

Article

# Design of an Epitope-Based Vaccine Ensemble for Animal Trypanosomiasis by Computational Methods

Lucas Michel-Todó <sup>1,\*</sup>, Pascal Bigey <sup>2,3</sup>, Pedro A Reche <sup>4</sup>, María-Jesus Pinazo <sup>1</sup>, Joaquim Gascón <sup>1</sup> and Julio Alonso-Padilla <sup>1,\*</sup> 

<sup>1</sup> Barcelona Institute for Global Health (ISGlobal), Hospital Clínic - University of Barcelona, 08036 Barcelona, Spain; mariajesus.pinazo@isglobal.org (M.-J.P.); quim.gascon@isglobal.org (J.G.)

<sup>2</sup> Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), Université de Paris, UTCBS, F-75006 Paris, France; pascal.bigey@parisdescartes.fr

<sup>3</sup> PSL University, ChimieParisTech, F-75005 Paris, France

<sup>4</sup> Laboratory of Immunomedicine, Faculty of Medicine, University Complutense of Madrid, Ave Complutense S/N, 28040 Madrid, Spain; parecheg@med.ucm.es

\* Correspondence: lucas.michel@isglobal.org (L.M.-T.); julio.a.padilla@isglobal.org (J.A.-P.)

Received: 11 December 2019; Accepted: 13 March 2020; Published: 16 March 2020



**Abstract:** African animal trypanosomiasis is caused by vector-transmitted parasites of the genus *Trypanosoma*. *T. congolense* and *T. brucei brucei* are predominant in Africa; *T. evansi* and *T. vivax* in America and Asia. They have in common an extracellular lifestyle and livestock tropism, which provokes huge economic losses in regions where vectors are endemic. There are licensed drugs to treat the infections, but adherence to treatment is poor and appearance of resistances common. Therefore, the availability of a prophylactic vaccine would represent a major breakthrough towards the management and control of the disease. Selection of the most appropriate antigens for its development is a bottleneck step, especially considering the limited resources allocated. Herein we propose a vaccine strategy based on multiple epitopes from multiple antigens to counteract the parasites' biological complexity. Epitopes were identified by computer-assisted genome-wide screenings, considering sequence conservation criteria, antigens annotation and sub-cellular localization, high binding affinity to antigen presenting molecules, and lack of cross-reactivity to proteins in cattle and other breeding species. We ultimately provide 31 B-cell, 8 CD4 T-cell, and 15 CD8 T-cell epitope sequences from 30 distinct antigens for the prospective design of a genetic ensemble vaccine against the four trypanosome species responsible for African animal trypanosomiasis.

**Keywords:** African animal trypanosomiasis; epitope-based vaccine; B-cell epitopes; CD4 T cell; CD8 T cell; pDNA; string-of-beads

## 1. Introduction

African animal trypanosomiasis (AAT) is a debilitating disease that affects wildlife and domestic animals in vast areas of Africa, Latin America, and Asia [1]. Known as nagana or surra, it can be caused by different protozoan parasite species of the genus *Trypanosoma* (class Kinetoplastea; family Trypanosomatidae), which are transmitted by insect vectors. Pathogens and vectors co-evolved and their distribution varies geographically. *Trypanosoma congolense* and *T. brucei brucei*, the most studied of the four, are transmitted by hematophagous tsetse flies (genus *Glossina*), and along with mechanically transmitted *T. vivax* are abundant in Africa [1]. *T. evansi* is also mechanically transmitted by fly bites and along with *T. vivax* can be found in America and Asia [1]. Nevertheless, they all thrive in mammalian reservoirs free-living in their bloodstream and manage to persist there causing chronic infections. In livestock, these infections produce anemia, poor condition and reproduction, and death, which

leads to huge economic losses [2]. In Africa, AAT exerts a devastating toll and it is considered the most important disease driven limitation to livestock production [3].

Infections can be treated with the drugs diminazene aceturate and isometamidium chloride, but the appearance of resistances is common [4,5]. A prophylactic vaccine used alongside vector control strategies would be a major breakthrough towards controlling the disease impact, but efforts in this direction have had limited success and there is none licensed (reviewed in [6]). The plethora of immune evasion mechanisms described in trypanosomes do not allow optimism in this respect [7], but naturally acquired immunity against trypanosomiasis has been reported in humans and animals [4,8]. In addition, some protection has been determined in AAT pre-clinical models of infection upon immunization with recombinant invariant parasite antigens like  $\beta$ -tubulin [9] or with a plasmid encoding an invariant surface glycoprotein (ISG) [10]. The immunization with irradiated trypanosomes or their variant surface glycoproteins (VSGs) has also been described to protect cattle against an homologous challenge (reviewed in [11]). Moreover, in a biologically relevant cow infection model, trypanotolerant N'Dama breeds have higher antibody levels in comparison to trypanosusceptible Boran cattle [11]. Thereby, it should be possible to elicit specific immune responses that protect against AAT infections, and if these immunizations were elicited by invariant conserved antigens, such responses could protect against heterologous infections, too.

In light of the high biological complexity of the parasites, and the very limited resources dedicated to the development of vaccines for AAT, a reverse vaccinology approach could be the most adequate strategy to follow. Herein we have used the genomic information available on AAT-causing *Trypanosome* spp. (downloaded from TriTrypDB release version 39) [12], and a series of tailor-made and publicly available immunoinformatic tools, to provide a list of predicted epitope sequences that could be the basis of an epitope-based vaccine. Such a vaccine could be delivered as a genetic construct at a fraction of the production and deployment costs of other vaccination strategies. The criteria employed to select the epitopes included a strict sequence conservation, sub-cellular location and annotation of their antigens of origin, high binding affinity to antigen presenting molecules, and lack of cross-reactivity to other proteins present in cows, sheep, goats, and pigs. As a result, we ultimately prioritized 31 B-cell, 8 CD4 T-cell, and 15 CD8 T-cell epitope sequences from 30 distinct antigens for the design of an ensemble pan-AAT vaccine. The rationale of choosing a plasmid DNA (pDNA) construct for its delivery is also discussed.

## 2. Materials and Methods

### 2.1. Collection of Protein Sequences from *Trypanosome* Spp. That Cause AAT

We downloaded the whole annotated proteomes of all *Trypanosome* spp. that have been related to AAT from TriTrypDB (release v39) [12]. We merged them to obtain a single file containing all annotated protein-coding sequences (CDS) available at the TriTrypDB resource from *T. brucei* Lister427, *T. brucei* TREU927, *T. congolense* IL3000, *T. evansi* STIB805, and *T. vivax* Y486.

### 2.2. Generation of Clusters

We used CD-HIT, running stand-alone with default settings, to reduce the redundancy of our *Trypanosome* spp. protein collection, clustering them with a shared identity >90% [13]. The resulting 28,709 clusters were filtered down by means of custom scripts to retain only those that contained at least one protein sequence from each of the 5 proteomes included in the analysis. This way we ensured that clusters considered for further analysis represented the protein diversity amongst all available sequences. Protein sequences within each of the remaining 115 clusters were aligned using MUSCLE (software version 3.8.31) [14] to obtain an equivalent number of multiple sequence alignments (MSA) with their corresponding consensus sequences. Nonetheless, instead of the consensus sequence that is built with pieces from different sequences and thus being unreal, we kept as a reference of each MSA the sequence in the alignment that was most similar to the consensus one.

### 2.3. Generation of the Invariable Proteome of Trypanosome Spp. Responsible for AAT

The Shannon entropy ( $H$ ) measure [15] was used to assess sequence variability on every position of the MSA according to the equation (Equation (1)):

$$H = - \sum_{i=0}^M P_i \log_2 P_i \quad (1)$$

where  $P_i$  is the frequency of an amino acid of type  $i$ , and  $M$  is the number of totally different amino acids (20).  $H = 0.0$  is equivalent to no variation in a given position among the studied protein sequences within the alignment, whereas higher values correspond to higher variation in that position [16]. Gaps were considered as data. Those residues at positions with an entropy value of  $H > 0.0$  were masked in the assigned protein reference sequence, substituting the residue symbol by an asterisk symbol (\*). As a result, we obtained a masked FASTA file which represented the conserved proteome of AAT trypanosomes, which we used to predict conserved epitopes within.

### 2.4. Prediction of B-Cell Epitopes

We identified the invariant regions of at least 15 residues long in the masked proteome FASTA file and created another FASTA file with the sequences of their corresponding antigens of origin. We then curated this list to retain only those fragments whose antigens of origin were expected to be surface-exposed accordingly to their annotation, as well as those whose antigens were not predicted with high reliability to be part on intracellular proteins. BepiPred2.0 was run on the whole protein sequences of the curated list with a threshold set at  $>0.6$  [17]. Finally, we crossed the results of this prediction with the conservation results obtained from the previous procedure (see Section 2.3). As a result, only those regions of at least 15 residues long, and that were predicted as epitopes by BepiPred2.0 and conserved accordingly to our analysis, were considered as putative epitopes for subsequent prioritization.

