CrossMark

Identifying and characterising PPE7 (Rv0354c) high activity binding peptides and their role in inhibiting cell invasion

Diana P. Díaz[1,2] · Marisol Ocampo[1,2] · Yahson Varela[1,2] · Hernando Curtidor[1,2] ·
Manuel A. Patarroyo[1,2] · Manuel E. Patarroyo[1,3]

Abstract This study was aimed at characterising the PPE7 protein from the PE/PPE protein family. The presence and transcription of the *rv0354c* gene in the *Mycobacterium tuberculosis* complex was determined and the subcellular localisation of the PPE7 protein on mycobacterial membrane was confirmed by immunoelectron microscope. Two peptides were identified as having high binding activity (HABPs) and were tested in vitro regarding the invasion of *Mycobacterium tuberculosis* H37Rv. HABP 39224 inhibited invasion in A549 epithelial cells and U937 macrophages by more than 50%, whilst HABP 39225 inhibited invasion by 40% in U937 cells. HABP 39224, located in the protein's *C*-terminal region, has a completely conserved amino acid sequence in *M. tuberculosis* complex species and could be selected as a base peptide when designing a subunit-based, anti-tuberculosis vaccine.

Keywords *Mycobacterium tuberculosis* H37Rv · PPE7 protein · Synthetic peptide · High activity binding peptide

✉ Marisol Ocampo
  Marisol.ocampo@urosario.edu.co

1 Fundación Instituto de Inmunología de Colombia (FIDIC), 111321 Bogotá, Colombia

2 Universidad del Rosario, 111321 Bogotá, Colombia

3 Universidad Nacional de Colombia, 11001 Bogotá, Colombia

Introduction

Tuberculosis (TB) is the disease caused by the bacillus *Mycobacterium tuberculosis* (*Mtb*); it is the main cause of death around the world when combining with the acquired immunodeficiency syndrome (AIDS) produced by the human immunodeficiency virus (HIV). The most recent statistics have reported 10.4 million new cases and 1.8 million directly related deaths in 2015, of which 1.4 million were HIV-negative individuals, meaning that it is vital to develop new and more efficient diagnostic methods, drugs and vaccines [1]. The pertinent characteristics obtained from the genome of pathogenic mycobacteria can provide important information for such development; this is the case for the PE/PPE protein family which constitutes around 10% of the genome [2]. It is characterised by the presence of proline-glutamic acid (PE) in positions 8 and 9 in the *N*-terminal domain and proline-proline-glutamic acid (PPE) in positions 7 and 9 in the highly conserved *N*-terminal extreme. This family of proteins could play an important biological function since they belong to the genus *Mycobacterium* and are only present in pathogenous mycobacteria [3]. It has been suggested that they could be a virulence factor [4, 5] and that they are involved in antigen variation, participating in avoiding the immune response; their cellular location is on mycobacterial membrane, wall, or culture supernatant [6–11]. They have been functionally linked in terms of pathogen–host interaction, based on the evidence that some of these proteins interact with host cell receptors, such as Toll-like receptor-2 [12–15].

In this respect, our studies have been focused on finding 20 amino acid-long high activity binding peptides (HABPs) derived from different *Mtb* H37Rv proteins which could be involved in the pathogen–host interaction. The concept of producing a subunit, chemically synthesised, multi-antigen

vaccine, derived from mycobacterial surface proteins' conserved regions was born in mind, following the methodology proposed by Patarroyo et al., for developing vaccines [16]. Our group has thus studied *Mtb* H37Rv surface proteins, finding sequences from peptides specifically binding to infection target cells and which impede bacillus invasion in in vitro assays, thereby representing promising candidates for designing an anti-TB vaccine [17].

The PPE7 protein is encoded by the *rv0354c* gene, has a 14.39 kDa molecular weight and a 141 amino acid (aa) sequence. It has been identified (by LC-MS) in the lungs of *Mtb*-infected guinea pigs 90 days after being exposed [18]. The *rv0354c* gene is expressed in response to antibiotics, identified by microarray studies as being downregulated after 96 h of nutrient starvation [19]. The PPE7 protein is different regarding the H37Ra and H37Rv strains due to nucleotide insertions, deletions and substitutions, meaning that the protein has 141 aa in the H37Rv strain but 181 aa in the H37Ra strain. Such changed aa composition between H37Rv and H37Ra can alter these proteins' solubility; a ~ 10% change in aliphatic index has been observed for PPE7 [20]; on the other hand, the *C*-terminal extension in H37Ra protein (MRA_0363) renders it less hydrophobic, as indicated by a decreased positive grand average hydropathy (GRAVY) value for the corresponding protein in H37Rv (Rv0354c).

This study was thus aimed at characterising the PPE7 protein from the host cell interaction point of view and identifying HABPs which could inhibit *Mtb* entry to target cells and therefore be used in a sequential search for different antigens derived from *Mtb* proteins which might be selected when designing a multi-epitope, anti-tuberculosis vaccine.

Materials and methods

In silico analysis of the PPE7 protein

The *Mycobacterium tuberculosis* H37Rv PPE7 protein sequence was obtained from the National Center of Biotechnology Information database http://www.ncbi.nlm.nih.gov/protein/CCP43084.1. The 141 amino acid sequence was recovered in FASTA format for identifying homologous sequences in different *Mycobacterium tuberculosis* complex (MTC) species and strains using the basic local alignment search tool (BLASTp) http://blast.ncbi.nlm.nih.gov/Blast.cgi [21].

The protein's subcellular location was predicted by the following bioinformatics tools: PSORTb v3.0 http://www.psort.org/psortb/results.pl [22], Gpos-mPLoc http://www.csbio.sjtu.edu.cn/cgi-bin/GposmPLoc.cgi [23], PA-SUB 2.5 http://pa.wishartlab.com/pa/pa/index.html [24] and

TBPred http://www.imtech.res.in/raghava/tbpred/ [25]. The following tools were used for ascertaining the presence of a signal sequence and transmembrane domains: SignalP 3.0 http://www.cbs.dtu.dk/services/SignalP-3.0/, [26], Phobius http://phobius.sbc.su.se/index.html [27] and TMHMM version 2.0 http://www.cbs.dtu.dk/services/TMHMM/ [28] and PRED-TMR http://athina.biol.uoa.gr/PRED-TMR/input.html [29]. The protein's secondary structure was predicted by PSIPRED http://bioinf.cs.ucl.ac.uk [30], SOPMA https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html [31] and I-TASSER http://zhanglab.ccmb.med.umich.edu/I-TASSER/ [32].

Mycobacterial species and strains

The following mycobacterial species and strains used here were obtained from ATCC: *M. tuberculosis* H37Rv (ATCC 27294), *M. tuberculosis* H37Ra (ATCC 25177), *M. bovis* (ATCC 19210) and *M. bovis* BCG (ATCC 27291, Pasteur substrain). All mycobacteria were grown for 5–30 days in Middlebrook 7H9 (Difco Laboratories, Detroit MI) supplemented with oleic acid, albumin, dextrose, NaCl (10% OADC) and incubated at their optimum temperature until cultures reached 0.5–1.0 OD_{600}. Mycobacteria were harvested at mid- to late-log phase culture by spinning at $12,500 \times g$ for 30 min at 4 °C, suspended in PBS and stored at −20 °C.

