



OPEN In silico design of a polypeptide as a vaccine candidate against ascariasis

Francisco M. D. Evangelista¹, Arnoud H. M. van Vliet¹, Scott P. Lawton² & Martha Betson¹✉

Ascariasis is the most prevalent zoonotic helminthic disease worldwide, and is responsible for nutritional deficiencies, particularly hindering the physical and neurological development of children. The appearance of anthelmintic resistance in *Ascaris* is a risk for the target of eliminating ascariasis as a public health problem by 2030 set by the World Health Organisation. The development of a vaccine could be key to achieving this target. Here we have applied an in silico approach to design a multi-epitope polypeptide that contains T-cell and B-cell epitopes of reported novel potential vaccination targets, alongside epitopes from established vaccination candidates. An artificial toll-like receptor-4 (TLR4) adjuvant (RS09) was added to improve immunogenicity. The constructed peptide was found to be non-allergic, non-toxic, with adequate antigenic and physicochemical characteristics, such as solubility and potential expression in *Escherichia coli*. A tertiary structure of the polypeptide was used to predict the presence of discontinuous B-cell epitopes and to confirm the molecular binding stability with TLR2 and TLR4 molecules. Immune simulations predicted an increase in B-cell and T-cell immune response after injection. This polypeptide can now be validated experimentally and compared to other vaccine candidates to assess its possible impact in human health.

Ascariasis is a disease caused by the soil-transmitted helminth (STH) *Ascaris*. This is the most prevalent STH worldwide with approximately 446 million people infected and is responsible for a loss of 754,000 disability adjusted life years (DALYs) in 2019¹. Ascariasis is classified as a Neglected Tropical Disease (NTD) and is typically associated with low- and middle-income regions, such as sub-Saharan Africa, South America, South- and South-east Asia^{2,3}. Concurrently, *Ascaris* is also one of the most prevalent intestinal nematode species in the domestic swine⁴. Ascariasis is similar in both humans and pigs, having a potential negative effect on growth, cognitive development and nutritional status^{5,6}. The two *Ascaris* species, *A. lumbricoides* and *A. suum*, are usually associated with infections in either humans or pigs, respectively. However, there is potential for zoonotic transmission between host species with *A. suum* infections identified in humans and *A. lumbricoides* infections reported in pigs^{3,7}. The two *Ascaris* species are also able to develop viable hybrids^{5,8}, reinforcing the need to control the parasite in both human and swine populations.

Preventive chemotherapy through Mass Drug Administration (MDA) is the main control method for ascariasis. This approach has been supported with large-scale screening protocols and improvements to water, sanitation and hygiene (WASH)^{2,6}. Recently, there have been increasing concerns about the potential development of resistance to anthelmintic treatment in *Ascaris*^{9,10}. Even though it has not become widespread, the appearance of resistance has led to a growing interest in other control/therapeutic methods, such as vaccines¹¹. At the point of writing, there is no vaccine that has gone through human clinical trials or that has been approved for commercial use in pigs. The development of a vaccine would help achieve the WHO 2030 Roadmap for NTD target of elimination of ascariasis as a public health problem².

Over the last few decades there have been a number of studies aiming to develop vaccines against *Ascaris*^{12–18}. The vaccines in these trials were tested mainly in mice and efficacies ranging from 38 to 77% reduction in larval burden. One of the issues with using mouse models is that *Ascaris* is not able to complete its life cycle from larvae to adult, so it is not possible to assess efficacy against the latter stage. These assays have used either *Ascaris* recombinant proteins^{12–14,16,18}, crude extracts¹⁷ or a subunit chimeric protein¹⁵. The latter proved to be the most efficient in reducing larval burden, with a 77%¹⁵ and 66% reduction¹⁹ when provided by subcutaneous injection or orally, respectively. This protein used a combination of in silico predicted B-cell epitopes from known vaccine

¹Department of Comparative Biomedical Sciences, School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7AL, UK. ²Centre for Epidemiology and Planetary Health, Department of Veterinary and Animal Sciences, Northern Faculty, Scotland's Rural University College (SRUC), An Lóchrán, Inverness IV2 5NA, UK. ✉email: m.betson@surrey.ac.uk

targets (As14, As16 and As37). More recently, using a mixed bioinformatics and in vitro methodology for B-cell epitope selection, ASCVac-1 was able to reduce larvae burden by 50%²⁰. Nevertheless, these values are still lower than the 94% efficacy with immunisation using UV-attenuated eggs²¹. Such differences led us to believe that there is still room for improvement and that other antigens or epitopes could be included in a multi-epitope vaccine.

The immune response in ascariasis is still not completely understood. It is however suggested that a T-helper type 2 (Th2) response with eosinophilia is mainly responsible for promoting resistance and controlling this parasite^{22–26}. The Th2 cytokines IL-4 and IL-10 are commonly elevated in humans and animals that have been previously exposed to *Ascaris* and are known to affect the recruitment of eosinophils^{24,27,28}. Furthermore, it is acknowledged that toll-like receptor (TLR)-2 and TLR-4 are important mediators in the adaptive immune response in helminth infections with cytokines and chemokines release^{24,29,30}. Animals without these receptors were found to be more prone to *Ascaris* reinfection due to lower eosinophil levels^{24,30}. It is also important to reinforce that antibodies, such as IgA, IgE and IgG, are also relevant in controlling *Ascaris*^{12,14,15,17,18,20,24,31–34}. Nevertheless, treatment and prevention of ascariasis requires a balance between the continuous local exposure to IgE and subsequent eosinophilia due to exacerbation of Loeffler syndrome, restrictive lung disease and gastrointestinal symptoms^{35–38}.

The aim of this study was to design and test an ascariasis vaccine using an in silico methodology. This approach has already used to develop vaccines for different pathogens including SARS CoV-2³⁹, *Mycobacterium tuberculosis*⁴⁰, *Schistosoma mansoni*⁴¹ and *Onchocerca volvulus*⁴². Here we have used antigens present in *A. lumbricoideis* and *A. suum* to devise a multi-epitope vaccine that could prevent infection in multiple hosts and interfere with zoonotic transfer of *Ascaris*.

Methods

Selection and retrieval of vaccination targets. The vaccination targets were selected based on the results and predictions of previous studies. A total of 7 different genes and respective proteins were selected: AgB13X_g094 (ATtype), AgR007_g063 (Apiezo), AgR007_g282 (Altype), AgB13X_g096 (Aproto)⁴³, AgB02_g183 (As37), AgR006_g148 (As16), and AgR006_g147 (As14)^{12,14,15,18}. The proteins Attype, Apiezo, Altype and Aproto were selected based on a reverse vaccinology framework⁴³, while the proteins As37, As16 and As14 are vaccine targets that have been successfully used in vaccination assays^{12,14,15,18}. The amino acid (aa) sequences were retrieved from the most recent *Ascaris* proteome (BioProject PRJNA62057) available in WormBase Parasite database⁴⁴. A workflow summary is available in Fig. 1.

Prediction and selection of helper T-cell epitopes and linear B-cell epitopes. Helper T-cell (HTC) epitopes for the Attype, Apiezo, Altype and Aproto proteins were retrieved from a previous study⁴³. HTC epitope predictions for the proteins As37, As16 and As14 were made using the IEDB MHC-II binding predictor v2.22.3 (MHCII-IEDB, available at <http://tools.iedb.org/mhcii/>). The IEDB-recommended 2.22 prediction Consensus method was used, encompassing the NN-align, SMM-align, CombLib and Sturniolo methods, or NetMHCIIpan⁴⁵. The peptide epitope length of 15 aa with a 9 aa core was selected and predictions were made against the 27 human leukocyte antigen (HLA) allele reference set⁴⁶. For each protein, we selected two non-redundant epitopes (without overlapping aa) that scored between 0 and 2 and were predicted to bind to the largest combined number of unique MHC-II alleles in MHCII-IEDB tool analysis. The selected epitopes were submitted to IL4pred (<https://webs.iitd.edu.in/raghava/il4pred/index.php>)⁴⁷ and IL-10Pred (<https://webs.iitd.edu.in/raghava/il10pred/>)⁴⁸ to assess their capacity to induce IL-4 and IL-10 production. In IL4pred we used the hybrid-based method in the Virtual Screening module with 0.2 SVM threshold. For IL-10 prediction we used the predict module with a threshold of -0.3 for the SVM based model.

B-cell epitopes for Attype, Apiezo, Altype and Aproto proteins were identified in our previous study⁴³. B-cell epitopes in As37, As16 and As14 were identified using Bepipred v2.0 (<http://www.cbs.dtu.dk/services/BepiPred/>)⁴⁹. The two highest scoring B-cell epitopes with a length between 8 and 40 aa were selected for each protein. These epitopes were then submitted to AbCPE webserver (<http://bioinfo.unipune.ac.in/AbCPE/Home.html>)⁵⁰ to predict which class of antibody they were capable of inducing.

