The Cell Surface 8 (2022) 100078

Contents lists available at ScienceDirect

The Cell Surface

journal homepage: www.sciencedirect.com/journal/the-cell-surface

In silico identification of *Theileria parva* surface proteins

Nitisha Gurav [a,1], Olivia J.S. Macleod [b], Paula MacGregor [c], R. Ellen R. Nisbet [a,*]

[a] School of Biosciences, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Leicestershire LE12 5RD, UK
[b] Department of Biochemistry, University of Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK
[c] School of Biological Sciences, University of Bristol, Bristol Life Sciences Building, 24 Tyndall Ave, Bristol BS8 1TQ, UK

ARTICLE INFO

Keywords:
Apicomplexa
Theileria
East Coast Fever
Parasitology

ABSTRACT

East Coast Fever is a devastating African cattle disease caused by the apicomplexan parasite, *Theileria parva*. Little is known about the cell surface, and few proteins have been identified. Here, we take an *in silico* approach to identify novel cell surface proteins, and predict the structure of four key proteins.

Introduction

Theileria parva is the causative agent of East Coast Fever (ECF) a lethal, tick-borne disease of cattle in sub-Saharan Africa. *T. parva* is an apicomplexan parasite, closely related to *Plasmodium*, the causative agent of malaria. Mortality levels vary from 3 to 80% depending on parasite strain and cattle breed, killing over one million cattle each year (Nene et al., 2016). The only drug licenced to treat *T. parva* is buparvaquone, which is over 30 years old. Although there is no reported resistance to buparvaquone in *T. parva*, there is rising levels of resistance in *Theileria annulata*, a related cattle pathogen (Mhadhbi et al., 2015).

The main mechanism for control is cattle dipping and vaccination. However, the current vaccination model is an 'infection and treat' model. Large numbers of infected ticks are produced, ground up, and frozen. The tick residue is transported in liquid nitrogen and injected to the cattle to be vaccinated. As this causes disease, the cattle are then treated with high dose of tetracycline to prevent the infection taking hold (MacGregor et al., 2021; Nene et al., 2016). This model of vaccination has significant drawbacks:

the transport of the vaccine in liquid nitrogen is impractical in the context of rural Africa,
it requires the use of high levels of antibiotics, thus driving the rise of antibiotic resistance,
it requires the presence of a veterinarian, which significantly increases costs,
regional vaccines are needed
vaccinated cattle remain lifelong asymptomatic carriers of disease and cannot be exported.
A modern vaccine is urgently required (Nene and Morrison, 2016).

As with many eukaryotic parasites, *T. parva* differentiate between a series of different life-stages, each characterised by distinct morphologies and patterns of gene expression. These include distinct protein composition of the parasite cell surface between life-stages. As putative vaccine targets there are two key life-stages of interest: (i) the free-living sporozoite stage, where the parasite is released from the tick's salivary glands and passes into the cattle bloodstream, and (ii) the schizont stage, where *T. parva* causes cancer-like cell proliferation within the cattle lymphocytes. An ideal vaccine would target the sporozoite and/or schizont stages, giving rise to a CD4 and CD8 T cell immune response (Morrison et al., 2021).

Previous works have used *in silico* analysis of gene sequences from the *T. parva* Muguga strain genome and proteome to identify putative surface proteins, containing both a signal peptide and putative GPI-anchor addition sequence (Nyagwange et al., 2018a). These can be considered as highly likely to reside on the cell surface, and not internal membranes, as utilization of a GPI-anchor for membrane attachment is a surface-specific feature. Here, we sought to identify a longer candidate list of putative surface proteins, using a similar theoretical approach, but to instead identify proteins that are attached to the plasma membrane with one or more transmembrane domains.

