Bioorganic & Medicinal Chemistry 26 (2018) 2401–2409

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Towards designing a synthetic antituberculosis vaccine: The Rv3587c peptide inhibits mycobacterial entry to host cells

Mary Lilian Carabali-Isajar [a,b], Marisol Ocampo [a,b,*], Deisy Carolina Rodriguez [a,b], Magnolia Vanegas [a,b], Hernando Curtidor [a,b], Manuel Alfonso Patarroyo [a,b], Manuel Elkin Patarroyo [a,c]

[a] Fundación Instituto de Inmunología de Colombia (FIDIC), Carrera 50 No. 26–20, Post Code: 111321, Bogotá, Colombia
[b] Universidad del Rosario, Carrera 24 No. 63C-69, Post Code: 111321, Bogotá, Colombia
[c] Universidad Nacional de Colombia, Carrera 45 No. 26-85, Post Code: 11001, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 23 December 2017
Revised 15 March 2018
Accepted 29 March 2018
Available online 4 April 2018

Keywords:
Mycobacterium tuberculosis
Rv3587c
Bioinformatics
HABPs
Invasion assay
Vaccine

ABSTRACT

Mycobacterium tuberculosis is considered one of the most successful pathogens in the history of mankind, having caused 1.7 million deaths in 2016. The amount of resistant and extensively resistant strains has increased; BCG has been the only vaccine to be produced in more than 100 years though it is still unable to prevent the disease's most disseminated form in adults; pulmonary tuberculosis. The search is thus still on-going for candidate antigens for an antituberculosis vaccine. This paper reports the use of a logical and rational methodology for finding such antigens, this time as peptides derived from the Rv3587c membrane protein. Bioinformatics tools were used for predicting mycobacterial surface location and Rv3587c protein structure whilst circular dichroism was used for determining its peptides' secondary structure. Receptor-ligand assays identified 4 high activity binding peptides (HABPs) binding specifically to A549 alveolar epithelial cells and U937 monocyte-derived macrophages, covering the region between amino acids 116 and 193. Their capability for inhibiting *Mtb* H37Rv invasion was evaluated. The recognition of antibodies from individuals suffering active and latent tuberculosis and from healthy individuals was observed in HABPs capable of avoiding mycobacterial entry to host cells. The results showed that 8 HABPs inhibited such invasion, two of them being common for both cell lines: 39265 (^{155}VLAAYVYSLDNKRLWSNLDT^{173}) and 39266 (^{174}APSNETLVKTFSPGEQVTTY^{192}). Peptide 39265 was the least recognised by antibodies from the individuals' sera evaluated in each group. According to the model proposed by FIDIC regarding synthetic vaccine development, peptide 39265 has become a candidate antigen for an antituberculosis vaccine.

1. Introduction

Being infected by *Mycobacterium tuberculosis* (*Mtb*) and its progression to active tuberculosis has caused the deaths of 1.7 million people worldwide.[1] The only vaccine approved by the World Health Organisation (WHO), the Calmette and Guerin attenuated bacillus (BCG, *Mycobacterium bovis*), is insufficient for preventing the disease in adults[2]; new vaccine candidates must thus be sought. Bearing in mind the pathogen's biology (so that the infection can affect a host cell) there must be an interaction between mycobacterial ligands and the receptors on a host cell[3–5]; our approach has thus been focused on the search for antigens which might be involved in such interaction for consideration as candidates when designing an antituberculosis vaccine.[6] Such antigens come from cell surface-located membrane proteins which bind specifically to a cell targeted for infection, thereby inhibiting mycobacterial entry *in vitro*. Our experience with *P. falciparum* has shown these antigens to be poorly immunogenic[7]; their structural characteristics should thus be modified for obtaining peptides having immunogenic and protection-inducing properties against tuberculosis. It is worth stating that our strategy consists of designing *Mtb* protein-derived peptide antigens for obtaining an effective synthetic antituberculosis vaccine.

The gene encoding the 264 amino acid-long Rv3587c protein forms part of the *Mtb* genome published in 1998[8]; found as a membrane protein, its function still remains unknown.[9] *rv3587c* has been reported as being a conserved gene in *M. tuberculosis* H37Rv, *M. leprae*, *M. avium*, *M. marinum* and *M. smegmatis* strains.[10] The protein contains a hydrophobic segment in its *N*-terminal region, indicating a signal peptide sequence, as predicted by Malen

* Corresponding author at: Fundación Instituto de Inmunología de Colombia (FIDIC), Carrera 50 No. 26–20, Post Code: 111321, Bogotá, Colombia.
E-mail address: marisol.ocampo@urosario.edu.co (M. Ocampo).

https://doi.org/10.1016/j.bmc.2018.03.044
0968-0896/

et al.[11] who identified the protein in mycobacterial culture filtrate, thereby indicating a possible secreted protein. It has also been identified by mass spectrometry in *Mtb* H37Rv culture filtrate and membrane protein fraction but not in whole cell lysate; it is enriched in the membrane fraction, has a predicted *N*-terminal signal peptide and is uncleaved.[12] Antibodies recognising this protein have also been found in tuberculosis patients' sera.[13] Himar1-based transposon mutagenesis sequencing has led to it being reported as an essential gene for *Mtb* H37Rv growth in *in vitro* culture.[14] Bearing such background in mind, the Rv3587c protein could be considered an interesting antigen regarding the search for anti-tuberculosis vaccine candidates; the present research was thus aimed at identifying Rv3587c-derived high activity binding peptides (HABPs) and their role in inhibiting *Mtb* entry to U937 alveolar macrophages and A549 epithelial cells. Peptide recognition by sera from patients having active and latent tuberculosis was also evaluated.

Our strategy for obtaining a synthetic anti-tuberculosis vaccine has been based on designing *Mtb* protein-derived peptide antigens; sequences are modified from functionally-important peptides in host-pathogen interaction, such as those reported here regarding Rv3587c.

2. Materials and methods

2.1. Bioinformatics approach in Rv3587c characterisation

The protein sequence reported in the TubercuList *Mycobacterium tuberculosis* database (http://tuberculist.epfl.ch/index.html) was used for assessing Rv3587c *in silico* characterisation. The basic local alignment search tool (BLAST, an algorithm for comparing primary biological sequence information) (https://blast.ncbi.nlm.nih.gov/Blast.cgi#)[15] was used for identifying the protein in *Mtb* complex (MTC) strains. TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)[16] and, HMMTOP (http://www.enzim.hu/hmmtop/) tools were then used for predicting transmembrane helices. SignalP 3.0 was used for predicting the signal peptide (http://www.cbs.dtu.dk/services/SignalP-3.0/)[17] and TBpred (http://www.imtech.res.in/raghava/tbpred/)[18] and Phobius (http://phobius.sbc.su.se/cgi-bin/predict.pl) were used for predicting subcellular location.

2.2. rv3587c gene presence and transcription

A widely-reported methodology involving genomic DNA (gDNA) extraction and polymerase chain reaction (PCR) assays[19–21] was used for detecting the gene encoding the Rv3587c protein in *Mtb* H37Rv (ATCC 27294), *Mtb* H37Ra (ATCC 25177), *M. bovis* (ATCC 19210), *M. bovis* BCG (ATCC 27291, Pasteur substrain) and *M. smegmatis* (ATCC 19420) strains.