Allergenicity, toxicity and epitope identity assessment. The retrieved HTC and B-cell epitopes were evaluated for allergenicity using Allertop 2.0 (<https://www.ddg-pharmfac.net/AllerTOP/index.html>)⁵¹ and toxicity with the ToxinPred web server (<https://webs.iitd.edu.in/raghava/toxinpred/index.html>)⁵². In ToxinPred, a 10 amino acid peptide fragment length was chosen, and toxicity was predicted using the SVM (Swiss-Prot) and Motif based methods, with 0.0 threshold and 10 e-value cut-offs respectively. Furthermore, epitopes were assessed for identity in both humans and pigs. These epitopes were submitted to a BLASTp (Basic Local Alignment Search Tool protein) search (available at <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). This was cross-checked by submitting the epitopes to the Multiple Peptide Match tool available in PIR (<https://research.bioinformatics.udel.edu/peptidematch/batchpeptidematch.jsp>)⁵³. Epitopes were matched against both human and pig proteomes using the UniProt/SwissProt option and including the different isoforms. Epitopes that were predicted to be toxic or were found in either humans or pigs were rejected, and a new epitope was selected for testing.

Design of a multi-epitope vaccine and immunological analysis. The epitopes identified in the previous section were fused together with flexible linkers and a TLR-4 agonist. The flexible linker GSGSG was selected to bind both HTC and B-cell epitopes^{15,54}. The synthetic TLR-4 adjuvant RS09 (Sequence: APPHALS) was added to the N-terminal of the vaccine construct using an EAAAK linker^{54,55}. The multi-epitope vaccine amino acid sequence was analysed for its immunological properties. Both AntigenPro (<http://scratch.proteomics.ics.uci.edu/>)⁵⁶ and VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>)⁵⁷ were uti-

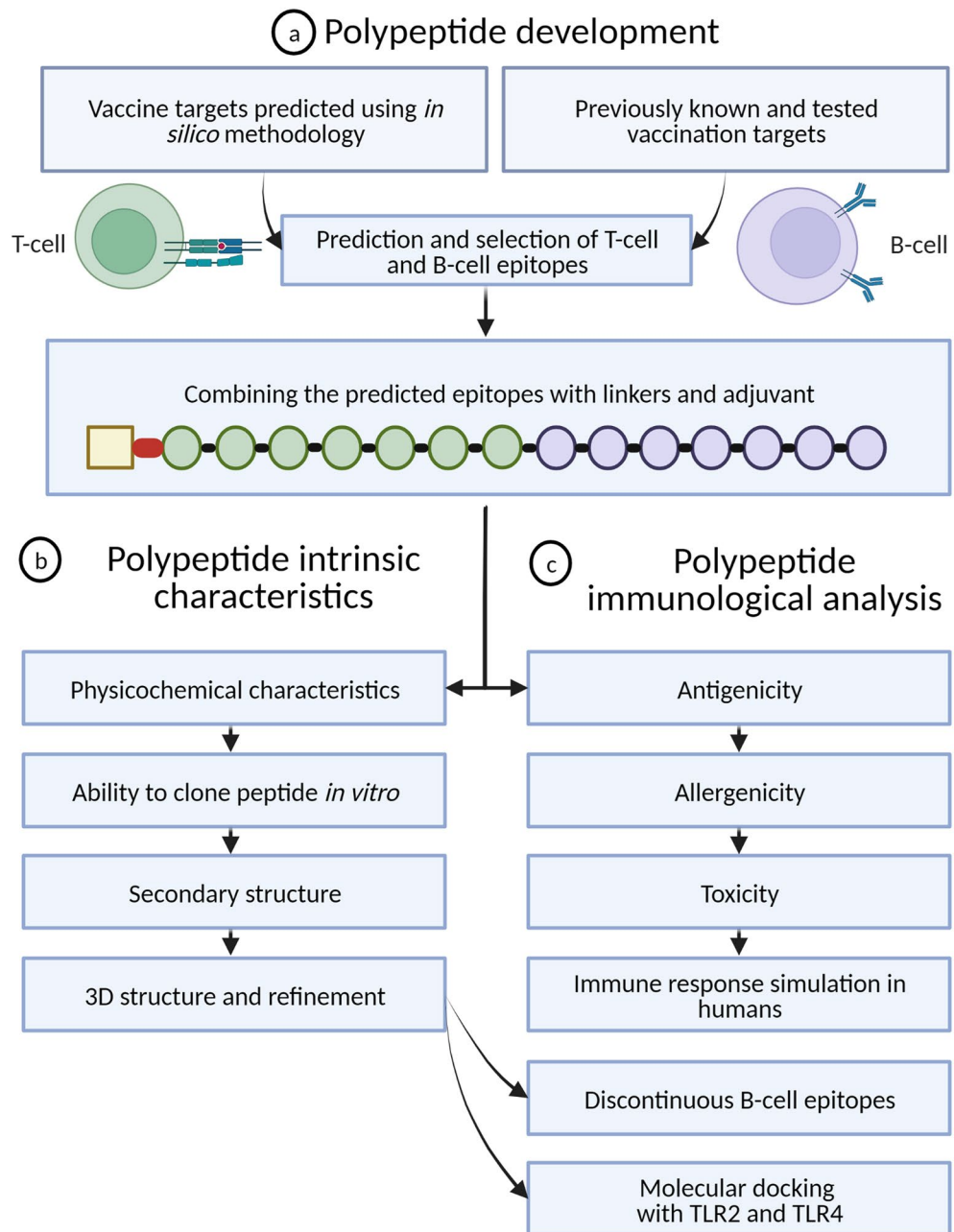


Figure 1. Summary workflow of the approach used to design and analyse the multi-epitope polypeptide. (a) Polypeptide development steps, including selection of vaccine targets and respective T-cell and B-cell epitopes. (b) Steps of the polypeptide intrinsic characteristics analysis. (c) Components of the immunological analysis.

used to assess the antigenicity of the multi-epitope vaccine construct. The VaxiJen v2.0 server was used with bacteria and parasite models, and a 0.7 threshold was used in both. Allergenicity was predicted using Allertop v2.0 and AllergenFP v1.0 (<http://www.ddg-pharmfac.net/AllergenFP/index.html>)⁵⁸. Vaccine toxicity assessment was made using the previously mentioned methodology with the ToxinPred webserver. The multi-epitope vaccine is expected to be antigenic, non-allergic and non-toxic. Furthermore, the coverage in the world population was assessed with the IEDB Population Coverage tool (<http://tools.iedb.org/population/>)⁵⁹. This was done by selecting “World” as target population, calculating MHC Class II separately and restricting the MHC alleles to the previously used 27 HLA reference set.

Immune simulation scenario in humans. The C-IMMSIM web server (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>)⁶⁰ was utilized for *in silico* simulations of human immune response to the multi-epitope vaccine. We considered two immunization schedules: one with two doses with a 4-week interval and a second scenario with a third dosage at the 8th week. In C-IMMSIM eight hours in real life correspond to one simulation step. Therefore,

the simulations were conducted with two injections at the time step 1 and 84 (vaccine injection happens during the first simulation step), with a third injection at step 168 for the second scenario. Both scenarios accounted for a total of 1096 simulation steps (one year). The injections in the tool were selected as vaccines with no LPS (as this is not present in the tested vaccine), random seed = 12,345, and the host HLA selection was modified to include the MHC-II alleles DRB1*07:01 and DRB1*15:01, as both are part of the HLA allele reference set and are known to react distinctly to different *Ascaris* antigens^{46,61}.

Prediction of physicochemical characteristics and secondary structure. The ProtParam web server (<http://web.expasy.org/protparam/>)⁶² was used to evaluate the physicochemical characteristics of the selected multi-epitope vaccine. The analysed characteristics included molecular weight, in vitro and in vivo half-life, instability index, theoretical isoelectric point (pI), aliphatic index and grand average of hydropathicity (GRAVY). The vaccine solubility upon overexpression in *Escherichia coli* was assessed with SOLPro (<http://scratch.proteomics.ics.uci.edu/>)⁶³. The vaccine secondary structure (2D) was predicted with the PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>)⁶⁴ and RaptorX Property Prediction (<http://raptorx.uchicago.edu/StructurePropertyPred/predict/>)⁶⁵ tools. These programs were utilised to evaluate the secondary structure of the vaccine regarding the presence of coils, α -helices, β -sheets and disordered domains with ~84% accuracy^{64,65}.

In-silico optimization and cloning of designed multi-epitope vaccine. The reverse translation and codon optimization of the multi-epitope vaccine were performed with the Java Codon Adaptation Tool (JCat) webserver (<http://www.jcat.de/>)⁶⁶ to prepare for future experimental work. This codon optimization was performed for expression in an *Escherichia coli* K12 derivative, while outputs of the Codon Adaptation Index (CAI) and the percentage of GC content were used to evaluate this step. The CAI provides an indication of how likely it is that an organism can express a heterologous gene. A good optimization has a CAI above 0.80 and a GC content between 30% and 70%, respectively. The NdeI and XhoI restriction sites were added to the N and C-terminus of the optimized nucleotide sequence, respectively. The final sequence was inserted into the pET30a (+) vector and tested for its predicted viability using the SnapGene software (from Insightful Science; available at <https://snapgene.com>).

