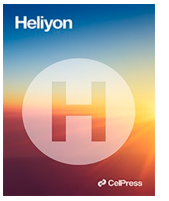
Heliyon 10 (2024) e37536

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

In silico design of *Mycobacterium tuberculosis* multi-epitope adhesin protein vaccines

Koobashnee Pillay [a], Thamsanqa E. Chiliza [b], Sibusiso Senzani [a], Balakrishna Pillay [b], Manormoney Pillay [a,*]

[a] *Discipline of Medical Microbiology, School of Laboratory Medicine and Medical Sciences, College of Health Science, University of KwaZulu-Natal, South Africa*
[b] *Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, South Africa*

ARTICLE INFO

Keywords:
Tuberculosis
Mycobacterium tuberculosis
Multi-epitope vaccine
Adhesin proteins
Immunoinformatics

ABSTRACT

Mycobacterium tuberculosis (Mtb) adhesin proteins are promising candidates for subunit vaccine design. Multi-epitope Mtb vaccine and diagnostic candidates were designed using immunoinformatic tools. The antigenic potential of 26 adhesin proteins were determined using VaxiJen 2.0. The truncated heat shock protein 70 (tnHSP70), 19 kDa antigen lipoprotein (lpqH), Mtb curli pili (MTP), and Phosphate transport protein S1 (PstS1) were selected based on the number of known epitopes on the Immune Epitope Database (IEDB). B- and T-cell epitopes were identified using BepiPred2.0, ABCpred, SVMTriP, and IEDB, respectively. Population coverage was analysed using prominent South African specific alleles on the IEDB. The allergenicity, physicochemical characteristics and tertiary structure of the tri-fusion proteins were determined. The in silico immune simulation was performed using C-ImmSim. Three truncated sequences, with predicted B and T cell epitopes, and without allergenicity or signal peptides were linked by three glycine-serine residues, resulting in the stable, hydrophilic molecules, tnlpqH-tnPstS1-tnHSP70 (64,86 kDa) and tnMTP-tnPstS1-tnHSP70 (63,96 kDa). Restriction endonuclease recognition sequences incorporated at the N- and C-terminal ends of each construct, facilitated virtual cloning using Snapgene, into pGEX6P-1, resulting in novel, highly immunogenic vaccine candidates (0,912–0,985). Future studies will involve the cloning, recombinant protein expression and purification of these constructs for downstream applications.

1. Introduction

Despite the past coronavirus pandemic, tuberculosis (TB) remains a global epidemic [1]. An estimated 10,6 million people were infected with TB in the year 2021 whilst a further ±1,5 million people lost their lives to this dreaded disease [2]. Africa accounts for 23 % of the global TB infections, with South Africa listed as a high-burden TB country. Despite the availability of an array of first- and second-line anti-TB drugs, the escalating drug resistance and the lack of rapid, accurate point of care diagnostic tests and effective vaccines have resulted in high disease burden [3,4].

Although TB has been present for centuries, only one vaccine, the Bacillus-Calmette Guerin (BCG) vaccine, which was initially

* Corresponding author. Discipline of Medical Microbiology School of Laboratory Medicine and Medical Sciences College of Health Science University of KwaZulu-Natal Nelson R. Mandela School of Medicine Campus, 719 Umbilo Road Congella, Durban, South Africa.
E-mail address: PILLAYC@ukzn.ac.za (M. Pillay).

https://doi.org/10.1016/j.heliyon.2024.e37536
Received 31 July 2024; Accepted 4 September 2024
Available online 7 September 2024

administered in 1921, remains approved for the prevention of this disease [5]. The BCG is an attenuated derivative of *Mycobacterium bovis*, in which the RD1 region was deleted [6]. The use of BCG in developing countries with a high prevalence of TB is strongly encouraged, despite its short-lived protection. BCG provides protection against adolescent, but not adult, pulmonary TB (pTB), as exposure to environmental mycobacteria renders the vaccine ineffective in the latter. Since inception, six strains of the BCG vaccine with phenotypic variations have been produced, that could potentially account for the variable protection observed in some population groups [7].

The temporary protection offered by the BCG and the potential risk of it becoming virulent, has created the need for a new and improved TB vaccine candidate that is effective on all age groups and populations [8]. Three approaches can be taken during vaccine development [9–12]: 1) an improvement of the current version of BCG vaccine which is safe and long-lasting; 2) a prime-boost strategy wherein the viral vectored or protein adjuvant is administered at a later stage as a booster dose; 3) immunotherapeutic vaccines which reduce the duration of TB therapy. The WHO prioritized the new vaccine candidates based on two strategies: 1) safe, effective, and affordable TB vaccine for adolescents and adults; 2) a vaccine for neonates and infants with improved safety and efficacy in comparison to the BCG [13]. Presently, the vaccine pipeline has approximately 29 new TB vaccine candidates in various phases of trials. Only six of these are in phase III, including three live attenuated mycobacterial vaccines: BCG travel vaccine MTBVAC and the VPM1002; one protein adjuvant vaccine: GamTBvac; and one inactivated mycobacterial vaccine: Immunvac (MIP) [14–16]. These trials attempt to measure either the prevention of infection (POI) against Mtb, or prevention of acquiring TB disease (POD), or prevention of recurrent (POR) TB disease. Vaccine candidates can be divided into whole cell vaccines and subunit vaccines [11]. Whole cell vaccines include live mycobacterial vaccines derived from live attenuated Mtb strains, *M. bovis* BCG or recombinant BCG and killed mycobacterial vaccines that could be formulated from other saprophytic mycobacterial species or Mtb [17]. Subunit vaccines contain Mtb antigens expressed as recombinant proteins formulated with different adjuvants [18]. These proteins are expressed by recombinant vectors that are used as vehicles for the administration of antigens. There are currently 9 subunit vaccine candidates with only one (GamTBvac) in the final phase.

Mycobacterium tuberculosis has distinct phases of growth which may be associated with active mycobacterial replication, persistence and dormancy [19]. Biomarkers are measurable indicators of the state of disease progression. Adhesin proteins (Table 1) display biomarker potential as they may be associated with active bacterial replication [20–22]. Subunit vaccines comprising Mtb proteins are lucrative candidates for vaccine design [23,24]. Adhesin proteins such as those belonging to the ESX-1 secretion system (Esx) and antigen 85 (Ag85) family, culture filtrate protein 10 (CFP-10) and heparin-binding hemagglutinin adhesin (HBHA) are highly immunogenic and have previously displayed protection in animal models [25,26]. Protection against TB requires a cell-mediated immune response which involves multiple components such as cluster of differentiation 4 and 8 ($CD4^+$ and $CD8^+$) T cells [27]. A multi-stage TB vaccine which comprises a fusion of adhesin proteins which are known to elicit an immune response during the various stages of TB is crucial to the control of the disease. Therefore, the selection of the adhesin protein within the fusion construct is critical. Adhesin proteins involved in the generation of an immune response during the early stages of TB infection are pivotal for the prevention and establishment of disease and should therefore be highly favoured in the design of a fusion construct [28].

2. Methodology

2.1. *Mycobacterium tuberculosis* adhesin protein sequences

A list of 26 Mtb adhesin proteins (Table 1) were identified in a review by Ramsugit et al. (2016) [22]. All information pertaining to the adhesin proteins were obtained from the Tuberculist GenoList Pasteur Database (http://genolist.pasteur.fr/TubercuList/). This

Table 1
Potential *Mycobacterium tuberculosis* adhesins proteins analysed for incorporation into the fusion construct [22].

Adhesin	Rv Number	Host Receptor	Reference
Alanine proline-rich antigen (Apa)	Rv1860	Macrophages	[29]
19-kDa Antigen (lpqH)	Rv3763	Monocytes and macrophages	[30]
Antigen 85 complex (Ag85)	Rv0129c, Rv1886c, Rv3803c, and Rv3804c	Fibronectin and elastin	[31]
Cpn60.2 molecular chaperone (Cpn60.2)	Rv0440	CD43 on Macrophages	[32]
Curli pili (MTP)	Rv3312A	Laminin, macrophages and epithelial cells	[33]
Early secreted antigen (ESAT-6)	Rv3874/Rv3875	Laminin	[34]
Glutamine synthetase A1	Rv2220	Fibronectin	[35]
Glyceraldehyde-3phosphate dehydrogenase (GAPDH)	Rv1436	Fibronectin	[36]
Heparin-binding hemagglutinin adhesin (HBHA)	Rv0475	Epithelial cells	[37]
L,D-transpeptidase	Rv0309	Fibronectin and laminin	[38]
Malate synthase (MS)	Rv1837c	Fibronectin, laminin, and epithelial cells	[39]
Membrane protein	Rv2599	Collagen, fibronectin and laminin	[38]
Molecular chaperone DnaK (HSP70)	Rv0350	Macrophages	[32]
Mycobacterium cell entry-1 protein	Rv0169	Epithelial cells	[40]
N-acetylmuromoyl-L-alanine amidase	Rv3717	Fibronectin and laminin	[38]
PE-PGRS proteins	Rv1759c, Rv1818c	Fibronectin and macrophages	[41]
PstS-1 (38-kDa antigen)	Rv0934	Macrophages	[30]
Type IV pili	Rv3654c-Rv3660c	Macrophage and epithelial cells	[42]

includes protein and DNA sequence, gene and protein size, location, Rv number and function.

2.2. Determination of vaccine potential of adhesin proteins

The protein sequences of each of the adhesins were input onto the VaxiJen 2.0 online tool (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [43]. This tool provides the antigenic potential of each sequence which is identified as a probable antigen if the bacterial threshold of 0.4 is exceeded. This tool was used to rank the 26 Mtb adhesin proteins.

2.3. Identification of the number of epitopes on adhesin proteins

The top ten ranked adhesin proteins mostly likely to generate an immune response were analysed on the Immune Epitope Database (IEDB) (https://www.iedb.org), an online tool used to enumerate and identify the epitope regions present on the protein [44]. The following filters were applied: organism: *Mycobacterium tuberculosis* (ID:1773); Antigen name; include Positive assays; Host: Homo sapiens (human); Disease data: infectious disease.

A literature search was conducted on adhesin proteins which displayed multiple epitope regions as identified by the IEDB, to aid in the selection of the proteins for the fusion molecule. Adhesins which induced cytokine and antibody responses detected in patient sera, were selected for further analysis. These included *Mycobacterium tuberculosis* curli pili (MTP); Phosphate-specific transporter protein PstS 1 (PstS1); 19 kDa Lipoprotein (lpqH) and Heat-shock protein 70 (HSP70). For the purposes of this study, the truncated HSP70 protein (tnHSP70), amino acids 359–610, was used for further analysis owing to its adjuvant properties.

2.4. Prediction of B-cell epitopes on protein sequences

B-cells are involved in the neutralization of pathogenic molecules using specific secreted or surfaced expressed receptors. Several online tools were explored to identify B-cell epitopes within the protein sequences of the selected adhesins. These include BepiPred2.0 (https://services.healthtech.dtu.dk/service.php?BepiPred-2.0), ABCpred server (https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html) and SVMTriP (http://sysbio.unl.edu/SVMTriP/prediction.php). BepiPred2.0 uses a random forest algorithm trained on new epitopes annotated from antibody-antigen protein structures [45]. The ABCpred server is based on recurrent machine-based techniques using fixed length patterns to predict B-cell epitopes on an antigenic sequence [46]. The SVMTriP tool employs super vector machine (SVM) in combination with tri-peptide similarity and propensity scores (SVMTriP) to better predict linear epitope regions [47].

2.5. Prediction of signal peptides within protein sequence

The SignalP 6.0 server (https://services.healthtech.dtu.dk/service.php?SignalP) was used to identify the presence of signal peptides on the protein sequences [48]. The server is able to predict the five types of signal peptides and their cleavage locations.

2.6. Prediction of T-cell epitopes

The Mtb adhesin proteins, tnHSP70, lpqH, MTP and PstS1 protein sequences lacking the signal peptide sequence were subjected to the MHCI and MHCII epitope prediction using the IEDB (http://tools.iedb.org/mhci/and http://tools.iedb.org/mhcii/) [49]. For the prediction of the MHC I binding sites, 54 HLA-A, HLA-B and HLA-C reference alleles were selected [50]. For the prediction of the MHC II binding sites, the reference set of 7 HLA-DR alleles (HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*03:01) with a default length of 15-mer were selected [51,52].

2.7. Population coverage for MCH class I and II alleles

The population coverage for MHC I and MHC II Mtb adhesin protein epitopes were determined using the IEDB analysis resource tool (http://tools.iedb.org/population/) [49]. A combination of MHCI and MHCII alleles which are prominent to the South African population were selected [53]. These include HLA-A*30:01, HLA-A*02:01, HLA-A*30:01, HLA-B*42:01, HLA-B*58:01, HLA-B*58:02, HLA-B*15:01, HLA-B*58:01, HLA-B*58:02, HLA-C*06:02, HLA-DPB1*13:01, HLA-DQB1*06:01, HLA-DRB1*11:01, HLA-DRB1*13:01, HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-C*07:01, HLA-C*07:02, HLA-DRB1*03:01.

2.8. Prediction of allergenicity of protein sequences

The potential of the selected proteins to cause an allergic response was determined using AlgPred (https://webs.iiitd.edu.in/raghava/algpred/submission.html) [54] and AllerTop Pro 2.0 https://www.ddg-pharmfac.net/AllerTOP/feedback.py) [55].

