Acta Biochim Biophys Sin 2022, 54(4): 474–481
https://doi.org/10.3724/abbs.2022033
Advance Access Publication Date: 6 April 2022
Original Article

Original Article

Crystal structure of the polyketide cyclase from *Mycobacterium tuberculosis*

Jie Zhuang[1], Shihui Fan[1], Chenyun Guo[2], Liubin Feng[1], Huilin Wang[2], Donghai Lin[2,*], and Xinli Liao[1,*]

[1]Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China, and [2]Department of Chemical Biology, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China
*Correspondence address. Tel: +86-592-2186078; E-mail: dhlin@xmu.edu.cn (D.L.) / E-mail: xlliao@xmu.edu.cn (X.L.)

Received 9 October 2021 Accepted 11 November 2021

Abstract

About 40% of proteins are classified as conserved hypothetical proteins in *Mycobacterium tuberculosis* (TB). Identification and characterization of these proteins are beneficial to understand the pathogenesis of TB and exploiting novel drugs for TB treatments. The polyketide cyclase, a protein from *M. tuberculosis* (*Mt*PC) has been annotated as a hypothetical protein in Uniprot database. Sequence analysis shows that the *Mt*PC belongs to the NTF2-like superfamily proteins with diverse functions. Here, we determined the crystal structure of *Mt*PC at a resolution of 2.4 Å and measured backbone relaxation parameters for the *Mt*PC protein. *Mt*PC exists as a dimer in solution, and each subunit contains a six-stranded mixed β-sheet and three α helixes which are arranged in the order α1-α2-β1-β2-α3-β3-β4-β5-β6. The NMR dynamics analysis showed that the overall structure of *Mt*PC is highly rigid on ps-ns time scales. Furthermore, we predicted the potential function of *Mt*PC based on the crystal structure. Our results lay the basis for further exploiting and mechanistically understanding the biological functions of *Mt*PC.

Key words *Mt*PC, crystal structure, NMR dynamics, *Mycobacterium tuberculosis*, NTF2-like superfamily protein

Introduction

Tuberculosis (TB), caused by mycobacterium, is a chronic infectious disease that seriously endangers the health of the people. Approximately 7.0 million people were newly infected with TB in 2018, and nearly half a million new cases of rifampicin-resistant TB were found [1]. The emergence of multidrug resistance continues to be a public health threat, which has prompted the exploration of new drug targets [2,3]. The complete genome sequence for the H37Rv strain of *Mycobacterium tuberculosis* published in 1998 [4] allows a better understanding of the factors contributing to antibiotic resistance at a molecular level. Additionally, about 40% of gene products were annotated as hypothetical proteins with unknown or uncertain function [5]. These hypothetical proteins are well conserved among organisms. Growing evidence shows that these proteins are substantially involved in crucial biochemical process, which can be exploited as potential targets for the development of new drugs against deadly *M. tuberculosis*.

*Mt*PC (polyketide cyclase from *Mycobacterium tuberculosis*) is encoded by *Rv3768* gene (Gene ID: 886102) and composed of 119 amino acids with a predicted molecular weight of 13.17 kDa. *Mt*PC is a conserved hypothetical protein never reported previously. Based on the predicted amino-acid sequence of *Mt*PC, the BLAST search shows that *Mt*PC contains a conserved SnoaL-like domain (residues 18–109), belonging to the nuclear transport factor 2 (NTF2) like superfamily. This superfamily is highly conserved in terms of structure but diverse in functions with diverse substrates [6]. Expectedly, clarification of the structural basis of *Mt*PC will be of great benefit to exploiting potential functions of this protein.

In the present study, we determined the crystal structure of *Mt*PC at 2.4 Å by molecular replacement using SeMet-*Mt*PC L66M as searching template. We also analyzed the oligomeric state of *Mt*PC in solution by performing analytical ultracentrifugation (AUC) experiments, and explored biological functions by performing a structural comparison between *Mt*PC and its homologs by DALI search [7]. Meanwhile, we conducted NMR relaxation experiments to clarify dynamics properties of *Mt*PC. Our work lays the structural basis for both identifying the substrate-binding sites and understanding the potential functions for *Mt*PC.

Materials and Methods

Cloning and mutation

The gene of *Mt*PC was synthesized commercially with codons optimized (GenScript, Nanjing, China) and inserted in fame with SUMO tag into pSMT3 expression plasmid, which was constructed based on pET-28b (a gift from Prof. Jixi Li, Fudan University, Shanghai, China), using *Bam*HI and *Xho*I cloning sites. The *Mt*PC L66M was generated by PCR mutagenesis using a forward primer (5′-ATATTCCGAACGCGATGTGGGACCT-3′) and a reverse primer (5′-GGTCCCACATCGCGTTCGGAATATC-3′), and wild-type *Mt*PC as a template. The mutation was verified by DNA sequencing.

Expression and purification

The pSMT3-*Mt*PC plasmids were transformed into *Escherichia coil* BL21 (DE3). The cells were grown at 37°C in LB medium until $OD_{600\ nm}$ = 0.6–0.8, then the expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cultures were incubated at 25°C for 10 h. For the expression of either the $^{13}C/^{15}N$- or the SeMet-labeled protein, cells were harvested by centrifugation at 1500 *g* for 4 min when OD_{600} reached 0.9. For $^{13}C/^{15}N$-labeled protein, the cell pellets were re-suspended in M9 medium containing 2 g of ^{13}C-labeled glucose and 0.5 g of ^{15}N-labeled ammonium chloride. For the SeMet-labeled protein, the cell pellets were re-suspended in M9 medium containing 50 mg each of SeMet (Sigma-Aldrich, St Louis, USA), Val, Leu, and Ile and 100 mg each of Phe, Lys and Thr. Then 0.5 mM IPTG was added to induce protein expression at 25°C for 10 h. The cells were harvested by centrifugation at 7600 *g* for 4 min.