The mycobacterial strains and species were cultured for 5–15 days in Middlebrook 7H9 medium, supplemented with oleic acid, catalase, bovine albumin fraction, dextrose and sodium chloride (10% OADC); catalase, bovine albumin fraction, dextrose and sodium chloride (10% ADC) were used as supplements for *M. smegmatis*. They were incubated at optimal growth temperature until reaching 0.5–1.0 OD_{600} optical density. The mycobacteria were harvested at logarithmic growth phase, spun at 13,000 rpm for 20 min at 4 °C, suspended in PBS and stored at −20 °C.

An UltraClean Microbial DNA isolation kit (MoBio Laboratories, Inc., Carlsbad, CA, USA) was used for isolating gDNA from mycobacteria, following the manufacturer's instructions. The PCR assay was carried out using a LABNET MultiGene Thermal Cycler (Woodbridge, NJ, USA), incubating 100 ng gDNA with a mixture of 1.25 units of GOtaq (Promega, Madison, USA), 1× Standard Taq Reaction Buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs and each primer (1 μM) in 10-μL final reaction volume. DNA quality was verified by amplifying a 439 bp fragment from the *hsp65* gene (as constitutive *Mtb* gene), using direct 5′-ACCAACGATGGTGTGTCCAT-3′ and reverse 5′-CTTGTCGAACCGCATACCCT-3′ primers. A set of *rv3587c*-D (5′-CTACGTTTACTCGCTGGAC-3′) and *rv3587c*-R (5′-GCCCAGTTGTACCACGAG-3′) primers was used for determining the presence of the gene encoding the protein of interest.

The PCR reaction was carried out in the following conditions: denaturing at 95 °C for 5 min, followed by 35 cycles lasting 30 s at 95 °C, 45 s at 54.7 °C and 30 s at 72 °C. A final extension step was carried out at 72 °C for 5 min. Nuclease-free water was used in each assay as negative control (using the same reaction conditions). PCR-amplified products were visualised on SYBR Safe stained 2% agarose gel.

Total RNA was isolated from the bacterial pellet by homogenisation in 1 mL Trizol reagent for determining *rv3587c* gene transcription, following the manufacturer's recommendations. DNAse I amplification grade was used for eliminating DNA during RNA purification. cDNA synthesis was carried out in RNase-free conditions using a Super Script III FirstStrand Synthesis System kit (Invitrogen, Carlsbad, CA, USA). An 11 μl reaction mixture was prepared, containing 8 μl RNA, 2 μl random primers (hexamers) and 1 μl dNTPs. The primer mixture was incubated at 65 °C for 5 min and then incubated on ice for 1 min; 2 μl 10× reaction buffer (250 mM Tris-HCl), 25 mM 4 μl $MgCl_2$, 1 μl RNaseOUT (20 U/μl) and 50 mM 2 μl dithiothreitol (DTT) were added to the mixture. Following a small pulse/burst of centrifugation, 1 μl (200 U/μl) Super Script enzyme was added. Negative synthesis control for each sample involved replacing SuperScript III with DEPC-treated water; 1 μl cDNA was used as template for PCR amplification, in the same conditions described for DNA. The *hsp65* gene was used as positive transcription control.

2.3. Peptide synthesis

The solid phase multiple peptide synthesis method,[22,23] involving MBHA resin (0.5 meq/g) and t-Boc amino acids (aa), was used for synthesising Rv3587c protein 20-mer peptides. This was followed by low (HF 35%, 0 °C) – high cleavage (HF 95%, temperature ramp from −10 to 0 °C)[24] and purified by semi-preparative reverse-phase, high-performance liquid chromatography (RP-HPLC) (Merck Hitachi, L-450 UV–vis detector). Peptide identity was verified by MALDI-TOF mass spectrometry (Bruker Daltonics, autoflex). Peptide purity was seen to be greater than 90%, determined by peak area analysis of purified peptide chromatogram. Peptides were coded from 39257 to 39270 (from protein *N*-terminal to *C*-terminal extreme); a tyrosine (Y) residue was added to peptides lacking it in their aa sequences to enable peptide radiolabelling.

2.4. Receptor-ligand assays

Monocyte-derived U937 macrophages (ATCC CRL-2367) and A549 alveolar epithelial cells (ATCC CLL-185) were used as host cells; these were cultured in RPMI-1640 supplemented with 10% inactivated bovine foetal serum (BFS) at 37 °C in a 5% CO_2 atmosphere. Fourteen Rv3587c-derived synthetic peptides were radiolabelled with $Na^{125}I$, as described previously.[6] The receptor-ligand assays consisted of incubating increasing radiolabelled peptide concentrations (0–950 nM) for 2 h with 1.2×10^6 A549 (37 °C) and U937 cells (4 °C) per well in 96-well dishes in triplicate to obtain total binding and with non-radiolabelled peptide (40 Mm) regarding inhibited binding; specific binding was calculated as the difference between total and inhibited binding. Cell-associated radioactivity was measured by gamma counter (Gamma Counter

Cobra II, Packard Instrument Co., Meriden, CT, USA). A HABP was defined as one whose specific binding curve slope was ≥0.01 (1%), according to established parameters.[25–27]

2.5. Invasion inhibition assays

A549 and U937 cell entry inhibition assays were carried out as reported beforehand.[28] Briefly, 2.5×10^5 cells from each line were placed in 24-well dishes and brought into contact with increasing concentrations of each HABP (2, 20 and 200 μM); a non-HABP was also tested for each cell line. A549 epithelial cells were incubated at 37 °C in 5% CO_2 for 2 h and U937 macrophages at 4 °C for reducing their phagocytic activity. *Mtb* H37Rv lysate (200 μg/ml) was used as inhibition control and untreated cells as infection control. All assays were carried out in triplicate. Cells were infected by *Mtb* H37Rv at 1:50 multiplicity of infection (MOI) and incubated overnight. Extracellular bacteria were removed and cells lysed with distilled water for quantifying intracellular mycobacteria; 50 μl of sample was sown in Middlebrook 7H10 agar. Inhibition percentage was calculated, taking control without any inhibitor as 100% infection.

A cytotoxicity assay with MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) was used for ascertaining whether peptides were toxic for cells. Briefly, 5×10^4 cells per well (from each cell line) were placed in 96-well dishes, left to adhere overnight in RPMI-1640 with 10% BFS (Hyclone, Logon, UT) in 5% CO_2 at 37 °C (200 μl final volume). Cells were then incubated for 2 h with peptides at 20 and 200 μM concentration in the same conditions used for the invasion inhibition assay (i.e. A549 cells at 37 °C and U937 cells at 4 °C). Cells without peptides were used as viability assay positive control and treatment with 30% dimethyl sulfoxide (DMSO) as cytotoxicity assay positive control. Once incubation time had elapsed, 100 μl of the medium was removed and 10 μl MTT added. The cells were incubated overnight at 37 °C in 5% CO_2. After formazan crystals became formed, 100 μl sodium dodecyl sulphate (SDS) was added and the cells incubated again for 12 h to allow the crystals to dissolve; supernatants were read at 570 nm wavelength.

