Crystallization and structure analysis of the core motif of the Pks13 acyltransferase domain from *Mycobacterium tuberculosis*

Mingjing Yu [1], **Chao Dou** [1], **Yijun Gu** [2], **Wei Cheng** Corresp. [1]

[1] Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China

[2] National Center for Protein Science, Shanghai, People's Republic of China

Corresponding Author: Wei Cheng
Email address: chwlab@163.com

Type I polyketide synthase 13 (Pks13) is involved in the final step of the biosynthesis of mycolic acid in Mycobacterium tuberculosis (Mtb). Recent articles have reported that Pks13 is an essential enzyme in the mycolic acid biosynthesis pathway, and it has been deeply studied as a drug target in TB. We report a high-resolution structure of the acyltransferase (AT) domain of Pks13 at 2.59 Å resolution. Structural comparison with the full-length AT domain (PDB code, 3TZW and 3TZZ) reveals a different orientation of the C-terminal helix and rearrangement of some conserved residues.

Crystallization and structure analysis of the core motif of the Pks13 acyltransferase domain from *Mycobacterium tuberculosis*

Mingjing Yu[a], Chao Dou[a], Yijun Gu[b] and Wei Cheng[a]*

a Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, 37 Guoxue Road, Chengdu, Sichuan, 610041, People's Republic of China

b National Center for Protein Science Shanghai China, People's Republic of China

*Correspondence email: cwlab@163.com

Abstract

Type I polyketide synthase 13 (Pks13) is involved in the final step of the biosynthesis of mycolic acid in Mycobacterium tuberculosis (Mtb). Recent articles have reported that Pks13 is an essential enzyme in the mycolic acid biosynthesis pathway, and it has been deeply studied as a drug target in TB. We report a high-resolution structure of the acyltransferase (AT) domain of Pks13 at 2.59 Å resolution. Structural comparison with the full-length AT domain (PDB code, 3TZW and 3TZZ) reveals a different orientation of the C-terminal helix and rearrangement of some conserved residues.

Introduction

Tuberculosis (TB) and its drug-resistant forms are still the primary causes of mortality, surpassing other infectious diseases*(Dande & Samant, 2018)* and emphasizing the unmet clinical need for new drugs with novel mechanisms**.** Owing to the indispensable and specific lipids forming the envelope of *Mycobacterium tuberculosis (Dubnau et al., 2000)*, targeting the synthesis and transport pathways of mycolic acids has always been the main route of TB drug discovery*(Bhatt et al., 2007; Brennan & Nikaido, 1995; North et al., 2014; Wilson et al., 2013)*.

Recently, powerful evidence has verified that Pks13 is an essential enzyme in the mycolic acid

biosynthesis pathway *(Gavalda et al., 2009; Portevin et al., 2004)*, and Pks13 has been extensively studied as a drug target for TB *(Aggarwal et al., 2017; Thanna et al., 2016)*. The type-1 polyketide synthase enzyme Pks13 consists of five domains. The medial three are mandatory polyketide synthase (PKS) domains, namely, the ketoacyl synthase (KS) domain, the acetyltransferase (AT) domain and the acyl carrier protein (ACP) domain. The other ACP domain is adjacent to the KS domain, and the thioesterase (TE) domain is the C-terminal portion of Pks13. The overall Pks13 topological structure has the order ACP-KS-AT-ACP-TE (**Fig. 1A**).

The residue Ser^{55} in the N-ACP domain has been identified as a very important active site for initializing the pathway. The sfp gene encodes Phosphopantetheinyl transferase (PPTase), which modifies ACPs by providing a P-pant arm for the general function of carrying the substrate acyl chain via a thioester bond involving its terminal thiol group*(Chalut et al., 2006; Gavalda et al., 2009; Wilson et al., 2013)*. The meromycoloyl chain on the N-ACP domain is transferred to the KS domain, and the intermediate product α-alkyl β-ketothioester is produced by a Claisen-type condensation reaction with another substrate, the carboxyacyl-CoA loaded by the AT domain. The mycolic acid precursor generated by the C-terminal ACP domain is then released by the TE domain *(Abrahams & Besra, 2016; Dubey et al., 2002)*.

Despite increasing insights into the mechanism of Pks13, no full-length structural information has been reported, except that the structures of a few domains belonging to Pks13 have been solved *(Bergeret et al., 2012; Herbst et al., 2016)*.

Here, we report a high-resolution structure of the core motif of the AT domain. First, the full-length Pks13 protein was successfully purified, and an extended crystal screening was performed, in which the initial crystal was obtained. While attempting to phase the diffraction data of the crystal, we found that the crystalized protein suggested a degraded fragment. Then, the crystals

were solved, and the N-terminal sequence was identified by mass spectrometry, the results of which were in line with the phase presented by the Se-Met crystal dataset. These results indicated that the crystalized protein was actually proteolyzed to become a fragment (Leu^{717} to Arg^{826}). The overall crystal structure displayed a fold similar to the reported AT domain, excluding several conformational changes relative to the reported AT domain (Protein Data Bank codes: 3TZW, 3TZZ). The structural alignment performed by the secondary structure matching (SSM) in Coot also showed a superimposition of the core motif and the AT domain with an r.m.s.d. of 1.33 Å, which was mainly attributed to the rearrangement of residues Ala^{796}-Ser^{801}. In addition, the position of residue Ser^{801} that is reported to be the catalytic residue was shifted away from the active site *(Bergeret et al., 2012; Gavalda et al., 2009)*. Furthermore, a highly conserved arginine residue, Arg^{826}, lost a hydrogen bond with the side chain of Gln^{773}, as observed in our structure. These features might all contribute to the unique state that survived proteolysis.

