less than 40 in MSA-1 and MSA-2c that infer these proteins are stable and (II) more than 40 in MSA-2b that indicate of unstable for this protein. The lower value of (GRAVY) in all proteins indicative of the possibility of better interaction with water. The anticipated half life time demonstrated that proteins were stable in the cell after synthesis. The server SOSUI classified all proteins as membrane proteins, and MSA-1 and MSA-2b proteins are primary in nature whereas MSA-2c is primary and secondary in nature. And the transmembrane region of all proteins is rich in hydrophobic amino acids. The helix of proteins is visualized using PepWheel. The phosphorylation site prediction showed serine is the most phosphorylated in all proteins. After detailed analysis of different proteins, we found the sequences belonging to the following families*: ASN_GLYCOSYLATION (*N-glycosylation site), Casein kinase II phosphorylation site *(CK2_PHOSPHO_SITE)*. Protein kinase C phosphorylation site (*PKC_PHOSPHO_SITE), MYRISTYL (*N-myristoylation site), Tyrosine kinase phosphorylation site (TYR_PHOSPHO_SITE) and *(CAMP_PHOSPHO_SITE)*. These families have a more or less significant relationship with the proteins roles in the cell signaling, regulation, and metabolism. The N-myristoylation is a form of lipid modification that targets a wide variety of eukaryotic proteins and plays important roles in cell physiology. In many instances, N-myristoylation alters the lipophilicity of the target proteins and facilitates its interaction with membranes, there by affecting its subcellular localization [12]. Phosphorylation by protein kinases is a major mechanism by which virtually every eukaryotic cellular activity is regulated, including proliferation, gene expression, metabolism, motility, membrane transport, and apoptosis [12]. In this work the secondary structure is predicted all proteins are a mixed protein having a composition of Helix, Extended strand and random coil, the outcome revealed that random helices were dominated among the secondary structure. Because only primary sequence information about MSA-1, MSA-2b and MSA-2c were available from NCBI, and no 3D structure information in the form of X-ray crystallographic data was available from the Protein Data Bank (PDB), Therefore, homology models for these proteins were developed. These structures will provide a good foundation for functional analysis of experimentally derived crystal structures and also for drug and vaccine designing. We found the quality of 3D structures less than 90% so we can say that the model is not good quality. However the lack of experimental structures of these proteins should be considered.

## 5. Conclusion

The knowledge of physicochemical properties, functional region, homology modeling and searching tool act as a fundamental way that will give various possible applications of protein. In addition, understand the structure of protein and its molecular organization is one of the requirements for developing effective vaccines and possible drugs against the disease.

In present study, the bioinformatics analysis of MSA-1, MSA-2b and MSA-2c proteins give the answers about the physicochemical properties, functional and structure of these proteins. Such information may aid in the development of new diagnostic tools, drugs and vaccines for treatment in endemic regions.

## References

[1] Hines S., McElwain T, Buening G, Palmer G. 1989. Molecular characterization of *Babesia bovis* merozoite surface proteins bearing epitopes immunodominant in protected cattle. Molecular Biochemical Parasitology. 37: 1–9.

[2] LeRoith T., Brayton A, Molloy J et al. 2005. Sequence variation and immunologic cross-reactivity among Babesia bovis merozoite surface antigen1 proteins from vaccine strains and vaccine breakthrough isolates. Infection And Immunity. 73: 5388–5394.

[3] Brown W., Palmer G. 1999. Designing blood-stage vaccines against Babesia bovis and Babesia bigemina. Parasitol today Journal. 15: 275–281.

[4] Mosqueda J., McElwain T, Palmer G. 2002. Babesia bovis merozoite surface antigen 2 proteins are expressed on the merozoite and sporozoite surface, and specific antibodies inhibit attachment and invasion of erythrocytes. Infection and Immunity. 70: 6448–6455.

[5] Florin-Christensen M., Suarez C, Hines S et al. 2002. The *Babesia bovis* merozoite surface antigen 2 locus contains four tandemly arranged and expressed genes encoding immunologically distinct proteins. Infection and Immunity journal. 70:3566–3575.

[6] Dubremetz J., Garcia-Reguet N, Conseil V, Fourmaux M. 1998. Apical organelles and host-cell invasion by Apicomplexa. International Journal of Parasitology. 28: 1007–1013.

[7] Suarez, C. E., Laughery, J. M., Bastos, R. G., Johnson, W. C., Norimine, J., Asenzo, G., Brown, W. C., Florin-Christensen, M. and Goff, W. L. 2011. A novel neutralization sensitive and subdominant RAP-1-related antigen (RRA) is expressed by Babesia bovis merozoites. Parasitology 18: 1–10.

[8] Carcy, B., Précigout, E., Schetters, T. and Gorenflot, A. 2006. Genetic basis for GPI-anchor merozoite surface antigen polymorphism of *Babesia* and resulting antigenic diversity. Veterinary Parasitology. 138: 33–49.

[9] Giancarlo R., Siragusa. A, Siragusa. E. et al. 2007. A Basic Analysis Toolkit for Biological Sequences, Algorithms for Molecular Biology.2: 404-406.

[10] Ashokan.K, Pillai. M. 2008. Characterization of biomimetc material, Mussel adhesive protein, using computer tools and servers. The Internet Journal of Bioengineering. 4: 1-8.

[11] Gay C.G, Zuerner R, Bannantine J.P, Lillehoj H.S, Zhu J.J, Green R. Pastoret P.P. 2007.Genomics and vaccine development Revue scientifique et technique (International Office of Epizootics). 26: 49-67.

[12] Sheoran. S, Pandey. B, Sharma. P, et al. 2013.In silico comparative analysis and expression profile of antioxidant proteins in plants Genetics and Molecular Research. 12: 537-551.

[13] Rost.B, Eyrich. V. 2001. Large-scale analysis of secondary structure prediction. Proteins 45 Suppl. 5: S192–S199.

[14] Prajapati. C, Bhagat. C.2012. In-Silico Analysis and Homology Modeling of Targets Proteins for Clostridium botulinum. International Journal Pharmaceutical sciences and research. 3: 2050-2056.

[15] Mundaganore S., Mundagnore D, Ashokan V. 2013. In Silico Validation OF Middle East Respiratory Syndrome (MERS) Virus Protins For Better Drug Development. International Journal of Applied Sciences and Biotechnology. 1: 272-278.

[16] K. Guruprasad, B. Reddy, M. Pandit. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering. 4: 155-161.

[17] Ikai A. 1980.Thermostability and aliphatic index of globular proteins. Journal of Biochemistry. 88: 1895-1898.

[18] Kyte J and Doolittle R. 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 157: 105-132.

[19] Castro E. de, Sigrist. C.J., Gattiker. A, et al. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Research.34: 362-365.

[20] Blom. N. Gammeltoft. S, Brunak. S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology. 294: 1351-1362.

[21] Ramachandran G.N, Sasiskharan. V. 1968. Conformation of polypeptides and proteins. Advances in Protein Chemistry 23: 283–437.

[22] Garnier, J., Gibrat, J.F. and Robson B. (1996). GOR secondary structure prediction method version IV. In: R.F. Doolittle (Ed.) Computer Methods for Macromolecular Sequence Analysis, Methods in Enzymology, vol. 266. Academic Press, USA. pp. 540-553.

[23] Mohammed A., Wasal A, Entssar S et al. 2015. Europe Academic of Research. Bioinformatics approach prediction of nonsynonymous and miRNA-binding site polymorphisms within human *SclA9* and *Scl22A12* genes. Vol II: 15412-15429.