Tertiary structure prediction, refinement, and validation. The three-dimensional (3D) model of the multi-epitope protein was created using the RaptorX Structure Prediction tool (<http://raptorx.uchicago.edu/ContactMap/>)⁶⁷. The 3D structures are predicted using a de novo deep-learning model that predicts a contact map based on a multiple sequence alignment from the primary sequence and is especially effective with proteins that do not possess various sequence homologues. The first-ranked balanced-energy protein model was retrieved and refined using GalaxyRefine web server (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>)⁶⁸. These tools were amongst the highest rated tools for protein 3D modelling and refinement in the CASP13 assessments^{69,70}. Both the unrefined and refined 3D models were validated using the ProSA-Web web server (<https://prosa.services.came.sbg.ac.at/prosa.php>)⁷¹ and the PROCHECK⁷², ERRAT⁷³, VERIFY 3D⁷⁴ and PROVE⁷⁵ programs available in SAVES v6.0 (<https://saves.mbi.ucla.edu/>). ProSA-Web analyses the protein 3D model by calculating an overall quality Z-score and comparing it to the score of protein structures obtained through X-ray analysis, NMR spectroscopy and theoretical calculations. The PROCHECK program validates a protein structure by using a Ramachandran plot and visualizing the percentage of aa residues that are in favoured, allowed and disallowed regions. While ERRAT is used to assess the overall model quality at a 0 to 100 scale using atomic composition, VERIFY 3D checks the model quality from the 3D model and compares it directly to its amino acid composition. Finally, PROVE assesses model quality by using an atom volume-based validation.

Discontinuous B-cell epitope prediction. Besides the linear B-cell epitopes already described, there was the need to assess the discontinuous B-cell epitopes as they are estimated to account for over 90% of the B-cell immunogenic response⁷⁶. The presence of discontinuous B-cell epitopes formed from protein folding was analysed with the ElliPro server (<http://tools.iedb.org/ellipro/>)⁷⁷. A minimum score of 0.5 and a maximum distance of 6 Angstrom was used and predicted discontinuous and linear epitopes were retrieved.

Molecular docking of designed vaccine with TLR2 and TLR4. Immune response against *Ascaris* has been reported to be supported by both TLR2 and TLR4 interactions^{24,30}. The ClusPro v2.0 web server (<https://cluspro.bu.edu/home.php>)⁷⁸ was used to dock the TLR2 (PDB ID: 6NIG)⁷⁹ and TLR4 (PDB ID: 4G8A)⁸⁰ proteins with the multi-epitope vaccine, individually. ClusPro protein docking was used with the vaccine as ligand and TLR2 and TLR4 as receptors. The highest ranked model for each docking prediction was retrieved and evaluated with PRODIGY (<https://wenmr.science.uu.nl/prodigy/>)⁸¹. PRODIGY was used to assess the binding affinity, dissociation constant, and the number of contacts created between the multi-epitope vaccine and both TLR2 and TLR4 at 37 °C (protein–protein complexes). The number and type of contacts was also predicted using the LigPlot + v2.2.7 DIMPLOT option⁸². The iMODS web server (<http://imods.chaconlab.org/>)⁸³ was used for a fast molecular dynamics study to define and calculate the protein flexibility following molecular docking. The Basic interface was used with the CA option for the Coarse Grain model representations.

Results

Selected vaccination targets and respective epitopes for vaccine design. A total of seven *Ascaris* amino acid protein sequences were retrieved from the WormBase database to design a multi-epitope vaccine against ascariasis. Four proteins highly expressed in adults (ATtype, APiezo, ALtype and AProto) and three proteins expressed in both larval and adult stages (As37, As16 and As14), were selected based on recent bioinformatics analysis⁴³ and previous immunization trials^{12,14,15,18}. According to transcriptomic and immunolocalization data, these proteins are mostly located in the muscle (APiezo and ALtype), muscle and hypodermis (As37), ovaries (AProto and ATtype), or are secreted (As16 and As14)^{8,13,14,84}.

As T-helper cell epitopes have already been described for AProto, ATtype, ALtype and APiezo⁴³, we selected two T-helper epitopes for As37, As16 and As14 based on the results of the MHCII-IEDB tool. A total of 14 HTC epitopes with a length of 15 aa were selected, two for each protein. These epitopes were predicted to be non-allergic, non-toxic and not present in both humans and pigs according to BLASTp and PIR (Table 1). Twelve of the fourteen HTC epitopes were predicted to induce IL-4 or IL-10 production (Supplementary Table S1).

As before, B-cell epitopes were already selected for some proteins. For As37, As16 and As14, the two highest scoring epitopes in Bepipred 2.0 with a length between 8 and 40 aa, were selected. As before, we retrieved 14 non-allergic, non-toxic B-cell epitopes that were not present in both humans and pigs (Table 1). All B-cell epitopes were predicted to bind to IgG.

Multi-epitope vaccine is predicted to have the intended immunological characteristics. The selected 14 T-helper epitopes were linked together with the 14 B-cell epitopes due to the flexible linker GSGSG. The TLR-4 synthetic adjuvant RS09 was included in the N-terminal of the vaccine, before HTC epitopes, thanks to an EAAAK linker^{54,55}. The final vaccine has 579 amino acids, and a schematic representation of its design is available in Fig. 2. The vaccine protein sequence can be seen in Supplementary Data S1.

The multi-epitope vaccine was predicted to be antigenic in both Vaxijen 2.0, bacteria and parasite models, and AntigenPro, with scores of 1.1752, 0.8079, and 0.729925, respectively. Both AllergenFP and Allertop 2.0 predicted the vaccine to be non-allergenic while the ToxinPred webserver found it to be non-toxic. Furthermore, the vaccine was calculated to be able to cover 99.98% of the world population.

Immune simulation highlights an immunological response consistent with actual vaccination outcomes. The immune simulation scenarios predicted by the C-ImmSim server predicted similar behaviours for the two different vaccination schedules (Fig. 3). Both scenarios were characterized by a secondary immune response with a rapid increase in the populations of B-cell and HTC, stimulating the high levels of IgM and IgG in response to vaccine injection. Following each injection, the levels of IgG increased proportionally higher than IgM, which further reinforces the characterization as a secondary immune response. Both memory B-cells and HTCs were especially elevated, suggesting long-lasting immune responses. When comparing both scenarios, the three-dose scenario had a significantly higher antibody titer for both IgM and IgG and a slightly higher memory B-cell count after one year. The total level of HTC, specifically memory HTC, was similar between the two injection schemes, both after injections and one year after. The levels of both IFN- γ and IL-2 were similar, independently to the number of boosters used.

Physicochemical and secondary structure of the designed polypeptide. The polypeptide molecular weight was predicted to be 59.7 kDa with a pI of 9.58. An instability index of 28.52 indicated the protein stability (proteins with an index lower than 40 are considered stable) while a GRAVY of -0.363 predicts the vaccine as hydrophilic in nature⁸⁵. The estimated aliphatic index of 67.94 points towards the thermostability of the multi-epitope protein⁸⁶. The multi-epitope protein's half-life was predicted as 4.4 h in in vitro mammalian reticulocytes, over 20 h in yeast, and around 10 h in *E. coli*. The final vaccine peptide secondary structure was estimated to be 55.1% coils, 30.6% α -sheets, and 14.3% β -strands (Fig. 4). The disordered domains are predicted to be 50% according to the RaptorX Property Prediction tool. A solubility score of 0.935892, indicating that

WormBase protein identifier	Protein name	Helper T-cell epitopes	B-cell epitopes
AgB13X_g094	ATtype	LRLRALRPLRVINR FKNFGMAFLTLFRIA	DATGVDMQPVENYN SIPPKSVSR
AgB13X_g096	AProto	HTFRRFITAISLLDR NQEGVVHILSRKIFD	LSQSDHHILPRFANFVDDRTESLRVITQLLCSLR RQQFTLTFPPYFSDGKFK
AgR007_g063	APiezo	NCLKYFANFFFYRFG SLFLRPMRVALALLN	LLSVHLKNDSDSIEST VDPSFDPVIPKEEVI
AgR007_g282	ALtype	NNNFHTFPAAILVLF ERSLLCLTLNPLRK	ALNDETHIHRNNN SNEEDRGPVYNA
AgB02_g183	As37	TTELKQDNRFSLRDL VMVMEFRAKSILKPT	PQGAPTFRKPKQILQKTSDSGD PLDDGADDA
AgR006_g148	As16	AEYEKAHAAAIAKFS YTNKFKAFKAELKAH	EGQTSPSRVPPF EDAKLNGIQKRQKIKETME
AgR006_g148	As14	KITSLQLSPAAVQA KVLIIFVAIVVIAFA	TQMQQGKARAEAA DNPNLKGREKQKITSL

Table 1. List of proteins and their respective epitopes used in the multi-epitope vaccine.

