



**HAL**  
open science

## **Detection of soluble co-factor dependent protein expression in vivo : Application to the 4'-phosphopantetheinyl transferase PptT from Mycobacterium tuberculosis**

Karine Rottier, Alexandre Faille, Thomas Prudhomme, Cécile Leblanc, Christian Chalut, Stéphanie Cabantous, Christophe Guilhot, Lionel Mourey, Jean-Denis Pedelacq

### ► To cite this version:

Karine Rottier, Alexandre Faille, Thomas Prudhomme, Cécile Leblanc, Christian Chalut, et al.. Detection of soluble co-factor dependent protein expression in vivo : Application to the 4'-phosphopantetheinyl transferase PptT from Mycobacterium tuberculosis. *Journal of Structural Biology*, 2013, 183 (3), pp.320 - 328. <10.1016/j.jsb.2013.07.010>. <hal-03002994>

**HAL Id: hal-03002994**

**<https://hal.science/hal-03002994v1>**

Submitted on 13 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

**Detection of soluble co-factor dependent protein expression *in vivo*: application to the 4'-phosphopantetheinyl transferase PptT from *Mycobacterium tuberculosis*.**

Karine Rottier<sup>1,2,\*</sup>, Alexandre Faille<sup>1,2,\*</sup>, Thomas Prudhomme<sup>1,2</sup>, Cécile Leblanc<sup>1,2</sup>, Christian Chalut<sup>1,2</sup>, Stéphanie Cabantous<sup>3,4,5</sup>, Christophe Guilhot<sup>1,2</sup>, Lionel Mourey<sup>1,2</sup>, and Jean-Denis Pedelacq<sup>1,2,•</sup>

<sup>1</sup>CNRS, IPBS (Institut de Pharmacologie et de Biologie Structurale), 205 route de Narbonne, BP 64182, F-31077 Toulouse, France; <sup>2</sup>Université de Toulouse, UPS, IPBS, F-31077 Toulouse, France, <sup>3</sup>INSERM UMR1037-Cancer Research Center of Toulouse, <sup>4</sup>Université de Toulouse, <sup>5</sup>Institut Claudius Regaud, 20-24 rue du Pont St-Pierre, 31052 Toulouse Cedex. \*These authors contributed equally to this work.

•Correspondence should be addressed to J.-D. Pedelacq ([Jean-Denis.Pedelacq@ipbs.fr](mailto:Jean-Denis.Pedelacq@ipbs.fr))

## Abstract

Ways to reduce the toxicity and improve the solubility of recombinant proteins in heterologous hosts have been the subject of prolific research over the past two decades. Major progresses have been made with the advent of structural genomics initiatives and the development of tools to circumvent this major obstacle. Available strains and expression vectors now make possible parallel approaches to increase the chances of isolating protein targets in a soluble form. In addition, DNA-shuffled and randomly truncated libraries of DNA fragments have revolutionized our approach to functional and structural characterization of recalcitrant proteins. Nevertheless, all these approaches suffer from the absence of early-on diagnostic tools to assess the folding and solubility of the expressed protein constructs *in vivo*. In this paper, we took advantage of the sub-picomolar detection level of the bipartite GFP1-10/GFP11 system to investigate the solubility of the *Mycobacterium tuberculosis* 4'-phosphopantetheinyl transferase PptT, an enzyme essential for the viability of the tubercle bacillus. *In vivo* and *in vitro* complementation assays clearly showed the improved solubility of the full-length PptT compared to its N-terminally and C-terminally truncated counterparts. Surprisingly, initial attempts to purify the full-length enzyme overexpressed in *E. coli* cells were hampered by aggregation issues that caused the protein to precipitate within hours. We have shown that the naturally occurring co-enzyme A and  $Mg^{2+}$  cofactors, essential for PptT to carry out its function, also play a major role in stabilizing the enzyme *in vitro*.

## Keywords

Split-GFP assay, domain trapping, 4'-phosphopantetheinyl transferase, polyketide, tuberculosis.

## Abbreviations

AcpS:	Acyl carrier protein synthase
AnTet:	Anhydrotetracycline
BSA:	Bovine serum albumin
CoA:	Co-enzyme A
DHFR:	Dihydrofolate reductase
DSF:	Differential scanning fluorimetry
DTT:	Dithiothreitol
GFP:	Green fluorescent protein
IMAC:	Immobilized metal ion affinity chromatography
IPTG:	Isopropyl $\beta$ -D-1-thiogalactopyranoside
Kan:	Kanamycin
LB:	Luria-Bertani
MBP:	Maltose binding protein
PCR:	Polymerase chain reaction
PDIM	Phtiocerol dimycocerosate
PKS:	Polyketide synthase
P-pant:	4'-Phosphopantetheine
SDS-PAGE:	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC:	Size-exclusion chromatography
Sfp:	Surfactin synthetase-activating enzyme
Spec:	Spectinomycin
SUMO:	Small ubiquitin-related modifier

## Introduction

High-level expression of a protein target in a soluble form is a prerequisite to its downstream functional and structural characterization. *Escherichia coli*, the most widely-used heterologous host for the production of recombinant protein, has incommensurable benefits over existing systems: well-characterized genetics, easy to transform, easy growth, and inexpensive. The diversity of *E. coli* strains and expression vectors publicly available now permits the production of most cytoplasmic proteins. Engineered *E. coli* strains resistant to the toxicity of membrane protein expression have been described<sup>1; 2; 3</sup>. Good progress has also been made in overexpressing eukaryotic proteins<sup>4</sup>, a major breakthrough considering the lack of post-translational machinery and the limiting codon usage in the *E. coli* host. Sometimes, proteins partition almost exclusively into “inclusion bodies” when overexpressed in heterologous hosts as a result of misfolding, aggregation, and intracellular accumulation. To circumvent this issue, cell-free protein synthesis systems and refolding strategies have been implemented<sup>5</sup>.

As an alternative to the time-consuming and rate-limiting steps that best characterize the search for optimized expression and solubility conditions in heterologous hosts, new strategies have emerged that specifically focus on modifying the underlying DNA sequence. With the advent of the error-prone PCR technique<sup>6</sup>, introducing mutations that enhance solubility and stability of a protein of interest while retaining its enzymatic activity has now become possible<sup>7; 8; 9</sup>. Technology improvements have also considerably reduced the cost of high-throughput screening of DNA fragment libraries thanks to the development of colony picking, arraying and liquid handling robots<sup>10</sup> coupled with

biotechnology tools. Strategies to filter out DNA fragments that do not encode authentic regions or domains of a protein of interest have been implemented<sup>11; 12</sup>.