2.9. Prediction of secondary and tertiary structure, and physicochemical characteristics of fusion proteins

The ProtParam server was used to predict the physicochemical characteristics of the tri-fusion proteins (https://web.expasy.org/cgi-bin/protparam/protparam) [56]. The parameters evaluated included the number of amino acids, molecular weight, theoretical

isoelectric point (pI), extinction coefficient, estimated half-life *in vitro* and *in vivo*, instability index, aliphatic index and Grand average of hydropathicity (GRAVY).

The D-I-TASSER (Deep learning-based Iterative Threading ASSEmbly Refinement) online tool (https://zhanggroup.org/D-I-TASSER/), an extension of the I-TASSER, was used to predict the tertiary structure and function of the tri-fusion protein based on the submission of the amino acid sequence whilst utilizing deep convolutional neural-network based distance-map predictions guiding the fragment assembly simulations of the I-TASSER [57,58]. In addition, the server provides an estimated accuracy of the predicted models including TM-score. The tool also predicts the secondary structure, solvent accessibility, enzyme classification and ligand-binding sites.

2.10. Refinement and validation of predicted tertiary structures

The first model of each trifusion protein construct predicted by the D-I-TASSER server was subjected to refinement using the web-based Galaxy Refine server (https://galaxy.seoklab.org) [59]. The first refined model generated was validated on the SAVESv6.0 server (https://saves.mbi.ucla.edu/) using the PROCHECK, ERRAT and Ramachandran verification tools.

2.11. In silico immune simulation

To characterize the immune response induced by the tri-fusion protein vaccine candidates, in silico immune simulation was performed using the C-ImmSim server (https://kraken.iac.rm.cnr.it/C-IMMSIM/) [60]. The server predicts immune interactions using position-specific scoring matrices derived from machine learning techniques for peptide prediction. It concurrently simulates three compartments representing three separate anatomical regions found in mammals: (i) the bone marrow, where hematopoietic stem cells were simulated to produce new lymphocytes and myeloid cells; (ii) the thymus, where naive T cells were selected to avoid autoimmunity; and (iii) the lymphatic organ such as lymph nodes. Generally, 2–3 doses within four week interval is recommended for most vaccines. Hence, three doses of 1000 vaccine units without lipopolysaccharide (LPS) were tested over four weeks in the immune simulation of these candidates. The parameters were as follows: three injections at time-step 1, 84, and 168, respectively; random seed: 12345; simulation volume of 10 and 1000 simulation steps [61].

2.12. In silico cloning of the tri-fusion proteins

Sequences were designed to incorporate BamH1 and Xho1 and EcoR1 and Xho1 restriction endonuclease recognition sequences at the N- and C-terminal ends of tri-fusion constructs tnlpqH-tnPstS1-tnHSP70 and tnMTP-tnPstS1-tnHSP70, respectively. This enables

Table 2
Ranking of adhesin protein according to vaccine potential using VaxiJen and their corresponding number of epitopes listed on The Immune Epitope Database (IEDB)

Rank	Adhesin Protein	Accession No.	Probability of Antigen (VaxiJen)	No. of epitopes (IEDB)
1	Wag22_ PE_PGRS	Rv1759c	2,0053	No data
2	PE-PGRS33	Rv1818c	1,6972	No data
3	19 kDa Antigen	Rv3763	1,15	14
4	mtp	Rv3312A	0,7609	No data
5	PstS-1	Rv0934	0,7312	48
6	DnaK/Hsp70	Rv0350	0,6417	69
7	N-acetylmuromoyl-L-alanine amidase	Rv3717	0,6327	No data
8	HupB	Rv2986c	0,6089	No data
9	HBHA	Rv0475	0,5986	3
10	Type IV pili	Rv3655c	0,5914	No data
11	Membrane protein	Rv2599	0,5883	–
12	Ag85B	Rv1886c	0,5842	–
13	Type IV pili	Rv3654c	0,5748	–
14	MCE1A	Rv0169	0,5666	–
15	Protein Kinase D	Rv0931c	0,5634	–
16	ESAT-6	Rv3875	0,5577	–
17	Cpn60,2	Rv0440	0,5421	–
18	Ag85A	Rv3804c	0,5259	–
19	Apa	Rv1860	0,5233	–
20	GlnA1	Rv2220	0,5087	–
21	L,D Transpeptidase	Rv0309	0,5044	–
22	Ag85C	Rv0129c	0,4795	–
23	MPT51/Ag85C	Rv3803c	0,4515	–
24	GAPDH	Rv1436	0,4123	–
25	Type IV pili	Rv3656c	0,3738	–
26	Malate Synthase	Rv1837c	0,3676	–

N.B: (-)Denotes that this adhesin protein was not investigated for this study and therefore data (if available) was not obtained.

cloning into the pGEX6P-1 vector via the BamH1 and Xho1 and EcoR1 and Xho1 restriction sites using the Snapgene software tool.

3. Results

3.1. Determination of vaccine potential and the identification of number of epitopes on adhesin proteins

The known Mtb adhesin proteins were ranked according to the vaccine potential using the VaxiJen online tool (Table 2). The VaxiJen tool identified any sequence that exceeds the threshold of 0.4 as a probable antigen. Of the 26 adhesin proteins (0,3676–2,0053), the top ten (0,5914–2,0053) were selected for further evaluation. The IEDB was used to identify the presence of epitope regions within the top ten adhesin protein sequences. Only four adhesin proteins: 19 kDa antigen lipoprotein (lpqH), Phosphate transport protein S 1 (PstS1), Heat-shock protein 70 (dnaK/HSP70) and the heparin-binding hemagglutinin (HBHA) adhesin protein, returned results with identified epitope regions. The lpqH, PstS1 and HSP70 adhesin proteins were selected for further analysis. The MTP adhesin protein which did not display any identified epitopes on the IEDB, but was selected since previous studies [42,62,63] on this protein displayed promising results as a diagnostic biomarker and was detected in patient serum samples indicating that it was capable of inducing an immune response. The C-terminal region of the HSP70 possesses natural adjuvant properties [64–67]. In addition, heat shock proteins such as HSP70 act as molecular chaperones aiding in the protein folding process which could prove beneficial when fused to other proteins [68]. Therefore, the tnHSP70 was utilized in the construction of the tri-fusion proteins in this study.

3.2. Prediction of B-cell epitopes on protein sequences

The sequences of the four selected adhesin proteins; tnHSP70, lpqH, MTP and PstS1 were further investigated for the presence of other B-cell epitope regions using online immunoinformatic tools. Fig. S1 displays the epitope regions identified for each of the adhesin proteins using the BepiPred 2.0 Tool. Any amino acid sequence exceeding the threshold of 0.5 was considered an epitope. The structural predictions of these epitope regions including their accessibility are noted based on the intensity of the gradients. Most of the epitopes on these proteins can be located on the coil and are exposed. The tnHSP70 adhesin protein contains three regions of continuous B-cell epitopes which are predominantly located on the coil region and are mostly buried. The lpqH protein contains five regions of continuous B-cell epitopes which are predominantly exposed on the coil, whilst few epitopes are located on the sheet and are buried. The MTP protein contains three regions of continuous B-cell epitopes, all of which are located on the coil and are either exposed or buried. Approximately 10 continuous B-cell epitope regions are located on the PstS1 protein with the majority being located on the coil where they are exposed, whilst other epitopes can be found on both the helix and sheet and are either buried or exposed.

The adhesin proteins were also subjected to other immunoinformatic tools such as the SvM TRiP and ABCpred to further identify B-cell epitopes which are displayed in Table 3. In addition to new epitope sequences, common regions were located among each of the sequences with the different tools used. The epitope regions identified by these tools were selected for the design of the tri-fusion protein constructs for use as a multi-epitope vaccine candidate and diagnostic biomarker. A continuous sequence encompassing all the epitopes was selected for ease of cloning rather than random epitope regions (Table S1). Two tri-fusion protein constructs were designed to contain truncated (tn) regions of each protein; tnlpqH-tnPstS1-tnHSP70 (tnLPH) and tnMTP-tnPstS1-tnHSP70 (tnMPH) (Fig. 1).

3.3. Prediction of signal peptides within protein sequence

The presence of signal peptides on proteins presents issues with downstream applications especially when attempting to clone the gene into another host [69]. The SignalP 6.0 online tool was used to identify the presence of these peptide regions within the selected adhesin proteins. All three proteins contained a signal peptide (Fig. 2). The tnHSP70 protein did not contain any signal peptide sequences (Fig. 2D). Both the lpqH and PstS1 adhesin proteins contain a lipoprotein signal peptide (Sec/SPII) with a probability of 1. The cleavage sites can be located between positions 21 and 22 of the lpqH adhesin protein (Fig. 3A) and at positions 23 and 24 on the PstS1 adhesin protein (Fig. 2C). The MTP protein contains a signal peptide (Sec/SPI) with a probability of 0.9979 with a cleavage site located between positions 31 and 32 (Fig. 2B). For the design of the tri-fusion protein construct, these sequences were excluded and their truncated sequences were used for further analysis (tnlpqH, tnMTP and tnPstS1).

3.4. Prediction of T-cell epitopes within protein sequence

T cells are able to recognize epitopes that are presented by the major histocompatibility molecules (MHC) which are either expressed on the surfaces of nucleated cells or on the surfaces of specialized antigen-presenting cells (APCs) [70]. The IEDB tool was used to determine the T cell epitopes present on each of the adhesin protein sequences. Table 4 displays the epitopes sequences binding to either MHC I or MHC II. These epitope regions were included in the trifusion protein construct sequences displayed on Table S1.

3.5. Population coverage for MCH class I and II alleles

The IEDB population coverage tool was used to investigate the population coverage of a combination of classes of the trifusion protein constructs using the alleles prominent to the South African population. This predicted a population of 91,81 % for both tnLPH

Table 3
Prediction of B-cell epitopes on the lpqH, MTP and PstS1 adhesin proteins using various immunoinformatic tools.

B Cell Epitope Prediction Tool	Predicted Epitope Sequence for each Protein and Score (if available)			
	tnHSP70	lpqH	MTP	PstS1
BepiPred	EVKDVLLLDVTPLSLGIETKGGVMTRLIERN TTIPTKRSETFTTADDNQPSVQIQVYQGEREI AAHNKLLGSF QIEVTFDIDANGIVHVTAKDKGTGKENTIRIQE GSGLSKEDIDRMIKDAEAHAEEDRKRREEA DVRNQAETLVYQ KEQREAEGGSKVPEDTLNKVDAAVAEAKA ALGGSDISAIKSAMEKLGQESQALGQAIYEAA QAASQATGAAHPGGEPGGAHP	SSNKSTTGSGETTTAAGTTASPGAASG GKDQNVT DGNPPEVKSVGLGNVN YTSGTGQGNASATK GVDMANPMSPVNKS	AAQTAPVPDY AWGPNWDPYTCHDDFHRDSDGPDHSRDYPGP GPVLDDPGAAPPPPAAG	GASDAYLSEGDMAAHK MNIALAISAQQVNYNLPGVSEHLKLNGKVLAA IKTWDDPQIAALNPGVNL SKQDPEGWGKSPGFGTTVDFPAVPGALGENGNGG QASQRGL SSGNFLLP FASKTPANQAISMIDGPAPD NRQKDAATAQ SFLD HFQP
ABCpred	EAHAEEDRKRREEADV (0.91) RGIPQIEVTFDIDANG (0.90) TKRSETFTTADDNQPS (0.86) AHPGGEPGGAHPGSAD(0.86) KGGVMTRLIERNTTIP (0.86) DRMIKDAEAHAEEDRK (0.86) GLSKEDIDRMIKDAE (0.83) KVPEDTLNKVDAAVAE (0.82)	VTGSVVCTTAAGNVNAIGGA (0.83) TTASPGAASGPKVVIDGKDQ (0.82) TGIAAVLTDGNPPEVKSVGL (0.81) GSGETTTAAGTTASPGAASG (0.80) KVVIDGKDQNVTGSVVCTTA(0.80) TLGYTSGTGQGNASATKDGS (0.79) TGQGNASATKDGSHYKITGT (0,76) AAGNVNAIGGAATGIAAVLT (0,74) GNPPEVKSVGLGNVNGVTLG (0.55)	YYWCPGQPFDPAWGP (0.94) QSAAQTAPVPDYYWCP (0.93) HRDSDGPDHSRDYPGP (0.91) GPILEGPVLDDPGAAP (0.87) AACILATGVAGLGVGA (0.85) GPNWDPYTCHDDFHRD (0.83) DDPGAAPPPPAAGGGA (0.82) PDHSRDYPGPILEGPV (0.82) TGVAGLGVGAQSAAQT (0.73)	DFPAVPGALGENGNGGMVTG (0.86) LHRSDGSGDTFLFTQYLSKQ (0.85) LSEGDMAAHKGLMNIALAIS (0.82) IINYEYAIVNNRQKDAATAQ (0.81) LGNSSGNFLLPDAQSIQAAA (0.81) YIGISFLDQASQRGLGEAQL (0.81) NGNGGMVTGCAETPGCVAYI (0.81) EHLKLNGKVLAAMYQGTIKT (0.80) GCAETPGCVAYIGISFLDQA (0.79)
ABCpred	FELTGIPPAPRGIPQI (0.80) EAEGGSKVPEDTLNKV (0.80) IERNTTIPTKRSETFT (0.79) GQAIYEAAQAASQATG (0.79) GQESQALGQAIYEAAQ (0.78) SLGIETKGGVMTRLIE (0.78) TFDIDANGIVHVTAKD (0.77) HVTAKDKGTGKENTIR (0.77) QAASQATGAAHPGGEP (0.75) GSDISAIKSAMEKLGQ (0.74) RKRREEADVRNQAETL (0.71) SVQIQVYQGEREIAAH (0.69) SAMEKLGQESQALGQA (0.68) FTTADDNQPSVQIQVY (0.67) ETLVYQTEKFVKEQRE (0.65) YQGEREIAAHNKLLGS (0.63)		RFACRTLMLAACILAT (0.54)	QIAALNPGVNLPGTAVVPLH (0.78) NYNLPGVSEHLKLNGKVLAA (0.78) LHWAITDGNKASFLDQVHFQ (0.78) AGFASKTPANQAISMIDGPA (0.77) LPGTAVVPLHRSDGSGDTFL (0.76) GTIKTWDDPQIAALNPGVNL (0.76) FLFTQYLSKQDPEGWGKSPG (0.74) VNNRQKDAATAQTLQAFLHW (0.72) PDAQSIQAAAAGFASKTPAN (0.72) QASQRGLGEAQLGNSSGNFL (0.72) DGPAPDGYPIINYEYAIVNN (0.68) LMNIALAISAQQVNYNLPGV (0.67) QDPEGWGKSPGFGTTVDFPA (0.66) KVLAAMYQGTIKTWDDPQIA (0.61)
SVM TRiP	HAEEDRKRREEADVRNQAET (1.000) DISAIKSAMEKLGQESQALG (0.563) IAAHNKLLGSFELTGIPPAP (0.435) AEGGSKVPEDTLNKVDAAVA (0.353) DIDANGIVHVTAKDKGTGKE (0.353) QAIYEAAQAASQATGAAHPG (0.203)	SGETTTAAGTTASPGAASGP (1.00) DGKDQNVTGSVVCTTAAGNV (0.998) DGNPPEVKSVGLGNVNGVTL (0.997)		DMAAHKGLMNIALAISAQQV (1.000) PLHRSDGSGDTFLFTQYLSK (0.998) SQRGLGEAQLGNSSGNFLLP (0.990) LPGVSEHLKLNGKVLAAMYQ (0.351) QAAAAGFASKTPANQAISMI (0.294)