### 2.5. Prediction of T-cell Epitopes

For the prediction of CD4 T-cell epitopes we used the IEDB recommended method, the MHC II binding predictions tool (<http://tools.iedb.org/mhcii/>) [18], over the invariant proteome file. Given that no bovine lymphocyte antigen (BoLA) alleles are available for this tool we ran the program against the set of human leukocyte antigens (HLAs), as described elsewhere [19]. We considered for further analysis only the 0.01% top scoring CD4 T-cell epitopes that were predicted on the *T. cruzi*-masked proteome file.

For de novo prediction of CD8 T-cell epitopes we used the IEDB recommended method, the MHC I binding predictions tool (<http://tools.iedb.org/mhci/>) [18]. We ran the program using the available set of BoLAs (see Supplementary File 1) alleles over the invariant proteome file. We considered for further analysis only the 0.1% top scoring predicted epitopes.

### 2.6. Blast Searches for Epitopes Identity and Other Analysis Procedures

In order to find out the epitopes identity (i.e., % identity over the queried sequence length) to proteins in cows, we blasted them against cow microbiome protein sequences obtained from the Cow Microbiome Project [20] and against the NCBI non redundant (nr-) collection of *Bos taurus* and *Bos indicus* proteins. We used the NCBI protein blast suite (BLASTP) with the default parameters, except for the PAM30 Scoring Matrix and an expectation value (E-value) of 10,000, as the epitope sequences to blast were shorter than 100 residues in all cases. Following the same procedure, we also blasted the epitopes sequences against the NCBI nr- collection of sheep (*Ovis aries*; tax id: 9940), goat (*Capra hircus*; tax id: 9925), and pig (*Sus scrofa*; tax id: 9825) proteins.

The prediction of patterns related with surface exposure of those antigens with predicted B-cell epitopes was made with SignalP, TargetP, and TMHMM [21]. The possible presence of glycosylphosphatidylinositol (GPI) anchor signals was predicted using the PredGPI website [22].

We used the RaptorX web portal to model the 3D structure of the vaccine ensemble peptide sequence [23]. Solvent accessibility (Acc) classifiers were obtained at RaptorX, too. PyMOL Molecular Graphics System, Version 1.8 Schrödinger, was used to visualize the ensemble 3D structure.

### 3. Results

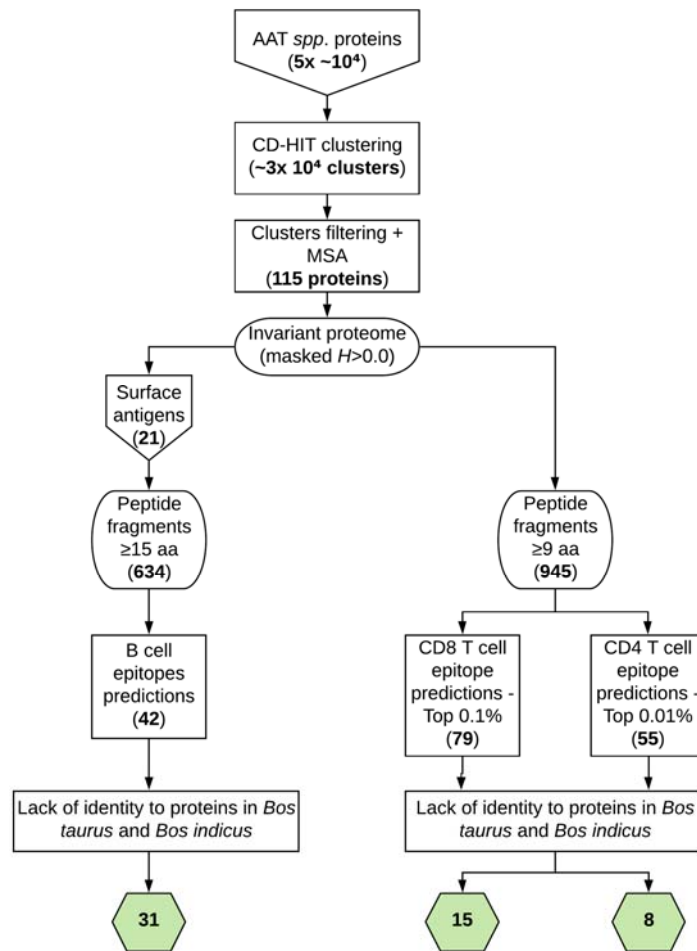
#### 3.1. Identification of the Invariant Proteome of AAT-Causing Trypanosomes

We used all the available annotated proteomes from *Trypanosome* spp. responsible for AAT to obtain the conserved “proteome”. We departed from the five available sequences at TriTryDB r39 (*T. brucei* Lister427, *T. brucei* TREU927, *T. congolense* IL3000, *T. evansi* STIB805, and *T. vivax* Y486; ~50,000 CDS) and imposed the strictest conservancy possible: any residue position that was not identical in all five of them would be masked and thus not accounted for in the forthcoming steps (Figure 1). The reason to be so rigid was due to the limited number of genomes available. Upon applying CD-HIT we obtained about 30,000 protein clusters (Figure 1), amongst which we progressed only those antigens that were represented in all five departing proteomes. Therefore, we ran a custom script to identify those clusters with at least one representative protein sequence from each of those five proteomes and ended up with just 115 clusters that were aligned with MUSCLE [14]. Each of the 115 MSA was guided by a consensus protein (Figure 1). At this point, instead of having as reference proteins the unreal consensus sequences yielded by MUSCLE, we rather chose as reference the “real” sequence amongst all present in the MSA that most closely resembled that of the consensus. As a consequence, upon calculating the variability in the MSA with a rigid Shannon entropy  $H > 0.0$  threshold, we obtained the invariant regions within those 115 potential antigens that were fully conserved in all five AAT-causing trypanosomes.

#### 3.2. Prediction of B-Cell Epitopes

It was not possible to find experimentally validated AAT-conserved epitopes at the IEDB. We thus had to perform de novo predictions with immunoinformatic tools to potentially identify peptides of interest. These predictions were made on the invariant protein sequences common to the five proteomes.

We used Bepipred2.0 to predict B-cell epitopes that were at least 15 residues long. Before applying it to the invariant proteome file we manually selected those antigens in it that were expected to be surface-exposed by the annotation terms (Figure 1). Bepipred2.0 was run onto the FASTAs of those 21 “surface” antigens. Whenever there was a coincidence between the Bepipred2.0 prediction and the fact that such region of the antigen was fully conserved in the invariant proteome, then we had a predicted epitope to be further progressed. This yielded a total of 42 predicted epitopes (Additional Table S1A). Subsequent selection steps entailed prospecting the epitopes’ potential cross-reactivity to other protein sequences present in cows, as well as in sheep, goat, and pig proteomes. Four peptides were found to have >70% identity to either *B. taurus* or *B. indicus* or the described cow microbiome, so we discarded them from further progression. None of the peptides showed >70% identity to sheep, goat, or pig proteins. Moreover, since immunogenic B-cell epitopes are generally located at the surface exposed antigens, we ran subcellular localization tools to find out which antigens of origin of those epitopes were expected to be expressed at the surface of the trypanosomes. Any peptides that were predicted to have a subcellular localization to the mitochondrion with a “reliability class” (RC) 1–3, as determined by [21], or mapped to the internal regions of the surface-located antigens, were dismissed (shown in light grey in Additional Table S1A). Upon applying the cross-reactivity and subcellular localization criteria, we selected 31 predicted B-cell epitopes that originate from 11 different antigens (Table 1).



**Figure 1.** Flowchart summary of the epitope identification strategy. Larger boxes show the steps and the hexagons the numbers of finally selected epitopes. This figure was made at <https://www.lucidchart.com>.