2.6. Establishing antigen recognition of Rv3587c protein peptides

Enzyme-linked immunosorbent (ELISA) assays (using sera collected years before) were carried out for ascertaining protein peptide antigen activity and kept in our Institute. Individuals were grouped into three categories: people suffering active tuberculosis (ATB, n = 6), people having latent tuberculosis infection (LTB, n = 7) and healthy individuals (HD, n = 7). An ATB diagnosis was based on *Mtb* sputum having a positive culture. LTB was diagnosed by positive interferon gamma release assay and HD by a negative interferon gamma release assay (QuantiFERON Gold Cellestis Limited, Valencia, USA) and negative *Mtb* sputum.

Peptides at 10 μg/ml concentration were immobilised on flat-bottomed polystyrene 96-well plates, in triplicate; they were incubated for 12 h at 4 °C. The plasma was used at 1:100 dilution for 2 h at 37 °C, thereby enabling antigen–antibody interaction. Peroxidase-conjugated anti-human IgG secondary antibody (Vector-Laboratories, Inc.) was then added at 1:5000 dilution and incubated for 1 h at 37 °C. Revealing solution (liquid substrate system for ELISA peroxidase substrate – TMB 3,3′,5,5′-tetramethylbenzidine) was added before incubating again for 15 min and absorbance was read at 620 nm.

2.7. Ascertaining synthetic peptide secondary structure

The peptides' structural element content was analysed by circular dichroism (CD). This was recorded at 20 °C using a Jasco J-810 spectropolarimeter at wavelengths ranging from 260 nm to 190 nm in 1.00-cm cuvettes.[29] Peptides were dissolved at 0.1 mM concentration in 30% TFE solutions and the spectrum was read three times. As the result is expressed as obtained ellipticity (θ), this is usually converted to molar ellipticity ([θ]) for comparing data (the units being degrees·centimeters squared per decimole according to $[\theta] = \theta\lambda/(100\boldsymbol{l}\text{cn})$. Measured ellipticity was θ at wavelength λ, cell path length was $\boldsymbol{l}$, $\boldsymbol{c}$ was peptide concentration and $\boldsymbol{n}$ the number of aa residues in a peptide sequence. SELCON3,[30] CONTINLL[31] and CDSSTR[32] were used for deconvolution calculations.

Self-optimised prediction method with alignment (SOPMA) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)[33] software was used for predicting protein secondary structure and I-TASSER (http://zhanglab.ccmb.med.umich.edu/),[34] Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index)[35] and nFOLD3 (http://www.reading.ac.uk/bioinf/nFOLD/nFOLD_form.html)[36] servers for predicting tertiary structure. The Swiss-Model server (http://swissmodel.expasy.org/workspace/index.php?)[37] was used for evaluating the models so produced and processed with Chimera software (https://www.cgl.ucsf.edu/chimera/).[38]

3. Results

3.1. The bioinformatics approach to Rv3587c characterisation

A bioinformatics approach was adopted for providing information about aspects of protein topology considered relevant in host-pathogen interaction. BLAST alignment gave 100% identity for the Rv3587c *M. tuberculosis* H37Rv protein with sequences from *M. tuberculosis* complex strains, such as *M. tuberculosis Beijing* strain, *M. tuberculosis* CCDC5180, *M. tuberculosis Erdman, M. tuberculosis ZMC13-88, M. tuberculosis XTB13-086 M. bovis, M. bovis* BCG *M. africanum* and *M. canetti* (Fig. S1). TMHMM 2.0 and HMMTOP analysis revealed a transmembrane helix between residues 20–42 or 21–41, a signal peptide region in the extreme intracellular *N*-terminal (1–19) and residues 43 to 264 in the extracellular region. SignalP 3.0 predicted a signal peptide cleavage site in positions 47 and 48 (0.651 max. cleavage site probability, 0.979 signal peptide probability).[39] TBpred and Phobius predicted an integral membrane protein.

3.2. The rv3587c gene was present and transcribed in all mycobacteria in the study

Following the prediction that the protein could be exposed on cell membrane, *rv3587c* gene presence and transcription (see Fig. 1) in *Mtb* H37Rv were analysed in normal culture conditions to confirm the results of the bioinformatics approach, taking into account previous reports about the presence of Rv3587c protein.[11,12] The *hsp65* gene was amplified as good quality control regarding gDNA isolated from mycobacteria; it had a 439 bp band (Fig. 1A). It was then found that the *rv3587c* gene was present in all mycobacteria involved in the study, including those forming part of the *Mycobacterium tuberculosis* complex such as *M. tuberculosis* H37Rv, *M. tuberculosis* H37Ra, *M. bovis*, *M. bovis* BCG and the *M. smegmatis* non-virulent strain (Fig. 1B). Transcription of the *hsp65* gene's 439 bp amplification fragment was evident for all mycobacteria studied here (Fig. 1C). Fig. 1D shows a 202 bp band for *rv3587c*, thereby highlighting its active transcription in bacteria in normal culture conditions.

3.3. Receptor-ligand assays identify HABPs for both cell lines

As the Rv3587c protein was found in *M. tuberculosis* H37Rv, the peptides included in the protein's complete sequences were