and tnMPH with an average number of epitope hits of 1,98. Using the same alleles, a world population coverage was also performed for a combination of both the classes. Both tnLPH and tnMPH displayed a coverage of 90,95 % with an average number of epitope hits of 1,97.

3.6. Prediction of allergenicity of protein sequences

Two online tools were used to determine if the truncated epitope regions of the selected adhesin proteins could induce an allergic response in the host. Allergenicity is the ability of the antigens to induce an abnormal immune response. The allergenicity results of the truncated proteins are listed in Table S2. All the truncated regions of the adhesin protein and the tri-fusion constructs, tnLPH and tnMPH, do not contain any experimental Immunoglobulin E (IgE) epitopes. The IgE is an allergen-specific antibody. The Motif Alignment and Search Tool (MAST) tool searches biological sequence databases for the sequences that contain one or more of a group of known motifs [54]. According to the output, all proteins are non-allergens since there were no matches to any motifs. The Support Vector Machine (SVM) can detect allergens based on either the amino acid sequence or the dipeptide composition of the protein. Based on amino acid composition, both the tri-fusion protein constructs and the tnlpqH, tnPstS1 proteins were identified as probable allergens whilst the tnMTP and tnHSP70 were identified as an allergen and potential allergen, respectively. Using the dipeptide composition, tnlpqH, tnPstS1 and tnLPH were identified as potential allergens, whereas the tnMTP was identified as a non-allergen and the tnHSP70 and tnMPH tri-fusion construct were identified as allergens. The BLAST results indicated that all individual truncated proteins were non-allergens, whereas a sequence present in the tri-fusion protein constructs was identified as an allergen. According to the hybrid approach, which considers the results from each of the preceding methods, all proteins were non-allergens. The AllerTop 2.0 online tool which is based on the amino acid composition of the protein and its structure-activity relationships (QSAR), identified all proteins as probable non-allergens, except tnMTP which was identified as a probable allergen.

3.7. Prediction of secondary and tertiary structure of fusion protein and its physicochemical characteristics

The physicochemical characteristics such as the amino acid composition, theoretical pI, protein half-life, aliphatic and instability index and the grand average of hydropathicity (GRAVY) of each of the tri-fusion proteins, determined by the ProtParam ExPasy tool, are displayed in Table 5. The instability index of both protein constructs was below 40 whilst the GRAVY index of the tri-fusion protein was calculated to be −0.29 and-0.406 for tnLPH and tnMPH, respectively.

The tertiary structure of tnLPH and tnMPH was predicted by the D-I-TASSER server and 5 models with TM-scores were obtained. A template modelling score (TM-score) is between 0 and 1 and is proportional to the relation of the prediction. Fig. 3A and B displays the first model of each of the trifusion proteins which displayed the highest TM-score. Model 1 of tnLPH had a TM-score of 0.65 (0.48–0.65) and the first model of tnMPH had a TM-score 0.51 (0.49–0.51).

3.8. Tertiary structure refinement and validation

The first model of the D-I-TASSER prediction was subjected to refinement and 5 models were generated by the Galaxy Refine server with the first model being highly favoured (Fig. 3C & D). Each model displays scores for the global distance test (GDT-HA), root mean square deviation (RMSD), MolProbity, Clash Score, Poor rotamers and Rama favoured score (Figs. S2A and B) . In comparison to the initial model of the tnLPH, the selected refined model displayed a reduced MolProbity (from 3.262 to 2.424) and Clash score (100.2–22.0) with an increased Rama favourability (74.8–91.5) and presence of poor rotamers (0–1.3). The selected refined tnMPH model displayed a reduced MolProbity (3.068–2.135) and Clash score (65.9–15.8) and an increased Rama favourability (76.4–93.3) and presence of poor rotamers (0–0.7). Both models displayed a GDT-HA (0.9367–0.9404) and RMSD (0.461–0.498) score indicating that the refined models closely resembled the native trifusion models.

The refined models were then subjected to validation on the SAVESv6.0 server. The ERRAT data suggests good overall quality of the refined models. The Ramachandran plot, generated by the PROCHECK programme, graphically displays the energetically allowed and forbidden regions for the dihedral angles. The tnLPH has 87,7% of residues in the favoured regions; 10% in the additional allowed regions, 0.8% in the generously allowed regions; and 1.4% in the disallowed regions. The tnMPH has 89.1% of residues in the favoured regions; 8.5 % in the additional allowed regions, 1 % in the generously allowed regions; and 1.4% in the disallowed regions.

3.9. In silico immune simulation

The in silico immune simulation indicated that both tri-fusion protein candidates were able to elicit an immune response. This was characterized by the heightened immunoglobulin M (IgM) and immunoglobulin G (IgG) concentration in response to the antigen when injected at all three timepoints (Fig. 4A and B). Furthermore, the antigen was depleted whilst the IgG concentration was maintained over a prolonged period, indicating that the immune system was able to eradicate the antigen from the system. In addition, there was an increase in natural killer (NK) cells, dendritic cell (DC), macrophages (MA) and epithelial cells (EP) (Figs. S3 and S4: I, J, K and L). Three peaks corresponding to the timepoint of injections were observed in B- cell population and T-helper (TH) population (Figs. S3 and S4:A and D). The cytokine profile induced by the immunization of tnLPH and tnMPH were analysed (Fig. 4C and D). Both candidates generated high levels of interferon gamma (IFN-g), transforming growth factor beta (TGF-b), interleukin-10 (IL-10) and interleukin-12 (IL-12) where three peaks were observed. The D levels remained low indicating that these candidates did not pose a danger to the immune system [60].

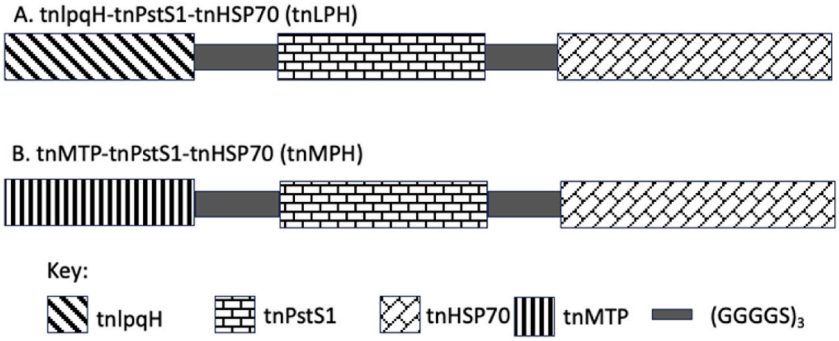


Fig. 1. Schematic representation of the trifusion protein constructs using immunodominant Mtb adhesin proteins. (A) tnLPH. (B) tnMPH.

3.10. In silico cloning of the tri-fusion proteins

Fig. 1 displays the order in which the trifusion protein constructs will be arranged. The DNA sequences of each of the selected proteins are highlighted in Table S1. Restriction endonucleases were included onto the N- and C-terminal ends of tri-fusion constructs tnlpqH-tnPstS1-tnHSP70 and tnMTP-tnPstS1-tnHSP70, respectively. Fig. 5 displays the recombinant vector map of the trifusion protein constructs, tnLPH and tnMPH, respectively. The Snapgene tool was further used to confirm that the incorporation of the restriction sites on the proposed protein constructs did not alter the reading frame. In addition, the PreScission Protease site remained intact which could be later be exploited for the cleavage of the glutathione-S-transferase (GST) tag.

4. Discussion

Diagnostic tests that are based on multiple, instead of a single antigen, have demonstrated improved sensitivity [71–74]. Recombinant subunit proteins constructed from many TB antigens have become suitable vaccine candidates owing to their potent immune response [75–79]. The epitope regions present on an antigen are responsible for the interaction with the host cells, triggering an immune response [80]. The removal of non-epitope regions from an antigen and the fusion of several epitope regions from the same, or different antigens, may greatly enhance the sensitivity of a diagnostic assay [81]. In addition, the fused epitopes may stimulate a heightened protection owing to the accessibility of these regions for the interaction with immune cells [64]. Adhesin proteins known to elicit an immune response are suitable candidates for the design of a multi-epitope fusion construct [82,83]. In silico and *in vivo* studies have demonstrated that the Heparin-binding hemagglutinin adhesin (HBHA) and the Antigen 85 Complex are well-studied adhesin proteins that have shown great potential as both diagnostic and vaccine candidates [84–89]. An immunodiagnostic test containing the Mtb HBHA induced interferon-gamma-release assay (IGRA) differentiated between active TB infection and latent TB infection but with reduced sensitivities in HIV infected patients [89]. A mucosal vaccine containing the fusion of Ag85B-Acr-HBHA which is absorbed on the surface of the inactivated *Bacillus subtilis* spores displayed reduced bacterial infection in the lungs and spleen of mice but not in non-human primates (NHP) when serving as a booster to the current vaccine, BCG [87]. A trial in South African adolescents (Aeras C-040-404) with BCG revaccination and vaccination with H4:IC31, a recombinant subunit vaccine comprising adhesins Ag85B and TB10.4, displayed efficacy in preventing sustained QuantiFERON (QFT) conversion [90,91]. The study showed highest vaccine-induced $CD4^+$ T-cell response rates were for those recognizing the Ag85B component. A virus-like particle (VLP) consisting of a fusion of the HBHA and MTP epitopes (LV20) was expressed in the Baculovirus-insect cell expression system [86]. When coupled with an adjuvant such as the DDA and Poly I:C (DP), this vaccine demonstrated both a cell-mediated and humoral response. Therefore, the fusion of many immunodominant adhesin proteins make for a more enhanced vaccine candidate.

In this study, two tri-fusion protein constructs comprising multiple epitopes from 3 immunodominant adhesin proteins were designed for use as potential serodiagnostic and vaccine candidates.

4.1. Four immunodominant Mtb adhesin proteins were selected

The VaxiJen software enabled the ranking of 26 known Mtb adhesin proteins based on their ability to act as potential antigens. Any antigen exceeding the bacterial threshold of 0.4 [43] was considered as a potential antigen. The suitability of the top ten ranked adhesin proteins as candidates for the tri-fusion protein construct was determined from published reports. The presence of epitope regions was determined using the IEDB, a database with previously studied experimental data on antibody and T cell epitopes. Six of the ten adhesin proteins did not return a positive result since no experimental data was available on the database at the time of analysis. This included the highly ranked Wag22_PE_PGRS and the PE_PGRS33 adhesin protein.

