Molecular Microbiology (2014) 94(5), 1006–1023 ■
doi:10.1111/mmi.12817
First published online 5 November 2014

Mycobacterium smegmatis BioQ defines a new regulatory network for biotin metabolism

Qing Tang,[1] Xinfeng Li,[1] Tingting Zou,[1] Huimin Zhang,[2] Yingying Wang,[1] Rongsui Gao,[2] Zhencui Li,[2] Jin He[1] and Youjun Feng[2]***

[1]*State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China.*
[2]*Department of Medical Microbiology & Parasitology, School of Basic Medical Sciences, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310058, China.*

Summary

Biotin (vitamin H), the sulfur-containing enzyme cofactor, is an essential micronutrient for three domains of life. Given the fact that biotin is an energetically expensive molecule whose *de novo* biosynthesis demands 20 ATP equivalents each, it is reasonable that bacteria have evolved diversified mechanisms in various microorganisms to tightly control biotin metabolism. Unlike the *Escherichia coli* BirA, the prototypical bi-functional version of biotin protein ligase (BPL) in that it acts as a repressor for biotin biosynthesis pathway, the BirA protein of *Mycobacterium smegmatis* (*M. smegmatis*), a closely relative of the tuberculosis-causing pathogen, *Mycobacterium tuberculosis*, lacked the DNA-binding activity. It raised a possibility that an alternative new regulator might be present to compensate the loss of regulatory function. Here we report that this is the case. Genomic context analyses of *M. smegmatis* detected a newly identified BioQ homolog classified into the TetR family of transcription factor and its recognizable palindromes. The *M. smegmatis* BioQ protein was overexpressed and purified to homogeneity. Size-exclusion chromatography combined with chemical cross-linking studies demonstrated that the BioQ protein had a propensity to dimerize. The promoters of *bioFD* and *bioQ/B* were mapped using 5′-RACE. Electrophoretic mobility shift assays revealed that BioQ binds specifically to the promoter regions of *bioFD* and *bioQ/B*. Further DNase I foot-printing elucidated the BioQ-binding palindromes. Site-directed mutagenesis suggested the important residues critical for BioQ/DNA binding. The isogenic mutant of *bioQ* (Δ*bioQ*) was generated using the approach of homologous recombination. The *in vivo* data from the real-time qPCR combined with the *lacZ* transcriptional fusion experiments proved that removal of *bioQ* gave significant increment with expression of *bio* operons. Also, expression of *bio* operons were repressed by exogenous addition of biotin, and this repression seemed to depend on the presence of BioQ protein. Thereby, we believed that *M. smegmatis* BioQ is not only a negative auto-regulator but also a repressor for *bioFD* and *bioB* operons involved in the biotin biosynthesis pathway. Collectively, this finding defined the two-protein paradigm of BirA and BioQ, representing a new mechanism for bacterial biotin metabolism.

Accepted 28 September, 2014. For correspondence. *E-mail fengyj@zju.edu.cn; Tel. (+86) 571 88208524; Fax (+86) 571 88208524; **E-mail hejin@mail.hzau.edu.cn; Tel. (+86) 27 87280670; Fax (+86) 27 87280670.

Introduction

Biotin (vitamin B_7 or vitamin H) is a coenzyme/micronutrition essential for three domains of life (Beckett, 2007; 2009). This requirement for biotin is linked to central metabolisms such as carboxylation, decarboxylation and trans-carboxylation (Knowles, 1989; Attwood and Wallace, 2002). It seemed reasonable that nearly all the organisms have retained the biotin protein ligase (BPL), a highly conserved enzyme for catalyzing the attachment of biotin to its cognate acceptor proteins like AccB/BCCP responsible for the first committed step of fatty acid synthesis (Chakravartty and Cronan, 2012; Feng *et al*., 2013a). We are aware that plants, most fungi and most bacteria synthesize biotin whereas mammals and birds cannot (Beckett, 2009; Feng *et al*., 2013b). Microorganisms have evolved two different mechanisms to meet the physiological demand for biotin, the prevalent one of which is the *de novo* synthesis pathway, the other is a scavenging route (Beckett, 2007). It is universal that the four genes (*bioF*/*bioA*/*bioD*/*bioB*) encoding the latter step of biotin synthesis are constantly present in any microbe with an ability to produce biotin (Fig. 1A) (Beckett, 2007; 2009). The paradigm mechanism for biotin synthesis in *Escherichia coli*

begins by hijacking a modified fatty acid biosynthesis pathway (Lin and Cronan, 2010; Lin *et al*., 2010) and then proceeds through a four-step path with pimeloyl-ACP as precursor (Lin and Cronan, 2010). To make/activate the pimelate thioester, BioC-BioH path is used in *E. coli* (Agarwal *et al*., 2012; Lin and Cronan, 2012), whereas a single enzyme BioW is engaged in *Bacillus* (Ploux *et al*., 1992; Bower *et al*., 1996; Manandhar and Cronan, 2013). In contrast, the bacteria (e.g. the zoonotic agent *Streptococcus*) without full functional *bio* operons exploited the transporter BioY to salvage the trace biotin from their growing environment/niche (Rodionov *et al*., 2002; Hebbeln *et al*., 2007).

Given the fact that biotin is an expensive molecule in that 20 ATP equivalents are consumed to synthesize one biotin, it seemed very true that tight/strict regulation of bacterial biotin metabolism is required physiologically. To the best of our knowledge, no less than two different regulatory machineries have been reported (Fig. 1B, C, E and F) (Beckett, 2007; Chakravartty and Cronan, 2012; Feng *et al*., 2013a,b). The classic biotin sensing system is an unusual *E. coli* BirA regulatory protein (Otsuka and Abelson, 1978; Beckett, 2007) in that this *bio* repressor has the activity of biotin protein ligase (Fig. 1B and E) (Rodionov *et al*., 2002; Beckett, 2009; Chakravartty and Cronan, 2012). Upon excess of biotin is present, the biotinoyl-5′-AMP-bound form of BirA protein occupies on the *bio* operator and then effectively represses transcription of biotin-related genes (Fig. 1H), *vice versa*. However, not all the BirA protein possesses the regulatory function in that the DNA-binding domain is lacking (Rodionov *et al*., 2002). Given this observation, Rodionov *et al*. (2002) predicted that some new transcription factors might compensate the loss of BirA in its regulatory function. As expected, we reported this is the case in the plant pathogen *Agrobacterium* (Feng *et al*., 2013b) and the human pathogen *Brucella* (Fig. 1C, F and I) (Feng *et al*., 2013a). In light that a new type of regulator BioR belonging to GntR family of transcription factor was found to act as a repressor for biotin metabolism (Fig. 1C), we raised a novel two-protein paradigm of BirA-BioR for biotin sensing (Fig. 1F and I) (Feng *et al*., 2013a,b). Unlike the *E. coli* BirA regulator whose function depends on the presence of the biotinoyl-5′-AMP ligand, it seemed likely that biotin-related metabolites have no roles in BioR-mediated regulation (Feng *et al*., 2013a). Using genomic context-based approach, Brune *et al*. (2012) discovered a TetR-like transcription factor, referred to BioQ (Fig. 1D), and preliminarily assigned it into biotin synthesis pathway of a biotin-auxotrophic *Corynebacterium glutamicum*. Somewhat, it further pinpointed the diversified regulation of biotin biosynthesis, and also extended our understanding the physiology and metabolism of biotin enzyme cofactor, a C7 fatty acid-derived vitamin.

Human tuberculosis is a leading serious bacterial disease in that the number of the patients worldwide is estimated to be around 1–2 billion, and no less than 2 million of death cases were recorded each year (Yang *et al*., 2012a). *Mycobacterium tuberculosis* is responsible for this infectious disease. It seemed likely that decoding genome sequence of *M. tuberculosis* H37Rv in 1998 had furthered our understanding the biology and pathogenesis of this bacterial pathogen (Cole *et al*., 1998; Domenech *et al*., 2001; Yang *et al*., 2012a). Although the feasibility/suitability of *M. smegmatis* as an alternative model for *M. tuberculosis* infection was in vigorous debate (Reyrat and Kahn, 2001), it still exhibited a series of advantages that included (i) fast-growing, (ii) avirulent (BSL-1) and (iii) bigger genome covering most of genes from the virulent species (Domenech *et al*., 2001; Tyagi and Sharma, 2002). Genomic context analyses of mycobacteria suggested the constant presence of a mono-functional BirA protein (Brune *et al*., 2012). Surprisingly, *M. smegmatis* (the closely relative of *C. glutamicum* within the phylum of *Actinobacteria*) possesses a *bioQ* homolog, whereas *M. tuberculosis* not (Domenech *et al*., 2001; Brune *et al*., 2012). Also the putative BioQ binding sites (referred to BioQ box) were localized to be in the inter-genic region between *bioQ* and *bioB* plus in front of *bioF* (Brune *et al*., 2012) (Fig. 2). However, all the predictions need experimental validations.

In this article, we integrated *in vitro* and *in vivo* evidence to prove the function of *M. smegmatis* BioQ protein and its recognizable sites. Different from an observation with *C. glutamicum*, the BioQ-mediated regulatory architecture is not involved in *bioYMN* of *M. smegmatis*. We anticipated our finding might represent a third regulatory mechanism for biotin metabolism: the two-protein paradigm of BirA and BioQ (Fig. 1D, G and J).

Results

BioQ is a TetR family transcriptional regulator ubiquitous in mycobacteria

The *bioQ* gene encodes a polypeptide that belongs to the TetR family of transcriptional regulators. The TetR family transcriptional regulators usually function as transcriptional repressors, and are particularly abundant in bacteria exposed to frequent environmental changes (Ramos *et al*., 2005). There are 51 and 113 members TetR family transcriptional regulators encoded in the genome of *M. tuberculosis* H37Rv and *M. smegmatis* MC^2155 respectively. BioQ in *C. glutamicum* was reported to negatively control the transcription of biotin metabolism relative genes (Brune *et al*., 2012). The *bioQ* gene was widely distributed in mycobacteria and colocalized with the biotin synthesis related gene cluster (Fig. 2A). Multiple sequence alignments reveled that BioQ orthologues in