Thereafter, the cells were re-suspended in lysis buffer (50 mM Na_2HPO_4-NaH_2PO_4, and 300 mM NaCl, pH 8.0) and extracted using the sonication method. After cell disruption, the supernatants were collected by centrifugation at 11,600 *g*, 4°C for 30 min, and then loaded onto the Ni-NTA column (Thermo Fisher, Waltham, USA), washed with a stepwise imidazole (0, 10, 20, 30 and 40 mM imidazole). After that, the target protein was eluted with 300 mM imidazole. The SUMO-*Mt*PC was dialyzed against the lysis buffer and then the SUMO tag was removed by SUMO enzyme. The *Mt*PC protein and the SUMO tag were separated using a Ni-NTA column. Next, the protein was further purified by using a size exclusion Superdex™ 75 10/300GL column (GE Healthcare, Bethesda, USA). The mutants were purified in the same way. The concentration of *Mt*PC protein was measured with NanPhotometer N50 (Implen GmbH, Munich, Germany), and the purity of *Mt*PC protein was verified by 15% SDS-PAGE. Similarly, the *Mt*PC L66M protein was purified and characterized.

Crystallization

Initial crystallization screening was performed at 22°C by sitting-drop vapor diffusion method in a 96-well plate using crystallization kits (Hampton Research, Aliso Viejo, USA). The protein was prepared at a concentration of 6 mg/mL for crystallization. The crystallization conditions of the *Mt*PC protein (0.1 M sodium acetate hydrate, pH 5.2, and 2.4 M ammonium sulfate) and SeMet-*Mt*PC L66M protein (2.4 M sodium malonate, pH 7.0) were different. The crystals were transferred into a cryoprotection solution containing 5% glycerol, and quickly cooled in liquid nitrogen.

Data collection and structure determination and refinement

The X-ray diffraction data were collected using a pilatus3 6M detector on the 19U1 beamline at Shanghai Synchrotron Radiation Facility (SSRF, Shanghai, China). The diffraction data were indexed, integrated and scaled by HKL2000 [8]. The crystal structure of SeMet-*Mt*PC L66M was determined by single-wavelength anomalous dispersion (SAD). The Crank2 [9] program in the CCP4 suite was used to solve initial phase by the SAD phasing method based on SeMet-derived data. The crystal structure of *Mt*PC was determined by molecular replacement using SeMet-*Mt*PC L66M as a template. Both model building and structural refinement were performed with Phenix [10] and Coot [11]. All structural figures were rendered in Pymol. The crystal structure of *Mt*PC was deposited in the PDB with an accession code of 7E85.

Analytical ultracentrifugation experiments

To determine the oligomeric state of the *Mt*PC protein, sedimentation velocity experiments were performed at 20°C and 23,480 *g* on the Beckman-Coulter XLA ultracentrifuge (Beckman, Pasadena, USA) with an AN-60 Ti rotor. Equilibrium at each rotor speed was reached after 20 h. Double sector cells equipped with sapphire windows were loaded with 450 μL of the buffer (50 mM Na_2HPO_4-NaH_2PO_4, and 300 mM NaCl, pH 8.0) and 450 μL of the *Mt*PC protein. The experiments were performed with two concentrations of the protein (6.0 and 10.6 mg/mL). The sedimentation was monitored by measuring the absorbance at 280 nm. Data analysis was performed with the SEDFIT software, using a continuous sedimentation coefficient distribution model c(S).

NMR experiments

The ^{15}N- and $^{13}C/^{15}N$-labeled NMR samples were prepared with 0.8 mM *Mt*PC in the NMR buffer (200 mM NaCl, 50 mM Na_2HPO_4-NaH_2PO_4, pH 6.2, 0.02% NaN_3, and 10% D_2O). All NMR experiments were performed at 298 K on a Bruker Avance III 850 MHz spectrometer (Bruker, Rheinstetten, Germany) equipped with a TCI triple-resonance cryoprobe. The two- and three-dimensional spectra of ^1H-^{15}N HSQC, ^1H-^{13}C HSQC, HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, and HN(CA)CO were recorded for backbone resonance assignments. For ^{15}N T_1 measurements, the delay times were set to 10, 50, 100 (×2), 200, 400, 600, 800 (×2), 1200, 1600, and 2000 ms. For ^{15}N T_2 measurements, the delay times were set to 17.0, 33.9 (×2), 50.9, 67.8, 84.8, 102, 119, 136 (×2), 153, and 170 ms. The repeated experiments were used to estimate uncertainties of T_1 and T_2 values. For $\{^1H\}$-^{15}N NOE experiments, a delay of 2 s was followed by ^1H saturation of 3 s, and in the control experiments without ^1H saturation, a total delay of 5 s was applied. All NMR spectra were processed using NMRPipe and analyzed using NMRFAM-SPARKY.

