Current Research in Structural Biology 3 (2021) 165–178

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology

Structure of the *Mycobacterium tuberculosis* cPknF and conformational changes induced in forkhead-associated regulatory domains

Sindy Cabarca [a,b,c], Maximilia Frazão de Souza [d], Andrew Albert de Oliveira [e], Gabriel S. Vignoli Muniz [f], M. Teresa Lamy [f], Caio Vinicius dos Reis [g], Jessica Takarada [g], Brian Effer [h], Lucas Santos Souza [b], Lilia Iriarte de la Torre [a,b,c], Rafael Couñago [g], Cristiano Luis Pinto Oliveira [d], Andrea Balan [b,*]

[a] Programa de Pós-graduação em Genética, Universidade Estadual de Campinas, Campinas, 13083-862, SP, Brazil
[b] Laboratório de Biologia Estrutural Aplicada LBEA, Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, 05508-000, São Paulo, SP, Brazil
[c] Grupo Investigaciones Biomédicas, Universidad de Sucre, Sincelejo, 700001, Sucre, Colombia
[d] Grupo de Fluidos Complexos, Departamento de Física Experimental, Instituto de Física, Universidade de São Paulo, São Paulo, 05508-090, SP, Brazil
[e] Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 14049-900, SP, Brazil
[f] Departamento de Física Geral, Instituto de Física, Universidade de São Paulo, São Paulo, 05508-090, Brazil
[g] Structural Genomics Consortium, Universidade Estadual de Campinas, Campinas, 13083-886, São Paulo, Brazil
[h] Center of Excellence in Traslational Medicine (CEMT) and Scientific and Technological Bioresource Nucleus (BIOREN), Universidad de la Frontera, Temuco, 01145, Chile

ARTICLE INFO

Handling Editor: Dr P Emmanouil

Keywords:
FHA domains
Membrane
Serine/threonine protein kinase
Phosphorylation
ABC transporter

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) has 11 Serine-Threonine Protein Kinases (STPK) that control numerous physiological processes, including cell growth, cell division, metabolic flow, and transcription. PknF is one of the 11 Mtb STPKs that has, among other substrates, two FHA domains (FHA-1 and FHA-2) of the ATP-Binding Cassette (ABC) transporter Rv1747. Phosphorylation in T152 and T210 located in a non-structured linker that connects Rv1747 FHA domains is considerate to be the regulatory mechanism of the transporter. In this work, we resolved the three-dimensional structure of the PknF catalytic domain (cPknF) in complex with the human kinase inhibitor IKK16. cPknF is conserved when compared to other STPKs but shows specific residues in the binding site where the inhibitor is positioned. In addition, using Small Angle X-Ray Scattering analysis we monitored the behavior of the wild type and three FHA-phosphomimetic mutants in solution, and measured the cPknF affinity for these domains. The kinase showed higher affinity for the non-phosphorylated wild type domain and preference for phosphorylation of T152 inducing the rapprochement of the domains and significant structural changes. The results shed some light on the process of regulating the transporter's activity by phosphorylation and arises important questions about evolution and importance of this mechanism for the bacillus.

1. Introduction

Reversible protein phosphorylation is the main cellular mechanism by which signals from the environment are transmitted to the interior of cells to be translated into responses, both in prokaryotes and eukaryotes (Av-Gay and Everett, 2000; Barthe et al., 2009). In the latter, these types of post-translational modifications play a crucial role in most cellular processes (Panni, 2019). Typically, phosphorylation activates the substrate to perform a specific activity or cellular localization and/or transfer the phosphate group to a subsequent effector, initiating a signaling cascade in response to the signal. The reverse reaction leaves the effector proteins to their initial state, ready to start a new signaling event. In this way, kinases and phosphatases act as molecular switches modulating specific signal transduction pathways (Canova and Molle, 2014). Therefore, Serine/Threonine protein kinases (STPKs) consist of important environmental sensing mechanism in bacteria, as it can not only alter the biological function of proteins, but also increase their ability to associate with other proteins through phosphorylation-dependent

* Corresponding author.
E-mail addresses: abalan@usp.br, abalan301@gmail.com (A. Balan).

https://doi.org/10.1016/j.crstbi.2021.07.001
Received 9 April 2021; Received in revised form 15 July 2021; Accepted 19 July 2021

interactions. In *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis, there are eleven STPKs that play a central role in the physiology and pathogenesis, acting over hundreds of substrates that participate in all biological processes of the bacterium (Prisic and Husson, 2014; Wagner et al., 2019). The maintenance of these genes during evolution and the up-regulation of at least eight members of this family during infection in *Mtb* reveal their relevance for the pathogenesis (Baer et al., 2014).

This work concerns the *Mtb* PknF STPK (Rv1746) that participates in several cellular processes of physiological importance due the diversity of substrates, such as cell division (HupB, ParB), arabino synthesis (EmbR and EmbR2), mycolic acid synthesis (FabD, FabH, KasA, KasB, HadAB/BC, InhA, PcaA), peptidoglycan synthesis (FhaA), TCA cycle (GarA), methionine cycle (SahH), signaling (PtkA), chaperone (GroE1) and transport (Rv1747) (Richard-Greenblatt and Av-Gay, 2017). A role for PknF was suggested in the regulation of growth, cellular morphology, glucose transport, sliding mobility and biofilm formation in *M. smegmatis* (Deol et al., 2005; Gopalaswamy et al., 2008), regulation of a cell wall transport component (Spivey et al., 2011; Richard-Greenblatt and Av-Gay, 2017) and phosphorylation of Rv0020 (Grundner et al., 2005) and MtFabH (Veyron-Churlet et al., 2009). Indeed, PknF phosphorylates the product of the closest gene present in the same operon, the non-canonical ATP-Binding Cassette (ABC) transporter Rv1747. Differently from ABC transporters described until now, Rv1747 has two FHA domains (FHA-1 and FHA-2) connected by a disordered region of ~100 amino acids that include two PknF target residues, T152 and T210 (Curry et al., 2005; Spivey et al., 2013). Replacement of these residues by alanine caused attenuation of *Mtb* growth in infection models (Spivey et al., 2011, 2013, 2013). Albeit function or transported substrate are not known yet, studies suggested the participation of Rv1747 in the translocation of bacterial cell wall components, possibly lipo-oligosaccharides or phosphatidyl-myo-inositol-mannosides (PIMs), and antibiotic extrusion. Indeed, four *Mtb* proteins with FHA domains are involved in processes connected to cell wall synthesis or remodeling (Pallen et al., 2002). Viability studies have shown that while the transporter is not essential for *in vitro* bacterial growth, its absence (Δ*rv1747* mutant) significantly decreased *Mtb* growth in *vivo* assays on bone marrow-derived macrophages and dendritic cells, associating its function with the infection and pathogenesis of the bacteria (Spivey et al., 2011, Curry et al., 2005). Phosphorylation of Rv1747 induces the phenomenon of phase separation and clustering of the transporter *in vitro* in membrane models and in live cells, as well as mediates the inter- and intra-molecular associations within the protein as a putative mechanism of Rv1747 transport activation (Heinkel et al., 2018, 2019).

In this work we carried out a biophysical characterization of the cytoplasmic catalytic domain of PknF (cPknF) and solved its three-dimensional structure in complex with the broad-spectrum human kinase inhibitor IKK16 (Waelchli et al., 2006; Chen et al., 2017). Small Angle X-Ray Scattering (SAXS) analysis showed the structural changes in cPknF in the non-phosphorylated and phosphorylated (cPknF*) states and its interaction with the FHA domains of Rv1747. We used the wild type (FHA_wt) and three phosphomimetic mutants that emulate phosphorylation in the target residues T152 (FHA_T152E), T210 (FHA_T210E) and in both (FHA_T152E/T210E). We showed that after the *in vitro* autophosphorylation, cPknF suffers structural changes and that its affinity for the FHA domains is dependent on their phosphorylation states. Similarly, the phosphorylation plays essential role for approximation of FHA-1 and FHA-2 domains and, in the stepwise of this movement, phosphorylation on T152 important for attraction of cPknF. Altogether, our data show the first structure of the *Mtb* cPknF and reveal important aspects of its interaction with Rv1747 FHA domains that might trigger the transport.

Table 1
Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

Parameters	cPknF-IKK16
Data collection	
Synchrotron and beamline	APS, BEAMLINE 24-ID-C
Temperature, K	80
Wavelength, Å	0.97
Resolution limit, Å	39.89–2.75 (2.848–2.75)
Space group	*P*2 (1)2 (1)2
Unit cell dimensions *a; b; c,* Å	100.13; 239.33; 53.98
α; β; γ, °	90.00; 90.00; 90.00
Total reflections	200514 (17870)
Unique reflections	34601 (3428)
Multiplicity	5.8 (5.2)
Completeness, %	99.41 (98.93)
Mean I/sigma(I)	11.14 (1.38)
Wilson B-factor, Å^2	62.42
R-merge	0.1453 (1.579)
R-meas	0.1599 (1.76)
R-pim	0.06569 (0.7616)
$CC_{1/2}$, %	0.997 (0.601)
CC*, %	0.999 (0.867)
Data refinement	
Reflections used in refinement	34543 (3412)
Reflections used for R-free	1685 (175)
R-work	0.2446 (0.3609)
R-free	0.2838 (0.4026)
CC(work)	0.925 (0.728)
CC(free)	0.897 (0.692)
Number of non-hydrogen atoms	7088
Macromolecules	6831
Ligands	140
Solvent	117
Protein residues	937
RMS(bonds)	0.013
RMS(angles)	1.72
Ramachandran favored, %	97.24
Ramachandran allowed, %	2.76
Ramachandran outliers, %	0.00
Rotamer outliers, %	0.31
Clashscore	2.62
Average B-factor, Å^2	72.16
Number of TLS groups	24

2. Material and methods

2.1. Phylogenetic analysis

Mycobacteria phylogenetic tree was based on 16S ribosomal RNA sequences obtained in the Mycobrowser website (https://mycobrowser.epfl.ch). The sequences represent species that belong to different complexes as defined for *Mycobacterium* genus. The evolutionary history was inferred using the Neighbor-Joining method and evolutionary analyses were conducted in MEGA X with 1000 bootstrap. The list of the species (NCBI reference) is showed in *SI Appendix,* Table A1. Searching for putative orthologs of Rv1746 and Rv1747 was performed using their protein sequences as query in BlastP in the non-redundant bank and using the parameters pre-established by the tool. Only sequences with coverage of 95 or higher and identity greater than 50 % were considered. Phylogenetic tree of the *Mtb* STPKs was obtained in ClustalW server (Larkin et al., 2007) after a multiple sequence alignment of the eleven *Mtb* sequences (function "build" of ETE3 v3.1.1) as implemented on the GenomeNet (https://www.genome.jp/tools/ete/). Branch supports were computed out of 100 bootstrapped trees.

2.2. Cloning and site-directed mutagenesis

The FHA domain of *rv1747* (residues 1 to 310, including FHA-1, linker and FHA-2) and the kinase domain of *pknF* (residues 1–299) were amplified from *Mtb* H37Rv chromosomal DNA by PCR using the two pairs of oligonucleotides: Fw_Rv1747_FHA_*Nco*I (5′ atataccatgggagtgccgatgagccaaccag 3′) and Rv_Rv1747_FHA_*Hin*dIII (5′ atatgaagctttcagttctcttcacggcgcg 3′) for *rv1747_FHA*, and Fw_PknF_*Eco*RI (5′ atatagaattcatgccgctcgcggaaggttcg 3′) and Rv_PknF_*Hin*dIII (5′

atataaagctttcacgcgggcgcggccacggcgaccggtt 3′) for *cpknF*. *rv1747_FHA* amplified fragment was cloned into the expression vectors pOP3BP (Addgene catalog number 112606) that adds a N-terminal His_8-tag and a GB1 protein Phusion. Similarly, *cpknF* fragment was cloned into pOP5GT vector (Addgene catalog number 112608) in fusion with a N-terminal His_8-tag and GST protein. The amplified PCR products and the expression vectors were digested at 37 °C for 2 h with the restriction enzymes *Nco*I and *Hind*III in the case of FHA, and *Eco*RI and *Hind*III for PknF. The digestion reaction was carried out following the guidelines of the Fast Digest enzyme manufacturer (Thermo Scientific, EUA). The fragments and vectors were purified from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Germany) and incubated in a 3:1 insert:vector molar ratio (150 ng of vector and 91 ng of FHA or 80 ng of PknF), with 5U of the T4 DNA ligase enzyme (Thermo Scientific, EUA) in the enzyme buffer in a 10 μl reaction mixture at 16 °C overnight according to the manufacturer's indications. The next steps were the transformation of electrocompetent DH5α cells, purification and analysis of these recombinant plasmids by colony PCR and restriction analysis with previously described protocols (Sambrook and Russell, 2001). All plasmids were confirmed by DNA sequencing.

Site-directed mutations of T152, T210 or both by glutamic acid were made in the coding region of Rv1747_FHA domain, to emulate different phosphorylation states of the domain. Three mutants of the FHA domain, FHA_T152E, FHA_T210E and FHA_T152E/T210E (double mutant), were produced. The recombinant plasmid containing the wild type domain (pOP3BP_$FHA_{1\text{-}310}$) was used as a template and the mutations were made using the Q5 Site-Directed Mutagenesis Kit (New England BioLabs, EUA) according to the instructions of the manufacturers. Mutations were confirmed using DNA sequencing.

2.3. Protein expression and purification

Overproduction of Rv1747_$FHA_{1\text{-}310}$ and cPknF$_{1\text{-}299}$ were performed, respectively, in *E. coli* strain BL21 (DE3) and *E. coli* BL21 Star (DE3). Luria-Bertani broth (LB) supplemented with ampicillin was inoculated with these cells harboring the recombinant plasmids and the cultures were incubated at 37 °C with shaking (180 r.p.m.) until the A_{600} reached 0.6. Protein expression induction was completed with addition of Isopropyl 1-thio-D-galactopyranoside (IPTG) at a final concentration of 1 mM for $FHA_{1\text{-}310}$ and 0.25 mM for cPknF$_{1\text{-}299}$. The bacterial growth was continued in shaker with 180 r.p.m. for 2 h at 37 °C and 18 h at 18 °C, respectively. Cells were harvested by centrifugation during 15 min at 7000 rpm at 4 °C. The cell pellets obtained during expression were resuspended in lysis buffer (100 mM Tris/HCl, pH 8.0, 100 mM NaCl, 5 % glycerol) containing DNase (5 μg/ml), RNase (5 μg/ml), lysozyme (1 mg/ml) and a mixture of protease inhibitors (Sigma Aldrich). Bacteria were disrupted by sonication (VibraCell, Sonics) on ice with 10 bursts of 20 s at amplitude 10. To obtain the supernatant, the lysate was centrifuged at 18 000 rpm at 4 °C for 45 min. The soluble extracts were applied to a HisTrap HP column 5 ml (GE Healthcare) equilibrated with binding buffer (100 mM Tris/HCl, pH 8.0, 100 mM NaCl, 5% Glycerol and 5 mM Imidazole) at flow rate of 1 ml/min^{-1}. The column was washed with 15 column volumes (CV) of binding buffer to remove the unbound material. Proteins were eluted by a linear elution gradient with Elution buffer (Binding buffer, with Imidazole at 500 mM) in an AKTA STAR® system (GE Healthcare). When the protein started to elute, the imidazole concentration was maintained. Flowthrough and eluted fractions were collected and analyzed by SDS-PAGE. Afterward, the GB1 and GST fusion proteins were removed by incubation overnight of the fractions with 3C and TEV protease, respectively, in a relation of 1:20 μM 3C/TEV:protein with addition of 1 mM EDTA and DTT. cPknF was separated of the GST protein fusion by passing again by the HisTrap 5 ml column at 0.5 ml min^{-1} and recuperation of the flow through. Before, imidazole, EDTA and DTT were remove of the sample by their dilution with binding buffer without imidazole. Finally, a size exclusion chromatography was performed to separate the proteins of interest from any remaining contaminant and eliminate the GB1 fusion protein from the FHA domain. It was used a pre-packed HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) equilibrated with GF Buffer (100 mM Tris/HCl, pH 8.0, 100 mM NaCl, 5% Glycerol) at flow rate 0.5 ml/min^{-1}. In the case of FHA domains, it was necessary to perform two steps of size exclusion chromatography, one after the affinity chromatography, and another after the protease cleavage.

2.4. Detection of the $FHA_{1\text{-}310}$ and cPknF$_{1\text{-}299}$ by SDS-PAGE and western blot assay

The expression and purification of recombinant proteins were monitored by SDS-PAGE electrophoresis and Western blot assays. Culture samples before induction (T0) and after induction of expression (T2), as well as samples of the various purification stages were mixed with protein loading buffer and then denatured at 95 °C for 5 min and loaded onto a 12 % SDS-PAGE gradient gel. Protein observation was performed by staining the gel with Coomasie brilliant blue. For the Western blot analysis, the proteins in the gel without staining were transferred to a polyvinylidene fluoride membrane (Merck Millipore, Massachusetts - USA) in electro-transfer buffer (25 mM Tris, 250 mM Glycine, 20 % Methanol) for 30 min at 100 V, then blocked overnight with 5 % skim milk in Tris-Buffered Saline with 0.05 % Tween 20 (TBS-T). For FHA and cPknF detection, monoclonal anti-His-Tag antibody produced in mouse (GE Healthcare) diluted at 1:20000 were used as primary antibody, followed by membrane incubation for 2 h. Afterward five washes with TBS-T, the membrane was incubated with the secondary antibody Anti-Mouse IgG (whole molecule) −Alkaline Phosphatase produced in goat (Sigma Aldrich, EUA) diluted 1:30 000 for 1 h. The membrane was revelated after previous washing with TBS-T and then adding the substrate for alkaline phosphatase BCIP/NBT reagent (Sigma-Aldrich, EUA).

2.5. Evaluation of the protein folding by Differential Scanning Fluorimetry (DSF)

DSF was used to evaluate the stability of proteins of interest based on the fluorescence emitted by SYPRO Orange dye (Sigma-Aldrich, Catalog Number S5692; 5000x stock solution; 10 mM) according to the stability of the tertiary structure. As proteins are denatured because of heating, will expose their hydrophobic regions, normally hidden. SYPRO Orange can interact with these hydrophobic surfaces and increases its quantum yield (Kroeger et al., 2017). The assays were carried out in triplicate and utilizing optical 96 well plates (Bio-Rad). The reaction was composed of 10 μM protein and 5x SYPRO Orange (10 μM) in GF buffer. DSF was measured in a real-time thermocycler (Bio-Rad CFX connect) with filters of excitation/detection from 450 to 580 nm and a temperature range from 25 °C to 95 °C with a linear increment of 0.5 °C by 30 s of dwell. The Melting Temperature (Tm) was determined by Bio-Rad CFX Manager software applying –d (RFU)/dT.