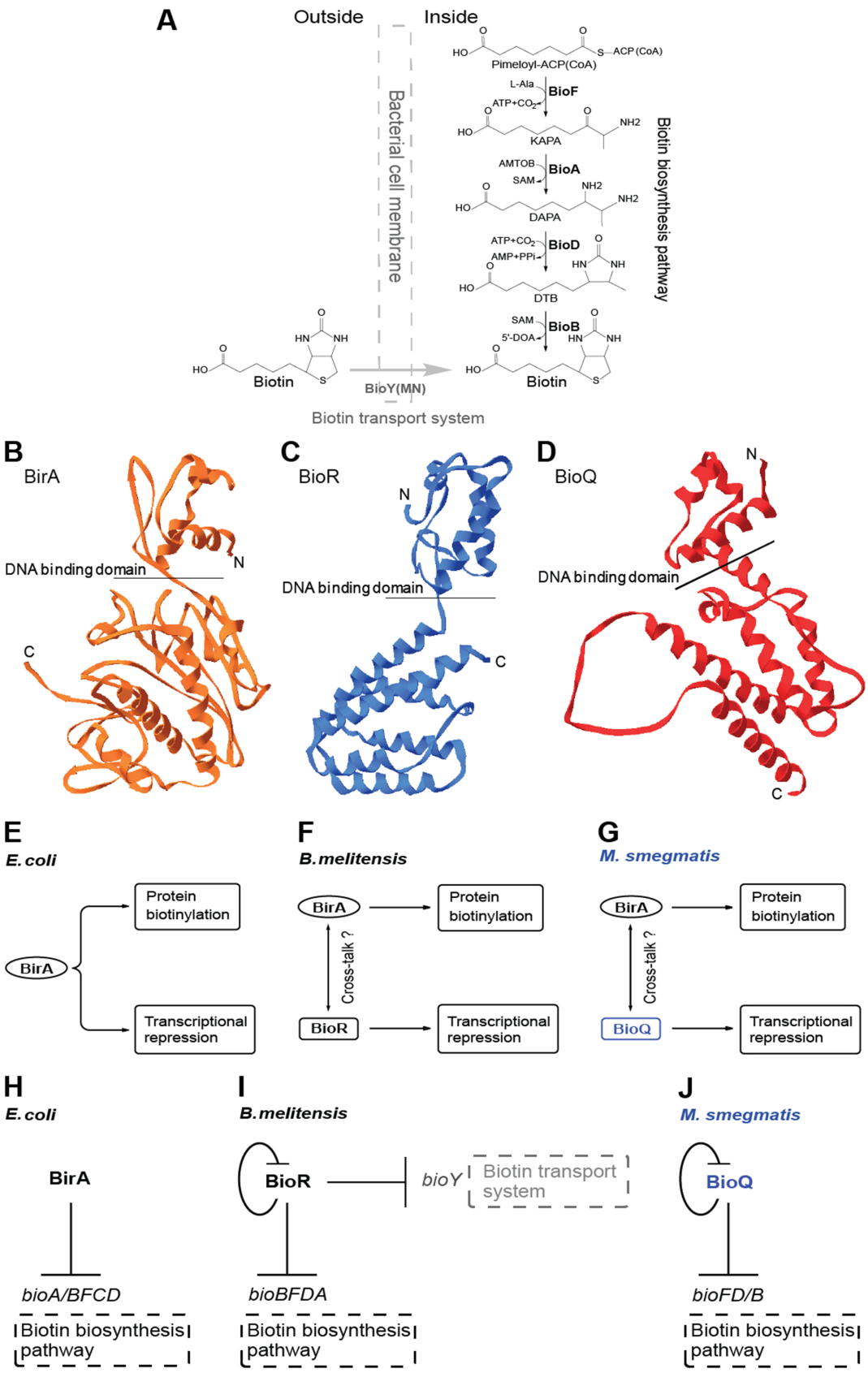
A
Outside
Inside
Bacterial cell membrane
Pimeloyl-ACP(CoA)
ACP (CoA)
L-Ala
ATP+CO2
BioF
KAPA
AMTOB
SAM
BioA
DAPA
ATP+CO2
AMP+PPi
BioD
DTB
SAM
5'-DOA
BioB
Biotin
BioY(MN)
Biotin
Biotin transport system
Biotin biosynthesis pathway
B
BirA
DNA binding domain
N
C
C
BioR
N
DNA binding domain
C
D
BioQ
N
DNA binding domain
C
E
E. coli
BirA
Protein biotinylation
Transcriptional repression
F
B. melitensis
BirA
Protein biotinylation
Cross-talk ?
BioR
Transcriptional repression
G
M. smegmatis
BirA
Protein biotinylation
Cross-talk ?
BioQ
Transcriptional repression
H
E. coli
BirA
bioA/BFCD
Biotin biosynthesis pathway
I
B. melitensis
BioR
bioY
Biotin transport system
bioBFDA
Biotin biosynthesis pathway
J
M. smegmatis
BioQ
bioFD/B
Biotin biosynthesis pathway

Fig. 1. Working model for bacterial biotin biosynthesis path and its diversified regulation.
A. A working model for bacterial biotin biosynthesis pathway.
B–D. Ribbon illustration for crystal structure of *E. coli* BirA protein (in panel B) and the modeled structures of both the GntR-like regulator BioR (in panel C) and the TetR family of transcriptional factor BioQ (in panel D).
E. *E. coli* BirA is an unusual regulator in that it has the activity of biotin protein ligase (BPL).
F. In *Brucella melitensis*, BirA is only a mono-functional BPL, and BioR compensates the loss of regulatory function.
G. In *M. smegmatis*, BirA is only a mono-functional BPL, and BioQ compensates the loss of regulatory function.
H. *E. coli* BirA represses expression of the *bioA/BFCD* operons.
I. *B. melitensis* BioR represses expression of both biotin uptake route and *de novo* biotin synthesis pathway.
J. *M. smegmatis* BioQ represses expression of *bioFD/B* and *bioQ*.

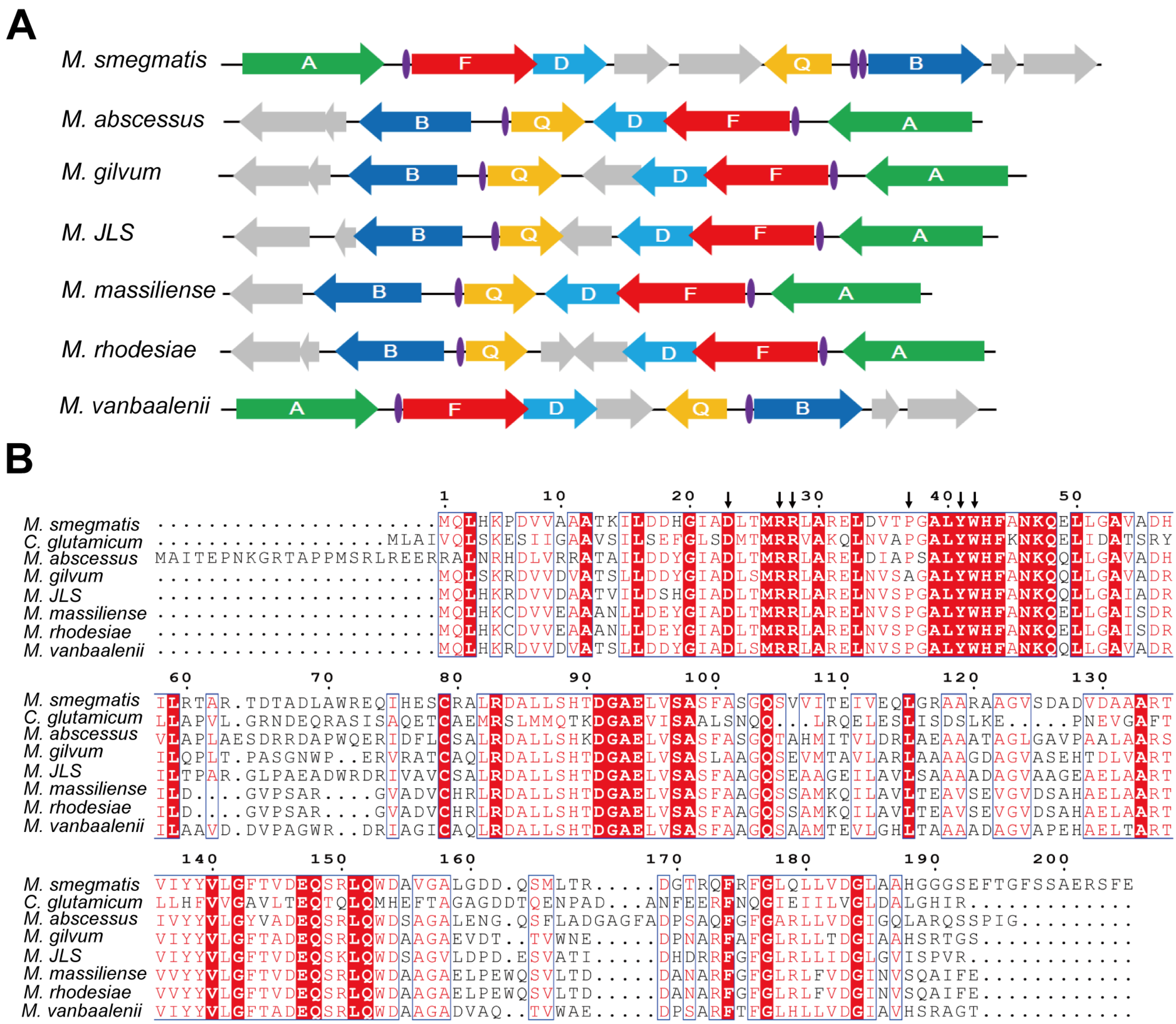

Fig. 2. Genetic organization and bioinformatics analyses for *bioQ* homologs from *Mycobacterium* species.
A. Genetic organization of *bioQ* and the putative biotin synthesis operons in mycobacteria. The white and grey arrows indicate the gene loci, and the putative BioQ binding sites were marked by grey bars.
B. Multiple sequence alignment of BioQ homologs. White letters shaded in red denote identical residues, red letters with white background are similar residues, black letters indicate variable residues and dots represent gaps. The six possible residues that are critical for DNA binding (D23, R27, R28, P37, 41Y and 42W) are highlighted with arrows. The amino acid sequences are from *M. abscessus* (Accession No.: YP_001703419.1), *M. gilvum* (Accession No.: ABP46099.1), *M. JLS* (Accession No.: ABN98870.1), *M. massiliense* (Accession No.: AFN63603.2), *M. rhodesiae* (Accession No.: AEV75711.1), *M. smegmaitis* (Accession No.: YP_887507.1), *M. vanbaalenii* (Accession No.: ABM13599.1), and *C. glutamicum* (Accession No.: cg2309).

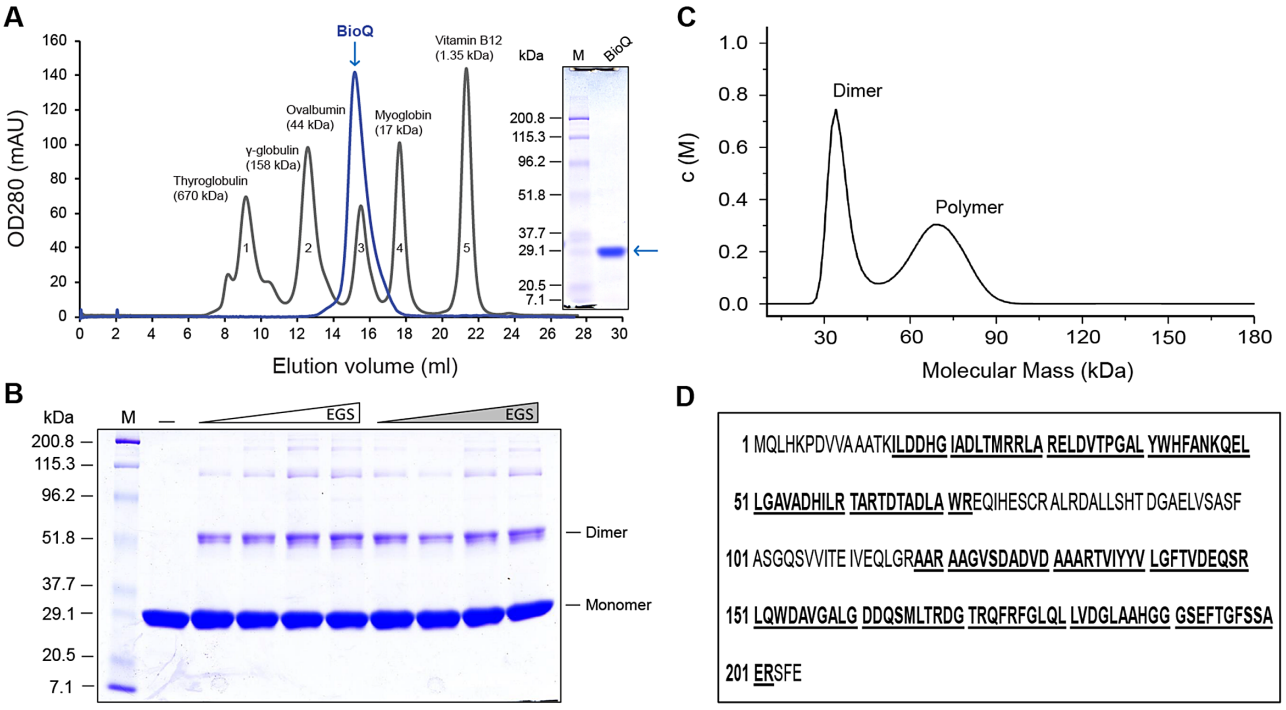

Fig. 3. Characterization of *M. smegmatis* BioQ protein.
A. Gel exclusion chromatographic profile of the *M. smegmatis* BioQ protein. The Superdex 200HR 10/30 column (GE Healthcare) was adjusted using the standard protein (Bio-Rad) consisting of Thymoglobin (670 kDa), γ-globulin (158 kDa), Ovalbumin (44 kDa), Myoglobin (17 kDa) and Vitamin B12 (1.35 kDa). The analytical FPLC analyses with this column showed that the expected peak of purified *M. smegmatis* BioQ was eluted at the position of ~ 15 ml (indicated with an arrow), suggesting its apparent molecular mass is little bit more than 44 kDa (Ovalbumin), but far less than 158 kDa (γ-globulin). Given the fact that the ideal molecular weight of recombinant *M. smegmatis* BioQ in monomer is about 23 kDa, we concluded that the prevalent form of BioQ in solution might be in dimer (~ 46 kDa). The inset gel is the SDS-PAGE profile of the purified *M. smegmatis* BioQ. OD280, optical density at 280 nm; mAU, milli-absorbance units.
B. Chemical cross-linking assay for the solution structure of *M. smegmatis* BioQ. M: Protein standard marker (Bio-Rad). Minus denotes no addition of the EGS chemical cross-linker. The level of EGS cross-linker illustrated with a triangle varies from 0.2, 0.5, 1.0, to 2.0 μM. In addition, the grey triangle on right hand also suggested addition of 100 μM biotin.
C. Sedimentation velocity analysis of BioQ protein. Sedimentation velocity analysis was carried out with an analytical ultracentrifuge (Beckman Coulter ProteomeLab XL-A) with an An-60 Ti rotor at 4°C. BioQ protein was prepared at in OD_{280} = 1 in a buffer containing 25 mM Tris (pH 8.0). Buffer alone was used as the reference solution. The molecular weights were calculated by the SEDFIT software, and the molecular weights estimated for each peak are stated.
D. MS identification of *M. smegmatis* BioQ. The matched amino acid residues are given bold and underlined type (67%).

different *Mycobacterium* species shared a high similarity with *C. glutamicum* BioQ (Fig. 2B). All the evidences suggested that BioQ could be a unique regulator involved in biotin metabolism in mycobacteria.