Therefore, despite being ranked 1 and 2, the adhesin proteins belonging to the PE_PGRS family were not selected owing to their inability to be expressed and purified outside of their host with ease [92,93], and the lack of available experimental data on the IEDB [94]. The HSP70, PstS1, 19 kDa lipoprotein and the HBHA adhesin proteins demonstrated the presence of known epitopes on the IEDB, ranging from 4 to 60 epitope regions, with the HBHA displaying the lowest number. The MTP adhesin protein was selected, despite not

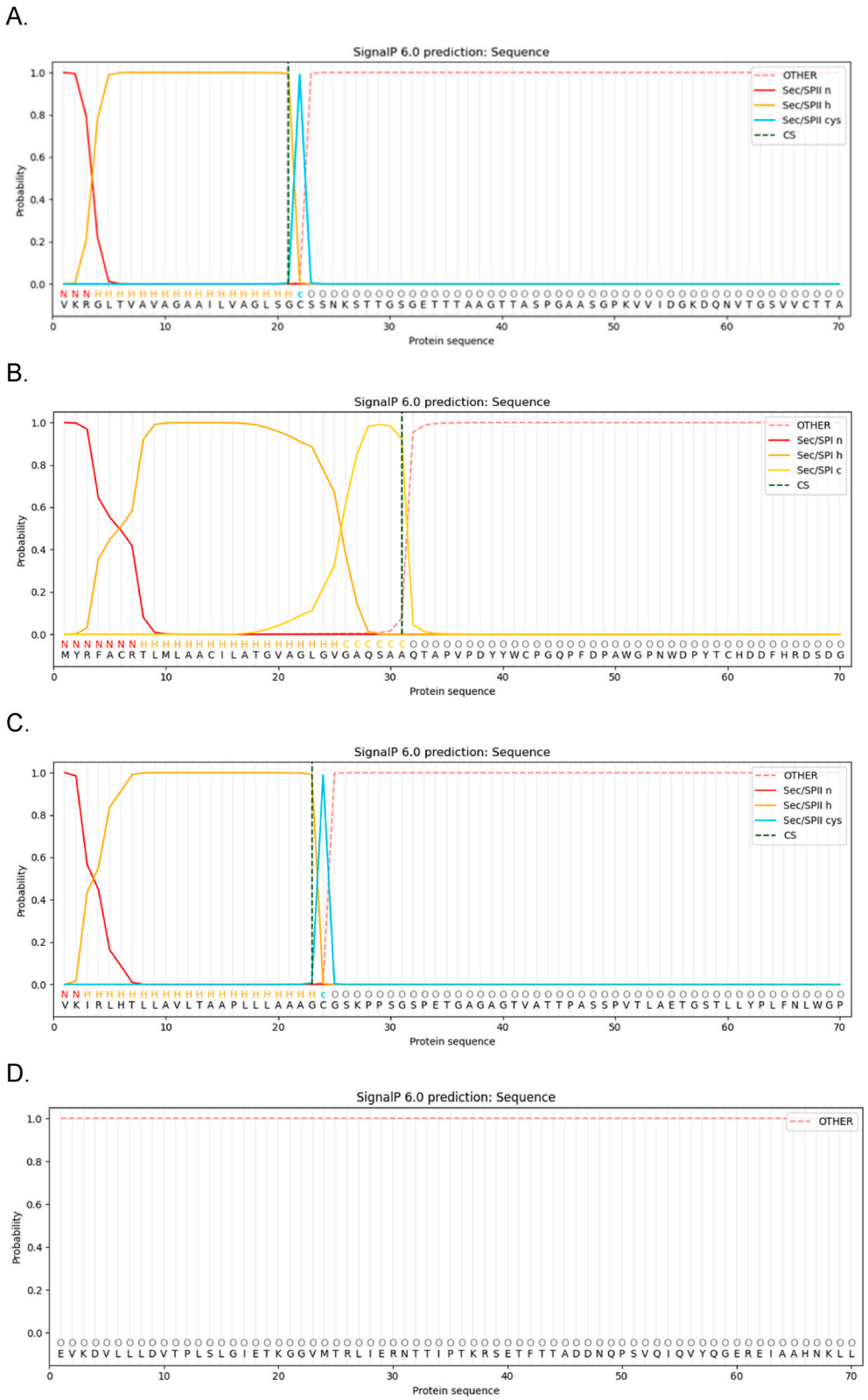


Fig. 2. Prediction of signal peptides on adhesin proteins (A) lpqH, (B) MTP, (C) PstS1 and (D) tnHSP70 using the Signal P 6.0 tool. N: n-terminal region of signal peptide; H: centre hydrophobic region of peptide; C: c-terminal region of signal peptide; O: no signal peptide.

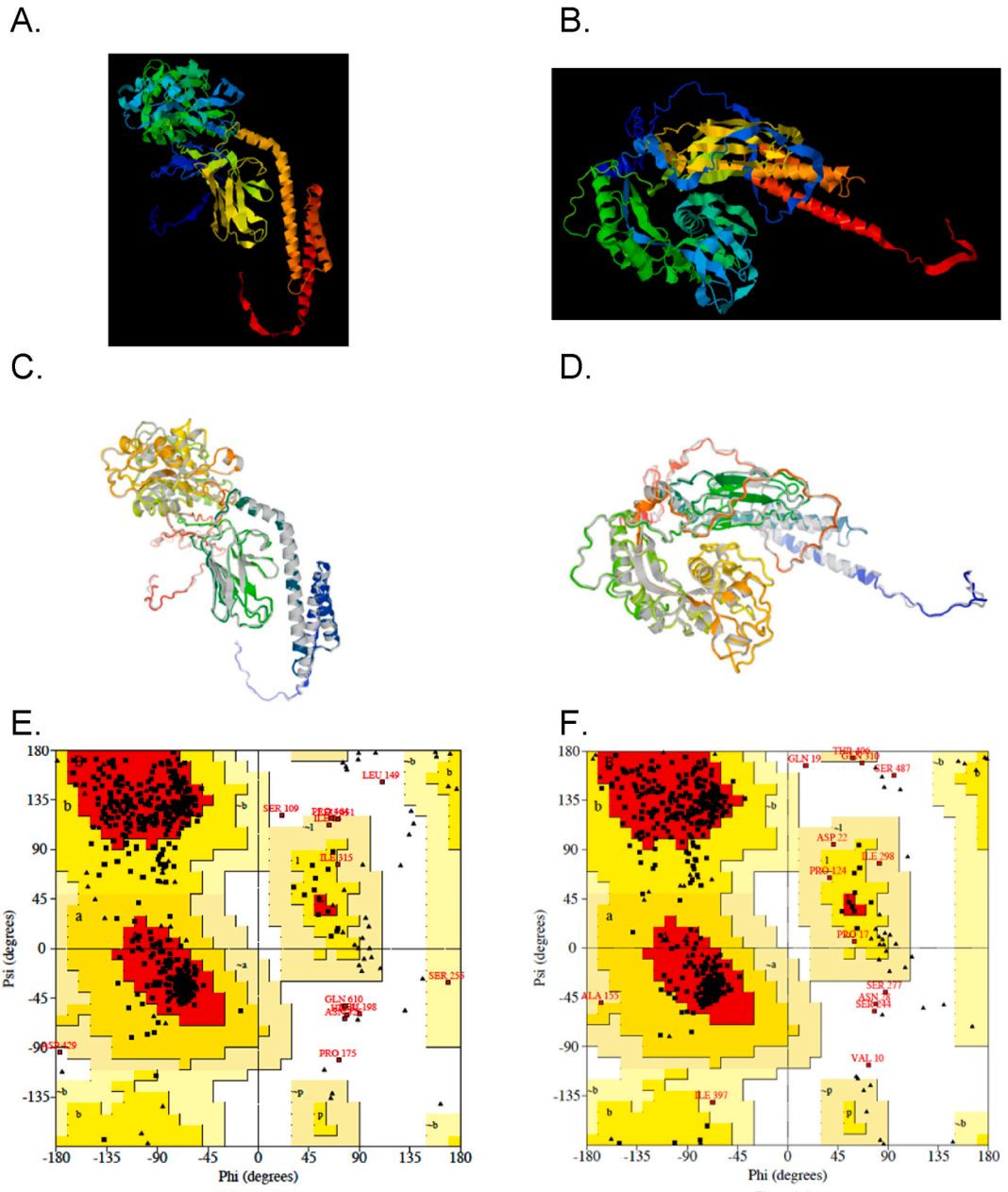


Fig. 3. Predicted 3D tertiary structures of the tri-fusion constructs (A) tnLPH and (B) tnMPH using the D-I-TASSER webserver; the refined tertiary structures of (C) tnLPH and (D) tnMPH modelled on the GALAXY Refine Server and the Ramachandran plots of tnLPH (E) and tnMPH (F) generated by the PROCHECK tool on the SAVEsv6.0 server.

yielding any saved epitope regions, since previous studies conducted by Alteri et al., 2007; Naidoo et al., 2018; Pillay et al., 2024 (unpublished data) demonstrated the presence of anti-MTP antibodies in patient serum samples [63,42,95]; indicating that MTP is capable of eliciting host immune response. The adhesin proteins tnHSP70, PstS1, 19 kDa and MTP were considered suitable candidates for use in the tri-fusion protein construct as discussed below.

4.1.1. Heat-shock protein 70 (HSP70)

Heat-shock proteins (hsp) have been shown to possess anti-inflammatory properties and function as immune modulators. Mycobacterial heat-shock proteins such as the 70kDa display an intense helper effect in cellular and humoral responses when coupled to other peptides in the absence of an adjuvant [96–98]. The HSP70 adhesin is known to induce the early production of IgG. The C-terminal domain of the Mtb HSP70 from amino acids 359–610 was shown to be a highly potent adjuvant in comparison to the amino

Table 4
Predicted MHC I and MHC II Binding Epitopes on tnHSP70, lpqH, MTP and PstS1 using the IEDB

Adhesin Protein	MHC I			MHC II		
	Allele	Epitope	Rank	Allele	Epitope	Rank
tnHSP70	HLA-B*40:01	REIAAHNKL	0.990	HLA-DRB3*01:0	EVTFDIDANGIVHVT	0.11
	HLA-A*68:01	ETKGGVMTR	0.982	HLA-DRB3*01:0	IEVTFDIDANGIVHV	0.13
	HLA-B*07:02	APRGIPQIEV	0.915	HLA-DRB3*02:0	TRLIERNTTIPTKRS	0.08
	HLA-A*03:01	RIQEGSGLSK	0.913	HLA-DRB3*01:0	QIEVTFDIDANGIVH	0.13
	HLA-A*02:03	RLIERNTTI	0.900	HLA-DRB5*01:0	ANGIVHVTAKDKGTG	0.13
lpqH	HLA-A*11:01	ATKDGSHYK	0.960	HLA-DRB1*03:01	ASGPKVVIDGKDQNV	2.50
	HLA-A*30:01	ATKDGSHYK	0.946	HLA-DRB1*03:01	SGPKVVIDGKDQNVT	2.50
	HLA-A*26:01	NVNGVTLGY	0.928	HLA-DRB1*03:01	AASGPKVVIDGKDQN	2.70
	HLA-A*03:01	ATKDGSHYK	0.872	HLA-DRB1*03:01	GPKVVIDGKDQNVTG	2.70
	HLA-A*68:02	TTAAGNVNI	0.815	HLA-DRB1*03:01	PKVVIDGKDQNVTGS	2.70
PstS1	HLA-B*57:01	QTLQAFLHW	0.996	HLA-DRB5*01:01	INYEYAIVNNRQKDA	0.60
	HLA-B*58:01	QTLQAFLHW	0.993	HLA-DRB5*01:01	IINYEYAIVNNRQKD	0.66
	HLA-B*57:01	LSKQDPEGW	0.980	HLA-DRB5*01:01	NYEYAIVNNRQKDAA	0.66
	HLA-B*07:02	TPASSPVTL	0.978	HLA-DRB5*01:01	PIINYEYAIVNNRQK	0.66
	HLA-B*58:01	LSKQDPEGW	0.975	HLA-DRB5*01:01	YEYAIVNNRQKDAAT	0.71
MTP	HLA-B*57:01	QTAPVPDYYW	0.994	HLA-DRB3*01:01	APVPDYYWCPGQPFD	5.80
	HLA-B*58:01	QTAPVPDYYW	0.990	HLA-DRB3*01:01	PVPDYYWCPGQPFDP	5.90
	HLA-A*01:01	QTAPVPDYY	0.977	HLA-DRB3*01:01	VPDYYWCPGQPFDPA	6.00
	HLA-A*26:01	QTAPVPDYY	0.827	HLA-DRB3*01:01	PDYYWCPGQPFDPAW	6.20
	HLA-A*30:02	QTAPVPDYY	0.805	HLA-DRB3*01:01	TAPVPDYYWCPGQPF	6.80

Table 5
Physicochemical and protein parameters of trifusion constructs identified using the ProtParam ExPasy Tool

Protein Parameters	tnLPH	tnMPH
Number of amino acids	636	617
Molecular weight	64865.80	63958.62
Theoretical pI	4.77	4.63
Total number of negatively charged residues (Asp + Glu)	73	79
Total number of positively charged residues (Asp + Glu)	45	43
Formula	$C_{2809}H_{4458}N_{798}O_{942}S_{12}$	$C_{2793}H_{4336}N_{782}O_{917}S_{13}$
Total number of atoms	9019	8841
Extinction coefficients all pairs of Cys residues form cystines all Cys residues are reduced	35995 35870	55600 55350
Estimated half-life	30 h (mammalian reticulocytes, in *vitro* >20 h (yeast, in *vivo*) >10 h. (*E. coli,* in *vivo*)	1 h (mammalian reticulocytes, in *vitro* 30 min (yeast, in *vivo*) >10 h. (*E. coli,* in *vivo*)
Instability Index	27.52	34.23
Aliphatic index	76.79	70.97
Grand average of hydropathicity (GRAVY)	−0.29	−0.406

acids located in the N-terminal domain when investigating the humoral response of the Hepatitis B vaccine [99]. The early secretory protein 6 (ESAT-6) fused to the C-terminal of the HSP70 stimulated Th1 mediated immunity and was shown to retain immunogenicity when fused together [66]. The use of the HSP70 as a fusion protein resulted in protein expression within both the soluble fraction and inclusion body, with the latter displaying a greater yield [66]. Since HSP70 is a molecular chaperone, it can assist in the refolding of proteins and prevent aggregation formation during stressful conditions [100]. It is anticipated that the use of the C-terminal of the $Hsp70_{359\text{-}610}$ will not only enhance the immune response attained by the tri-fusion protein vaccine candidate but will also promote the expression of the recombinant protein within the soluble fraction. Further analysis revealed that this region contains several B- and T-cell epitope sequences.