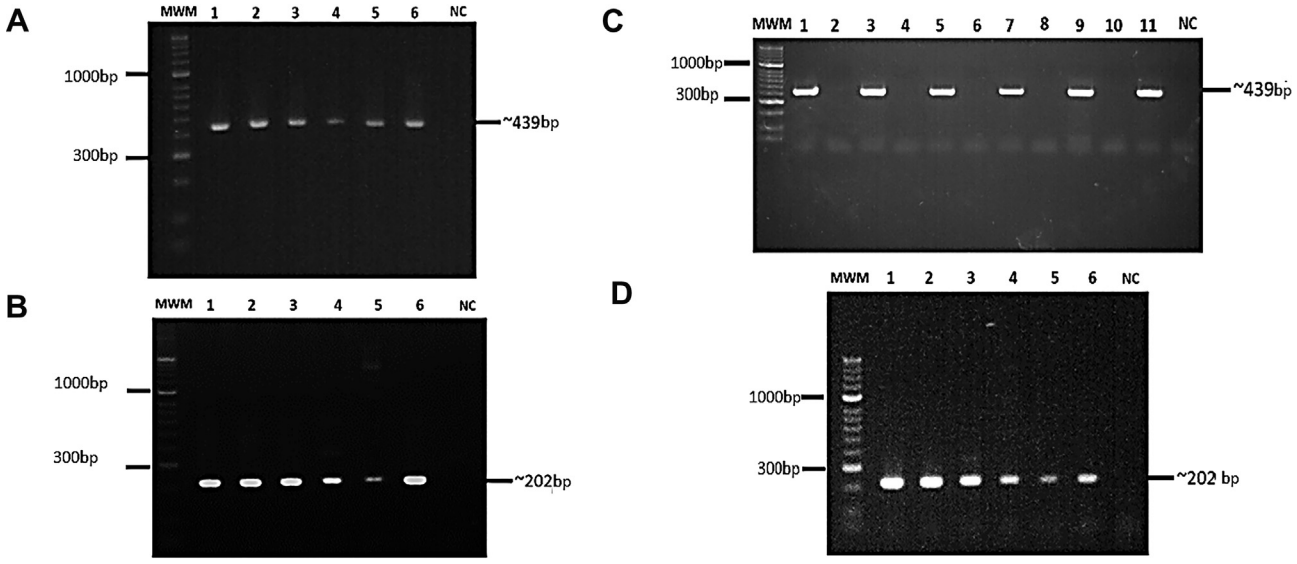

Fig. 1. *rv3587c* gene presence and transcription. (A) *hsp65* gene amplification from gDNA isolated from: 1. *M. tuberculosis* H37Rv; 2. *M. tuberculosis H37Ra*; 3. *M. bovis*; 4. *M. bovis* BCG; 5. *M. smegmatis*; 6. Positive PCR control; NC: Negative PCR control. MWM: 50 bp molecular weight marker used (HyperLadder II). (B) *rv3587c* gene amplification from gDNA isolated from: 1. *M. tuberculosis* H37Rv; 2. *M. tuberculosis H37Ra*; 3. *M. bovis*; 4. *M. bovis* BCG; 5. *M. smegmatis*; 6. Positive PCR control; NC: Negative PCR control. (C) *hsp65* gene amplification from cDNA from: 1. *M. tuberculosis* H37Rv plus synthesis (positive control); 2. *M. tuberculosis* H37Rv minus synthesis (negative control); 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. *M. smegmatis* plus synthesis; 10. *M. smegmatis* minus synthesis; 11. Positive PCR control *(M. tuberculosis* H37Rv gDNA*)*; NC: Negative PCR control; MWM: 50 bp molecular weight marker. (D) *rv3587c* gene amplification from cDNA from: 1. *M. tuberculosis* H37Rv; 2. *M. tuberculosis* H37Ra; 3. *M. bovis*; 4. *M. bovis* BCG; 5. *M. smegmatis*; 6. Positive PCR control; NC: Negative PCR control. MWM: 50 bp molecular weight marker.

synthesised; receptor–ligand assays were used for determining sequences specifically binding target cells. Fig. S2 gives the binding curves for peptides binding to each target cell. Greater than 1% specific binding determined a HABP; receptor-ligand assays led to ascertaining that the region between residues ^{116}Leu and ^{193}Thr, containing peptides 39263 to 39266, were HABPs regarding both cell lines (Fig. 2). It is worth noting that peptide 39265 had high specific binding values for both the A549 epithelial cell line (9.1%) and U937 macrophages (4.4%). These peptides might share a common receptor for both cell lines.

Peptide 39260 (^{59}QSHPGSPAPQAPQPAGQTEY^{77}) was only determined to be a HABP for the A549 cell line (1.1% specific binding) whilst peptides 39267 (3.7% specific binding) (^{193}AVTWTGMGSAPRCPLPRPAY^{211}), 39269 (3.1%) and 39270 (2.7%) from the protein's *C*-terminal extreme (^{232}PFILNQPPPPPGPVPAPGPAQAPPPESPAQGGY^{264}) were HABPs for alveolar macrophages. Receptor-ligand assays highlighted 5 HABPs for A549 epithelial cells and 8 HABPs for U937 monocyte cells. These cells represent a recognised model for studying *Mtb* infection, given that alveolar macrophages and lung epithelial cells are the cells involved in the initial encounter with mycobacterial organisms after pathogen inhalation into the alveolar space in the lungs, which the pathogen invades and uses to replicate itself. Although the lungs are the primary tuberculosis infection site, *Mtb* can cause infection at other sites and certain strains can cause infection at specific extra-pulmonary sites, suggesting that the development of extra-pulmonary tuberculosis depends on the microbe's ability to pass effectively through the alveolar epithelial

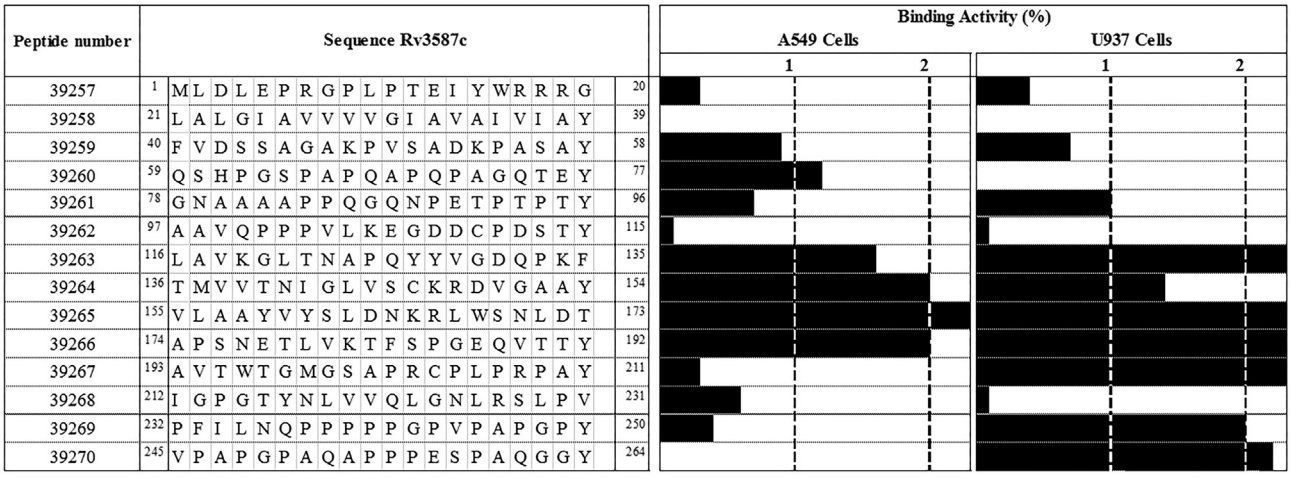

Fig. 2. Profile for peptide binding to A549 and U937 cell lines. FIDIC's synthetic peptide serial numbers and amino acid sequences are shown on the left of the figure. Horizontal black bars represent specific binding capacity, expressed as percentages for each peptide regarding both cell lines tested here. Vertical dotted lines show the 1% specific binding cut-off above which a peptide was considered to be a HABP.

lining and its affinity for cells other than those in pulmonary tissue. Future studies should be orientated towards defining the nature of the receptors used for peptides having high specific binding capability.

A saturation assay was used for determining physicochemical constants for characterising the peptide-host cell interaction regarding HABPs from both cell lines. This involved maintaining constant receptor concentration in addition to increasing radiolabelled peptide concentrations over a wider range than that used for the initial peptide screening until specific binding became saturated, for which the presence or absence of non-radiolabelled peptide is also required. Table 1 shows that dissociation constants (K_d) were slightly higher for alveolar macrophages compared to those for epithelial cells regarding peptide 39263 (^{116}LAVKGLT NAPQYYVGDQPKF^{135}); the same constant was obtained for both cell lines with 39264 (^{136}TMVVTNIGLVSCKRDVGAAY^{154}). Peptide 39265 (^{155}VLAAYVYSLDNKRLWSNLDT^{173}) had greater affinity for epithelial cells than for macrophages and peptide 39266 had less affinity for alveolar cells than macrophages.

The forgoing results suggested that HAPBs from both cell lines have different affinity for the receptors on both types of cell used as the model for target cell infection; peptide 39264 was an exception. The Hill coefficient (n_H) for A549 cells and U937 macrophages was >1 in all assays, denoting positive cooperativity, meaning that a peptide's initial binding to its receptor could facilitate other molecules' association. This assay also enabled calculating binding sites per cell; peptide 39265 had the most sites for both types of cell, thereby explaining the higher specific binding observed in the initial screening assay.