Characterization of M. smegmatis *BioQ protein*

To explore the putative function of BioQ, the recombinant BioQ-His_6 protein was purified from *E. coli* BL21 (DE3) harboring pET28a-*bioQ* using Ni-NTA agarose chromatography, and further fractionated by Superdex 200 gel filtration chromatography (Fig. 3A). FPLC profile suggested that BioQ protein can form dimer (Fig. 3A), which is further validated by the result of analytical ultracentrifugation (Fig. 3C). Apparently, our observation was consistent with the fact that the TetR family regulator usually appears as a homo-dimer (Ramos *et al.*, 2005). Using chemical cross-linking assays, we also proved that besides the monomer, the dimer form of BioQ also occurred in the solution. The solution structure of BioQ was independent of biotin as no difference was observed when increasing amount of biotin was added (Fig. 3B). The purified protein was then subjected to 12% SDS-PAGE to test its homogeneity (Fig. 3A). The recombinant protein loaded on SDS-PAGE exhibited a clear band about 23 kDa which corresponded to the predicted size. The band was then cut and subjected to liquid chromatography mass spectrometry, the results proved that the protein indeed come from *M. smegmatis* BioQ with a corresponding coverage score of 67% (Fig. 3D).

Genetic organization of bioQ and the putative biotin synthesis operons

The *bioQ* and biotin synthesis-related genes were neighbor-adjacent in the genome, but the genetic organi-

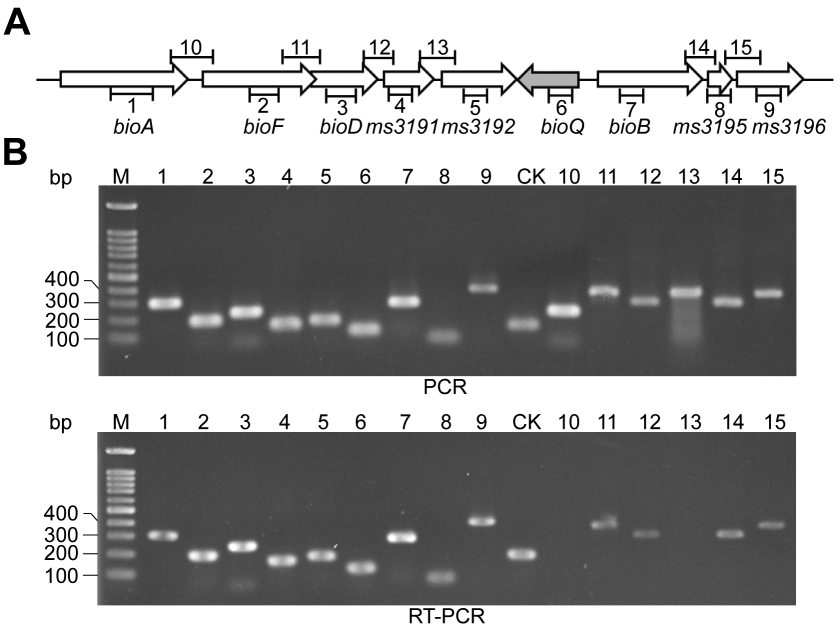

Fig. 4. Transcriptional analyses of the *bioQ* and biotin synthesis-related genes.
A. Genetic organization of the *bioQ* gene and its neighboring biotin synthesis-relevant operons. The white and grey arrows indicate the gene loci. The numbered short lines (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15) represent the specific PCR amplicons that could be observed in following PCR and RT-PCR assays (in panel B).
B. PCR and RT-PCR analyses of the *bioQ* and the putative biotin synthesis-related loci. The primer numbering was identical to that seen in panel A and CK (control) denotes the *16S* rDNA.

zation and regulatory mechanism remained unclear. Both PCR and RT-PCR assays were performed using multiple sets of primers to test if the genes constitute an operon (Fig. 4A and B). First, PCR analyses were conducted to validate the primers. All the intragenic amplifications (1, 2, 3, 4, 5, 6, 7, 8 and 9) were observed in RT-PCR assays, verifying that all the nine genes had detectable transcription. Positive amplifications of intragenic regions 11, 12, 14 and 15 were obtained by RT-PCR analyses (Fig. 4B). Therefore, *bioF*, *bioD* and *MSMEG_3191*, excluding *bioA* and *MEMEG_3192*, were in an operon. In addition, *bioB*, *MMEG_3195* and *MSMEG_3196* also comprised a transcription unit and hence an operon. Interestingly, *MSMEG_3195* and *MSMEG_3196* might also be controlled by other promoters as these genes showed much higher levels of the transcription than *bioB* illustrated by RNA-seq data of *M. smegmatis* (not shown).

The isolated gene *bioA* coding for an adenosylmethionine-8-amino-7-oxononanoate amino-transferase played a key role in the biotin biosynthetic pathway (Feng *et al*., 2013a). It was possible that *bioA* was a constitutively expressed gene or regulated by other transcription regulators. The function of *MSMEG_3191*, *MSMEG_3195* and *MSMEG_3196* coding for hypothetical proteins still required further investigation, and they might make contribution to the biotin metabolism in an obscure mechanism. The transcription initiation sites of the two operons and *bioQ* gene were identified by 5′-RACE experiments (Fig. 5). We therefore mapped the promoters of *bioQ* and the biotin synthesis operons.

Characterization of BioQ binding sites

To determine whether BioQ function as a direct regulator of biotin synthesis-related genes, electrophoretic mobility shift assays (EMSAs) were performed with all potential promoters to screen direct targets of BioQ. The promoter region of *bioF* and the upstream sequence shared by *bioQ* and *bioB* were found to efficiently bind with BioQ (Fig. S1). And then the promoter regions of *bioF* and *bioQ/B* labeled with fluorescence (5′-FAM and 3′-HEX) were used to precisely determine the BioQ binding sites by DNase I footprinting assays. The results from electropherograms of both strands at various BioQ concentrations showed that the region around TGAACGGTGTTCA in *bioF* promoter was obviously protected (Fig. 6A). This sequence was a perfect inverted repeat sequence which in accordance with the BioQ binding site TGAACN$_3$GTTCA predicted by Brune *et al*. (2012). Similarly, the region around TGAACACCGTTCAAGTAACCTGAACACCGTTCA in the promoter sequence of *bioQ/B* was also bound by BioQ (Fig. 6B). This region contained two palindromic motifs (TGAACACCGTTCA), which further proved TGAACN$_3$GTTCA as the binding site of BioQ experimentally.

According to the sequence analyses, the binding sites of BioQ in the promoter region of *bioF* and *bioB* were overlapped with the transcriptional start sites. Therefore, the binding of BioQ might repress the transcription of *bioF* and *bioB* (Fig. 6A and B). We searched the BioQ binding sites in the genomes of the mycobacteria by Blastn, and six more *Mycobacterium* species were found to contain two or more BioQ-recognizable sites in their genomes (Fig. 2A). Those sites were located in the

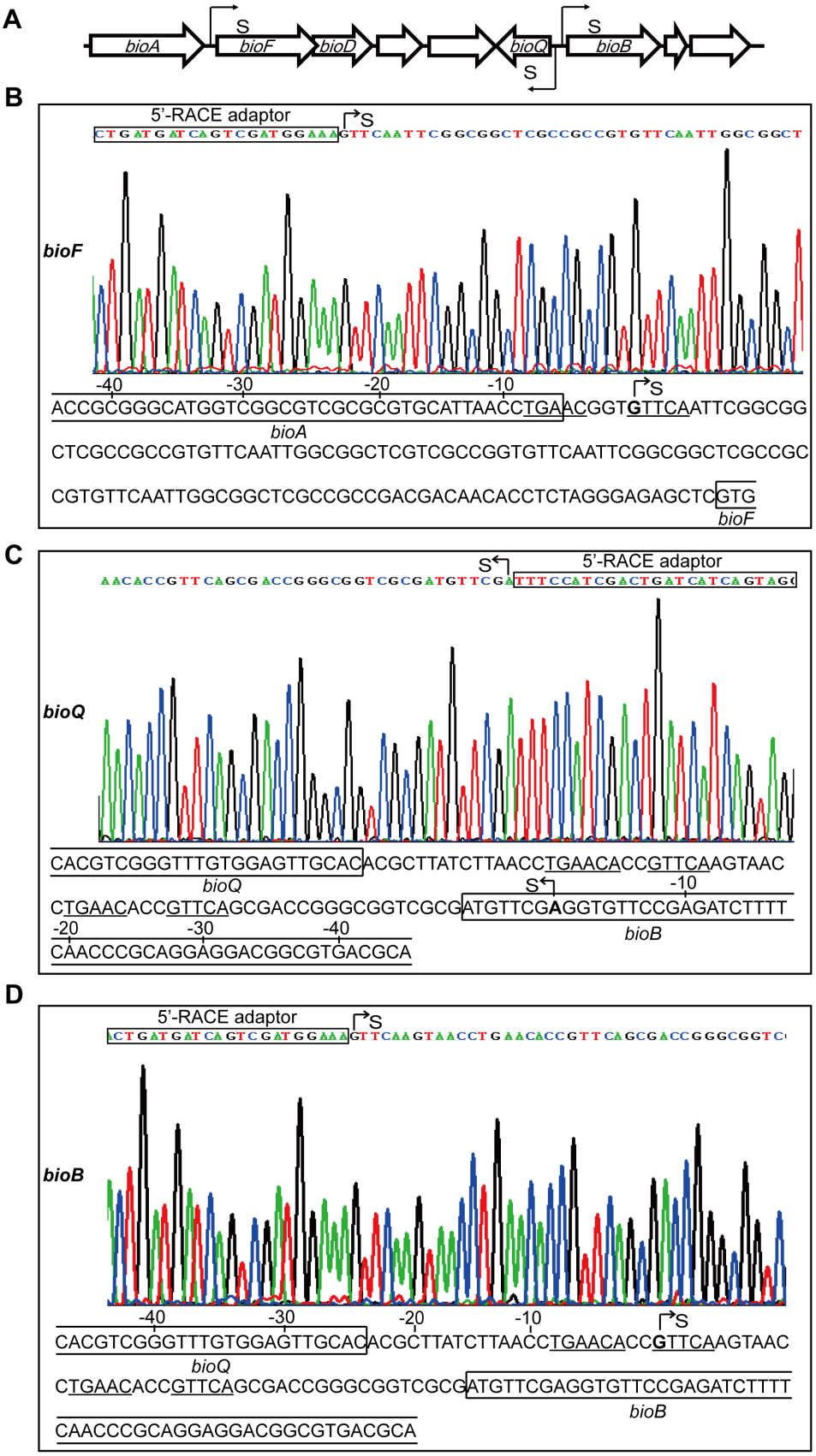

Fig. 5. Mapping the transcription start sites of *bioQB* and *bioF*.
A. Genetic organization of the *bio* operon.
B. 5′-RACE analyses for the *bioF* transcription start site.
C. 5′-RACE-based mapping of the *bioQ* transcription start site.
D. 5′-RACE-based mapping of the *bioB* transcription start site.
Designations: S, transcriptional start site.