4.1.2. 19 kDa lpqH lipoprotein

Lipoproteins are involved in the virulence and immunoregulatory processes of TB pathogenesis. The lpqH adhesin protein interacts with the mannose receptor to promote the phagocytosis of the Mtb bacilli. In addition to inducing a T cell mediated immunity, it has been shown to down-regulate the antigen presentation to T cells [101–103]. The stage-specific expression promotes such adhesin

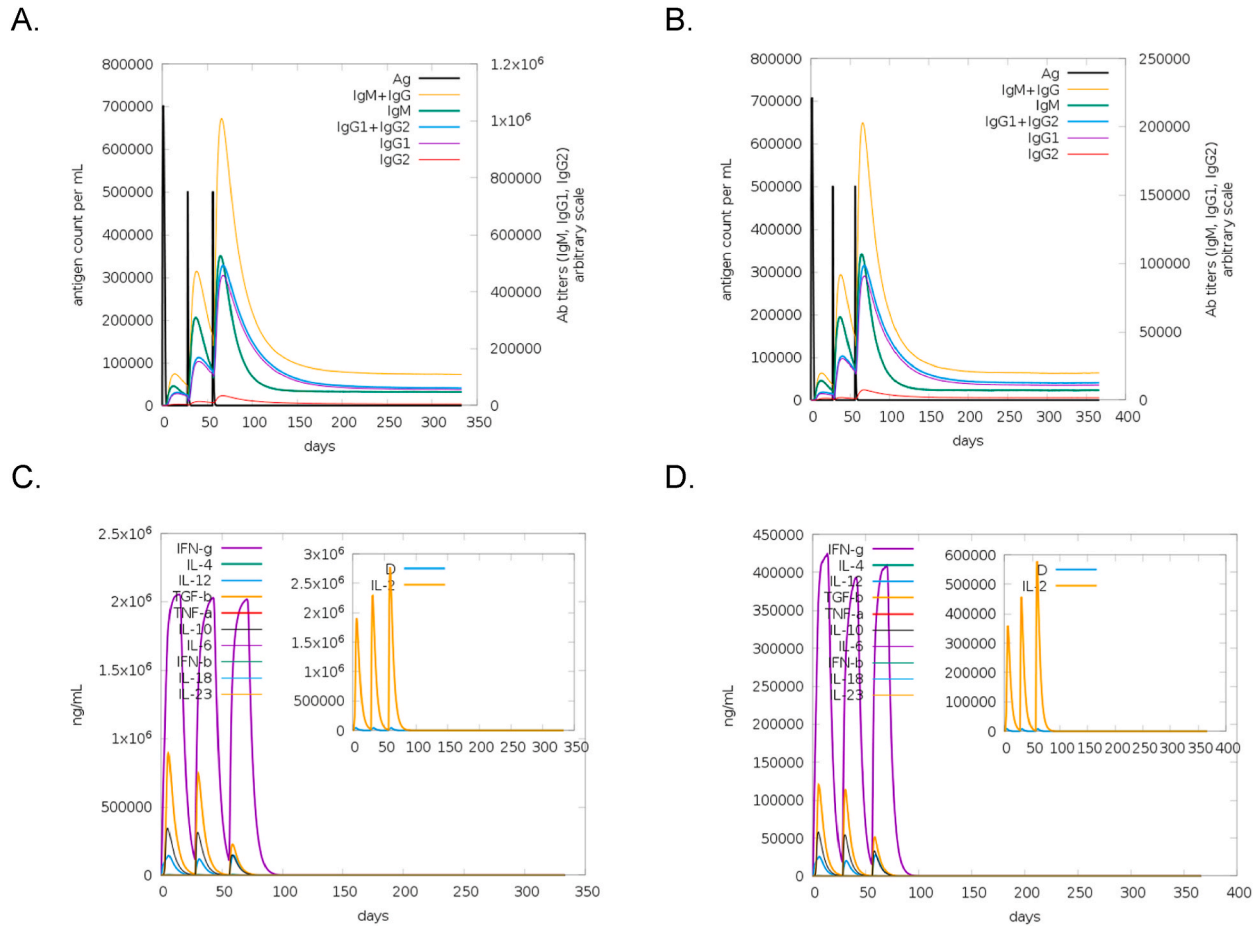


Fig. 4. In silico analysis of antibody and cytokine immune responses induced by the tnLPH and tnMPH vaccine candidates using the C-ImmSim server. The response of immunoglobulin production when exposed to the antigen (Ag) tnLPH (A) and tnMPH (B). The production and concentration of cytokines and interleukins after immunizations with tnLPH (C) and tnMPH (D). The D curve, shown in the small box within the graph, represents the diversity and is a signal of danger.

proteins as suitable vaccine candidates. The DNA vaccine constructed with lpqH did not provide protection in mice against TB [104, 105]. In contrast, the expression of the lpqH in strains such as *Mycobacterium smegmatis* and *Mycobacterium vaccae* demonstrated a reduced protection in mice when challenged with Mtb [104]. This indicates that the use of a live form of mycobacteria displays reduced vaccine efficacy and should therefore not be considered for testing the lpqH. A DNA vaccine comprising lpqH fused to a marked protein of the autophagosome, a microtubule associated protein light chain-3 (CL3), lowered bacterial load in the spleen and lungs of challenged mice [106]. This indicates an improved Th1 response and protection when two or more proteins are fused in the construction of vaccine candidates. In animals, lpqH is regarded as a plasma biomarker which can differentiate between paratuberculosis positive animals or paratuberculosis vaccinated cows from the TB positive cows [107].

4.1.3. Phosphate transport protein 1 (PstS1)

Another lipoprotein, PstS1, a phosphate-binding transporter protein, is a glycosylated protein with mannose residues located at the N-terminal region [108]. The study of the three PstS proteins (PstS1, PstS2 and PstS3) as DNA vaccine candidates demonstrated that the PstS1 did not confer a protective effect, whilst the PstS3 induced the highest cytokine secretion [109]. However, PstS1 displayed a strong antibody response suggesting that it may be used as a serodiagnostic biomarker [30]. It has also been shown to differentiate between active and latent TB and is most likely associated with active TB [110]. The use of a truncated PstS1 antigen was shown to improve the diagnostic efficiency of the assay [111]. The removal of several amino acid sequences from both the N- and C-terminal domains unmasked the epitopes, leading to greater exposure for antibody-antigen interaction. The authors further showed that the fusion of the truncated PstS1 to the EchA1 (Rv0222) proteins displayed an improved serodiagnosis of TB [112].

4.2. Two trifusion protein constructs with multi-epitope regions lacking a signal peptide with good population coverage

Common B-cell epitope regions (Table 3) predicted by the three online tools were incorporated into the design of the tri-fusion

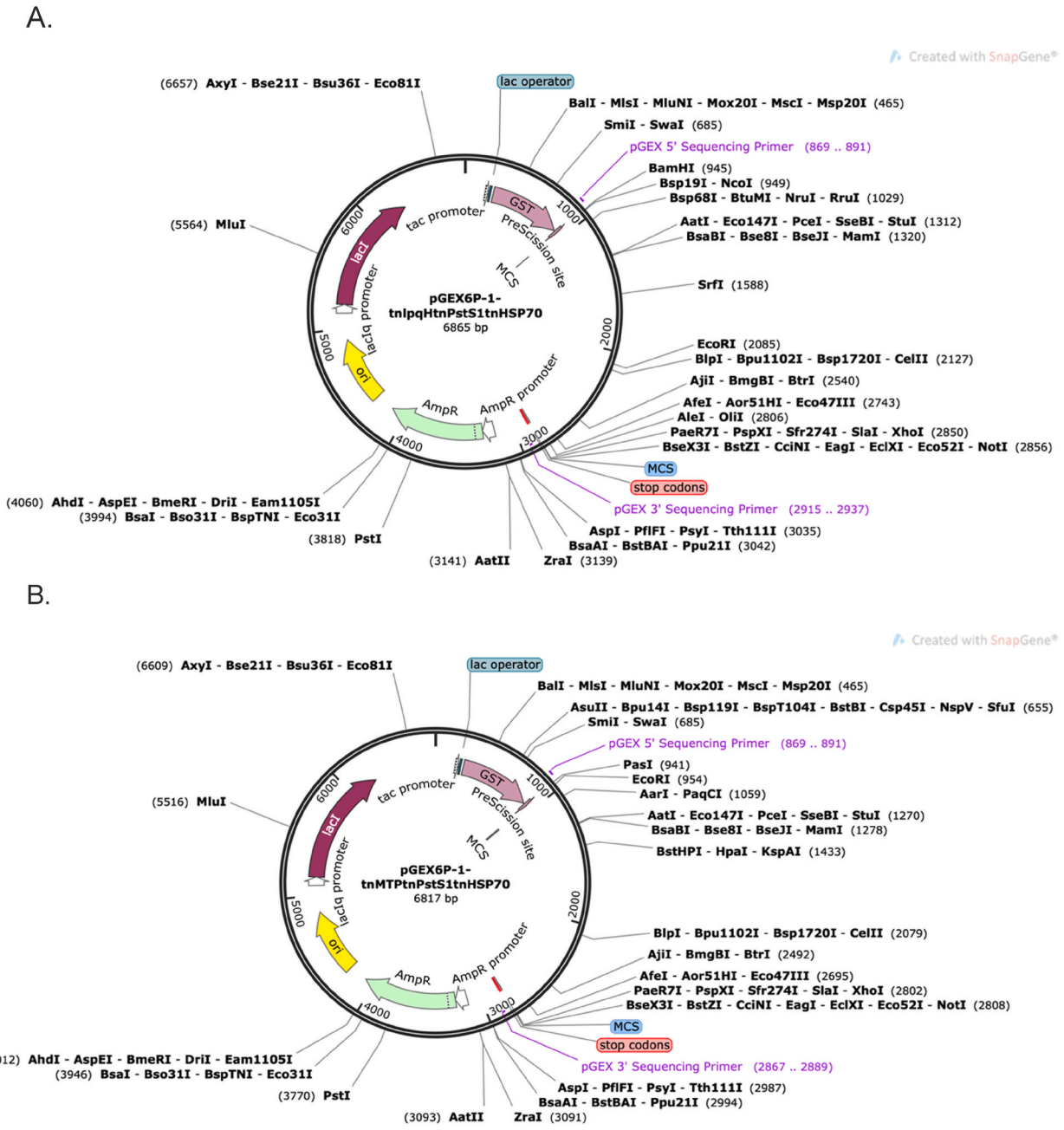


Fig. 5. Restriction map of recombinant pGEX6P-1- tnLPH (A) and Restriction map of recombinant pGEX6P-1- tnMPH (B) generated using the Snapgene software.

protein. T-cell epitope regions were identified using IEDB where reference alleles for both the MHC I and MHC II binding sites were selected. The epitope regions were also included in the tri-fusion protein constructs. Since cloning was to be performed in-house, a continuous region of the epitopes on the gene sequence was used. Three adhesin proteins, lpqH, PstS1 and MTP were identified by the signal peptide prediction tool to contain a signal peptide at the N-terminal domain. For the design of the tri-fusion protein construct, the signal peptides of these proteins were removed to optimize the expression of the recombinant protein [113]. The truncated gene sequence encoding the adhesin proteins that were used in the design of the tri-fusion protein are shown in Table S1 with the schematic of the design of the trifusion constructs shown in Fig. 2. The population coverage of both the tnLPH and tnMPH trifusion protein constructs in the South African population was 91,81 % for a combination of MHC alleles with an average of 1,98 epitopes recognized by the population. In addition, a world coverage of 90,95 % was obtained. This indicates that the potential vaccine candidates can be recognized by a vast number of people and can therefore elicit an immune response.

4.3. AllerTop 2.0 and the hybrid approach indicate that both trifusion protein constructs are non-allergens

The ability of the antigens to induce an abnormal immune response was predicted using AlgPred and AllerTop 2.0 (Table S2). The AllerTop server predicts the allergenicity of a protein based on the physicochemical characteristics of the protein sequence [55,114]. AlgPred integrates the results of various methods to predict an allergen [54], however, the allergenicity results of the truncated proteins obtained were contradictory in each of the methods. According to the IgE epitope mapping, Motif Alignment and Search Tool (MAST), all the truncated regions of the adhesin protein and the tri-fusion constructs, tnLPH and tnMPH are non-allergens. The MAST tool displayed the highest sensitivity of 93.94% with a specificity of 33.34%, whilst the highest specificity of 98.14% with a sensitivity of 17.47% was achieved by the IgE mapping. However, The Support Vector Machine (SVM) based on amino acid composition which has an accuracy of 85.02% predicted that both the tri-fusion protein constructs and the tnHSP70, tnlpqH, tnPstS1 and tnHSP70 proteins are probable allergens, whilst the tnMTP is identified as an allergen. Using the dipeptide composition with an accuracy of 85%, the tnHSP70, tnlpqH, tnPstS1 and tnLPH were identified as potential allergens, whereas the tnMTP was identified as a non-allergen and the tnMPH tri-fusion construct was identified as an allergen. The BLAST results predicted that all individual truncated proteins were non-allergens but a sequence within the tri-fusion protein constructs was identified as an allergen. However, the Hybrid Approach predicted that all proteins were non-allergens. This approach is based on the combination of two or more methods and can therefore be considered highly likely. The AllerTop 2.0 online tool predicted that all proteins are non-allergens except for tnMTP, which was identified as a probable antigen. The AllerTop 2.0 tool has been shown to display the highest sensitivity (94%) when compared to other tools whilst the AlgPred showed the highest specificity (100%) [114]. Therefore, all results were taken into consideration for finalizing the design and use of the tri-fusion protein construct.