* Corresponding author.
E-mail address: ellen.nisbet@nottingham.ac.uk (R.E.R. Nisbet).
[1] Now at: Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, United Kingdom.

https://doi.org/10.1016/j.tcsw.2022.100078
Received 23 February 2022; Received in revised form 19 May 2022; Accepted 20 May 2022
Available online 21 May 2022

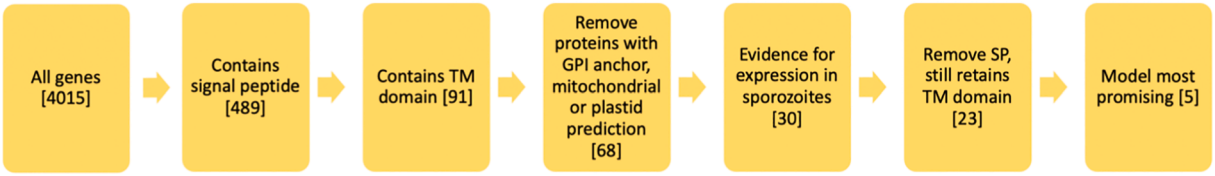

Fig. 1. Flow chart showing workflow to identify putative *T. parva* surface membrane proteins.

Table 1
Putative membrane proteins. Shortlist of putative membrane proteins with signal peptide, with experimental evidence for expression in *Theileria parva* sporozoites. * Sporozoite proteome: figure given is emPAI, a measure of the relative protein abundance in sample.

Gene name	Gene ID	Product name	Number of TM domains	TMs wo/signal peptide	Size/ kDa	Sporozoite proteome *
TpMuguga_01g00016	XP_765543.1	Vacuolar protein sorting/targeting protein 10	1	1	96	0.71
TpMuguga_01g00326	XP_765853.1	emp24/gp25L/p24 family/GOLD family protein	1	1	30	1.2
TpMuguga_01g00509	XP_766029.1	putative integral membrane protein	1	1	66	1.69
TpMuguga_01g00620	XP_766141.1	putative integral membrane protein	1	0	124	0.16
TpMuguga_01g00921		HAD ATPase P-type IC family protein	11	12	166	0.91
TpMuguga_01g00939	XP_766460.1	putative integral membrane protein	1	0	36	3.65
TpMuguga_01g01013	XP_766534.1	Sugar efflux transporter for intercellular exchange family protein	7	7	42	0.4
TpMuguga_01g01069	XP_766590.1	Sugar (and other) transporter family protein	12	11	52	0.43
TpMuguga_01g01091	XP_766612.1	Tp12	1	0	64	12.02
TpMuguga_01g01195	XP_766716.1	emp24/gp25L/p24 family/GOLD family protein	2	1	24	1.15
TpMuguga_02g00330	XP_764896.1	emp24/gp25L/p24 family/GOLD family protein	1	1	24	1.15
TpMuguga_02g00538	XP_765104.1	Sel1 repeat family protein	1	1	177	1.34
TpMuguga_02g00543	XP_765109.1	putative integral membrane protein	3	3	54	0.3
TpMuguga_02g00602	XP_765168.1	Thioredoxin family protein	1	1	25	19.43
TpMuguga_02g02055		putative integral membrane protein	1	0	117	4.39
TpMuguga_03g00168	XP_763186.1	putative integral membrane protein	1	1	269	4.18
TpMuguga_03g00175	XP_763193.1	putative integral membrane protein	2	1	36	2.68
TpMuguga_03g00264	XP_763282.1	S1/P1 Nuclease family protein	2	2	45	1.05
TpMuguga_03g00419	XP_763440.1	Thioredoxin family protein	1	0	62	33.82
TpMuguga_04g00068		putative integral membrane protein	1	0	141	1.32
	XP_763703.1	putative integral membrane protein	1	1	27	1.39
TpMuguga_04g00399	XP_764034.1	putative integral membrane protein	2	1	41	7.78
TpMuguga_04g00649	XP_764284.1	GPI transamidase subunit Gpi16	1	1	68	0.62
TpMuguga_04g00668	XP_764304.1	unspecified product	1	1	46	0.83
TpMuguga_04g00917	XP_764554.1	SVSP family protein	1	0	60	0.48

Methodology

The *T. parva* Mugaga strain genome (causative agent of ECF) (Hayashida et al., 2013; Tretina et al., 2020) was computationally screened for all proteins with a putative signal peptide using SignalP 5.0 (Almagro Armenteros et al., 2019) plus one or more transmembrane domain using HMMer (Finn et al., 2011). 91 protein-encoding genes with a signal peptide and a transmembrane domain were identified, Fig. 1.