Results

Expression, purification and characterization of *Mt*PC protein

The purity of the *Mt*PC protein was more than 90% after purification with Ni-NTA affinity column (Lanes 11,12, Figure 1A). The *Mt*PC protein was mostly eluted at 11.75 mL (Figure 1B). Based on our home-made calibration curve: lgMr (kDa) = –0.18446V (mL) + 3.63546, the molecular weight (MW) of *Mt*PC was estimated to be 29.98 kDa. This value was approximately twice of the size expected for the theoretical MW of one subunit (13.17 kDa), implying that the protein existed mainly as a dimer in solution.

Overall structure of *Mt*PC

The purified *Mt*PC and SeMet-labeled *Mt*PC L66M proteins were crystallized in different conditions. Both protein crystals displayed different space groups. The crystal structure of *Mt*PC was determined at 2.4 Å resolution using SeMet-labeled *Mt*PC L66M as searching template by molecular replacement and the space group is $P4_132$. The X-ray diffraction data and refinement statistics are shown in Table 1. The analysis of Ramachandran plot with Coot showed that most of the modeled residues were in preferred and allowed regions. The *Mt*PC molecule was composed of three nearly identical subunits (A–C) and displayed a "V"-like shape (Figure 2A). The RMSD deviation of a superposition of all main-chain atoms

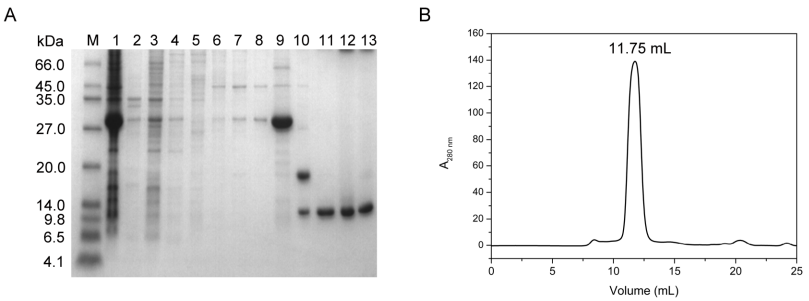

Figure 1. Purification and characterization of *Mt*PC (A) Purification of *Mt*PC. M: marker, Lane 1: supernatant, Lane 2: precipitation, Lane 3: flow-through, Lanes 4–8: washing with 0, 10, 20, 30, and 40 mM imidazole buffer, respectively, Lane 9: elution, Lane 10: protein samples after SUMO enzyme digestion, Lanes 11,12: purified protein after SUMO enzyme digestion by Ni-NTA, Lane 13: purified protein by size exclusion chromatography. (B) Size exclusion chromatography of *Mt*PC.