**Table 1.** List of the B-cell epitopes predicted in the invariant AAT-spp. proteome selected for the ensemble.

Epitope	Length	Antigen <sup>1</sup>	Annotation <sup>1</sup>	TargetP <sup>2</sup>	<i>Bos taurus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>	<i>Bos indicus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>
MQYGSTPKDIRYGIE	15	TcIL3000.11.15250.1	chaperonin HSP60, mitochondrial precursor	Other (5)	NP_001029859.1	28.6	XP_019842974.1	31.6
GFTSPYFVTNTKSQKC	16	TcIL3000.11.15250.1	chaperonin HSP60, mitochondrial precursor	Other (5)	AAI02078.1	56.3	XP_019829938.1	56.3
VFTGAQMISEDGLSLDQS	19	TcIL3000.11.15250.1	chaperonin HSP60, mitochondrial precursor	Other (5)	XP_024847167.1	23.3	XP_019820980.1	47.6
TISRDECILMEGGGSAIAVE ERVQMIRDMIAAEDHEYNRERL	42	TcIL3000.11.15250.1	chaperonin HSP60, mitochondrial precursor	Other (5)	XP_024847167.1	23.3	XP_019820980.1	47.6
RIEDRGLEDKEKREGLN	17	TcIL3000.11.15250.1	chaperonin HSP60, mitochondrial precursor	Other (5)	XP_024847167.1	23.3	XP_019811187.1	47.1
MRKRVNESQAPLPAL	15	TcIL3000.11.15250.1	chaperonin HSP60, mitochondrial precursor	Other (5)	NP_001029859.1	28.6	XP_019842974.1	31.6
KCIYYVTGDSKKKLETSPFIEQAKRRG	27	TcIL3000_0_26140.1	Heat shock protein 83, putative	Other (1)	NP_001073105.1	48.1	CCA61548.1	48.1
EYIPRAFVPKSTTGL	15	TcIL3000_10_3450.1	hypothetical protein, conserved	M (4)	XP_024851069.1	65.4	XP_019820927.1	65.4
ECQDDWCTLEKRYFW	15	TcIL3000_10_3450.1	hypothetical protein, conserved	M (4)	XP_024851069.1	65.4	XP_019820927.1	65.4
LETAQKWNEWRFWELDPQVKKVAREDQNIGR EGLGYNSPWEQVVREDFDKRKKALTQEEMAE LKRMDTERMARETAAYKERKDRIRDDLEKARG	95	TcIL3000_10_3450.1	hypothetical protein, conserved	M (4)	NP_001029859.1	28.6	XP_019842974.1	31.6
KFLNKRFAVDKDLQRMQPGKRYSGKTAD	28	TcIL3000_10_3450.1	hypothetical protein, conserved	M (4)	NP_001029859.1	28.6	XP_019834515.1	28.6
PFGRFAHTPTVLPDSSIDLSYEVPPWW	26	TcIL3000_10_510.1	hypothetical protein, conserved	Other (3)	NP_001029859.1	28.6	XP_019842974.1	31.6
DSLCTLNMVPGVSVYGEKRVEVGATQ	26	TcIL3000_10_6460.1	fibrillarlin, putative	Other (3)	XP_024851069.1	65.4	XP_019820927.1	65.4
TGQFEKAGEQKEREKGKH	17	TcIL3000_10_970.1	intraflagellar transport protein 172, putative	M (5)	XP_024847167.1	23.3	XP_019820980.1	47.6

Table 1. Cont.

Epitope	Length	Antigen <sup>1</sup>	Annotation <sup>1</sup>	TargetP <sup>2</sup>	<i>Bos taurus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>	<i>Bos indicus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>
DFDCTDFPKKYPMPKSSS	18	TcIL3000_10_970.1	intraflagellar transport protein 172, putative	M (5)	-	-	-	-
ADDGYVGYDSVPFHRYNR	18	TcIL3000_7_2400.1	hypothetical protein, conserved	M (4)	-	-	-	-
LGCATCKMPNDINEA	15	TcIL3000_7_510.1	hypothetical protein, conserved	Other (3)	NP_001029859.1	28.6	XP_019842974.1	31.6
IIKGCNVFELDGMSDVHQSI	21	TcIL3000_7_510.1	hypothetical protein, conserved	Other (3)	AAI02078.1	56.3	XP_019829938.1	56.3
TSAAGLSPDWLEAFFTNVAYNT	22	TcIL3000_7_510.1	hypothetical protein, conserved	Other (3)	XP_024847167.1	23.3	XP_019820980.1	47.6
KSLRMNFFTVCERCVLKEM	19	TcIL3000_7_510.1	hypothetical protein, conserved	Other (3)	NP_001029859.1	28.6	XP_019842974.1	31.6
EKGPSTLPPLEHLFVASVYLAAQRQFTNLFFF	32	TcIL3000_7_510.1	hypothetical protein, conserved	Other (3)	XP_024851069.1	65.4	XP_019820927.1	65.4
DLKEWCKQYEQLTNDMLRLR	20	TcIL3000_7_510.1	hypothetical protein, conserved	Other (3)	-	-	-	-
ELPGAITHKSILELR	15	TcIL3000_8_4670.1	hypothetical protein, conserved	Other (2)	XP_024851069.1	65.4	XP_019820927.1	65.4
QSQQQLASTRGGMPAR	17	TcIL3000_8_4670.1	hypothetical protein, conserved	Other (2)	XP_024847167.1	23.3	XP_019816904.1	23.3
SGVRHLRMAGDGTVGQN	17	TcIL3000_9_6470.1	hypothetical protein, conserved	M (5)	XP_024847167.1	23.3	XP_019820980.1	47.6
DWLERQFIDNCATPERDP	18	TcIL3000_9_6470.1	hypothetical protein, conserved	M (5)	XP_024851069.1	65.4	XP_019820927.1	65.4
LLDWEDFGIPREDLYR	16	TcIL3000_9_6470.1	hypothetical protein, conserved	M (5)	NP_001029859.1	28.6	XP_019842974.1	31.6
QQTRFDNTTEEKLRSL TYTQTDKTVDEYYG	30	TevSTIB805.5.5170-t26_1	hypothetical protein, conserved	S (3)	XP_024847167.1	23.3	XP_019816904.1	23.3
FPDPPFNFSAVVPER	15	TevSTIB805.5.5170-t26_1	hypothetical protein, conserved	S (3)	XP_024851069.1	65.4	XP_019820927.1	65.4
SGVKANDTGVGPNTTNTAGGA	21	TevSTIB805.5.5170-t26_1	hypothetical protein, conserved	S (3)	XP_024847167.1	23.3	XP_019820980.1	47.6
CVDIIDWRDLDEMLNRTDEVVEKSL	26	TevSTIB805.5.5170-t26_1	hypothetical protein, conserved	S (3)	-	-	-	-

<sup>1</sup> Antigen name and annotation are those reported in the parasite proteome of origin from TriTrypDB [12]. <sup>2</sup> Subcellular localization of antigens predicted with TargetP1.1 [21]: M, mitochondrial; S, secreted; number in brackets stands for the prediction “Reliability class” (RC), which ranges from 1 to 5 where 1 indicates the strongest prediction. <sup>3</sup> *Bos taurus* and *Bos indicus* non-redundant protein collection protein sequences are identified by their accession reference number at NCBI. <sup>4</sup> Percentage of identity (number of identical residues per queried length) of each epitope to its corresponding blasted hit (NCBI Accession name). The “-” means there is no hit. The Id % to the cow microbiome proteins of the B-cell epitopes selected can be checked in Additional Table S1A.

### 3.3. Prediction of T-Cell Epitopes

T-cell epitopes also had to be predicted. We used the IEDB Analysis Resource MHC-II and MHC-I binding tools to respectively predict CD4 and CD8 T-cell epitopes. The MHC-II binding tool at IEDB does not allow to specifically select class II bovine leukocyte antigens for the binding predictions so we used the available HLAs and the IEDB recommended set of tools [18]. Although the validated data are limited, MHC-II predictions using human alleles have been shown to be helpful towards the identification and validation of *Mycobacterium bovis* epitopes in cows [19]. We applied these MHC binding tools onto the invariant proteome of AAT-causing trypanosomes. No size limit was imposed to the predicted CD4 T-cell epitopes, whereas only nonameric epitopes were chosen for the prediction binding of CD8 T-cell epitopes. We respectively selected the 0.01% and 0.1% top-scoring epitopes amongst the CD4 T cell and CD8 T cell overall predictions. This way we ended up with 55 CD4 T-cell and 79 CD8 T-cell predicted epitopes, respectively (see Additional Tables S1B and S1C). Upon applying the cross-reactivity threshold to proteins present in cows, as well as against the sheep, goat, and pig proteomes, we disregarded from further analysis 47 CD4 T-cell and 64 CD8 T-cell potential epitopes that had a >70% sequence identity to those. Therefore, we selected for inclusion in the vaccine ensemble 8 CD4 T-cell-predicted epitopes (Table 2) and 15 CD8 T-cell-predicted epitopes (Table 3). The CD4 T-cell peptides were predicted to bind to five different HLAs (Table 2), whereas the CD8 T-cell peptides were predicted to bind to 23 distinct BoLa alleles (Table 3). The percentage of identity (Id %) of the selected CD4 and CD8 T-cell epitopes to cow microbiome proteins as well as to protein sequences in sheep, goat and pig species can be checked in Additional Tables S1B and S1C, respectively.

**Table 2.** List of the CD4 T-cell epitopes predicted in the invariant AAT-spp. proteome that were selected for the ensemble.