We believe that comprehensive structural studies of Pks13 will pave the way for structure-based antimycobacterial drug design and drug screening.

Materials & Methods

Cloning, Over-expression, and Purification

The codon-optimized gene encoding the full-length Pks13 protein originating from *Mycobacterium tuberculosis* was ligated into the *Nde* l and *Xho* l sites of the pET-28b expression plasmid (Novagen, Madison, WI, USA). The sfp gene, which encodes the P-pant transferase that serves as a kind of cofactor to modify Ser^{55} in the N-ACP domain of Pks13, from *Bacillus subtilis* str.168*(Chalut et al., 2006)* was also ligated into the *Nde* l and *Xho* l sites of the pET-21b expression plasmid (Novagen, Madison), and a terminator codon was added to the C-terminal end.

The detailed information on these constructs is shown in **Table 1**. All constructed plasmids were verified by sequencing.

The constructed plasmid pks13-pET-28b was cotransformed with sfp-pET-21b into *E. coli* strain BL21 (DE3). The bacteria containing these recombinant plasmids were grown at 310 K in M9 medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, and 0.4% glucose) supplemented with 0.05 g/L kanamycin and 0.1 g/L ampicillin. When the OD600 reached 0.5, the medium was supplemented with amino acids (0.1 g/L L-lysine, L-phenylalanine, and L-threonine; 0.05 g/L L-isoleucine, L-leucine, and L-valine; and 0.1 g/L L-Se-methionine). In addition, the protein was overexpressed after the addition of 0.3 mM IPTG at 289 K for approximately 16 h. Cell pellets were harvested by 4,000 rpm centrifugation for 10 min and suspended in a solution of 1 mM PMSF, 150 mM NaCl, and 25 mM Tris/HCl (pH 8.0) suspension buffer. After sonication, we clarified the cell lysate by centrifugation at 15,000 g for 30 min. The supernatant containing the modified protein was applied to a nickel-affinity column (Ni-NTA; GE Healthcare) preequilibrated with suspension buffer.

The resin was gradient washed with ice-cold washing buffer (25 mM Tris/HCl (pH 8.0) and 150 mM NaCl) containing 20, 30, and 40 mM imidazole, and the proteins were eluted with elution buffer (25 mM Tris/HCl pH 8.0, 150 mM NaCl, and 250 mM imidazole). Before loading onto an anion exchange column (Source Q; GE Healthcare), the eluate with 250 mM imidazole was diluted by half with buffer A (25 mM Tris/HCl (pH 8.0) and 3 mM DTT). Subsequently, the peak fractions were collected for further purification by size-exclusion chromatography (Superdex 200 10/300; GE Healthcare) in 10 mM Tris/HCl (pH 8.0) buffer containing 100 mM NaCl. The purity of the protein was determined by 12% SDS-PAGE gels stained by Coomassie brilliant blue. The eluted

protein was concentrated by a 10 kDa centrifugal filter and flash-frozen in liquid nitrogen for crystallization.

Crystallization

The protein encoded by the constructed plasmid and labeled with Se-Met was concentrated to 12 mg/ml. Index (Hampton Research) and PEG/ION (Hampton Research) kits were used for the initial crystallization trials at 293 K by the sitting-drop vapor-diffusion method*(Luft & Detitta, 1995)*. Each drop contained 1 μL of protein solution and an equal volume of reservoir solution.

The initial crystal was obtained from a solution of 300 mM KAc, pH 8.1 and 20% PEG 3,350. Further crystal optimization experiments were performed by systematic variation of the precipitant concentration. Ultimately, the best crystals were screened in a solution consisting of 300 mM KAc, pH 8.1, and 25% PEG 3,350. The crystals grew to full size in 10 days and were flash-frozen in liquid nitrogen with 10% glycerol added as a cryoprotectant before X-ray diffraction.

Data collection

X-ray diffraction data were collected at 100 K using a Pilatus3 6M detector. All the datasets were obtained at beamline BL19U1 of the Synchrotron Radiation Facility in Shanghai*(Wang et al., 2016)*. A total of 360 images were recorded with 0.5 s exposure at a crystal-to-detector distance of 450 mm, and a total rotation range of 360° was covered using 1.0 oscillation.

Protein N-termial sequence based on mass spectrometry

Regarding the dataset of the crystalized pks13, the initial trial did not seem to provide a structure with all of the residues because of the insufficient density for many residues. After X-ray diffraction, the crystals were collected together and analyzed with SDS-PAGE gels stained by Coomassie brilliant blue. The gel with a single low molecular line was processed with the standard

in-gel digestion for mass spectrometric characterization to identify the actual location of the degraded fragment in Pks13*(Shevchenko et al., 2006)*.

Data refinement

All datasets were processed by HKL-2000*(Brodersen et al., 2006)*. The crystal structure of the motif was solved by single-wavelength anomalous dispersion (SAD) phasing using the anomalous data collected from the Se-Met crystal. The final model was manually built in Coot*(Emsley et al., 2010)* and refined in PHENIX*(Adams et al., 2010)*. The final models were validated by MolProbity and deposited in the Protein Data Bank (PDB code 5XUO).

Results

Purification and crystallization of Pks13

The full-length Pks13 protein was successfully overexpressed in *E. coli* BL21 (DE3), and the initial crystal condition (300 mM KAc, pH 8.1 and 20% PEG 3,350) was screened. The mature lump-like crystals were optimized after a series of crystal optimization experiments, including crystallization with different detergents and additives.

Data collection

X-ray diffraction datasets for the Se-Met-labeled crystals were obtained at beamline BL19U1 of the Synchrotron Radiation Facility in Shanghai with a wavelength of 0.97852 Å. Diffraction images for the crystals were processed using HKL-2000.