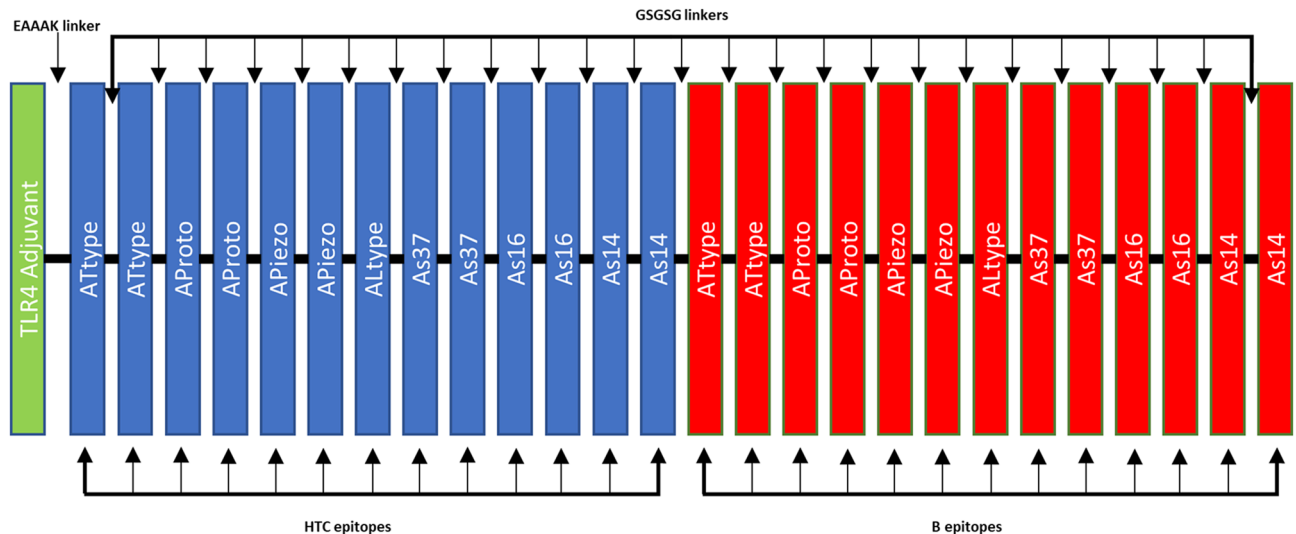


Figure 2. Schematic representation of the final multi epitope vaccine. The sequence contains 579 amino acids with the RS09 TLR4 adjuvant (green) linked at the N-terminal through an EAAAK linker. Both Helper T-cell (blue) epitopes and B-cell epitopes (red) are linked with a GSGSG linker.

the vaccine was soluble after overexpression in *E. coli*. The JCat tool was able to generate an optimized DNA sequence for the vaccine sequence to be expressed in *E. coli* (strain k12). The obtained codon sequence has 1737 nucleotides with a predicted GC content of 51.58% and a CAI of 1, optimal for protein expression. The DNA sequence content is available in Supplementary Data S2 and S3. The optimized codon sequence was successfully inserted into the pET30a (+) expression vector using SnapGene (Fig. 5). Both solubility scores and successful sequence insertion reinforce the likelihood of good protein expression when using *E. coli* production model.

Vaccine 3D model. The first 3D protein model predicted by RaptorX Structure Prediction also had the lowest RMSD of 13.039 and was selected for refinement (Fig. 6a). When compared to the original model, the first refined model provided by GalaxyRefine had an improved Clash Score from 20.8 to 13.4, a Rama favoured value increase from 86.1 to 92.0, MolProbity decrease from 2.4521 to 2.119, RMSD from 0 to 0.444, GDT-HA from 1 to 0.9404, and an increase in poor rotamers from 0.5 to 0.7 (Fig. 6b). This improved model was validated through a lower Z-score in ProSa-Web, -7.87 to -7.9, and improved residue positions in the Ramachandran plot in Procheck (Fig. 6c,d). The Procheck tool calculated that the proportion of residues in favoured and allowed regions improved from 71.9% and 26.1% to 80.8% and 17.3%, while the residues in disallowed regions remained 2.0%. While ERRAT failed the model with a 76.52% score and PROVE failed the model due to 11.5% of the atoms being buried in an outlier position, VERIFY 3D passed the model with a score of 86.53%. This model was then used in ElliPro's prediction for discontinuous B-cell epitopes. In this, the polypeptide is expected to have eight discontinuous epitopes that scored between 0.518 and 0.782. These epitopes have a total of 320 amino acids and their size ranged between seven and 115 residues (Supplementary Table S2).

Designed vaccine is predicted to interact and bind with TLR2 and TLR4. We selected the first-ranked complex models between the vaccine and both TLR4 and TLR2, as predicted by ClusPro 2.0 (Fig. 7a,b). The PRODIGY server estimates a strongest binding affinity between the vaccine and TLR4 (-19.0 kcal/mol) than with TLR2 (-12.6 kcal/mol), which is supported by a lower dissociation constant (4.20×10^{-14} M and 1.30×10^{-09} M, respectively) and a higher number of contacts, 177 between the vaccine and TLR4 and 108 with TLR2 (Supplementary Table S3). LigPlot+ corroborated the PRODIGY contact results by predicting more hydrogen bonds (22 and 20), non-bonded contacts (212 and 150) and salt bridges (4 and 2) between the vaccine and TLR4 than with TLR2 (Supplementary Tables S4–S9). The iMODS molecular dynamics study results for TLR4 and TLR2 vaccine-binding complexes can be observed in Figs. 8 and 9, respectively. The normal mode analysis (NMA) analysed the complex mobility where the areas highlighted in blue colour represent the areas with the lowest mobility and red the areas with the highest mobility (observed in Figs. 8a and 9a). The B-factor value scores corroborate the NMA analysis (Figs. 8b and 9b). The deformability index graph highlights that the areas in the beginning of the TLR molecules and in the centre of the polypeptide molecule are the ones with the higher capacity to deform (Figs. 8c and 9c). Furthermore, the calculated eigenvalues of $5.103652e-06$ and $1.591150e-06$ for TLR4 and TLR2 complexes, respectively, indicate a slightly higher stiffness in the TLR4-vaccine complex (Figs. 8d and 9d). The variance graph is inversely related to the eigenvalues and highlights in red and green the individual and cumulative variances, respectively (Figs. 8e and 9e). The covariance map (Figs. 8f and 9f) and the elastic network (Figs. 8g and 9g) underpin that the areas with correlated motions correspond to the areas with higher stiffness within the docking complexes. These results suggest that both complexes are stable, with the TLR4-vaccine complex requiring higher energy to be deformed and, therefore, aligning with the previous PRODIGY results.

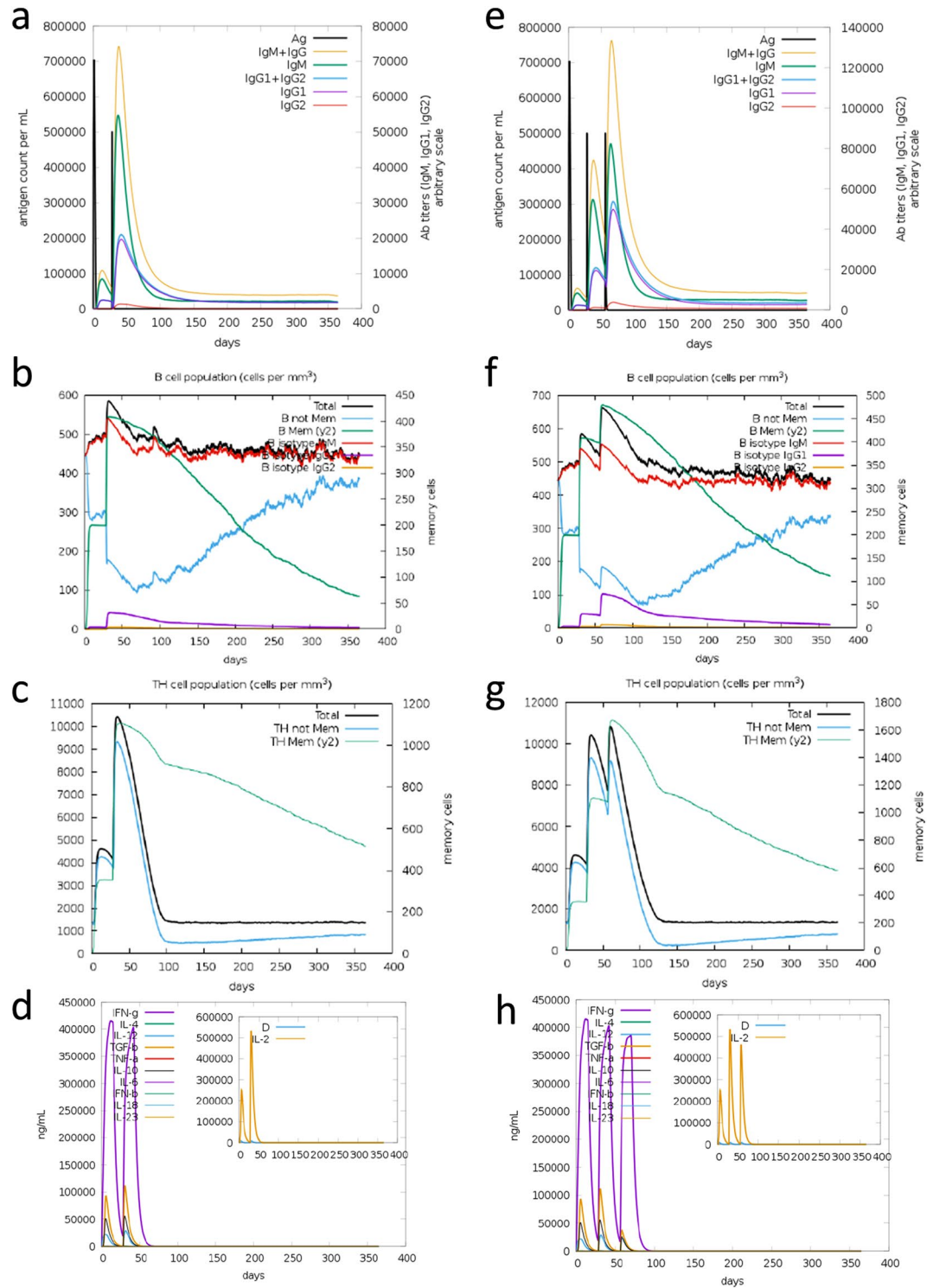


Figure 3. C-ImmSim in silico immune simulation with the designed vaccine complex. **(a–d)** Immune simulation with one booster after four weeks. **(e–h)** Immune simulation with two boosters at four and eight weeks. **(a, e)** Antibody titer following vaccine injection. **(b, f)** B-cell population upon vaccine injection. **(c, g)** The evolution of T-helper cell population. **(d, h)** Cytokines and interleukins concentration.