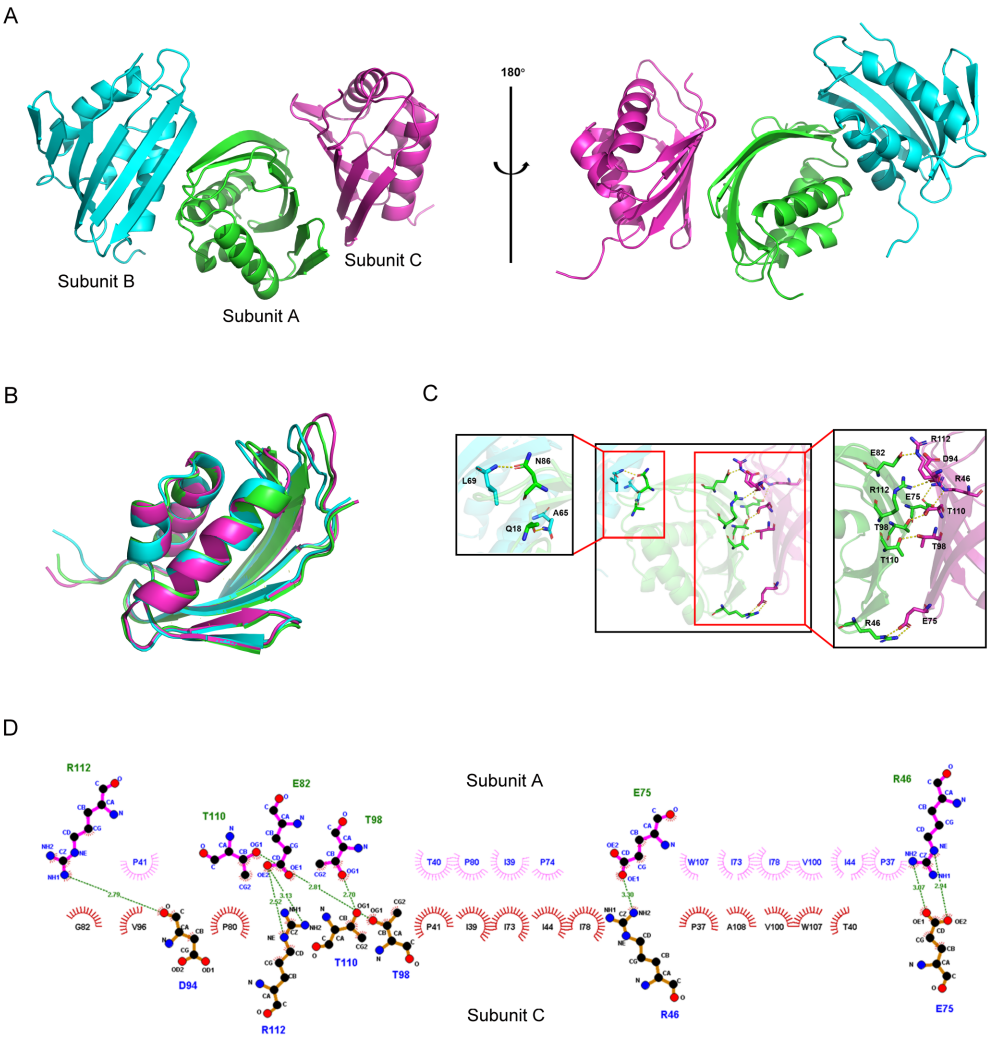

Figure 2. Overall structure of *Mt*PC protein and interactions between subunits (A) Overall structure. (B) Superimposed of three subunits. (C) Hydrogen bonds (yellow dotted line) between the subunits. (D) The hydrogen bond (green dash line) and hydrophobic interactions (arcs with spokes) between subunit A and subunit C.

Table 1. Data collection and refinement statistics

	SeMet-*Mt*PC L66M	*Mt*PC
Data collection		
Wavelength (Å)	0.97930	0.97852
Detector distance (mm)	450	350
a, b, c (Å)	66.703, 66.703, 148.841	147.798, 147.798, 147.798
α, β, γ (°)	90, 90, 120	90, 90, 90
Resolution range (Å)	50-2.40	50-2.42
Space group	$P6_222$	$P4_132$
Unique reflections	14836	21418 (2101)
Total observations	226367	1602165
Completeness (%)	99.03% (92.50%)	99.39 (100.00)
I/σ	43.33	36.00
CC1/2	0.989	0.939
Refinement		
Resolution range (Å)[a]	26.461-2.377 (2.462–2.377)	28.99-2.424 (2.511–2.424)
Reflections used in refinement	14788	21415 (2101)
Reflections used for R-free	1451	1065 (102)
Average B-factor	48.40	53.68
Wilson B-factor		44.73
R-work (%)[b]	0.2193 (0.2576)	0.2014 (0.2263)
R-free (%)[b]	0.2731 (0.3132)	0.2565 (0.3026)
Macromolecules	870	2691
Solvent	42	191
Protein residues	112	342
RMS (bonds)	1.04	0.008
RMS (angles)	0.008	0.098
Clashscore	7.64	7.60
Macromolecules	48.30	53.99
Solvent	50.60	49.20
Ramachandran plot		
Favored (%)	98.15	95.83
Allowed (%)	1.85	4.17
Outliers (%)	0.00	0.00

[a]Values in parentheses are for the highest resolution shell. [b]R-work was calculated with 95% of the unique reflections used for refinement, whereas R-free was calculated with the remaining 5% of the unique reflections.

between subunit A and B, A and C, B and C was 0.518, 0.422 and 0.344, respectively, indicating that these subunits were not significantly different in structure (Figure 2B).

We analyzed the interfaces between neighboring subunits and tested the probable quaternary structures of *Mt*PC by using the PISA server [12] (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). The surface area buried between the A and B subunits was 543.2 Å^2, and that between the A and C subunits was 957.0 Å^2. However, no surface area was identified between the B and C subunits. The inter-subunit interactions were primarily contributed by hydrogen bonds and several salt bridges. Side chains of the interfacing residues were mapped in the crystal structure of the *Mt*PC molecule (Figure 2C). Six interfacing residues (R46, E75, E82, D94, T98, T110, and R112), located in β-strand secondary structural elements, were identified between the A and C subunits. On the A-C subunit interface, four inter-subunit hydrogen bonds were formed between E82 and R112, R112 and D94, E75 and R46, T98 and T110. Moreover, two inter-subunit salt bridges were formed between E82 and R112, E75 and R46. However, only four interfacing residues (Q18, A65, L69, and N86) were identified between the A and B subunits, and two inter-subunit hydrogen bonds were formed between A65 and Q18, N86 and L69. Thus, it could be expected that the *Mt*PC protein exited as a dimer in solution. To gain more detail information of the A and C subunits, we used the Ligplot + software to analyze the hydrophobic interaction between the two subunits. Figure 2D showed that 11 residues in subunit A and 13 residues in subunit C participated in hydrophobic interactions.

Subunit structure of *Mt*PC

The *Mt*PC protein consists of 113 residues (3–116) in each monomer. The main features of the *Mt*PC monomer structure are a six-stranded mixed β-sheet and three α helixes arranged in the order α1-α2-β1-β2-α3-β3-β4-β5-β6 (Figure 3A). Three α-helixes (α1: 9–22, α2: 25–31, α3: 37–40) are packed on one side of the crystal structure and