3.4. HABPs inhibited Mtb H37Rv invasion of alveolar macrophages and alveolar epithelial cells

Inhibition assays regarding *Mtb* entry to epithelial cells and macrophages were performed after establishing HABPs for both cell lines and the characteristics concerning their interaction ascertained by saturation assays. It was found that peptide 39265 inhibited mycobacterial entry to both cell lines in a concentration-dependent way (Fig. 3); this peptide had the highest HABP specific binding (Fig. 2). Three of the five peptides had inhibitory activity regarding A549 cells. Interestingly, it was observed that a peptide having low specific binding (39262) had concentration-dependent inhibition regarding epithelial cells, its percentage binding being very similar to that shown by complete *Mtb* H37Rv lysate (close to 80%) (Fig. 3A), whilst HABPs 39260, 39263 and 39364 did not inhibit invasion. Regarding U937, invasion inhibition was seen in six out of the seven HABPs, particularly concerning peptides 39266 and 39267; this was concentration-dependent (Fig. 3B) and the latter inhibited at 200 μM, more so than mycobacterial lysate control. It is worth noting that peptide 39270 only inhibited at maximum concentration and that peptide 39260 (a non-HABP) was capable of inhibiting *Mtb* entry; peptide 39262 (non-HABP) did not inhibit mycobacterial entry to U937 cells, unlike its behaviour regarding A549 cells.

The cytotoxicity assays highlighted the fact that the peptides were not toxic for macrophages or epithelial cells.

3.5. HABP 39265 had poor antigen recognition

Bearing in mind that our approach is aimed at the search for antigens to be included in a multi-epitope anti-tuberculosis vaccine, it had to be ascertained whether Rv3587c protein peptides were recognised by individuals' sera. Fig. 4 shows high *Mtb* lysate recognition of ATB individuals' sera, having significant differences regarding recognition by sera from both LTB and HD individuals. The most antigenic peptides for all individuals came from the region between C-terminal extreme ^{175}Ala and ^{251}Pro, but were not associated with peptide binding or inhibition capability. HABP 39265 (which was conserved amongst mycobacterial strains and inhibited *Mtb* H37Rv entry in both cell lines, Fig. S1) had very low antigenicity, thereby agreeing with that found in experiments when developing a synthesised antimalarial vaccine, i.e. conserved HABPs are inadequate for eliciting any humoral or cellular protective immune responses in immunised animals; such sequences must thus be modified.[7] The *N*-terminal region between ^{1}Met and ^{115}Thr was not recognised by any of the sera whilst all Rv3587c peptides (except 39257 and 39260) were recognised by animal hyperimmune sera obtained by inoculation with *Mtb* lysate.

3.6. Towards synthetic peptide secondary structure

The presence of secondary structure elements in peptides was evaluated, bearing in mind that peptides having a poly l-proline structure can fit better into major histocompatibility complex type II (MHC-II), resulting in better peptide-MHC-II stability regarding their presentation and thereby eliciting a protective immune response.[40] Circular dichroism deconvolution analysis revealed secondary structure elements, most being different to those predicted by self-optimised prediction from multiple alignment (SOPMA) (Fig 5A).

SELCON3, CONTINLL and CDSSTR's lowest normalised root-mean-square deviation (NRMSD) values were used for selecting a suitable deconvolution programme; such analysis highlighted 11 peptides as having helical structure elements whilst random coil-type was mainly predicted *in silico* (Fig. 5A and B). However, analysing the circular dichroism curves showed that peptides 39257, 39264, 39265, 39266 and 39268 had characteristic α-helix peaks (192, 209 and 222 nm). The other peptides had characteristic random coil curves in spite of deconvolution software mainly identifying helical elements. Peptides 39261 and 39269 could not be analysed by any of the deconvolution programmes; 39261 curve seemed to indicate poly l-proline elements and 39269 random coil ones. Differences could have been due to SOPMA considering all interactions within the protein, while experimental data obtained by CD concerning individual synthetic peptides did not consider intramolecular interactions with other regions within the same protein. Future Magnetic Nuclear Resonance studies would give more precise elucidation of the peptides' "true" structure.

Table 1
HABP-cell interaction physicochemical constants.

Peptide number	A549 cells			U937 cells		
	K_d (nM)	n_H	Binding sites ($\times 10^6$)	K_d (nM)	n_H	Binding sites ($\times 10^6$)
39263	3500	1.5	8.1	4100	1.3	3.7
39264	3500	1.1	0.7	3500	1.5	6.2
39265	1500	1.5	11.1	3600	1.3	10.6
39266	4700	1.4	4.0	3600	1.2	9.3

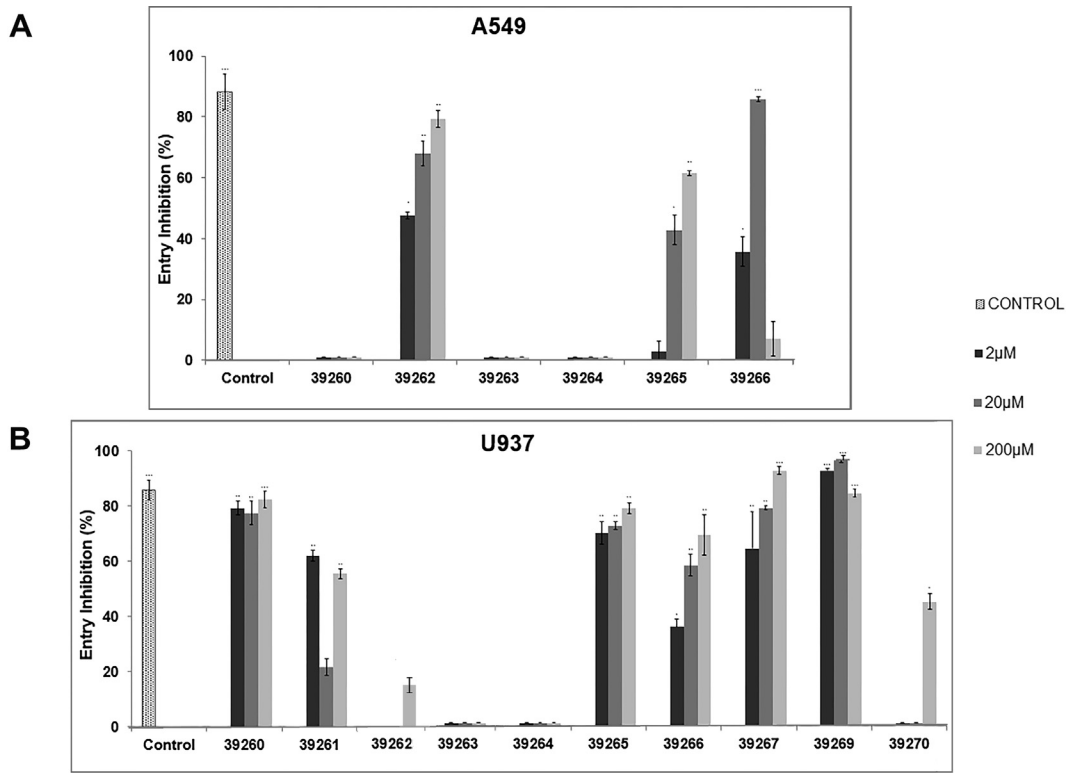

Fig. 3. Inhibition assay regarding *Mtb* invasion of A549 and U937 cells; 2, 20 and 200 μM concentrations were used for both cell lines. *Mtb* H37Rv lysate at 200 μg/ml concentration was used as inhibition control. (A) Inhibiting entry to A549 cells and (B) U937 cells. Student's *t*-test results are represented by dots above each bar: ·≤0.05, ··≤0.01 and ···≤0.001.