2.6. cPknF activity assay and mass spectrometry analysis

Presence of phosphorylated sites in cPknF was confirmed by intact mass spectrometry analysis considering that each phosphate group added approximately 80 Da to the protein. The analyses were performed at the Proteomics Service of the Central Laboratory of High-Performance Technologies in Life Sciences (LaCTAD) of the State University of Campinas (UNICAMP, Campinas, SP, Brazil) in the equipment LC-MS/MS coupled to a mass spectrometer with a quadrupole-time-of-flight (QTof) analyzer model Waters XEVO G2 XS, equipped with an ESI type ionization source. 1 μL of each purified sample of cPknF, FHA and cPknF + FHA_wt (3 μM each) were submitted to the mass intact spectrometry analysis before and after the *in vitro* phosphorylation reaction in buffer containing Tris-HCl pH 8.0100 mM, NaCl 100 mM, DTT 1 mM, EDTA 5 mM, $MgCl_2$ 5 mM and ATP 5 mM for 30 min, at room temperature.

2.7. Circular dichroism (CD) and fluorescence measurements

CD assays were performed in the range of 190–260 nm on the Jasco J-810 spectropolarimeter (JASCO, Easton, MD USA) using the facility of the Laboratory of Biotechnology at Institute Butantan, São Paulo, Brazil. Data were collected with a 50 nm/min scan with 20 accumulations for the sample at 5 μM and 10 accumulations for the buffer (20 mM Tris-HCl pH 8.0 and 10 mM NaCl). The CD spectra for each sample were obtained using the CAPITO online tool (Wiedemann et al., 2013). On the other hand, UV–visible absorption spectra were obtained with an UV–visible spectrophotometer (Varian Cary, Santa Clara, CA). In addition, the fluorescence emission spectrum was recovered with a Varian Cary Eclipse fluorimeter (Santa Clara, CA). Samples were placed in quartz cuvettes, with an optical pathway of 4 mm (10 × 4 mm) and the experiments performed at room temperature (22.5 °C), with 1 ml samples of proteins (10 μM) in GF buffer. Emission spectra were obtained using an excitation beam light at 295 nm. No inner filter correction was necessary as Absorbance value at 295 nm was found to be smaller than 0.1 (Albrecht, 2008).

2.8. Crystallization and structure determination

Non-phosphorylated protein (confirmed by mass spectrometry) was used in the crystallographic experiments at 14 mg/ml (426 μM). To improve the protein stability and chances of obtaining crystals, the competitive inhibitors of kinases, JNJ-7706621, IKK16 and AZD 5438 (SGC library, UNICAMP, Campinas, SP, Brazil), diluted to 10 mM 100 % DMSO, as well as the non-hydrolysable ATP analog, ATP-γ-S, were added at a final concentration of 174 μM in buffer GF containing 29 μM cPknF (0.96 mg/ml). A sample containing protein and ligand in a molar ratio 1:6 was incubated for 30 min in ice and concentrated using the Amicon Ultra-0.5 Centrifugal Filter Unit up to 14 mg/ml of protein, now with the bound compound. cPknF:inhibitor crystals appeared. Crystallization trials of cPknF:inhibitor were screened in 96-well plates against sparse matrix crystallization commercial kits LFSG, HCS3, and SALT-RX2 (Molecular Dimensions, Maumee, OH, USA) using the sitting-drop vapor diffusion process. Crystals of cPknF with IKK16 were obtained after three days when maintained at constant temperature of 18 °C in 0.1 M Citrate pH 5.5, 200 mM Ammonium acetate and 30 % (w/v) polyethylene glycol (PEG) 4000. The diffraction dataset was collected at Advanced Photon Source (APS) 24 ID-C beamline, by fine ϕ-slicing (0.2°), and achieved 2.75 Å maximum resolution after processing with autoPROC (Vonrhein et al., 2011). Phaser (McCoy et al., 2007) was used for molecular replacement (MR), applying the *Mtb* kinase domain - cPknI (PDB 5M07) as assemble. ClustalW was used to align the accessible sequences of homologous PDB protein structures. Homology modelling using MODELLER (Eswar et al., 2006), together with the structural alignment of kinase domains from PknA, PknB, PknE and PknI (r.m.s.d. = 0.295) were used for the strategic removal of flexible loops and helped to establish a logical sub-domain division between N-lobe and C-lobe modules for better MR solutions and model building. C-lobe module of cPknF was used for building initial search. After several modelling cycles in Coot (Emsley et al., 2010) and autoBuild (Terwilliger et al., 2008), the satisfactory location of the two other modules in liaison to the p-lobe validated our possible MR solution. Reciprocal space refinement was carried out by phenix.refine (Adams et al., 2010) and BUSTER-TNT (Blanc et al.,

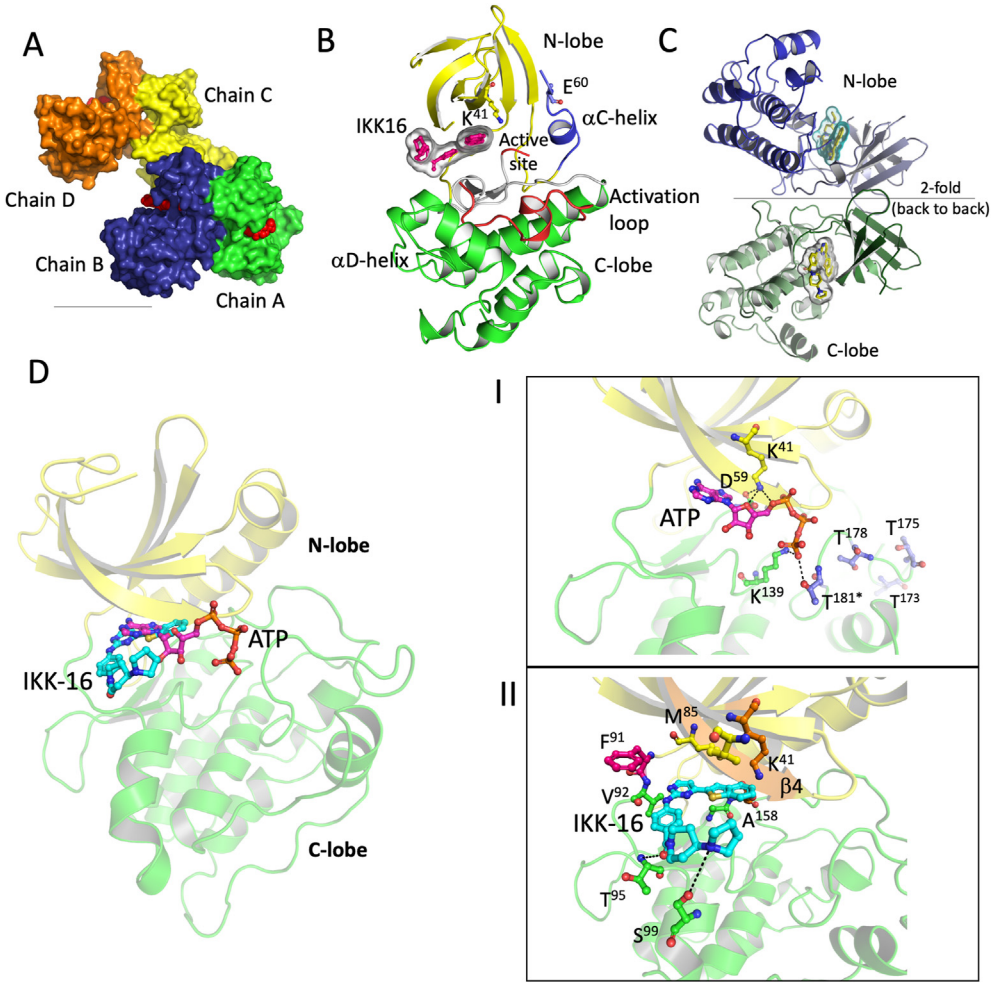


Fig. 1. The three-dimensional structure of *Mtb* cPknF domain and active site. (*A*) Surface representation of the four molecules of the crystal asymmetric unit (chain A to D) organized in two dimers. The IKK16 inhibitor is shown as red spheres. (*B*) A single molecule of cPknF with the N and C-terminal lobes in yellow and green cartoon, respectively. N-terminal lobe consist of six β antiparallel twisted sheet (β1-β6) and C-terminal is an a-helix bundle (C1–C6). The active site lies on the cleft between the two lobes, on which is bound the IKK16 kinase inhibitor (in surface and pink sticks). The αC-helix (blue) and activation loop (red) are the important secondary structures for catalysis. (*C*) Dimeric state of cPknF with the N-terminal back-to-back interface. The two molecules are shown in pale blue and green cartoon, respectively. IKK16 inhibitor is represented in yellow sticks and surface. (*D*) Positioning of IKK16 and ATP in the active site of *Mtb* cPknF. N and C lobes are shown in yellow and green cartoon, respectively, with the ligands in the active site. Details of the cPknF with ATP and IKK16 are shown in the frames I and II, respectively, evidencing the main residues for interactions.

2004) supported by the ligand restraints, which were generated using geometry and coordinates analysis in Mogul (Bruno et al., 2004; Cottrell et al., 2012). Control of the model quality was performed with MOLPROBITY (Williams et al., 2018) and the Ramachandran statistics are reported in Table 1. The PISA (Krissinel, 2010) webserver was used for the recognition of surface contacts area that is associated with various subunits, hydrophobic or polar regions. Further electrostatic measurements were accomplished with APBS Tools (Baker et al., 2001).

2.9. Small-angle X-ray scattering experiments (SAXS)

Small angle X-ray scattering experiments were performed on a lab-based Xenocs-Xeuss 2.0 equipment operating with CuKα, $\lambda = 1.52$ Å, radiation from a GENIX 3D source with optics arrangement of FOX 3D-Xenocs beam focus and two sets of scatterless slits to collimation. SAXS data were carried out to samples of FHA wild type, FHA_T152E, FHA_T210E, FHA double mutant, cPknF in non-phosphorylated and phosphorylated states (cPknF*) using samples concentration of 1.0, 0.6, 0.6, 1.2, 1.0, respectively for FHA domains, and 1.7 mg/ml for cPknF. Mixed samples of FHA wild type + cPknF and FHA wild type + cPknF*, each protein in the concentration of 0.5 mg/ml, also were submitted to SAXS analyses. All the samples and each buffer were injected in a Xenocs low noise flow cell and exposed time of 30 min to 2D SAXS data collections recorded on a photon-counting Pilatus 300k detector at a sample-to-detector distance of 1.2 m, in a total of 3 frames per sample. The 2D isotropic scattering imagens were submitted to azimuthal integration resulting the 1D SAXS profile intensity as function of momentum transfer vector modulus (q) defined by $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle. Afterward, the data treatment to 1D SAXS profile were performed with the subtraction of respectively buffer from each sample, finally the water and empty cell SAXS experiments performed in the same condition was used to determine the final samples SAXS profile in absolute scale. All the data treatment procedure was performed using the program package SUPERSAXS (Oliveira and Pedersen, unpublished). The SAXS data profiles were fitted using GNOM program (Semenyuk and Svergun, 1991) where the pair distribution function $[p(r)]$ is calculated using Indirect Fourier Transformation (IFT) procedures. The maximum diameter (D_{max}) and the radius of gyration (R_g) were derived from the $p(r)$ and GNOM parameters calculation results.

2.10. Structural modelling and dynamics analysis of FHA domain and docking simulations of cPknF*-FHA complex

The investigation of the full-length FHA interdomain orientation was performed using the ensemble optimization method (EOM) program (Tria et al., 2015), applying different tools of the EOM routine. First, a random pool of 10 000 structures of FHA-1 and FHA-2 connected by a long linker of 95 residues were built to using the RANCH program with input of the crystal model to FHA-1 (PDB 6CCD) and NMR model to FHA-2 (PDB 6CAH), including no symmetry constraint. Afterward, the pool was used as input to GAJOE program, which selected an optimized ensemble that best fit the collected SAXS data to FHA wild type and mutant samples, individually. This step was based on the standard EOM configuration process for the method applied in the genetic algorithm (GA).

Docking simulations between the cPknF* domain and both crystal structure of FHA-1 (PDB 6CCD) and NMR ensemble of FHA-2 (PDB 6CAH) were performed using the HADDOCK2.4 webserver (de Vries et al., 2010). To further investigate the interaction between the kinase and its substrate, the missing activation loop region of cPknF* (A158-T181) was modeled with a LoopModel protocol from MODELLER and the residues T173 and T175 substituted by their phosphorylated state. Prior to docking simulations, the cPknF* structure was submitted to energy minimization using steepest descent method from GROMACS (Berendsen et al., 1993). Both phosphorylated residues were defined as the active residues of cPknF, and the motifs GR, SRxH, and SxNG were set as the active residues of FHA domains. For both receptor and ligand, passive residues were automatically assigned around active residues. To improve docking prediction, the intrinsically disordered region of the kinase activation loop was set as a full-flexible segment. The top 400 solutions generated were clustered by the fraction of common contacts with a cutoff of 0.75 and scored by HADDOCK score. The four representative models of the top 2 clusters were further ensembled and sent to the HADDOCK refinement function. The final docking solutions were analyzed by PISA (Krissinel, 2010).

2.11. Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) experiments were performed at 20 °C with 74 μM of FHA wild type (FHA_wt) and mutant samples (FHA_T152E, FHA_T210E and double mutant) and 132 μM of cPknF*, in a MicroCal ITC200 (GE Healthcare). Samples of cPknF* were accommodated into the calorimeter cell and the syringe was filled with FHA samples in the same buffer. The ITC thermograms were collected in a total of 56 injection with the first injection of FHA sample of 0.1 μl, followed by 0.7 μl for the remaining injections with a time interval of 150s between each. ITC experiments were carried out to solution buffer of cPknF* with the same configuration to each FHA sample titration, then used as background enthalpy variation to FHA sample injection. The Origin 7.0 software provided by the instrument was used to integrate the thermogram picks and to analyze the colorimetry final profile, assuming a stoichiometric binding of N sites with same affinity, where the final dissociation constant, K_d, binding enthalpy (ΔH) and entropy (ΔS) were determined.

3. Results

3.1. rv1746/rv1747 operon is highly conserved in the Mtb complex

The *rv1746/rv1747* operon encodes the STPK PknF and the ABC transporter Rv1747, respectively. The predicted PknF structure indicates that its cytoplasmic catalytic domain (cPknF) is connected to an extracellular sensor domain (ePknF) by a transmembrane helix. It is expected that the dimerization in the membrane is needed for activation of these proteins (Alber, 2009). The ABC transporter Rv1747 forms a homodimer consisting of two transmembrane domains (TMDs) with predicted six α-helices each, two nucleotide-binding domains (NBDs, residues 330–559) and two extraordinary FHA domains (FHA-1 and FHA-2, residues 1–310), which are connected by a disordered region where the phosphorylation targets T152 and T210 are localized (*SI Appendix,* Fig. A1*A*). The *rv1746/rv1747* operon is spread among species of biological and medical relevance in the *Mycobacteria* genus (*SI Appendix,* Table A1) and identical in those that belong to the *Mtb* complex, including the human pathogens *Mtb* H37Rv, *M. bovis*, *M. caprae* and *M. microti* and the opportunistic *M. canetti* (*SI Appendix,* Fig. A1*B*). The presence of operons with relative high similarity when compared to *rv1746/rv1747* and that encode a putative STPK in tandem with transmembrane transporters is also observed in species such as *M. smegmatis* group, *M. marinum*, *M. ulcerans, M. kansasii* group and *M. avium* complex, suggesting that the corresponding proteins have an important role in the genus and highlights the importance of the phosphorylation as a regulation mechanism of the transport. Species that belong to the *Mtb* complex also conserved the eleven genes encoding the STPKs. The phylogenetic relationship among these *Mtb* kinases as previously described (Narayan et al., 2007) indicated that PknF is closely related to PknJ and PknI. With the exception of PknG and PknK that show large cytoplasmic domains, all the other kinases of *Mtb* have similar structural organization with a cytoplasmic catalytic domain, a transmembrane helix and an extracellular sensor domain (*SI appendix,* Fig. A2). The cPknF shared between 28 % and 37 % of amino acid sequence identity with the corresponding domains of the *Mtb* kinases, including PknB, PknD and PknE, enzymes that also phosphorylate Rv1747 and other FHA domains