Epitope	Antigen <sup>1</sup>	Annotation <sup>1</sup>	HLAs (human) <sup>2</sup>	<i>Bos taurus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>	<i>Bos indicus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>
AQIHDIYRMIPPLQVVLVSAT	TcIL3000.11.9080.1	ATP-dependent RNA helicase FAL1, putative	HLA-DRB3*02:02	NP_001039653	66.7	XP_019837752	66.7
FKAQIHDIYRMIPPLQVVLVS	TcIL3000.11.9080.1	ATP-dependent RNA helicase FAL1, putative	HLA-DRB1*09:01	NP_001039653	66.7	XP_019837752	66.7
GLVIERRLSDKHVFV	TcIL3000_0_22360.1	NADH-dependent fumarate reductase	HLA-DRB3*01:01	3FE5_A	46.7	XP_019825021	46.7
RINLVVQFDMASDADSYLHRV	TcIL3000_10_400.1	ATP-dependent RNA helicase SUB2, putative	HLA-DRB3*01:01	NP_001029924	57.1	XP_019819542	57.1
DGRIGIILMDNITEVQSGQK	TcIL3000_7_6050.1	ATP synthase F1, alpha subunit, putative	HLA-DRB3*01:01	XP_005203166	40.0	XP_019836600	40.0
AKNKFFYMYVQELNYLIRF	TcIL3000_7_6050.1	ATP synthase F1, alpha subunit, putative	HLA-DPA1*02:01_DPBI*14:01	XP_005205520	36.8	XP_019814319	36.8
ASYANVWEMDDPYRFLQTEQD	TcIL3000_9_4570.1	Chromosome passenger complex (CPC) protein INCENP N terminal, putative	HLA-DRB3*01:01	DAA31134	38.1	XP_019813076	38.1
DYTNRIIRQMLHNVAALSCNK	TevSTIB805.10.4720-t26_1	Pumillo RNA binding protein PUF1	HLA-DRB4*01:01	XP_002697522	38.1	XP_019840994	38.1

<sup>1</sup> Antigen name and annotation are those reported in the parasite proteome of origin from TriTrypDB [12]. <sup>2</sup> HLA alleles to which the IEDB tool “MHC-II Binding Prediction” (<http://tools.iedb.org/mhcii/>) reported a high binding prediction. <sup>3</sup> *Bos taurus* and *Bos indicus* non-redundant protein collection protein sequences are identified by their accession reference number at NCBI. <sup>4</sup> Percentage of identity (number of identical residues per queried length) of each epitope to its corresponding blasted *Bos taurus* and *Bos indicus* hit.



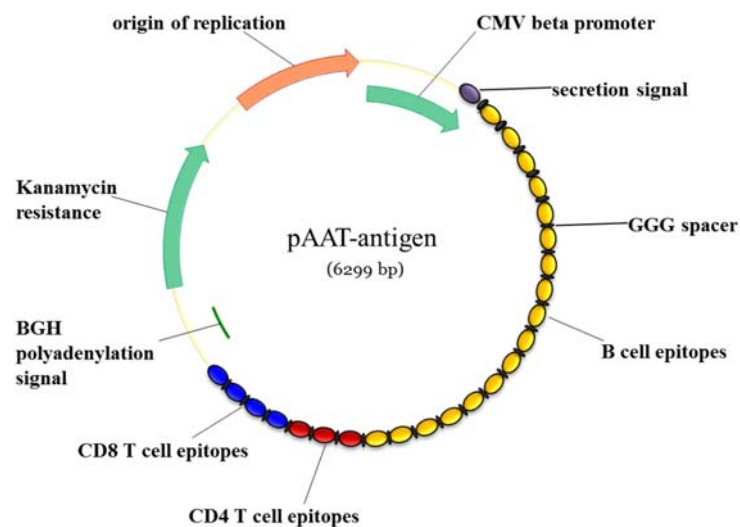
**Table 3.** List of the CD8 T-cell epitopes predicted in the invariant AAT-spp. proteome that were selected for the ensemble.

Epitope	Antigen <sup>1</sup>	Annotation <sup>1</sup>	BoLA <sup>2</sup>	<i>Bos taurus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>	<i>Bos indicus</i> Hit <sup>3</sup>	Id (%) <sup>4</sup>
MMHAYASRY	TcIL3000.11.1250.1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 7, mitochondrial	BoLA-N:05501, BoLA-N:02401	DAA15150.1	66.7	XP_019842744.1	66.7
YPAVRVYPV	TcIL3000.11.9220.1	NADH dehydrogenase subunit NB6M, putative	BoLA-N:00501	DAA32829.1	66.7	XP_019825300.1	66.7
FGPWVFEV	TcIL3000_0_05230.1	proteasome beta 3 subunit, putative	BoLA-N:05301, BoLA-N:00402, BoLA-N:00401, BoLA-N:00102, BoLA-N:00103, BoLA-N:00101	NP_789861.1	55.6	XP_019823249.1	55.6
VTPSIHYTM	TcIL3000_0_22360.1	NADH-dependent fumarate reductase	BoLA-N:05301, BoLA-N:03801, BoLA-N:00402, BoLA-N:00401	XP_024843166.1	66.7	XP_019834535.1	66.7
ATFEVFHTI	TcIL3000_10_13200.1	methyltransferase domain containing protein, putative	BoLA-N:05401	XP_005225549.2	66.7	XP_019808611.1	66.7
LQYHKYGCL	TcIL3000_10_3580.1	ubiquinol-cytochrome c reductase complex 14kD subunit, putative	BoLA-HD6	NP_001179015.1	66.7	XP_019841558.1	66.7
YGPSWHCVM	TcIL3000_10_3900.1	dynein light chain, putative	BoLA-N:05301, BoLA-N:04001, BoLA-N:00402, BoLA-N:00401, BoLA-N:00102, BoLA-N:00103, BoLA-N:00101	XP_015328853.2	55.6	XP_019837337.1	55.6
DMYLHQHEF	TcIL3000_10_970.1	intraflagellar transport protein 172, putative	BoLA-NC1:00101, BoLA-NC1:00201, BoLA-NC1:00301	XP_002693758.2	55.6	XP_019830009.1	55.6
EEMKYVAGL	TcIL3000_10_970.1	intraflagellar transport protein 172, putative	BoLA-N:04101	NP_001192842.1	66.7	XP_019822633.1	66.7
RKYETTWEM	TcIL3000_6_4030.1	hypothetical protein, conserved	BoLA-D18.4	AAI42017.1	55.6	XP_019813499.1	55.6
EHNWMAEI	TcIL3000_9_4570.1	Chromosome passenger complex (CPC) protein INCENP N terminal, putative	BoLA-N:05001	XP_024855981.1	55.6	XP_019826963.1	55.6
SEMRAWYWK	TcIL3000_9_4570.1	Chromosome passenger complex (CPC) protein INCENP N terminal, putative	BoLA-N:04201	NP_001095539.1	55.6	XP_019817650.1	66.7
AMWSQDSPY	TcIL3000_9_6470.1	hypothetical protein, conserved	BoLA-N:00901	NP_001193092.1	66.7	XP_019825232.1	66.7
ARYEYFLAY	TevSTIB805.10.7310-t26_1	Sterol methyltransferase, putative	BoLA-N:02201	XP_005212827.1	55.6	XP_019825865.1	55.6
THETHSFLY	TvY486_1100540:mRNA	dynein light chain, putative	BoLA-N:04801	XP_024852897.1	66.7	XP_019822516.1	66.7

<sup>1</sup> Antigen name and annotation are those reported in the parasite proteome of origin from TriTrypDB [12]. <sup>2</sup> BoLa alleles to which the IEDB “MHC-I Binding Prediction” (<http://tools.iedb.org/mhci/>) reported a high binding prediction. <sup>3</sup> *Bos taurus* and *Bos indicus* non-redundant protein collection protein sequences are identified by their accession reference number at NCBI. <sup>4</sup> Percentage of identity (number of identical residues per queried length) of each epitope to its corresponding blasted *Bos taurus* and *Bos indicus* hit.

### 3.4. Design of a Genetic Vaccine Ensemble

The most economic scaffold to administer the vaccine ensemble would likely be genetic delivery of the epitopes as plasmid DNA (pDNA). We therefore designed a genetic construct that encompasses the 54 selected epitopes shaped as a “string-of-beads” [24]. Linkers of three glycine (GGG) residues were included to separate the epitopes and limit the formation of neo-epitopes [25,26]. It must be noted that besides experimentally validating the predicted epitopes, their optimum ordering within the ensemble and the spacers between them will have to be empirically determined. Nevertheless, in an attempt to anticipate a pDNA design we placed B-cell epitopes first right after a secretory signal to ensure their adequate processing to the surface, because B-cell epitopes are expected to be of utmost importance for eliciting a protective immunological response against AAT free-living trypanosomes [8] (Figure 2). Following the B-cell epitopes are the eight CD4 T-cell epitopes (Figure 2). These need to be physically connected to the B-cell epitope sequences in order to promote the correct processing and reactivity against the latter. Finally, the fifteen prioritized CD8 T-cell epitopes are placed towards the 3'-end of the “string-of-beads”, also separated by GGG spacers, for their optimal proteasome processing (Figure 2).