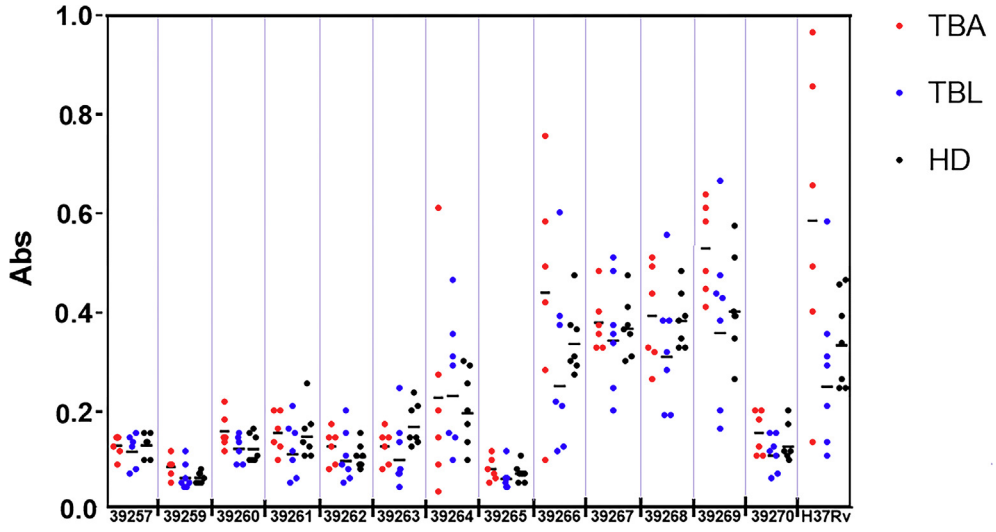

Fig. 4. Antigen recognition of Rv3587c peptides. Evaluating sera from individuals having active (TBA) and latent tuberculosis (TBL) and healthy individuals (HD). 1× PBS was used as negative control regarding recognition and sera obtained from immunising a rabbit with total *Mtb*H37Rv lysate as positive control.

Fig. 5C shows Rv3587c tertiary structure prediction; the sequence was submitted to I-TASSER, Phyre2 and nFOLD3 servers. Qualitative model energy analysis (QMEAN) value (0–1) was considered as part of the criteria for choosing the best model (1 signifying the highest similarity with other proteins' folding resulting from combining geometric patterns comparing the model with similar crystallised structures). The nFOLD3 server provided the best QMEAN (0.693), compared to Iterative Threading ASSEmbly Refinement (I-TASSER) (0.345) and Phyre2 (0.275). Ramachamdran plot located 91.3% of residues in favoured regions (more than 90% would be expected in a model of good quality), 7.8% in allowed regions and 0.8% in disallowed regions. Qmean Z-score indicated that the model's absolute quality given by nFOLD3 was at least one standard deviation between prediction and possible native structure. Qmean Z-score was obtained by comparison with high resolution X-ray structures from the same dimension (i.e. measuring compatibility between sequence and structure) (Fig. 5C). HABPs colocalisation regarding the Rv3587c protein's proposed structure revealed accessibility for interacting with both cell lines, using peptides 39263, 39264, 39265 and 39266. It has been

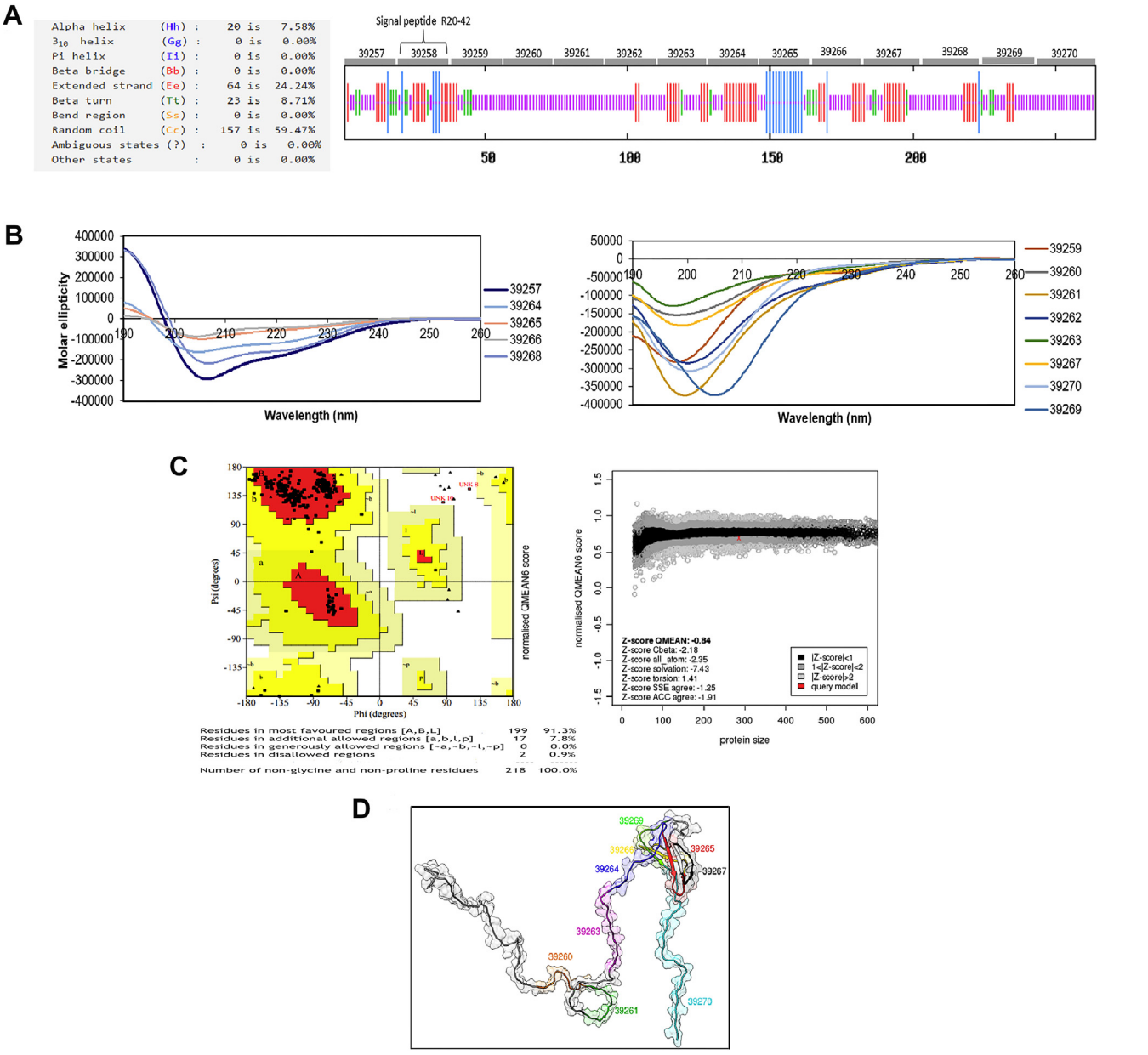

Fig. 5. Rv3587c structural aspects. (A) Bioinformatics analysis involved obtaining a Sopma-predicted (self-optimised prediction) secondary structure model of Rv3587c secondary structure, representing each residue and signal peptide. (B) Dichroism of Rv3587c 20 aa residues. (C) Predicted Rv3587c tertiary structure. A Ramachandran plot was used for evaluating tertiary structure; red marks the most favoured (core) regions, yellow the allowed regions and pink the disallowed regions. Qualitative model energy analysis (QMEAN) is also shown; it was used for assessing protein model quality, combining composite scoring with information regarding structural density. (D) Colocalisation of peptides on predicted tertiary structure.

established that the structure found for an individual peptide has been the same as that observed in the protein from which it was derived.