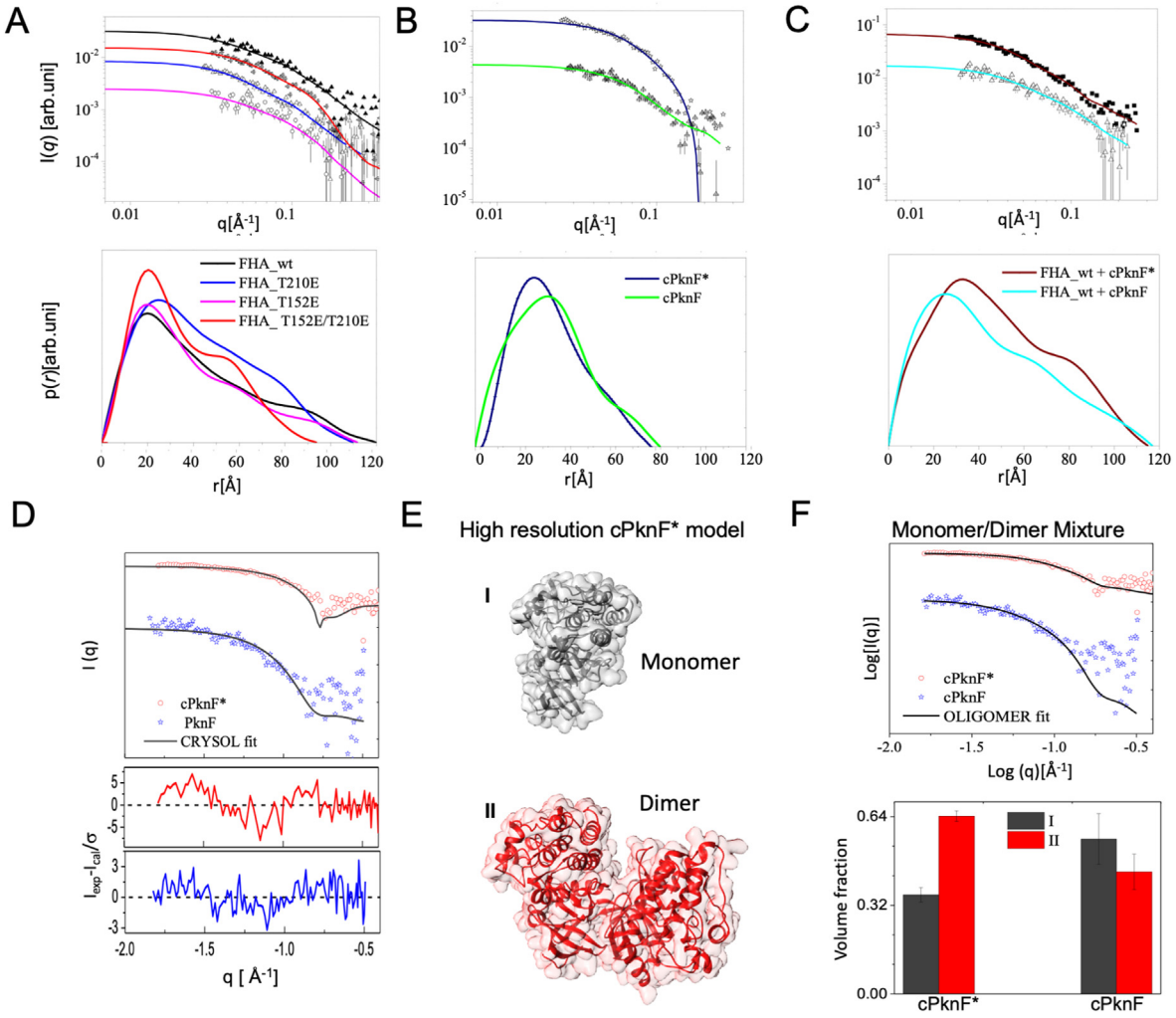


Fig. 2. SAXS profile fit and oligomerization studies of Mtb Rv1747 FHA domains and cPknF. SAXS experimental intensity as a function of the modulus of the scattering vector ($q = 4\pi \sin\theta/\lambda$) (symbol) and fitting results from EOM program using IFT method (line) (upper graphs) and pair distribution functions, p(r), respectively to each fitted sample (bottom graphs). (A) SAXS data to FHA wild type and the mutants; (B) SAXS data to cPknF (non-phosphorylated) and cPknF* (phosphorylated) in buffer solution. (C) SAXS data to the mixture samples of FHA wild type and cPknF* and cPknF. (D) Oligomerization states of cPknF* and cPknF. Superposition of theoretical SAXS profile calculated using CRYSOL program and the experimental data from cPknF samples in the phosphorylated and non-phosphorylated states as monomeric state as input. (E) High-resolution three-dimensional structures of cPknF in the monomeric (gray cartoon and surface) and dimeric (red cartoon and surface) are shown in I and II, respectively. (F) Upper graph: SAXS profile analysis assuming a mixture of monomers and dimers in solution using the program OLIGOMER and the models in (E) as input. Bottom graph: Plot of the volume fraction of cPknF* and cPknF assuming a mixture of PknF* in monomeric (I: black) and dimer (II: red) states.

(Grundner et al., 2005). The available three-dimensional structures of extracellular domains of PknD (PDB 1RWL), PknH (PDB 4ESQ) and PknI (PDBs 5CLL and 5XLM) show they have differences in the folding. An analysis of the genomic context of the kinase genes in *Mtb* genome show they are close to genes that are involved in a diversity functions, which is in accordance with their multiplicity of targets (*SI appendix,* Table A2). Genes *pknG, pknB, pknA* and *pknD* have in common with *pknF* the vicinity of membrane proteins. *pknD* is located in a cluster that encodes components of an ABC transporter dedicated to the phosphate uptake (Braibant et al., 2000; Oliveira and Balan, 2020) and *pknG* is close to genes *glnH* (*rv0411*) and *rv0412,* which encode a periplasmic glutamine-binding protein and a membrane protein, respectively. *pknB* and *pknA* share the same operon that encodes a penicillin-binding protein (PBP, *rv0016c*) and the RodA protein involved with cell division (*rv0017c*) (*SI appendix,* Table A2). Resuming, our data showed that the *rv1746/rv1747* operon is highly conserved in *Mycobacterium* species suggesting the relevance of the encoded proteins for physiological processes. The maintenance of the kinase gene to always close transporter genes also reinforce its role in the transport regulation.

3.2. Overall three-dimensional structure of catalytic domain of Mtb PknF

The structure of the catalytic domains of *Mtb* PknA, PknB, PknE, PknG and PknI have been determined, but thus far there was no structural information for the corresponding domain of PknF. To gain insight on how it exerts its action, we pursued protein crystallization and structure determination of cPknF (residues 1–299). The domain was produced in *Escherichia coli* BL21 (DE3) Star overexpressing the *E. coli* lambda phosphatase to obtain a non-phosphorylated protein and to reduce sample heterogeneity. The recombinant cPknF was obtained after purification and cleavage with TEV protease with a molecular mass of 32.8 kDa. The autophosphorylated cPknF (cPknF*) was obtained in the presence of ATP and confirmed by intact mass spectrometry. *Mtb* cPknF* incorporated up to seven phosphate groups into its structure and presented molecular weight of 33 378 Da (*SI Appendix,* Fig. A3).

Attempts to obtain crystals of apo cPknF were not successful. Protein kinases have a conserved structural architecture composed of two lobes linked by a flexible hinge region. In solution, both lobes are thought to be free to move around the protein hinge. Binding of ATP (or ATP-

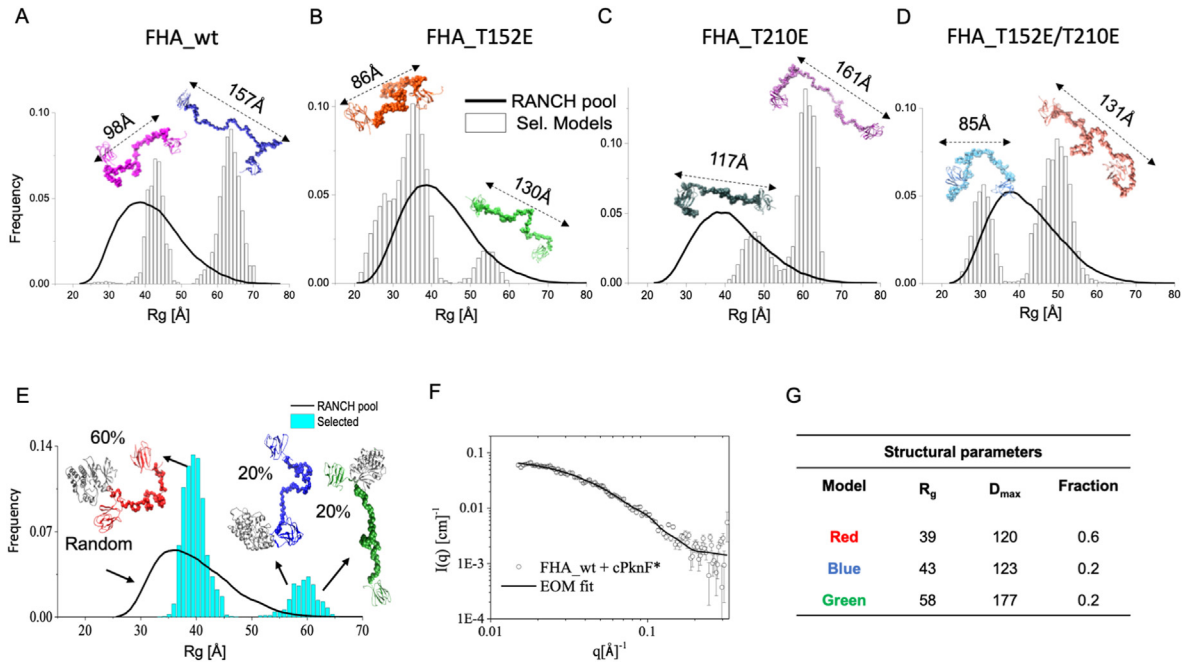


Fig. 3. FHA-1 and FHA-2 orientation assessment and cPknF*+FHA mixture by SAXS analysis. For the FHA analysis, we assumed an intrinsically flexibility of the domains due the long-disordered linker between them. Then, a construction of a pool of structures with randomized orientation between FHA-1 and FHA-2 using the RANCH program (internal EOM routine) was performed and followed by selection of sub-ensemble that better describe the experimental SAXS data. The radius of gyration (Rg) distribution for RANCH models build randomly to FHA-1 and FHA-2, connected by a long randomize linker of 98 residues, are showed as solid line. The selected EOM models from the RANCH pool to fit the SAXS profile using an ensemble of optimized fraction of models are represented as bar symbol for each analyzed sample. The representative models of these samples are shown in cartoon and the built loop as surface representation. (A) FHA_wt, (B) FHA_T152E, (C) FHA_T210E, (D) FHA_T152E/T210E. (E) EOM analysis for the sample FHA + cPknF*. Frequency of populations in samples of the mixture and EOM fit. The radius of gyration (Rg) distribution for RANCH models build to FHA + cPknF* is showed on black solid line. The selected EOM models from the RANCH pool to fit the SAXS profile using an ensemble optimized fraction of models are shown in the bar symbol for each analyzed sample. (F) For each sample, the representative models were obtained and presented. cPknF and FHA domains are in cartoon and the built loop as surface representation. (G) Structural parameters obtained for the most significant models.

competitive ligands) engages residues from both lobes and stabilizes the kinase domain in a "closed" conformation around the hinge. We thus used thermal-shift assay to identify a ligand that could stabilize the cPknF and facilitate its crystallization from a library of commercially available ATP-competitive kinase inhibitors. The unphosphorylated protein and ligands showing the largest thermal shifts were used in co-crystallization experiments. Co-crystals of cPknF and IKK-16, an inhibitor originally designed for the human IκK complex (composed of inhibitory-κB kinases α and β), produced an X-ray diffraction dataset that allowed us to determine the crystal structure of cPknF bound to IKK-16 at 2.75 Å resolution. The crystallographic phase problem was solved by molecular replacement method using the structural coordinates of the *Mtb* PknI (PDB 5M07) as search model (Data collection and crystallographic refinement statistics are shown in Table 1). Four cPknF monomers with similar structure were present in the crystal asymmetric unit (root mean square deviation, r.m.s.d. < 0.5 Å) (Fig. 1*A*). Overall, cPknF displayed the typical fold seen for other Serine/Threonine protein kinases that consisted of bilobed architecture, in which N- and C-terminal lobes were connected by a short hinge region forming a deep cleft where ATP can bind and where the inhibitor IKK-16 was located (Fig. 1*B*). Poor electron density coincided with regions of the kinase domain known to be quite plastic (Taylor and Kornev, 2011) and prevented us from confidently placing the following residues in the final crystallographic model (numbering for the *Mtb* cPknF protomer in chain A): 1–3, 21–23 (within the protein P-loop), 46–57 (connecting strand β4 to helix αC), 161–180 (part of the kinase activation loop) and 277–290. In the *Mtb* cPknF, the conserved GxGxxG motif or P-loop, known to be quite dynamic and to fold over the phosphates of ATP during catalysis, folded over the benzothiphene ring of the ligand. Likewise, in kinases, αC helix (showed in blue in cPknF) is quite dynamic and can swing towards ("in") or away ("out") from the N-lobe. The "αC-in" conformation facilitates the formation of a salt bridge between a conserved Glutamic acid in this helix (E60 in cPknF, blue stick) and a conserved Lysine in the β3 strand (K41, yellow stick) (Fig. 1*B*). Formation of this salt bridge is one of the hallmarks of the kinase active state and allows the Lysine to engage the oxygen atoms of the α and β phosphates of ATP via hydrogen bonds (Taylor and Kornev, 2011). In the crystal, cPknF αC adopted an outwards conformation, preventing the formation of this crucial salt bridge, and indicating an inactive conformation. At this point it is unclear if the non-phosphorylated state of cPknF or if IKK-16 (in pink) played a role in stabilizing cPknF in an inactive conformation, as other *Mtb* STPKs that co-crystallized in the presence of ATP-competitive ligands adopted active conformations (αC-in) (Wehenkel et al., 2006; Mieczkowski et al., 2008; Wang et al., 2017; Wlodarchak et al., 2018).

Finally, the kinase domain activation loop is a highly dynamic stretch of 20–30 residues between the conserved DFG and APE motifs (residues 159–161 and 186–188 in *Mtb* cPknF, respectively). In our crystal structure, most of the residues in the activation loop could not be built into the final crystallographic model, indicating the plasticity of this region. In eukaryotic protein kinases, phosphorylation of one or more sites within the activation loop reorganizes it into a binding platform for the substrate and is crucial for kinase activity. Nevertheless, even when phosphorylated, the activation loops of *Mtb* PknA, PknB and PknE remained largely disordered (Young et al., 2003; Ortiz-Lombardía et al., 2003; Wehenkel et al., 2006; Gay et al., 2006; Lombana et al., 2010). Thus, it is likely that a similarly disordered activation loop would be observed in crystals of phosphorylated cPknF. Recently, crystallographic studies showed that the formation of a docking platform for a peptide substrate in *Mtb* PknB required the interaction between the kinase phosphorylated activation loop and the FHA domain in *Mtb* GarA (Wagner et al., 2019). Further, the formation of this docking platform displaced the kinase αC towards the N-lobe, compared to the position of this helix in the unbound *Mtb* PknB

structure. Crucial to these rearrangements is the interaction of a phospho-Threonine (pT) residue from *Mtb* PknB activation loop and the phosphate-binding pocket within GarA FHA domain. This same pT residue engages an Arginine in the αC helix of *Mtb* PknB, contributing to the displacement of this helix towards the kinase N-lobe (Wagner et al., 2019). T173 and R59 of cPknF are structurally equivalent residues to the PknB activation loop pT and αC Arginine residues, respectively. It is thus likely that the binding of cPknF to FHA domain-containing proteins, such as *Mtb* Rv1747, would also reorganize the kinase phosphorylated activation loop into a docking site for a peptide substrate and swing cPknF αC into an active conformation.

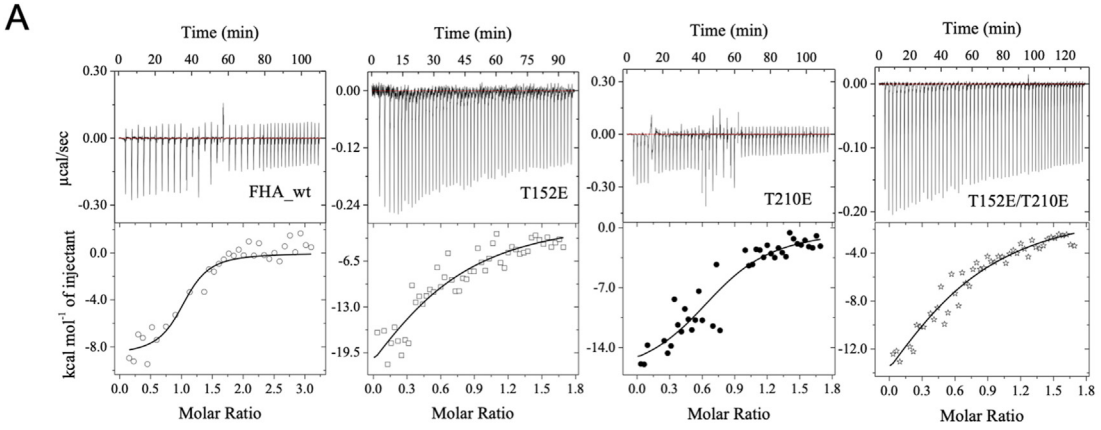
Dimer formation also plays a crucial role in regulating the activation of *Mtb* STPKs. The kinase domains of *Mtb* PknB and PknE crystallized as structurally similar ("back-to-back") dimers (Young et al., 2003; Ortiz-Lombardía et al., 2003; Wehenkel et al., 2006). Further biochemical studies revealed this organization to be important for the allosteric activation of the kinase activity in PknB and PknD (Greenstein et al., 2007; Lombana et al., 2010). Likewise, the four protomers within *Mtb* cPknF:IKK-16 co-crystals formed two "back-to-back" homodimers (Fig. 1*C*). The dimerization interface in cPknF occluded ~1070 Å^2 of solvent accessible surface area and, as seen for PknB and PknE, involved only residues within the kinase N-lobe domain. Nevertheless, dimer formation in cPknF was not sufficient to stabilize the kinase in an active conformation. These observations further suggest that, as for PknB, interaction with FHA domain-containing proteins is required for the activation of cPknF.

These results showed that cPknF, now one of the six *Mtb* protein kinases with available structure, was solved in the inactive state in the presence of an inhibitor. The enzyme showed the typical bilobed architecture observed in the orthologs as well as the conservation of functional motifs.

3.3. The Mtb cPknF ATP-binding site

The *Mtb* cPknF conserves 13 identical residues present in 36 unique kinase domains (*SI Appendix,* Fig. A4). Seven of them are related to the ATP catalysis, whereas the other six residues maintain the C-terminal lobe tertiary structure. To compare IKK-16 and ATP binding modes in cPknF, we built a molecular model of the kinase domain containing the unmodeled secondary structures and docked the ATP into the active site. The ATP-binding mode in docked cPknF:ATP is comparable to cPknB, with similar conformations of several well-conserved residues in this region of ATP-binding kinases (these include D127, K129, and Q132). The lysine residue (K41) in strand β3 forms an ion pair with a glutamic acid presented in the DFG-STPK signature (D159). This lysine also contacts the α-phosphate of ATP and the ribose O1 oxygen, suggesting that the lysine-glutamate pair may be involved in stabilizing or protonating the product ADP by acid-base function (Fig. 1*D*-I). At the ATP γ-phosphate tail, a similar complex repeats in contacting the K139 and T181 to that phosphate group, where we predict the cPknF phosphorylation starts. The salt bridge formed between ATP γ-phosphate tail and K139 helps the phosphate group transference to T181, which also performs interactions mediated by hydrogen bonds. Consequently, the Threonine position and proximity to the T178/T173/T175 at the P-loop might enable the phosphate group transference to the FHA domain at the Rv1747 transporter (Fig. 1*D*-I).