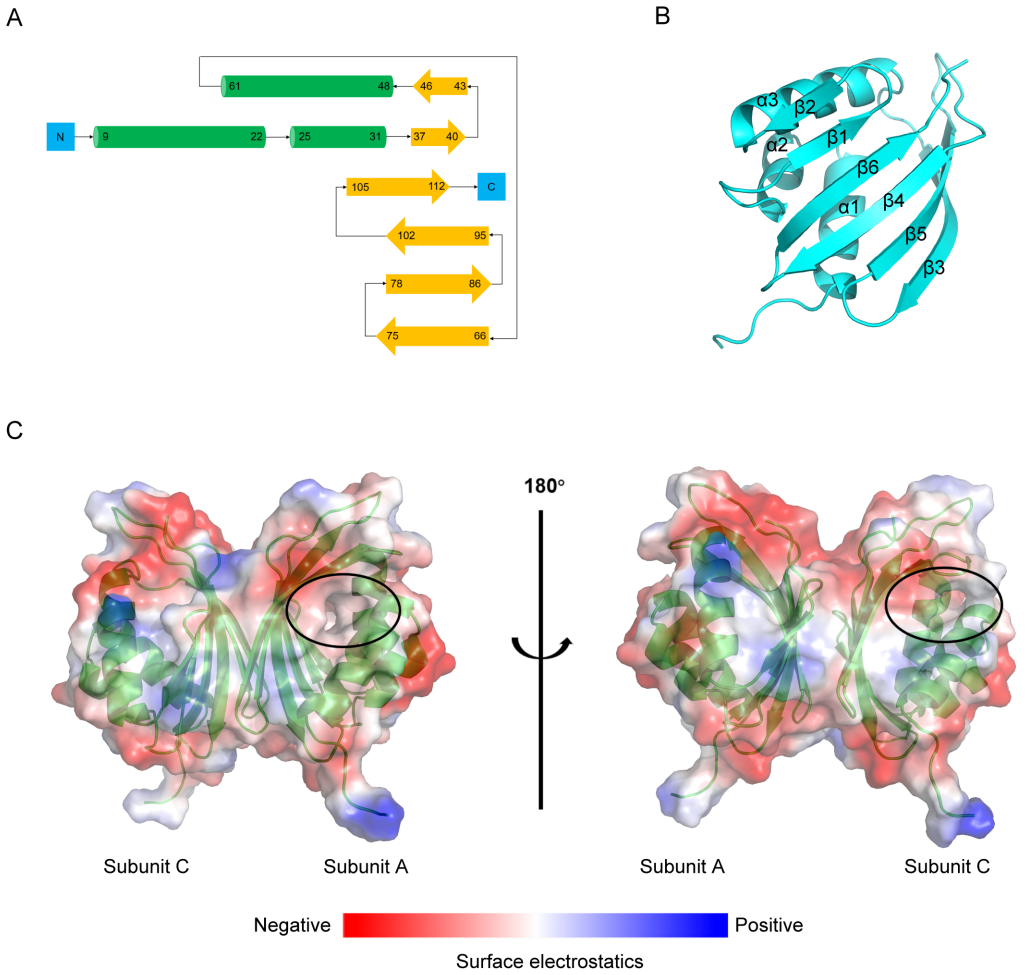

Figure 3. Subunit structure of *Mt*PC (A) Topology diagram of one *Mt*PC subunit (subunit A). (B) Subunit A. (C) The electrostatic surface of subunit A and subunit C. The proposed ligand-binding pocket is highlighted by a black ellipse.

six curved β strands (β1: 37–40, β2: 43–46, β3: 66–75, β4: 78–86, β5: 95–102, β6: 105–112) lie on the opposite side. The overall subunit resembles a cone with an inner cavity (Figure 3B). The N- and C-termini of each subunit is located in the α1, and C-termini and in the β6, respectively. The β1 and β2 are arranged anti-parallelly and the β2 is connected to the α3 helix by a loop. Then, the C-terminal of α3 is connected to another anti-parallel β sheet formed by β3, β4, β5 and β6. The electrostatic surface of subunit A and C is shown in Figure 3C. The surface of the dimer looks like a butterfly and each subunit has a pocket which is surrounded by α1, α3 together with β3, β4, β5, β6. We speculate that this pocket may potentially bind with ligands, which looks like a "stick" inserted into the pocket and each ligand binds with each subunit.

Backbone relaxation measurements of *Mt*PC

NMR dynamics experiments provide the general characteristics of protein in solution as well as internal motions on picosecond to nanosecond timescales of a biomolecule. Some insights into the specific structural changes or configurational associated with function could be obtained from its results. Therefore, we characterized the dynamics of *Mt*PC by liquid NMR spectroscopy, which is helpful for reflecting the structure and functions of the protein at molecular level. In this study, we labeled *Mt*PC with ^{13}C and ^{15}N isotopes, assigned more than 97% backbone resonances (BMRB accession number: 28132) and measured backbone ^{1}H-^{15}N relaxation parameters of R_1, R_2 and NOE (Figure 4). The secondary structures were predicted by TALOS-N based on the assigned chemical shifts. The *Mt*PC protein in solution state exhibited the similar

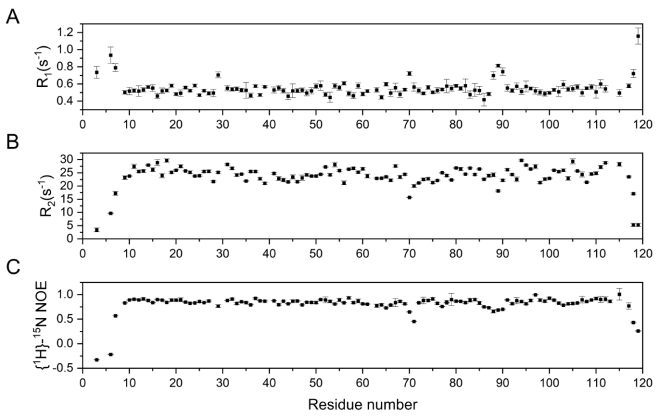

Figure 4. The ^{1}H-^{15}N relaxation parameters (bottom) of *Mt*PC (A) Longitudinal relaxation rate (R1). (B) Transverse relaxation rate (R2). (C) ^{1}H-^{15}N steady-state NOEs value.