Protein N-terminal sequence

The prepared gel was digested by trypsin, and the digestion was purified into freeze-dried peptide powder. Then, the peptide was resolved by an Orbitrap Elite LC-MS/MS for analysis. The sequenced peptides were blasted within the full-length pks13 protein, and the crystalized fragment protein was located in the range from Ala^{717} to Arg^{826} (**Table 2**).

Data refinement

The crystal belonged to the space group R32, with asymmetric unit cell parameters of a=93.694, b=93.694, c=97.908, α=β=90, and γ=120. Additionally, the phases were determined by the SAD method. The final model was manually built in Coot and refined in PHENIX to an R_{free} of 26.05% with good stereochemistry. The collected and processed data are presented in **Table 3**.

Overall architecture and Superimposition with AT domain

The overall structure of the core motif contains a long α helix, five short α helixes and two short η turns, in the order of α1-α2-α3-α4-η1-η2-α5-α6, which constitutes a compact motif (Fig. 1B). The long α helix, α4, distributes in the middle and is surrounded by the other five short α helixes and two short η turns (Fig. 2A). Superimposition with the reported structure of the AT domain (PDB code 3TZZ)*(Bergeret et al., 2012)* suggested that the core motif was located in the central region of the AT domain (Fig. 2B). The crystallized core motif ranging from Leu^{717} to Arg^{826} represents approximately one-third of the AT domain, and the overall crystal structure displays a fold similar to the reported AT domain (Fig. 2C).

Although sequence alignment showed 100% identity between the core motif and the AT domain, the secondary structure elements presented a slight conformational change from residues Ala^{796} to Ser^{801}, for which refinement indicated two η turns instead of the β strand highlighted by red dashed square line (Fig. 3A). According to the structure of the AT domain reported by Bergeret et al., there was a parallel six-stranded β-sheet (β13-β12-β4-β5-β10-β11) along with the active site in the reported AT domain, while only the central β strand, β5, was presented in the motif structure and was refined as a completely different secondary element (Fig. 3B, 3C). Previous studies suggested that the conserved Ser^{801} and Arg^{826} could serve as a catalytic residue and binding site, respectively. The active site Ser^{801} of the AT domain is located in the nucleophilic elbow between

β5 and helix α10 and could directly contact the lipid substrate. Additionally, the active cite constituted the part of the highly conserved consensus sequence Gly-X-Ser-X-Gly that stabilizes the β5 strand shape*(Bergeret et al., 2012; Serre et al., 1995)*. In our work, the topographic conformation of Ala^{796} to Ser^{801} was transformed into two relatively disordered η turns, along with the conformational change of the position of Ser^{801} dislocating from the substrate. Furthermore, the side chain of the binding site Arg^{826} was also stretched to the reverse side of α5 and lost its interaction with Gln^{773}. However, this side chain formed direct hydrogen bonds with the negative side chain of the lipid substrate, and the conformation was held in position through a strong hydrogen bond interaction with the side chain of Gln^{773} in the AT domain (Fig. 4A).

The structural alignment performed by SSM in Coot*(Emsley & Cowtan, 2004)* also showed that the superimposition of the core motif and the AT domain had an r.m.s.d. of 1.33 Å; along with the conformational changes, this alignment might also suggest a more compact crystal packing state than that of the AT domain. According to a close view of the superimposition, some particular apolar contacts among α1 (His^{723}, Leu^{730}) and the long α4 (Gln^{773}, Ile^{779}, Gln^{780}, and Leu^{783}) and α5 (Ile^{823}) residues all contribute to the stabilization of the unique state (Fig. 4B). Electrostatic calculations of the AT domain (Protein Data Bank code 3TZZ) revealed the presence of an electropositive area corresponding to the floor of the active site cavity due to the presence of Ser^{801} and Arg^{826} (Fig. 4C). Comparison of the electrostatic potential surface presentation of the motif indicated that the surface of the active site cavity was transformed to an electronegative state (Fig. 4D)

Discussion

The synthesis and transport pathways of mycolic acids in *Mycobacterium tuberculosis* have always been a critical drug target. These mycolic acids serve as the primary defense to counteract

185 the low permeability of the envelop to many hydrophilic molecules. Many biochemical and
186 structural studies have sought to elucidate the participation of Pks13 in the synthesis of the lipid
187 complex. Obtaining the structure of Pks13 is of great significance in drug screening, as many
188 inhibitors have been reported to target Pks13 or its individual domains.

189 The structure of the fragment of the AT domain provides a relatively new perspective of a unique
190 state that can evade proteolysis. We have determined the 2.59 Å high-resolution crystal structure
191 of a partial AT domain from the *Mycobacterium tuberculosis* Pks13 protein. The overall structure
192 of the core motif of the AT domain is similar to the corresponding part of the reported AT domain,
193 with slight conformational differences. Some conserved residues showed a completely different
194 secondary structure. Residues Ala^{796}, Val^{797}, Ile^{798}, Gly^{799}, Gln^{800}, and Ser^{801} formed a β strand in
195 the previously reported AT domain (PDB codes 3TZW and 3TZZ), which instead refined as a
196 flexible loop conformation in the motif structure. In contrast to the typical structure of the whole
197 AT domain containing a palm-shaped parallel six-stranded β sheet, in which β5 is located in the
198 middle of a connection with the other five β strands. In our work, the β-sheet structure was
199 disrupted along with loosing connections among these β strands due to the conformational
200 changes. Actually, there was less possibility of the AT domain remaining the same because of the
201 conformational changes from Ala^{796} to Ser^{801}, which tend to confirm the speculation that the
202 conformational changes are a tactic to evade proteolysis. With the structural alignment performed
203 by SSM in Coot, the superimposition of the core motif and the AT domain shows an r.m.s.d. of
204 1.33 Å. The novel packed structure formed by these bundles seems tighter than the AT domain,
205 which is especially reflected in the apolar contacts among α1 (His^{723}, Leu^{730}) and the long α4
206 (Gln^{773}, Ile^{779}, Gln^{780}, and Leu^{783}) and α5 (Ile^{823}) residues. These apolar contacts among the

residues might strengthen the interactions of α4 with other helixes to form a more stable packing state.