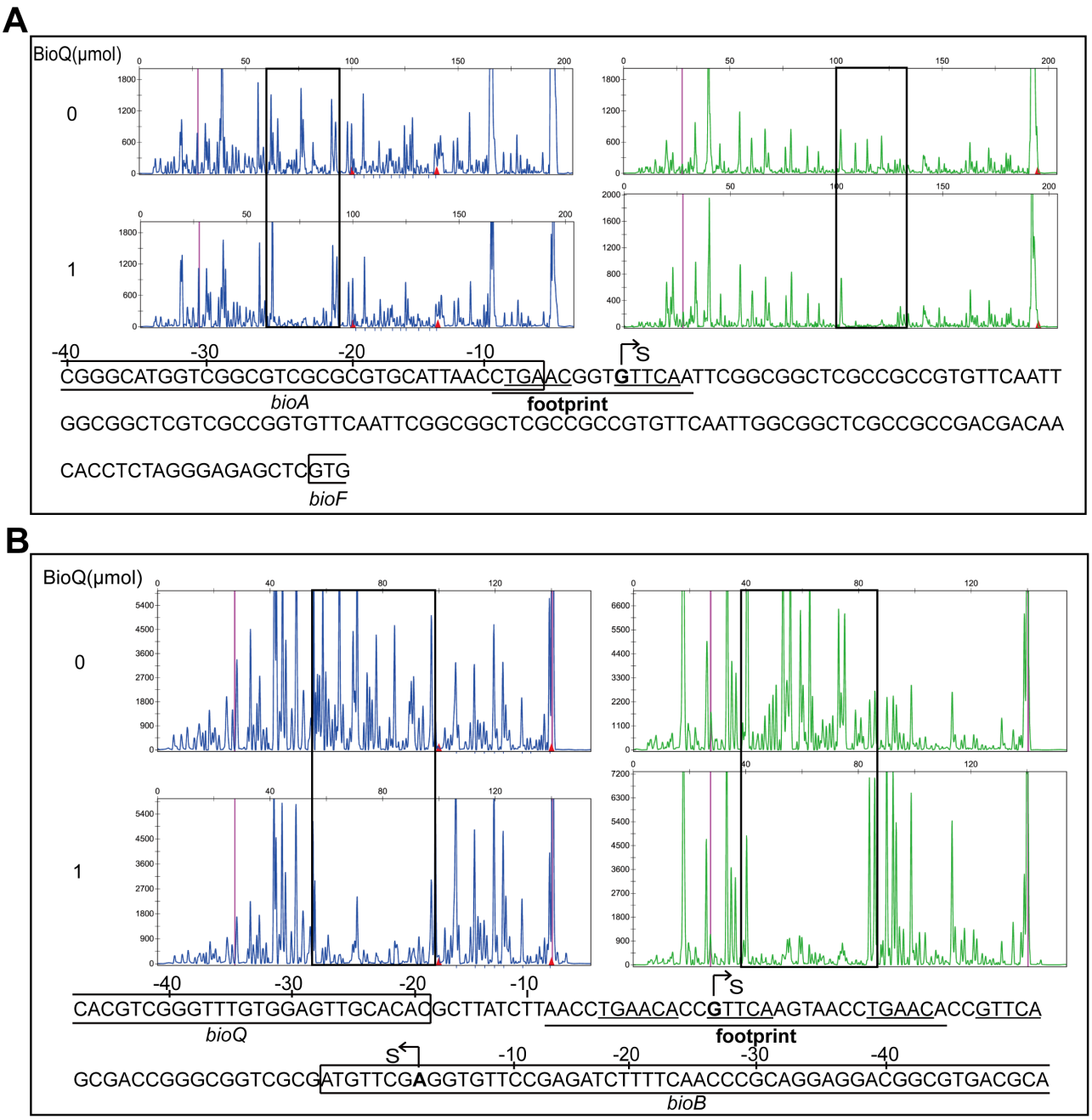

Fig. 6. DNase I foot-printing assays for the *bioF* and *bioQB* loci. Electropherogram-based visualization for protection pattern of *bioF* (A) and *bioQB* (B) promoter after digestion with DNase I following incubation with (in the bottom, 1 μmol) or without (on the top, 0 μmol) BioQ protein. The fluorescence signal of the labeled DNA fragments (6-FAM-labeled sense strand and HEX-labeled antisense strand) is plotted against the sequence length of the fragment. The DNase I foot-printing analysis data of both strands are showed below and the protected region containing inverted repeat (IR) is underlined. The transcriptional start site (S) is indicated in bold.

promoter region of *bioF* or *bioB* without any exception. Multiple sequence alignment showed that the inverted repeats in BioQ binding sites in different *Mycobacterium* species were strictly conserved and the three nucleotides separated the inverted repeats were quite variable (Fig. 7C).

BioQ binding to the promoter regions of biotin synthesis-related genes

Further EMSA assays were conducted to confirm the binding specificity of BioQ with the palindromic motif. The promoter region of *bioF* (designated as *bioF* probe) and

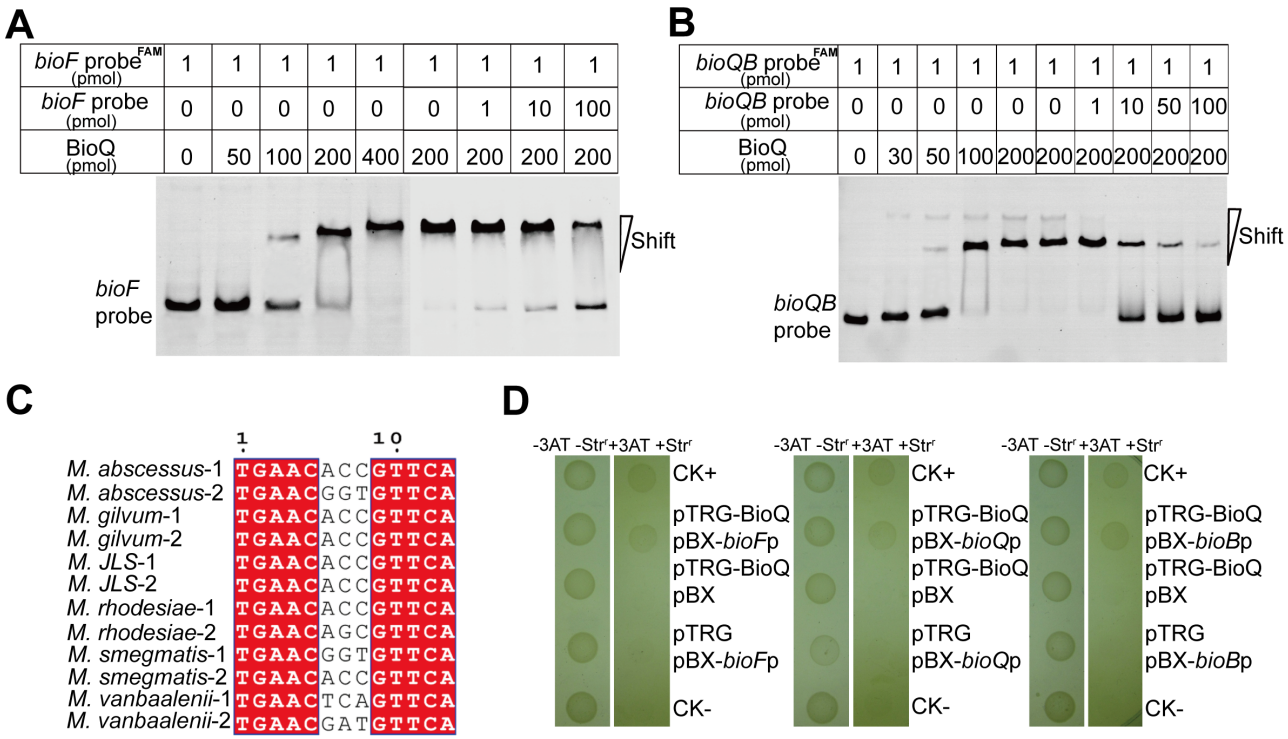

Fig. 7. Interplay between BioQ protein and *bioF/bioQB* operons in *Mycobacterium smegmatis*.
A and B. FAM-labeled *bioF* probe (A) and *bioQB* probe (B) both interact with BioQ protein. DNA probes were incubated with different amounts of BioQ as shown in the above table. Competitive experiments were performed using unlabeled/cold probes.
C. Sequence alignment of the putative BioQ binding sites. White letters shaded in red denote strictly conserved bases.
D. Use of bacterial one-hybrid system to assay the interaction of BioQ with *bioF* and/or *bioQB*. The *bioQ* gene was cloned into pTRG vector and the promoters of *bioF* and/or *bioQB* were cloned into pBX. A pair of pBX/pTRG plasmids was co-transformed into the reporter strain BioQ/*bioFD*p-hybrid, BioQ/*bioQ*p-hybrid and BioQ/*bioB*p-hybrid, and then its growth was monitored together with the self-activation controls on selective medium containing 3-AT, Strep, Kan, Str and Chlr as described in *Experimental procedures*. Each unit represents the corresponding co-transformant in the plates. CK+: co-transformant containing pBX-*Rv2031*p and pTRG-*Rv3133*c as a positive control (Guo *et al.*, 2009). CK–: co-transformant containing pBX and pTRG as a negative control.

bioQ/B (designated as *bioQB* probe) were found to efficiently bind BioQ protein. As shown in Fig. 7A and B, the shifted bands were clearly observed when either *bioF* or *bioQB* probe was incubated with increasing amounts of BioQ. In competition assays, the shifted bands progressively disappeared when fixed amounts of BioQ and labeled probe were incubated with increasing amounts of non-labeled probe. The non-labeled probe competitively bound BioQ to form a DNA–protein complex, resulting in increasing amounts of free labeled probe. These results demonstrated that BioQ can bind to the promoter region of *bioF* and *bioQ*/*B* specifically.

Also, bacterial one-hybrid assays confirmed the specificity of the interactions between BioQ and the promoter regions. Generally, bacterial one-hybrid technique is regarded as an efficient strategy to determine the specific interaction between target DNA sequences and their corresponding transcription factors (Guo *et al.*, 2009). In these assays, the co-transformants with pTRG-*bioQ*/pBX-*bioF*p, pTRG-*bioQ*/pBX-*bioQ*p, pTRG-*bioQ*/pBX-*bioB*p as well as the positive control grew very well in the screening medium. In contrast, the co-transformants with pTRG-*bioQ*/pBX or pTRG/pBX-promoter which were served as self-activated controls and the negative control couldn't grow in screening medium (Fig. 7D). These results indicated that BioQ could specifically bind to the promoter region of *bioF* and *bioQ*/*B* which were cloned into vector pBX, then activated the expression of *HIS3-aadA* reporter cassette resulting in the growth of the co-transformants in the screening medium containing 3-amino-1, 2, 4-triazole (3-AT) and streptomycin.