**Figure 2.** Map of the proposed pDNA vaccine construct.

Upstream of the epitope “string-of-beads” there is a cytomegalovirus (CMV) promoter sequence that governs its transcription, whereas at the string’s 3'-end there is a bovine growth hormone (BGH) poly-adenylation (polyA) signal to tag the growing mRNA molecule with a polyA tail. These two adjoining sequences are widely used within mammalian expression plasmids and they have been shown to work in the context of bovine cells, for instance [27,28].

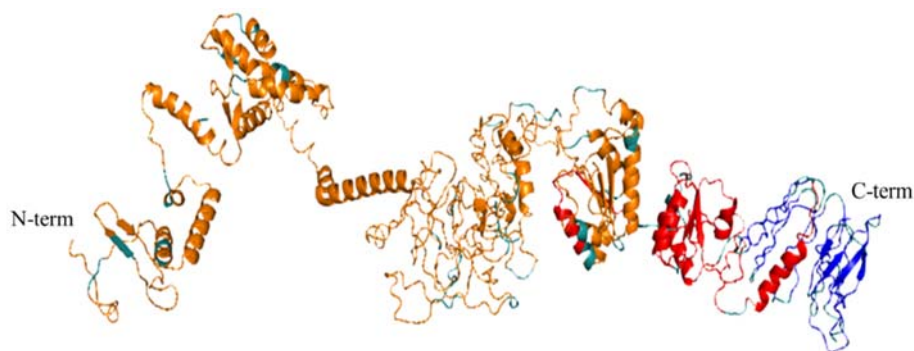
Thinking of a delivery of the vaccine ensemble to cows, the full peptide sequence of the ensemble (Supplementary File 2) was subjected to reverse translation using the *Bos taurus* code and its most likely abundant codons. As can be seen in Supplementary File 3, the major codon dictionary used in *Bos taurus* differs in some residues with respect to that of AAT spp. However, plain reverse translation does not account for the optimal codon usage. This is a complex process where not only the codon usage is important, as it involves many other parameters, such as those facilitating transcription (GC content, CpG dinucleotides content, cryptic splicing sites, and occurrence of TATA boxes) or those increasing translation efficiency (codon usage bias, mRNA secondary structure, premature polyA sites, and RNA instability motifs). Thereby, we used an in vitro-translated mRNA (IVT) codon optimization tool [29] to optimize the reverse translated “string-of-beads” to be included in the pDNA. Alignment of the nucleotide sequence upon reverse translating the epitopes and linkers (with the *Bos taurus* codon usage; [30]) with the optimized nucleotide sequence for its expression in cows showed that

both sequences share a high percentage of identity (77.9%), but many triplets were indeed optimized accordingly to the IVT sequence optimization tool [29] (Supplementary File 4).

Taking into account that the “string-of-beads” itself is 3498 nucleotides long, and that other regulatory sequences must go within the pDNA, such as the origin of replication (Ori) and the antibiotic resistance gene for propagation and selection (Kanamycin resistance in this case), we estimated a pDNA length of ~6.3 Kb. Such length can be now fully synthesized in vitro with a cow-optimized codon usage pattern.

### 3.5. 3D Modelling of the Peptide Vaccine Ensemble

The pDNA construct will be eventually expressed within cow cells to yield the poly-peptide epitopes. We used RaptorX to obtain a 3D model of what the vaccine ensemble would look like upon being expressed. Notably, RaptorX also displays solvent accessibility (Acc) per position of all the residues within the submitted ensemble assigning a classification of buried (“B”), medium (“M”), or exposed (“E”) to each of these residues [23]. The retrieved 3D model of the AAT vaccine ensemble showed that regions with B-cell and CD4 T-cell epitopes mostly contained  $\alpha$ -helices as secondary structure, whereas the majority of the modelled  $\beta$ -sheets were in the CD8 T-cell epitopes C-term side of the ensemble (Figure 3). Moreover, large fragments of the B-cell epitope-containing N-term of the poly-peptide were modelled as disordered structures, which would theoretically make them more accessible for recognition by B-cell receptors and antibodies (Figure 3). Nonetheless, it must be noticed that the 3D model provided is just a model, and that the “combined protein” may just be one of the multiple possible combinations and not necessarily the best one.



**Figure 3.** 3D model of the vaccine ensemble generated by RaptorX, respectively encompassing from the N-term to the C-term: B-cell epitopes (orange), CD4 T-cell epitopes (red), and CD8 T-cell epitopes (blue). GGG linkers are shown in gray.

If we take into account the Acc classification by RaptorX, it appeared that the highest percentage of “B” residues matched to CD4 and CD8 T-cell epitopes (35.2% and 33.3, respectively; Supplementary File 5), whilst the percentage of B-cell epitopes “E” residues was 44.5% (Supplementary File 5). The GGG spacers were mostly modelled as disordered structures (Figure 3), and thus 71.0% of them were tagged as “E” (Supplementary File 5).

## 4. Discussion

In light of the lack of continuity in vector control programs and the increasing reports of drug resistance occurrence, vaccination appears to be a sought-after solution for the control of AAT impact. But the achievement of a functional vaccine product is in doubt due to the parasites ability to dampen any immune reactivity against them. Hence, alternative animal production methods, such as farming trypanotolerant breeds (e.g., N'Dama cows), have been proposed [4]. It must be noted though that these breeds are still sensitive to the disease in areas of high transmission rate, and there is the risk that trypanosome infections still pose an immunosuppressive toll, making cattle more vulnerable to other

pathogens [4]. This is indeed an important obstacle to exploring trypanotolerance in most livestock production systems in sub-Saharan Africa, and enforces seeking a vaccine-based intervention. Hitherto, a certain level of protection against a challenge has been described in a variety of immunizing protocols in mouse models of trypanosome infection, including the inoculation of *T. congolense* recombinant transialidase [31] or immunization with other invariant antigens [9,10]. Furthermore, in a larger and more biologically relevant cow preclinical model of infection, inoculation of animals with irradiated *T. congolense* or *T. brucei* or with purified VSGs protected against the homologous challenge. Besides, trypanotolerant breeds have been described to have higher levels of anti-trypanosome immunoglobulins than trypanosusceptible cattle (reviewed in [11]). It should thus be possible to develop a vaccine for AAT. Nonetheless, since studies relying on classical approaches have mostly failed, innovative strategies might be required to overcome the challenge, e.g., a multi-antigenic epitope-based vaccine ensemble as the one proposed here. The increasing availability of genomic and other omics data on pathogens, allows researchers to use immune-informatics approaches towards a reverse vaccinology driven development [32]. Some of these tools are already being applied for the development of vaccines against cattle-relevant pathogens [33–35] or the search of new diagnostics for *T. vivax* [36].

We used publicly available and tailor-made computational methods to identify the epitopes presented in this article. Unfortunately, we did not find any validated AAT epitopes at the IEDB repository that complied with the conservation criteria, so we had to run de novo predictions. Therefore, the next step to take will be to experimentally validate that the predicted epitopes are presented and recognized in the cow immunological context, and that they are capable of eliciting immunogenic responses. Despite being based on predictions, the goodness of currently available tools for this purpose has been acknowledged by several authors and the use of prediction tools is seen as a great aid for the identification of new epitopes [19,36–38]. Since the protein space of trypanosomes, particularly *T. congolense*, *T. evansi*, and *T. vivax* remains largely unexplored, the majority of the antigens of origin of the predicted epitopes were either uncharacterized proteins or had low annotation scores inferred by homology. However, the antigens from where the B-cell epitope DSLCTLNMVPGVSVYGEKRVEVGATQ; CD4 T-cell epitopes AQIHDIYRMIPPLQVVLVSAT, DGRIGIILMDNITEVQSGQK, and AKNKFFYMYVQELNYLIRF; and CD8 T-cell epitopes MMHAYASRY, FGPWFVEPV, and EHNWMFAEI emerged, had *T. brucei brucei* homologues counting with medium to high annotation scores and experimental evidence of expression at either the transcriptional and/or protein level. The aforementioned B-cell and first CD4 T-cell epitope sequences, respectively matched to fibrillarlin and a putative ATP-dependent RNA helicase. The other two CD4 T-cell epitopes matched to the *T. congolense* homologue of a well-characterized *T. b. brucei* mitochondrial ATP synthase subunit alpha (UniProt ref.: Q9GS23). This protein had experimental evidence provided by three different laboratories and it has been shown to be very important for the correct growth of the parasite [39]. Each of the three CD8 T-cell epitopes referred above was from a different intracellularly located antigen. The first two were respectively inferred by homology to *T. b. brucei* NADH dehydrogenase and proteasome subunit beta type-3; whereas the last one was an uncharacterized *T. b. brucei* protein with a >90% identity to the *T. equiperdum* chromosome passenger complex (CPC) protein. The uncharacterized *T. b. brucei* homologue and one of the ATP synthase subunit alpha had 3D crystal structures, and thus experimental evidence at the protein level.