Additionally, the active site Ser^{801}, which plays a critical role in catalytic activity, was dislocated away from the substrate cavity to the inner position of the core motif. The nucleophilic elbow of α10 and β5 also transformed from an electropositive state to an electronegative state which indicates an unsuitable state to absorb a substrate. In summary, the conformational change of residues from Ala^{796} to Ser^{801} and the rearrangement of residues Gln^{773}, Ser^{801} and Arg^{826} might all suggest that the degraded fragment formed a unique crystal packing state to survive proteolysis. In other words, the fragment forms a relatively stable state in contrast to the AT domain in such conditions. This work might provide new insight into the core motif of the AT domain. Our work also provides a structural basis for protein engineering.

However, the overall structure of Pks13 is still unrevealed, and its mechanism is unknown yet. More work should be performed, and we hope that our present work will provide some assistance.

Acknowledgments: We gratefully acknowledge the assistance of the staff of BL19U1 at the Shanghai Synchrotron Radiation Facility (SSRF) for their assistance with X-ray diffraction data collection.

References

Abrahams KA, and Besra GS. 2016. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. *Parasitology*:1-18.

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, and Grossekunstleve RW. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**:213-221.

Aggarwal A, Parai MK, Shetty N, Wallis D, Woolhiser L, Hastings C, Dutta NK, Galaviz S, Dhakal RC, and Shrestha R. 2017. Development of a Novel Lead that Targets M. tuberculosis Polyketide Synthase 13. *Cell* **170**:249-259.

Bergeret F, Gavalda S, Chalut C, Malaga W, Quémard A, Pedelacq JD, Daffé M, Guilhot C, Mourey L, and Bon C. 2012. Biochemical and structural study of the atypical acyltransferase

domain from the mycobacterial polyketide synthase Pks13. *Journal of Biological Chemistry* **287**:33675-33690.

Bhatt A, Molle V, Besra GS, Jr JW, and Kremer L. 2007. The Mycobacterium tuberculosis FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Molecular Microbiology* **64**:1442–1454.

Brennan PJ, and Nikaido H. 1995. The Envelope of Mycobacteria. *Annual Review of Biochemistry* **64**:29-63.

Brodersen DE, La Fortelle ED, Vonrhein C, Bricogne G, Nyborg J, and Kjeldgaard M. 2006. Applications of single-wavelength anomalous dispersion at high and atomic resolution. *Acta Crystallographica* **56**:431-441.

Chalut C, Botella L, De S-DAC, Houssin C, and Guilhot C. 2006. The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of Mycobacteria. *Proc Natl Acad Sci U S A* **103**:8511-8516.

Dande P, and Samant P. 2018. Acquaintance to Artificial Neural Networks and use of artificial intelligence as a diagnostic tool for tuberculosis: A review. *Tuberculosis* **108**:1.

Dubey VS, Sirakova TD, and Kolattukudy PE. 2002. Disruption of msl3 abolishes the synthesis of mycolipanoic and mycolipenic acids required for polyacyltrehalose synthesis in Mycobacterium tuberculosis H37Rv and causes cell aggregation. *Molecular Microbiology* **45**:1451-1459.

Dubnau E, Chan J, Raynaud C, Mohan VP, Lanéelle MA, Yu K, Quémard A, Smith I, and Daffé M. 2000. Oxygenated mycolic acids are necessary for virulence of Mycobacterium tuberculosis in mice. *Molecular Microbiology* **36**:630-637.

Emsley P, and Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica* **60**:2126-2132.

Emsley P, Lohkamp B, Scott WG, and Cowtan K. 2010. Features and development of Coot. *Acta Crystallographica* **66**:486-501.

Gavalda S, Léger M, Rest BVD, Stella A, Bardou F, Montrozier H, Chalut C, Burletschiltz O, Marrakchi H, and Daffé M. 2009. The Pks13/FadD32 Crosstalk for the Biosynthesis of Mycolic Acids in Mycobacterium tuberculosis. *Journal of Biological Chemistry* **284**:19255-19264.

Herbst DA, Jakob RP, Zähringer F, and Maier T. 2016. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **531**:533.

Luft JR, and Detitta GT. 1995. Chaperone salts, polyethylene glycol and rates of equilibration in vapor-diffusion crystallization. *Acta Crystallographica* **51**:780–785.

North EJ, Jackson M, and Lee RE. 2014. New Approaches to Target the Mycolic Acid Biosynthesis Pathway for the Development of Tuberculosis Therapeutics. *Current Pharmaceutical Design* **20**:-.

Portevin D, De SAC, Houssin C, Grimaldi C, Chami M, Daffé M, and Guilhot C. 2004. A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci U S A* **101**:314-319.

Serre L, Verbree EC, Dauter Z, Stuitje AR, and Derewenda ZS. 1995. The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component. *Journal of Biological Chemistry* **270**:12961-12964.

Shevchenko A, Tomas H, Havli J, Olsen JV, and Mann M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1**:2856-2860.

Thanna S, Knudson SE, Grzegorzewicz A, Kapil S, Goins CM, Ronning DR, Jackson M, Slayden RA, and Sucheck SJ. 2016. Synthesis and evaluation of new 2-aminothiophenes against Mycobacterium tuberculosis. *Organic & Biomolecular Chemistry* **14**:6119.