Asp23 and Arg27 are critical for DNA binding

The TetR regulators usually consist of a DNA-binding domain at the N-terminal and a C-terminal regulatory core domain involved in dimerization and ligand binding (Ramos *et al.*, 2005). Given the multiple sequence alignment and structural modeling of BioQ homologs (Figs 2B and 8A and B), we predicted the putative DNA binding sites located in DNA-binding domain. To identify the amino acid residues essential for DNA binding in BioQ, 6 BioQ mutants

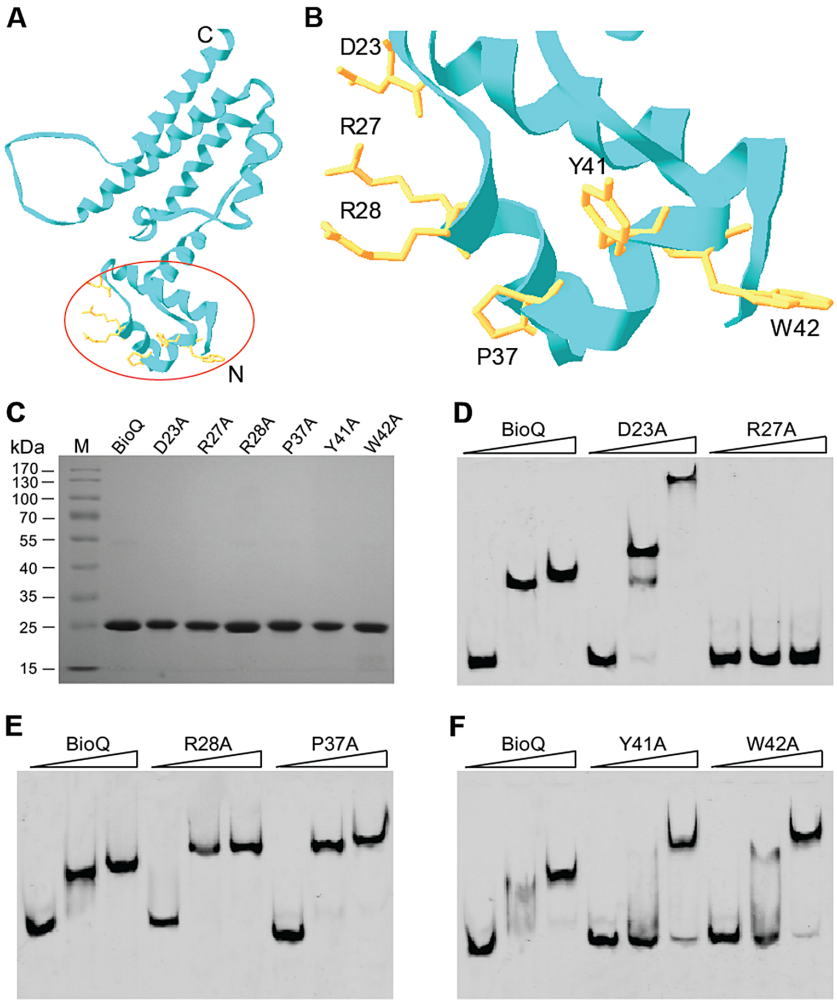

Fig. 8. Determination of critical residues of BioQ for DNA binding.
A. Structural analyses for the predicted DNA-binding residues through structural modeling of *M. smegmatis* BioQ.
B. The enlarged view of the six critical residues of BioQ protein.
C. SDS-PAGE profile of the purified BioQ and its mutants.
D. EMSA experiments revealed that super-shift is observed with the point mutant of BioQ (D23A), whereas no DNA binding ability is detected with the other mutant R27A.
E. The two BioQ mutants (R28A and P37A) exhibited decreased affinity of DNA binding.
F. No apparent role of the two residues (Y41 and W42) in BioQ binding the *bioQB* probe.

(D23A, R27A, R28A, P31A, Y41A and W42A) were conducted by Over-lapping PCR. The hexahistidine-tagged proteins were purified and judged by SDS-PAGE (Fig. 8C), then the purified proteins were adjusted to the same concentration for EMSA assays. The DNA-binding ability of BioQ was fully disrupted when Arginine 27 was replaced with Alanine (Fig. 8D–F). Surprisingly, when the *bioQB* probe was incubated with BioQ-D23A, a super-shifted band was observed at a higher protein concentration (Fig. 8D). It seemed likely that the mutation in Asp23 might promote the protein to form multimer instead of dimer.

BioQ is a functional repressor

To examine the regulatory role of BioQ, we constructed a Δ*bioQ* mutant strain using the strategy of alleic exchange (Fig. S2). A comparison of the expressions of *bioF*, *bioD* and *bioB* genes in both wild-type and Δ*bioQ* mutant mycobacterial strains were conducted using quantitative real-time PCR (qRT-PCR) assays (Fig. 9A). Relative to the wild-type strain, the expression of *bioA*, *bioF*, *bioD* and *bioB* was strikingly upregulated in the Δ*bioQ* mutant strain. However, the expression of *bioA* gene was not affected significantly in that its promoter lacks the BioQ binding site. To verify that the change in expression was result from the deletion of *bioQ*, a functional complementation was constructed by electroporation of the recombinant plasmid pMV261-p-*bioQ* into the Δ*bioQ* strain. As expected, the expression of *bioF*, *bioD* and *bioB* was restored clearly in the complementary strain when compared to the wild-type.

A series of pMV261-borne transcriptional *lacZ* fusions were constructed to confirm the repression of BioQ on the target genes (Fig. S3). As shown in Fig. 9B, the strong promoter *hsp60* strikingly promoted the expression of *lacZ* and the promoterless-*lacZ* plasmid barely showed the activity of β-galactosidase in both the wild-type and the Δ*bioQ* mutant, suggesting that the assays worked well. Compared to wild-type, the *lacZ* expression was sharply promoted under the promoter of *bioFD* in the

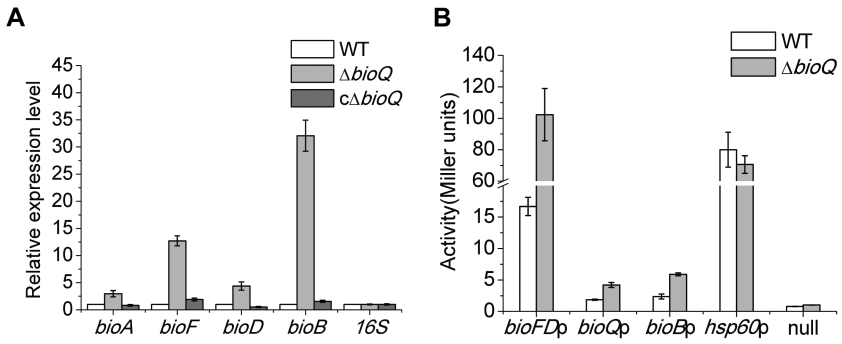

Fig. 9. Repression of *bio* operon by BioQ in *M. smegmatis*.
A. qRT-PCR analyses of differential expression profile of *bio* operon in the Δ*bioQ* mutant in comparison with the wild-type (WT) and the complementary strain (cΔ*bioQ*). The qRT-PCR data were from no less than three independent tests and are expressed in means ± SD.
B. Use of the plasmid-borne *lacZ* transcriptional fusions to investigate the effect of BioQ protein on expression of *bio* operon in *M. smegmatis*. The promoter activity of *bioFD*, *bioQ*, and *bioB* were tested in wild-type and *bioQ*-deleted strain respectively. A truncated LacZ and the one fusion with *hsp60* promoter were set as negative and positive control. The data of β-gal were collected from no less than three independent tests and were expressed in means ± SD.

Δ*bioQ* strain. The promoter of *bioB* showed a moderate activity in the Δ*bioQ* strain but almost had no activity in wild-type strain (Fig. 9B and Fig. S3). This phenomenon probably due to the transcription sites were strongly bound by BioQ in the wild-type strain. In contrast, the transcription could be easily initiated when *bioQ* was knocked out. The *bioQ* promoter was partly overlapped with the *bioB* promoter, resulting in the *bioQ* promoter in the Δ*bioQ* strain also showed a relatively higher activity than in the wild-type. These results strongly proved that BioQ indeed functions as a transcriptional repressor in the context of modulating the expression of biotin synthesis operons.

BioQ senses biotin or a biotin metabolite

Given the scenario seen upon the addition of exogenous biotin at various level into cultures, the transcription of *bioF*, *bioD*, and *bioB* in the wild-type strains were repressed in a concentration-dependent manner. The expression level was dramatically downregulated even at 10 nM biotin, which was expected to be the physiological concentration (Feng *et al.*, 2013a) (Fig. 10A). As expected, biotin synthesis-related genes in *bioQ*-deficient strain showed no apparent altered expression even the level of biotin is as high as 1 mM (Fig. 10B). A moderate decrease of β-galactosidase activity was also observed in the wild-type cultures when the concentration of exogenous supplied biotin gradually increased, whereas the β-galactosidase activity of Δ*bioQ* strain remained unchanged (not shown).

Subsequently, we carried out EMSA assays to probe if the function of BioQ as a repressor requires its binding to biotin or its related-metabolite biotinoyl-5′-AMP (Fig. 11A and Fig. S4). First, we employed Streptavidin with high affinity to biotin for removing the trace mount of biotin bound to BioQ protein. Consequently, we did observe that Streptavidin-treated BioQ protein remains DNA binding activity (Fig. 11A), In fact, the supply of biotin in an increasing amount did not alter interaction of BioQ protein with *bioQB* probe (Fig. S4). It seemed likely that biotin does not play any detectable role in BioQ binding DNA palindromes. Second, we set up BirA-catalyzed enzymatic reaction to synthesize the biotinylated-BCCP protein *in vitro* (Fig. 11B). However, we failed to observe that biotinylated-BCCP protein exert any effect on BioQ interplay with cognate DNA (Fig. 11B). Third, we employed the almost

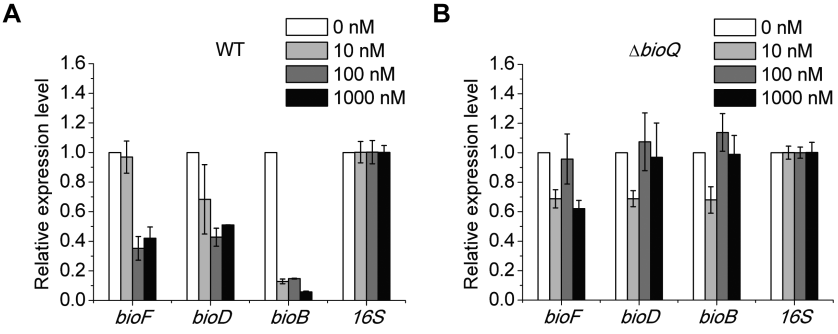

Fig. 10. Repression of expression of *bio* operon by exogenous addition of biotin and depends on presence of BioQ. Real-time qPCR analyses for effect of biotin on expression of *bioF*, *bioD*, *bioB* expression in the wild-type strain (A) and the Δ*bioQ* mutant (B). The wild-type strain was grown in HdB media supplemented with different concentrations of biotin. qRT-PCR experiments were performed in no less than three times and the data were expressed in means ± SD.