On the other hand, the inhibitor IKK-16 lies on the site entrance and hydrophobically interacts with residues V26, M89, V92, and A158. The benzothiophene planar ring extends deep within the cPknF catalytic site, connecting the N1 of the pyrimidine ring to the main chain residues from the nucleotide-binding region (V26, M89, F91, V92). Other polar contacts are formed between T95 and S99 with the piperidine and pyrrolidine rings, respectively. Conventional hydrogen bonds performed with T95, D96, and S99 help on the benzamide and piperidine ring



FHA	*N*	K_d (μM)	Δ*H* (kcal/mol)	Δ*S* (cal/mol/deg)
wild type	1.0±0.1	0.41±0.08	-8.7±0.6	-0.5
T210E	0.8±0.1	2.0±0.6	-17±1.9	-34
T152E	0.6±0.2	8.9±0.2	-42±19	-121
T152E/T210E	0.6±0.1	9.4±0.4	-27±6.6	-70

Fig. 4. Binding affinity of cPknF* to FHA wild type and mutants evaluated by isothermal titration calorimetry (ITC). (A) ITC profiles of cPknF* and FHA_wt, FHA_T152E, FHA_T210E and FHA_T152E/T210E. (B) Overall thermodynamics parameters from ITC fitted profile of FHA samples titrated with cPknF*. The assumed model was of N binding sites of cPknF* and FHA with the same equilibrium dissociation constant (K_d). (C) Representation of cPknF* affinity for FHA domains and correlation with radius of gyration (Rg).

stabilization. The ligand oddly forms an intramolecular interaction between the thiophene S1 and centroid benzamide that reduces the polar interactions by repelling D159 into the αC helix direction. As a motion result, the β4 sheet (in orange) from the N-terminal to the activation loop, the kinase domain, thus free to maintain a salt bridge between K41 and D159. In the absence of a nucleotide-binding connection, the active site pocket is occupied by the F91 side chain, as an active and inactive regulatory element (Fig. 1*D*-II).

In summary, the comparison of the binding mode of ATP and IKK-16 in cPknF shows that both ligands are positioned on the cleft that forms the active site, ATP in the centre, coordinated by conserved residues, and IKK-16 displaced in a region not so conserved, which can be explored for developing of specificity inhibitors.

3.4. Secondary structural changes induced in cPknF and Rv1747-FHA domains

The secondary structural content of the recombinant cPknF and cPknF* was evaluated by circular dichroism (CD) (*SI Appendix,* Fig. A5*A*-I) and revealed that both domains presented an α/β folding as indicated by two negative peaks of 208 nm and 222 nm, respectively, and no significant changes in the profiles. However, experiments of Differential Scanning Fluorimetry (DSF) showed that the phosphorylation induced changes in the thermal stability with reduction of the at least 3 °C in the T_m value of cPknF* (*SI Appendix,* Fig. A5*A*-II). The tryptophan fluorescence of cPknF* was quenched in comparison with the non-phosphorylated protein, without displacement of the maximum peak (*SI Appendix,* Fig. A5*A*-III). The cPknF structure has two Tryptophan, W185 close to ATP and W86 facing T8 and T13 in the N-lobe, which could be the most affected (*SI Appendix,* Fig. A5-*C*).

To study the effect of the phosphorylation on Rv1747 FHA domains we made four constructs: the non-phosphorylated wild type FHA domain (FHA_wt, residues 1 to 310), and three mutants mimicking different states of FHA phosphorylation (FHA_T152E, FHA_T210E and the double mutant FHA_T152E/T210E). The overexpression of all domains in fusion with GB was obtained from *E. coli* BL21 (DE3) cells in the soluble extracts of cultures upon induction with IPTG. The four FHA domains showed bands of 33.5 kDa in the SDS-PAGE, and they were obtained firstly by immobilized affinity nickel chromatography (IMAC) followed by size-exclusion chromatography and then, by cleavage with 3C protease for removing of GB1 fusion (*SI Appendix,* Fig. 6A to 6C). Biophysical assays were first performed to confirm their folding and stability. Circular dichroism profiles of the four domains indicated the expected characteristic β-sandwich fold with non-structured regions. No significant differences were seen among the four domains indicating that phosphorylation does not affect the secondary structural content. On the other hand, the insertion of point mutations in the FHA_wt domain had an impact on its thermal stability that gradually decreased with mutation of T152, T210 or both residues. Differences depending on the phosphorylation position were also identified during tryptophan intrinsic fluorescence assays. The maximum fluorescence peak at 337 nm identified in FHA_wt sample was shifted to 340 nm in the mutated domains indicating changing in the tryptophan environment. Also, it was observed a quenching of the fluorescence in the samples FHA_T210E, FHA_T152 and FHA_T152E/T210E when compared to the wild type. The maximum quenching evidenced in the FHA_T152 and FHA_T152E/T210E mutants might be due the proximity of W175 and W125 to T152 (*SI Appendix,* Fig. A.5*B*). The results of CD and fluorescence for the different states of cPknF and FHA suggest that the phosphorylation of these domains, although did not alter their secondary structural content, promote changes that change the stability, favoring their thermal destabilization.

Moreover, we tested if the cPknF* could phosphorylate the FHA_wt domain *in vitro*. Samples of cPknF* and FHA_wt were incubated together and submitted to mass spectrometry analysis. The results showed the calculated mass of 33.357 for FHA_wt non-phosphorylated and 33.515 and 33.594 Da for FHA_wt after incubation with cPknF*. The increase in the molecular weight of FHA_wt is interpreted as incorporation of phosphate groups in the FHA domain as a result of cPknF-mediated phosphorylation. Since the molecular weight of each phosphate group is approximately 80 Da, these results suggest the addition of two and three these groups in the domain (*SI Appendix,* Fig. A7).

3.5. Structural changes induced by recombinant cPknF* on FHA domains

Small Angle X-ray Scattering (SAXS) analysis were performed to evaluate the global arrangement of FHA-1 and FHA-2 domains in the wild type and mutants, the behavior of cPknF in comparison with cPknF*, and the interaction between cPknF* with FHA domains (Fig. 2). The pair distribution function *p*(r) results to FHA_wt and the mutant samples showed a characteristic profile of a two-domain protein with two peaks. A common peak at approximately 30 Å is present in the four samples but significant differences were observed in the *p*(r) profile of FHA_wt in comparison to FHA mutants (Fig. 2*A*) regarding the second peak. The mutant curves presented a displacement showing a correlation of pairs in a shorter distance, which indicated a more compact orientation between the domains after phosphorylation, clearly highlighted in the profile of the double mutant (red line). Still, the *p*(r) profiles of the mutants revealed a second shoulder in FHA_T152E (pink line) closer to the position of FHA_T152E/T210E at ~60 Å than FHA_T210E (blue line). In contrast, the maximum observed for the second shoulder of FHA_wt is shifted to ~90 Å. These results showed that mimetic phosphorylation in Threonine present in the FHA linker led to the compaction of the FHA-1 and FHA-2 domains, with a suggestive participation of T152 in the process. Based on GNOM analysis to obtain structural parameters, the higher values of radius of gyration (R_g) and maximum distance (D_{max}) corresponded to FHA_wt, which molecular weight (*MW*) suggested a monomeric state in solution to all samples (*SI Appendix,* Table A4). The lower values of R_g and D_{max} observed for the mutants corroborated the previous observations that the phosphorylation has an important function to approximate the domains (Spivey et al., 2011; Heinkel et al., 2018, 2019, 2019).

We also used SAXS analyses to compare cPknF and cPknF* (Fig. 2B, green and black lines, respectively). According to the *p*(r) profiles of the two samples, the phosphorylation induced conformational changes in cPknF structure, suggesting that phosphorylation is important to trigger the active state of the kinase. Indeed, structural parameters for cPknF and cPknF* samples indicate that phosphorylated domain tends to acquire a more compact conformation (*SI Appendix,* Table A4). Finally, we evaluated the behavior of the FHA_wt in the presence of cPknF and cPknF* in solution. The mixture of FHA_wt with cPknF or cPknF* shows distinct *p*(r) profiles, both with two peaks (Fig. 2*C*). The profile of the FHA_wt with the non-phosphorylated kinase shows a larger peak with a maximum of 22 Å and a smaller peak with a maximum of 66 Å. In the presence of cPknF* this profile of two peaks is conserved, however, there is greater definition of the peaks, with a displacement of approximately 15 Å in the maximum value of each one (37 Å and 83 Å, respectively for the largest and smallest). The profile of FHA_wt + PknF* seems a combination of the profiles of the double mutant FHA_T152E/T210E (Fig. 2A) and cPknF* (Fig. 2B). The GNOM analysis of the mixtures showed reduction in the values of D_{max} of FHA_wt in presence of cPknF* and an increasing of the *MW* suggesting a possible formation of a complex FHA/kinase and phosphorylation of FHA_wt (*SI Appendix,* Table A4). Altogether, the results showed that cPknF suffered changes in the structure after phosphorylation, which probably is important for its affinity and activity over the FHA domains. The SAXS data for wild type and phosphomimetic mutants in the presence of cPknF* showed that the enzyme only approaches the domains if they are partially or totally non-phosphorylated, and that upon phosphorylation, FHA domains approximate. cPknF* showed no affinity for the double mutant that mimics the phosphorylation of both Threonine residues. The lack of affinity between the kinase and the domains, also is evidenced when the cPknF is not phosphorylated, once the enzyme would be in the inactive state.

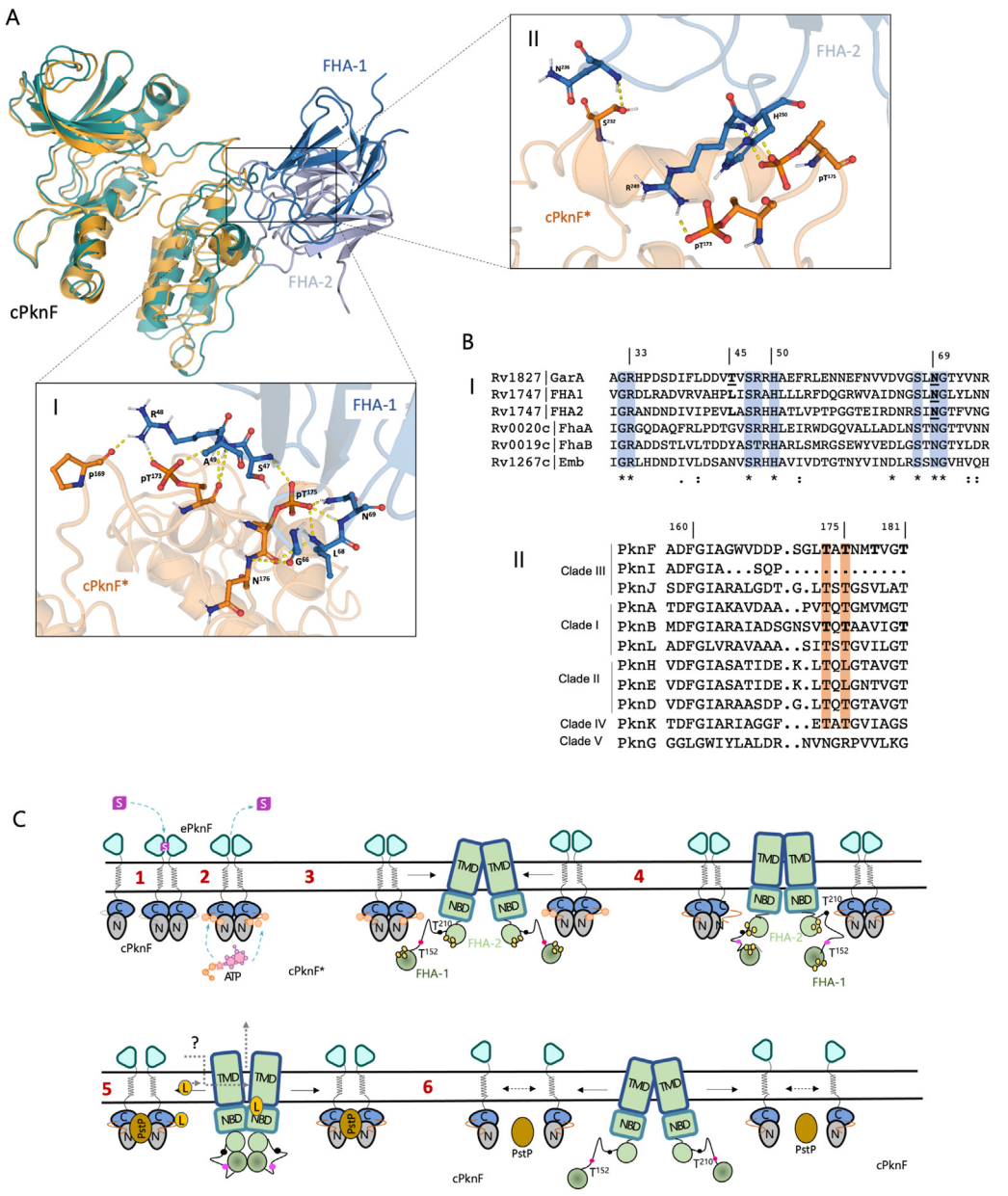


Fig. 5. Putative mechanism of interaction and modulation of Rv1747 FHA domains by *Mtb* cPknF. (*A*) Prediction of the interaction between Rv1747 FHA domains and cPknF. Docking simulations between phosphorylated cPknF and the individual domains FHA-1 and FHA-2 was performed based on the available structures of *Mtb* GarA and PknB (PDB 6I2P). The figure shows cPknF* (orange cartoon) superimposed to cPknB (green cartoon) in contact with FHA-1 (dark blue) and FHA-2 (clear blue). The details of the interaction with each FHA domain are shown in (I) and (II). The involvement of two phosphorylation sites in the activation loop could increase the phosphorylation potential and activation of the kinase. (*B*) Structural-based amino acid sequence alignment of all *Mtb* FHA domains and STPKs using as reference the interface region of GarA and PknB. (I) The *Mtb* FHA domains conserve most residues of the canonical recognition of pThr by FHA domains – $G^{32}R$, $S^{47}RxH$, and $S^{67}xNG$ (blue bars; reference number from Rv1747 FHA-1; x representing any residue). (II) The alignment of the *Mtb* STPKs activation loops reveals that cPknF conserves the two phosphorylated Threonine when compared with PknB and other kinases. (*C*) Schematic model of the structural changes induced by *Mtb* PknF in the ABC transporter Rv1747. In this model, PknF would assume a dimeric structure after recognition of a signal (S, magenta) by the PknF extracellular sensor domain (ePknF, in cyan) (step 1); in this state occurs the autophosphorylation (step 2) and the kinase approaches the Rv1747 transporter, which structure is expected to be in inward state with NBDs and FHA-1 and FHA-2 apart (step 3). The process of Rv1747 phosphorylation would involve the interaction between PknF activation loop and Rv1747_FHA. We speculate that FHA-2 (T210, black sphere in the linker) could be the first FHA domain to be phosphorylated due its proximity to the kinase. According to our data, when T210 is phosphorylated but not T152 (pink sphere), the cPknF affinity for FHA domains is increased and that might be a way to approximate the FHA-1 domain (step 4). The movement of FHA-1 and FHA-2 and the intra and intermolecular interactions lead to structural changes in the transporter, including NBDs approximation. It is also expected that the transporter acquires intermediate states (not all represented) that would increase the affinity for the ligand(s) (L, yellow sphere). Gray arrows indicate the possible pathways for ligand entry. In step 5, a set of features such as condensation of the FHA domains of a monomer with the domains of the second monomer, break of ATP, and structural changes in the permeases would finally lead to the activation of the

transporter, ligand interaction and its exclusion. PstP phosphatase (in brown) might have participation in the kinase-transporter complex rupture and dissociation of the PknF dimer, allowing the transporter to return to resting state (step 6).

3.6. cPknF in solution is a mixture of dimers and monomers

To evaluate the behavior of cPknF in solution, we first calculated the theoretical profile of SAXS for the enzyme using the full structural model of cPknF as input to the CRYSOL program (Svergun et al., 1995). The results were compared with the experimental data collected for the cPknF* and cPknF and showed that the model fit the SAXS profile, however, for small values of the scattering vector modulus, q, we have an indication of possible oligomeric arrangements of the protein in size larger than the monomer (Fig. 2*D* and *E*). In this case, to characterize the possible arrangement of cPknF in the sample we used the OLIGOMER program (Konarev et al., 2003), assuming a mixture of monomers and dimers and applying the high-resolution models of cPknF* (Fig. 2*E* and *F*). A nice agreement was obtained for the samples, $\chi^2_{\mathrm{PknF(n)^*}} = 5.6$ e $\chi^2_{\mathrm{PknF(n)^*}} = 1.7$, indeed indicating a mixture of dimers and monomers, respectively.

The volume fraction that best fit the data was composed of 64 % of dimers and 36 % of monomers for cPknF*, and 44 % of dimers and 56 % of monomers for cPknF (Fig. 2F). Although these results suggested predominance of dimers favored by phosphorylation, the analysis showed that the system is quite dynamic and that the fraction correspondent to the monomers is predominant (*SI Appendix*, Table A4).

3.7. FHA-1/FHA-2 orientation assessment in the wild type and phosphomimetic mutants

In order to investigate the effect that mutations caused on the orientation of FHA domains (FHA-1 and FHA-2) we used the ensemble optimization method (EOM) to SAXS data, assuming the presence of dynamics (EOM fit for each sample is shown in *SI Appendix*, Fig. A8). Initially, a set of 10.000 structures was created without assuming restrictions for the construction of disordered loops and treating arrangements with known structures as rigid bodies. The superposition of the R_g distribution of the FHA domains with random pool orientation and the selected ensemble by EOM were analyzed (Fig. 3). The FHA wild type sample showed oscillation between two main populations, which R_g were centered at 43 Å and 63 Å with $R_g = 56$ Å to the best ensemble of FHA_wt (Fig. 3*A*). So, in absence of phosphomimetic mutations, FHA-1 and FHA-2 domains showed higher flexibility. On the other hand, the analysis for individual mutants showed that FHA_T152E became majority more compact with a large fraction of conformers at $R_g = 35$ Å, and FHA_T210E presented a shift to larger R_g (59 Å) (Fig. 3*B* and *C*). These data indicated that the mimetic phosphorylation in T152 increased the attraction of FHA domains in comparison with T210E. Still, the results obtained for the double mutant revealed an intermediate behavior between the single mutants and the FHA_wt, represented by two more compact populations and an ensemble conformation with $R_g = 49$ Å (Fig. 3*D*). Altogether, these data indicated that the phosphorylation in T152 and T210 induced conformational changes and rapprochement of FHA-1 and FHA-2, but the T152 appears to play a significant role in the transportation regulation process.

Following the same methodology used for the previous analysis, we monitored the behavior of FHA_wt + cPknF* mixture in solution (Fig. 3*E*). The spatial coordinates of cPknF* and FHA-1 were maintained as a single domain and the orientation with FHA-2 was free to optimize, building the waste linker as a disordered native arrangement. In this analysis, three well-defined populations arrangements were determined with an optimal fit to the SAXS data, two of more compact (Rg ~ 40 Å) and one more open (Rg ~ 60 Å), suggesting a dynamic in the interaction between the kinase and FHA wild type domain. The EOM fit and structural parameters of this analysis are shown in Fig. 3*F* and *G*.