The *rv0354c* gene presence and transcription

Genomic DNA (gDNA) was isolated from mycobacterial species and strains using an Ultra Clean Microbial DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA), following the manufacturer's instructions. Extracted DNA quality was assessed by amplifying a 439 base pair (bp) fragment from the *hsp65* gene using forward primer Tb11 (5′-ACCAACGATGGTGTGTCCAT-3′) and reverse primer Tb12 (5′-CTTGTCGAACCGCATACCCT-3′). PCR amplification was used for assessing *rv0354c* gene presence and transcription in in vitro culture conditions in all mycobacteria tested, using the following primers: *rv0354*-sense: 5′-GGGATTCTTCAACTCGACC-3′ and rv0354-anti-sense: 5′-CGAAGTTTTGGAAGCCCG-3′, which amplified a 178 bp fragment.

The PCR assay was carried out in a thermal cycler (Gene Amp PCR system 9600, Perkin-Elmer) by incubating gDNA with a PCR mixture containing: 1 unit of MangoTaq DNA polymerase (Bioline, London, UK), 1X Taq polymerase reaction buffer, 1.5 mM $MgCl_2$, 1mM dNTPs and 1 μM of each primer in a final 25 μL reaction volume. The reaction was carried out in the following conditions: an initial denaturing step at 95 °C for 5 min followed by 35 cycles consisting of: 30 s at 52 °C, 30 s at 72 °C and 30 s at

95 °C. A final extension cycle was performed at 72 °C for 5 min. DNase- and RNase-free water was used as negative PCR control for all reactions. All amplifications were visualised on 1.5% agarose gel stained with SYBR Safe (Invitrogen, Carlsbad, CA).

Total RNA was isolated from the bacterial pellet by homogenisation in 1 mL Trizol reagent (Invitrogen), following the manufacturer's recommendations. 5 µg RNA was treated with 1 unit of amplification grade deoxyribonuclease I (DNase I) (Invitrogen), according to the manufacturer's recommendations. Complementary DNA (cDNA) was synthesised using SuperScript III reverse transcriptase (Invitrogen), using 250 ng random hexamers for increasing PCR product yield (following the manufacturer's recommendations). SuperScript III enzyme was replaced by DEPC-treated water as negative synthesis control and included for each sample. Two µL cDNA were used as template for PCR amplification following the conditions previously described for DNA. The *hsp65* gene was used as positive transcription control.

Mtb PPE7 protein localisation

Two New Zealand rabbits were inoculated with 800 µg of PPE7 protein peptide–synthetic polymer mixture; 39212 (CGHRAAGAGRRQRRRRSGDGQWRCG) and 39214 (CGGFFNSTTTPSSGFFNSGAGGGC), selected by Bepipred 1.0 http://www.cbs.dtu.dk/services/BepiPred/ [33] as B-cell epitopes when analysing such protein's sequence. The peptides were mixed with incomplete Freund's adjuvant (Sigma, St. Louis, US) (1:1) and administered three times, with a 20-day interval between each inoculation. Polyclonal serum was collected from the animals on day 0 (Pre-immune) and post-third inoculation on day 60.

The PPE7 protein was localised in the mycobacteria by immunoelectron microscopy (IEM), using the antibodies so produced. The *Mtb* H37Rv bacillus, from log phase culture, was fixed (4%-p-formaldehyde-0.5% glutaraldehyde), embedded in LR White acrylic resin and cut into 400 nm strips. The strips were incubated with the primary antibody (polyclonal serum anti-peptide) and then washed (0.5% BSA 0.4% Tween in PBS 1X) and incubated with secondary antibody [10 nm colloidal gold particles coupled IgG anti-rabbit (1:200)] (Sigma, St. Louis, US); 6% uranyl acetate was used as contrast stain in IEM and samples were visualised by transmission microscope. Pre-immune and hyperimmune sera were used as control [11, 17, 34, 35].

Peptide synthesis and radiolabelling

Twenty amino acid-long synthetic peptides, covering the whole PPE7 protein sequence, coded 39219–39225, were obtained by solid phase synthesis [36], using t-Boc aa and MBHA resin (0.5 meqv/g). The peptides were deprotected and cleaved by low–high HF technique, purified by reverse phase, high-performance liquid chromatography (RP-HPLC) and characterised by matrix-assisted laser desorption–ionisation time-of-flight (MALDI-TOF) mass spectrometry, using α-cyano-4-hydroxycinnamic acid (α-CCA) as matrix.

Each peptide was radiolabelled with 5 µL NaI^{125} (100 mCi/mL) and 15 µL chloramine T (2.8 mg/mL); the reaction was stopped with 15 µL sodium metabisulfite (2.3 mg/mL). The mixture was separated by gel filtration on a Sephadex G-10 column; the eluted fractions corresponding to the radiolabelled peptides were analysed on a gamma counter (Gamma Counter Cobra II, Packard Instrument Co., Meriden, CT, USA). Tyrosine was added to the amino terminal of those which lacked this residue in their sequence, thereby enabling radiolabelling.

PPE7 peptide binding to A549 and U937 cells

A549 alveolar epithelial cells (ATCC CLL-195) and U937 monocyte-derived macrophages (ATCC CRL-2367), at 1.2×10^6 cells/well concentration, in 96-well microplates, were incubated with radiolabelled peptide concentrations ranging from 0 up to −950 nM for 90 min at 4 °C. Total binding was equivalent to radiolabelled peptide–cell binding; inhibited binding was obtained by adding an excess (40 µM) of the same non-radiolabelled peptide. Specific binding was defined as the difference between binding total and inhibited binding. Cell-associated radioactivity was measured by gamma counter. A HABP was considered to be a peptide having high specific binding activity (i.e. when the slope of the line on the specific binding graph regarding added peptide was greater than or equal to one) [17, 32, 35].

The saturation assay, incubating 1×10^6 cells/well with increasing amounts of radiolabelled HABP (0–8000 nM) in the absence or presence of non-radiolabelled peptide, led to determining the dissociation constants (K_D), the Hill coefficients (n_H) and the binding sites per cell for some HABPs. The binding and saturation assays were done in triplicate.

The peptides' structural determination

Circular dichroism (CD) was used for PPE7 protein peptide conformational analysis. Peptide spectra at 0.1 mM concentration in 30% trifluoroethanol (TFE) (v/v) were taken at 260–190 nm wavelength at 0.2 nm intervals on a JASCO J-810 spectropolarimeter. The data were acquired in ellipticity, given the symbol θ, and measured in millidegrees (mdeg); θ ellipticity was converted to molar ellipticity to analyse this data and given the symbol $[\theta]$ (with units of

degrees.cm squared. per decimole). SELCON3, CDSSTR and CONTIN-LL [38] informatics softwares were used for data deconvolution, thereby allowing each peptide's structural elements to be estimated, expressed as percentages of α-helices, β-sheets and random coils.

HABP ability to inhibit target cell invasion

The U937 and A549 cell lines (2.5×10^5 cells/well) were incubated with each HABP at 2, 20 and 200 μM concentrations for 2 h at 37 °C and 5% CO_2. Cells without HABPs were taken as invasion controls and cytochalasin (30 μM) as inhibition control (all assays done in triplicate). The macrophages were incubated at 4 °C for reducing their phagocytic action. They were infected with *Mtb* H37Rv using a 1:10 *Mtb*/cell multiplicity of infection (MOI); 16 h after being infected they were washed three times with Hank's balanced salt solution (HBSS), eliminating extracellular bacteria, and the cells were lysed with water for 20 min. The intracellular bacteria from each well were sown in Middlebrook agar 7H10, in triplicate, for CFU count. The count was made 20 days later for each HABP concentration and compared to invasion and inhibition controls [38].