In this paper, we detail a new application of the split-GFP complementation assay<sup>13</sup> as a diagnostic tool to investigate the solubility of co-factor dependent proteins expressed *in vivo*. We used this system onto the 4'-phosphopantetheinyl transferase PptT from *Mycobacterium tuberculosis*, a coenzyme A (CoA) binding enzyme essential for the survival of mycobacteria<sup>14; 15; 16</sup>. *E. coli* assays indicated that the full-length enzyme was the most soluble among all the variants tested. We have shown that the natural endogenous CoA and divalent metal ion Mg<sup>2+</sup>, which play a crucial role for PptT to carry out its function, are also a prerequisite for its stability overtime. This easy-to-use split-GFP system can be generalized to detect soluble co-factor dependent protein expression *in vivo* whenever the presence of cofactors is required for folding and stability.

## Materials and Methods

***PptT small scale expression and solubility tests.*** The *pptT* gene from *M. tuberculosis* was amplified from genomic DNA using top and bottom primers listed in supplementary table 1. The NdeI/BamHI digested *pptT* was then inserted into commercially available pET28 vectors (Novagen, Madison, WI, USA) with no 6His tag, N-terminal or C-terminal 6His tag. pET vectors allowing the production of PptT with MBP<sup>17</sup> or SUMO (Invitrogen, Carlsbad, CA, USA) fused to its N-terminus were also tested. Ligated plasmids were transformed into chemically competent *E. coli* BL21(DE3) cells (Invitrogen, Carlsbad, CA, USA). Transformed cells were plated onto Luria-Bertani (LB)

agar plates containing 35 mg/ml kanamycin, allowing them to grow overnight at 32°C. The resulting clones were grown at 37°C in 1ml cultures using 35 mg/ml kanamycin. Cells were induced in exponential phase with 1mM IPTG for 3 h. Cell culture pellets of 1 ml of each fragment were separately resuspended in 40 ml TNG buffer (150 mM NaCl, 100 mM Tris-HCl pH=7.5, 10% (v/v) glycerol) and sonicated. The lysate was fractionated by centrifugation to yield the soluble and pellet fractions. The pellet fraction was washed twice with 100 ml TNG buffer, centrifuged and resuspended in the same starting volume. Samples corresponding to the soluble (S) and pellet (P) fractions were resolved on a 4–20% gradient Criterion SDS-PAGE gel (Bio-Rad, Hercules, CA, USA). Protein samples were stained using Gel Code Blue stain reagent (Pierce, Rockford, IL) and imaged using a GS-800 Calibrated Densitometer (Biorad, Hercules, CA, USA).

***Cloning and in vivo split-GFP solubility screen.*** Full-length and truncated variants of the *M. tuberculosis PptT* gene were amplified by PCR using a series of forward and reverse primers listed in supplementary table 1. Cleaned inserts from *NdeI/SpeI* restriction digests were ligated into the pTET CoIE1 GFP11 vector and transformed into chemically competent BL21(DE3) cells containing the pET GFP 1–10 plasmid<sup>13; 18</sup>. Frozen cells were used to grow 1 ml Luria-Bertani (LB) medium in a 96-well deep-well plate containing 35 µg/ml kanamycin and 112 µg/ml spectinomycin until OD<sub>600</sub> reached ~1.0 at 37°C. 50 µl of two successive 400-fold dilutions of the expressed constructs were plated onto a compartmentalized nitrovivose membrane and grew overnight at 32°C. The membrane was transferred to an LB/Agar plate containing the same antibiotics supplemented with 250 ng/ml anhydrotetracycline (AnTET) for 2 h at 32°C, and then

moved back onto the original plate for 1 h at the same temperature. Finally, the membrane was transferred onto a plate with same antibiotics and 1 mM IPTG to induce the expression of the GFP 1-10 fragment at 32°C. Colonies were illuminated using an Illumatool Bright light System LT-9900 (<http://www.lighttools.com/>) at excitation and emission wavelengths of 470 nm and 515 nm, respectively. Pictures were taken after 30 min complementation at 32°C.

***In vitro split-GFP complementation assay.*** 5 ml LB-Spec cultures were grown at 37°C until OD<sub>600</sub> reached 0.5-0.7. After induction at 32°C with AnTET at a final concentration of 8 µg/ml and growth for an additional 4 h, cells were harvested by centrifugation at 20000g for 20 min. 20 µl of PptT-GFP11 soluble fractions were mixed with 180 µl of 0.35 mg/ml refolded GFP1-10 in a 96 well microplate (Nunc-Immuno<sup>TM</sup>plate, Nunc, Rochester, NY), as previously described<sup>18</sup>. Fluorescence kinetics ( $\lambda_{exc}=488$  nm /  $\lambda_{em}=530$  nm) were monitored with a FL600 Microplate Fluorescence Reader (Bio-Tek, Winooski, VT), at 3 min intervals, for 15 h. The background fluorescence of a blank sample (20 µl *E. coli* lysate expressing irrelevant protein, 100 µl of 0.5 mg/ml GFP1-10 and 100 µl 0.5% BSA in TNG buffer) was subtracted from final fluorescence values. The blank was less than 30% the signal from the lowest target concentration (0.1 pmol SR-S11).

***Full-length PptT production and purification.*** Frozen cells from the full-length PptT constructs ligated into a pTet vector missing the GFP11 tag were used to start an overnight 3 ml LB-spectinomycin (112 µg/ml) culture at 32°C prior to inoculation in baffled flasks containing 500 ml of the same media. Cells were allowed to grow for



approximately 2 h at 37°C before temperature was dropped from 37°C to 20°C. When OD<sub>600</sub> reached 0.5-0.7, cells were induced with AnTET at a final concentration of 8 µg/ml and grown for an additional 18 h prior to harvesting by centrifugation at 4000 g for 30 min and storage at -80°C. The pelleted cells were suspended in 40 ml of buffer A (100 mM Tris pH 8.0, 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 50 µM CoA) and lysed by sonication (6 cycles of 30 s pulse, 50% amplitude, power 5) prior to centrifugation at 20,000 g for 1 h. The supernatant was filtered (0.2 µm) and purified on an ÄKTA purifier system (GE Healthcare) following a two-step procedure combining an immobilized metal ion affinity chromatography (IMAC) and a size-exclusion chromatography (SEC). The lysate was loaded onto a 1 ml HisTrap HP (GE Healthcare) affinity column. The N-terminally His tagged PptT was eluted from the column with a step of buffer A supplemented with 120 mM imidazole. The eluted protein fractions were pooled and concentrated in Vivaspin® 20 ml centrifugal down to 2 ml prior to injection into a HiLoad 16/60 Superdex 200 (GE Healthcare) pre-equilibrated with buffer B (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 50 µM CoA).