secondary structures as those in crystal state. Sequence variations of the relaxation parameters were observed for residues K70 and T71 in β3, the loop between β4 to β5 (from N86 to S90), the loop in the N-terminal of the protein (from G2 to P9) and the loop in the C-terminal (from H116 to T119). These regions were highly flexible with ^{1}H-^{15}N NOE values less than 0.6. The secondary structural element β3 covering K70 and T71 had ^{1}H-^{15}N NOE values less than 0.5. Moreover, these regions showed significantly higher R_1 and lower R_2 values than most of general regions, indicating that they are involved in ps-ns internal motion.

Analysis of macromolecular rotational diffusion

Given that limited dynamics information was directly obtained from the NMR relaxation measurements of R1, R2 and ^{1}H-^{15}N NOEs, we carried out model-free analysis of dynamics data by using the FAST-model-free software (fastMF) [13,14], to obtain more internal motion information for the *Mt*PC protein. As the ratio of the parallel component to the perpendicular component for the rotation diffusion tensor D ($D_{ratio} = D_{par}/D_{per}$) was about 1.124, we were thus able to perform the model-free analysis based on an axial-symmetric rotation diffusion model.

The fitted values of generalized order parameter (S^2) in all amino acid residues were higher than 0.7, indicating that *Mt*PC was relatively rigid in solution. Conformational chemical exchange rates (R_{ex}) were obtained for several residues including H16, Q18, R52, K57, L58, D60, D94, D97, and Y113, which were involved in ms-μs molecular motion (Figure 5). The dynamics properties of these residues might be related to the functions of *Mt*PC. The overall correlation time of the molecule (τ_m) was 15.296 ns. Compared with other proteins with molecular weights of about 13 kDa, the measured τ_m value of *Mt*PC appeared relatively higher, supporting our speculation that the *Mt*PC protein may exist as a dimer in solution state. Furthermore, four residues showed ps-ns internal motion with effective correlation times (τ_e), including I62 and A88 in random coil, and I28 and S90 in secondary structure elements.

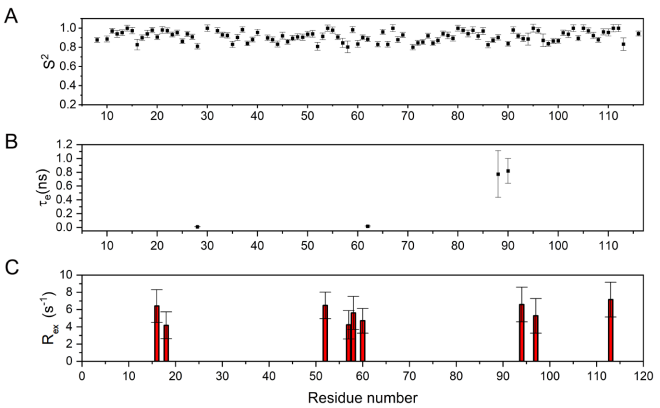

Figure 5. Graphical plots of internal motion parameters of *Mt*PC obtained by using FAST-model-free Optimized model-free parameters S^2 (A), τ_e (B) and R_{ex} (C) obtained from the model-free analysis of NMR relaxation data of ^{15}N-R1, -R2 and ^{1}H-^{15}N NOE.

*Mt*PC protein forms a dimer in solution

We further verified the oligomeric state of the *Mt*PC protein in solution. Analytical ultracentrifugation (AUC) analysis was performed on *Mt*PC at two concentrations which were used for protein crystallization experiments (6 mg/mL) and NMR experiments (10.6 mg/mL), respectively. The continuous c(S) distributions (Figure 6) implied that the *Mt*PC protein at the two different concentrations mostly existed as a dimer in solution (~99% of the total mass), as indicated by the areas under the c(S) curves. These results are consistent with those from the gel filtration experiment, confirming the dimeric form of *Mt*PC in solution.

Discussion

Tuberculosis is a chronic infectious disease that seriously endangers people's health. Although many proteins that may be related to the drug resistance of *M. tuberculosis* have been resolved, about 40% of gene products were annotated as hypothetical proteins with unknown or uncertain function. The growing evidence shows that these proteins are involved in important biochemical process. *Mt*PC is a conserved hypothetical protein with unknown function from *M. tuberculosis*, which will provide new mechanism to treat tuberculosis.

We herein determined the crystal structure of *Mt*PC at 2.4 Å and clarified its backbone dynamics properties. The crystal structure of the *Mt*PC protein showed that the asymmetric unit of the space group contains three molecules, but the backbone dynamics analysis and AUC result suggested that *Mt*PC exists as a dimer in solution. Combined with the predicted result from the PISA program (Figure 2C), we revealed that the A and C subunits substantially form a dimer through the subunit-interaction. Of note, the crystal packing potentially contributes to the subunit-interaction between the A and B subunits.