4. Discussion

Considering that FIDIC's proposed methodology involves a systematic search for sequences which might be implicated in host–pathogen interactions, this work presents the results obtained for the Rv3587c protein, described as probably being a conserved membrane protein having a yet-as-unknown function but seeming to play a role in mycobacterial invasion of host cells.

Bioinformatics analysis predicted that more than 80% of Rv3587c would not be located on mycobacterial cell membrane, even though its fragments have been identified in *Mtb* H37Rv filtrate culture and *Mtb* H37Rv membrane protein fraction having a predicted *N*-terminal signal peptide,[11,12] meaning that it might be available for host–pathogen interactions. Such findings are relevant as our approach to designing an antituberculosis vaccine is aimed at identifying surface proteins which can interact with host cells.[41–44]

The *rv*3587c gene presence and transcription were ascertained, finding it in all *M. tuberculosis* complex bacterial strains evaluated here; it was also observed in *M. smegmatis* in normal culture conditions. Such experimental evidence supported the aforementioned bioinformatics analysis predicting it inside and outside the MTB complex and as an essential gene for *Mtb*H37Rv *in vitro* growth.[14] This gene could play an essential role in other mycobacterial strains and has also been reported in mycobacteria from macrophages infected with *Mtb* H37Rv and H37Ra.[45]

Once the *rv*3587c gene presence and transcription were ascertained (finding it in all *M. tuberculosis* complex bacterial strains evaluated), receptor-ligand assays were carried out using synthetic peptides covering the whole Rv3587 protein sequence. These assays form an important part of our approach to identifying HABPs in the search for candidate vaccine antigens and developing pharmaceutical and biological matrices.[46] These assays led to 9 HABPs being identified in the Rv3587c sequence; 4 of them (from the ^{116}Leu to ^{193}Thr region) had high specific binding to the cell lines used as infection target. These HABPs had intermediate affinity constants, similar to those in antigen-antibody interaction (K_D: 10^{-6} M), positive cooperativity and a significant amount of binding sites. Bearing in mind that HABPs can inhibit mycobacterial invasion of host cells (since these peptides can occupy binding sites on target cells usually occupied by mycobacteria for invading them), it was found that most HABPs inhibited *Mtb*H37Rv entry by more than 50%. It was also found that peptides 39265 and 39266 (conserved amongst mycobacterial strains) could significantly inhibit mycobacterial entry to both target cells (i.e. as HABPs) in a concentration-dependent way which depended on blocking target cells' specific receptor sites. These have been the only Rv3587c peptides of interest to date in designing a multi-epitope synthetic vaccine due to their binding, inhibition and secondary structure characteristics which can be modified to produce inmunogenic sequences.

Some peptides which inhibited mycobacterial entry were not taken into account when designing an anti-tuberculosis vaccine, as they did not have specific infection target cell binding. An inhibitory effect due to cytotoxicity affecting target cells was ruled out; such sequences could be of interest in other studies searching for drugs. Further assays are required for ascertaining such interaction and effect on mycobacteria.

The aforementioned bioinformatics approach predicted Rv3587c peptides' secondary and tertiary structures; however, such predictions for various peptides had little similarity regarding dichroism-related experimental evidence, possibly due to the predictive methods[33,36] for secondary structure being based on homology regarding complete protein sequences rather than on their fragments.

It has been reported that the humoral immune response seems to have an important role in tuberculosis.[47] Few antibodies recognised synthetic peptides in all the groups of individuals evaluated here; however, recognition in the region between aa ^{175}A-^{251}P was similar to that observed for *Mtb* lysate. Although HD donors were negative for Quantiferon and *Mtb* presence tests, they could have been in contact with other types of mycobacteria (environmental and *M. bovis* BCG (obligatory anti-tuberculosis vaccine in Colombia)) and have had antibodies capable of recognising peptide 39266 to 39269 sequences in Rv3587c which is conserved in many *Mycobacterium* strains and species, as shown by the BLAST results.

HABP 39265 (having high mycobacterial entry inhibition capability *in vitro*) was poorly recognised by individuals' sera when evaluated by ELISA assays; there was thus no relationship between target cells' binding characteristics, inhibiting *Mtb* entry and/or peptides' antigenicity, a phenomenon we have called conserved antigens' immunological code of silence.[7,48,49] Binding characteristics can be explained if the approach to protein tertiary structure is taken into account; it was observed that HABPs are exposed in a protein's external region. Peptide 39265 must be modified to endow it with the structural characteristics needed to elicit a protective immune response, according to our experimental model.

5. Conclusion

This work has thus highlighted Rv3587c HABPs being able to inhibit *Mtb* invasion of macrophages and epithelial cells (i.e. a fundamental event in developing tuberculosis). Peptides 39265 and 39266 having high A549 cell and U937 macrophage binding capacity which can inhibit mycobacterial entry can be considered part of a synthetic antituberculosis vaccine; however, taking that learned in developing a synthetic antimalarial vaccine into account, such HABPs must be modified to promote MCH-peptide-T-cell receptor complex formation and improve their capability of producing protective immunity against *M. tuberculosis*.

Competing interests

All authors declare they have no actual or potential competing interests.

Funding

This research was supported by the Colombian Science, Technology and Innovation Institute "Francisco Jose de Caldas" (COLCIENCIAS) through contract 0732016.

Acknowledgements

We would like to thank Jason Garry for translating the manuscript.

Author contributions

All authors have read and approved the manuscript. Experiments were conceived and designed by M.C., M.O., M.A.P., M.E.P. Acquisition of data: M.C., D.R., M.V. and H.C. Analysis, interpretation of data and drafting of manuscript: M.C., M.O.

A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.bmc.2018.03.044.