Inclusion of proteins that have putative transmembrane domains will inevitably include proteins that are internal (e.g. in the endosomal pathway) and so not exposed to the external environment. We therefore removed all proteins that were predicted to localise to the mitochondria or apicoplast (with bipartide targeting) using TargetP, Fig. 1 (Armenteros et al., 2019). Proteins with a putative GPI-anchor addition sequence were also excluded as these have been considered previously (Nyagwange et al., 2018a; Pierleoni et al., 2008). This led to a final longlist of 68 proteins, shown in Supplementary Table 1. These represent a maximal list of proteins of interest.

Next, we looked for proteins with evidence of expression in the sporozoite. A whole-cell proteomic dataset is available for the *T. parva* sporozoite stage, providing evidence for 2007 proteins, representing about 50% of the total predicted genes (Nyagwange et al., 2018b). Of the 68 proteins of interest (Suppl. Table 1), a total of 30 were present in the proteome dataset, Table 1.

To further refine our search, we removed all proteins predicted to contain a single transmembrane domain where it is located within the first 30 aa; these most likely represent false positive results, where the signal peptide domain is incorrectly identified as a containing a transmembrane domain. However, we retained TpMuguga_01g01091 (Tp12), a known antigen (Morrison et al., 2015).

Following analysis of localization and function of homologous proteins in other species, and selection for proteins with the largest predicted external size, we selected the five most promising candidates, plus Tp12 (TpMuguga_01g01091). For each protein, we removed the signal peptide and predicted the structure using Phyre2, Swiss-model and AlphaFold (Jumper et al., 2021; Kelley et al., 2015; Waterhouse et al., 2018). The results are shown in Table 2. TpMuguga_03g00168 was too large for AlphaFold, giving rise to no hits in Phyre2 and no hits in Swiss-model. Modelling of TpMuguga_01g01091 (Tp12) gave rise to a disordered protein with little confidence in AlphaFold. The four protein structures that were able to be modelled are shown in Fig. 2.

Two proteins (TpMuguga_01g01013 and TpMuguga_01g01069) are predicted sugar transporters, and TpMuguga_01g00921 is predicted to be a cation transporting ATPase. The remaining membrane protein TpMuguga_03g00168 has unknown function. All four have external facing regions that may offer target epitopes for novel vaccines. To test this, we utilized the epitope prediction tool NetMHC 4.0 which predicts peptide-MHC Class 1 binding for known proteins (Jurtz et al., 2017; Nielsen et al., 2018). Each of the four proteins were analyzed for the bovine BoLA-T2a allele, with peptide length between 8 and 14 aa. The results are shown in Table 3, giving the numbers of peptides identified as strong or weak binders. These indicate that these proteins are highly likely to contain epitopes that will be candidates for vaccination.

As many cell surface proteins contain post-translational modifications, we used bioinformatic tools to search for glycosylation (Steentoft et al., 2013) and phosphorylation sites (Wang et al., 2020), as shown in

Table 2
Short list of the six most promising membrane proteins. The results of structure prediction by AlphaFold, Phyre2 and Swiss-model are shown, together with the final model choice for modelling in Fig. 2.

Long list of IDsPredicted mature protein (aa#)		ColabFold: AlphaFold2 using MMseqs2	Region (aa#)	Average pLDDT	PDB Phyre top Function top hit (confidence %)		Region (aa#)	PDB Swissmodel top hit (GMQE)	Function top hit	Region (aa#)	Final model
TpMuguga_01g00921	18–1450	not performed (>1000aa)	–	–	7OP1 (100%)	cation transporting ATPase	233–1449	7OP1 (0.41)	cation transporting ATPase	172–1428	Phyre2
TpMuguga_01g01013	16–379	model model	16–379	67.9	5XPD (100%)	sugar transporter sugar	147–367	5XPD (0.38) 6 M20	sugar transporter	144–356	AlphaFold2
TpMuguga_01g01069	20–474		20–474	90.7	6RW3 (100%)	transporter	21–465	no model	sugar transporter	23–472	AlphaFold2 not used
TpMuguga_01g01091	23–568	model	23–568	56.3	no model	–	–		–	–	
TpMuguga_02g00543	19–472	model	19–472	61.5	no model	–	–	no model	–	–	AlphaFold2
TpMuguga_03g00168	19–2361	not performed (>1000aa)	–	–	no model	–	–	no model	–	–	–

Table 3. These indicated that all four proteins are glycosylated, while two of the four also contain phosphorylation sites. It should be noted that the training datasets for post-transcriptional modification prediction software were not obtained from *T. parva*, so these results will need to be validated experimentally.