Discussion

In this work, we have designed a multi-epitope polypeptide as a vaccine component against *Ascaris* using an in silico approach. The protein was predicted to be antigenic, non-allergenic and non-toxic, while also being stable



Figure 4. The vaccine secondary structure as estimated by PSIPRED 4.0. The predicted 2D structure contains coils (55.1%), α -sheets (30.6%) and β -strands (14.3%).

and soluble. Immune simulations and molecular docking predicted its ability to promote the development of B-cell and memory T-helper cells, IgG, and the binding to both TLR2 and TLR4.

Despite the improvements to control strategies and their implementation, *Ascariis* remains the most prevalent STH worldwide¹. Currently, there is no vaccine developed against *A. lumbricoides* or *A. suum* for use in either humans or pigs^{11,28}. Recently, helminth vaccine research has included the development of subunit vaccines^{11,87}. The polypeptide developed here using an in silico approach included epitopes from previously

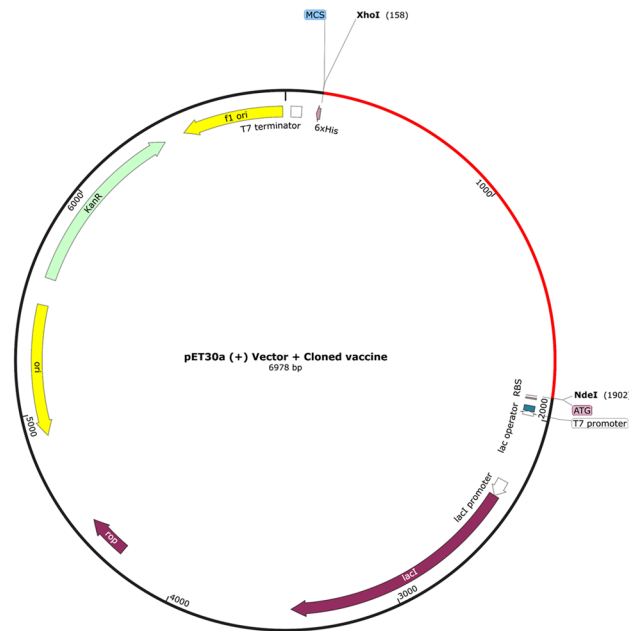


Figure 5. In silico cloning into the pET30a(+) expression vector. The red circle component represents the vaccine coding sequence while the black represents the vector backbone.

tested *Ascaris* antigens and proteins recently identified as potential vaccine candidates using bioinformatics methodologies^{12,14,15,18,43}. This antigen selection differed from previous studies that chose proteins based only on a bioinformatics approach or known vaccine targets^{15,88}. The selected proteins were present in both *A. lumbricoides* and *A. suum*, due to the known potential for host cross-infection and hybridization between both species^{8,89}. Having a vaccine that targets both species is increasingly relevant in areas where both humans and pigs co-exist^{3,7,90}. This contrasts to the approach used in previous in silico designed vaccines that focused on antigens only present in *A. lumbricoides* proteome⁸⁸. The epitope selection in the present study also differs from the previous ones. Both B-cell and HTC epitopes were selected while the chimeric vaccine was created solely based on B-cell epitopes¹⁵. Nonetheless, there is overlap between our selected B-cell epitopes for As14 and As37 and the ones in the aforementioned study. There was no overlap between As16 selected epitopes between the studies that could be explained by the different program version used (Bepipred 1.0 and Bepipred 2.0). The HTC epitope selection in the present study was based on binding to a 27 MHC-II human reference allele set, contrasting to the two-allele set used by the previous in silico study⁸⁸, and were present in extracellular domains of the selected antigens. MHC-II molecules appear to have a prominent role in the control of nematode infections in mammals by presenting antigenic peptides (through their epitopes) to HTC^{11,91}. This makes the discovery of epitopes that bind to these molecules a priority for the design of multi-epitope/subunit-based vaccines. The role of B-cells in *Ascaris* infection is still not fully understood but it is known to reduce larvae burden through antibody production^{15,17,23,92}. The selection of non-allergic, non-toxic, and proteins and epitopes without orthologues in humans or pigs should minimise the risk of adverse reactions to the vaccine. It is important to note that the selected B-cell epitopes were also predicted to only induce IgG production. Even though IgG is known to be a major component in most *Ascaris* vaccination assays^{12,13,15,17,32}, both IgE and IgA also play a role in disease control^{12,17,24,33}. For this reason, exploring other epitopes could be an interesting proposal for further optimizing a multi-epitope *Ascaris* vaccine.

The development of a subunit vaccine requires the selection of appropriate linkers between the selected epitopes as epitopes tend to not promote an immunological response by themselves⁵⁴. These linkers not only improve antigen processing and presentation, they should also provide the needed structural rigidity and flexibility to the vaccine^{54,93}. The selection of EAAAK and GSGSG epitope linkers was made based on previous nematode subunit vaccines studies^{15,41,42}. The EAAAK linker was used to fuse the vaccine with the RS09 TLR-4 agonist. TLR-4 mediates immune responses to *Ascaris* and the use of a synthetic agonist is known to be a safe alternative to natural agonists^{24,30,55,94}. The use of other adjuvants, such as Monophosphoryl lipid A (MPLA™), as already been proven useful in controlling *Ascaris* infection and should be considered in future assays^{15,95}. The high antigenicity, non-allergenicity and non-toxicity scores strengthen the potential usefulness as a good vaccination candidate. The multi-epitope vaccine was predicted to be both stable (with a half-life of 10 h) and soluble upon expression *E. coli*, supporting its potential for production in these commonly used expression bacteria systems. This was also reinforced by the ability to optimize the peptide coding sequence for expression in *E. coli* k12 using the pET30a (+) vector. *E. coli* is the most popular system for the production of recombinant proteins and has been optimized for its fast growth and high protein production output⁹⁶.

Knowledge of vaccine secondary and tertiary structure is a staple in understanding their potential usefulness⁹⁷. The 2D structure analysis indicated an existence of 55.1% coils and 30.6% α -sheets, allowing the protein to be