Taking into account that (i) mixed infections with various trypanosome species have been reported [40]; (ii) the very limited resources available to develop a vaccine specific for each AAT-causing pathogen; and (iii) the fact that evolutionary conserved CDS regions of an organism are likely to have fundamental biological functions, sequence conservation was a key criteria to consider in the process of prediction of the epitopes that encompass the proposed vaccine ensemble. In fact, we asked for strict conservation between the protein coding regions of all five genomes from the AAT parasites, upon the hypothesis that immune responses elicited by invariant conserved antigens could protect against heterologous infections, too.

Trypanosome parasites responsible of AAT persist in the bloodstream thanks to a series of evolved mechanisms that allow them to escape innate immune controls and survive the successive waves of adaptive immune responses. At the initial phases of the infection, it has been described in mouse experimental models that *T. brucei* dampens macrophage pro-inflammatory responses, and both *T. brucei* and *T. congolense* can modulate in their favor the T-cell response functionality [8]. However, a series of studies in mice described that VSG-specific CD4<sup>+</sup> T-cell responses can provide resistance to *T. brucei* infection [41]. Moreover, IFN $\gamma$  secreted by Th1 cells was shown to activate macrophages, leading to the release of trypanosome killing factors and thus mediating resistance to *T. congolense* infection [42]. These Th responses were mainly mapped to epitopes in the variable regions of VSG antigens [41]. Nonetheless, the immunogenicity to CD4<sup>+</sup> T cells specific of a VSG conserved peptide has also been described [43]. Thereby, the feasibility of the stimulation of peptide-specific immune responses in AAT infections with a series of MHC-II binding CD4 T-cell epitopes from conserved protein regions could support prophylactic cross-protection as well as the induction of functional memory cells.

In comparison to the suggested protective role of CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells were described to have a parasite-promoting effect in a mouse model of experimental *T. brucei* infection [44]. But CD8<sup>+</sup> T cell depletion in trypanosusceptible Boran cattle naturally infected with *T. congolense* through tsetse bites did not lead to any effect in the parasitemia or anemia, which is the main determinant of survival [3]. Similarly, no immune-suppression caused by CD8<sup>+</sup> T cells was observed in *T. evansi* experimentally infected sheep [45]. Thereby, the role played by CD8<sup>+</sup> T cells in AAT infections is still obscure, and the potential usefulness of their inclusion in a vaccine product should not be disregarded. In fact, CD8<sup>+</sup> T cells might promote certain level of protection yet to be discovered, since they are exacerbated during the natural infection process [46].

Nevertheless, being overall extracellular pathogens, the humoral response against AAT trypanosomes is likely to be the most important towards ultimately controlling the infections [47]. In order to counteract B-cell mediated responses these parasites pose complex and thick surface coats packed with homodimers of the VSG family, elevated on GPI anchors to conceal other “more sensible” antigens such as those within the ISG family [48]. Moreover, AAT parasites have evolved to antigenically switch their VSGs upon immune pressure in order to escape antibody-mediated destruction [48]. From the perspective of the modulation of host factors, mouse models of infection with *T. brucei* and *T. congolense* have shown that these trigger a polyclonal B-cell activation, which may contribute to take the immune attention away from relevant antigens, like VSGs (see [8] for a review). Besides, the two former trypanosome species and *T. vivax* have been described to disrupt B-cell lymphopoiesis in primary and secondary lymphoid organs, which affects the establishment of pools of memory B cells [8]. In sight of the major relevance of the antibody response to control parasitemia, a vaccine strategy that promotes a protective response against conserved regions of surface protein sequences may contribute to control AAT, especially if it is accompanied of CD4<sup>+</sup> T cells inducing epitopes and the adjuvant used to exacerbate the exerted response biases it towards a Th1 profile. Whether such an adjuvant is sourced outside the vaccine construct or can be encoded within it or co-immunized in another construct [49], remains to be evaluated. Adjuvants will be required for effective vaccination and the latter choice seems more advisable as it will contribute to save costs.

Here we provide a tentative design for a pDNA ensemble, thinking of its likely reduced development and production costs in comparison to other delivery platforms. Following previously described constructs [27,28], we have included a CMV promoter signal and a BGH-pA, respectively upstream and downstream of the epitope “string-of-beads”. We did not include any extra immune enhancing signals in the Figure 2 map, but this type of booster could also be included within the pDNA as far as there is room for them. Nonetheless, it is advisable to keep the length of the pDNA under 7.5 kb long, as it has been shown that the bigger the plasmid the less efficient will be the cell transfection, and there is an inverse correlation between the construct size and the final expression of the gene as the bigger the plasmid the less it can enter the nucleus [50].

At present there are five DNA vaccines already approved for veterinary use [51]. Remarkable advantages of a pDNA approach for the vaccination of large animals in tropical regions are that it can be easily produced at high quantities and does not require cold-chain. On the other side, it must be considered that effective immunization of large animals often fails [51], and adjuvants and/or prime-boost strategies should be pursued if a pure DNA vaccine is not potent enough. Immunization of large animals may face difficulties to reach sufficient concentration of antigens upon pDNA immunization, and thus this approach might work better if the animals are not too big, i.e. immunizing calves early in life with vaccine doses appropriately spaced [52].

Potency and longevity of vaccine-exerted immune response will be fundamental and will have to be determined experimentally. In this respect, the questioned poor immunogenicity concerns on pDNA vaccination might be circumvented by including codon optimization and administering the construct in combination with the appropriate adjuvant. Besides, the pDNA delivery methodology and the prime-boost regime followed can be decisive [49]. Bearing in mind the results by Tabel et al. the route and dose of the antigen to be delivered will be key [53]. In this respect, electroporation could be used to enhance the exerted immune reactivity against pDNA immunization [28] and their field use considered [54].

## 5. Conclusions and Limitations

Despite there are two drugs available for the treatment of AAT the appearance of resistances is common due to poor attachment to treatment. Besides, the implementation and maintenance of vector control programs, crucial to control the disease, is irregularly performed as they are linked to economic and political stability, which may not be assured in the affected regions. For these reasons, the availability of a vaccine that could prevent the infection would mean a major breakthrough to limit the impact of AAT. Some efforts have been initiated but the biological complexity of the parasites and the very limited funding available for research on the subject has precluded its development.

We propose an alternative vaccine approach based on epitopes from a range of different parasite antigens to cope with the parasites antigenic complexity. The epitopes were prioritized in terms of their absolute conservation among all the sequenced genomes of the distinct AAT-causing spp. The main rationale, besides taking advantage of the biological relevance of highly conserved protein motifs for the parasites' biology, is to ultimately provide a single vaccine ensemble to tackle all ATT-causing trypanosomes. Progressing a single vaccine product would contribute to save developmental costs. Moreover, we propose a pDNA vaccine construct to carry the epitope-based "string-of-beads" because it could be synthesized at a fraction of the cost of producing recombinant protein subunits or virus vectored vaccines. Remarkably, it is also a platform capable of yielding the high amounts of product needed to immunize large animals, which can be lyophilized and suitably used in tropical regions. All epitopes in the ensemble were prioritized in agreement with their lack of identity to any protein sequences present in cows, and in the proteomes of sheep, goat, and pig species, so as to avoid the appearance of cross-reactivity reactions, thus favoring the safety of the vaccine but also likely promoting its immunogenicity.

A major limitation of the study is that we did not find experimentally validated conserved AAT epitopes in the IEDB repository database, so we had to rely on de novo predictions of epitopes derived from the computed invariant proteome file. We applied a very conservative threshold and only top qualifying predicted epitopes were selected for progression. Nonetheless, whether they are presented and recognized in the course of an infection will need to be experimentally demonstrated. Appropriate processing is a key feature towards the immunogenicity of the epitopes, and this will have to be thoroughly studied in relation to the delivery of the epitopes in the form of a genetic construct. Furthermore, their optimum ordering and spacing within the ensemble will also require to be adjusted experimentally as well as the selection of the most adequate adjuvant. Thereby, the pDNA ensemble provided here must just be considered as a theoretical scheme, subject to any required reformulations upon validation of the predicted epitopes and the study of their immunogenicity and

capacity to induce protective immune responses. Other acknowledged limitations of the reverse vaccinology strategy we have followed are the inability to provide glycan-based epitopes and the absence of conformational B-cell epitopes, since these cannot be isolated from their context and thus we only focused on linear ones.

In an attempt to anticipate a vaccine candidate, we placed the B-cell epitope component first as it is expected that an antibody-mediated response will play a major role in controlling free-living AAT infections. In the construct these are immediately followed by CD 4 T-cell epitopes, which will be fundamental for the correct response to all the former, but also because there is evidence of the importance of CD4<sup>+</sup> T cells against some AAT-causing trypanosomes. Finally, since the functionality of CD8<sup>+</sup> T cells is still under debate, we placed them at the end of the string but decided to keep them in the proposed ensemble, as CD8<sup>+</sup> T cells may play a role in eliminating infected macrophages and, therefore, freeing the suppressive control they have on AAT immunity.