Conclusions

Recent advances in *Plasmodium* cell biology have identified numerous surface proteins which could be part of a future multi-valent subunit vaccine against malaria. In contrast, very few surface proteins have been characterized in *T. parva*. However, many surface proteins that have been identified give rise to neutralizing antibodies, suggesting that they would be potential vaccine targets (Musoke et al., 1992; Nyagwange et al., 2018a).

In the absence of a proteome-based experimental approach to identify further surface proteins, a recent *in silico* study identified 21 putative GPI-anchored surface proteins (Nyagwange et al., 2018a). Of the six expressed GPI anchored surface proteins, four gave rise to sporozoite neutralizing antibodies.

Here, we took an *in silico* approach to identify transmembrane domain proteins on the *T. parva* cell surface, thus further increasing the numbers of putative surface proteins for future vaccinology attempts. Of the six most promising candidate proteins, three (TpMuguga_01g01013, TpMuguga_01g00921 and TpMuguga_01g01069) are likely to be involved in nutrient or cation uptake. This is unsurprising, as the parasite must interact with the environment.

Two of these proteins (TpMuguga_01g01069 and TpMuguga_01g00921) also have high levels of expression in the schizont stage, suggesting that in addition, they play a role in nutrient acquisition during this intracellular stage (Tonui et al., 2018). The schizont stage must have very high requirement for glucose, due to the rapid rate of cell proliferation. In *Plasmodium*, chemical inhibition of a key hexose transporter suppresses the growth of the parasite (Jiang et al., 2020). As TpMuguga_01g01069 is the only sugar transporter predicted to have a signal peptide and thus a surface localization in *T. parva*, it may also be a good drug target. A fourth protein (TpMuguga_03g00168) is a clear membrane protein with no predicted function. An experimental based approach will be required to obtain structures of TpMuguga_03g00168 and TpMuguga_01g01091 (Tp12).

Epitope prediction software suggests that these four proteins may be immunogenic, with predicted strong binding peptides by NetMHC (Jurtz et al., 2017). Although laboratory-based methodologies will be required to test these predictions, the results suggest that these proteins have potential as vaccine targets.

One of the challenges of *T. parva* research is the absence of a system for stable genetic manipulation, so it is not possible to confirm if these four proteins are essential, nor confirm the localization of proteins with tagging. An alternative approach would be to take a whole-cell spatial proteomic localization technique which would give localization information for all cellular proteins (Lundberg and Borner, 2019). An analysis of the related apicomplexan parasite *Toxoplasma* using hyperLOPIT (hyperplexed localisation of organelle proteins by isotope tagging) identified 110 integral surface proteins and 71 peripheral surface proteins (Barylyuk et al., 2020). A similar proteomic-based approach of *T. parva* would transform our understanding of this parasite, especially at the cell surface, and provide a massive leap forward in the quest to develop a modern *T. parva* vaccine.

NG was funded by a University of Nottingham Developing Solutions Masters Scholarship and by the J N Tata endowment, India.