3.8. cPknF binding affinity for FHA_wt and phosphomimetic mutants

Isothermal Titration Calorimetry (ITC) experiments were carried out to investigate the affinity of cPknF* to FHA wild type and mutants (Fig. 4). The analysis of the isothermal profile as function of the molar ratio to each experiment showed different behavior, respectively to individually titration (Fig. 4*A*). To describe the ITC binding curves with the determined thermodynamic parameters, it was used a stoichiometric model assuming N as individual sites with same binding affinity. The interaction between FHA_wt and cPknF* was described by a stoichiometric ratio of 1:1 ($N = 1$) with an equilibrium dissociation constant (K_d) of approximately 0.4 μM, indicating a high affinity of the kinase for its substrate when it is not phosphorylated. Indeed, we observed that the smallest affinity was calculated for the interaction between cPknF* and the double phosphomimetic mutant. The results of the interaction of cPknF* with the different phosphomimetic mutants of FHA showed a smaller order of magnitude in binding affinity and a likely lower efficiency of interaction, indicated by $N < 1$ (Fig. 4*B*). cPknF* showed higher affinity for FHA_T210E than FHA_T152E, indicating that the non-phosphorylated T152 is more attractive to cPknF than T210, corroborating our previous results. cPknF affinity is inverse to the compact state of the domains evaluated by the *Rg*, since the closer they are (lower *Rg*), the less the tendency of the enzyme to associate. A parallel between the affinity and the *Rg* is showed in Fig. 4*C* showing that after the double phosphorylation event the kinase's affinity is decreased.

3.9. The merging of the FHA and cPknF SAXS data shows good agreement with the SAXS data collected from the sample of the FHA and cPknF mixture

ITC data did not show a detected sign of interaction between FHA wild type and non-phosphorylated cPknF, indicating that there was no complex formation in this case. With this information, we assumed that the intensity collected for the mixture FHA wild type and cPknF [$I_{mix}(q)$] could be described by the linear combination between the intensity profile of the FHA-wt sample and the cPknF (*Equation (1)*):

$$I_{mix}(q) = A\, I_{FHA}(q) + BI_{PknF(n)^*}(q) + C \quad (1)$$

where A, B and C are optimized constants when combining spread profiles.

The result obtained from this analysis showed an excellent agreement between the scattered intensity collected for the mixture of FHA wild type and cPknF, with the data of the junction of the individual experimental scattered intensity of each protein (*SI Appendix*, Fig. A9). This result corroborated that the interaction between these two proteins was possibly not observed in the SAXS experiments, as well as in the calorimetry data, highlighting the importance of phosphorylation of cPknF for the beginning of the transport regulation process.

3.10. A model for PknF and Rv1747 interaction

To better understand how cPknF could interact with Rv1747 FHA domains, we performed docking simulations between the three-dimensional structures of Rv1747 FHA-1 (PDB 6CCD), FHA-2 (PDB 6CAH), and cPknF (Fig. 5) through the HADDOCK2.4 webserver (de Vries et al., 2010). The simulations generated 400 final structures for each complex, which were further clustered and ranked based on the HADDOCK score. The top 2 clusters demonstrated lower electrostatic and desolvation energy terms for cPknF/FHA-2 than the cPknF/FHA-1 complex (*SI Appendix,* Table A4). However, these values by themselves cannot be directly correlated to the binding free energy of the systems. When only the residues at the contact interface of both complexes were considered, we observed RMSD values not much higher than 3 Å, demonstrating a low dispersity of docked solutions (*SI Appendix,* Fig. A10).

To characterize the binding interface, the four representative structures of the top 2 clusters were refined and the solutions with lower HADDOCK score submitted to PISA (*SI Appendix,* Table A5). The analysis revealed an interface area of 497.4 $Å^2$ and 655.5 $Å^2$ for cPknF/FHA-1 and cPknF/FHA-2, respectively, demonstrating similar results to the interface of the homologous cPknB/GarA structure (PDB 6I2P) that was 539.7 $Å^2$. The interaction between cPknF and FHA-1 counts with 11 hydrogen bonds while in FHA-2 binding this bond type falls off to only four (Fig. 5*A* and *B*). These interactions mainly occurred with the phosphorylated Threonine residues in the activation loop of the kinase (pT173 and pT175) and some other present in the loops between β-strands of FHA domains. In the same way as in the PknB/GarA complex, PnkF/FHA-1 demonstrated a canonical recognition of pThr residues through the conserved motifs S^{47}RxH and S^{67}xNG, but no interaction was found with the G^{32}R sites. Additionally, the cPknF/FHA-1 complex was also stabilized by the hydrogen bond between the main-chain nitrogen atom of N176 and the oxygen of the carboxyl group of G66. In a reduced number, cPknF-FHA2 also presented contact predictions between both pT and the S^{248}RxH, but the other motifs characterized in the canonical activation of kinase by FHA domains did not participate in the complex formation. Interestingly, besides the activation loop, the FHA-2 domain also interacts, through N236, with the S232 residue located at the C-terminal of the α-G helix of cPknF. In opposite to PknB-GarA, neither FHA-1 nor FHA-2 presented salt bridges stabilizing the interaction with cPknF. The involvement of two phosphorylation sites in the activation loop could increase the phosphorylation potential and activation of the kinase. In this way, the phosphorylated kinase would assume an active dimeric state that increases the affinity for FHA domains, specifically domain FHA-1 and T152 of the linker region, as suggested in the analyses of the SAXS and ITC data. Therefore, we hypothesize that the flexibility that the linker confers to FHA-1 and FHA-2 is important for fitting of both domains in the phosphorylated kinase.

Based on previous results and data from this work, we propose a model of the modulation of Rv1747 transporter activity by *Mtb* PknF kinase (Fig. 5*D*). In this model, an external signal sensed by the extracellular domain PknF (ePknF) would trigger conformational changes, autophosphorylation, and dimerization of PknF increasing its affinity for Rv1747 transporter. According to docking results and the higher relevance that mutation on T152 revealed in comparison with T210, we might expect that cPknF* would be attracted to the "relaxed" non-phosphorylated FHA-1 domain by means of T152 phosphorylation, which would trigger significant conformational changes that would allow the approximation of FHA-1 to FHA-2 and consequent phosphorylation of T210. The flexibility of the linker and the higher affinity of cPknF to T152 in this situation would help the rapprochement movement of FHA-1. Because of both phosphorylation, FHA domains assume a compact structure closer to the NBDs cascading conformational changes and activation of the transporter (Fig. 5*C*). Although the specific substrate is not known, this model would satisfy the Rv1747 interaction with membrane components, lipids or drugs, as previously suggested. cPknF affinity for FHA-1/FHA-2 domains reduces after double phosphorylation and it is expected that the enzyme to leave the complex. The role of PstP phosphatase, although not studied in this work, could be related to the complex dissolution and PknF monomerization.

4. Discussion

PknF is one of the 11 STPKs from *Mtb* and consists of two domains connected by a TM helix: the extracellular sensor domain (ePknF) and the cytoplasmic kinase domain (cPknF). PknF has been related to phosphorylation of FHA domain of the universal stress protein Rv2623 and mainly the FHA domains of Rv1747 (FHA-1 and FHA-2), an ABC transporter, which gene is located downstream of *pknF*. The conservation of the operon *rv1746/rv1747* in almost all species evaluated, more significantly evidenced in the *Mtb* complex group, suggests that the set "transporter and kinase" participate in an important physiological process such as maintenance of the bacillus viability, which is in accordance with the putative role of Rv1747 in the membrane recycling and glycolipids transport (Deol et al., 2005; Curry et al., 2005; Gopalaswamy et al., 2008; Spivey et al., 2013).

Studies using super-resolution single-molecule localization microscopy showed that increasing amounts of cPknF* and consequent phosphorylation of T152 and T210, but not only (other Threonine present in the ID linker also showed phosphorylation), led to the oligomerization and phase-separation in models of membranes *in vitro* and in live cells with impact on the activity of the transporter (Heinkel et al., 2018, 2019). In this work, we explored the structural features of cPknF and their interaction with the full regulatory module in different mimetic phosphorylation states, including the non-phosphorylated wild type, FHA-1/linker/FHA-2. The recombinant cPknF*, when in presence of the three mutants, showed preferences indicating a stepwise mechanism for the compaction of the Rv1747 regulatory module. The enzyme affinity suffered direct influence of the position of phosphorylation, state of phosphorylation, flexibility of the linker and the conformation of the domains (higher affinity, higher *Rg*). The lower values of R_g and D_{max} observed for the mutants in comparison with the wild type FHA corroborated the previous observation that the phosphorylation has an important function to approximate the domains with further consequences to the transporter and regulation (Spivey et al., 2011; Heinkel et al., 2018, 2019, 2019). On the other hand, the results obtained for the double mutant showed an intermediate behavior between the single mutants and the FHA_wt, represented by two more compact populations and an ensemble conformation with $R_g = 49$ Å. FHA-2 is less stable and has higher affinity for p-Thr peptides than FHA-1 (Heinkel et al., 2018), which might explain the behavior of FHA_T152E mutant. Studies have shown that only the FHA-1 domain is essential for transporter function, since mutations in this domain involved in the interaction with cPknF, avoid the entire phosphorylation of Rv1747 by cPknF*. The mutant strain carrying the plasmid with these mutations is attenuated in *vivo* models, in contrast to the mutations involved in FHA-2 (Curry et al., 2005; Spivey et al., 2013). One explanation for this may be the proximity of FHA-1 domain to pT152 that, as we have shown, has a greater effect in tertiary structural changes than phosphorylation of T210. SAXS analysis revealed that in solution, the formation of the complex is dynamic and is described by different populations. There are few examples in the literature of the effects of protein conformation after phosphorylation in bacteria. Some of the proteins with FHA domains are OdhI and GarA, in which the phosphorylation of a Threonine residue at the amino-terminal end mediates their self-inhibition (Nott et al., 2009). In contrast, phosphorylation of the FHA domain of the Rv0020c (Fha-A) by PknB did not induce structural changes linked to a self-inhibition mechanism, as seen in GarA (Roumestand et al., 2011). A study conducted with β-arrestin-1 and β-arrestin-2 evaluated the effects of phosphorylation on these domains through mutations of Serine and Threonine residues by aspartic acid, followed by the Hydrogen-Deuterium exchange-mass spectrometry

(HDX-MS) technique (Kang et al., 2020). The results showed the phosphomimetic mutations of T276D and S14D induced changes in the HDX-MS profile when compared to the wild type protein, suggesting that the loop harboring the mutation became less dynamic or less exposed to the buffer. Additional results showed a synergistic effect on conformational changes compared to single phosphorylation states with simultaneous phosphorylation of S14 and T276 (Kang et al., 2020).

Altogether, the results obtained in this work showed how the Rv1747 FHA domains behavior after phosphorylation by cPknF* and allowed the design of a model for the transport regulation mechanism. However, many structural and functional questions are still unclear. Heinkel and collaborators (2018) have proposed that the phosphorylation of FHA domains and intra molecular interactions would lead Rv1747 clustering as a mechanism to increase the transporter activity. It is possible that the oligomerization of Rv1747 after phosphorylation could help the formation of membrane structural organization of active macromolecular complexes such as lipid rafts (Sviridov et al., 2020). This structure would facilitate the transport of components of the *Mtb* membrane, which should adapt to the new macrophage environment. In addition, questions regarding the role of substrate and NBDs during this process, or how they act in the clustering process or even what the stoichiometry of the complex is, still need to be evaluated. Similarly, what is the role of PstP during the process and what are the stimuli for its activation are still unknown.

Finally, the three-dimensional structure of cPknF is conserved when compared to the available *Mtb* STPKs structures of PknA, PknB, PknE, PknG and PknI, but interestingly, the pocket where the inhibitor is located in cPknF, has distinct residues that can be explored for the development of PknF specific inhibitors, aiming the control of bacillus infection.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Research grant 2018/20162–9 (A.B.), Conselho Nacional de Desenvolvimento Científico (CNPq) Universal grant 401505/2016–2 (A.B.), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Administrative Department of Science, Technology and Innovation COLCIENCIAS (PhD fellowships for S.P.C.B. and L.I.D.T.) and Universidad de La Frontera (Resolución exenta 2960, for B.E.).

This article contains an Appendix with Supplementary Data online.

Data availability

Atomic coordinates for the cPknF model have been deposited in the Protein Data Bank (PDB) under accession number 7NAA.

CRediT authorship contribution statement

Sindy Cabarca: Produced all the proteins and domains, performed all the experiments and conducted, Formal analysis, Investigation, Writing – original draft. **Maximilia Frazão de Souza:** Conducted SAXS experiments, Data curation, Formal analysis, Writing – original draft. **Andrew Albert de Oliveira:** resolved the PknF 3D structure, Writing – original draft. **Gabriel S. Vignoli Muniz:** Conducted Fluorescence experiments, Data curation, Formal analysis, Writing – original draft. **M. Teresa Lamy:** Conducted Fluorescence experiments, Data curation, Formal analysis, Writing – original draft. **Caio Vinicius dos Reis:** conducted protein crystallization. **Jessica Takarada:** performed mass spectroscopy, Formal analysis. **Brian Effer:** performed cloning and expression. **Lucas Santos Souza:** conducted the molecular docking and simulations. **Lilia Iriarte de la Torre:** conducted, Investigation, and protein production. **Rafael Couñago:** conducted the crystallization trials, Formal analysis, Writing – review & editing. **Cristiano Luis Pinto Oliveira:** Conducted SAXS experiments, Data curation, Formal analysis, Resources, Writing – review & editing. **Andrea Balan:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

We thank Professor Luís Carlos de Souza Ferreira from Vaccine Development Laboratory at Department of Microbiology (University of São Paulo) for the facility assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2021.07.001.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221. https://doi.org/10.1107/S0907444909052925.

Alber, T., 2009. Signaling mechanisms of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinases. Curr. Opin. Struct. Biol. 19, 650–657. https://doi.org/10.1016/j.sbi.2009.10.017.

Albrecht, C., 2008. Joseph R. Lakowicz: Principles of Fluorescence Spectroscopy, 3rd Edition, Analytical and Bioanalytical Chemistry. https://doi.org/10.1007/s00216-007-1822-x.

Av-Gay, Y., Everett, M., 2000. The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. Trends Microbiol. 8, 238–244. https://doi.org/10.1016/S0966-842X(00)01734-0.

Baer, C.E., Iavarone, A.T., Alber, T., Sassetti, C.M., 2014. Biochemical and spatial coincidence in the provisional Ser/Thr protein kinase interaction network of *Mycobacterium tuberculosis*. J. Biol. Chem. 289, 20422–20433. https://doi.org/10.1074/jbc.M114.559054.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. 98, 10037–10041. https://doi.org/10.1073/pnas.181342398.

Barthe, P., Roumestand, C., Canova, M.J., Kremer, L., Hurard, C., Molle, V., Cohen-Gonsaud, M., 2009. Dynamic and structural characterization of a bacterial FHA protein reveals a new autoinhibition mechanism. Structure 17, 568–578. https://doi.org/10.1016/j.str.2009.02.012.

Berendsen, H.J.C., van der Spoel, D., van Drunen, R., 1993. Gromacs: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 91, 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., Bricogne, G., 2004. Refinement of severely incomplete structures with maximum likelihood in BUSTER–TNT. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2210–2221. https://doi.org/10.1107/S0907444904016427.

Braibant, M., Gilot, P., Content, J., 2000. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. FEMS Microbiol. Rev. 24, 449–467. https://doi.org/10.1016/S0168-6445(00)00034-6.

Bruno, I.J., Cole, J.C., Kessler, M., Luo, J., Motherwell, W.D.S., Purkis, L.H., Smith, B.R., Taylor, R., Cooper, R.I., Harris, S.E., Orpen, A.G., 2004. Retrieval of crystallographically-derived molecular geometry information. J. Chem. Inf. Comput. Sci. 44, 2133–2144. https://doi.org/10.1021/ci049780b.

Canova, M.J., Molle, V., 2014. Bacterial Serine/Threonine protein kinases in host-pathogen interactions. J. Biol. Chem. 289, 9473–9479. https://doi.org/10.1074/jbc.R113.529917.

Chen, J., Kieswich, J.E., Chiazza, F., Moyes, A.J., Gobbetti, T., Purvis, G.S.D., Salvatori, D.C.F., Patel, N.S.A., Perretti, M., Hobbs, A.J., Collino, M., Yaqoob, M.M., Thiemermann, C., 2017. IkB kinase inhibitor attenuates sepsis-induced cardiac dysfunction in CKD. J. Am. Soc. Nephrol. 28 (1), 94–105. https://doi.org/10.1681/ASN.2015060670.

Cottrell, S.J., Olsson, T.S.G., Taylor, R., Cole, J.C., Liebeschuetz, J.W., 2012. Validating and understanding ring conformations using small molecule crystallographic data. J. Chem. Inf. Model. 52, 956–962. https://doi.org/10.1021/ci200439d.

Curry, J.M., Whalan, R., Hunt, D.M., Gohil, K., Strom, M., Rickman, L., Colston, M.J., Smerdon, S.J., Buxton, R.S., 2005. An ABC transporter containing a forkhead-associated domain interacts with a Serine-Threonine protein kinase and is required for growth of *Mycobacterium tuberculosis* in mice. Infect. Immun. 73, 4471–4477. https://doi.org/10.1128/IAI.73.8.4471-4477.2005.

de Vries, S.J., van Dijk, M., Bonvin, A.M.J.J., 2010. The HADDOCK web server for data-driven biomolecular docking. Nat. Protoc. 5, 883–897. https://doi.org/10.1038/nprot.2010.32.

Deol, P., Vohra, R., Saini, A.K., Singh, A., Chandra, H., et al., 2005. Role of *Mycobacterium tuberculosis* ser/thr kinase PknF: implications in glucose transport and cell division. J. Bacteriol. 20, 3415–3420. https://doi.org/10.1128/JB.187.10.3415-3420.2005.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0907444910007493.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. In: Current Protocols in Bioinformatics. NIH Public Access. https://doi.org/10.1002/0471250953.bi0506s15. Unit-5.6.

Gay, L.M., Ng, H.L., Alber, T., 2006. A conserved dimer and global conformational changes in the structure of apo-PknE Ser/Thr protein kinase from *Mycobacterium tuberculosis*. J. Mol. Biol. 360, 409–420. https://doi.org/10.1016/j.jmb.2006.05.015.