4.4. Hydrophilic, thermo-stable tri-fusion protein constructs identified by ExPasy ProtParam with a confident prediction of tertiary structure by D-I-TASSER

The physicochemical properties of each tri-fusion protein were computed by ExPasy ProtParam tool. These included the amino acid composition, molecular weight, theoretical isoelectric points (pI), extinction coefficient, instability index, aliphatic index, grand average of hydropathicity (GRAVY), and total number of negatively and positively charged residues (Table 5). The theoretical pI was 4.77 and 4.63, for tnLPH and tnMPH, respectively. The theoretical pI is the pH at which a particular molecule or surface carries no net electrical charge, and it is useful for understanding the protein charge stability. The extinction coefficient is defined as a measurement of how strongly a protein absorbs light at 280 nm and may aid in the study of the interactions of protein-protein and protein-ligand [115]. The extinction coefficients for tnLPH and tnMPH were determined to be 35870 and 55350, respectively. A high extinction coefficient is associated with the high concentration of cysteine residues which implies a more stable protein structure [116,117]. The instability index is an estimation of the stability of a protein experimentally, where a value smaller than 40 is considered as stable. The instability index values for tnLPH and tnMPH is 27.52 and 34.23, respectively, indicating that both tri-fusion proteins are stable. The half-life of tnLPH and tnMPH were estimated at 30 h and 1 h in mammalian reticulocytes respectively. This implies that the tnLPH may actually be able to trigger a better immune response *in vivo* rather than the tnMPH since it will persist longer within the host. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains, alanine, valine, isoleucine and leucine. The aliphatic index for tnLPH and tnMPH is 76.79 and 70.97, respectively, indicating that the tri-fusion protein is thermo-stable over a wide temperature range. The GRAVY value was calculated as the sum of hydropathy values of all of the amino acids divided by the number of residues in the sequence. A low GRAVY value indicates the possibility of the protein being a hydrophilic rather than hydrophobic protein. The GRAVY index of the tri-fusion protein was calculated to be −0.29 and-0.406 for tnLPH and tnMPH, respectively. This implies that the tri-fusion protein may be hydrophilic in nature and maybe more amenable to purification strategies.

The D-I-TASSER online software was used to predict the tertiary structure of both trifusion constructs which proposed 5 models with varying TM-scores. TM-score is a more accurate measure of global similarity between two structures in comparison to the RMSD [58]. The TM-score closest to 1 implies a perfect match between the protein structures and the first models for both constructs displayed the highest TM-score (Fig. 3A and B).

This models were subjected to refinement to optimize the predicted structures. The Galaxy Refine server generated five models with the first model being the most favoured. The Galaxy Refine displays six factors relating to the original model submission, including the GDT-HA, RMSD, MolProbity, Clash score, Poor Rotamers and Rama favoured [59]. The GDT-HA and RMSD pertains to the similarity of the two structures and this scores indicated that the refined model and the original model were highly similar. The MolProbity score should lie between −4 and 2. The MolProbity score for the original models of both proteins were >3 and the reduction observed is an indication of a refined model with a closer to normal structure. This is further confirmed by the reduction in clash scores observed [118]. The poor rotamer score refers to the geometry of the amino acid side chains and its presence in the poor favoured category. Despite an increase in comparison to the original model, the refined models are within the accepted range since a value greater than 1.5% is a cause for concern. The Rama favourability of these models is linked to the proposed statistical distribution of the combinations of the residues on the backbone dihedral angles. This is graphically illustrated by the Ramachandran plots generated by the PROCHECK validation server [119,120]. The refined tnLPH and tnMPH models contain 87.7–89.1% of residues in the favoured regions; 10–8.5% in the additional allowed regions, 0.8–1% in the generously allowed regions; and 1.4% in the disallowed regions. The residues in disallowed region should not exceed 2% as this would indicate steric hindrance between the side chain and the main atoms resulting in a poor protein structure. Since both constructs are below this value, the refined structure is of good quality. The ERRAT program verifies protein structure that is determined by crystallography and provides an overall measure of quality and resolution of structure [121]. A good model should display an overall quality of 95% but anything above 50% is accepted. The overall quality of the

tnLPH and tnMPH was 76.51% and 85.25%, respectively. Collectively, the refinement and validation results indicate that the predicted structures for the tnLPH and tnMPH trifusion proteins are good. The in silico cloning of both trifusion constructs into the pGEX6P-1 vector indicated that the reading frame of the sequence is not altered nor does its incorporation affect the PreScission Protease sequence on the vector.

4.5. In silico immune simulation shows that both trifusion constructs are immunogenic

In silico analysis of the tri-fusion constructs indicated that the vaccine candidates may be immunogenic since the design was based on B and T cell epitopes which are needed to stimulate both a humoral and adaptive immune response. This is further evidenced by the in silico immune simulation which revealed that both the tnLPH and tnMPH elicited both an adaptive and innate immune response characterized by heightened concentrations of TH cells and B-cells during immunization. Furthermore, both candidates stimulated the production of the IFN-g, TGF-b IL-10 and IL-12 cytokines which are essential for the protection of the host (Fig. 4C and D). The presence of immunoglobulins remained after the eradication of the antigen (Fig. 4A and B). The use of a GGGGS linker sequence between each of the truncated adhesin protein will provide accessibility and flexibility to the epitope regions. The analysis of the physicochemical characteristics of the tri-fusion construct indicates that both candidates when expressed, should be hydrophilic and stable but this is yet to be experimentally confirmed. Therefore, the present findings should be followed up with the synthesis of the proposed trifusion constructs, cloned into a suitable vector and transformed into an appropriate bacterial host such as the *Escherichia coli* BL21. The recombinant protein should be expressed and purified in sufficient quantity for the immunization of mice to further validate the findings of the in silico analysis. Furthermore, the recombinant proteins can be evaluated in an enzyme-linked immunosorbent assay (ELISA) to determine the diagnostic potential of these two tri-fusion proteins.

5. Conclusions

The use of in silico immunoinformatic tools has been shown to be vital to the generation of potential vaccine candidates, diagnostic biomarkers and therapeutic drug targets. It has greatly assisted in the control of coronavirus pandemic and has extensively been used for the design of new TB vaccine candidates. Two potential subunit vaccines comprising of truncated epitope regions of immunodominant Mtb adhesin proteins were designed. The phosphate-binding transport protein PstS1, C-terminal region of heat-shock protein 70, and either the *Mycobacterium tuberculosis* curli pili or the 19 kDa lipoprotein, were incorporated in the each of the trifusion protein constructs. The analysis revealed that the potential vaccine candidates are immunogenic, thermostable, hydrophilic and are structures of good quality. However, these claims are yet to be validated by *in vitro* and *in vivo* experiments.

Ethical considerations

This study was approved by the Biomedical Research Ethics Committee (BREC) of the University of KwaZulu-Natal (BREC/00001444/2020).

CRediT authorship contribution statement

Koobashnee Pillay: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Thamsanqa E. Chiliza:** Writing – review & editing. **Sibusiso Senzani:** Writing – review & editing. **Balakrishna Pillay:** Conceptualization, Writing – review & editing. **Manormoney Pillay:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

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

Acknowledgements

This research was funded by the project running costs of MP's UKZN Research Flagship "Afrocentric Precision Approach to Control Health Epidemics (APACHE)" grant award as Co-PI. KP was supported by the National Research Foundation (NRF)- German Academic Exchange Service (DAAD) PhD Scholarship (grant ID 117866).

This research was presented at the University of KwaZulu-Natal's College of Health Science School of Laboratory Medicine and Medical Sciences Research Day on the October 4, 2023 in Westville, South Africa.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37536.

References

[1] K. Dheda, T. Perumal, H. Moultrie, R. Perumal, A. Esmail, A.J. Scott, et al., The intersecting pandemics of tuberculosis and COVID-19: population-level and patient-level impact, clinical presentation, and corrective interventions, The Lancet Respiratory 10 (2022) 603–622, https://doi.org/10.1016/S2213-2600(22)00092-3.

[2] World Health Organization, Global tuberculosis report 2022. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022, 2022. (Accessed 23 April 2023).

[3] G. Voss, D. Casimiro, O. Neyrolles, A. Williams, S.H.E. Kaufmann, H. McShane, et al., Progress and challenges in TB vaccine development, F1000Res 7 (2018) 199, https://doi.org/10.12688/f1000research.13588.1.

[4] H. Mutavhatsindi, G.D. van der Spuy, S.T. Malherbe, J.S. Sutherland, A. Geluk, H. Mayanja-Kizza, et al., Validation and optimization of host immunological bio-signatures for a point-of-care test for TB disease, Front. Immunol. 12 (2021) 194, https://doi.org/10.3389/FIMMU.2021.607827/BIBTEX.

[5] S. Luca, T. Mihaescu, Maedica-a Journal of Clinical Medicine History of BCG Vaccine, vol. 8, 2013.

[6] G.G. Mahairas, P.J. Sabo, M.J. Hickey, D.C. Singh, C.K. Stover, Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*, J. Bacteriol. 178 (1996) 1274, https://doi.org/10.1128/JB.178.5.1274-1282.1996.

[7] A. Keyser, J.L.M. Troudt, J.L. Taylor, A.A. Izzo, BCG sub-strains induce variable protection against virulent pulmonary *Mycobacterium tuberculosis* infection, with the capacity to drive Th2 immunity, Vaccine 29 (2011) 9308, https://doi.org/10.1016/J.VACCINE.2011.10.019.

[8] E. Nemes, H. Geldenhuys, V. Rozot, K.T. Rutkowski, F. Ratangee, N. Bilek, et al., Prevention of M. Tuberculosis infection with H4:IC31 vaccine or BCG revaccination, N. Engl. J. Med. 379 (2018), https://doi.org/10.1056/NEJMoa1714021.

[9] R.G. White, W.A. Hanekom, J. Vekemans, R.C. Harris, The way forward for tuberculosis vaccines. https://doi.org/10.1093/cid/ciy835, 2019.

[10] S.H.E. Kaufmann, H.M. Dockrell, N. Drager, M.M. Ho, H. McShane, O. Neyrolles, et al., TBVAC2020: advancing tuberculosis vaccines from discovery to clinical development, Front. Immunol. 8 (2017) 1203, https://doi.org/10.3389/fimmu.2017.01203.

[11] G. Voss, D. Casimiro, O. Neyrolles, A. Williams, S.H.E. Kaufmann, H. Mcshane, et al., Progress and challenges in TB vaccine development, F1000 Research 7 (2018) 1–14, https://doi.org/10.12688/f1000research.13588.1.

[12] B. Zhu, H.M. Dockrell, T.H.M. Ottenhoff, T.G. Evans, Y. Zhang, Tuberculosis vaccines: opportunities and challenges, Respirology (2018), https://doi.org/10.1111/resp.13245.

[13] World Health Organization, WHO preferred product characteristics for new tuberculosis vaccines department of immunization, Vaccines And Biologicals (2018).

[14] TB Vaccine Clinical Pipeline - Working Group on New TB Vaccines n.d. https://newtbvaccines.org/tb-vaccine-pipeline/(accessed January 6, 2024).

[15] C. Martin, N. Aguilo, D. Marinova, J. Gonzalo-Asensio, Update on TB vaccine pipeline, Appl. Sci. 10 (2020), https://doi.org/10.3390/app10072632.

[16] Pipeline of vaccines - TBVI n.d. https://www.tbvi.eu/what-we-do/pipeline-of-vaccines/(accessed September 17, 2022).

[17] T.J. Scriba, S.H.E. Kaufmann, P.H. Lambert, M. Sanicas, C. Martin, O. Neyrolles, Vaccination against tuberculosis with whole-cell mycobacterial vaccines, J. Infect. Dis. 214 (2016) 659–664, https://doi.org/10.1093/INFDIS/JIW228.

[18] V.T. Duong, M. Skwarczynski, I. Toth, Towards the development of subunit vaccines against tuberculosis: the key role of adjuvant, Tuberculosis 139 (2023) 102307, https://doi.org/10.1016/J.TUBE.2023.102307.

[19] M. Gengenbacher, S.H.E. Kaufmann, *Mycobacterium tuberculosis*: success through dormancy, FEMS Microbiol. Rev. 36 (2012) 514, https://doi.org/10.1111/J.1574-6976.2012.00331.X.

[20] K. Wright, K. Plain, A. Purdie, B.M. Saunders, K. De Silva, Biomarkers for detecting resilience against mycobacterial disease in animals, Infect. Immun. 88 (2020), https://doi.org/10.1128/IAI.00401-19.

[21] V.S. Govender, S. Ramsugit, M. Pillay, *Mycobacterium tuberculosis* adhesins: potential biomarkers as anti-tuberculosis therapeutic and diagnostic targets, Microbiology (N. Y.) 160 (2014) 1821–1831.

[22] S. Ramsugit, M. Pillay, Identification of *Mycobacterium tuberculosis* adherence-mediating components: a review of key methods to confirm adhesin function, Iran J Basic Med Sci 19 (2016) 579–584.

[23] Q. Xin, H. Niu, Z. Li, G. Zhang, L. Hu, B. Wang, et al., Subunit vaccine consisting of multi-stage antigens has high protective efficacy against *Mycobacterium tuberculosis* infection in mice, PLoS One 8 (2013) e72745, https://doi.org/10.1371/JOURNAL.PONE.0072745.

[24] C. Bellini, K. Horváti, Recent advances in the development of protein- and peptide-based subunit vaccines against tuberculosis, Cells 9 (2020), https://doi.org/10.3390/CELLS9122673.

[25] W. Yuan, N. Dong, L. Zhang, J. Liu, S. Lin, Z. Xiang, et al., Immunogenicity and protective efficacy of a tuberculosis DNA vaccine expressing a fusion protein of Ag85B-Esat6-HspX in mice, Vaccine 30 (2012) 2490–2497, https://doi.org/10.1016/j.vaccine.2011.06.029.