***ACP fragment cloning and purification.*** The *ppsC* gene from *Mycobacterium tuberculosis* was cloned into the NdeI/SpeI sites of a pET26b plasmid (Novagen, Madison, WI) and PCR amplified using gene specific forward 5'-GATATACCATATGACCGCAGCGACACCAGATCG-3' and reverse 5'-AATTCACTAGTTGACTCGCCTCGCGTCGCAGC-3' primers. DNA fragmentation, cloning of the DNA library into the DHFR ORF filter and the split-GFP solubility reporter<sup>18</sup> were conducted as previously described<sup>11</sup>. Briefly, a 400 to 850 bp DNA

library was created by mechanical shearing of the PCR amplified *ppsC* gene. Blunt fragments were ligated in a *Stu*I-digested insertion DHFR pET vector to select for in-frame fragment in the presence of 6 µg/ml trimethoprim. Inverse PCR was performed with fragments ligated into the pTET ColE1 GFP 11 vector using phosphorylated forward (5'-CGACTCGCTGATGGGCCTGGAATTGCGCAATC-3') and reverse (5'-AGTCCCAGGGTTTCCAGCGGTCGGTGGTGATC-3') primers. After transformation into chemically competent BL21(DE3) pET GFP1-10 cells, ninety-six individually picked clones were grown overnight at 30°C in a 96-well tissue culture plate containing 7.5% glycerol in LB-Kan-Spec medium prior to sequencing. DNA sequences were analyzed using BioEdit® software and aligned onto the full-length parent gene to determine the exact boundaries. Based on the *in vitro* solubility assays<sup>18</sup>, fragments were color-coded black, dark gray, and light gray, where the black side of the spectrum identifies the least soluble ACP fragments and the light gray side corresponds to the top 25% most soluble ones. A DNA fragment encompassing the ACP domain and linker regions of PpsC was amplified by conventional PCR and was cloned into the *Nde*I/*Bam*HI restriction sites of a pET28a plasmid (Novagen, Madison, WI) using GATATACATATGCATGACTCGGCGGCCCGCAAAA and AATTCGGATCCTGACTCGCCTCGCGTCGCAGCTT as forward and reverse primers, respectively. The resulting plasmid was transformed into chemically competent BL21(DE3) *E. coli* cells. Cells expressing the ACP fragment were grown to OD<sub>600</sub> ~ 0.6, induced with 0.25 mM IPTG for 4 h at 30°C, pelleted by centrifugation at 4000 g for 30 min and stored at -80°C. The frozen pelleted cells were resuspended in 40 ml of buffer A and lysed by sonication (6 cycles of 30 s pulse, 50% amplitude, power 5) prior to

centrifugation at 45,000 *g* for 30 min. 4 ml of Talon Superflow beads (Talon resin, Clontech, Palo Alto, CA) conditioned in buffer A were added to the supernatant for 5 min prior to centrifugation at 1,000 *g* for 5 min. The unbound fraction was removed and the beads were washed twice with 50 ml of buffer A then 30 ml of buffer A supplemented with 10 mM imidazole to eliminate non-specific binding to the beads. The N-terminally His-tagged protein was eluted with 5 ml of buffer A containing 250 mM imidazole. The protein elution sample was resolved on a 4-20% gradient SDS-PAGE gel and stained using a Coomassie® Blue R-250 staining solution. The protein solution was stored at -20°C after dialysis overnight against 25 mM Tris pH 8.0, 25 mM NaCl with 10 % glycerol.

***DSF experiments.*** Fluorescence of the SYPRO® Orange Protein Gel Stain (Invitrogen, USA) was measured using the CFX96™ real-time PCR detection System (Bio-Rad, Hercules, CA) from 20°C to 70°C in 0.3°C increments of 3 seconds. PptT was concentrated to 400 μM in the presence of buffer A (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 50 μM CoA) prior to a 10-fold dilution with either buffer A or buffer B (50 mM Tris pH 8.0, 50 mM NaCl). 20 μl mixtures containing 2 μl PptT 40 μM and 1 μl Sypro Orange 200X were placed in 96-well clear-bottom Hard-Shell® 96-Well Skirted PCR Plates, Low-Profile (BioRad). Addition of 17 μl of buffer A or B completes the reaction mixture. Each condition was realized in triplicate. Melting temperatures were then calculated by the BioRad CFX Manager software.

**Activity assays.** PptT at 50 nM was incubated at 30°C with 10 μM of ACP domain of PpsC. 20 μM CoA, 25 mM DTT and 10 mM MgCl<sub>2</sub> were also present for the reaction to take place. Reaction was then stopped at different times with the addition of 100 mM EDTA. The same experiment was conducted in parallel using Sfp from *B. subtilis* (NEB, Ipswich, MA) at the same concentration instead of PptT. Samples were then loaded on a 10% polyacrylamide gel supplemented with 2.5 M urea. Positive and negative controls for the transfer of the Ppant moiety consisted in the incubation at 30°C with the same reactants for one hour with or without the presence of 200 nM Sfp as the only enzyme, respectively.

## Results

**Expression of PptT in different bacterial vectors.** We first cloned *pptT* into pET28 derived vectors prior to transformation into classical BL21(DE3) cells. Only in a situation where the histidine tag is present at the N-terminal extremity of the protein that overexpression of PptT can be achieved. The construct is poorly soluble and expresses mainly as inclusion bodies after IPTG induction (**Figure 1**). To maximize the amount of soluble protein, one strategy was to screen conditions where the inclusion bodies could be refolded in a stable and active conformation. This was achieved in the presence of 100 mM MES at pH 5.5. Unfortunately, analytical gel filtration experiments indicated an aggregated form of the enzyme detected in the exclusion volume of the column and the activity of the enzyme could not be confirmed.

As an alternative, we cloned *pptT* into pET vectors with fusion partners MBP (Maltose binding protein)<sup>19</sup> and SUMO (Small ubiquitin-related modifier)<sup>20</sup>, previously described as solubility enhancers of passenger proteins (see Materials and Methods). On the one hand, SUMO had no effect on the solubility of the fusion. On the other hand, the PptT-MBP fusion was mostly soluble and could be purified easily following a two-step chromatographic procedure using metal-affinity resin followed by size exclusion chromatography. Unfortunately, the enzyme precipitates massively after proteolytic cleavage of the C-terminal MBP tag, which indicates that the MBP fusion maintains PptT in a soluble form. This MBP-PptT fusion was used for the implementation of an enzymatic assay suitable for high-throughput screening of small molecule libraries<sup>17</sup>.