Percentage invasion inhibition was calculated by comparing average UFC in infection control (estimated with 100% invasion) to the triplicate of each peptide concentration used.

PPE7 protein HABPs' cytotoxic effect was also evaluated for cell viability using an MTT kit (MTT cell proliferation assay, ATCC). U937 and A549 cells (at 50,000 cells/well concentration in 96-well microplates) were incubated at 37 °C, 5% CO_2 for 2 h with each peptide at 20 and 200 μM concentration, using 1% Triton X-100 as toxicity control. Following the incubation period, 10 μL MTT was added to the cells with the peptide and incubated for 4 h. The formazan crystals were then dissolved with detergent (sodium dodecyl sulphate) overnight; readings were taken at 570 nm wavelength.

Results

Bioinformatics analysis of the protein

BLAST was used for aligning the *Mycobacterium tuberculosis* (*Mtb*) H37Rv PPE7 protein aa sequence (Gen-Bank: CCP43084.1), the reference strain used in our laboratory, giving 99% protein identity with *Mycobacterium tuberculosis* complex strains. *M. tuberculosis* CDC1551, F11, T85, GM1503, T17 and T9 sequences had 99% identity, as did other strains, such as *M. bovis, M. africanum, M. canetti, M caprae* and *M. orygis*, (Supplementary Material 1). The identity data refer to 88% query cover,

excluding the *C*-terminal region. The PPE7 protein had no homology with non-pathogenous strains, thereby coinciding with an important characteristic of the PE/PPE protein family which has only been found in pathogenic mycobacteria.

The Rv0354c protein's *in silico* subcellular localisation was found using different bioinformatics servers. The PSORTb server gave an extracellular localisation (8.91 score; cut-off above 7.5) like Gpos-mPLoc which predicted an extracellular localisation. PA-SUB and TBpred predicted that the protein is bound to a membrane by lipid anchorage. SignalP predicted that the protein had no signal peptide; however, the Secretome 2.0 score (0.7253) indicated that it is secreted by non-classical route [39]. Some results for this protein have already been reported by Vizcaíno et al. [40].

Regarding transmembrane domain prediction, the servers showed that the protein is totally exposed to the exterior and that no transmembrane regions were found.

rv0354c gene presence and transcription

All samples had good gDNA integrity, as shown by the *hsp65* gene amplification at 439 bp (Fig. 1A). A 178 bp amplification product indicated *rv0354c* gene presence in *M. tuberculosis* H37Rv, *M. tuberculosis* H37Ra, *M. bovis* and *M. bovis BCG* (Fig. 1B).

Amplification of the *hsp65* constitutive gene from plus synthesis cDNA was used as transcription control for each mycobacterial strain; the absence of product in minus synthesis confirmed that there was no gDNA contamination (Fig. 1C). Furthermore, the 178 bp amplification fragment observed in plus synthesis for cDNA indicated that *rv0354c* was being transcribed in *M. tuberculosis* H37Rv, *M. tuberculosis* H37Ra, *M. bovis* and *M. bovis BCG* strains in Middlebrook 7H9 culture medium conditions (Fig. 1D).

PPE7 protein immunolocalisation

The PPE7 protein's experimental subcellular localisation was made by immunoelectron microscopy. Figure 2 (upper panels) shows that colloidal gold particles coupled to secondary antibody bound to anti-PPE7 polyclonal antibody which was recognised on the external surface of *Mycobacterium tuberculosis*, thereby supporting the hypothesis that this protein is expressed on mycobacterial surface, as predicted in *in silico* analysis and as is usual in the PE/PPE protein family. No coupled particles were observed in preimmune sera, whilst recognition throughout the bacteria was observed in hyperimmune sera (lower panels).
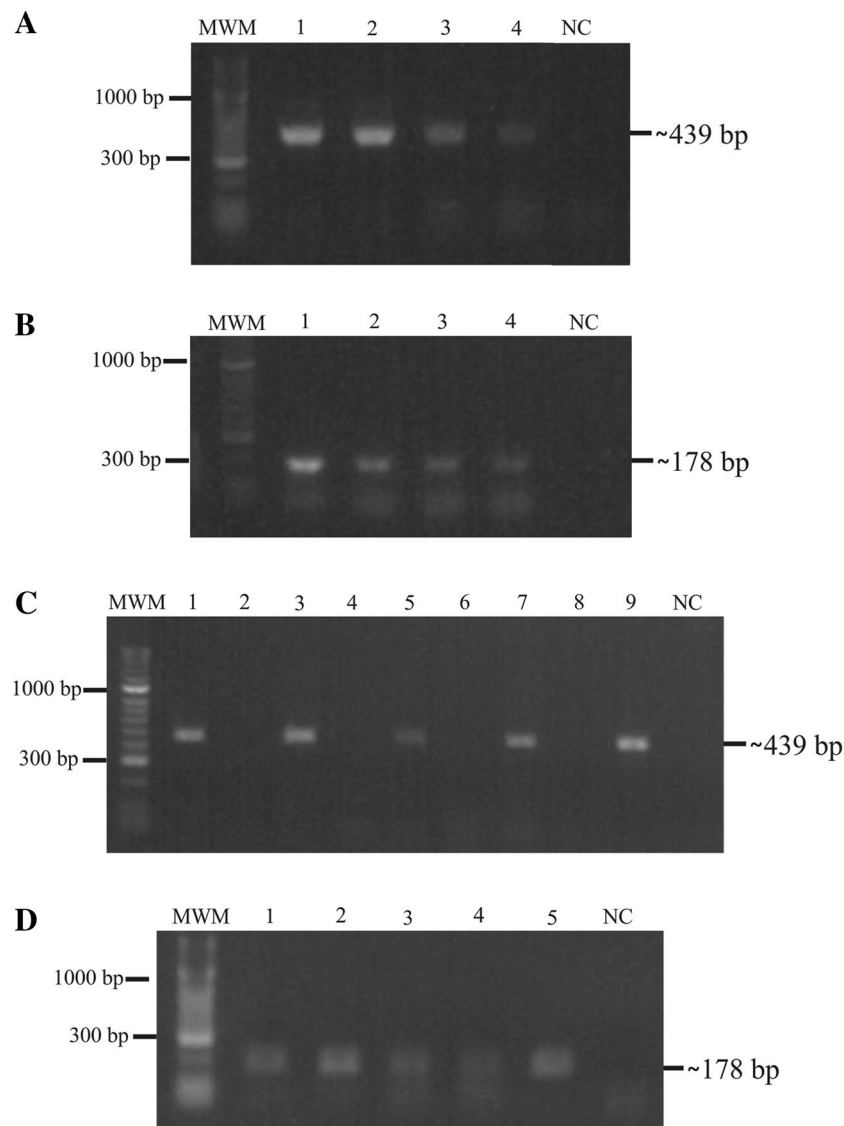
Fig. 1 The *rv0354* gene presence and transcription. **A** *hsp65* gene amplification from gDNA isolated from the species and strains included in this study. *MWM* 1000 bp molecular weight marker; *1 M. tuberculosis* H37Rv; *2 M. tuberculosis* H37Ra; *3 M. bovis; 4 M. bovis BCG;* NC: Negative PCR control. **B** *rv0354c* PCR product amplified from gDNA isolated from different mycobacterial species and strains. *MWM* 1000 bp molecular weight marker; *1 M. tuberculosis* H37Rv; *2 M. tuberculosis* H37Ra; *3 M. bovis; 4 M. bovis BCG; NC* negative PCR control. **C** *hsp65* gene amplification of cDNA from the samples included in this study. *MWM* 1000 bp molecular weight marker; *1 M. tuberculosis* H37Rv plus synthesis; *2 M. tuberculosis* H37Rv minus synthesis; *3 M. tuberculosis* H37Ra plus synthesis; *4 M. tuberculosis* H37Ra minus synthesis; *5 M. bovis* plus synthesis; *6 M. bovis* minus synthesis; *7 M. bovis BCG* plus synthesis; *8 M. bovis BCG* minus synthesis; *9* Positive PCR control (*M. tuberculosis* H37Rv gDNA); *NC* negative PCR control. **D** *rv0354c* gene amplification of cDNA from the samples included in this study. *MWM* 100 bp molecular weight marker; *1 M. tuberculosis* H37Rv plus synthesis; *2 M. tuberculosis* H37Ra plus synthesis; *3 M. bovis* plus synthesis; *4 M. bovis BCG* plus synthesis; *5* Positive PCR control (*M. tuberculosis* H37Rv gDNA); *NC* negative PCR control