References

1. WHO, World Health Organisation. Global tuberculosis report 2017, 2017.
2. Moliva JI, Turner J, Torrelles JB. Prospects in *Mycobacterium bovis* Bacille Calmette et Guerin (BCG) vaccine diversity and delivery: why does BCG fail to protect against tuberculosis? *Vaccine*. 2015;33:5035–5041.
3. Sharma AK, Dhasmana N, Dubey N, et al. Bacterial virulence factors: secreted for survival. *Indian J Microbiol*. 2017;57:1–10.
4. Awuh JA, Flo TH. Molecular basis of mycobacterial survival in macrophages. *Cell Mol Life Sci*. 2017;74:1625–1648.
5. Ernst JD. Macrophage receptors for *Mycobacterium tuberculosis*. *Infect Immun*. 1998;66:1277–1281.
6. Ocampo M, Patarroyo MA, Vanegas M, Alba MP, Patarroyo ME. Functional, biochemical and 3D studies of *Mycobacterium tuberculosis* protein peptides for an effective anti-tuberculosis vaccine. *Crit Rev Microbiol*. 2014;40:117–145.
7. Patarroyo ME, Patarroyo MA. Emerging rules for subunit-based, multiantigenic, multistage chemically synthesized vaccines. *Acc Chem Res*. 2008;41:377–386.
8. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393:537–544.
9. Song H, Sandie R, Wang Y, Andrade-Navarro MA, Niederweis M. Identification of outer membrane proteins of *Mycobacterium tuberculosis*. *Tuberculosis*. 2008;88:526–544.
10. Marmiesse M, Brodin P, Buchrieser C, et al. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology*. 2004;150:483–496.
11. Malen H, Berven FS, Fladmark KE, Wiker HG. Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics*. 2007;7:1702–1718.
12. de Souza GA, Leversen NA, Malen H, Wiker HG. Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *J Proteomics*. 2011;75:502–510.
13. Malen H, Softeland T, Wiker HG. Antigen analysis of *Mycobacterium tuberculosis* H37Rv culture filtrate proteins. *Scand J Immunol*. 2008;67:245–252.
14. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog*. 2011;7:e1002251.

15. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct*. 2012;7:12.
16. Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 2001;17:646–653.
17. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–786.
18. Rashid M, Saha S, Raghava GP. Support Vector Machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs. *BMC Bioinf*. 2007;8:337.
19. Caceres SM, Ocampo M, Arevalo-Pinzon G, Jimenez RA, Patarroyo ME, Patarroyo MA. The *Mycobacterium tuberculosis* membrane protein Rv0180c: Evaluation of peptide sequences implicated in mycobacterial invasion of two human cell lines. *Peptides*. 2011;32:1–10.
20. Rodriguez DC, Ocampo M, Varela Y, Curtidor H, Patarroyo MA, Patarroyo ME. Mce4F *Mycobacterium tuberculosis* protein peptides can inhibit invasion of human cell lines. *Pathog Dis*. 2015;73.
21. Rodriguez DM, Ocampo M, Curtidor H, Vanegas M, Patarroyo ME, Patarroyo MA. *Mycobacterium tuberculosis* surface protein Rv0227c contains high activity binding peptides which inhibit cell invasion. *Peptides*. 2012;38:208–216.
22. Houghten RA. 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. *PNAS*. 1985;82:5131–5135.
23. Merrifield RB. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem Soc*. 1963;85:2149–2154.
24. Tam JP, Heath WF, Merrifield RB. SN 1 and SN 2 mechanisms for the deprotection of synthetic peptides by hydrogen fluoride. *Studies to minimize the tyrosine alkylation side reaction, Int J Pept Protein Res*. 1983;21:57–65.
25. Ocampo M, Aristizabal-Ramirez D, Rodriguez DM, et al. The role of *Mycobacterium tuberculosis* Rv3166c protein-derived high-activity binding peptides in inhibiting invasion of human cell lines. *Protein engineering, design & selection: PEDS*. 2012;25:235–242.
26. Rodriguez DC, Ocampo M, Reyes C, et al. Cell-Peptide Specific Interaction Can Inhibit *Mycobacterium tuberculosis* H37Rv Infection. *J Cell Biochem*. 2016;117:946–958.
27. Vera-Bravo R, Torres E, Valbuena JJ, et al. Characterising *Mycobacterium tuberculosis* Rv1510c protein and determining its sequences that specifically bind to two target cell lines. *Biochem Biophys Res Commun*. 2005;332:771–781.
28. Diaz DP, Ocampo M, Pabon L, et al. *Mycobacterium tuberculosis* PE9 protein has high activity binding peptides which inhibit target cell invasion. *Int J Biol Macromol*. 2016;86:646–655.
29. Provencher SW, Glockner J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*. 1981;20:33–37.
30. Sreerama N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem*. 1993;209:32–44.
31. Sreerama N, Venyaminov SY, Woody RW. Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein science: a publication of the Protein Society*. 1999;8:370–380.
32. Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins*. 1999;35:307–312.
33. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. *Trends Biochem Sci*. 2000;25:147–150.
34. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf*. 2008;9:40.
35. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–858.
36. Jones DT, Bryson K, Coleman A, et al. Prediction of novel and analogous folds using fragment assembly and fold recognition. *Proteins*. 2005;61:143–151.
37. Benkert P, Schwede T, Tosatto SC. QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Struct Biol*. 2009;9:35.
38. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–1612.
39. Vizcaino C, Restrepo-Montoya D, Rodriguez D, et al. Computational prediction and experimental assessment of secreted/surface proteins from *Mycobacterium tuberculosis* H37Rv. *PLoS Comput Biol*. 2010;6:e1000824.
40. Patarroyo ME, Bermudez A, Moreno-Vranich A. Towards the development of a fully protective *Plasmodium falciparum* antimalarial vaccine. *Expert Rev Vaccines*. 2012;11:1057–1070.
41. Curtidor H, Patarroyo ME, Patarroyo MA. Recent advances in the development of a chemically synthesised anti-malarial vaccine. *Expert Opin Biol Ther*. 2015;15:1567–1581.
42. Moreno-Vranich A, Patarroyo ME. Steric-electronic effects in malarial peptides inducing sterile immunity. *Biochem Biophys Res Commun*. 2012;423:857–862.
43. Patarroyo ME, Bermudez A, Moreno-Vranich A. Towards the development of a fully protective *Plasmodium falciparum* antimalarial vaccine. *Expert Rev Vaccines*. 2012;11:1057–1070.
44. Patarroyo ME, Bermudez A, Patarroyo MA. Structural and immunological principles leading to chemically synthesized, multiantigenic, multistage, minimal subunit-based vaccine development. *Chem Rev*. 2011;111:3459–3507.
45. Li AH, Waddell SJ, Hinds J, et al. Contrasting transcriptional responses of a virulent and an attenuated strain of *Mycobacterium tuberculosis* infecting macrophages. *PLoS ONE*. 2010;5:e11066.
46. de Jong LA, Uges DR, Franke JP, Bischoff R. Receptor-ligand binding assays: technologies and applications. *J Chromatogr B*. 2005;829:1–25.
47. Rao M, Valentini D, Poiret T, et al. B in TB: B Cells as mediators of clinically relevant immune responses in tuberculosis. *Clin Infect Dis*. 2015;61(Suppl. 3): S225–S234.
48. Garzon-Ospina D, Forero-Rodriguez J, Patarroyo MA. Inferring natural selection signals in *Plasmodium vivax*-encoded proteins having a potential role in merozoite invasion. *Infect Genet Evol*. 2015;33:182–188.
49. Buitrago SP, Garzon-Ospina D, Patarroyo MA. Size polymorphism and low sequence diversity in the locus encoding the *Plasmodium vivax* rhoptry neck protein 4 (PvRON4) in Colombian isolates. *Malar J*. 2016;15:501.