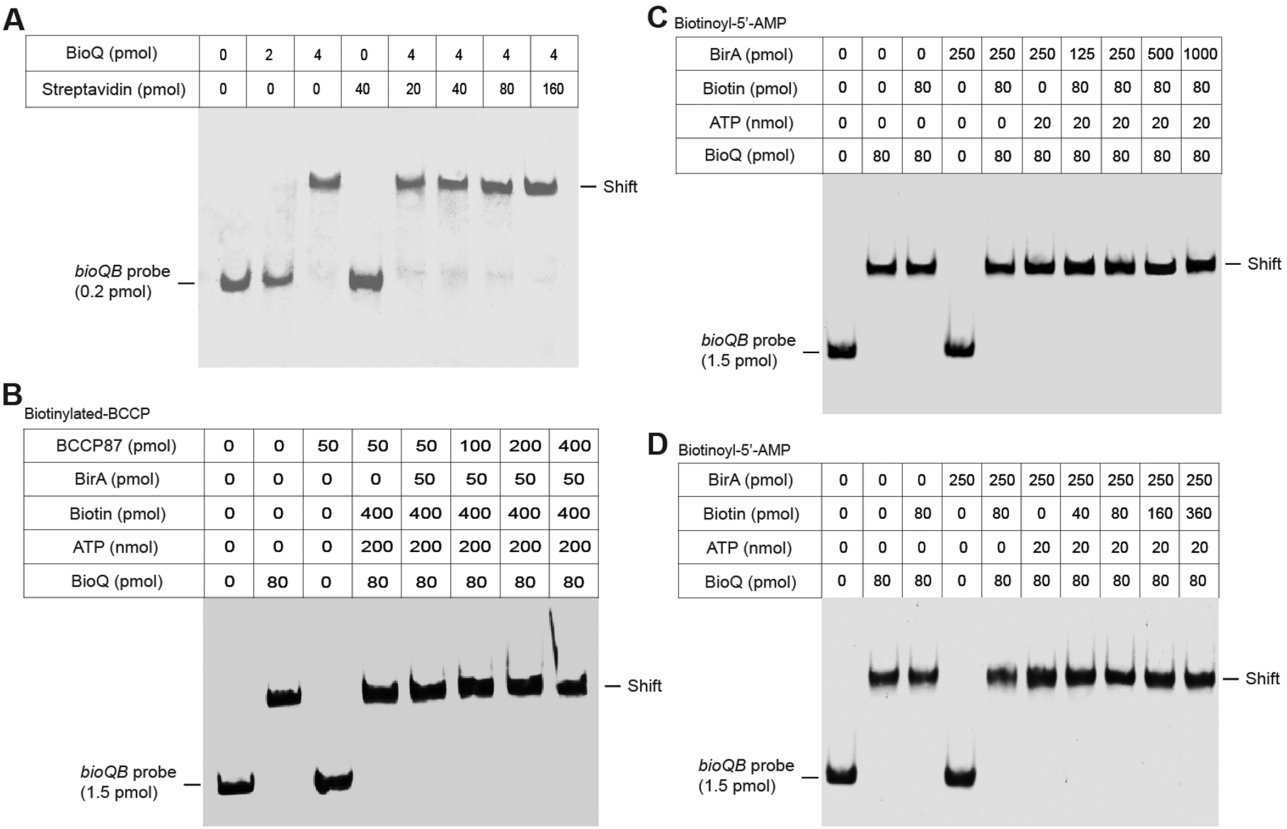

Fig. 11. Probing of possible roles of biotin/biotinoyl-5′-AMP/biotinylated-BCCP in BioQ binding cognate DNA.
A. Streptavidin-treated BioQ protein remains DNA binding activity. To rule out the possibility of biotin contamination in present during the protein preparation, BioQ protein was incubated with streptavidin at various levels for 30 min. In EMSA assays, 0.2 pmol *bioQB* probe was supplied.
B. Biotinylated-BCCP does not interfere BioQ binding cognate DNA. In panel B, various amounts of *E. coli* BCCP87 protein was incubated with BirA (50 pmol), biotin (400 pmol) and ATP (nmol) for 30 min to *in vitro* synthesize biotinylated BCCP.
C and D. EMSA-based evaluation of potential effects of biotinoyl-adenylate on BioQ binding activity. In panel C, various amounts of BirA was incubated with biotin (80 pmol) and ATP for 30 min to *in vitro* synthesize biotinoyl-5′-adenylate (biotinoyl-5′-AMP). In panel D, various amounts of biotin was incubated with BirA enzyme (250 pmol) and ATP for 30 min to produce biotinoyl-5′-AMP. In EMSA experiments, 1.5 pmol *bioQB* probes were incubated with 80 pmol BioQ protein.

identical enzymatic reaction to prepare the intermediate biotinoyl-5′-AMP, in which the level of either biotin (Fig. 11C) or BirA protein (Fig. 11D) is fixed. The EMSA assay coupled with the above enzymatic reaction revealed that biotinoyl-5′-AMP has not any obvious role in direct binding of BioQ protein to cognate DNA (Fig. 11C and D). Probing of possible roles of biotin/in BioQ binding target DNA.

Very recently, we determined a new regulator BioR, which is the *bio* operons repressor in α-proteobacteria (Feng *et al.*, 2013a). Extensive investigation allows us to exclude the possibility that the dethiobiotin, the precursor of biotin, as the ligand of BioR (Feng *et al.,* 2013a). Therefore, we concluded that BioQ could regulate the expression of its target genes upon sensing biotin indirectly, and the BioQ regulatory ligand might be one of its precursors or degradation products.

Discussion

The data showed here defined a new regulatory mechanism for bacterial biotin metabolism in which the two major players included the biotin protein ligase BirA and the newly identified BioQ repressor (Fig. 1). Somewhat it is similar to the two-protein paradigm of BirA and BioR we recently proposed (Feng *et al.*, 2013a,b). It seemed likely that the two-protein paradigm of BirA and BioR is exclusively restricted to α-proteobacteria (Rodionov and Gelfand, 2006; Feng *et al.*, 2013a,b), whereas the other model of BirA-BioQ only belongs to *Actinobacteria* (Brune *et al.*, 2012). In the above two models, the two BirA proteins of different origins are mono-functional BPL. Clearly, the loss of BirA regulatory roles was compensated by BioR and BioQ respectively. Unlike the BioR with an assignment into a family of GntR-type transcription factor, BioQ belongs to

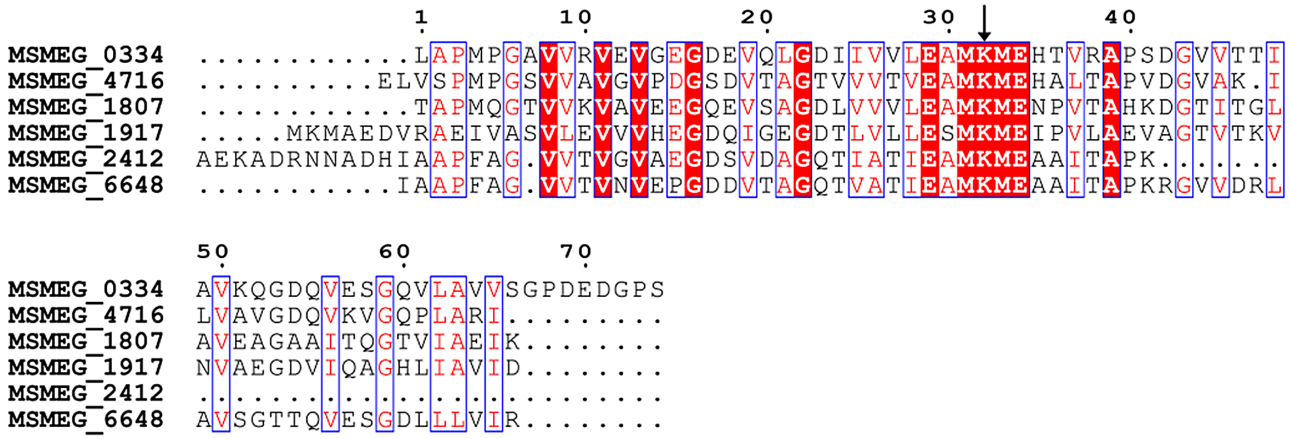

Fig. 12. Sequence comparison of the predicted biotinylated domains of six putative biotin-dependent enzymes from *M. smegmatis*. Genome searches indicated that six putative biotin-dependent enzymes included MSMEG_1807 (acetyl-/propionyl-coenzyme A carboxylase subunit alpha), MSMEG_4716 (acetyl-/propionyl-coenzyme A carboxylase subunit alpha), MSMEG_1807 (acetyl-/propionyl-coenzyme A carboxylase subunit alpha), MSMEG_1917 (acetyl-CoA carboxylase biotin carboxyl carrier protein subunit), MSMEG_2412 (pyruvate carboxylase), and MSMEG_6648 (pyruvate carboxylase) respectively. The conserved 'MKME' motif where 'K' is the predicted site of biotin attachment indicated with an arrow.

a group of the TetR-like regulators (Brune *et al*., 2012). Although *bioR* can duplicate in some cases, like *Paracoccus denitrificans* (Feng *et al*., 2013a), we failed to observe the similar scenario in any species of *Actinobacteria* (Brune *et al*., 2012). Although the BioQ box (BioQ recognizable site) seemed pretty conserved (TGAACN$_3$GTTCA) in the phylum of *Actinobacteria*, the number of this site varies significantly in different species (Brune *et al*., 2012). Even in the species of *Mycobacterium* we examined, the organization of *bio* operons and its regulatory signals are not fully identical (Fig. 2). Undoubtedly, our observation further validated the unexpected complexity and diversity of regulatory mechanisms in the context of biotin metabolism.

Here we prepared the recombinant protein of *M. smegmatis* BioQ, and reported its solution structure of dimer. EMSA-based experiments proved its specificity of BioQ protein to all the predicted DNA sites we examined (Fig. 7). Also, our *in vivo* data from real-time PCR combined with β-gal assay of the transcriptional *lacZ* fusion demonstrated that *M. smegmatis* BioQ is a functional repressor (Fig. 9). Given the fact that *M. smegmatis bioQ* gene is different from its counterpart of *C. glutamicum* in that the former has a BioQ box, while not for the latter, we thus believed that *M. smegmatis* BioQ is an auto-regulator, whereas *C. glutamicum bioQ* itself is unregulated. Considering the functional replacement of the other *bio* repressor BioR with different origins (Feng *et al*., 2013a,b), it would be of much interest to test whether *M. smegmatis* BioQ protein and its recognizable signals are functionally exchanged with the cousins of *C. glutamicum*. Although that the biotin uptake route of *C. glutamicum* is negatively regulated by BioQ in that *bioYMN* operon contained a functional BioQ box, *M. smegmatis* seemed to be not the case which is due to lacking of biotin scavenging pathway (Figs 1 and 2). Even in the other species of mycobacteria remaining a *bioYMN* operon, the BioQ box also is cryptic. The fact that the biotin uptake path is exempt/away from BioQ-mediated regulation in such cases, probably represent a physiological requirement/advantage during the long-term evolution under the pressure of environmental selection. Among all the seven species of mycobacteria we checked (Fig. 2), three of four genes that claimed assembling the double rings of biotin molecule (*bioB* and *bioFD*) are consistently found to retain the BioQ box, whereas *bioA* not. During the four-step process of closing the double rings of biotin molecule, BioF is catalyzed the first step, BioA mediates the second step, whereas the BioB is responsible for the last reaction. Probably, in this case, the regulated expression of at least the enzymes involved in the first/last committed steps is enough to ensure appropriate production of biotin. Genome-wide mining revealed that six putative biotin-dependent enzymes [MSMEG_1807 (acetyl-/propionyl-coenzyme A carboxylase subunit alpha), MSMEG_4716 (acetyl-/propionyl-coenzyme A carboxylase subunit alpha), MSMEG_1807 (acetyl-/propionyl-coenzyme A carboxylase subunit alpha), MSMEG_1917 (acetyl-CoA carboxylase biotin carboxyl carrier protein subunit), MSMEG_2412 (pyruvate carboxylase), and MSMEG_6648 (pyruvate carboxylase)] are present in *M. smegmatis* (Fig. 12). It is somewhat unexpected, but not without precedent in that the similar scenario was recently seen in the case of *Agrobacterium* (Feng *et al*., 2013b). In contrast, the unregulated expression of *M. smegmatis bioA* is opposite to the BioQ-regulated expression of *bioAD* in *C. glutamicum*. It would be of particular note that although

the *bio* operons encoding the biotin synthesis pathway is present, no *bioQ* homolog can be detected in the tuberculosis-causing pathogen, *M. tuberculosis*. We anticipated it might be in part the patho-physiological consequence for *M. tuberculosis* pathogen to adapt/survive in the severely stressful condition of its infected human host. Given the fact that the availability of trace/limited host-derived biotin excluded the necessity of BioQ-mediated regulation of biotin, it is rational that *M. tuberculosis* un-lock its biotin synthesis machinery to satisfy the basal requirement for biotin with the destination to protein biotinylation of central metabolic enzymes. Hopefully, it could act as a physiological advantage for bacterial survival in the host environment.

Although it is well known that biotinoyl-5′-AMP is small molecule effector for BirA, the paradigm regulator of biotin synthesis (Beckett, 2007), no evidence accumulated can answer the question whether the newly identified BioR repressor needs some physiological ligands to exert its regulatory effects (Feng *et al*., 2013a,b). The same issue also remains enigmatic in the context of BioQ-mediated regulation. In fact, we had no success in 3 years of search for possible physiological ligand through testing a series of metabolite intermediates of biotin biosynthesis (Fig. 11). Of particular note, exogenous addition of biotin does not given any obvious effect on BioQ binding target DNA sites (Fig. 11 and Fig. S4), which is quite similar our recent observation seen with BioR regulator (Feng *et al*., 2013a). We also failed to note any effects of biotinylated-BCCP/biotinoyl-5′-AMP exerted on BioQ binding cognate DNA (Fig. 11B–D). We are quite confident that resolving crystal structures of BioQ protein alone and its complex with DNA target could represent a straight-forward way to gain a glimpse of this unanswered question in the near future. Thanks to the excellent performance of BioQ protein *in vitro,* we have luck to almost arrive at the end of the way to this question. However we still have not taken the same fortune to handle the weird BioR protein in that it easily precipitates *in vitro* during the process of preparation (Feng *et al*., 2013a).