Peptides 39224 and 39225 were HABPs

Figure 3A shows the U937 and A549 cell binding activity of peptide sequences forming the PPE7 protein (represented by black bars), their code and position within the protein.

Peptides coded 39224 and 39225 had high specific binding according to the binding assay where two curves were obtained: total binding and inhibited binding. Regarding added peptide and the difference between them (the specific binding line on the graph), slope values greater than 1% enabled the selection of peptides that were considered HABPs. Peptide 39224 ([101]YAAAVSGLGNVFTETSGFFNA[121])

Fig. 2 Immunolocalisation Photographs of *Mycobacterium tuberculosis* immunoelectron microscopy. The PPE7 protein can be observed on *Mtb* surface. The two upper images show post-third inoculation recognition with peptide 39214. The anti-rabbit IgG antibody associated with 10 nm colloidal gold particles is shown by *arrows*. The *lower images* show pre-immune (*lower left-hand image*) and hyperimmune serum controls (*lower right-hand image*)
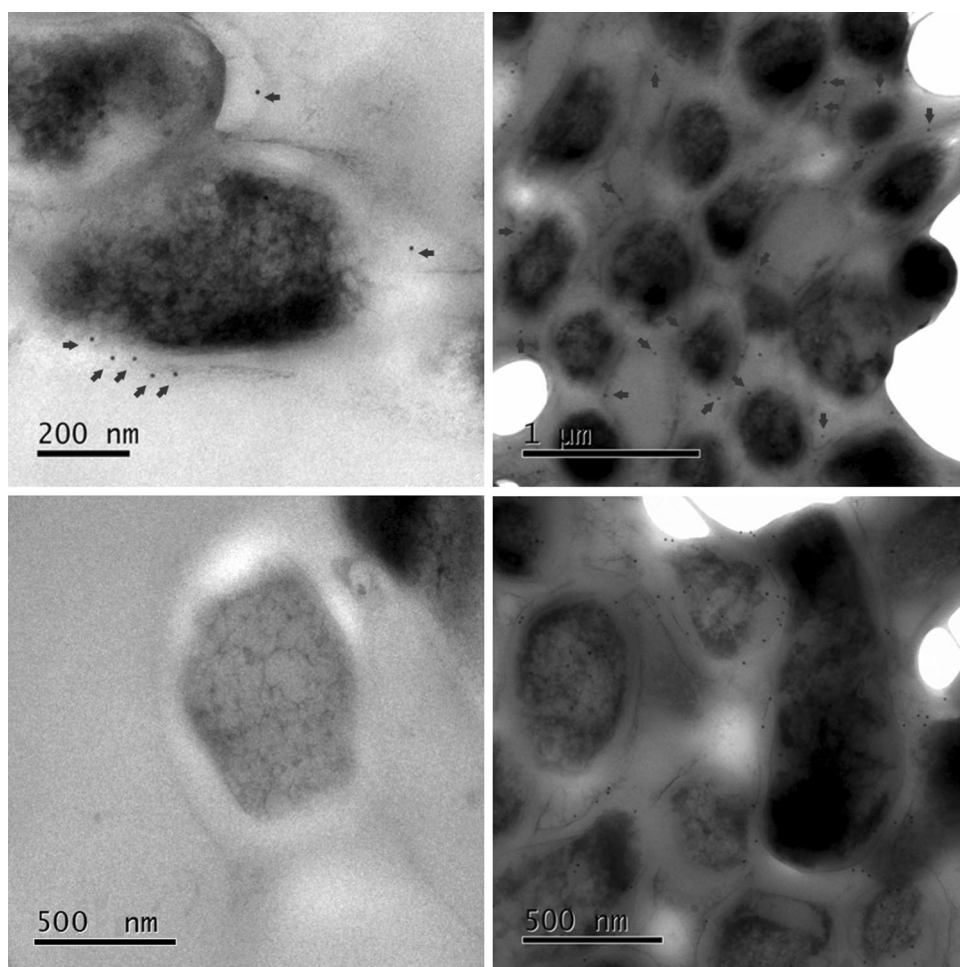
had close to 3% specific binding activity value for both cell types, meaning that it was a HABP for both cell lines (A549 epithelial cells and U937 macrophages: Fig. 3B, left-hand panel). High specific binding for both cell types, probably meant recognition by a common receptor on the cell surface of both cell lines, whilst peptide 39225 (^{122}YGGVGI-RASKTSATTTRAGRT^{141}) was only a HABP for U937 cells, having greater than 3% binding activity (Fig. 3C, left-hand panel).

Saturation assays were used for determining the physicochemical constants regarding peptide–cell interaction for each HABP. In this assay, just as in screening assay, the receptor concentration was constant; increasing concentrations of radiolabelled peptide were added in the presence or absence of an excess of non-radiolabelled peptide. Cell binding sites became occupied by the peptide and specific binding tended to increase to the point where all binding sites were occupied and saturation would be reached; a curve was thus obtained for each peptide with each cell line (Fig. 3B, C, right-hand panel). A 3600 nM dissociation constant (K_D) was determined for peptide 39224 in the interaction with A549 cells whilst this was 1650 nM for

U937 cells, suggesting greater peptide affinity for the latter cells. HABP 39225 had a 2400 nM dissociation constant for U937 cells. The Hill coefficient for the three interactions was greater than 1, indicating positive cooperativity where the peptide's initial binding facilitated the binding of other molecules from the same peptide. Regarding the amount of binding sites per cell, this was 4.64×10^7 for HABP 39224 in A549 and 1.08×10^6 in U937 cells, possibly due to receptors which were different in nature and amount. HABP 39225 had less binding sites per cell for the U937 cell line (8.43×10^5).

The secondary structure of the peptides forming the protein

SELCON 3, CONTIN-LL and CDSSTR deconvolution softwares were used for analysing the peptides' CD far-UV spectra (Fig. 4A). CONTIN-LL software gave the lowest normalised root-mean-square deviation (NRMSD) value (<0.2) when predicting each peptide's structural elements (Supplementary Table 1) and was used for analysing the data.