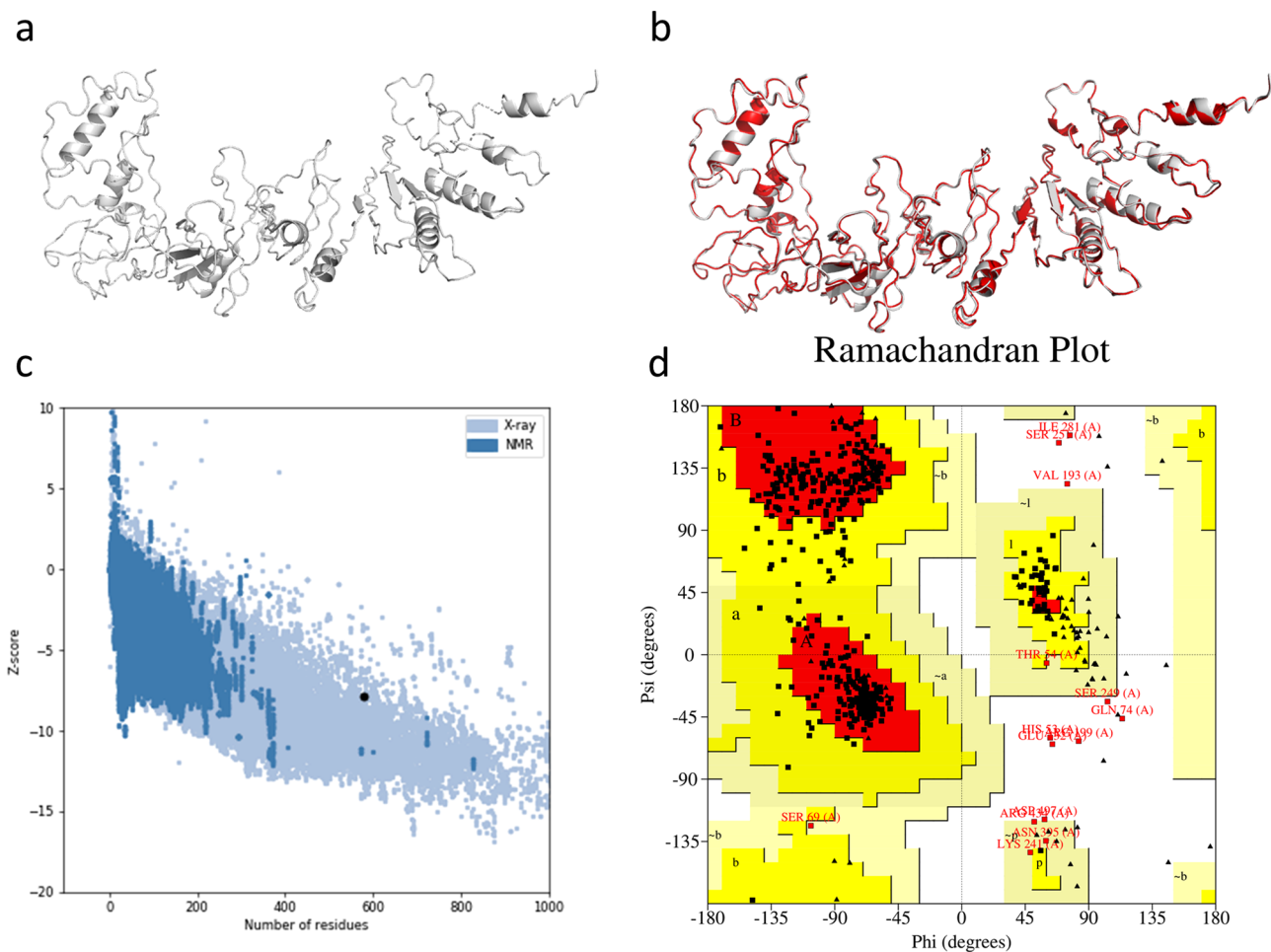


Figure 6. Vaccine 3D modelling, refinement, and assessment. **(a)** The 3D model produced by RaptorX (gray). **(b)** The refined 3D model with GalaxyRefine (red) overlapped with the original RaptorX 3D model (gray). **(c)** The Z-score (-7.9) obtained with ProSa-Web. **(d)** The Procheck Ramachandran plot showing 80.8% of the residues in favoured regions, 17.3% in allowed regions, and 2% in disallowed regions.

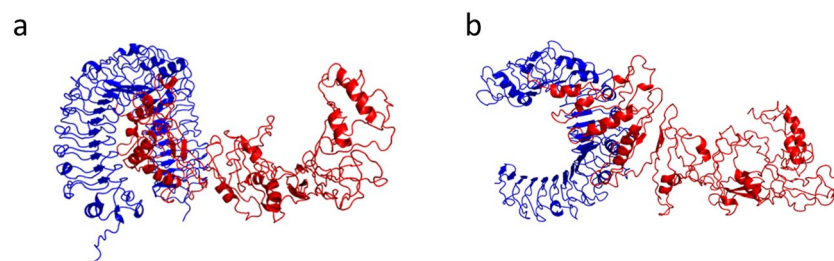


Figure 7. Molecular docking complexes predicted by Cluspro 2.0. **(a)** TLR4 (blue) and vaccine (red). **(b)** TLR2 (blue) and vaccine (red).

recognized by antibodies⁹⁷. This structure is similar to what was predicted for other *in silico* vaccine studies^{40,42,95}. The predicted protein tertiary structure improved after refinement. The Ramachandran plot highlighted the presence of 98% of the residues in favoured and allowed regions, indicating the high-quality of the produced model⁷². Although analysis by ERRAT and PROVE were negative, the analysis with ProSA-Web, VERIFY 3D and PROCHECK validate the protein 3D model. These results go in line to what was found in other *in silico* vaccine design studies, where residues in disallowed regions using Ramachandran plot were between zero and five per cent^{40–42,98}. A high-quality 3D model of a polypeptide is needed to assess the presence of discontinuous B-cell epitopes and to accurately predict its binding to proteins responsible for the hosts immune response. Even though the designed vaccine was constructed with linear B-cell epitopes, these are only responsible for 10% of the B-cell response⁷⁶. With the prediction of eight discontinuous B-cell epitopes, it is foreseen to be able to stimulate B-cell binding and recognition. Alongside B-cells, both TLR2 and TLR4 are known to promote immune response

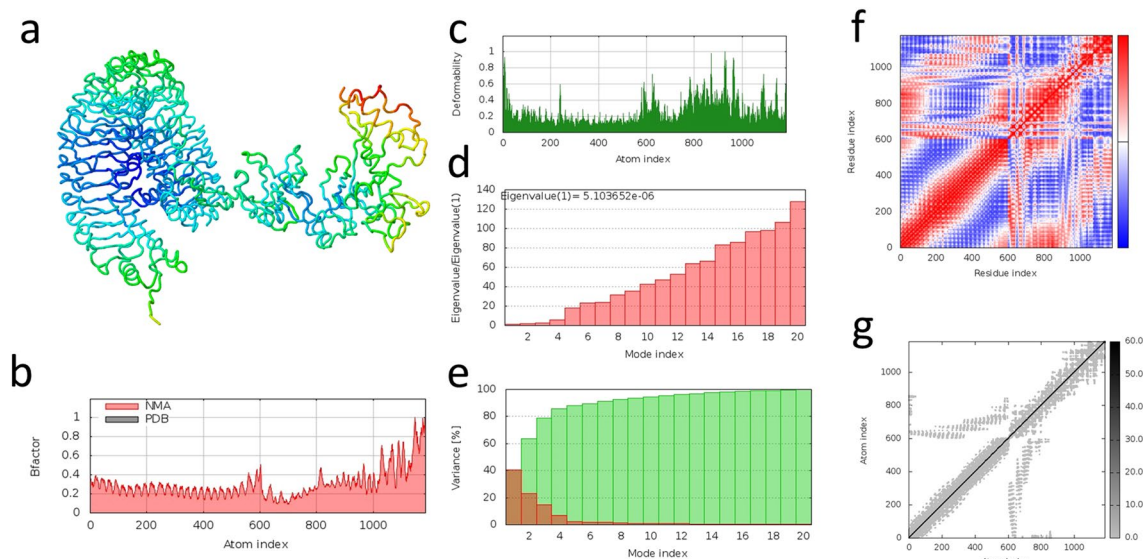


Figure 8. Molecular dynamics analysis of the TLR4 and vaccine complex by the iMODS server (<https://imods.iqfr.csic.es/>). **(a)** TLR4 and vaccine docking complex Normal Mode Analysis (NMA) highlighting areas of low mobility (blue) to high mobility (red). **(b)** B-factor values with regional mobility scores from 0 (lowest) to 1 (highest). These values are experimentally similar to NMA mobility analysis in a. **(c)** Deformability index of each residue in the complex with higher values reflecting a higher capacity to deform. **(d)** Eigenvalue where the lower it is, the easier it is to deform the complex. **(e)** Variance is inversely related to the eigenvalue. Green indicates cumulative variances while red indicates individual variances. **(f)** Covariance map with correlated (red), uncorrelated (white) or anti-correlated (blue) motions. **(g)** Elastic network where darker grey areas indicate higher stiffness.

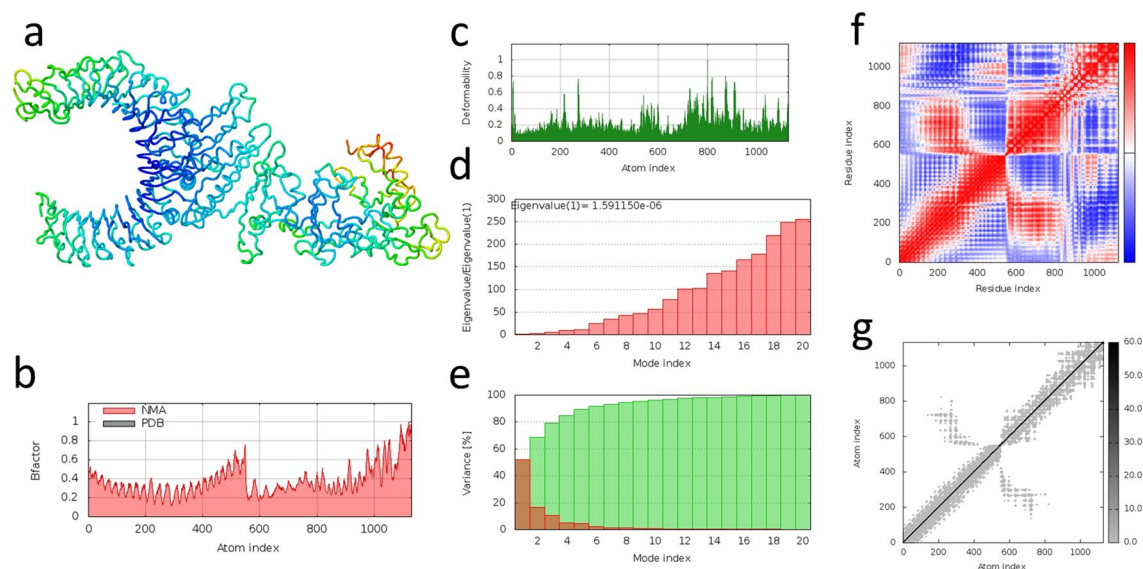


Figure 9. Molecular dynamics analysis of the TLR2 and vaccine complex by the iMODS server (<https://imods.iqfr.csic.es/>). **(a)** TLR2 and vaccine docking complex Normal Mode Analysis (NMA) highlighting areas of low mobility (blue) to high mobility (red). **(b)** B-factor values with regional mobility scores from 0 (lowest) to 1 (highest). These values are experimentally similar to NMA mobility analysis in a. **(c)** Deformability index of each residue in the complex with higher values reflecting a higher capacity to deform. **(d)** Eigenvalue. The lower the value, the easier it is to deform the complex. **(e)** Variance is inversely related to the eigenvalue. Green indicates cumulative variances while red indicates individual variances. **(f)** Covariance map with correlated (red), uncorrelated (white) or anti-correlated (blue) motions. **(g)** Elastic network where darker grey areas indicate higher stiffness.

against *Ascaris* and play a vital role in innate immune responses^{24,30}. The docking study with Cluspro 2.0 was able to predict binding between the designed vaccine and both TLR2 and TLR4. Even though the polypeptide binding was higher to TLR4 than to TLR2, this could be explained due to the addition of the synthetic TLR4 agonist to the N-terminal of the peptide.