This work lies at a very preliminary stage of the long vaccine development path. Notwithstanding, one of the advantages of an *in silico* vaccinology approach is that it can contribute to saving time and resources. Hopefully, the epitopes here selected can reach experimental studies in the near future and prove they can indeed elicit immunogenic responses against infections with AAT trypanosomes.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-393X/8/1/130/s1>, Tables S1A: Conserved B-cell epitopes predicted in the invariant AAT-spp. proteome. Tables S1B: Conserved CD 4 T-cell epitopes predicted (0.01% top scoring) in the invariant AAT-spp. proteome. Tables S1C: Conserved CD 8 T-cell epitopes predicted (0.1% top scoring) in the invariant AAT-spp. proteome. File S1: Full list of BoLa alleles available at IEDB MHC-I binding tool. All were used for the prediction of CD8 T-cell epitopes in the invariant AAT-causing spp. proteome. File S2: Full peptide sequence of the vaccine ensemble encompassing the selected B cell, CD4 T-cell and CD8 T-cell epitopes with “GGG” spacers placed in between them. File S3: Distinct codon usage triplets for the translation of nucleic acids into amino acids in the vaccine ensemble target species (*Bos taurus*) and in the different AAT-causing species from which there was information available [30]. Those triplets which translation differs between *Bos taurus* and the depicted AAT-spp. are circled. File S4: Alignment of the reverse translated vaccine ensemble sequence (named “*Bos taurus*”) with the codon optimized “string-of-beads” (named “*Optimized*”). Asterisks (\*) indicate equal residues. File S5. “String-of-beads” full peptide sequence aligned with RaptorX provided solvent accessibility per position. B, buried; M, medium; E, exposed. B-cell epitopes are depicted in yellow; CD4 T-cell epitopes in red; and CD8 T-cell epitopes in blue.

**Author Contributions:** Conceptualized the study, its methodology, and performed its formal analysis and data curation, L.M.-T. and J.A.-P.; provided the developed software, L.M.-T. and P.A.R.; contributed to the methodology followed, P.A.R.; contributed to the methodology, investigation and visualization of the results, P.B.; provided the required resources, M.-J.P. and J.G.; wrote the original draft, L.M.-T. and J.A.-P. All authors reviewed and approved the manuscript before submission.

**Funding:** JAP was funded by a Juan de la Cierva—Incorporación contract from the Spanish Science Ministry. This work was also supported by the Departament d’Universitats i Recerca de la Generalitat de Catalunya, Spain (AGAUR; 2017SGR00924) and by the Instituto de Salud Carlos III RICET Network for Cooperative Research in Tropical Diseases (ISCIII; RD12/0018/0010) and FEDER funds. PAR acknowledges BIO2014:54164R support. MJP research is supported by the Ministry of Health, Government of Catalunya (PERIS 2016-2010 SLT008/18/00132). We acknowledge support from the Spanish Ministry of Science and Innovation through the “Centro de Excelencia Severo Ochoa 2019-2023” Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program.

**Acknowledgments:** The authors want to thank the help support provided by IEDB-AR Solutions Centre to all the queries raised throughout the study.

**Conflicts of Interest:** The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. The authors declare no conflict of interests.

**Data Availability Statement:** The data used to support the findings of this study are included within the article or within the supplementary information file(s). The scripts developed are available from the corresponding author upon request.

## References

1. Morrison, L.J.; Vezza, L.; Rowan, T.; Hope, J.C. Animal African trypanosomiasis: Time to increase focus on clinically relevant parasite and host species. *Trends Parasitol.* **2016**, *32*, 599–607. [CrossRef] [PubMed]

2. Black, S.J.; Mansfield, J.M. Prospects for vaccination against pathogenic African trypanosomes. *Parasite Immunol.* **2016**, *38*, 735–743. [[CrossRef](#)] [[PubMed](#)]
3. Sileghem, M.; Naessens, J. Are CD8 T cells involved in control of African trypanosomiasis in a natural host environment? *Eur. J. Immunol.* **1995**, *25*, 1965–1971. [[CrossRef](#)] [[PubMed](#)]
4. Yaro, M.; Munyard, K.A.; Stear, M.J.; Groth, D.M. Combatting African animal trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance. *Vet. Parasitol.* **2016**, *225*, 43–52. [[CrossRef](#)] [[PubMed](#)]
5. Mekonnen, G.; Mohammed, E.F.; Kidane, W.; Nesibu, A.; Yohannes, H.; Van Reet, N.; Büscher, P.; Birhanu, H. Isometamidium chloride and homidium chloride fail to cure mice infected with Ethiopian *Trypanosoma evansi* type A and B. *PLoS Negl. Trop. Dis.* **2018**, *12*, e0006790. [[CrossRef](#)] [[PubMed](#)]
6. Magez, S.; Caljon, G.; Tran, T.; Stijlemans, B.; Radwanska, M. Current status of vaccination against African trypanosomiasis. *Parasitology* **2010**, *137*, 2017–2027. [[CrossRef](#)]
7. La Greca, F.; Magez, S. Vaccination against trypanosomiasis: Can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? *Hum. Vaccines* **2011**, *7*, 1225–1233. [[CrossRef](#)]
8. Stijlemans, B.; Radwanska, M.; De Trez, C.; Magez, S. African trypanosomes undermine humoral responses and vaccine development: Link with inflammatory responses? *Front. Immunol.* **2017**, *8*, 582. [[CrossRef](#)]
9. Tewari, A.K.; Kurup, S.P.; Baidya, S.; Barta, J.R.; Sharma, B. Protective antibody and cytokine responses in mice following immunization with recombinant beta-tubulin and subsequent *Trypanosoma evansi* challenge. *Parasites Vectors* **2015**, *8*, 580. [[CrossRef](#)]
10. Lança, A.S.; De Sousa, K.P.; Atouguia, J.; Prazeres, D.M.F.; Monteiro, G.A.; Silva, M.S. *Trypanosoma brucei*: Immunisation with plasmid DNA encoding invariant surface glycoprotein gene is able to induce partial protection in experimental African trypanosomiasis. *Exp. Parasitol.* **2011**, *127*, 18–24. [[CrossRef](#)]
11. Taylor, K.A. Immune responses of cattle to African trypanosomes: Protective or pathogenic? *Int. J. Parasitol.* **1998**, *28*, 219–240. [[CrossRef](#)]
12. Aslett, M.; Aurecochea, C.; Berriman, M.; Brestelli, J.; Brunk, B.P.; Carrington, M.; Depledge, D.P.; Fischer, S.; Gajria, B.; Gao, X.; et al. TriTrypDB: A functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **2010**, *38*, D457–D462. [[CrossRef](#)] [[PubMed](#)]
13. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
14. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
15. Shannon, C. The mathematical theory of communication. *Bell Syst. Technol. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
16. Stewart, J.J.; Lee, C.Y.; Ibrahim, S.; Watts, P.; Shlomchik, M.; Weigert, M.; Litwin, S. A Shannon entropy analysis of immunoglobulin and T cell receptor. *Mol. Immunol.* **1997**, *34*, 1067–1082. [[CrossRef](#)]
17. Jespersen, M.C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **2017**, *45*, W24–W29. [[CrossRef](#)]
18. Dhanda, S.K.; Mahajan, S.; Paul, S.; Yan, Z.; Kim, H.; Jespersen, M.C.; Jurtz, V.; Adreatta, M.; Greenbaum, J.A.; Marcatili, P.; et al. IEDB-AR: Immune epitope database-analysis resource in 2019. *Nucleic Acids Res.* **2019**, *47*, W502–W506. [[CrossRef](#)]
19. Farrell, D.; Jones, G.; Pirson, C.; Malone, K.; Rue-Albrecht, K.; Chubb, A.J.; Vordermeier, M.; Gordon, S.V. Integrated computational prediction and experimental validation identifies promiscuous T cell epitopes in the proteome of *Mycobacterium bovis*. *Microb. Genom.* **2016**, *2*, e000071. [[CrossRef](#)]
20. Stewart, R.D.; Auffret, M.D.; Warr, A.; Wiser, A.H.; Press, M.O.; Langford, K.W.; Lichko, I.; Snelling, T.J.; Dewhurst, R.J.; Walker, A.W.; et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **2018**, *9*, 870. [[CrossRef](#)]
21. Emanuelsson, O.; Brunak, S.; Von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953–971. [[CrossRef](#)]
22. Pierleoni, A.; Martelli, P.; Casadio, R. PredGPI: A GPI-anchor predictor. *BMC Bioinform.* **2008**, *9*, 392. [[CrossRef](#)]
23. Peng, J.; Xu, J. RaptorX: Exploiting structure information for protein alignment by statistical inference. *Proteins* **2011**, *79* (Suppl. S10), 161–171. [[CrossRef](#)]
24. Suhrbier, A. Multi-epitope DNA vaccines. *Immunol. Cell Biol.* **1997**, *75*, 402–408. [[CrossRef](#)]