***Split-GFP solubility screen of PptT variants.*** Single amino-acid truncation at the N-terminal or C-terminal extremity of a protein of interest often has a dramatic effect onto the solubility level of the expressed variant<sup>11; 21</sup>. With this in mind, we explored the expression and solubility levels of an ensemble of 48 constructs *in vivo* (**Figure 2a,b**), corresponding to the full-length PptT along with N- and C-terminally truncated variants, using the split-GFP based assay developed by Cabantous and colleagues<sup>13</sup>. In this system, two separately inducible expression systems are present in the same *E. coli* cell, one tetracycline-inducible vector bearing the small GFP11 tag N-terminally fused to the protein of interest while the large complementary fragment GFP1-10 is located onto a separate IPTG-inducible expression vector. After sequential induction, only soluble GFP11 fusions can spontaneously bind to GFP1-10, thus restoring the fluorescence of the full-length GFP. In this system, fluorescent complementation is directly proportional to

the amount of soluble protein present in the living cells or purified from cell crude extracts, with detection levels within the sub-picomole range. *In vivo*, the highest fluorescence levels could be attained for the full-length enzyme and N-terminally truncated variants whereas truncation of the last C-terminal residue L227 was sufficient to negatively impact the fluorescence intensity of the colonies (**Figure 2b**). To confirm this result, an *in vitro* split-GFP complementation kinetic assay was performed by adding a large excess of GFP1-10 to the N-terminally and C-terminally truncated variants in the soluble fraction. After fifteen hours of complementation at room temperature, fluorescence intensity measured for the  $\Delta 227$  deletion mutant was reduced by ~75% compared to the full-length enzyme (**Figure 2c**). Fluorescence intensity levels measured for the other six C-terminally truncated variants, from position 226 to position 220, remained low and comparable.

***Both CoA and MgCl<sub>2</sub> are essential for the stability of PptT.*** *In vivo* expression and solubility levels of the recombinant His-PptT-GFP11 fusion were assessed by comparing the fluorescence intensities of co-induced and sequentially induced *E. coli* cell cultures, as previously described<sup>18</sup>. Approximately one third of the enzyme could be expressed in a soluble form, a result that was further confirmed *in vitro* by quantification of the corresponding band intensities on a SDS-PAGE gel (data not shown). A two-step purification procedure combining immobilized metal ion affinity chromatography (IMAC) and size-exclusion chromatography (SEC) indicated that the protein was mostly aggregated and could precipitate within hours.

Knowing the essentiality of CoA and  $Mg^{2+}$  for PptT to carry out its function *in vivo*<sup>22</sup>, we were wondering if these compounds could also help maintaining the enzyme in a stable conformation *in vitro*. Only in a situation where CoA and  $MgCl_2$  are present in a large molar excess over PptT in the cell lysate and all along the two-step purification strategy that high yields of soluble recombinant protein can be attained. We then conducted a series of experiments using the Differential Scanning Fluorimetry (DSF) method<sup>23</sup> to confirm the importance of the two protagonists onto the stability of PptT. We used Sypro Orange to detect the buried hydrophobic regions of the enzyme that become progressively exposed to solvent during a linear temperature ramp from 20°C to 70°C. The low initial fluorescence background and the high signal-to-noise fluorescence ratios have facilitated an accurate measurement of the transition midpoint ( $T_m$ ) values. High  $T_m$  values around 50°C were measured in the presence of CoA and  $MgCl_2$  at a final concentration of 50  $\mu$ M and 10 mM, respectively, corresponding to the purification conditions. Diluting PptT so that CoA concentration was reduced to trace amount (0.5  $\mu$ M final) and a 25-fold molar excess of  $MgCl_2$  (100  $\mu$ M final) over PptT profoundly impacted the stability of the full-length enzyme as no  $T_m$  values could be measured.

***In vitro enzymatic activity of PptT.*** PptT catalyzes the covalent transfer of the 4'-phosphopantetheine (P-pant) group from CoA onto the ACP domain of various type I polyketide synthases (PKS)<sup>15; 17</sup>. Issues related to heterologous host expression, folding and solubility of these large-size multi-domain enzymes forced us to identify a more compact and stand-alone soluble fragment encompassing the functional ACP domain. We used our recently published *domain trapping* strategy<sup>11</sup> to generate a 400-850 bp fragment library centered onto the ACP domain of PpsC, a 2188 amino-acid enzyme

involved in the biosynthesis of phthiocerol dimycocerosates (PDIM). These lipids play a major role in the virulence of this microbial pathogen<sup>24</sup>. Our approach combines a two-body dihydrofolate reductase (DHFR) scaffolding for selecting in frame DNA sequences from a random library of the fragmented *ppsC* gene, with the split-GFP technology<sup>18</sup> to identify soluble candidates centered onto the ACP domain (**Figure 3a**). The main advantage of our approach lies in the efficiency of the DHFR step to filter-out ~95% of the constructs present in the original DNA library and the use of an inverse PCR step to select for fragments encompassing a specific region or domain of a given protein. The implementation of the split-GFP assay to distinguish colonies on the basis of their fluorescence intensities completes this highly reliable and efficient strategy. We picked a total of 96 clones with fluorescence intensity levels varying from bright to faint depending on the solubility of the expressed fragment. Activity assays were carried out in the presence of the most soluble and compact ACP fragment 2042-2188 using either PptT or its closest homologue Sfp from *Bacillus subtilis*. We then visualized differences in the migration of the ACP domain on a 10% native polyacrylamide gel supplemented with 2.5 M urea (**Figure 3b**). Knowing that the reaction catalyzed by Sfp is magnesium-<sup>25</sup> or manganese-dependent<sup>26</sup> whereas the human AcpS X-ray structure revealed the presence of calcium in its active site<sup>27</sup>, we then explored the role of all three divalent metal ions onto the activity of PptT. No activation of the ACP domain was detected in the presence of 10 mM CaCl<sub>2</sub> with either PptT or Sfp. When replacing CaCl<sub>2</sub> with MnCl<sub>2</sub>, activation of the ACP was noticeable after 30 min incubation, only in the presence of PptT. At 10 mM MgCl<sub>2</sub>, approximately half of the ACP domain was activated by the transfer of the Ppant moiety after 5 and 20 minutes incubation with PptT and Sfp, respectively. In



conclusion,  $Mg^{2+}$  resulted in the highest activity among the divalent ions tested followed by  $Mn^{2+}$ , while  $Ca^{2+}$  has no effect on the activity of the enzyme activity.