CRediT authorship contribution statement

Nitisha Gurav: Methodology, Investigation. **Olivia J.S. Macleod:** Methodology, Investigation. **Paula MacGregor:** Conceptualization, Methodology, Writing – review & editing. **R. Ellen R. Nisbet:**

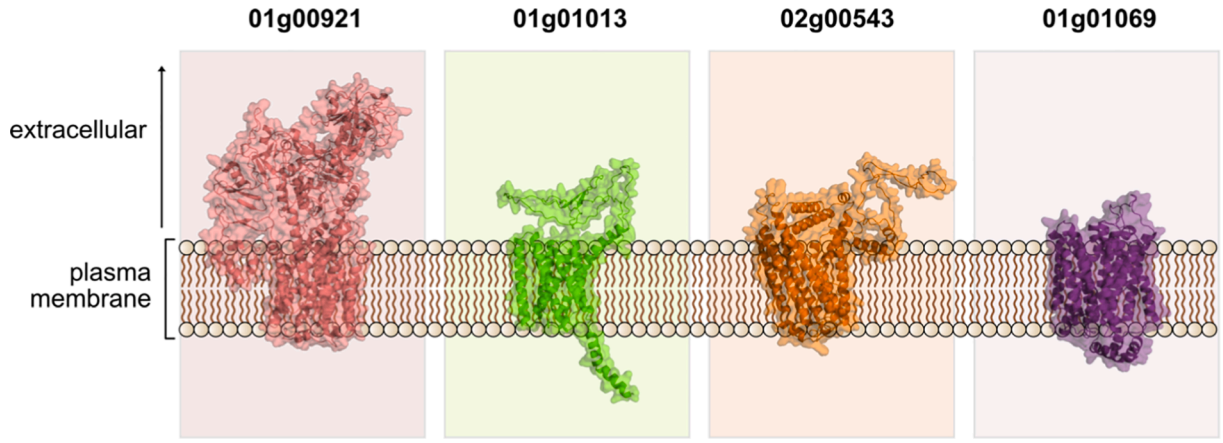

Fig. 2. Predicted protein structures. 01g00921 encodes a cation transporter ATPase (red; Phyre2 model shown), 01g01013 encodes a sugar transporter (green; AlphaFold2 model shown), the function of 02g00543 is unknown (orange) and 01g01069 encodes a sugar transporter (purple, modelled with AlphaFold2).

Table 3
Epitope and post-transcriptional modification. Epitope prediction was carried out using NetMHC 4.0, with the bovine BoLA-T2a antigen. The number of strong binding (SB) epitopes (with a rating < 0.5) and weak binding (WB) epitopes (with a rating < 2) is given for each protein. Identification of putative glycosylation was carried out using NetOGlyc 4.0, and phosphorylation sites by MusiteDeep.

Protein	SB	WB	glycosylation sites	phosphorylation Sites
Tp01g00921	43	206	5	27
Tp01g01013	7	37	21	0
Tp02g00543	20	53	7	7
Tp01g01069	13	51	1	0

Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tcsw.2022.100078.