We performed structural comparison between *Mt*PC and structural homologs to gain more detail information of *Mt*PC structure and its potential biological functions. As annotated in the NCBI database, the *Mt*PC protein contains a predicted SnoaL-like domain

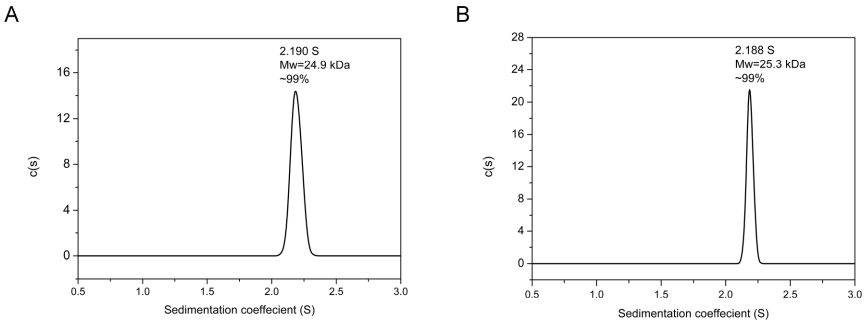

Figure 6. Sedimentataion velocity experiments of *Mt*PC (A) 6 mg/mL. (B) 10.6 mg/mL.

(residues 18–109). A previous study demonstrated that residues P15, I51 and D121 in the SnoaL domain play crucial roles in the protein functions [15]. However, *Mt*PC does not contain these three conserved residues. The sequence identity between *Mt*PC and SnoaL was calculated to be only 12.6% with ClustalW [16]. These results implied that *Mt*PC does not share similar functions to SnoaL.

A DALI search for structural homologs of *Mt*PC identified several proteins with similar structural folds to *Mt*PC. The top six protein structures ordered by Z-score are showed in Table 2. All the top five proteins are hypothetical proteins without characterized properties. They share less than 20% sequence identities with *Mt*PC. Interestingly, steroid delta-isomerase (KSI) adopts a 3D structure similar to *Mt*PC with an RMSD of 1.8 Å. Although the KSI protein in *Comamonas testosteroni* (*Ct*KSI, PDB: 1ohp) has a Z-score of 14.8 which is lower than the other five homologous protein, it shares the largest sequence identity (24%) with *Mt*PC. Moreover, D99 in *Ct*KSI is conserved in *Mt*PC, which acts as a crucial residue for the functions of *Ct*KSI. However, this residue is not conserved in other structural homologs (Figure 7).

Here, we focused on the structural comparison between *Ct*KSI and *Mt*PC. Both proteins belong to the NTF2-like superfamily with similar structural features. Previous studies clarified the catalytic mechanism of *Ct*KSI and identified its essential catalytic residues (Y14, D38 and D99) [17–19]. Both Y14 and D38 form hydrogen bonds directly with the O-3 atom of the substrate steroid. The catalytic residue D38 acts as the base responsible for transferring the proton in the steroid. The residue D97 in *Mt*PC is equivalent to D99 in *Ct*KSI (Figure 8), which is oriented into the active site. However, the two residues Y14 and D38 in *Ct*KSI are not conserved in *Mt*PC. Indeed, the binding assay with 5AND (5-androstene-3,17-dione) did not exhibit the activity of the *Ct*KSI protein (data no shown). We also applied the ProFunc program [20] (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc) to predict the biological functions of *Mt*PC based on its crystal structure. The results showed that the *Mt*PC protein might be involved in several metabolic processes such as lipid metabolism and steroid metabolism.

While the biological functions of the *Mt*PC protein are currently unclear, its crystal structure shows that several hydrophobic residues are situated in the pocket of each subunit. The structural comparison of *Mt*PC with its homologs suggests that *Mt*PC might be involved in steroid metabolism. Notably, several hydrophobic residues are located in the pocket, including F13, L20, L57, W67, L69, L81, W83, F99 and V111, which potentially contribute to the binding of *Mt*PC with hydrophobic ligands. In the next step, we will select some compounds such as steroids to conduct functional assays according to the sizes of the hydrophobic pockets. Further-

Table 2. DALI structural homologs search

Rank	PDB code	Z-score	RMSD(Å)	Number of residues	Identity(%)	Description
1	5evh-A	15.5	1.7	121	13	Uncharacterized protein
2	5tgn-A	15.4	1.9	109	14	Uncharacterized protein
3	4u13-A	15.3	2.0	109	15	Putative polyketide cyclase sma1630
4	3bb9-F	14.9	1.8	120	12	Putative orphan protein
5	3f8x-D	14.8	2.2	132	12	Putative delta-5-3-ketosteroid
6	1ohp-A	14.8	1.8	125	24	Steroid delta-isomerase