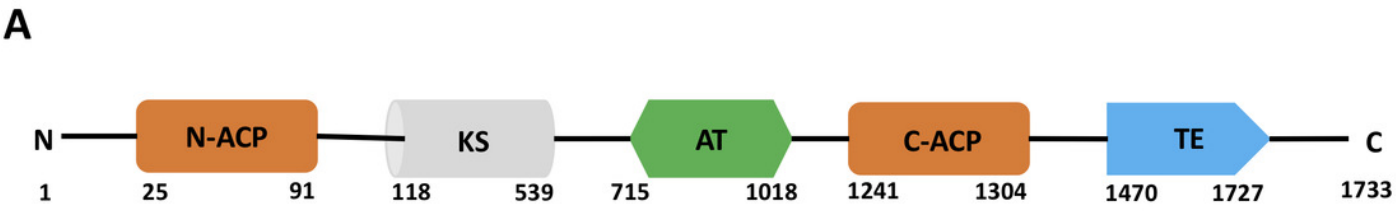
Wang Z, Pan Q, Yang L, Zhou H, Xu C, Yu F, Wang Q, Huang S, and He J. 2016. Automatic crystal centring procedure at the SSRF macromolecular crystallography beamline. *Journal of Synchrotron Radiation* **23**:1323–1332.

Wilson R, Kumar P, Parashar V, Vilchèze C, Veyronchurlet R, Freundlich JS, Barnes SW, Walker JR, Szymonifka MJ, and Marchiano E. 2013. Antituberculosis thiophenes define a requirement for Pks13 in mycolic acid biosynthesis. *Nature Chemical Biology* **9**:499.

Figure 1

The overall Pks13 domain structure has the order ACP-KS-AT-ACP-TE.

A. The leading domain N-ACP (25-91) is colored orange. The medial three domains are mandatory PKS domains, including the KS domain (118-539), the AT domain (715-1018) and the C-ACP domain (1241-1304), colored cyan, green and orange, respectively. The TE domain (1470-1727) is located in the C terminus and colored blue. Residue numbers are given below for each domain boundary. B. The motif resides in the AT domain ranging from Ala717 to Arg826. The whole topographic structure is composed of 6 α helixes and 2 short η turns, in the order of N terminus-α1-α2-α3-α4-η1-η2-α5-α6-C terminus.



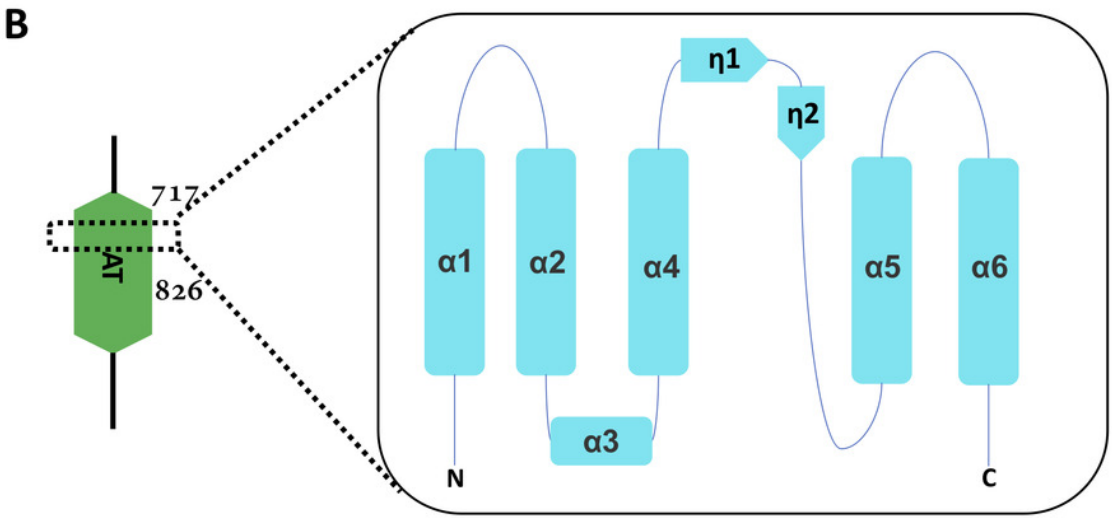


Figure 2

The architecture of the core motif from the AT domain and structural comparison with the AT domain (PDB code 3TZW)

A. The overall topological structure of the core motif contains a long α helix in the middle, five short α helixes and two short η turns around the long helix, which constitutes a compact motif. The top and bottom views with a rotation of 90° are exhibited on the right. B. The structures of the motif and the AT domain (PDB code 3TZZ) are superimposed together and colored blue and gray, respectively. The aligned region is zoomed in for clear observation. C. The superimposition of the motif and the region of the AT domain that could be aligned is shown in two orthogonal views. The secondary elements are labeled in the picture.