References

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 374 (37), 420–423.
Armenteros, J.J.A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance 2, e201900429.
Barylyuk, K., Koreny, L., Ke, H., Butterworth, S., Crook, O.M., Lassadi, I., Gupta, V., Tromer, E., Mourier, T., Stevens, T.J., Breckels, L.M., Pain, A., Lilley, K.S., Waller, R. F., 2020. A Comprehensive Subcellular Atlas of the *Toxoplasma* Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. Cell Host Microbe 28, 752–766.
Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.
Hayashida, K., Abe, T., Weir, W., Nakao, R., Ito, K., Kajino, K., Suzuki, Y., Jongejan, F., Geysen, D., Sugimoto, C., 2013. Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. DNA Res. 20, 209–220.
Jiang, X., Yuan, Y., Huang, J., Zhang, S., Luo, S., Wang, N., Pu, D., Zhao, N., Tang, Q., Hirata, K., Yang, X., Jiao, Y., Sakata-Kato, T., Wu, J.W., Yan, C., Kato, N., Yin, H., Yan, N., 2020. Structural Basis for Blocking Sugar Uptake into the Malaria Parasite *Plasmodium falciparum*. Cell 183, 258–268.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nat 2021 5967873 596, 583–589.
Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., Nielsen, M., 2017. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J. Immunol. 199, 3360–3368.
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 106 (10), 845–858.
Lundberg, E., Borner, G.H.H., 2019. Spatial proteomics: a powerful discovery tool for cell biology. Nat. Rev. Mol. Cell Biol. 20, 285–302.
MacGregor, P., Nene, V., Nisbet, R.E.R., 2021. Tackling protozoan parasites of cattle in sub-Saharan Africa. PLoS Pathog. 17, e1009955.
Mhadhbi, M., Chaouch, M., Ajroud, K., Darghouth, M.A., BenAbderrazak, S., 2015. Sequence polymorphism of cytochrome b gene in *Theileria annulata* Tunisian isolates and its association with buparvaquone treatment failure. PLoS ONE 10 (6), e0129678.
Morrison, W.I., Aguado, A., Sheldrake, T.A., Palmateer, N.C., Ifeonu, O.O., Tretina, K., Parsons, K., Fenoy, E., Connelley, T., Nielsen, M., Silva, J.C., 2021. CD4 T Cell Responses to *Theileria parva* in Immune Cattle Recognize a Diverse Set of Parasite Antigens Presented on the Surface of Infected Lymphoblasts. J. Immunol. 207, 1965–1977.
Morrison, W.I., Connelley, T., Hemmink, J.D., MacHugh, N.D., 2015. Understanding the Basis of Parasite Strain-Restricted Immunity to *Theileria parva*. Annu. Rev. Anim. Biosci. 3 (1), 397–418.
Musoke, A., Morzaria, S., Nkonge, C., Jones, E., Nene, V., 1992. A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proc. Natl. Acad. Sci. U.S.A. 89, 514–518.
Nene, V., Kiara, H., Lacasta, A., Pelle, R., Svitek, N., Steinaa, L., 2016. The biology of *Theileria parva* and control of East Coast fever – Current status and future trends. Ticks Tick Borne Dis. 7, 549–564.
Nene, V., Morrison, W.I., 2016. Approaches to vaccination against *Theileria parva* and *Theileria annulata*. Parasite Immunol. 38, 724–734.
Nielsen, M., Connelley, T., Ternette, N., 2018. Improved Prediction of Bovine Leucocyte Antigens (BoLA) Presented Ligands by Use of Mass-Spectrometry-Determined Ligand and in Vitro Binding Data. J. Proteome Res. 17, 559–567.
Nyagwange, J., Nene, V., Mwalimu, S., Henson, S., Steinaa, L., Nzau, B., Tijhaar, E., Pelle, R., 2018a. Antibodies to in silico selected GPI-anchored *Theileria parva* proteins neutralize sporozoite infection in vitro. Vet. Immunol. Immunopathol. 199, 8–14.
Nyagwange, J., Tijhaar, E., Ternette, N., Mobegi, F., Tretina, K., Silva, J.C., Pelle, R., Nene, V., 2018b. Characterization of the *Theileria parva* sporozoite proteome. Int. J. Parasitol. 48, 265–273.
Pierleoni, A., Martelli, P., Casadio, R., 2008. PredGPI: a GPI-anchor predictor. BMC Bioinf. 9, 392.
Steentoft, C., Vakhrushev, S.Y., Joshi, H.J., Kong, Y., Vester-Christensen, M.B., Schjoldager, K.T., Lavrsen, K., Dabelsteen, S., Pedersen, N.B., Marcos-Silva, L., Gupta, R., Bennett, E.P., Mandel, U., Brunak, S., Wandall, H.H., Levery, S.B., Clausen, H., 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 32 (10), 1478–1488.
Tonui, T., Corredor-Moreno, P., Kanduma, E., Njuguna, J., Njahira, M.N., Nyanjom, S.G., Silva, J.C., Djikeng, A., Pelle, R., 2018. Transcriptomics reveal potential vaccine antigens and a drastic increase of upregulated genes during *Theileria parva* development from arthropod to bovine infective stages. PLoS ONE 13, e0204047.
Tretina, K., Pelle, R., Orvis, J., Gotia, H.T., Ifeonu, O.O., Kumari, P., Palmateer, N.C., Iqbal, S.B.A., Fry, L.M., Nene, V.M., Daubenberger, C.A., Bishop, R.P., Silva, J.C., 2020. Re-annotation of the *Theileria parva* genome refines 53% of the proteome and

uncovers essential components of N-glycosylation, a conserved pathway in many organisms. BMC Genom. 21, 279.
Wang, D., et al., 2020. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res. 48. W140–W146.
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, W296–W303.