25. Wang, Y.; Wang, M.; Wang, G.; Pang, A.; Fu, B.; Yin, H.; Zhang, D. Increased survival time in mice vaccinated with a branched lysine multiple antigenic peptide containing B- and T-cell epitopes from *T. gondii* antigens. *Vaccine* **2011**, *29*, 8619–8623. [[CrossRef](#)]
26. Choudhury, S.; Ganguly, A.; Chakrabarti, K.; Sharma, R.K.; Gupta, S.K. DNA vaccine encoding chimeric protein encompassing epitopes of human ZP3 and ZP4: Immunogenicity and characterization of antibodies. *J. Reprod. Immunol.* **2009**, *79*, 137–147. [[CrossRef](#)]
27. Da’Dara, A.A.; Li, C.; Yu, X.; Zheng, M.; Zhou, J.; Shollenberger, L.M.; Li, Y.S.; Harn, D.A. Prime-boost vaccine regimen for SjtPI and Sjc23schistosome vaccines, increase efficacy in water buffalo in a field trial in China. *Front. Immunol.* **2019**, *10*, 284. [[CrossRef](#)]
28. Tollefsen, S.; Vordermeier, M.; Olsen, I.; Storset, A.K.; Reitan, L.J.; Clifford, D.; Lowrie, D.B.; Wiker, H.G.; Huygen, K.; Hewison, G.; et al. DNA injection in combination with electroporation: A novel method for vaccination of farmed ruminants. *Scand. J. Immunol.* **2003**, *57*, 229–238. [[CrossRef](#)]
29. Integrated DNA Technologies IDT—Codon Optimization Tool. Available online: <https://eu.idtdna.com/CodonOpt> (accessed on 20 November 2019).
30. Stothard, P. Sequence Manipulation Suite (SMS): Reverse Translate. Available online: [http://www.bioinformatics.org/sms2/rev\\_trans.html](http://www.bioinformatics.org/sms2/rev_trans.html) (accessed on 20 November 2019).
31. Coustou, V.; Plazolles, N.; Guegan, F.; Baltz, T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. *Cell. Microbiol.* **2012**, *14*, 431–445. [[CrossRef](#)]
32. Bragazzi, N.L.; Gianfredi, V.; Villarini, M.; Rosselli, R.; Nasr, A.; Hussein, A.; Martini, M.; Behzadifar, M. Vaccines meet big data: State-of-the-art and future prospects. From the classical 3Is (“isolate-inactivate-inject”) Vaccinology 1.0 to Vaccinology 3.0, vaccinomics, and beyond: A historical overview. *Front. Public Health* **2018**, *6*, 62. [[CrossRef](#)]
33. Saadi, M.; Karkhah, A.; Nouri, H.R. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infect. Genet. Evol.* **2017**, *51*, 227–234. [[CrossRef](#)]
34. Maritz-Olivier, C.; Van Zyl, W.; Stutzer, C. A systematic, functional genomics, and reverse vaccinology approach to the identification of vaccine candidates in the cattle tick, *Rhipicephalus microplus*. *Ticks Tick Borne Dis.* **2012**, *3*, 179–187. [[CrossRef](#)]
35. Kar, P.P.; Srivastava, A. Immuno-informatics analysis to identify novel vaccine candidates and design of a multi-epitope based vaccine candidate against *Theileria* parasites. *Front. Immunol.* **2018**, *9*, 2213. [[CrossRef](#)]
36. Guedes, R.L.M.; Rodrigues, C.M.F.; Coatnoan, N.; Cosson, A.; Cadioli, F.A.; Garcia, H.A.; Gerber, A.L.; Machado, R.Z.; Minoprio, P.M.C.; Teixeira, M.M.G.; et al. A comparative in silico linear B-cell epitope prediction and characterization for South American and African *Trypanosoma vivax* strains. *Genomics* **2019**, *111*, 407–417. [[CrossRef](#)]
37. Pandya, M.; Rasmussen, M.; Hansen, A.; Nielsen, M.; Buus, S.; Golde, W.; Barlow, J. A modern approach for epitope prediction: Identification of foot-and-mouth disease virus peptides binding bovine leukocyte antigen (BoLA) class I molecules. *Immunogenetics* **2015**, *67*, 691–703. [[CrossRef](#)]
38. Nielsen, M.; Connelley, T.; Ternette, N. Improved prediction of bovine leukocyte antigens (BoLA) presented ligands by use of mass-spectrometry-determined ligand and in vitro binding data. *J. Proteome Res.* **2018**, *17*, 559–567. [[CrossRef](#)]
39. Zikova, A.; Schnauffer, A.; Dalley, R.A.; Panigrahi, A.K.; Stuart, K.D. The F(0)F(1)-ATP synthase complex contains novel subunits and is essential for procyclic *Trypanosoma brucei*. *PLoS Pathog.* **2009**, *5*, e1000436. [[CrossRef](#)]
40. Simwango, M.; Ngonyoka, A.; Nnko, H.J.; Salekwa, L.P.; Ole-Neselle, M.; Kimera, S.I.; Gwakisa, P.S. Molecular prevalence of trypanosome infections in cattle and tsetse flies in the Maasai Steppe, northern Tanzania. *Parasites Vectors* **2017**, *10*, 507. [[CrossRef](#)]
41. Dagenais, T.R.; Demick, K.P.; Bangs, J.D.; Forest, K.T.; Paulnock, D.M.; Mansfield, J.M. T-cell responses to the trypanosome variant surface glycoprotein are not limited to hypervariable subregions. *Infect. Immun.* **2009**, *77*, 141–151. [[CrossRef](#)]
42. Magez, S.; Radwanska, M.; Drennan, M.; Fick, L.; Baral, T.N.; Brombacher, F.; Baetselier, P.D. Interferon- $\gamma$  and nitric oxide in combination with antibodies are key protective host immune factors during *Trypanosoma congolense* Tc13 infections. *J. Infect. Dis.* **2006**, *193*, 1575–1583. [[CrossRef](#)]

43. Mansfield, J.M.; Paulnock, D.M.; Hedberg, G.M. Bridging Innate and Adaptive Immunity in African Trypanosomiasis. In *Trypanosomes and Trypanosomiasis*; Springer-Nature: London, UK, 2013; pp. 89–114.
44. Liu, G.; Sun, D.; Wu, H.; Zhang, M.; Huan, H.; Xu, J.; Zhang, X.; Zhou, H.; Shi, M. Distinct contributions of CD4+ and CD8+ T cells to pathogenesis of *Trypanosoma brucei* infection in the context of gamma interferon and interleukin-10. *Infect. Immun.* **2015**, *83*, 2785–2795. [[CrossRef](#)]
45. Onah, D.N.; Hopkins, J.; Luckins, A.G. Effects of the depletion of CD8+ T cells and monocytes on the proliferative responses of peripheral blood leucocytes from *Trypanosoma evansi*-infected sheep. *Vet. Parasitol.* **2000**, *92*, 25–35. [[CrossRef](#)]
46. Flynn, J.N.; Sileghem, M.; Williams, D.J. Parasite-specific T-cell responses of trypanotolerant and trypanosusceptible cattle during infection with *Trypanosoma congolense*. *Immunology* **1992**, *75*, 639–645.
47. Radwanska, M.; Guirnalda, P.; De Trez, C.; Ryffel, B.; Black, S.; Magez, S. Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. *PLoS Pathog.* **2008**, *4*, e1000078. [[CrossRef](#)]
48. Horn, D. Antigenic variation in African trypanosomes. *Mol. Biochem. Parasitol.* **2017**, *195*, 123–129. [[CrossRef](#)]
49. Villarreal, D.O.; Talbott, K.T.; Choo, D.K.; Shedlock, D.J.; Weiner, D.B. Synthetic DNA vaccine strategies against persistent viral infections. *Expert Rev. Vaccines* **2013**, *12*, 537–554. [[CrossRef](#)]
50. Yin, W.; Xiang, P.; Li, Q. Investigations of the effect of DNA size in transient transfection assay using dual luciferase system. *Anal. Biochem.* **2005**, *346*, 289–294. [[CrossRef](#)]
51. Jazayeri, S.D.; Poh, C.L. Recent advances in delivery of veterinary DNA vaccines against avian pathogens. *Vet. Res.* **2019**, *50*, 78. [[CrossRef](#)]
52. Zhao, Z.; Liu, H.; Zhang, K.; Xue, Q.; Chen, K.; Xue, Y. Minimum dose, antigen content, and immunization duration of a trivalent vaccine of inactivated *Haemophilus parasuis* serovars 4, 5, and 12 against Glässer's disease in pigs. *Can. J. Vet. Res.* **2016**, *80*, 287–293.
53. Tabel, H.; Wei, G.; Bull, H.J. Immunosuppression: Cause for failures of vaccines against African trypanosomiasis. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2090. [[CrossRef](#)]
54. Young, J.L.; Dean, D.A. Electroporation-mediated gene delivery. *Adv. Genet.* **2015**, *89*, 49–88.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).