## Discussion

Endogenous or exogenous ligands are often needed to stabilize a protein of interest in a particular conformation to carry out its function. Hosts such as *E. coli* have a great reservoir of co-factors that may be requisitioned for the correct folding and stability of the nascent protein. Although co-factors are produced in relatively small amounts compared to the overexpressed protein, early detection assay with sub-picomolar sensitivity, such as the one provided by the split-GFP system, can be very helpful to identify those proteins amenable to downstream functional and structural characterization. Optimizing the amount of soluble protein expressed can be attained using a systematic truncation strategy to cherry pick the best candidates.

In this study, we have applied the split-GFP system to the 4'-phosphopantetheinyl transferase (PptT) from *M. tuberculosis*, an enzyme required for growth and persistence of the bacteria *in vivo*<sup>17</sup>. PptT is responsible for the covalent transfer of the P-pant group of CoA to a conserved serine residue onto the acyl carrier protein (ACP) domain of various type I PKS. This post-translational modification converts PKS into functional *holo* enzymes that play a key role in the biosynthesis of various lipids in the cell envelope<sup>15</sup>. The *pptT* gene consists of 681 bp with a GC content of ~66% similar to all mycobacterial species. *pptT* also contains a total number of ten rare leucine, arginine and proline codons (<http://nihserver.mbi.ucla.edu/RACC/>). Despite the limited number of

tRNA for rare codons in the BL21(DE3) *E. coli* strain, expression levels remained high. Nevertheless, *E. coli* expression of PptT using pET vectors was hampered by solubility issues that caused the protein to partition exclusively into inclusion bodies, even after cleavage of the MBP solubility enhancer (**Figure 1**). To the opposite, its closest homologue, the surfactin synthetase-activating enzyme (Sfp) from *B. subtilis*, yields ~10 mg of pure protein from a 400 ml BL21(DE3) cell culture<sup>28</sup>.

We took advantage of the weak AnTet-inducible promoter of the GFP11 vector<sup>13</sup> to investigate the solubility levels of an ensemble of forty-eight protein constructs *in vivo*, corresponding to the full-length enzyme along with N- and C-terminally truncated variants. Surprisingly, the full-length PptT appeared to be the most soluble protein construct after sequential induction *in vivo* (**Figure 2**). Although the *in vitro* experiments followed the same trend, aggregation of the protein over time ultimately led to its precipitation within hours. In an attempt to evaluate the contribution of the cellular cofactors  $Mg^{2+}$  and CoA on the stability of PptT *in vitro*, a series of experiments using the DSF method was performed. High quality and reproducible fluorescent melting curves with  $T_m$  values of 50°C could be measured only when CoA and  $MgCl_2$  were maintained in a large molar excess over PptT, at concentrations equivalent to the purification conditions. Essentiality of these two cofactors for PptT to carry out its function was also demonstrated using a soluble fragment encompassing the ACP domain of the polyketide synthase PpsC (**Figure 3**). Transfer of the P-pant moiety onto the ACP domain could be observed in the presence of  $Mg^{2+}$  and, to a lesser extent, in the presence of  $Mn^{2+}$ . Surprisingly, activity of the *B. subtilis* Sfp was not detected in the presence of  $Mn^{2+}$ , a

result not in accordance with the work published by Mofid and colleagues<sup>26</sup>, although the reaction mixture was incubated at 37°C instead of 30°C in our assay.

Fluorescence levels of clones obtained from N-terminally truncated constructs are similar to the full-length enzyme (**Figure 2**), thus indicating that residues at the N-terminal end do not play a critical role on the stability of the enzyme. We then generated a structure-based sequence alignment between structural homologues, the *B. subtilis* Sfp (PDB code 1QR0,<sup>25</sup>) and human AcpS (PDB code 2c43,<sup>29</sup>), that we further extended to PptT with respect to the conservation of sequence motifs with key residues important for the activity and stability of CoA (**Figure 4a**). In the Sfp X-ray structure, N-terminal residues 2 to 7 form the outer most-strand of a three-stranded antiparallel  $\beta$ -sheet, at the opposite side of CoA (**Figure 4b**). This N-terminal  $\beta$ -strand also interacts with the 26-residue long C-terminal region through hydrogen bond interactions that involve the main-chain nitrogen and oxygen atoms of Ile3 and Gly5. Tyr4 also contributes to the stabilization of the adjacent helix through interactions with Arg57. Disrupting all these interactions would have a dramatic impact onto the stability of the enzyme. In the human AcpS structure, the N-terminal domain is mostly disordered and unstructured with the first residue identified at position 8 in the X-ray structure and strand  $\beta$ 1 starting at position 17 (**Figure 4c**). We then anticipated that the N-terminal region of PptT adopts an unstructured conformation similar to the one observed for the human AcpS, in accordance with the sequence alignment presented in Figure 4a. To the opposite, deletion of C-terminal residues has a dramatic effect on the solubility of PptT, as illustrated by the fluorescence drop to 75% for the truncated variant missing the C-terminus residue L227. In the Sfp and AcpS structures, this highly variable and poorly structured region wraps

around the  $\alpha/\beta$  core of the protein (**Figure 4b,c**). Stabilization occurs through numerous hydrophobic and van der Waals contacts and a constellation of hydrogen bond interactions that involve the main-chain and side-chain atoms of residues all along the C-terminal region. At this stage, considering the low level of amino-acid sequence similarity between the three enzymes, questions about the role of L227 onto the stability of PptT are difficult to answer as no NMR or X-ray structure is yet available. In the short run, we are confident that our efforts to overcome this major solubility bottleneck will help determining the three-dimensional structure of PptT. In the long run, the high-throughput screening of small molecule libraries will serve as a basis for the structure-based rational design of inhibitors with confirmed activity against *M. tuberculosis*.