The programmes gave six possible structures: alpha helix, regular (α_R), alpha helix, distorted (α_D), beta sheet, regular (β_R), beta sheet, distorted (β_D), turn and unordered. Total alpha helix and beta structure values were used for comparing and elucidating the peptides' secondary structure. Peptide 39219 ($_1$MSVTVIYIPFKGTVKHVSVT$_{20}$) had deconvolution percentages throughout all structural elements: 38% beta structure, 13.6% beta turn and 44% random coil, thereby coinciding with the bioinformatics prediction (Fig. 4B); the spectrum was possibly the sum of various structural components (Fig. 4A). Peptides 39220 ($_{21}$IPITTEHLGPYEIDASTINPD$_{41}$), 39221 ($_{42}$YQPID-TAFTQTLDFAGSGTVG$_{61}$) and 39224 ($_{101}$YAAAVSGL-GNVFTETSGFFNA$_{121}$) had alpha helix structure elements (around 90%) as well as characteristic spectra having two minima at 208 and 222 nm and a maxima at 192 nm [41]. Peptide 39222 ($_{62}$YAFPFGFGWQQSPGFFNSTTT$_{81}$) whose deconvolution percentages were close to 70% in alpha helix elements had a particular spectrum, having a strong negative band close to 200 nm and weak positive band close to 226 nm, this being particular to poly-l-proline type II structures (PPII) [42, 43]. Peptide 39225 ($_{122}$YGG-VGIRASKTSATTTRAGRT$_{141}$) had an indefinite spectrum very similar to that of a random coil structure; however, it had 69% alpha helix and 21% random coil.

CD plays an important role in proteins' structural determination and it has been found that synthetic peptides' CD, separately, provides an approach to their structure and the region where they are located. The SOPMA and I-TASSER servers predicted the secondary structure, indicating a disordered tendency throughout the whole sequence and an alpha helix towards the *C*-terminal extreme (Fig. 4B). The Psipred and RaptorX servers confirmed a mostly random coil structure, with some beta sheets throughout the sequence.

The experimental CD results obtained for each PPE7 protein peptide did not agree with the bioinformatics prediction of protein secondary structure which mostly gave random coil elements and some beta sheets (Fig. 4B). Even though no studies have described PPE7 protein 3D structure to date, it is quite probable that the conformation adopted by the peptides together in their native form is very different from when they are folded, as happens with individual peptides.

Mycobacterium tuberculosis invasion of A549 epithelial cells and U937 macrophages

Peptides 39224 and 39225 were tested in each cell line to which they had high specific binding; A549 alveolar epithelial cells and U937 macrophages were infected with *Mtb* H37Rv in the presence of HABPs at increasing concentrations (2, 20 and 200 μM) as well as a cytochalasin inhibition control (30 μM).

Inhibition control for both cell lines inhibited around 80% of mycobacterial entry to the cells; cytochalasin inhibited phagocytosis by actin microfilament polymerisation (Fig. 5). Peptide 39224 in the A549 cell line inhibited mycobacterial entry depending on the concentration; as peptide concentration increased, percentage inhibition also increased, close to 30% for the 20 μM concentration and 60% for the 200 μM concentration. The same peptide in the U937 cell line gave close to 70% invasion inhibition for the two highest concentrations. This peptide's high affinity for both cell lines and the inhibition results suggested that this peptide is fundamental for mycobacterial entry. HABP 39225 had 40% inhibition at 20 μM concentration in U937 cells but did not inhibit at 2 and 200 μM concentrations.

Non-HABP peptide 39223 was used as control in the same assay; it did not have high specific binding activity to either cell line. The result with A549 cells was less than 14% inhibition and did not exceed 20% with U937 cells; this peptide's concentration thus had no direct relationship with inhibition.

Regarding peptide cytotoxicity (Fig. 6), it was seen that peptides 39223 and 39225 had the highest cytotoxicity percentages (close to 20%) in the A549 cell line. However, these did not represent high cytotoxicity values which could affect previous results. The Triton X-100 used as control for the cytotoxicity assay is a non-ionic tensioactive agent which is widely used for membrane solubilisation and cell membrane permeabilisation, thereby leading to rapid cell death by necrosis [44].
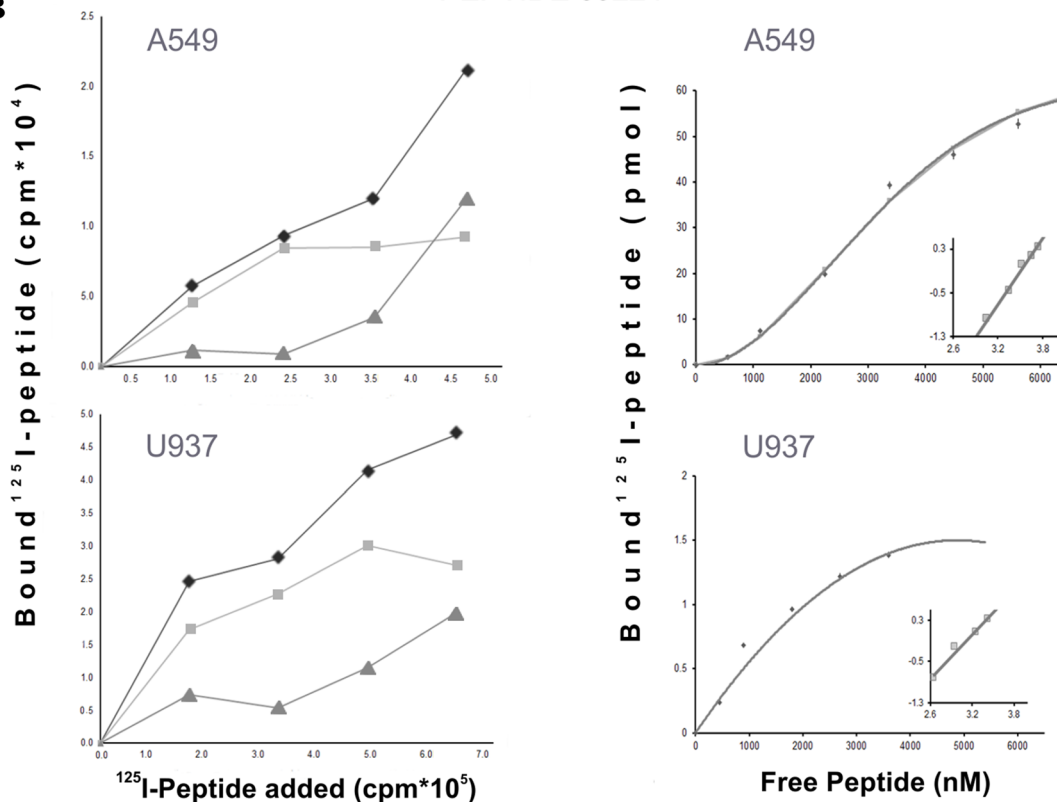
Discussion

The Bacillus Calmette–Guerin (BCG) vaccine has been the only anti-TB vaccine available for almost 100 years now; its limited efficacy and variability regarding the protection it confers on different populations are yet to be explained. The delay concerning diagnosis and the appearance of strains which are resistant to known drugs make developing new ones an urgent need for avoiding this disease's dispersion. The publication of the *Mtb* H37Rv genome [2] led to hopes of great advances regarding knowledge concerning the mycobacteria which would have led to improving diagnostic methods for treatment and developing an effective vaccine against TB. However, the function of many of the genome's proteins is yet to be defined. Those proteins directly implicated in pathogen–host interaction must be recognised in our approach when designing a subunit vaccine against TB [16, 32]. This is why we have characterised a series of *Mtb* H37Rv proteins by identifying regions having a high capability of specifically binding to target cells
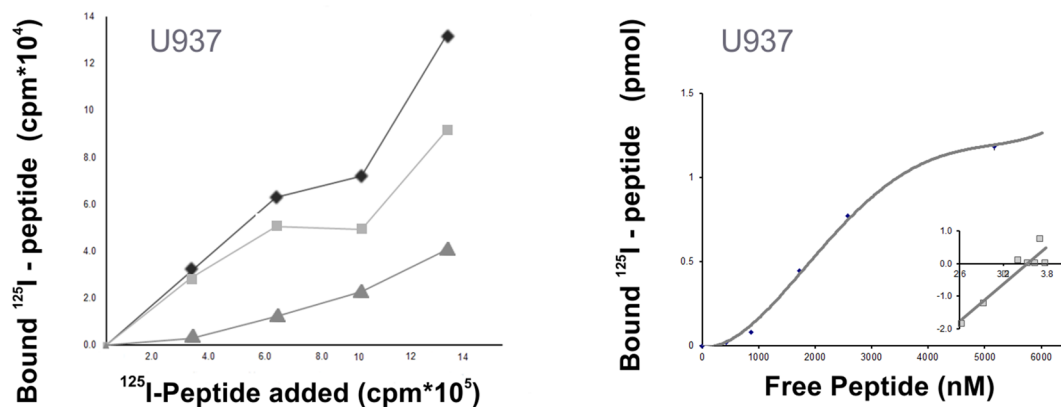
A