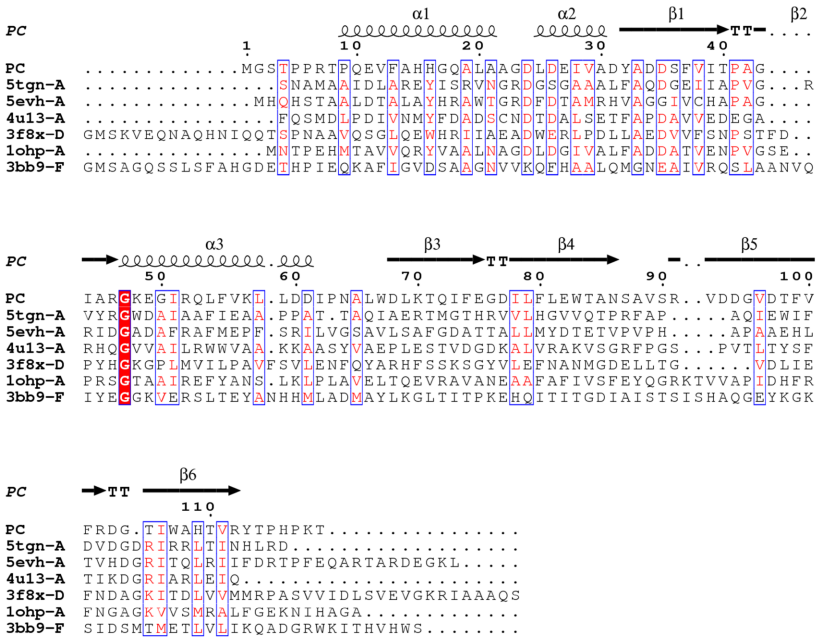

Figure 7. Sequence alignment of *Mt*PC from *M. tuberculosis* (PDB code 7E85) with other homologs PDB 5evh, Uncharacterized protein; PDB 5tgn, Uncharacterized protein; PDB 4u13, Putative polyketide cyclase sma1630; PDB 3bb9, Putative orphan protein; PDB 3f8x, Putative delta-5-3-ketosteroid; PDB 1ohp, Steroid delta-isomerase.

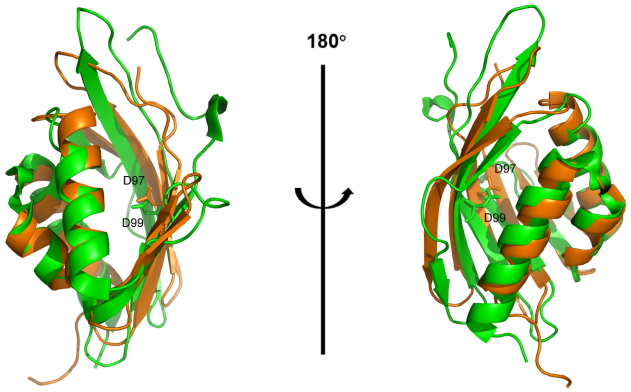

Figure 8. Structural alignments of *Mt*PC (orange) and KSI (green)

more, we will apply drug discovery software to screen lead compounds, and use molecular modeling approaches to exploit the potential ligands of *Mt*PC based on the crystal structure of *Mt*PC. In addition, we will also use bioinformatics tools to further predict the biological functions of *Mt*PC, including crucial residues, conservative residues and motifs, as well as domains, with necessary experimental confirmations based on the amino acid sequence of this protein.

In conclusion, our work may be beneficial to mechanistically understand the *Mt*PC protein and other hypothetical proteins in *M. tuberculosis*.

Funding

This work was supported by the grants from the National Key Research and Development Project of China (No. 2016YFA0500600).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Harding E. WHO global progress report on tuberculosis elimination. *Lancet Respiratory Med* 2020, 8: 19
2. Koul A, Arnoult E, Lounis N, Guillemont J, Andries K. The challenge of new drug discovery for tuberculosis. *Nature* 2011, 469: 483–490
3. Walter ND, Strong M, Belknap R, Ordway DJ, Daley CL, Chan ED. Translating basic science insight into public health action for multidrug- and extensively drug-resistant tuberculosis. *Respirology* 2012, 17: 772–791
4. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998, 393: 537–544
5. Camus JC, Pryor MJ, Médigue C, Cole ST. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 2002, 148: 2967–2973
6. Eberhardt RY, Chang Y, Bateman A, Murzin AG, Axelrod HL, Hwang WC, Aravind L. Filling out the structural map of the NTF2-like superfamily. *BMC BioInf* 2013, 14: 327
7. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993, 233: 123–138
8. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997, 276: 307–326
9. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, *et al.* Overview of the *CCP* 4 suite and current developments. *Acta Crystlogr D Biol Crystlogr* 2011, 67: 235–242
10. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystlogr D Biol Crystlogr* 2002, 58: 1948–1954
11. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystlogr D Biol Crystlogr* 2004, 60: 2126–2132
12. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007, 372: 774–797
13. Cole R. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J Biomol NMR* 2003, 26: 203–213
14. Dosset P, Hus JC, Blackledge M, Marion D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* 2000, 16: 23–28
15. Sultana A, Kallio P, Jansson A, Wang JS, Niemi J, Mäntsälä P, Schneider G. Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *EMBO J* 2004, 23: 1911–1921
16. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* 1994, 22: 4673–4680
17. Cho HS, Choi G, Choi KY, Oh BH. Crystal structure and enzyme mechanism of Δ^5-3-ketosteroid isomerase from *Pseudomonas testosteroni*. *Biochemistry* 1998, 37: 8325–8330
18. Choi G, Ha NC, Kim SW, Kim DH, Park S, Oh BH, Choi KY. Asp-99 donates a hydrogen bond not to Tyr-14 but to the steroid directly in the catalytic mechanism of Δ^5-3-ketosteroid isomerase from *Pseudomonas putida* biotype B. *Biochemistry* 2000, 39: 903–909
19. Hawkinson DC, Pollack RM, Ambulos Jr. NP. Evaluation of the internal equilibrium constant for 3-Oxo-.DELTA.5-steroid isomerase using the D38E and D38N mutants: the energetic basis for catalysis. *Biochemistry* 1994, 33: 12172–12183
20. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 2005, 33: W89–W93