## **Acknowledgments**

This work was supported by European Structural Funds (FEDER) and the Région Midi-Pyrénées (MYCA, grants 34249 and 09005193; TUBCAN, grant 11052698) and the Agence Nationale de la Recherche (XPKS-MYCO, grant 09-BLAN-0298-01)

## References

1. Buyske, D. A., Handshumacher, R. E., Schilling, E. D. & Strong, F. M. (1954). The Stability of Coenzyme A. *J. Am. Chem. Soc.* **76**, 3575-7.
2. Narayanan, A., Ridilla, M. & Yernool, D. A. (2011). Restrained expression, a method to overproduce toxic membrane proteins by exploiting operator-repressor interactions. *Protein science : a publication of the Protein Society* **20**, 51-61.
3. Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology* **260**, 289-98.
4. Ito, T. & Wagner, G. (2004). Using codon optimization, chaperone co-expression, and rational mutagenesis for production and NMR assignments of human eIF2 alpha. *Journal of biomolecular NMR* **28**, 357-67.
5. Chow, M. K., Amin, A. A., Fulton, K. F., Fernando, T., Kamau, L., Batty, C., Louca, M., Ho, S., Whisstock, J. C., Bottomley, S. P. & Buckle, A. M. (2006). The REFOLD database: a tool for the optimization of protein expression and refolding. *Nucleic acids research* **34**, D207-12.
6. Mokbel, K. M., Parris, C. N., Ghilchik, M., Amerasinghe, C. N. & Newbold, R. F. (2000). Telomerase activity and lymphovascular invasion in breast cancer. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* **26**, 30-3.
7. Pedelacq, J. D., Piltch, E., Liong, E. C., Berendzen, J., Kim, C. Y., Rho, B. S., Park, M. S., Terwilliger, T. C. & Waldo, G. S. (2002). Engineering soluble proteins for structural genomics. *Nature biotechnology* **20**, 927-32.
8. Yang, J. K., Park, M. S., Waldo, G. S. & Suh, S. W. (2003). Directed evolution approach to a structural genomics project: Rv2002 from *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 455-60.
9. Keenan, R. J., Siehl, D. L., Gorton, R. & Castle, L. A. (2005). DNA shuffling as a tool for protein crystallization. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8887-92.
10. Yumerefendi, H., Tarendeau, F., Mas, P. J. & Hart, D. J. (2010). ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *Journal of structural biology* **172**, 66-74.
11. Pedelacq, J. D., Nguyen, H. B., Cabantous, S., Mark, B. L., Listwan, P., Bell, C., Friedland, N., Lockard, M., Faille, A., Mourey, L., Terwilliger, T. C. & Waldo, G. S. (2011). Experimental mapping of soluble protein domains using a hierarchical approach. *Nucleic acids research* **39**, e125.
12. Hart, D. J. & Waldo, G. S. (2013). Library methods for structural biology of challenging proteins and their complexes. *Current opinion in structural biology*.
13. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature biotechnology* **23**, 102-7.
14. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C. & Walsh, C. T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & biology* **3**, 923-36.

15. Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C. & Guilhot, C. (2006). The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of Mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8511-6.
16. Chen, H., O'Connor, S., Cane, D. E. & Walsh, C. T. (2001). Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chemistry & biology* **8**, 899-912.
17. Leblanc, C., Prudhomme, T., Tabouret, G., Ray, A., Burbaud, S., Cabantous, S., Mourey, L., Guilhot, C. & Chalut, C. (2012). 4'-Phosphopantetheinyl Transferase PptT, a New Drug Target Required for Mycobacterium tuberculosis Growth and Persistence In Vivo. *PLoS pathogens* **8**, e1003097.
18. Cabantous, S. & Waldo, G. S. (2006). In vivo and in vitro protein solubility assays using split GFP. *Nature methods* **3**, 845-54.
19. Baba Moussa, L., Werner, S., Colin, D. A., Mourey, L., Pedelacq, J. D., Samama, J. P., Sanni, A., Monteil, H. & Prevost, G. (1999). Discoupling the Ca(2+)-activation from the pore-forming function of the bi-component Panton-Valentine leucocidin in human PMNs. *FEBS letters* **461**, 280-6.
20. Mossessova, E. & Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Molecular cell* **5**, 865-76.
21. Listwan, P., Pedelacq, J. D., Lockard, M., Bell, C., Terwilliger, T. C. & Waldo, G. S. (2010). The optimization of in vitro high-throughput chemical lysis of Escherichia coli. Application to ACP domain of the polyketide synthase ppsC from Mycobacterium tuberculosis. *Journal of structural and functional genomics* **11**, 41-9.
22. Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. (1998). Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chemistry & biology* **5**, 631-45.
23. Niesen, F. H., Berglund, H. & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2**, 2212-21.
24. Astarie-Dequeker, C., Le Guyader, L., Malaga, W., Seaphanh, F. K., Chalut, C., Lopez, A. & Guilhot, C. (2009). Phthiocerol dimycocerosates of M. tuberculosis participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. *PLoS pathogens* **5**, e1000289.
25. Reuter, K., Mofid, M. R., Marahiel, M. A. & Ficner, R. (1999). Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily. *The EMBO journal* **18**, 6823-31.
26. Mofid, M. R., Finking, R., Essen, L. O. & Marahiel, M. A. (2004). Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from Bacillus subtilis: carrier protein recognition and reaction mechanism. *Biochemistry* **43**, 4128-36.
27. Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J. & Somers, W. S. (2000). Crystal structures of substrate binding to

- Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883-95.
28. Mofid, M. R., Marahiel, M. A., Ficner, R. & Reuter, K. (1999). Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases. *Acta crystallographica. Section D, Biological crystallography* **55**, 1098-100.
  29. Bunkoczi, G., Pasta, S., Joshi, A., Wu, X., Kavanagh, K. L., Smith, S. & Oppermann, U. (2007). Mechanism and substrate recognition of human holo ACP synthase. *Chemistry & biology* **14**, 1243-53.
  30. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic acids research* **31**, 3320-3.

## Figure Legends

**Figure 1. SDS-PAGE of soluble (S) and pellet (P) fractions of PptT constructs of *E. coli* BL21(DE3) cell cultures carrying a pET28-PptT plasmid allowing expression of PptT (1) with no 6His tag, (2) with N-terminal and (3) C-terminal 6His tag, (4) with N-terminally fused SUMO and (5) MBP proteins. Molecular weight markers (Mw) are also indicated.**

**Figure 2. PCR-directed truncations of *pptT* and split-GFP complementation reaction.** (a) Schematic representation of the 48 constructs aligned onto the PptT amino-acid sequence. Fragments are organized in groups of 6 with identical N-terminal positions and decreasing C-terminal positions. (b) Solubility screen of *E. coli* BL21(DE3) cells expressing the corresponding fragments in fusion with GFP11 following sequential complementation with GFP1-10. Most soluble fragments correspond to the full-length and N-terminally truncated constructs. (c) Superimposition of progress curves for complementation of PptT variants as a function of the fluorescence intensity versus time in minutes.