PEPTIDE NUMBER	Rv0354c SEQUENCE		Specific Binding Activity (%)								
			U937				A549				
			1	2	3	4	1	2	3	4	
39219	1	MSVTVIYIPFKGTVKHVSVT	20								
39220	21	IPITTEHLGPYEIDASTINPD	41								
39221	42	YQPIDTAFTQTLDFAGSGTVG	61								
39222	62	YAFPFGFGWQQSPGFFNSTTT	81								
39223	82	PSSGFFNSGAGGASGFLND	100								
39224	101	YAAAVSGLGNVFTETSGFFNA	121								
39225	122	YGGVGIRASKTSATTTRAGRT	141								

B

PEPTIDE 39224

C

PEPTIDE 39225

◄**Fig. 3** PPE7 protein synthetic peptides' specific binding. **A** Percentage of specific binding to U937 and A549 cells. The *black bars* in front of each peptide sequence show the percentage of the protein's synthetic peptide binding to the cells. Peptides having ≥1 binding were considered HABPs. Tyrosine residues were added to the *N*-terminal of peptides which did not contain this amino acid in their sequence (to enable radiolabelling). **B** Graphical representation of HABP 39224 binding to A549 and U937 cells (*left-hand panel*) and HABP 39224 saturation curves for A549 and U937 cells (*right-hand panel*). **C** Graphical representation of HABP 39225 binding to U937 cells (*left-hand panel*) and HABP 39225 saturation curves for U937 cells (*right-hand panel*). In the binding graph, total binding (*filled diamond*) means cells incubated with ^{125}I-peptide, non-specific binding (*filled square*) means cells incubated with ^{125}I-peptide with an excess of non-radiolabelled peptide and specific binding (*filled triangle*) means the difference between the two previous ones. The smallest graph within the saturation curves is a representation of the Hill plots. The *X*-axis (log F) shows free peptide and the *Y*-axis [log $(b/B-b)$] shows F is free peptide, b the amount of bound peptide and B the maximum amount of bound peptide

inhibiting infection in vitro and which might form the basis for designing immunogenically active molecules [11, 17].

The PE/PPE complex is one of the proteins in which research is becoming increasingly interesting for developing drugs or vacci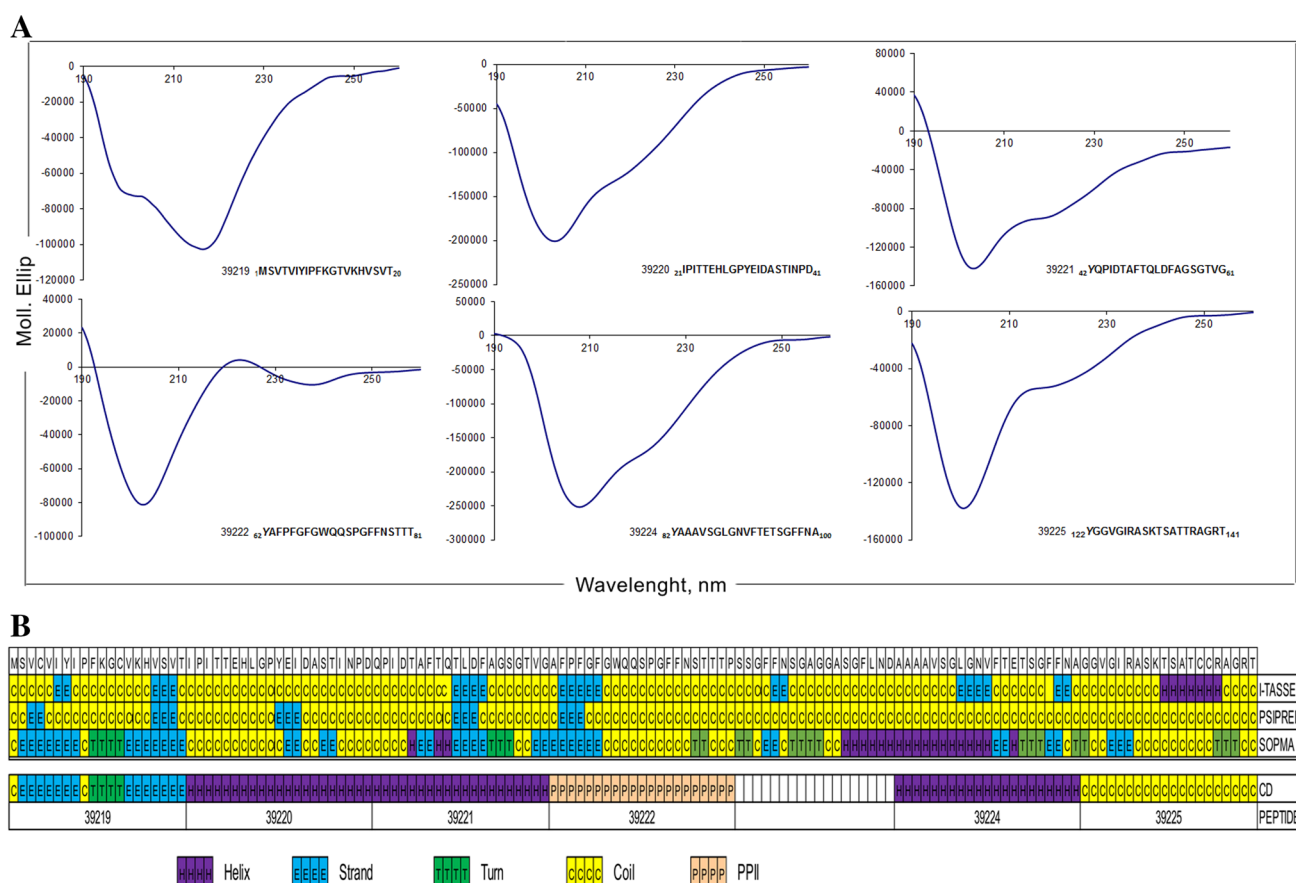nes against TB due to its function continuing to remain unclear, in spite of it being especially abundant in pathogenic mycobacteria. According to research results, the PPE7 protein does have the characteristics which are most common amongst the proteins from its family which are exclusive to pathogenic mycobacteria, having a subcellular location and interaction with host components. The presence of the *rv0354c* gene encoding the PPE7 protein has been found experimentally by amplifying the specific 178 bp segment in strains from the MTC complex (*Mtb* H37Rv, *Mtb* H37Ra, *M. bovis* and *M. bovis* BCG). This gene has also been transcribed in standard culture conditions, thereby agreeing with bioinformatics analysis and indicating that the PPE7 protein is conserved and that its biological function in the PE/PPE complex could be related to *Mtb* pathogenicity.