The immune simulation scenarios suggested the development of a favourable immune response after immunization. This immune simulation server delivers the vaccine through injection. It has been shown that it be worth considering other delivery methods as they could be more effective¹⁹. Despite a similar increase in both B-cell and HTC populations in both scenarios, antibody titer was higher in the three-dose scenario than the two-dose scenario, suggesting that it could be more useful to prevent further larvae infection²⁴. The predicted increase in IgG production is a desirable scenario for ascariasis as it is associated with larvae reduction and protection against infection^{15,17}. IL-4 and IL-10 stimulation was observed in both scenarios, with the three-dose scenario having a smaller peak after the second booster. Nevertheless, the differences observed might not prove to be enough to justify a second booster based on economic or immunological considerations. The immune simulation reinforces the epitope analysis data that these would induce the production of IL-4, IL-10 and IgG. These results are similar to what was found in previous *Ascaris* vaccination assays, including the most recent ones that used a multi-epitope vaccine^{14,15,17,20,32}.

Alongside with other recently published vaccines, the one developed in this study should be ready to be expressed and studied both in vitro and in vivo using animal models^{15,88}. These studies should be able to provide a clear understanding of the vaccine capacity to stimulate an adequate immune response and if this could be improved by a new combination of epitopes or antigens.

Conclusion

In this study we have proposed a multi-epitope polypeptide, based on a variety of in silico analyses that jointly investigate important immunological properties that such a vaccine should include. This strategy will allow rapid prediction of the potential of vaccines before they are tested experimentally. The developed vaccine uses B-cell and HTC epitopes from proteins existent in both *A. lumbricoides* and *A. suum* and is predicted to be able to develop protection against these zoonotic species. The present study provides an alternative development approach to vaccines against *Ascaris* that could prove useful in future studies.

Data availability

The datasets used and analysed in this study are available in the article/Supplementary information. For further information do contact the corresponding author.

Received: 10 November 2022; Accepted: 23 February 2023

Published online: 02 March 2023

References

- Institute for Health Metrics and Evaluation (IHME). *GBD 2019 Cause and Risk Summary*, vol. 393 http://www.healthdata.org/results/gbd_summaries/2019 (2020).
- World Health Organization. *Ending the Neglect to Attain the Sustainable Development Goals: A Road Map for Neglected Tropical Diseases 2021–2030*. (2020).
- Kajero, O. T. *et al.* Co-infection of intestinal helminths in humans and animals in the Philippines. *Trans. R. Soc. Trop. Med. Hyg.* **1**, 002. <https://doi.org/10.1093/trstmh/trac002> (2022).
- Thamsborg, S. M., Nejsum, P. & Mejer, H. Impact of *Ascaris suum* in livestock. In *Ascaris: The Neglected Parasite* (ed. Holland, C.) 363–381 (Elsevier, 2013). <https://doi.org/10.1016/B978-0-12-396978-1.00014-8>.
- Dold, C. & Holland, C. V. *Ascaris* and ascariasis. *Microb. Infect.* **13**, 632–637 (2011).
- Ásbjörnsdóttir, K. H., Means, A. R., Werkman, M. & Walson, J. L. Prospects for elimination of soil-transmitted helminths. *Curr. Opin. Infect. Dis.* **30**, 1–10 (2017).
- Miller, L. A. *et al.* Ascariasis in humans and pigs on small-scale farms, Maine, USA, 2010–2013. *Emerg. Infect. Dis.* **21**, 332 (2015).
- Easton, A. *et al.* Molecular evidence of hybridization between pig and human *Ascaris* indicates an interbred species complex infecting humans. *Elife* **9**, e61562 (2020).
- Krücken, J. *et al.* Reduced efficacy of albendazole against *Ascaris lumbricoides* in Rwandan schoolchildren. *Int. J. Parasitol. Drugs Drug Resist.* **7**, 262–271 (2017).
- Furtado, L. F. V. *et al.* First identification of the benzimidazole resistance-associated F200Y SNP in the beta-tubulin gene in *Ascaris lumbricoides*. *PLoS ONE* **14**, e0224108 (2019).
- Zawawi, A. & Else, K. J. Soil-transmitted helminth vaccines: Are we getting closer?. *Front. Immunol.* **11**, 2426 (2020).
- Tsuji, N. *et al.* Intranasal immunization with recombinant *Ascaris suum* 14-kilodalton antigen coupled with cholera toxin B subunit induces protective immunity to *A. suum* infection in mice. *Infect. Immun.* **69**, 7285–7292 (2001).
- Tsuji, N. *et al.* Recombinant *Ascaris* 16-kilodalton protein-induced protection against *Ascaris suum* larval migration after intranasal vaccination in pigs. *J. Infect. Dis.* **190**, 1812–1820 (2004).
- Versteeg, L. *et al.* Protective immunity elicited by the nematode-conserved As37 recombinant protein against *Ascaris suum* infection. *PLoS Negl. Trop. Dis.* **14**, e0008057 (2020).
- de Castro, J. C. *et al.* Vaccination with chimeric protein induces protection in murine model against ascariasis. *Vaccine* **39**, 394–401 (2021).
- Chen, N. *et al.* *Ascaris suum* enolase is a potential vaccine candidate against ascariasis. *Vaccine* **30**, 3478–3482 (2012).
- Gazzinelli-Guimarães, A. C. *et al.* IgG induced by vaccination with *Ascaris suum* extracts is protective against infection. *Front. Immunol.* **9**, 25–35 (2018).
- Wei, J. *et al.* Yeast-expressed recombinant As16 protects mice against *Ascaris suum* infection through induction of a Th2-skewed immune response. *PLoS Negl. Trop. Dis.* **11**, 1–20 (2017).
- Castro, J. C. *et al.* Bioaccessibility and oral immunization efficacy of a chimeric protein vaccine against *Ascaris suum*. *Microb. Infect.* **25**, 105042. <https://doi.org/10.1016/j.micinf.2022.105042> (2022).
- Gazzinelli-Guimarães, A. C. *et al.* ASCVac-1, a multi-peptide chimeric vaccine, protects mice against *Ascaris suum* infection. *Front. Immunol.* **12**, 5318 (2021).