**Figure 3. Activation time-course of the ACP domain by PptT and the Sfp from *Bacillus subtilis*.** (a) Schematic representation of the “domain trapping” strategy that integrates a DHFR filter to eliminate the out-of-frame constructs and a split-GFP solubility screen to identify those soluble ones. 450-850bp DNase I generated library of fragments was inserted at position 86/87 of the DHFR. Inverse PCR focused sub-libraries of fragments helped identifying a soluble fragment (indicated by an arrow) centered onto the ACP domain of the *M. tuberculosis* polyketide synthase PpsC. We used the color scheme gray - light gray - bright gray, to distinguish between the least and most soluble ACP fragments. (b) The transfer of the P-pant group from CoA to the ACP domain of PpsC is visualized on a 10% native polyacrylamide gel supplemented with 2.5 M urea in the presence of CaCl<sub>2</sub> (left), MnCl<sub>2</sub> (middle) and MgCl<sub>2</sub> (right) at different time intervals (5, 10, 15, 20, 30 and 60 min). Negative (-) and positive (+) controls correspond to the inactivated and activated forms of the ACP domain, respectively.

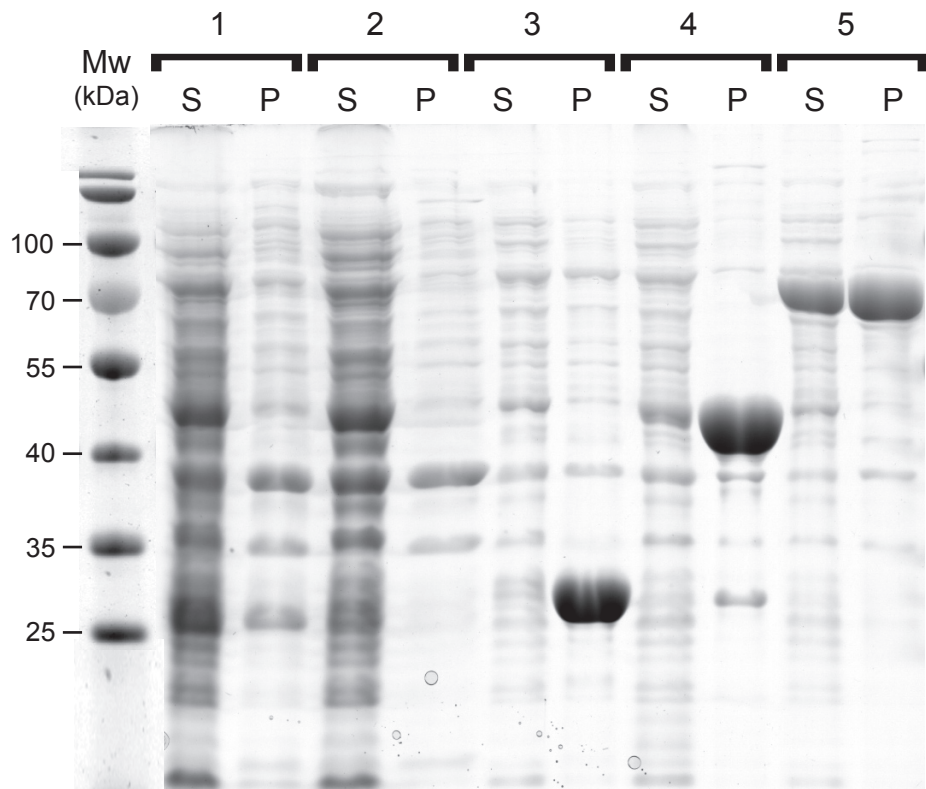
**Figure 4. Sequence alignment of PptT with known homologues.** (a) Structure-based sequence alignment of the *B. subtilis* Sfp (Uniprot code P39135) and the human AcpS (Uniprot code Q9NRN7) extended to PptT with respect to the conservation of sequence motifs with key residues important for the activity and interactions with CoA. Secondary structure elements deduced from the X-ray structure of Sfp (PDB code 1QR0) and the human AcpS (PDB code 2C43) are shown at the bottom of the aligned sequences in light and dark gray, respectively. Sequence similarities are highlighted in red; sequence identities are shown as white letters on a red background. Strictly conserved catalytic residues D114 and E157 are indicated with brown triangles. The figure was produced with ESPript<sup>30</sup>. Ribbon representation of the X-ray structure of the *B. subtilis* Sfp (b) and the human AcpS (c). CoA is shown in a ball-and-stick representation. N-terminal and C-terminal ends are colored in green and orange in the X-ray structures and the sequence alignment. The side-chains of residues Tyr4 and Arg57 are indicated.



**Supplementary Table 1. PCR primers used for *pptT* truncations with NdeI/BamHI restrictions sites labeled in bold and underlined.**

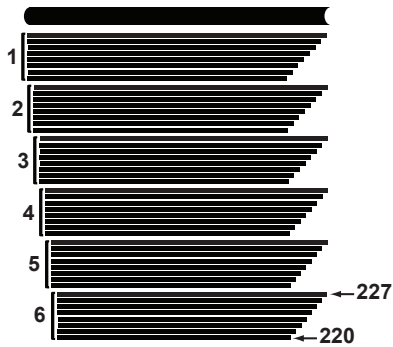
	# aa	# DNA	
Top primers	1	1	5'-GATATAC <b><u>CATATG</u></b> ACGGTAGGCACGCTGGTGGCGT-3'
	2	4	5'-GATATAC <b><u>CATATG</u></b> GTAGGCACGCTGGTGGCGTCGG-3'
	3	7	5'-GATATAC <b><u>CATATG</u></b> GTAGGCACGCTGGTGGCGTCGG-3'
	4	10	5'-GATATAC <b><u>CATATG</u></b> ACGCTGGTGGCGTCGGTGTGTC-3'
	5	13	5'-GATATAC <b><u>CATATG</u></b> CTGGTGGCGTCGGTGTGCCGG-3'
	6	16	5'-GATATAC <b><u>CATATG</u></b> GTGGCGTCGGTGTGCCGGCGA-3'
Bottom primers	219	660	5'-AATTC <b><u>GGATCCC</u></b> ACCAGTCCGCGCTCAACCGACC-3'
	220	663	5'-AATTC <b><u>GGATCCC</u></b> AGCACCAGTCCGCGCTCAACCG-3'
	221	666	5'-AATTC <b><u>GGATCC</u></b> GGTCAGCACCAGTCCGCGCTCAA-3'
	222	669	5'-AATTC <b><u>GGATCCC</u></b> GCGGTCAGCACCAGTCCGCGCT-3'
	223	672	5'-AATTC <b><u>GGATCC</u></b> GATCGCGGTCAGCACCAGTCCGC-3'
	224	675	5'-AATTC <b><u>GGATCCC</u></b> ACGATCGCGGTCAGCACCAGTC-3'