Experimental procedures

Bacterial strains and growth conditions

The bacteria used here included *E. coli* strains and *Mycobacterium* (Table 1). *M. smegmatis* MC^2 155 was grown at 37°C in Middle brook 7H9 broth (BD, USA) or Hartmans-de Bont (HdB) medium [30 μM EDTA, 500 μM $MgCl_2$, 7 μM $CaCl_2$, 0.8 μM, $NaMoO_4$, 1.68 μM $CoCl_2$, 5.49 μM $MnCl_2$, 6.95 μM $ZnSO_4$, 20 μM $FeSO_4$, 0.8 μM $CuSO_4$, 6 μM K_2HPO_4, 6 μM NaH_2PO_4, 15 mM $(NH_4)_2SO_4$, pH 7.0] (Smeulders *et al*., 1999) supplemented with 0.2% gycerol and 0.05% Tween 80. *E. coli* DH5α and *E. coli* BL21(DE3) were grown in Luria–Bertani (LB) medium at 37°C, and *E. coli* XR in Bacterial one-hybrid assays was grown in M9 minimal media containing 47.7 mM $Na_2HPO_4{\cdot}7H2O$, 22 mM KH_2PO_4, 8.56 mM NaCl, 18.7 mM NH_4Cl and 111 mM glucose. If required, antibiotics were added as follows: tetracycline, 25 μg ml^{-1}; chloramphenicol, 25 μg ml^{-1}; hygromycin, 100 μg ml^{-1}; and kanamycin, 50 μg ml^{-1}.

DNA manipulations

Overlap PCR was used for site-directed mutagenesis of *bioQ* gene using the designed primers (Table 2) (Ho *et al*., 1989). The corresponding genes (wild-type and its mutant versions) were cloned into pET28a via BamHI and HindIII restriction sites and transformed into *E. coli* DH5α and *E. coli* BL21 (DE3). All the constructs were verified by direct DNA sequencing.

Expression, identification and characterization of BioQ protein

The hexahistidine-tagged protein was separated and purified using Ni-NTA affinity column, then dialyzed overnight to remove the residual imidazole. The purity was judged by 12% SDS-PAGE and quantified by Coomassie Brilliant Blue assay (Yang *et al*., 2012b). The identity of the acquired proteins was determined using the liquid chromatography quadrupole time-of-flight mass spectrometry (Feng and Cronan, 2011b; Feng *et al*., 2013b). The solution structure of the recombinant protein was verified using gel filtration (Feng and Cronan, 2011b), analytical ultracentrifugation and chemical cross-linking (Feng and Cronan, 2010; 2011b).

Electrophoretic mobility shift assays

DNA probes were amplified using PCR with specific fluorescence-labeled primers (Table 2) from *M. smegmatis* genomic DNA and purified with a Nucleic Acid Purification kit (Axygen). Labeled probes were incubated with different amounts of BioQ proteins in EMSA buffer (50 mM Tris-HCl, pH 7.5; 10 mM $MgCl_2$; 1 mM DTT; and 100 mM NaCl) at 25°C for 2 h. The mixtures were then subjected to 6% native polyacrylamide gel electrophoresis at 150 V for 1 h. Unlabeled cold probes used for competitive experiments were amplified from *M. smegmatis* genomic DNA (Table 2) and purified as described previously. The labeled and unlabeled probes in different ratio were mixed with BioQ in EMSA buffer, then incubated at room temperature for 2 h and separated by 6% polyacrylamide gel electrophoresis. Images of gels were obtained using Typhoon Scanner (GE Healthcare).

DNase I foot-printing assays

The fluorescence-labeled probes and the reaction system were the same as in EMSA. The mixtures were treated with DNase I (0.25 unit, Fermentas) at 25°C for 5 min. The reaction was stopped by adding 0.25 M EDTA and incubating in water bath at 75°C for 15 min. The digested DNA fragments were purified with the Nucleic Acid Purification kit (Axygen) and eluted in 50 μl distilled water. The purified DNA was mixed with HiDi formamide (Applied Biosystems) and GeneScan-500 LIZ size standards (Applied Biosystems),

Table 1. Strains and plasmids used in this study.

Bacteria or plasmids	Relevant characteristics	Origins
Bacterial strains		
E. coli DH5α	A cloning host	Lab stock
DH5α (λ-*pir*)	Δ*lac* host for pAH125 and its derivatives	Feng and Cronan (2009), Haldimann and Wanner (2001)
Topo10	F^-, Δ*lac*X74, a cloning *h*ost for recombinant plasmids	Invitrogen
E. coli BL21(DE3)	Protein expression host	Lab stock
FYJ256	BL21(DE3) with pET28a-*birAat*	Feng *et al.* (2013b)
BL21-pET-*bioQ*	BL21(DE3) with pET28a-*bioQ*	This work
BL21-pET-*bioQ*-D23A	BL21(DE3) with pET28a-*bioQ*-D23A	This work
BL21-pET-*bioQ*-R27A	BL21(DE3) with pET28a-*bioQ*-R27A	This work
BL21-pET-*bioQ*-R28A	BL21(DE3) with pET28a-*bioQ*-R28A	This work
BL21-pET-*bioQ*-P37A	BL21(DE3) with pET28a-*bioQ*-P37A	This work
BL21-pET-*bioQ*-Y41A	BL21(DE3) with pET28a-*bioQ*-Y41A	This work
BL21-pET-*bioQ*-W42A	BL21(DE3) with pET28a-*bioQ*-W42A	This work
E. coli XR	A recipient strain of *E. coli* for bacterial single-hybrid system	Stratagene
BioQ/*bioFD*p-hybrid	XR with pTRG-*bioQ* and pBX-*bioFD*p	This work
BioQ/*bioQ*p-hybrid	XR with pTRG-*bioQ* and pBX-*bioQ*p	This work
BioQ/*bioB*p-hybrid	XR with pTRG-*bioQ* and pBX-*bioB*p	This work
pTRG/*bioFD*p-hybrid	XR with pTRG and pBX-*bioFD*p, self-activated control	This work
pTRG/*bioFD*p-hybrid	XR with pTRG and pBX-*bioQ*p, self-activated control	This work
pTRG/*bioFD*p-hybrid	XR with pTRG and pBX-*bioB*p, self-activated control	This work
BioQ/pBX-hybrid	XR with pTRG-*bioQ* and pBX, self-activated control	This work
pTRG/pBX-hybrid	XR with pTRG and pBX, negative control	This work
*Rv2031*p/*Rv3133c*-hybrid	XR with pBX-*Rv2031*p and pTRG-*Rv3133c*, positive control	Guo *et al.* (2009)
M. smegmatis MC^2155	A wild-type strain	ATCC
Δ*bioQ*	Δ*bioQ* mutant of *M. smegmatis* MC^2155	This work
Δ*bioQ*/pMV261	Δ*bioQ* mutant with pMV261	This work
Δ*bioQ*/pMV261-p-*bioQ*	The complementary strain of Δ*bioQ*	This work
MC^2155/pMV261	*M. smegmatis* MC^2155 with pMV261	This work
Plasmids		
pET28a(+)	T7-driven expression vector, Km^R	Novagen
pET28a-*bioQ*	*bioQ* in BamHI and HindIII sites of pET28a	This work
pET28a-*bioQ*-D23A	*bioQ*-D23A BamHI and HindIII sites of pET28a	This work
pET28a-*bioQ*-R27A	*bioQ*-R27A BamHI and HindIII sites of pET28a	This work
pET28a-*bioQ*-R28A	*bioQ*-R28A BamHI and HindIII sites of pET28a	This work
pET28a-*bioQ*-P37A	*bioQ*-P37A BamHI and HindIII sites of pET28a	This work
pET28a-*bioQ*-Y41A	*bioQ*-Y41A BamHI and HindIII sites of pET28a	This work
pET28a-*bioQ*-W42A	*bioQ*-W42A BamHI and HindIII sites of pET28a	This work
pBX	Chl^R, p15A replicon, lac-UV5 promoter, used for bacterial one-hybrid assays	Guo *et al.* (2009)
pBX-*bioFD*p	*bioFD*p in EcoRI and XbaI sites of pBX	This work
pBX-*bioQ*P	*bioQ*p in EcoRI and XbaI sites of pBX	This work
pBX-*bioB*p	*bioB*p in EcoRI and XbaI sites of pBX	This work
pTRG	Tet^R, ColE1 replicon, lpp/lac-UV5 promoter, used for bacterial one-hybrid assays	Guo *et al.* (2009)
pTRG-*bioQ*	*bioQ* in EcoRI and XbaI sites of pTRG	This work
pMV261	Kan^R, pAL5000 replicon, colE1 replicon, *hsp60* promoter, expression vector	Guo *et al.* (2009), Yang *et al.* (2012b)
pMV261-p-*bioQ*	*bioQ* and its promoter in XbaI and NheI sites of pMV261	This work
pMV261-null-*lacZ*	*lacZ* in XbaI and NheI sites of pMV261	This work
pMV261-hsp60-*lacZ*	*lacZ* in HindIII and NheI sites of pMV261	This work
pMV261-*bioFD*p-*lacZ*	*lacZ* with the promoter of *bioFD* in XbaI and NheI sites of pMV261	This work
pM261-*bioQ*p-*lacZ*	*lacZ* with the promoter of *bioQ* in XbaI and NheI sites of pMV261	This work
pM261-*bioB*p-*lacZ*	*lacZ* with the promoter of *bioB* in XbaI and NheI sites of pMV261	This work
pMind	Kan^R, Hyg^R, pAL5000 replicon ,colE1 replicon, used for gene-knockout	Parish and Stoker (2000)
pGoAL17	Amp^R, pBR322 replicon, used for gene-knockout	Blokpoel *et al.* (2005)
pMind-*bioQ*U	Intermediate vector in gene-knockout experiments	This work
pMind-*bioQ*UD	Intermediate vector in gene-knockout experiments	This work
pMInd-*bioQ*UD-*sacB*-*lacZ*	Suicide plasmid used for gene-knockout	This work
pMD19-T simple vector	Intermediate cloning vector	Takara

then assayed with Applied Biosystems 3730XL DNA analyzer (manufactured by Tsingke Company, Wuhan). Electropherograms were analyzed using the GENEMAPPER software (Applied Biosystems).

Bacterial one-hybrid assays

A bacterial one-hybrid reporter system which is based on the *HIS3-aadA* reporter cassette was applied to investigate the

Table 2. Primers used in this study.