21. Urban, J. F. J. & Tromba, F. G. An ultraviolet-attenuated egg vaccine for swine ascariasis: Parameters affecting the development of protective immunity. *Am. J. Vet. Res.* **45**, 2104–2108 (1984).
22. Turner, J. D. *et al.* Th2 cytokines are associated with reduced worm burdens in a human intestinal helminth infection. *J. Infect. Dis.* **188**, 1768–1775 (2003).
23. Harris, N. & Gause, W. C. To B or not to B: B cells and the Th2-type immune response to helminths. *Trends Immunol.* **32**, 80–88 (2011).
24. Nogueira, D. S. *et al.* Eosinophils mediate SIgA production triggered by TLR2 and TLR4 to control *Ascaris suum* infection in mice. *PLOS Pathog.* **17**, e1010067 (2021).
25. Measure, D. *et al.* A role for eosinophils in the intestinal immunity against infective *Ascaris suum* larvae. *PLoS Negl. Trop. Dis.* **7**, e2138 (2013).
26. Measure, D. *et al.* The intestinal expulsion of the roundworm *Ascaris suum* is associated with eosinophils, intra-epithelial T cells and decreased intestinal transit time. *PLoS Negl. Trop. Dis.* **7**, e2588 (2013).
27. Coakley, G. *et al.* Immune serum-activated human macrophages coordinate with eosinophils to immobilize *Ascaris suum* larvae. *Parasite Immunol.* **42**, e12728 (2020).
28. Gazzinelli-Guimarães, A. C., Gazzinelli-Guimarães, P. & Weatherhead, J. E. A historical and systematic overview of *Ascaris* vaccine development. *Parasitology* **148**, 1795–1805 (2021).
29. Lafferty, E. I., Qureshi, S. T. & Schnare, M. The role of toll-like receptors in acute and chronic lung inflammation. *J. Inflamm.* **7**, 57 (2010).
30. Aguirre-García, M. M., Rojas-Bernabé, A., Gómez-García, A. P. & Escalona-Montaño, A. R. TLR-mediated host immune response to parasitic infectious diseases. In *Toll-like Receptors* (ed. Rezaei, N.) (IntechOpen, 2020). <https://doi.org/10.5772/intechopen.84679>.
31. Oliveira, L. M. *et al.* Genetic background affects the mucosal secretory IgA levels, parasite burden, lung inflammation, and mouse susceptibility to *Ascaris suum* infection. *Infect. Immun.* **90**, e00595–e621 (2022).
32. Vlamincx, J. *et al.* Community rates of IgG4 antibodies to *Ascaris* haemoglobin reflect changes in community egg loads following mass drug administration. *PLoS Negl. Trop. Dis.* **10**, e0004532–e0004532 (2016).
33. Tsuji, N. *et al.* Mice intranasally immunized with a recombinant 16-kilodalton antigen from roundworm *Ascaris* parasites are protected against larval migration of *Ascaris suum*. *Infect. Immun.* **71**, 5314–5323 (2003).
34. McSharry, C., Xia, Y., Holland, C. V. & Kennedy, M. W. Natural immunity to *Ascaris lumbricoides* associated with immunoglobulin E antibody to ABA-1 allergen and inflammation indicators in children. *Infect. Immun.* **67**, 484–489 (1999).
35. Vlamincx, J. *et al.* Immunizing pigs with *Ascaris suum* haemoglobin increases the inflammatory response in the liver but fails to induce a protective immunity. *Parasite Immunol.* **33**, 250–254 (2011).
36. Magalhães, L. *et al.* Immunological underpinnings of *Ascaris* infection, reinfection and co-infection and their associated comorbidities. *Parasitology* **148**, 1764–1773 (2021).
37. Weatherhead, J. E. *et al.* Host immunity and inflammation to pulmonary helminth infections. *Front. Immunol.* **11**, 594520 (2020).
38. Yantiss, R. K. Eosinophils in the GI tract: How many is too many and what do they mean?. *Mod. Pathol.* **28**, S7–S21 (2015).
39. Singh, A., Thakur, M., Sharma, L. K. & Chandra, K. Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Sci. Rep.* **10**, 16219 (2020).
40. Bibi, S. *et al.* In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Sci. Rep.* **11**, 1–16 (2021).
41. Sanches, R. C. O. *et al.* Immunoinformatics design of multi-epitope peptide-based vaccine against *Schistosoma mansoni* using transmembrane proteins as a target. *Front. Immunol.* **12**, 490 (2021).
42. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* **9**, 4409 (2019).
43. Evangelista, F. M. D., van Vliet, A. H. M., Lawton, S. P. & Betson, M. A reverse vaccinology approach identifies putative vaccination targets in the zoonotic nematode *Ascaris*. *Front. Vet. Sci.* <https://doi.org/10.3389/fvets.2022.1014198> (2022).
44. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite: A comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2–10 (2017).
45. Fleri, W. *et al.* The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front. Immunol.* **8**, 278 (2017).
46. Greenbaum, J. *et al.* Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different super-types and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* **63**, 325–335 (2011).
47. Dhanda, S. K., Gupta, S., Vir, P. & Raghava, G. P. S. Prediction of IL4 inducing peptides. *Clin. Dev. Immunol.* **2013**, 263952 (2013).
48. Nagpal, G. *et al.* Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci. Rep.* **7**, 42851 (2017).
49. Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-20: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **45**, W24–W29 (2017).
50. Kadam, K. *et al.* Antibody class(es) predictor for epitopes (AbCPE): A multi-label classification algorithm. *Front. Bioinform.* **1**, 709951 (2021).
51. Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP vol 2: A server for in silico prediction of allergens. *J. Mol. Model.* **20**, 2278 (2014).
52. Gupta, S. *et al.* Peptide toxicity prediction BT. In *Computational Peptidology* (eds Zhou, P. & Huang, J.) 143–157 (Springer, 2015). https://doi.org/10.1007/978-1-4939-2285-7_7.
53. Wu, C. H. *et al.* The protein information resource. *Nucleic Acids Res.* **31**, 345–347 (2003).
54. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
55. Shanmugam, A. *et al.* Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PLoS ONE* **7**, e30839–e30839 (2012).
56. Magnan, C. N. *et al.* High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* **26**, 2936–2943 (2010).
57. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4 (2007).
58. Dimitrov, I., Naneva, L., Doytchinova, I. & Bangov, I. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* **30**, 846–851 (2013).
59. Bui, H.-H. *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform.* **7**, 153 (2006).
60. Rapin, N., Lund, O., Bernaschi, M. & Castiglione, F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE* **5**, e9862 (2010).
61. Ebner, F. *et al.* CD4+ Th immunogenicity of the *Ascaris* spp. secreted products. *NPJ Vaccines* **5**, 25 (2020).
62. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005). <https://doi.org/10.1385/1-59259-890-0-571>.
63. Magnan, C. N., Randall, A. & Baldi, P. SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* **25**, 2200–2207 (2009).

64. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
65. Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: A web server for protein structure property prediction. *Nucleic Acids Res.* **44**, W430–W435 (2016).
66. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, W526–W531 (2005).
67. Xu, J. Distance-based protein folding powered by deep learning. *Proc. Natl. Acad. Sci.* **116**, 16856–16865 (2019).
68. Heo, L., Park, H. & Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **41**, W384–W388 (2013).
69. Hou, J., Wu, T., Cao, R. & Cheng, J. Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. *Proteins Struct. Funct. Bioinform.* **87**, 1165–1178 (2019).
70. Read, R. J., Sammito, M. D., Kryshchuk, A. & Croll, T. I. Evaluation of model refinement in CASP13. *Proteins Struct. Funct. Bioinform.* **87**, 1249–1262 (2019).
71. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–W410 (2007).
72. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
73. Colovos, C. & Yeates, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993).
74. Lüthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85 (1992).
75. Pontius, J., Richelle, J. & Wodak, S. J. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**, 121–136 (1996).
76. Barlow, D. J., Edwards, M. S. & Thornton, J. M. Continuous and discontinuous protein antigenic determinants. *Nature* **322**, 747–748 (1986).
77. Ponomarenko, J. *et al.* ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* **9**, 514 (2008).
78. Kozakov, D. *et al.* The ClusPro web server for protein–protein docking. *Nat. Protoc.* **12**, 255–278 (2017).
79. Su, L. *et al.* Structural basis of TLR2/TLR1 activation by the synthetic agonist diprovocim. *J. Med. Chem.* **62**, 2938–2949 (2019).
80. Ohto, U., Yamakawa, N., Akashi-Takamura, S., Miyake, K. & Shimizu, T. Structural analyses of human toll-like receptor 4 polymorphisms D299G and T399I*. *J. Biol. Chem.* **287**, 40611–40617 (2012).
81. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: A web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).
82. Laskowski, R. A. & Swindells, M. B. LigPlot+: Multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).
83. López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S. & Chacón, P. iMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Res.* **42**, W271–W276 (2014).
84. Wang, J. *et al.* Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* **27**, 2001–2014 (2017).
85. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
86. Ikai, A. Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895–1898 (1980).
87. Noon, J. B. & Aroian, R. V. Recombinant subunit vaccines for soil-transmitted helminths. *Parasitology* **144**, 1845–1870 (2017).
88. Kaur, R. *et al.* Immunoinformatics driven construction of multi-epitope vaccine candidate against *Ascaris lumbricoides* using its entire immunogenic epitopes. *Expert Rev. Vaccines* **1**, 1–13. <https://doi.org/10.1080/14760584.2021.1974298> (2021).
89. Sadaow, L. *et al.* Molecular identification of *Ascaris lumbricoides* and *Ascaris suum* recovered from humans and pigs in Thailand, Lao PDR, and Myanmar. *Parasitol. Res.* **117**, 2427–2436 (2018).
90. Katakam, K. K., Thamsborg, S. M., Dalsgaard, A., Kyvsgaard, N. C. & Mejer, H. Environmental contamination and transmission of *Ascaris suum* in Danish organic pig farms. *Parasit. Vectors* **9**, 80 (2016).
91. Garamszegi, L. Z. & Nunn, C. L. Parasite-mediated evolution of the functional part of the MHC in primates. *J. Evol. Biol.* **24**, 184–195 (2011).
92. Zaini, A., Good-Jacobson, K. L. & Zaph, C. Context-dependent roles of B cells during intestinal helminth infection. *PLoS Negl. Trop. Dis.* **15**, e0009340 (2021).
93. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng. Des. Sel.* **14**, 529–532 (2001).
94. Coler, R. N. *et al.* A synthetic adjuvant to enhance and expand immune responses to influenza vaccines. *PLoS ONE* **5**, e13677 (2010).
95. Rehman, A. *et al.* Integrated core proteomics, subtractive proteomics, and immunoinformatics investigation to unveil a potential multi-epitope vaccine against schistosomiasis. *Vaccines* **9**, 658 (2021).
96. Bhatwa, A. *et al.* Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Front. Bioeng. Biotechnol.* **9**, 65 (2021).
97. Corradin, G., Villard, V. & Kajava, A. V. Protein structure based strategies for antigen discovery and vaccine development against malaria and other pathogens. *Endocr. Metab. Immune Disord. Drug Targets* **7**, 259–265 (2007).
98. Arumugam, S. & Varamballi, P. In-silico design of envelope based multi-epitope vaccine candidate against Kyasanur forest disease virus. *Sci. Rep.* **11**, 17118 (2021).

Author contributions

Concept and design, data interpretation, and critical review and editing of the manuscript: F.E., A.V., S.L., and M.B. Data acquisition and analysis and manuscript draft: F.E. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the Doctoral College Studentship Award (DCSA3) from the University of Surrey awarded to Francisco Miguel Dias Evangelista.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-30445-x>.

Correspondence and requests for materials should be addressed to M.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023