Bioinformatics tools TBPred and PA-SUB gave the PPE7 protein a subcellular localisation, anchored to the membrane by lipids; this was confirmed by immunoelectron microscopy using polyclonal antibodies. Such subcellular localisation and the bioinformatics prediction suggested a high probability of the protein's peptide sequence interacting with infection target cells. The protein's sequence was synthesised in 20 aa peptides, two peptides being found

Fig. 4 PPE7 protein secondary structure. **A** CD spectra for the peptides forming the PPE7 protein. Molar ellipticity compared to wavelength. **B** Bioinformatics analysis of the protein's secondary structure, compared to the results obtained by CD
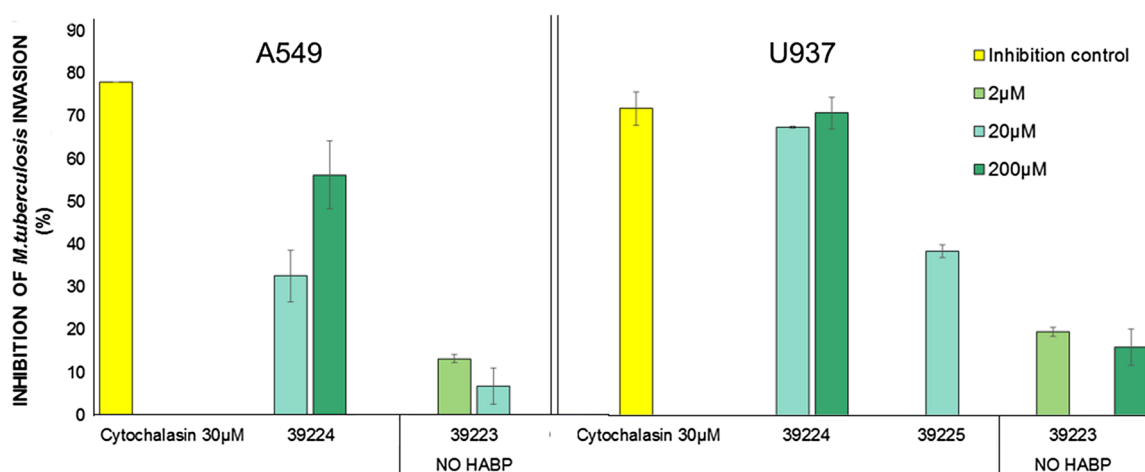
Fig. 5 Inhibiting invasion mediated by HABPS. Percentage inhibition for *Mycobacterium tuberculosis* H37Rv invasion of A549 and U937 cells at increasing HABP 39224 and 39225 concentrations (2, 20 and 200 µM). Cytochalasin D (30 µM) was used as inhibition control, as well as a non-HABP control (peptide 39223)

which had high binding activity (HABPs) to epithelial cells and macrophages, located in the PPE7 protein's carboxyl terminal region, between aa 101 and 141. This region could be essential for mycobacteria regarding their interaction with cell invasion since these HABPs' binding to cells indicated specificity in interactions between the PPE7 protein and cell receptors.

Both HABPs had micromolar order dissociation constants (K_D) which was comparable to those obtained in antigen–antibody interaction, having a greater amount of binding sites for A549 epithelial alveolar cells (10^7) compared to that for the binding of U937 monocyte-derived macrophages (10^5–10^6). HABP 39224 (having an

alpha helix secondary structure identified by CD) was a HABP for A549 epithelial cells and U937 macrophages and inhibited mycobacterial invasion of cells by more than 50%, whilst peptide 39225 (having an undefined structure) was only a HABP for A549 and inhibited invasion of this cell line by 40% at 20 µM concentration. The inhibition assay result could indicate that peptides 39224 and 39225 are expressed in *Mtb* H37Rv. The 200 µM concentration possibly caused damage to the mycobacterial membrane; this could have been related to percentage cytotoxicity. On the other hand, the variability found regarding peptide 39225 is a characteristic which excludes this peptide being considered for designing an
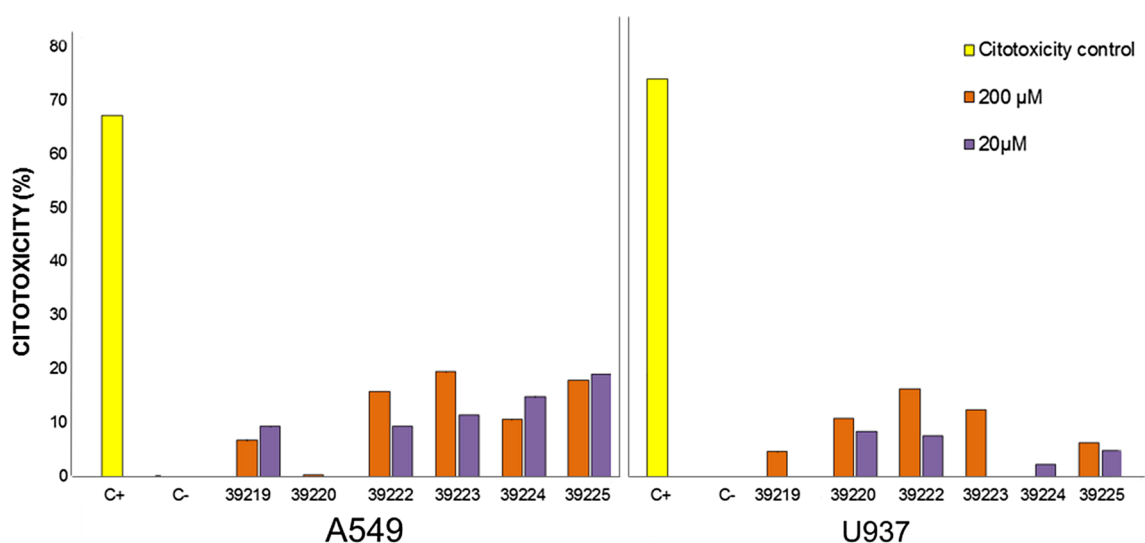

Fig. 6 The Rv0354c protein peptides' cytotoxicity. Each peptide's percentage cytotoxicity regarding A549 and U937 cells is shown at two concentrations (20 and 200 µM). Triton x-100 was used as cytotoxic control and one triplicate was incubated without peptides as a negative control

effective vaccine against different agents causing TB in humans.

Each PPE7 protein peptide's secondary structure did not have a direct relationship with cell binding for selecting HABPS or the inhibitor effect; further studies should be aimed at establishing the immunogenic characteristics and structural modifications needed for producing an immune response in the host.

The results given here form a part of an on-going series of work seeking to define peptide sequences enabling *Mycobacterium tuberculosis* interaction with infection target cells [11, 17] as a novel approach in the search for antigen candidates for developing a multi-epitope vaccine against tuberculosis and explores the advantages of using synthetic peptides (i.e. stability, purity and low cost compared to recombinant proteins), a methodology which has already been seen to be valid in developing a multi-epitope, subunit-based, chemically synthesised vaccine against malaria [16, 45].

Conclusion

An on-going study of the PE/PPE protein complex which, due to its particularities, has become the target for studies and its potential for being included in new drugs or in an anti-tuberculosis vaccine, led to revealing two peptides from the PPE7 membrane protein located in the *C*-terminal extreme having specific high binding to target cells. The inhibitory effect of conserved peptide coded HABP 39224 in in vitro infection assays showed it to be promising for inclusion when designing a multi-antigen, chemically synthesised vaccine.