Gopalaswamy, R., Narayanan, S., Jacobs, W.R., Av-Gay, Y., 2008. *Mycobacterium smegmatis* biofilm formation and sliding motility are affected by the Serine/Threonine protein kinase PknF. FEMS Microbiol. Lett. 278, 121–127. https://doi.org/10.1111/j.1574-6968.2007.00989.x.

Greenstein, A.E., Echols, N., Lombana, T.N., King, D.S., Alber, T., 2007. Allosteric activation by dimerization of the PknD receptor Ser/Thr protein kinase from *Mycobacterium tuberculosis*. J. Biol. Chem. 282, 11427–11435. https://doi.org/10.1074/jbc.M610193200.

Grundner, C., Gay, L.M., Alber, T., 2005. *Mycobacterium tuberculosis* serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains. Protein Sci. 14, 1918–1921. https://doi.org/10.1110/ps.051413405.

Heinkel, F., Shen, L., Richard-Greenblatt, M., Okon, M., Bui, J.M., Gee, C.L., Gay, L.M., Alber, T., Av-Gay, Y., Gsponer, J., McIntosh, L.P., 2018. Biophysical characterization of the tandem FHA domain regulatory module from the *Mycobacterium tuberculosis* ABC Transporter Rv1747. Structure 26, 972–986. https://doi.org/10.1016/j.str.2018.04.018 e6.

Heinkel, F., Abraham, L., Ko, M., Chao, J., Bach, H., Hui, L.T., Li, H., Zhu, M., Ling, Y.M., Rogalski, J.C., Scurll, J., Bui, J.M., Mayor, T., Gold, M.R., Chou, K.C., Av-Gay, Y., McIntosh, L.P., Gsponer, J., 2019. Phase separation and clustering of an ABC transporter in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. 116, 16326–16331. https://doi.org/10.1073/pnas.1820683116.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I., 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36, 1277–1282. https://doi.org/10.1107/S0021889803012779.

Krissinel, E., 2010. Crystal contacts as nature's docking solutions. J. Comput. Chem. 31, 133–143. https://doi.org/10.1002/jcc.21303.

Kroeger, T., Frieg, B., Zhang, T., Hansen, F.K., Marmann, A., Proksch, P., Nagel-steger, L., Groth, G., Smits, S.H.J., Gohlke, H., 2017. EDTA Aggregates Induce SYPRO Orange-Based Fluorescence in Thermal Shift Assay 1–21.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., et al., 2007. Clustal W and clustal X version 2.0. Bioinformatics 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404.

Lombana, T.N., Echols, N., Good, M.C., Thomsen, N.D., Ng, H.-L., Greenstein, A.E., Falick, A.M., King, D.S., Alber, T., 2010. Allosteric activation mechanism of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase. PknB. Structure 18, 1667–1677. https://doi.org/10.1016/j.str.2010.09.019.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674. https://doi.org/10.1107/S0021889807021206.

Mieczkowski, C., Iavarone, A.T., Alber, T., 2008. Auto-activation mechanism of the *Mycobacterium tuberculosis* PknB receptor Ser/Thr kinase. EMBO J. 27, 3186–3197. https://doi.org/10.1038/emboj.2008.236.

Narayan, A., Sachdeva, P., Sharma, K., Saini, A., 2007. Serine threonine protein kinases of mycobacterial genus: phylogeny to function. Physiol. Genom. 29, 66–75. https://doi.org/10.1152/physiolgenomics.00221.2006.

Nott, T.J., Kelly, G., Stach, L., Li, J., Westcott, S., Patel, D., Hunt, D.M., Howell, S., Buxton, R.S., O'Hare, H.M., Smerdon, S.J., 2009. An intramolecular switch regulates phosphoindependent FHA domain interactions in. *Mycobacterium tuberculosis*. Sci. Signal. 2. https://doi.org/10.1126/scisignal.2000212.

Oliveira, M.C.B., Balan, A., 2020. The ATP-binding cassette (ABC) transport systems in *Mycobacterium tuberculosis*: structure, function, and possible targets for therapeutics. Biology 9, 443. https://doi.org/10.3390/biology9120443.

Ortiz-Lombardía, M., Pompeo, F., Boitel, B., Alzari, P.M., 2003. Crystal structure of the catalytic domain of the PknB Serine/Threonine kinase from. *Mycobacterium tuberculosis*. J. Biol. Chem. 278, 13094–13100. https://doi.org/10.1074/jbc.M300660200.

Pallen, M., Chaudhuri, R., Khan, A., 2002. Bacterial FHA domains: neglected players in the phospho-Threonine signalling game? Trends Microbiol. 10, 556–563. https://doi.org/10.1016/S0966-842X(02)02476-9.

Panni, S., 2019. Phospho-peptide binding domains in *S. cerevisiae* model organism. Biochimie 163, 117–127. https://doi.org/10.1016/j.biochi.2019.06.005.

Prisic, S., Husson, R.N., 2014. *Mycobacterium tuberculosis* Serine/Threonine protein kinases. Microbiol. Spectr. 2. https://doi.org/10.1128/microbiolspec.mgm2-0006-2013.

Richard-Greenblatt, M., Av-Gay, Y., 2017. Epigenetic phosphorylation control of *Mycobacterium tuberculosis* infection and persistence. Microbiol. Spectr. 5. https://doi.org/10.1128/microbiolspec.tbtb2-0005-2015.

Roumestand, C., Leiba, J., Galophe, N., Margeat, E., Padilla, A., Bessin, Y., Barthe, P., Molle, V., Cohen-Gonsaud, M., 2011. Structural insight into the *Mycobacterium tuberculosis* Rv0020c protein and its interaction with the PknB kinase. Structure 19, 1525–1534. https://doi.org/10.1016/j.str.2011.07.011.

Semenyuk, A.V., Svergun, D.I., 1991. GNOM. A program package for small-angle scattering data processing. J. Appl. Crystallogr. 24, 537–540. https://doi.org/10.1107/S002188989100081X.

Spivey, V.L., Molle, V., Whalan, R.H., Rodgers, A., Leiba, J., Stach, L., Walker, K.B., Smerdon, S.J., Buxton, R.S., 2011. Forkhead-associated (FHA) domain containing ABC transporter Rv1747 is positively regulated by Ser/Thr phosphorylation in *Mycobacterium tuberculosis*. J. Biol. Chem. 286, 26198–26209. https://doi.org/10.1074/jbc.M111.246132.

Spivey, V.L., Whalan, R.H., Hirst, E.M.A., Smerdon, S.J., Buxton, R.S., 2013. An attenuated mutant of the Rv1747 ATP-binding cassette transporter of *Mycobacterium tuberculosis* and a mutant of its cognate kinase, PknF, show increased expression of the efflux pump-related iniBAC operon. FEMS Microbiol. Lett. 347, 107–115. https://doi.org/10.1111/1574-6968.12230.

Svergun, D., Barberato, C., Koch, M.H., 1995. Crysol - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. 28, 768–773. https://doi.org/10.1107/S0021889895007047.

Sviridov, D., Mukhamedova, N., Miller, Y.I., 2020. Lipid rafts as a therapeutic target. J. Lipid Res. 61, 687–695. https://doi.org/10.1194/jlr.TR120000658.

Taylor, S.S., Kornev, A.P., 2011. Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem. Sci. 36, 65–77. https://doi.org/10.1016/j.tibs.2010.09.006.Protein.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., Adams, P.D., 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr. Sect. D Biol. Crystallogr. 64, 61–69. https://doi.org/10.1107/S090744490705024X.

Tria, G., Mertens, H.D.T., Kachala, M., Svergun, D.I., 2015. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ 2, 207–217. https://doi.org/10.1107/S205225251500202X.

Veyron-Churlet, R., Molle, V., Taylor, R.C., Brown, A.K., Besra, G.S., Zanella-Cléon, I., Fütterer, K., Kremer, L., 2009. The *Mycobacterium tuberculosis* β-ketoacyl-acyl carrier protein synthase III activity is inhibited by phosphorylation on a single Threonine residue. J. Biol. Chem. 284, 6414–6424. https://doi.org/10.1074/jbc.M806537200.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., Bricogne, G., 2011. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 293–302. https://doi.org/10.1107/S0907444911007773.

Waelchli, R., Bollbuck, B., Bruns, C., Buhl, T., Eder, J., Feifel, R., Hersperger, R., Janser, P., Revesz, L., Zerwes, H.G., Schlapbach, A., 2006. Design and preparation of 2-benzamido-pyrimidines as inhibitors of IKK. Bioorg. Med. Chem. Lett 16, 108–112. https://doi.org/10.1016/j.bmcl.2005.09.035.

Wagner, T., André-Leroux, G., Hindie, V., Barilone, N., Lisa, M.-N., Hoos, S., Raynal, B., Vulliez-Le Normand, B., O'Hare, H.M., Bellinzoni, M., Alzari, P.M., 2019. Structural insights into the functional versatility of an FHA domain protein in mycobacterial signaling. Sci. Signal. 12. https://doi.org/10.1126/scisignal.aav9504.

Wang, T., Bemis, G., Hanzelka, B., Zuccola, H., Wynn, M., Moody, C.S., Green, J., Locher, C., Liu, A., Gao, H., Xu, Y., Wang, S., Wang, J., Bennani, Y.L., Thomson, J.A., Müh, U., 2017. Mtb PknA/PknB dual inhibition provides selectivity advantages for inhibitor design to minimize host kinase interactions. ACS Med. Chem. Lett. 8, 1224–1229. https://doi.org/10.1021/acsmedchemlett.7b00239.

Wehenkel, A., Fernandez, P., Bellinzoni, M., Catherinot, V., Barilone, N., Labesse, G., Jackson, M., Alzari, P.M., 2006. The structure of PknB in complex with mitoxantrone, an ATP-competitive inhibitor, suggests a mode of protein kinase regulation in mycobacteria. FEBS Lett. 580, 3018–3022. https://doi.org/10.1016/j.febslet.2006.04.046.

Wiedemann, C., Bellstedt, P., Görlach, M., 2013. Capito - a web server-based analysis and plotting tool for circular dichroism data. Bioinformatics 29, 1750–1757. https://doi.org/10.1093/bioinformatics/btt278.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315. https://doi.org/10.1002/pro.3330.

Wlodarchak, N., Teachout, N., Beczkiewicz, J., Procknow, R., Schaenzer, A.J., Satyshur, K., Pavelka, M., Zuercher, W., Drewry, D., Sauer, J.D., Striker, R., 2018. In silico screen and structural analysis identifies bacterial kinase inhibitors which act with β-lactams to inhibit mycobacterial growth. Mol. Pharm. 15, 5410–5426. https://doi.org/10.1021/acs.molpharmaceut.8b00905.

Young, T.A., Delagoutte, B., Endrizzi, J.A., Falick, A.M., Alber, T., 2003. Structure of *Mycobacterium tuberculosis* PknB supports a universal activation mechanism for Ser/Thr protein kinases. Nat. Struct. Biol. 10, 168–174. https://doi.org/10.1038/nsb897.

Appendix A

Supplementary Information for

Structure of the *Mycobacterium tuberculosis* *c*PknF and conformational changes induced in Forkhead-Associated regulatory domains

Sindy Paola Cabarca Barreto[a,b,c], Maximilia Frazão de Souza[d], Andrew Albert de Oliveira[e], Gabriel S. Vignoli Muniz[f], M. Teresa Lamy[f], Caio Vinicius dos Reis[g], Jessica Takarada[g], Rafael Couñago[g], Brian Effer[h], Lucas Santos Souza[b], Lilia Iriarte de la Torre[a,b,c] Cristiano Luis Pinto Oliveira[d] and Andrea Balan[b*]

*Corresponding author: Andrea Balan
Email: abalan@usp.br

This PDF file includes:

Tables A.1 to A.5
Figures A.1 to A.10

Supplementary Tables

Table A.1 *Mycobacterium* spp. species cited in this work.

***Mycobacterium* species**
Mycobacteroides chelonae CCUG 47445
Mycobacteroides abscessus ATCC 19977
Mycobacteroides immunogenum FLAC016
Mycolicibacterium fortuitum CT6
Mycolicibacterium smegmatis MC2 155
Mycolicibacterium wolinskyi ATCC 700010
Mycolicibacter terrae NCTC10856
Mycolicibacter nonchromogenicus NCK 8460
Mycobacterium triplex DSM 44626
Mycobacterium florentinum JCM 14740
Mycobacterium xenopi JCM 15661T
Mycobacterium shimoidei P7336
Mycobacterium fragae
Mycobacterium celatum NCTC 12882
Mycobacterium marinum E11
Mycobacterium ulcerans SGL03
Mycobacterium canettii CIPT 140010059
Mycobcaterium tuberculosis var. *bovis* Danish 1331
Mycobcaterium tuberculosis var.*caprae* Allgae
Mycobacterium tuberculosis var. *microti* 12
Mycobacterium tuberculosis var. *africanum* GM041182
Mycobacterium gastri DSM 43505
Mycobacterium kansasii ATCC 12478
Mycobacterium haemophilum DSM 44634
Mycobacterium avium subsp. *paratuberculosis* DSM44135
Mycobacterium colombiense CECT 3035
Mycobacterium marseillense FLAC0026
Mycobacterium intracellulare ATCC 13950
Mycobacterium chimaera AH16
Mycobacterium leprae TN

Table A.2 List of up- and downstream genes that are located close to the *Mtb* genes encoding serine/threonine kinase proteins.

Genes up and downstream of the kinase genes	Protein Function
pknG	Serine/threonine protein kinase
glnH (rv0411c)	Glutamine-binding protein
rv0412c	Membrane protein
rv3079c	Hypothetical protein
pknH	Serine/threonine protein kinase
pknB (rv0014c)	Serine/threonine protein kinase
pknA (rv0015c)	Serine/threonine protein kinase
pbpA (rv0016c)	Penicillin-binding protein
rodA (rv0017c)	Cell division protein
rv2175c	DNA-binding protein
pknL (rv2176c)	Serine/threonine protein kinase
rv2177c	Transposase
pstS3 (rv0928)	Phosphate substrate-binding Lipoprotein
pstC2 (rv0929)	Phosphate ABC transporter permease
pstA1 (rv0930)	Phosphate ABC transporter permease
pknD (rv0931c)	Serine/threonine protein kinase
pstS2 (rv0932c)	Phosphate substrate-binding Lipoprotein
pstB (rv0933)	Phosphate ATP-binding protein
pstS1 (rv0934)	Phosphate substrate-binding Lipoprotein
pstC1 (rv0935)	Phosphate ABC transporter permease
pstA2 (rv0936)	Phosphate ABC transporter permease
rv1265	Hypothetical protein
pknH (rv1266c)	Serine/threonine protein kinase
embR (rv1267c)	Transcriptional regulator
rv1740 (vapB34)	Antitoxin
rv1741 (vapC34)	Ribonuclease
rv1742	Hypothetical protein
rv1743	Hypothetical protein
pknE	Serine/threonine protein kinase
rv2086	Hypothetical protein
pknJ (rv2088)	Serine/threonine protein kinase
pepE (rv2089c)	Dipeptidase
pknF (rv1746)	Serine/threonine protein kinase
rv1747	

	ABC transporter ATP-binding protein/permease
rv2912c	TetR family HTH-type transcriptional regulator
rv2913c	D-amino acid aminohydrolase
pknI (rv2914c)	Serine/threonine protein kinase
rv2915c	Hypothetical protein
ffh (rv2916c)	signal recognition particle protein

Table A.3 Structural parameters calculated from SAXS data analyses of overall samples of FHA, PknF and the mixtures of FHA + PknF. R_g: radius of gyration, MW: molecular weight and (D_{max}): maximum dimension.

Sample	R_g (Å)	MW (kDa)	D_{max} (Å)
FHA			
FHA_wt	35±1	29±4	120
T210E	33±0.8	28±3	110
T152E	34±0.7	29±3	110
Double mutant	28±0.7	29±3	94
PknF			
cPknF*	27±0.3	32±3	78
cPknF	28±1	30±1	90
Mixture			
FHA_wt + cPknF*	38±0.5	56±6	115
FHA_wt + cPknF	36±0.7	47±5	117

Table A.4 Clustering of the cPknF-FHA solutions proposed by HADDOCK2.4 webserver (17). Z-score relies on how many standard deviations from the average the cluster score is. HADDOCK scores are given in arbitrary units and energy terms in kcal/mol.

Cluster	HADDOCK score	Size	RMSD	Evdw	Eelec	Edesolv	Z
				cPknF/FHA-1			
1	-64.3 +/- 3.9	103	5.3 +/- 0.2	-15.0 +/- 4.1	-273.1 +/- 38.2	3.7 +/- 2.0	-0.7
2	-62.7 +/- 1.9	67	5.3 +/- 0.6	-19.0 +/- 2.0	-247.1 +/- 20.4	4.9 +/- 1.3	-0.4
3	-61.2 +/- 2.6	39	4.9 +/- 0.9	-14.1 +/- 3.8	-268.4 +/- 9.7	3.4 +/- 0.7	-0.1
4	-63.8 +/- 12.8	21	5.5 +/- 1.1	-7.6 +/- 1.7	-305.1 +/- 65.5	3.6 +/- 1.1	-0.6
7	-64.8 +/- 7.0	12	4.0 +/- 0.8	-26.8 +/- 8.2	-210.7 +/- 79.4	3.9 +/- 1.4	-0.8
8	-55.5 +/- 4.9	10	10.8 +/- 0.3	-8.8 +/- 7.8	-270.9 +/- 60.3	5.0 +/- 1.3	1.0
9	-65.3 +/- 19.1	9	5.1 +/- 0.3	-24.2 +/- 4.4	-228.4 +/- 99.8	4.2 +/- 1.6	-0.9
11	-52.1 +/- 5.3	7	3.9 +/- 0.9	-25.2 +/- 4.0	-148.6 +/- 8.1	2.0 +/- 0.9	1.7
12	-52.1 +/- 9.8	6	4.5 +/- 0.2	-22.0 +/- 0.8	-165.9 +/- 49.5	2.3 +/- 1.1	1.7
13	-66.4 +/- 10.5	6	11.4 +/- 0.3	-23.3 +/- 5.0	-262.1 +/- 43.0	8.0 +/- 1.1	-1.1
				cPknF/FHA-2			
1	-70.4 +/- 2.7	36	7.1 +/- 0.8	-15.9 +/- 2.4	-283.4 +/- 28.5	1.8 +/- 2.1	0.9
2	-71.8 +/- 6.0	28	6.5 +/- 0.5	-21.0 +/- 5.1	-264.2 +/- 19.7	0.9 +/- 0.5	0.6
4	-85.3 +/- 11.6	13	6.9 +/- 0.4	-16.6 +/- 3.3	-349.9 +/- 46.5	0.8 +/- 1.7	-2.2
9	-78.3 +/- 16.5	9	2.7 +/- 1.6	-20.9 +/- 5.5	-297.9 +/- 88.5	1.8 +/- 1.7	-0.8
11	-71.0 +/- 6.0	7	5.5 +/- 0.9	-18.0 +/- 4.5	-293.6 +/- 19.0	2.6 +/- 1.8	0.7
17	-74.2 +/- 7.9	5	7.3 +/- 0.3	-13.4 +/- 2.3	-328.3 +/- 43.4	1.4 +/- 1.1	0.1
21	-70.4 +/- 16.3	4	4.1 +/- 0.2	-24.9 +/- 6.0	-233.4 +/- 74.3	0.9 +/- 2.4	0.9
24	-69.8 +/- 19.9	4	6.9 +/- 0.2	-21.4 +/- 2.1	-269.3 +/- 88.0	3.0 +/- 0.9	1.0
26	-75.6 +/- 6.0	4	7.7 +/- 0.3	-27.1 +/- 3.4	-266.5 +/- 12.4	3.7 +/- 2.4	-0.2
27	-79.3 +/- 11.1	4	3.7 +/- 1.1	-11.4 +/- 4.8	-361.6 +/- 66.2	0.5 +/- 1.3	-1.0

Table A.5 Interface prediction of PknF-FHA domains performed through PISA server (19).