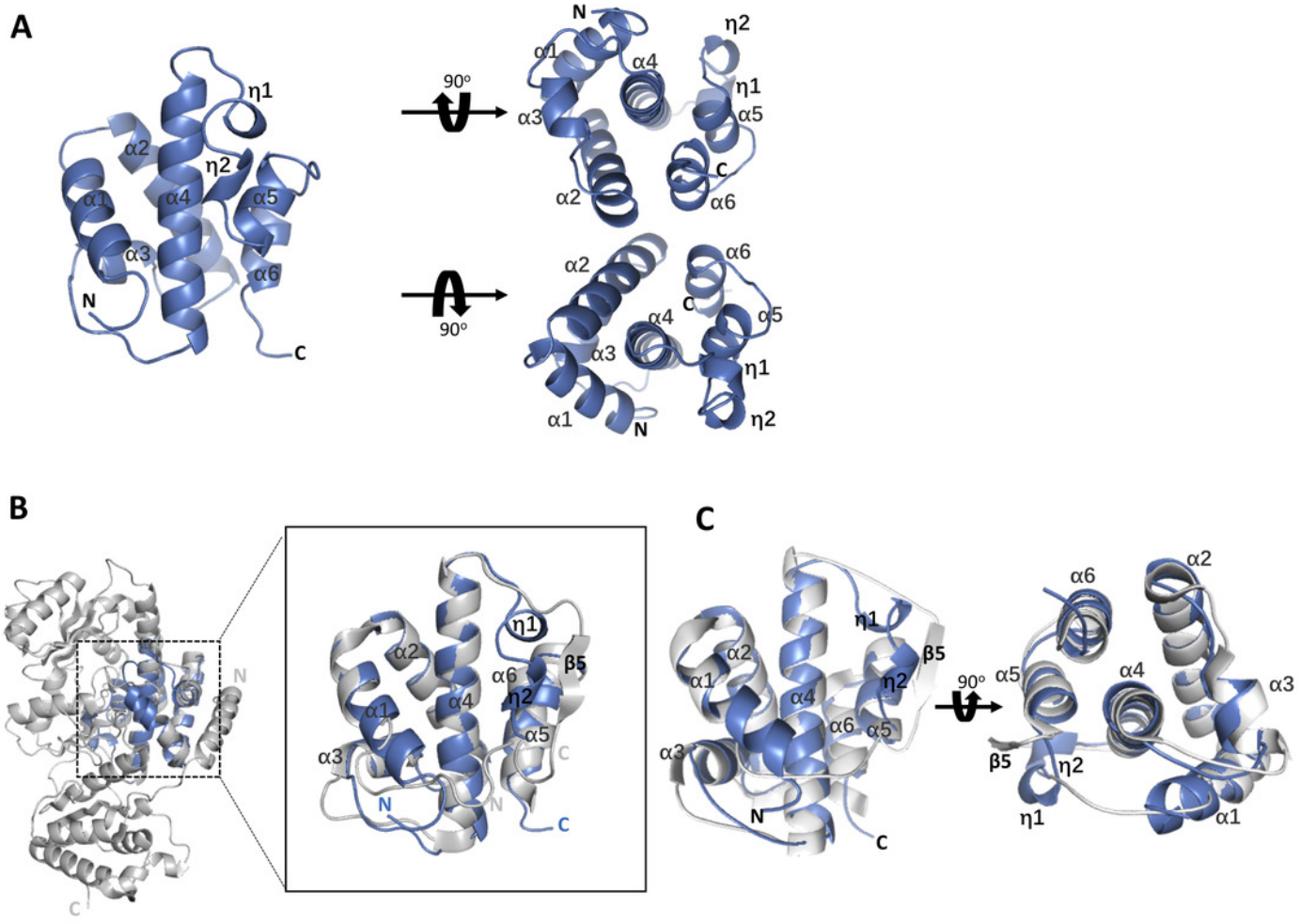


Figure 3

Structural and sequence alignments.

A. Structure-based sequence alignment with the whole AT domain. The secondary structural elements of the motif are given along the top of the alignment; the secondary elements of the AT domain (Protein Data Bank codes 3TZZ and 3TZW) are shown below. The difference between the structures is circled by the red dashed line. B. The six β-sheets (β13-β12-β4-β5-β10-β11) (Protein Data Bank codes 3TZW and 3TZZ) are presented in the figure, and β5 is aligned with the η turns of the motif. C. Some conserved residues show a totally different topographic structure, and residues Lys793, Pro794, Ala795, Ala796, Val797, Ile798, Gly799, Gln800, and Ser801 are shown as sticks.