[26] J. Fulkerson, M. Brennan, K. Velmurugan, C. Locht, Recombinant Mycobacterium Encodinga Heparin-Binding Hemagglutinin (Hbha) Fusion, 2012. US 2018/0305416 A1.

[27] M. de Martino, L. Lodi, L. Galli, E. Chiappini, Immune response to *Mycobacterium tuberculosis*: a narrative review, Front Pediatr 7 (2019), https://doi.org/10.3389/FPED.2019.00350.

[28] F. Squeglia, A. Ruggiero, A. De Simone, R. Berisio, A structural overview of mycobacterial adhesins: key biomarkers for diagnostics and therapeutics, Protein Sci. 27 (2018) 369–380, https://doi.org/10.1002/pro.3346.

[29] A. Ragas, L. Roussel, G. Puzo, M. Rivière, The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A, J. Biol. Chem. 282 (2007) 5133–5142, https://doi.org/10.1074/jbc.M610183200.

[30] M. Esparza, B. Palomares, T. García, P. Espinosa, E. Zenteno, R. Mancilla, PstS-1, the 38-kDa *Mycobacterium tuberculosis* glycoprotein, is an adhesin, which binds the macrophage mannose receptor and promotes phagocytosis, Scand. J. Immunol. 81 (2015) 46–55, https://doi.org/10.1111/sji.12249.

[31] D.R. Ronning, V. Vissa, G.S. Besra, J.T. Belisle, J.C. Sacchettini, *Mycobacterium tuberculosis* antigen 85A and 85C structures confirm binding orientation and conserved substrate specificity, J. Biol. Chem. 279 (2004) 36771–36777, https://doi.org/10.1074/jbc.M400811200.

[32] T. Hickey, H.J. Ziltener, D.P. Speert, R.W. Stokes, *Mycobacterium tuberculosis* employs Cpn60. 2 as an adhesin that binds CD43 on the macrophage surface, Cell Microbiol. 12 (2010) 1634–1647.

[33] N. Naidoo, M. Pillay, Bacterial pili, with emphasis on *Mycobacterium tuberculosis* curli pili: potential biomarkers for point-of care tests and therapeutics, Biomarkers 22 (2017) 93–105, https://doi.org/10.1080/1354750X.2016.1252960.

[34] P.S. Renshaw, K.L. Lightbody, V. Veverka, F.W. Muskett, G. Kelly, T.A. Frenkiel, et al., Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6, EMBO J. 24 (2005) 2491–2498, https://doi.org/10.1038/sj.emboj.7600732.

[35] W. Xolalpa, A.J. Vallecillo, M. Lara, G. Mendoza-Hernandez, M. Comini, R. Spallek, et al., Identification of novel bacterial plasminogen-binding proteins in the human pathogen *Mycobacterium tuberculosis*, Proteomics 7 (2007) 3332–3341, https://doi.org/10.1002/pmic.200600876.

[36] V. Pancholi, G.S. Chhatwal, Housekeeping enzymes as virulence factors for pathogens, International Journal of Medical Microbiology 293 (2003) 391–401.

[37] F.D. Menozzi, R. Bischoff, E. Fort, M.J. Brennan, C. Locht, Molecular Characterization of the Mycobacterial Heparin-Binding Hemagglutinin, a Mycobacterial Adhesin, vol. 95, Proceedings of the National Academy of Sciences, 1998, pp. 12625–12630.

[38] S. Kumar, B.L. Puniya, S. Parween, P. Nahar, S. Ramachandran, Identification of novel adhesins of M. tuberculosis H37Rv using integrated approach of multiple computational algorithms and experimental analysis, PLoS One 8 (2013) e69790, https://doi.org/10.1371/journal.pone.0069790.

[39] A.G. Kinhikar, D. Vargas, H. Li, S.B. Mahaffey, L. Hinds, J.T. Belisle, et al., *Mycobacterium tuberculosis* malate synthase is a laminin†€ binding adhesin, Mol. Microbiol. 60 (2006) 999–1013.

[40] S. Chitale, S. Ehrt, I. Kawamura, T. Fujimura, N. Shimono, N. Anand, et al., Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry, Cell Microbiol. 3 (2001) 247–254.

[41] M. Singh, A. Martens, J. Campuzano, A. Amador, M. Mondragón- Palomino, J.P. Laclette, et al., The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis* : a new family of fibronectin-binding proteins? Microbiology (N. Y.) 145 (1999) 3487–3495, https://doi.org/10.1099/00221287-145-12-3487.

[42] C.J. Alteri, J. Xicohténcatl-Cortes, S. Hess, G. Caballero-Olin, J.A. Girón, R.L. Friedman, *Mycobacterium tuberculosis* produces pili during human infection, Proc Natl Acad Sci U S A 104 (2007) 5145–5150, https://doi.org/10.1073/PNAS.0602304104.

[43] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (2007) 1–7, https://doi.org/10.1186/1471-2105-8-4/TABLES/2.

[44] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, et al., The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (2019) D339–D343, https://doi.org/10.1093/NAR/GKY1006.

[45] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res. 45 (2017) W24, https://doi.org/10.1093/NAR/GKX346.

[46] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins 65 (2006) 40–48, https://doi.org/10.1002/PROT.21078.

[47] B. Yao, L. Zhang, S. Liang, C. Zhang, SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity, PLoS One 7 (2012), https://doi.org/10.1371/JOURNAL.PONE.0045152.

[48] F. Teufel, J.J. Almagro Armenteros, A.R. Johansen, M.H. Gíslason, S.I. Pihl, K.D. Tsirigos, et al., SignalP 6.0 predicts all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (40) (2022) 1023–1025, https://doi.org/10.1038/s41587-021-01156-3, 7 2022.

[49] S.K. Dhanda, S. Mahajan, S. Paul, Z. Yan, H. Kim, M.C. Jespersen, et al., IEDB-AR: immune epitope database—analysis resource in 2019, Nucleic Acids Res. 47 (2019) W502–W506, https://doi.org/10.1093/nar/gkz452.

[50] B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, Nucleic Acids Res. 48 (2020) W449–W454, https://doi.org/10.1093/nar/gkaa379.

[51] P. Wang, J. Sidney, C. Dow, B. Mothé, A. Sette, B. Peters, A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach, PLoS Comput. Biol. 4 (2008), https://doi.org/10.1371/JOURNAL.PCBI.1000048.

[52] P. Wang, J. Sidney, Y. Kim, A. Sette, O. Lund, M. Nielsen, et al., Peptide binding predictions for HLA DR, DP and DQ molecules, BMC Bioinf. 11 (2010) 1–12, https://doi.org/10.1186/1471-2105-11-568/TABLES/5.

[53] M. Tshabalala, J. Mellet, M.S. Pepper, Human leukocyte antigen diversity: a southern african perspective, J Immunol Res 2015 (2015), https://doi.org/10.1155/2015/746151.

[54] S. Saha, G.P.S. Raghava, AlgPred: prediction of allergenic proteins and mapping of IgE epitopes, Nucleic Acids Res. 34 (2006), https://doi.org/10.1093/NAR/GKL343.

[55] I. Dimitrov, I. Bangov, D.R. Flower, I. Doytchinova, AllerTOP v.2—a server for in silico prediction of allergens, J. Mol. Model. 20 (2014), https://doi.org/10.1007/S00894-014-2278-5.

[56] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, et al., Protein identification and analysis tools on the ExPASy server, The Proteomics Protocols Handbook (2005) 571–607, https://doi.org/10.1385/1-59259-890-0, 571.

[57] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinf. 9 (2008) 1–8, https://doi.org/10.1186/1471-2105-9-40/FIGURES/4.

[58] W. Zheng, Y. Li, C. Zhang, X. Zhou, R. Pearce, E.W. Bell, et al., Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14, Proteins: Struct., Funct., Bioinf. 89 (2021) 1734–1751, https://doi.org/10.1002/prot.26193.

[59] L. Heo, H. Park, C. Seok, GalaxyRefine: protein structure refinement driven by side-chain repacking, Nucleic Acids Res. 41 (2013) W384–W388, https://doi.org/10.1093/nar/gkt458.

[60] N. Rapin, O. Lund, M. Bernaschi, F. Castiglione, Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, PLoS One 5 (2010) e9862, https://doi.org/10.1371/JOURNAL.PONE.0009862.

[61] H. Al Tbeishat, Novel in Silico mRNA vaccine design exploiting proteins of M. tuberculosis that modulates host immune responses by inducing epigenetic modifications, Sci. Rep. 12 (2022), https://doi.org/10.1038/s41598-022-08506-4.

[62] N. Naidoo, S. Ramsugit, M. Pillay, *Mycobacterium tuberculosis* pili (MTP), a putative biomarker for a tuberculosis diagnostic test, Tuberculosis 94 (2014) 338–345, https://doi.org/10.1016/j.tube.2014.03.004.

[63] N. Naidoo, B. Pillay, M. Bubb, A. Pym, T. Chiliza, K. Naidoo, et al., Evaluation of a synthetic peptide for the detection of anti-*Mycobacterium tuberculosis* curli pili IgG antibodies in patients with pulmonary tuberculosis, Tuberculosis 109 (2018) 80–84, https://doi.org/10.1016/j.tube.2018.01.007.

[64] A.P. Junqueira-Kipnis, L.M. Marques Neto, A. Kipnis, Role of fused *Mycobacterium tuberculosis* immunogens and adjuvants in modern tuberculosis vaccines, Front. Immunol. 5 (2014), https://doi.org/10.3389/FIMMU.2014.00188.

[65] F. Fu, H. Tian, X. Li, Y. Lang, G. Tong, S. Liu, et al., C-terminal heat shock protein 70 of *Mycobacterium tuberculosis* as a molecular adjuvant for DNA vaccination with the porcine circovirus type 2 ORF2 (capsid) gene in mice, Vet. J. 195 (2013) 244–247, https://doi.org/10.1016/J.TVJL.2012.06.005.

[66] M. Tebianian, A.Z. Hoseini, S.M. Ebrahimi, A. Memarnejadian, A.R. Mokarram, M. Mahdavi, et al., Cloning, expression, and immunogenicity of novel fusion protein of *Mycobacterium tuberculosis* based on ESAT-6 and truncated C-terminal fragment of HSP70, Biologicals 39 (2011) 143–148, https://doi.org/10.1016/j.biologicals.2011.02.002.

[67] Lehner T, Wang Y, Whittall T, Mcgowan E, Kelly CG, Singh M. Heat Shock Proteins and Modulation of Cellular Function Functional Domains of HSP70 Stimulate Generation of Cytokines and Chemokines, Maturation of Dendritic Cells and Adjuvanticity. n.d.

[68] J. Ellis, Proteins as molecular chaperones, Nature 328 (1987) 378–379, https://doi.org/10.1038/328378A0.

[69] R. Freudl, Signal peptides for recombinant protein secretion in bacterial expression systems, Microb. Cell Factories 17 (2018) 52, https://doi.org/10.1186/S12934-018-0901-3.

[70] A.L. Schaap-Johansen, M. Vujović, A. Borch, S.R. Hadrup, P. Marcatili, T cell epitope prediction and its application to immunotherapy, Front. Immunol. 12 (2021) 2994, https://doi.org/10.3389/FIMMU.2021.712488/BIBTEX.

[71] Z. Dai, Z. Liu, B. Xiu, X. Yang, P. Zhao, X. Zhang, et al., A multiple-antigen detection assay for tuberculosis diagnosis based on broadly reactive polyclonal antibodies, Iran J Basic Med Sci 20 (2017) 360, https://doi.org/10.22038/IJBMS.2017.8575.

[72] B. Pathakumari, M. Prabhavathi, D. Anbarasu, P. Paramanandhan, A. Raja, Dynamic IgG antibody response to immunodominant antigens of M. tuberculosis for active TB diagnosis in high endemic settings, Clin. Chim. Acta 461 (2016) 25–33, https://doi.org/10.1016/J.CCA.2016.06.033.

[73] D. Jaganath, J. Rajan, C. Yoon, R. Ravindran, A. Andama, L. Asege, et al., Evaluation of multi-antigen serological screening for active tuberculosis among people living with HIV, PLoS One 15 (2020), https://doi.org/10.1371/JOURNAL.PONE.0234130.

[74] P.D. Burbelo, J. Keller, J. Wagner, J.S. Klimavicz, A. Bayat, C.S. Rhodes, et al., Serological diagnosis of pulmonary *Mycobacterium tuberculosis* infection by LIPS using a multiple antigen mixture, BMC Microbiol. 15 (2015), https://doi.org/10.1186/s12866-015-0545-y.

[75] X. Fan, X. Zhao, R. Wang, M. Li, X. Luan, R. Wang, et al., A novel multistage antigens ERA005f confer protection against *Mycobacterium tuberculosis* by driving Th-1 and Th-17 type T cell immune responses, Front. Immunol. 14 (2023) 1276887, https://doi.org/10.3389/FIMMU.2023.1276887.

[76] W. Gong, Y. Liang, J. Mi, Z. Jia, Y. Xue, J. Wang, et al., Peptides-based vaccine MP3RT induced protective immunity against *Mycobacterium tuberculosis* infection in a humanized mouse model, Front. Immunol. 12 (2021), https://doi.org/10.3389/fimmu.2021.666290.

[77] P. Cheng, Y. Xue, J. Wang, Z. Jia, L. Wang, W. Gong, Evaluation of the consistence between the results of immunoinformatics predictions and real-world animal experiments of a new tuberculosis vaccine MP3RT, Front. Cell. Infect. Microbiol. 12 (2022), https://doi.org/10.3389/fcimb.2022.1047306.