##	cPknF	Dist. [Å]	FHA-1
1	A:ASN 176 [N]	3.87	B:GLY 66 [O]
2	A:TPO 175 [O1P]	2.89	B:SER 47 [N]
3	A:TPO 173 [O]	3.86	B:ARG 48 [N]
4	A:PRO 169 [O]	1.95	B:ARG 48 [HH11]
5	A:TPO 173 [O1P]	1.71	B:ARG 48 [HH12]
6	A:TPO 173 [O]	3.19	B:ALA 49 [N]
7	A:TPO 173 [O2P]	3.23	B:ALA 49 [N]
8	A:TPO 175 [O]	3.31	B:LEU 68 [N]
9	A:TPO 175 [O3P]	2.83	B:LEU 68 [N]
10	A:TPO 175 [O3P]	2.72	B:ASN 69 [N]
11	A:TPO 175 [O3P]	1.86	B:ASN 69 [HD21]
##	**cPknF**	**Dist. [Å]**	**FHA-2**
1	A:SER 232 [OG]	2.96	B:ASN 236 [N]
2	A:TPO 175 [O1P]	2.98	B:ARG 249 [N]
3	A:TPO 173 [O2P]	2.04	B:ARG 249 [HH12]
4	A:TPO 175 [O2P]	2.64	B:HIS 250 [N]

Supplementary Figures

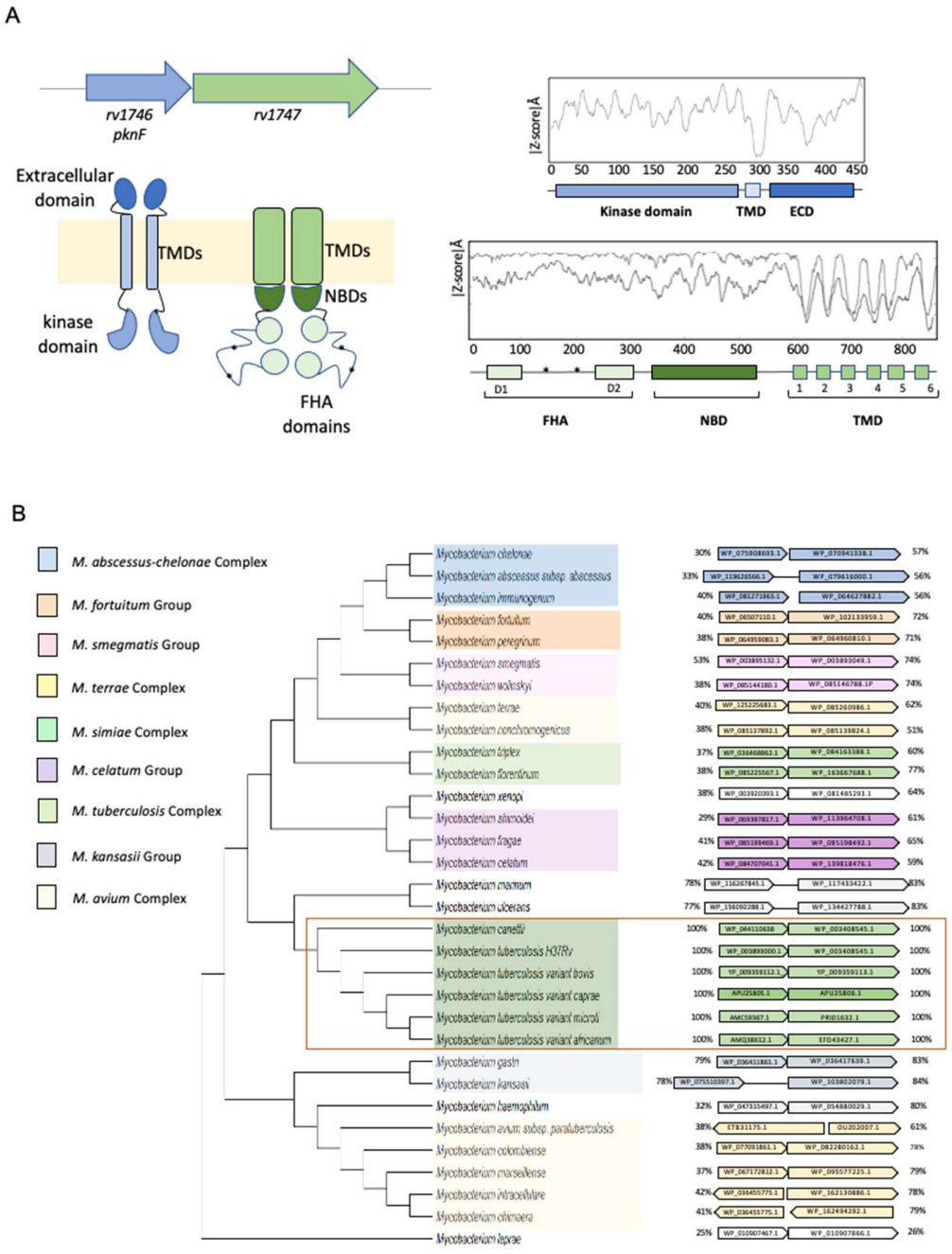


Figure A.1 Structural organization and conservation of the *Mtb rv1746/rv1747* operon components in different species of *Mycobacterium spp.* (*A*) *rv1746/rv1747* operon encodes the STPK PknF and the ABC transporter Rv1747. The prediction of transmembrane helices and structural organization of the proteins obtained with TOPCONS revealed that both are organized into the inner membrane. PknF has one TM helix that connects the extracellular domain (ECD) to the kinase domain (cPknF) located in the cytoplasm. The kinase is shown as a dimer. *rv1747* gene encodes the three

components of the transporter, which is organized as a homodimer. Each monomer consists of two Fork Head Associated domains (FHA-1 and FHA-2) in the N-terminal, two nucleotide-binding domains (NBDs) and a TMD that consists of six TM helices. FHA-1 and FHA-2 are connected by a disordered region of 150 amino acids where the two phosphorylated threonine residues (T152 and T210) are present (indicated with asterisks). (*B*) Conservation of the *rv1746/rv1747* operon in different species of the *Mycobacterium* genus and amino acid sequence identity of putative orthologs. The species segregate in different groups (showed in different colors) and the correlation with the presence of *pknF* (*rv1746*) and *rv1747* genes is presented as small and large arrows, respectively. The corresponding amino acid sequence identity of the encoded proteins is indicated in the side of the arrows and correspond to 100% of coverage for all the sequences. Clear arrows indicate the absence of sequences with significant similarity. The phylogenetic tree was based on RNA 16s of different species of mycobacteria. The evolutionary history was inferred using the Neighbor-Joining method and evolutionary analyzes were conducted in MEGA 6.0 with 1000 bootstrap. The list of the species (NCBI reference) is showed in *SI Appendix,* Table A.1.

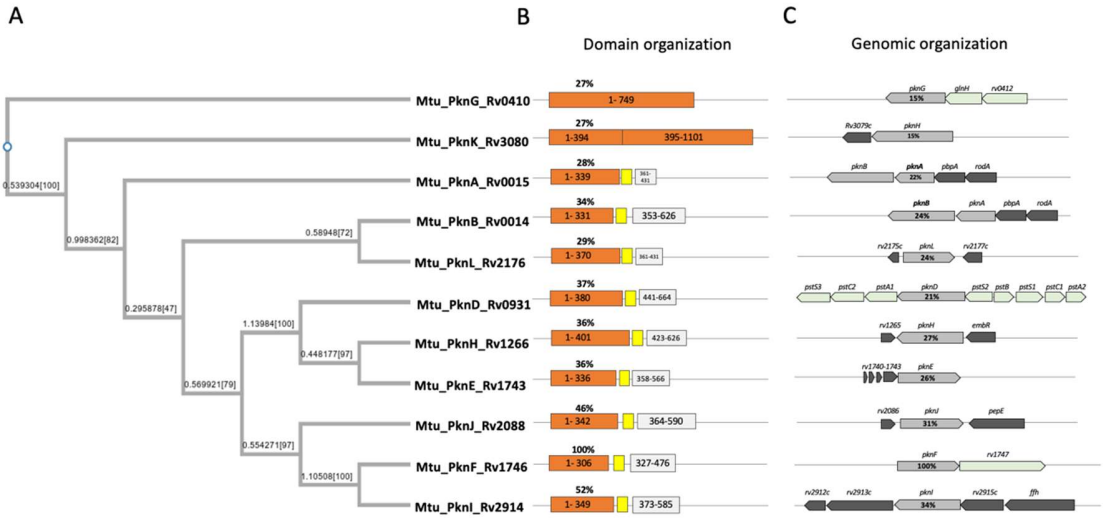


Figure A.2 The serine/threonine kinases of *Mtb.* (*A*) Phylogenetic relationship among the STP kinases. (*B*) Organization of the domains in the kinases. Kinase, transmembrane and extracellular domains are show as boxes in orange, yellow and clear gray, respectively. Percentages of amino acid sequence identity for kinase domains are shown above the boxes. (*C*) Genome localization of genes encoding the STPK in the genome of *Mtb*. The corresponding kinase genes are shown in gray arrows and the up- and downstream genes in black. Genes in green correspond to membrane proteins and associated domains. PknA and PknB belong to the same operon. The list of genes, references and their functions are shown in *SI Appendix,* Table A.2.

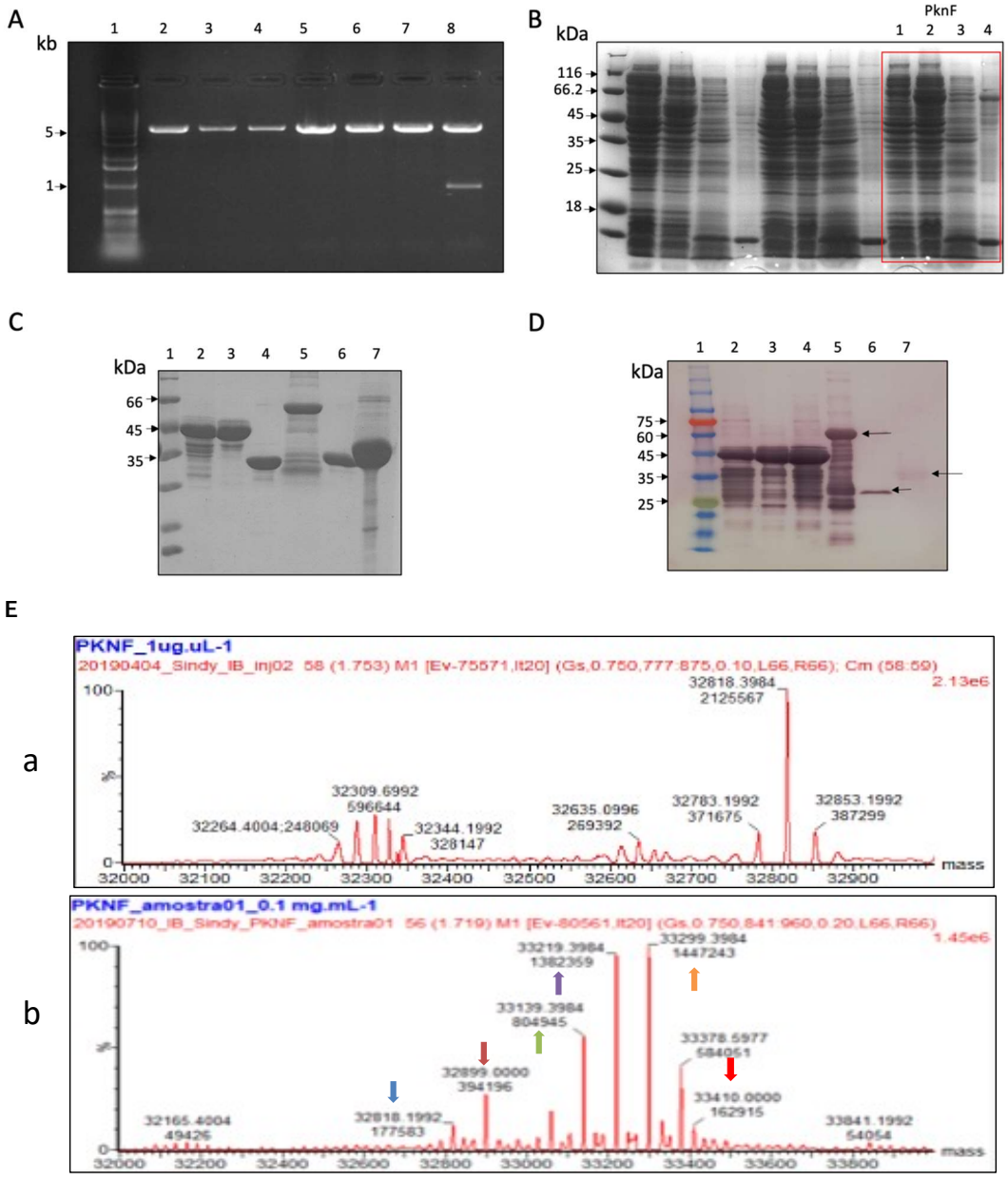


Figure A.3 Production of *Mtb* cPknF. (*A*) Restriction analysis of pOP5GT vectors extracted from transformed *E. coli* DH5a cells. 1: DNA ladder 1 Kb; 2-6: linearized pOP5GT vectors after digestion with *Eco*RI and *Hin*dIII endonucleases; 7: digested pOP5GT vector liberating the corresponding *cpknF* 911 bp fragment. (*B*) Expression analysis of the cPknF domain in fusion with GST after overnight induction in *E. coli* BL21 (D3) cells at 18°C, with 0.25 mM of IPTG and 200 rpm. 1: non induced extracts (T0); 2: induced extracts (T2); 3: soluble fraction; 4: insoluble fraction. (*C*) Purified recombinant FHA and cPknF domains after the two steps of purification and cleavage with protease. 1: molecular weight marker; 2 and 3: FHA wild type fused with GB1 protein obtained by nickel affinity chromatography (IMAC) and size-exclusion chromatography, respectively. 4: FHA wild type cleaved with the 3C protease and used to the further analyses. 5: cPknF fused with GST purified by immobilized metal affinity chromatography (expected mass 60 KDa); 6: cPknF after

cleavage with TEV protease (in non-phosphorylated state) with the GST removed (expected mass 32.8 KDa); 7: cPknF* (phosphorylated) after GST removal (expected mass 32.8 KDa). (*D*) Western blotting assay for detection of cPknF using anti-his-tag antibody (diluted 1: 20.000). 1, 2 and 3: Eluted fractions from IMAC of the FHA domain in fusion with His8-tag and GB1 recognized by the anti-His-tag antibody. 4: cPknF in fusion with His8-tag and GST after IMAC recognized by the anti-His-tag antibody; 5: IMAC cPknF fraction after TEV cleavage. 6: cPknF domain after cleavage and reverse affinity chromatography. As in the previous line, it is not possible to observe the band corresponding to the protein (expected size of 32.8 KDa) because after cleavage with TEV, the histidine tag was removed. (*E*) Analysis of cPknF samples using mass spectrometry. (I) Mass spectrum of the purified cPknF with mass of 32,818 Da (highest peak and absence of post-translational modification). (II) Mass spectra of cPknF after autophosphorylation in presence of 5 mM ATP for 30 min at room temperature (cPknF*). The narrows indicate the intact mass of the cPknF (blue) and several peaks corresponding to increments of 80,80 Da (orange), 321,19 Da (gray), 401,19 Da (yellow), 481,19 Da (green) and 560,59 da (red), which suggested incorporation of one, four, five, six and seven phosphate groups, respectively.

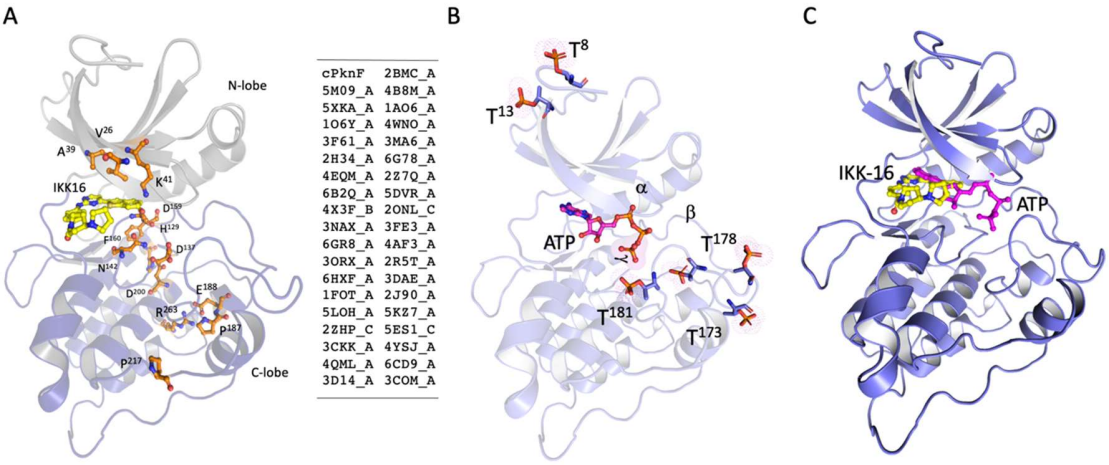


Figure A.4 Conservation of residues in the *Mtb* cPknF structure in comparison with orthologs. (*A*) The conserved residues in cPknF compared to available structures of TSPKs in PDB. The N and C lobes are coloured in grey and blue, respectively. Conserved residues are showed in orange sticks and the list of all TSPK used in the comparison is shown in the side table as PDB entries. (*B*) Putative phosphorylation sites in cPknF. Threonine residues are shown in blue sticks and the phosphor group in orange. The ATP is shown in the active site between lobes N and C. (*C*) Three-dimensional structure of cPknF shown in cartoon evidenced the localization of ATP (pink sticks) and IKK16 (yellow sticks) inhibitor.