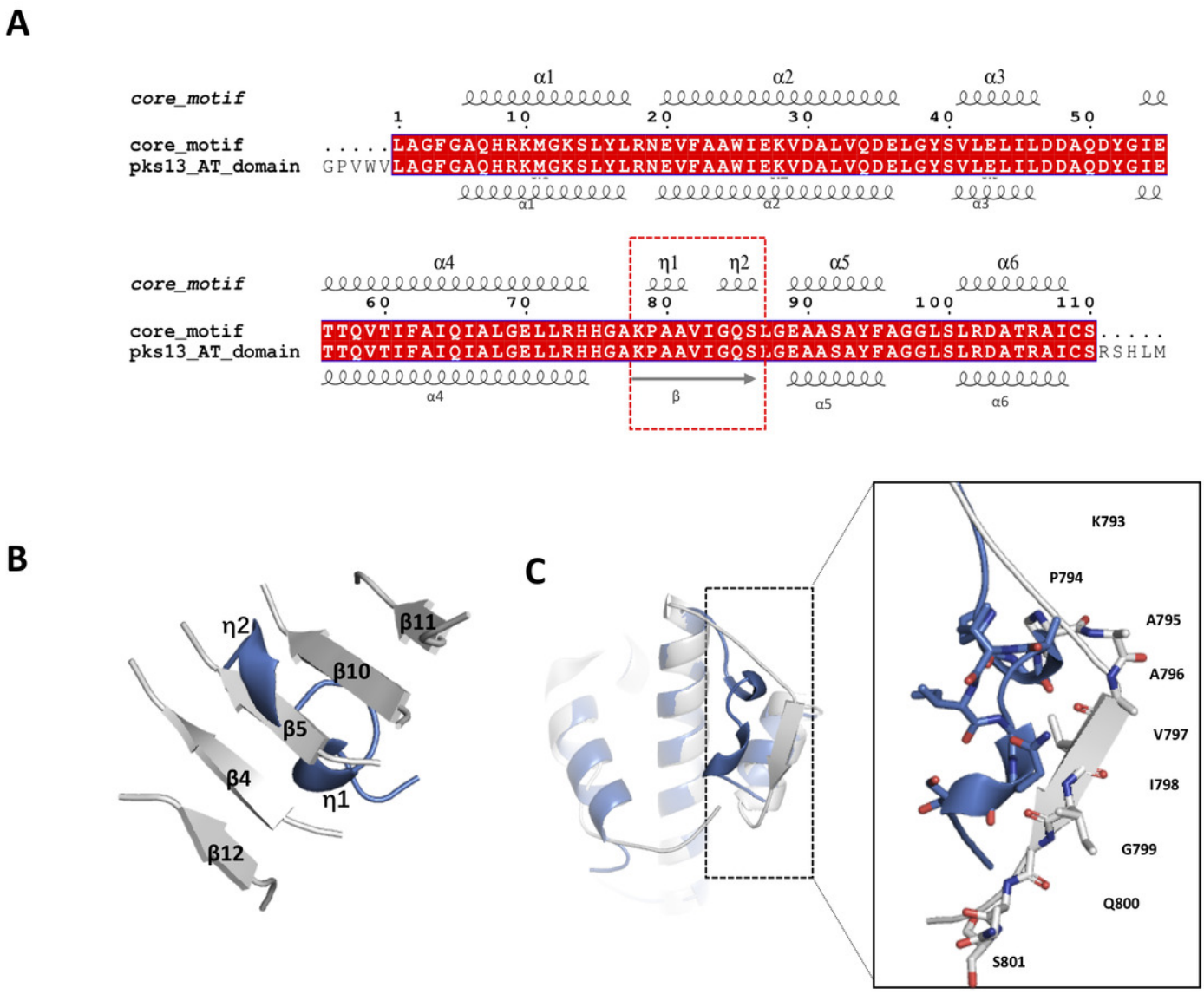

A
core_motif
α1
α2
α3
1
10
20
30
40
50
core_motif .....LAGFGAQHRKMGKSLYLRNEVFAAWIEKVDALVQDELGYSVLELILDDAQDYGIE
pks13_AT_domain GPVWVLAGFGAQHRKMGKSLYLRNEVFAAWIEKVDALVQDELGYSVLELILDDAQDYGIE
α1
α2
α3
core_motif
α4
η1
η2
α5
α6
60
70
80
90
100
110
core_motif TTQVTIFAIQIALGELLRHHGAKPAAVIGQSLGEAASAYFAGGLSLRDATRAICS.....
pks13_AT_domain TTQVTIFAIQIALGELLRHHGAKPAAVIGQSLGEAASAYFAGGLSLRDATRAICSRSHLM
α4
β
α5
α6
B
η2
β11
β10
β5
β4
η1
β12
C
K793
P794
A795
A796
V797
I798
G799
Q800
S801

Figure 4

Structural comparison between different states of AT domain.

A. Detailed description of the active site of the motif compared with the AT domain. The nucleophilic elbow comprising strand β5 and helix α10 in the AT domain corresponding to helix α5 and two helical η turns, respectively, is shown in cartoon representation. Important residues defining the active site are shown and labeled. Hydrogen bonds are represented by orange dotted lines. The lipid substrate colored orange is shown as sticks. B. Some apolar contacts among α1 (His723, Leu730) and the long α4 (Gln773, Ile779, Gln780, Leu783) and α5 (Ile823) residues are shown and labeled. C. Electrostatics calculations for the AT domain (Protein Data Bank code 3TZZ) revealed the presence of an electropositive area corresponding to the floor of the active site cavity bound with a lipid substrate. The surface representation was generated by PyMOL and colored according to its electrostatic potential (positive potential, blue; negative potential, red). The substrate cavity was highlighted by a dotted red circle and zoomed at the right panel. D. Electrostatics calculations for the motif in this work revealed the electrostatic potential transformation from an electropositive state to an electronegative state. The substrate originating from the AT domain (Protein Data Bank code 3TZZ) was docked on the floor of the catalytic cavity and was highlighted with a red dotted circle.