References

1. WHO (2016) Global tuberculosis report 2016. World Health Organization, Geneva
2. Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S, Eiglmeier K, Gas S, Barry CR (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544
3. van Pittius NCG, Sampson SL, Lee H, Kim Y, Van Helden PD, Warren RM (2006) Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol Biol 6:95
4. Akhter Y, Ehebauer MT, Mukhopadhyay S, Hasnain SE (2012) The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? Biochimie 94:110–116
5. Rindi L, Lari N, Garzelli C (1999) Search for genes potentially involved in *Mycobacterium tuberculosis* virulence by mRNA differential display. Biochem Biophys Res Commun 258:94–101
6. Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jiménez C, Parra M, Cadieux N, Brennan MJ, Appelmelk BJ (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. Mol Microbiol 73:329–340
7. Sampson S, Lukey P, Warren R, van Helden P, Richardson M, Everett M (2001) Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c. Tuberculosis 81:305–317
8. Sani M, Houben E, Geurtsen J, Pierson J, de Punder K, van Zon M, Wever B, Piersma SR, Jiménez CR, Daffé M (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. PLoS Pathog 6:e1000794
9. Demangel C, Brodin P, Cockle PJ, Brosch R, Majlessi L, Leclerc C, Cole ST (2004) Cell envelope protein PPE68 contributes to *Mycobacterium tuberculosis* RD1 immunogenicity independently of a 10-kilodalton culture filtrate protein and ESAT-6. Infect Immun 72:2170–2176
10. Le Moigne V, Robreau G, Borot C, Guesdon J-L, Mahana W (2005) Expression, immunochemical characterization and localization of the *Mycobacterium tuberculosis* protein p27. Tuberculosis 85:213–219
11. Diaz DP, Ocampo M, Pabon L, Herrera C, Patarroyo MA, Munoz M, Patarroyo ME (2016) Mycobacterium tuberculosis PE9 protein has high activity binding peptides which inhibit target cell invasion. Int J Biol Macromol 86:646–655. doi:10.1016/j.ijbiomac.2015.12.081
12. Nair S, Ramaswamy PA, Ghosh S, Joshi DC, Pathak N, Siddiqui I, Sharma P, Hasnain SE, Mande SC, Mukhopadhyay S (2009) The PPE18 of *Mycobacterium tuberculosis* interacts with TLR2 and activates IL-10 induction in macrophage. J Immunol 183:6269–6281
13. Tiwari B, Ramakrishnan UM, Raghunand TR (2015) The *Mycobacterium tuberculosis* protein pair PE9 (Rv1088)–PE10 (Rv1089) forms heterodimers and induces macrophage apoptosis through Toll-like receptor 4. Cell Microbiol 17:1653–1669
14. Bansal K, Sinha AY, Ghorpade DS, Togarsimalemath SK, Patil SA, Kaveri SV, Balaji KN, Bayry J (2010) Src homology 3-interacting domain of Rv1917c of *Mycobacterium tuberculosis* induces selective maturation of human dendritic cells by regulating PI3K-MAPK-NF-κB signaling and drives Th2 immune responses. J Biol Chem 285:36511–36522
15. Bansal K, Elluru SR, Narayana Y, Chaturvedi R, Patil SA, Kaveri SV, Bayry J, Balaji KN (2010) PE_PGRS antigens of Mycobacterium tuberculosis induce maturation and activation of human dendritic cells. J Immunol 184:3495–3504
16. Patarroyo ME, Bermudez A, Patarroyo MA (2011) Structural and immunological principles leading to chemically synthesized, multiantigenic, multistage, minimal subunit-based vaccine development. Chem Rev 111:3459–3507
17. Ocampo M, Patarroyo MA, Vanegas M, Alba MP, Patarroyo ME (2014) Functional, biochemical and 3D studies of *Mycobacterium tuberculosis* protein peptides for an effective anti-tuberculosis vaccine. Crit Rev Microbiol 40:117–145

18. Kruh N, Troudt J, Izzo A, Prenni J, Dobos K, Aziz RK (2010) Portrait of a pathogen: The Mycobacterium tuberculosis Proteome. Vivo. PLoS One 5:e13938

19. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K (2002) Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol Microbiol 43:717–731

20. Kohli S, Singh Y, Sharma K, Mittal A, Ehtesham NZ, Hasnain SE (2012) Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H37Rv and H37Ra reveal novel and interesting differences with implications in virulence. Nucleic Acids Res 40:7113–7122. doi:10.1093/nar/gks465

21. Boratyn GM, Schaffer A, Agarwala R, Altschul SF, Lipman DJ, Madden TL (2012) Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12

22. Nancy YY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615

23. Shen H-B, Chou K-C (2009) Gpos-mPLoc: a top-down approach to improve the quality of predicting subcellular localization of Gram-positive bacterial proteins. Protein Pept Lett 16:1478–1484

24. Szafron D, Lu P, Greiner R, Wishart DS, Poulin B, Eisner R, Lu Z, Anvik J, Macdonell C, Fyshe A (2004) Proteome Analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. Nucleic Acids Res 32:W365–W371

25. Rashid M, Saha S, Raghava GP (2007) Support Vector Machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs. BMC Bioinform 8:337

26. Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795

27. Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res 35:W429–W432

28. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Ismb 6: 175–182

29. Pasquier C, Promponas V, Palaios G, Hamodrakas J, Hamodrakas S (1999) A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. Protein Eng 12:381–385

30. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–405

31. Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684.

32. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738

33. Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2:2

34. Rodriguez DM, Ocampo M, Curtidor H, Vanegas M, Patarroyo ME, Patarroyo MA (2012) Mycobacterium tuberculosis surface protein Rv0227c contains high activity binding peptides which inhibit cell invasion. Peptides 38:208–216. doi:10.1016/j.peptides.2012.08.023

35. Rodriguez DC, Ocampo M, Varela Y, Curtidor H, Patarroyo MA and Patarroyo ME (2015) Mce4F *Mycobacterium tuberculosis* protein peptides can inhibit invasion of human cell lines. Pathog Dis 73(3): ftu020. doi:10.1093/femspd/ftu020

36. Houghten RA (1985) General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. Proc Natl Acad Sci USA 82:5131–5135

37. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287:252–260

38. Bermudez LE, Goodman J (1996) *Mycobacterium tuberculosis* invades and replicates within type II alveolar cells. Infect Immun 64:1400–1406

39. Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S (2004) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 17:349–356

40. Restrepo-Montoya D, Vizcaíno C, Niño LF, Ocampo M, Patarroyo ME, Patarroyo MA (2009) Validating subcellular localization prediction tools with mycobacterial proteins. Bmc Bioinform 10:134

41. Holzwarth G, Doty P (1965) The ultraviolet circular dichroism of polypeptides1. J Am Chem Soc 87:218–228

42. Berova N, Nakanishi K (2000) Circular dichroism: principles and applications. Wiley, New York

43. Bochicchio B, Tamburro AM (2002) Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. Chirality 14:782–792

44. Jones M (1999) Surfactants in membrane solubilisation. Int J Pharm 177:137–159

45. Patarroyo ME, Bermudez A, Alba MP, Vanegas M, Moreno-Vranich A, Poloche LA, Patarroyo MA (2015) IMPIPS: the immune protection-inducing protein structure concept in the search for steric-electron and topochemical principles for complete fully-protective chemically synthesised vaccine development. PLOS ONE 10:e0123249. doi:10.1371/journal.pone.0123249