# Figure 1

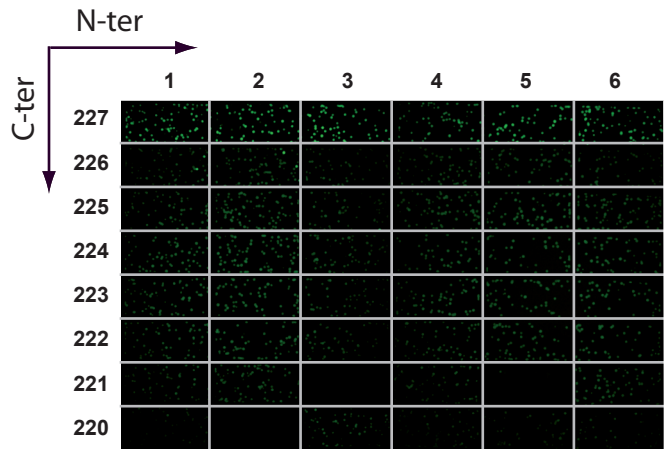


# Figure 2

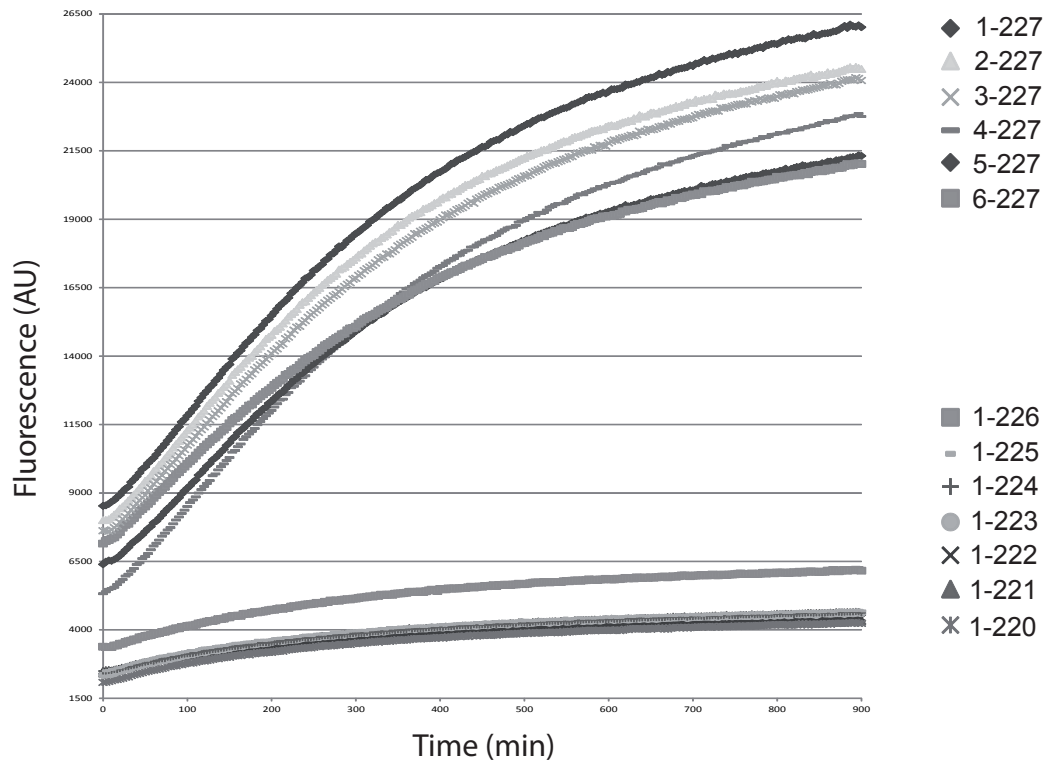
a



b



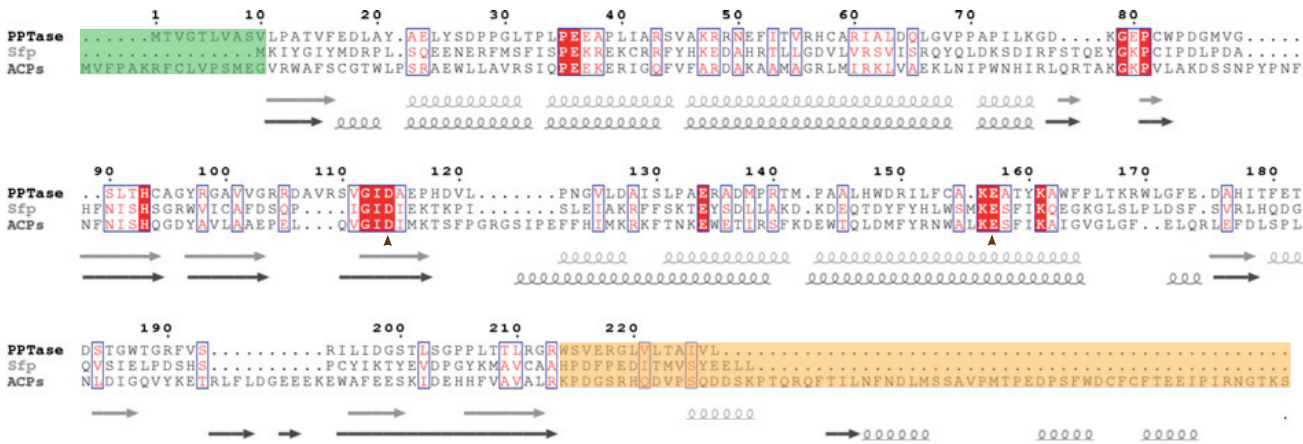
c



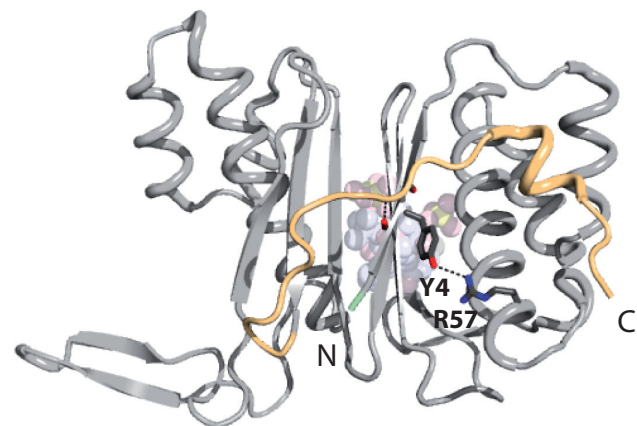


# Figure 4

a



b



c