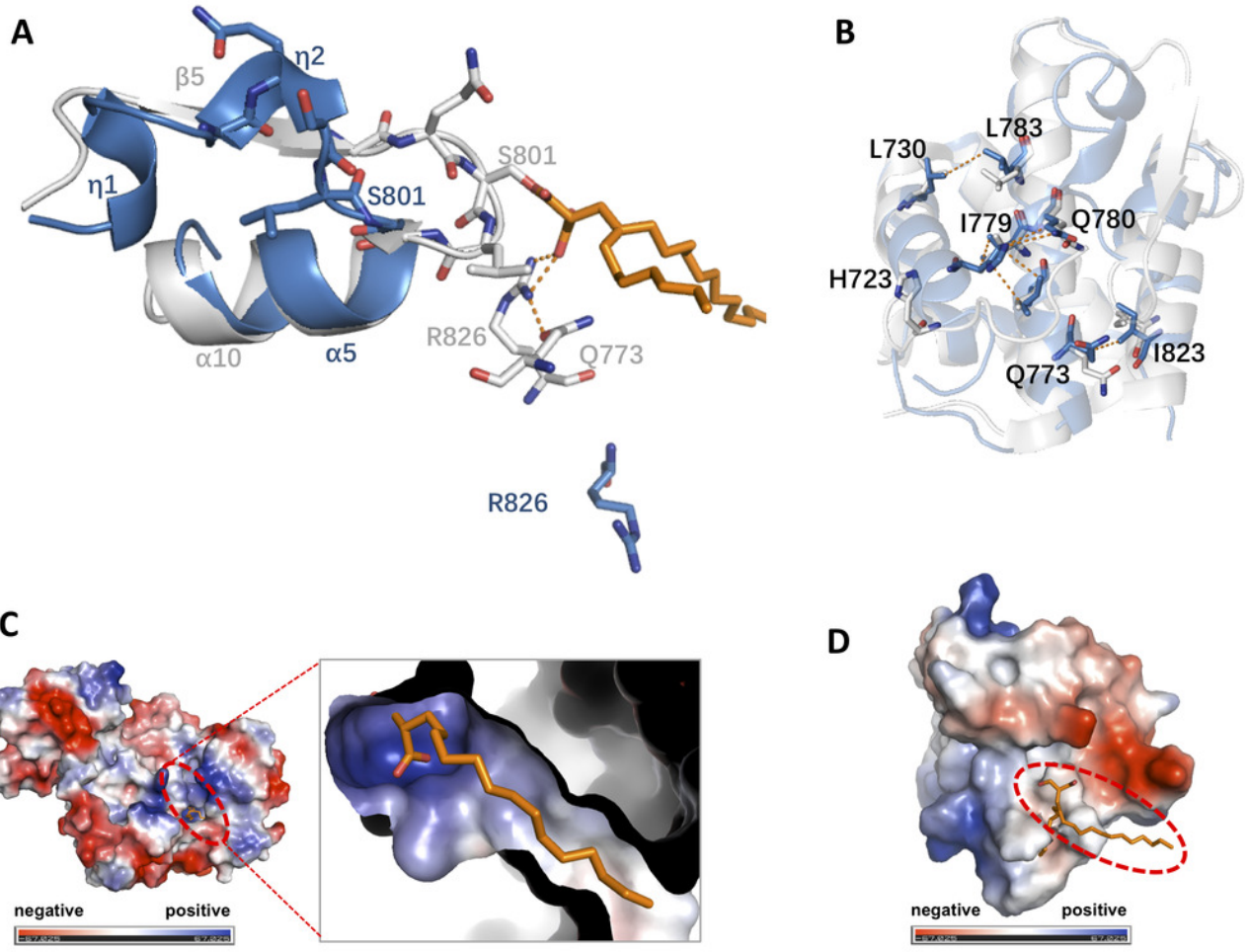

A
β5
η2
η1
S801
S801
α10
α5
R826
Q773
R826
B
L730
L783
I779
Q780
H723
Q773
I823
C
negative
positive
D
negative
positive

Table 1(on next page)

Macromolecule production information

1

2 Table 1 Macromolecule production information

Source organism	*Mycobacterium tuberculosis*(H37Rv)	*Bacillus subtilis str.168*
DNA source	Full-length Pks13	Sfp (P-pant transferase)
Forward primer	5-ggaattccatatgatggcagatgtggccg-3	5-ggaattccatatgaagatttacggaa -3
Reverse primer	5-ccgctcgagctgtttaccaacctcg-3	5-ccgctcgagtcaagcggaagcgata-3
Cloning vector	pET-28b	pET-21b
Expression vector	pET-28b	pET-21b
Expression host	*E. coli* strain(DE3)	*E. coli* strain(DE3)
Complete amino acid sequence of the construct produced	MADVAESQENAPAERA……IEADRTSEV GKQLE	MKIYGIYMDRPLSQEENERF MSFISPEKREKCR……PGYK MAVCAAHPDFPEDITMVSY EELL

3

4

Table 2(on next page)

Mass spectrum based on protein N-terminal sequencing

PEP, Posterior Error Probability of the identification. This value essentially operates as a p-value, where smaller is better.

Table 2 Mass spectrum based on protein N-terminal sequencing

Sequence	Length	Mass	Charges	PEP	Score
AGFGAQHR	8	842.9	2;3	0	201.07
AGFGAQHRK	9	971.07	3	0.020363	8.308
HHGAKPAAVIGQSLGEAASAYFAGGLSLR	29	2835.478	2;3;4;5	2.88E-48	83.418
HHGAKPAAVIGQSLGEAASAYFAGGLSLRDATR	33	3278.6909	3;4;5	0.012023	14.817
KMGKSLYLR	9	1094.627	2;3	0.0011449	41.427
MGKSLYLR	8	966.53207	2	5.16E-09	75.109
MGKSLYLRNEVFAAWIEK	18	2154.1296	3;4	1.24E-11	40.856
NEVFAAWIEK	10	1205.6081	2	8.84E-19	89.08
PAAVIGQSLGEAASAYFAGGLSLR	24	2305.2066	2;3	1.27E-303	269.26
PAAVIGQSLGEAASAYFAGGLSLRDATR	28	2748.4195	3	6.29E-159	160
SLYLRNEVFAAWIEK	15	1837.9727	2;3	7.27E-68	127.02
SSGLVPR	7	714.40244	2	3.61E-08	74.191

Table 3(on next page)

X-ray data collection and refinement statistics

1

2

3 Table 3. X-ray data collection and refinement statistics.

Data Set	Core motif of AT domain
Data collection	
X-ray source	SSRF BEAMLINE BL19U1
Space group	R32
Wavelength(Å)	0.97852
Resolution Range(Å)	50-2.587
Total No. of reflections	90229
No. of unique reflections	5358(524)
Completeness (%)	99.70
Redundancy	16.999
$R_{r.i.m.}$	0.026
$I/\sigma(I)$	41.27(2.10)
Refinement	
Resolution range	50-2.59
Reflections: working/test	5083/276
Final R_{cryst}	23.69%

Final R_{free}	26.05%
rotamer outliers	2.4%
Ramachandran plot:	
favored/allowed/outliers (%)	93.58/4.59/1.83
Rmsd bonds(Å)	0.003
Rmsd angles (°)	0.540
PDB accession code	5XUO