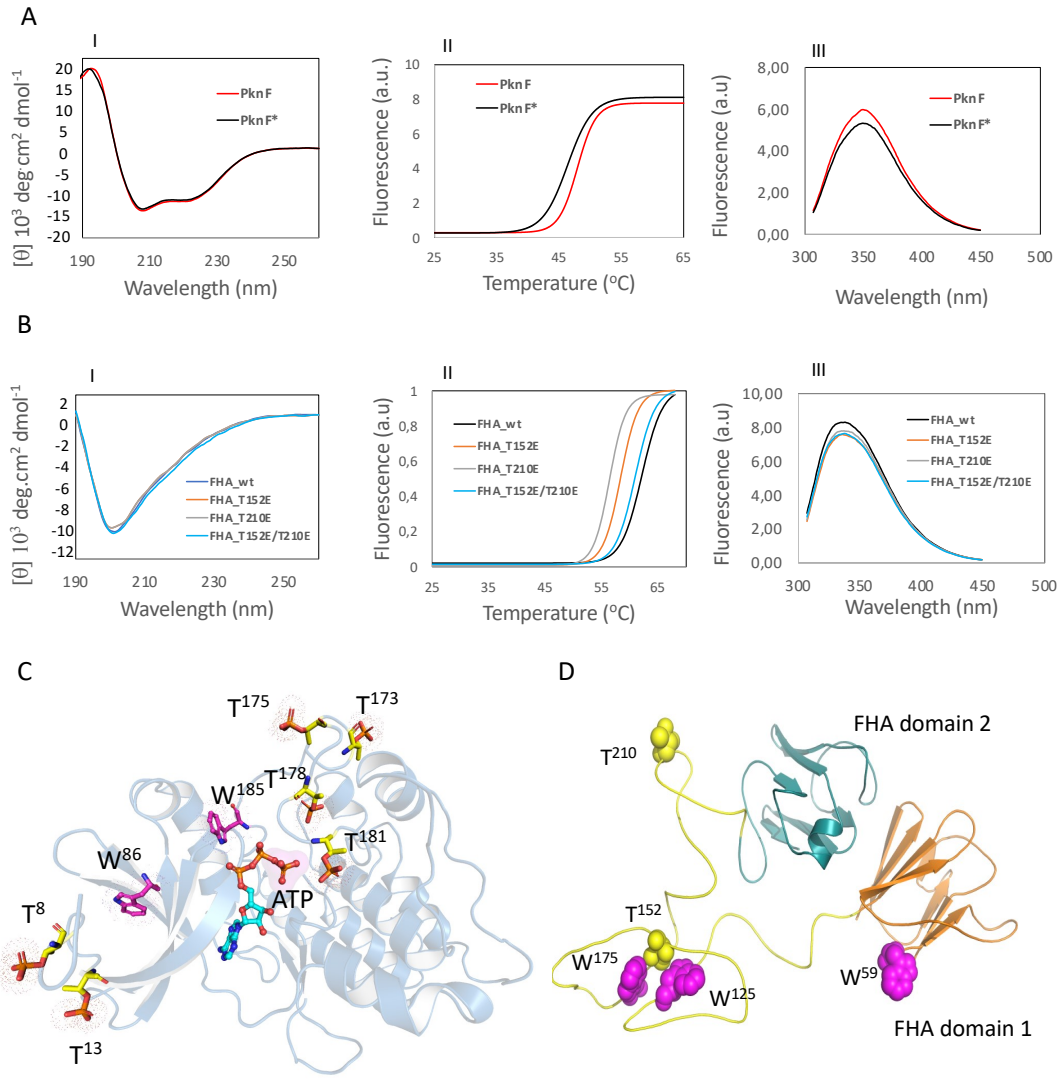


Figure A.5 Secondary structural profile and structural changes in the recombinant *Mtb* cPknF and Rv1747-FHA domains. (*A*) Biophysical analysis of cPknF. I: Secondary structural profile of non-phosphorylated (cPknF) and phosphorylated kinase (cPknF*) measured by dichroism circular. A typical spectrum of a protein with a high a-helix content is evidenced by the prominent negative peaks at 208 nm and 222 nm and a positive peak close to 190 nm. This spectrum remains unchanged after protein phosphorylation indicating that the content of secondary structure is the same in both samples. II: Thermal stability analysis of recombinant cPknF and cPknF* by DSF. The cPknF* shows a decrease in the melting temperature (Tm) when compared to the sample non-phosphorylated. III: Structural changes upon the phosphorylation of the cPknF domain measured by intrinsic fluorescence of tryptophan. The phosphorylation in cPknF* cause the quenching in the intensity of the emitted fluorescence. (*B*) Changes assessment in the secondary and tertiary structures of the FHA domain and its mutants. I: The Circular Dichroism spectrum for FHA and its phosphomimic mutants shows the profile characteristic of b-sandwich folding and a disorderly region with a of great magnitude around 200 nm. Phosphorylation does not alter the profile or content of secondary structures in the protein domains. II: Importance of the phosphorylation for the thermal stability of FHA domains measured by DFS. After the phosphorylation of one or both

threonine, the FHA domains suffer a loss of thermal stability evidenced by the decrease in the melting temperature. III: Analysis of the intrinsic tryptophan fluorescence of FHA domains. Phosphorylation causes conformational changes in the domain, as evidenced by the quenching of the tryptophan fluorescence after emulation of phosphorylation. (*C*) Structure of cPknF with phosphorylation sites and positioning of tryptophan. ATP, tryptophan and phospho-threonine are shown in cyan, magenta and yellow sticks, respectively. (*D*) Structural model of FHA domains indicating the positioning of the tryptophan (magenta spheres) and the threonine (yellow spheres).

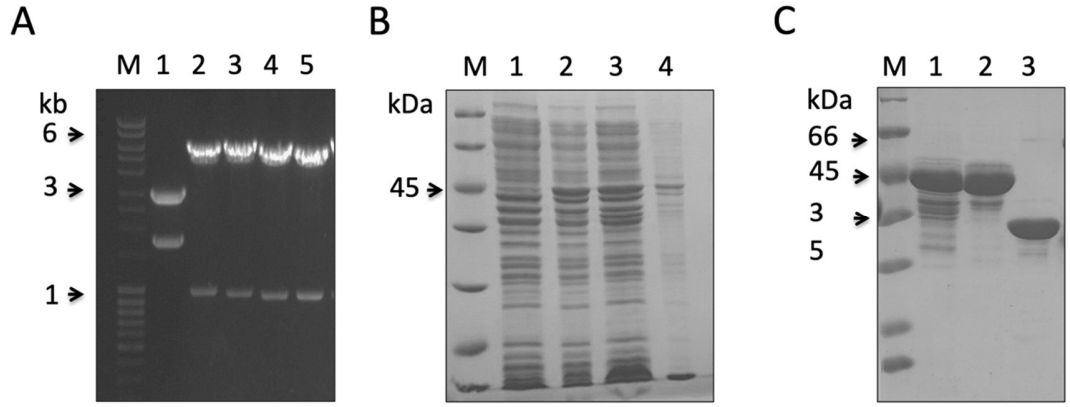


Figure A.6 Cloning, expression and purification of *Mtb* FHA domains. (*A*) Restriction analysis of clones of pOP3BP with FHA_wt and mutant domains. M: DNA ladder 1 Kb. 1: positive control of the digestion reaction; 2 to 5: pOP3BP_FHA wild type, pOP3BP_FHA_T152E, pOP3BP_FHA_T212E, and pOP3BP_FHA_T152E/T212E, after digestion with *Nco*I and *Hin*dIII endonucleases. (*B*) Expression analysis of FHA_wt in fusion with GB1 after induction for 2 hours in *E. coli* BL21 (D3) cells at 37°C, with 1 mM of IPTG and 200 rpm. 1: non induced extracts (T0); 2: induced extracts (T2); 3: soluble fraction; 4: insoluble fraction. (*C*) Purified recombinant proteins after the two steps of purification. M: molecular weight marker; 1 and 2: FHA wild type fused with GB1 protein obtained by nickel affinity chromatography (IMAC) and size-exclusion chromatography, respectively. 3: FHA wild type cleaved with the 3C protease and used to the further analyses. All the mutants produced exactly similar profiles and are not shown.

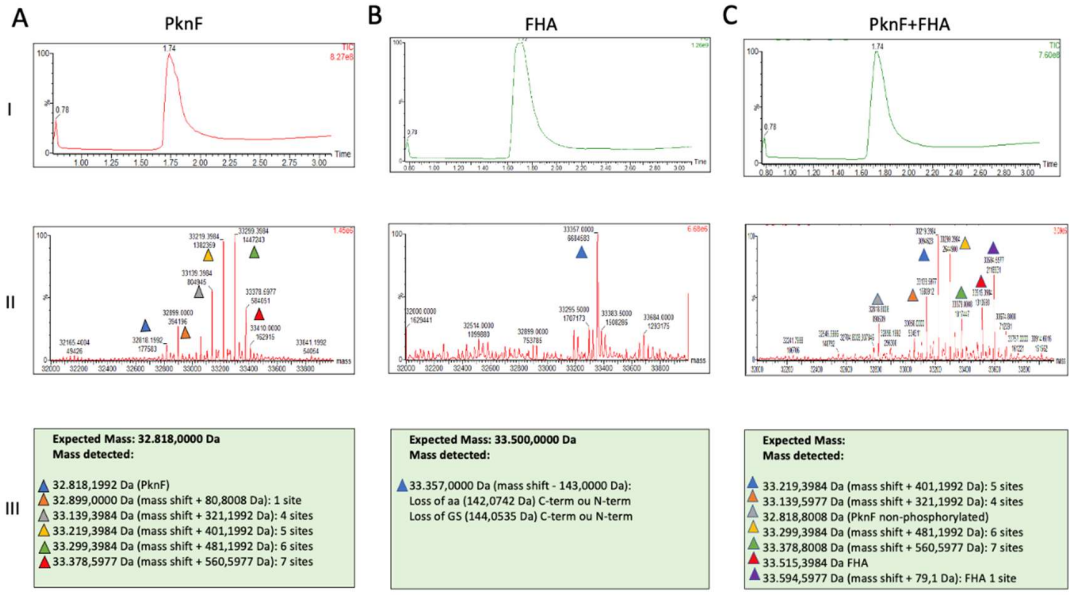


Figure A.7 Mass spectrometry of the *Mtb* FHA and cPknF domains. Samples of cPknF, FHA_wt and cPknF* + FHA_wt (equimolar ratio 1:1) were prepared in 50 mL of Tris-HCl pH 8.0 100 mM, NaCl 100 mM, DTT 1 mM, EDTA 5 mM, $MgCl_2$ 5 mM and ATP 5 mM, concentrated to 3 µM, incubated at room temperature for 30 min and 1 µL of each sample injected on an LC-MS/MS equipment (coupled to a mass spectrometer with a quadrupole-time-of-flight (QTof) analyzer model Waters XEVO G2 XS, equipped with an ESI type ionization source). (*A*) cPknF. (*B*) FHA_wt domain. (*C*) cPknF* + FHA_wt upon incubation at room temperature for 30 min. I: Total ion chromatograms; II: Deconvoluted spectra; III: expected and obtained masses in the assays for the three different domains.

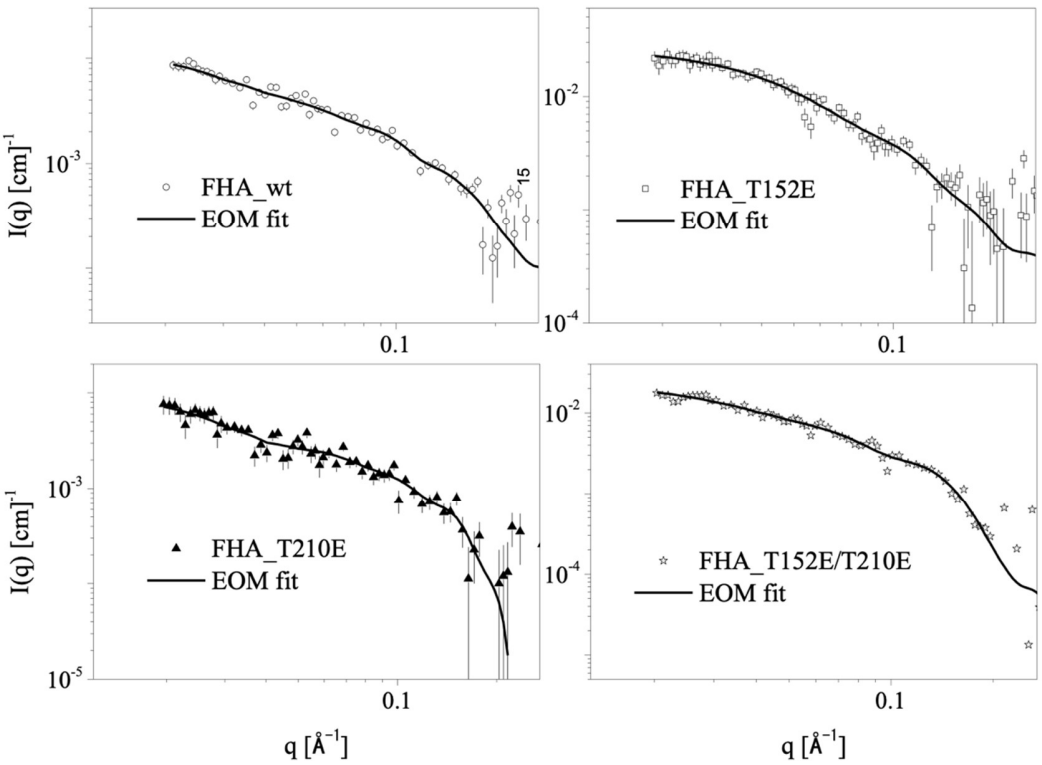


Figure A.8 EOM fit to FHA-1/FHA-2 full length constructs (wild type and mutants) measured by SAXS experiment. RANCH program was used for building of a pool of 10 000 structures of FHA-1/FHA-2 connected by a linker of 95 residues, based on the crystal model to FHA-1 (PDB 6CCD) and NMR model to FHA-2 (PDB 6CAH). The pool was used as input to GAJOE program that selected an optimized ensemble that best fit the SAXS data collected to FHA wild type and the mutant samples, individually, using the standard EOM configuration process for the method applied in the genetic algorithm (GA), as described in Material and methods.

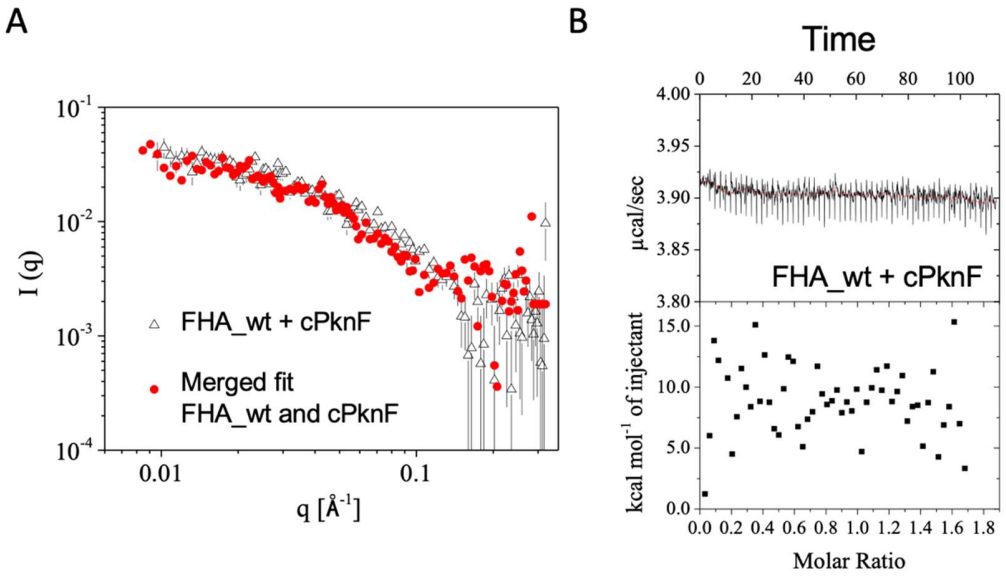


Figure A.9 Merging of SAXS data from isolated FHA and cPknF (red symbols) and the recorded SAXS data from the mixture sample of FHA and cPknF (square symbols).

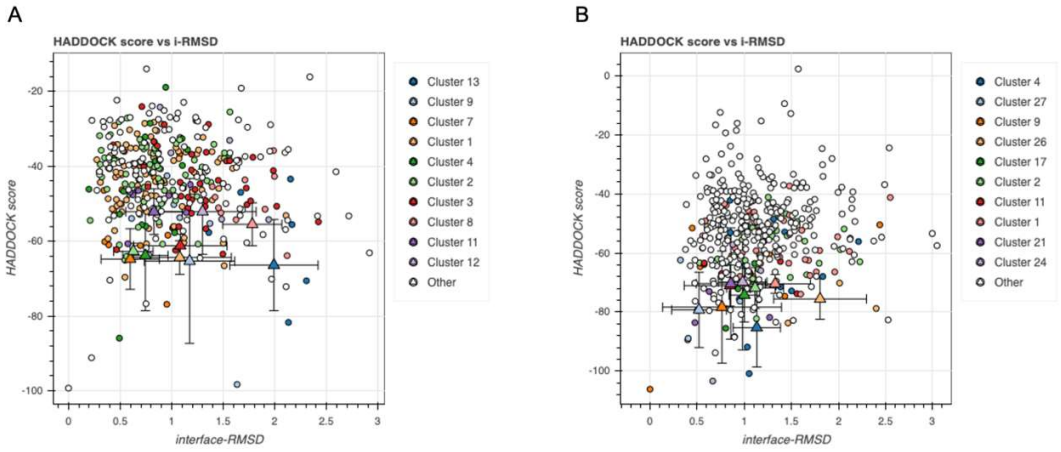


Figure A.10 HADDOCK score vs. i-RMSD plot of the complexes (A) cPknF-FHA1 and (B) cPknF-FHA2. The results demonstrate the dispersity between the clustered solutions not much higher than 3Å, suggesting reliability of the docking data.