[78] P. Cheng, F. Jiang, G. Wang, J. Wang, Y. Xue, L. Wang, et al., Bioinformatics analysis and consistency verification of a novel tuberculosis vaccine candidate HP13138PB, Front. Immunol. 14 (2023), https://doi.org/10.3389/fimmu.2023.1102578.

[79] F. Jiang, C. Peng, P. Cheng, J. Wang, J. Lian, W. Gong, PP19128R, a multiepitope vaccine designed to prevent latent tuberculosis infection, induced immune responses in silico and in vitro assays, Vaccines (Basel) 11 (2023) 856, https://doi.org/10.3390/vaccines11040856.

[80] P. Andersen, P. Andersen, Host responses and antigens involved in protective immunity to *Mycobacterium tuberculosis*, Scand. J. Immunol. 45 (1997) 115–131, https://doi.org/10.1046/J.1365-3083.1997.D01-380.X.

[81] J. Zhou, J. Chen, Y. Peng, Y. Xie, Y. Xiao, A promising tool in serological diagnosis: current research progress of antigenic epitopes in infectious diseases, Pathogens 11 (2022) 1095, https://doi.org/10.3390/PATHOGENS11101095, 2022;11:1095.

[82] W. Gong, C. Pan, P. Cheng, J. Wang, G. Zhao, X. Wu, Peptide-based vaccines for tuberculosis, Front. Immunol. 13 (2022), https://doi.org/10.3389/FIMMU.2022.830497.

[83] V. Vinod, S. Vijayrajratnam, A.K. Vasudevan, R. Biswas, The cell surface adhesins of *Mycobacterium tuberculosis*, Microbiol. Res. 232 (2020) 126392, https://doi.org/10.1016/j.micres.2019.126392.

[84] B.T. Andongma, Y. Huang, F. Chen, Q. Tang, M. Yang, S.-H. Chou, et al., In silico design of a promiscuous chimeric multi-epitope vaccine against *Mycobacterium tuberculosis*, Comput. Struct. Biotechnol. J. 21 (2023) 991–1004, https://doi.org/10.1016/j.csbj.2023.01.019.

[85] M. Keikha, M. Karbalaei, K. Ghazvini, In silico design of multi-epitope ESAT-6:Ag85b:fcγ2a fusion protein as a novel candidate for tuberculosis vaccine, Arch Clin Infect Dis 15 (2020), https://doi.org/10.5812/archcid.90449.

[86] J. Wang, T. Xie, I. Ullah, Y. Mi, X. Li, Y. Gong, et al., A VLP-based vaccine displaying HBHA and MTP antigens of *Mycobacterium tuberculosis* induces protective immune responses in M. tuberculosis H37Ra infected mice, Vaccines (Basel) 11 (2023), https://doi.org/10.3390/vaccines11050941.

[87] A.D. White, A.C. Tran, L. Sibley, C. Sarfas, A.L. Morrison, S. Lawrence, et al., Spore-FP1 tuberculosis mucosal vaccine candidate is highly protective in Guinea pigs but fails to improve on BCG-conferred protection in non-human primates, Front. Immunol. 14 (2023) 1246826, https://doi.org/10.3389/fimmu.2023.1246826.

[88] A. Forouharmehr, Whole proteome screening to develop a potent epitope-based vaccine against Coxiella burnetii: a reverse vaccinology approach, J. Biomol. Struct. Dyn. (2024) 1–13, https://doi.org/10.1080/07391102.2024.2326198.

[89] J. Tang, Y. Huang, Z. Cai, Y. Ma, Mycobacterial heparin-binding hemagglutinin (HBHA)-induced interferon-γ release assay (IGRA) for discrimination of latent and active tuberculosis: a systematic review and meta-analysis, PLoS One 16 (2021) e0254571, https://doi.org/10.1371/journal.pone.0254571.

[90] E. Nemes, H. Geldenhuys, V. Rozot, K.T. Rutkowski, F. Ratangee, N. Bilek, et al., Prevention of M. tuberculosis infection with H4:IC31 vaccine or BCG revaccination, N. Engl. J. Med. 379 (2018) 138–149, https://doi.org/10.1056/NEJMoa1714021.

[91] R. Lai, A.F. Ogunsola, T. Rakib, S.M. Behar, Key advances in vaccine development for tuberculosis-success and challenges, NPJ Vaccines 8 (2023) 158, https://doi.org/10.1038/s41541-023-00750-7.

[92] F. Guo, J. Wei, Y. Song, B. Li, Z. Qian, X. Wang, et al., Immunological effects of the PE/PPE family proteins of *Mycobacterium tuberculosis* and related vaccines, Front. Immunol. 14 (2023), https://doi.org/10.3389/FIMMU.2023.1255920.

[93] M. Strong, M.R. Sawaya, S. Wang, M. Phillips, D. Cascio, D. Eisenberg, Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*, Proc Natl Acad Sci U S A 103 (2006) 8060–8065, https://doi.org/10.1073/PNAS.0602606103/SUPPL_FILE/02606FIG9.JPG.

[94] W. Fleri, K. Vaughan, N. Salimi, R. Vita, B. Peters, A. Sette, The immune epitope database: how data are entered and retrieved, J Immunol Res 2017 (2017), https://doi.org/10.1155/2017/5974574.

[95] K. Pillay, T. Coetzer, C. Connolly, B. Pillay, T. Chiliza, K. Naidoo, et al., Diversity of IgG Antibody Response to Mycobacterium tuberculosis Curli Pili (MTP) in HIV-Positive and Negative Patients from Different Geographical Regions in Sub-saharan Africa, 2024.

[96] J. Jiang, D. Xie, W. Zhang, G. Xiao, J. Wen, Fusion of Hsp70 to Mage-A1 Enhances the Potency of Vaccine-specific Immune Responses, 2013.

[97] W.S. Kim, J.-S. Kim, H.M. Kim, K.W. Kwon, S.-Y. Eum, S.J. Shin, Comparison of immunogenicity and vaccine efficacy between heat-shock proteins, HSP70 and GrpE, in the DnaK operon of *Mycobacterium tuberculosis*, Sci. Rep. 8 (2018) 14411, https://doi.org/10.1038/s41598-018-32799-z.

[98] W.S. Kim, J.-S. Kim, H.M. Kim, K.W. Kwon, S.-Y. Eum, S.J. Shin, Comparison of immunogenicity and vaccine efficacy between heat-shock proteins, HSP70 and GrpE, in the DnaK operon of *Mycobacterium tuberculosis*, Sci. Rep. 8 (2018) 14411, https://doi.org/10.1038/s41598-018-32799-z.

[99] X. Li, X. Yang, L. Li, H. Liu, J. Liu, A truncated C-terminal fragment of *Mycobacterium tuberculosis* HSP70 gene enhanced potency of HBV DNA vaccine, Vaccine 24 (2006) 3321–3331, https://doi.org/10.1016/J.VACCINE.2006.01.012.

[100] C. Ghazaei, Role and mechanism of the Hsp70 molecular chaperone machines in bacterial pathogens, J. Med. Microbiol. 66 (2017) 259–265, https://doi.org/10.1099/JMM.0.000429/CITE/REFWORKS.

[101] H. Diaz-Silvestre, P. Espinosa-Cueto, A. Sanchez-Gonzalez, M.A. Esparza-Ceron, A.L. Pereira-Suarez, G. Bernal-Fernandez, et al., The 19-kDa antigen of *Mycobacterium tuberculosis* is a major adhesin that binds the mannose receptor of THP-1 monocytic cells and promotes phagocytosis of mycobacteria, Microb. Pathog. 39 (2005) 97–107, https://doi.org/10.1016/J.MICPATH.2005.06.002.

[102] E.H. Noss, R.K. Pai, T.J. Sellati, J.D. Radolf, J. Belisle, D.T. Golenbock, et al., Toll-like receptor 2-dependent inhibition of macrophage class II MHC expression and antigen processing by 19-kDa lipoprotein of *Mycobacterium tuberculosis*, J. Immunol. 167 (2001) 910–918, https://doi.org/10.4049/JIMMUNOL.167.2.910.

[103] A. Sánchez, P. Espinosa, T. García, R. Mancilla, The 19 kDa *Mycobacterium tuberculosis* lipoprotein (LpqH) induces macrophage apoptosis through extrinsic and intrinsic pathways: a role for the mitochondrial apoptosis-inducing factor, Clin. Dev. Immunol. 2012 (2012) 11, https://doi.org/10.1155/2012/950503.

[104] V.V. Yeremeev, I.V. Lyadova, B.V. Nikonenko, A.S. Apt, C. Abou-Zeid, J. Inwald, et al., The 19-kD antigen and protective immunity in a murine model of tuberculosis, Clin. Exp. Immunol. 120 (2000) 274, https://doi.org/10.1046/J.1365-2249.2000.01212.X.

[105] F.A. Post, C. Manca, O. Neyrolles, B. Ryffel, D.B. Young, G. Kaplan, *Mycobacterium tuberculosis* 19-kilodalton lipoprotein inhibits Mycobacterium smegmatis-induced cytokine production by human macrophages in vitro, Infect. Immun. 69 (2001) 1433–1439, https://doi.org/10.1128/IAI.69.3.1433-1439.2001.

[106] D. Hu, J. Wu, R. Zhang, L. Chen, Z. Chen, X. Wang, et al., Autophagy-targeted vaccine of LC3-LpqH DNA and its protective immunity in a murine model of tuberculosis, Vaccine 32 (2014) 2308–2314, https://doi.org/10.1016/j.vaccine.2014.02.069.

[107] A. Palacios, L. Sampedro, I.A. Sevilla, E. Molina, D. Gil, M. Azkargorta, et al., *Mycobacterium tuberculosis* extracellular vesicle-associated lipoprotein LpqH as a potential biomarker to distinguish paratuberculosis infection or vaccination from tuberculosis infection, BMC Vet. Res. 15 (2019), https://doi.org/10.1186/S12917-019-1941-6.

[108] S. Wehmeier, A.S. Varghese, S.S. Gurcha, B. Tissot, M. Panico, P. Hitchen, et al., Glycosylation of the phosphate binding protein, PstS, in Streptomyces coelicolor by a pathway that resembles protein O-mannosylation in eukaryotes, Mol. Microbiol. 71 (2009) 421–433, https://doi.org/10.1111/J.1365-2958.2008.06536.X.

[109] A. Tanghe, P. Lefevre, O. Denis, S. D'Souza, M. Braibant, E. Lozes, et al., Immunogenicity and protective efficacy of tuberculosis DNA vaccines encoding putative phosphate transport receptors, J. Immunol. 162 (1999) 1113–1119, https://doi.org/10.4049/JIMMUNOL.162.2.1113.

[110] A. Davidow, G.V. Kanaujia, L. Shi, J. Kaviar, X.D. Guo, N. Sung, et al., Antibody profiles characteristic of *Mycobacterium tuberculosis* infection state, Infect. Immun. 73 (2005) 6846, https://doi.org/10.1128/IAI.73.10.6846-6851.2005.

[111] S. Khurshid, R. Khalid, M. Afzal, M. Waheed Akhtar, Truncation of PstS1 Antigen of *Mycobacterium tuberculosis* Improves Diagnostic Efficiency, 2013, https://doi.org/10.1016/j.tube.2013.07.005.

[112] S. Khurshid, M. Afzal, R. Khalid, I.H. Khan, M. Waheed Akhtar, Improving sensitivity for serodiagnosis of tuberculosis using TB16.3-echA1 fusion protein, Tuberculosis 94 (2014) 519–524, https://doi.org/10.1016/j.tube.2014.06.006.

[113] C. Peng, C. Shi, X. Cao, Y. Li, F. Liu, F. Lu, Factors influencing recombinant protein secretion efficiency in gram-positive bacteria: signal peptide and beyond, Front. Bioeng. Biotechnol. 7 (2019) 448926, https://doi.org/10.3389/FBIOE.2019.00139/BIBTEX.

[114] I. Dimitrov, D.R. Flower, I. Doytchinova, AllerTOP–a server for in silico prediction of allergens, BMC Bioinf. 14 (Suppl 6) (2013), https://doi.org/10.1186/1471-2105-14-S6-S4.

[115] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326, https://doi.org/10.1016/0003-2697(89)90602-7.

[116] J.L. Meitzler, S. Hinde, B. Bánfi, W.M. Nauseef, P.R.O. De Montellano, Conserved cysteine residues provide a protein-protein interaction surface in dual oxidase (DUOX) proteins, J. Biol. Chem. 288 (2013) 7147, https://doi.org/10.1074/JBC.M112.414797.

[117] A.A. Bhopatkar, V.N. Uversky, V. Rangachari, Granulins modulate liquid–liquid phase separation and aggregation of the prion-like C-terminal domain of the neurodegeneration-associated protein TDP-43, J. Biol. Chem. 295 (2020) 2506, https://doi.org/10.1074/JBC.RA119.011501.

[118] D. Gront, S. Kmiecik, M. Blaszczyk, D. Ekonomiuk, A. Koliński, Optimization of protein models, WIREs Computational Molecular Science 2 (2012) 479–493, https://doi.org/10.1002/wcms.1090.
[119] Laskowski RomanA, Rullmann JantoonC, MacArthur MalcolmW, R. Kaptein, JanetM. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR 8 (1996), https://doi.org/10.1007/BF00228148.
[120] M. Aarthy, S.K. Singh, Envisaging the conformational space of proteins by coupling machine learning and molecular dynamics, in: Advances in Protein Molecular and Structural Biology Methods, Elsevier, 2022, pp. 467–475, https://doi.org/10.1016/B978-0-323-90264-9.00028-3.
[121] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (1993) 1511–1519, https://doi.org/10.1002/pro.5560020916.