Primers	Primer sequences (5′–3′)
pBX-*bioFD*p-F	CCGGAATTCTTCATCTGCACGCCCGAGGA
pBX-*bioFD*p-R	TGCTCTAGAGAGCTCTCCCTAGAGGTGTTGT
pBX-*bioQ*p-F	CCGGAATTCCTGGTCGAGGCCGACGCCGCG
pBX-*bioQ*p-R	TGCTCTAGAACGCTTATCTTAACCTGAACACCGTTC
pBX-*bioB*p-F	CCGGAATTCTCGTCAGGTCGGCGATGCCGTGGTCGTC
pBX-*bioB*p-R	TGCTCTAGACGCGACCGCCCGGTCGCTGAACGGTGTT
pTRG-*bioQ*-F	CGGGATCCGTGCAACTCCACAAACCCGACGT
pTRG-*bioQ*-R	TGCTCTAGACTACTCGAAAGAGCGCTCCGCAGAA
1-F	GCACACATCTGGCATCCCTA
1-R	AAGAACACCGTGTCCAGACC
2-F	TCCAACGACTATCTGGGCCT
2-R	GGTGTAGCCCGACGAAAAGA
3-F	CTCAACCACACCGCATTGAC
3-R	CTCAGGCTCGTGATCCAGTC
4-F	CACTCCGTCGAACTCGTCTT
4-R	CAGGATCTGTTCGACCTCGG
5-F	GTCGAGCAATCGAAACCGTC
5-R	ATCAACGCCTCGTAGGTCTG
6-F	GCGCTCCGCAGAAGAAAATC
6-R	ACGATCAGTCGATGCTCACC
7-F	TGCTCGCAGTCAGGACTTTT
7-R	TCCAGGTTGTGGTTGTAGCG
8-F	TGGTGATGATCGAGGAAACCC
8-R	GCCGGTGTAGACGTTGTAGA
9-F	GTTCATGTTCACTGGCCTGC
9-R	GATTCACGGGCGGATAGACC
10-F	GCGCTCGCGCACGGGGTGTGGC
10-R	GTCGGGGTGCTGCGACAGGCCCA
11-F	GTGCGCGTGGGCTGCTTCCGCCC
11-R	GTGCCAGGGGCTCGGGGTAACGCCA
12-F	CGCTCGCGCCAGTGCGTGCCGTGC
12-R	GTTCGACCTCGGCGTCGATGTGCTGGGC
13-F	GCCGCGTCGCCCCGACGCGAC
13-R	GCCGCGACGGTTTCGATTGCTCGACCGC
14-F	AGGGCATCCTCGGCGGCATCAACGC
14-R	CGATCATCCGGCGGCCGCACTCG
15-F	AACGTCTACACCGGCGGTCCGGCG
15-R	CGTGGATCGGGGGCGCCAGGAC
16S-F	AGGGTCCGAGCGTTGTC
16S-R	GCTCGTTGCGGGACTTA
bioQ-U-F	CCTTAATTAAATGCCTGCCTCGCGGACCATGTTGA
bioQ-U-R	CTAGACTAGTACGCTTATCTTAACCTGAACACCGTTCAAG
bioQ-D-F	CCCAAGCTTTCACGGCATACCGCGGGAGACCATGGCAT
bioQ-D-R	CTAGCTAGCGTGGCATCTGACCGAGGTGCACCTGTTCT
PbioQ-F	CTAGTCTAGAGTGGGCCAGCGAGAGCAGG
bioQ-NheI-R	ATGGCTAGCCTAGTGGTGGTGGTGGTGCTCGAAAGA GCGCTCCGCAG
bioQB-FAM-F	AGGATCTTCGTCGCCGCA
bioQB-HEX-R	GTTGAAAAGATCTCGGAACACC
bioF-FAM-F	TTCATCTGCACGCCCGAGGA
bioF-HEX-R	GAGCTCTCCCTAGAGGTGTTGT
5′-RACE Outer	CATGGCTACATGCTGACAGCCTA
5′-RACE Inner	CGCGGATCC ACAGCCTACTGATGATCAGTCGATG
bioF-race-outer	AGGTGCGCAATGCCTCGACG
bioF-race-inner	ACCCGCAGTTCACGGCGCA
bioQ-race-outer	CGCAGAATGTGGTCGGCCAC
bioQ-race-inner	CCAACAGTTCCTGTTTGTTGGCG
bioB-race-outer	AGCGAGAGCAGGTCGTCGAG
bioB-race-inner	CGCGCCACGTCCAGTACATCT
bioQ-F	CGCGGATCCGTGCAACTCCACAAACCCGACGTGGT
bioQ-R	CCCAAGCTTCTACTCGAAAGAGCGCTCCGCAGAAGA
bioQ-D23A-F	ACCACGGCATCGCCGCCCTGACGATGCGG
bioQ-D23A-R	CCGCATCGTCAGGGCGGCGATGCCGTGGT
bioQ-R27A-F	GACCTGACGATGGCCCGCCTGGCGCGC
bioQ-R27A-R	GCGCGCCAGGCGGGCCATCGTCAGGTC
bioQ-R28A-F	CTGACGATGCGGGCCCTGGCGCGCGAG
bioQ-R28A-R	CTCGCGCGCCAGGGCCCGCATCGTCAG
bioQ-P37A-F	AGCTCGACGTCACGGCGGGCGCCCTGTACT
bioQ-P37A-R	AGTACAGGGCGCCCGCCGTGACGTCGAGCT
bioQ-Y41A-F	CCGGGCGCCCTGGCCTGGCATTTCGCCAA
bioQ-Y41A-R	TTGGCGAAATGCCAGGCCAGGGCGCCCGG
bioQ-W42A-F	CCGGGCGCCCTGTACGCGCATTTCGCCAACAAA
bioQ-W42A-R	TTTGTTGGCGAAATGCGCGTACAGGGCGCCCGG

protein–DNA interactions as described by Guo *et al*. (2009). The *bioQ* gene and the promoter regions of *bioFD*, *bioQ*, *bioB* were cloned into the modified pTRG and pBX plasmid respectively. A pair of pTRG/pBX plasmids was co-transformed into *E. coli* XR and spotted onto M9 agar plates with appropriate antibiotics. 20 mM 3-amino-1, 2, 4-triazole and 25 μg ml^{-1} streptomycin were added to screen the co-transformant in which the DNA–protein interaction triggered the transcriptional activation of the *HIS3-aadA* reporter gene in pBX. A co-transformant with pBX-*Rv2031*p plus pTRG-*Rv3133c* was used as a positive control (Guo *et al*., 2009), whereas the transformant carrying the empty vector pBX and pTRG was used as a negative control. The plates were incubated at 30°C for 2–3 days.

Construction of the ΔbioQ mutant and functional complementation

The knockout strain of the *M. smegmatis bioQ* (*MSMEG_3193*) gene was conducted using the method of homologous recombination as Yang and co-workers described (Yang *et al*., 2012b). A suicide plasmid was constructed by cloning the upstream and downstream fragments' arm of *bioQ* gene into pMind plasmid flanking the hygromycin resistance gene, and a *sacB-lacZ* cassette was inserted as a selection marker. The recombinant plasmid was electroporated into competent cells of *M. smegmatis* and plated on Middle brook 7H10 medium containing 100 μg ml^{-1} hygromycin, 10% sucrose and 1 mg ml^{-1} X-gal. The allelic-exchange mutant strain was determined by multiplex-PCR and direct DNA sequencing. The DNA fragment covering the coding sequence and its promoter region of *bioQ* were amplified by PCR and cloned into pMV261 plasmid to create the expression plasmid pMV261-p-*bioQ*. The plasmid was then electroporated into the Δ*bioQ* mutant, and its expression in the complementary strain was detected using both RT-PCR and Western blot.

Analysis of β-galactosidase activity

β-Gal activity was assayed with *M. smegmatis* carrying the plasmid pMV261-borne transcriptional *lacZ* fusions (Yang *et al*., 2014). The positive control is the pMV261 plasmid containing the *hsp*60 promoter-driven *lacZ* gene. The reporter plasmids are derivatives of pMV261 plasmid in which the *hsp60* promoter was replaced by the promoter regions of either *bioFD*, *bioQ* or *bioB.* The negative control is the promoter-less *lacZ* gene (Fig. S3).These reporter plasmids were transformed into the wild-type and/or Δ*bioQ* mutant giving the reporter strains. All strains were grown in HdB broth containing 0.05% Tween-80 and 50 μg ml^{-1} kanamycin at 37°C for 24 h (Mid-log phase), and biotin at various level was supplied where indicated. The collected culture was assayed for β-gal activity (Bharati *et al*., 2013).

RNA isolation, RT-PCR and real-time PCR

Mid-log-phase cells of *M. smegmatis* grown in either Middle brook 7H9 broth or HdB medium were prepared to isolate

total RNA using TRIzol reagent (life technologies, USA). The final total RNA was analyzed by 1% agarose gel electrophoresis and quantified by NanoDrop 2000 (Thermo Scientific, USA). First-strand cDNAs were synthesized using PrimeScript RT reagent Kit with gDNA Eraser (Takara Biotechnology, Japan) according to the manufacturer's instructions. The final cDNAs were diluted and served as template for PCR amplification of the biotin synthesis-related genes using specific primers (Table 2) on the Bio-RAD Thermal Cycler. The PCR products were separated on 1% agarose gel electrophoresis. For real-time PCR analysis, gene-specific primers (Table 2) were used and cDNA were the same in RT-PCR experiments. Each PCR reaction (10 μl) contained 5 μl of 2× SYBR Green Master Mix Reagent (Applied Biosystems), 0.65 μl diluted cDNA sample, 150 nM gene-specific primers and 5 μl distilled water. The reactions were performed in an ABI ViiA™ 7 real-time PCR machine using following thermo-cycling condition: 95°C for 10 min, and 40 cycles at 95°C for 15 s, 60°C for 15 s and 72°C for 30 s. The *16S* rRNA gene was served as internal reference and the degrees of expression change were calculated using the $2^{-\triangle\triangle Ct}$ method.

5′-RACE

To determine the transcription start site, 5′-RACE analysis was performed using a tobacco acid pyrophosphatase-based 5′-RACE kit (Takara Biotechnology, Japan). TAP-treated RNA was ligated to RNA adapter, then purified by phenol-chloroformisoamyl-alcohol (25:24:1) extraction and ethanol precipitation. The treated RNA samples were used for cDNA synthesis using random hexameric oligonucleotide primers. Nested PCR reactions were applied using the 5′-RACE Outer Primers plus gene-specific outer primers and the 5′-RACE Inner Primers plus gene-specific inner primers (Table 2). pMD19-T Simple Vector (Takara Biotechnology, Japan) was used for further sequencing of the PCR products (Feng and Cronan, 2010; 2011a,b). The first nucleotide adjacent to the 5′-RACE adaptor was supposed to be the transcriptional start site of gene (Feng *et al.*, 2013b).

Bioinformatic analyses

BioQ homologs from different *Mycobacterium* species were found by Blastp (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi). The multiple alignments of either BioQ proteins or BioQ binding sites were conducted using the ClustalW2 program (http://www.ebi.ac.uk/Tools/clustalw2/index.html), and final output was processed by the ESPript 2.2 server (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) (Feng and Cronan, 2010).

Acknowledgements

This work was supported by the start-up package of Zhejiang University (Youjun Feng), the National High-tech R&D Program of China (863 Program, grant 2011AA10A205, Jin He), the National Natural Science Foundation of China (No. 31270105, Jin He) and the National Basic Research Program of China (973 Program, grant 2010CB126105, Jin He). Dr. Feng is a recipient of the 'Young 1000 Talents' Award. We especially thank Zhengguo He from Huazhong Agricultural University for providing us the *M. smegmatis* strain, the expression vector pMV261, the original plasmids of bacterial one-hybrid reporter system and the vectors involved in gene knockout experiments.

Author contributions

Y. F. and J. H. conceived and designed this project and experiments. Y. F., Q. T., X. L., T. Z., H. Z., R. G. and Z. L. performed the experiments and contributed to the development of the figures and tables. J. H., Y. F. and Y. W. analyzed the data. Y. F., Q. T., and J. H. wrote this manuscript.

Conflict of interests

We declare that no conflict of interest is present.

References

Agarwal, V., Lin, S., Lukk, T., Nair, S.K., and Cronan, J.E. (2012) Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc Natl Acad Sci USA* **109:** 17406–17411.

Attwood, P.V., and Wallace, J.C. (2002) Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc Chem Res* **35:** 113–120.

Beckett, D. (2007) Biotin sensing: universal influence of biotin status on transcription. *Annu Rev Genet* **41:** 443–464.

Beckett, D. (2009) Biotin sensing at the molecular level. *J Nutr* **139:** 167–170.

Bharati, B.K., Swetha, R.K., and Chatterji, D. (2013) Identification and characterization of starvation induced msdgc-1 promoter involved in the c-di-GMP turnover. *Gene* **528:** 99–108.

Blokpoel, M.C., Murphy, H.N., O'Toole, R., Wiles, S., Runn, E.S., Stewart, G.R., *et al.* (2005) Tetracycline-inducible gene regulation in mycobacteria. *Nucleic Acids Res* **33:** e22.

Bower, S., Perkins, J.B., Yocum, R.R., Howitt, C.L., Rahaim, P., and Pero, J. (1996) Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *J Bacteriol* **178:** 4122–4130.

Brune, I., Gotker, S., Schneider, J., Rodionov, D.A., and Tauch, A. (2012) Negative transcriptional control of biotin metabolism genes by the TetR-type regulator BioQ in biotin-auxotrophic *Corynebacterium glutamicum* ATCC 13032. *J Biotechnol* **159:** 225–234.

Chakravartty, V., and Cronan, J.E. (2012) Altered regulation of *Escherichia coli* biotin biosynthesis in BirA superrepressor mutant strains. *J Bacteriol* **194:** 1113–1126.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:** 537–544.

Domenech, P., Barry, C.E., 3rd, and Cole, S.T. (2001) *Mycobacterium tuberculosis* in the post-genomic age. *Curr Opin Microbiol* **4:** 28–34.

Feng, Y., and Cronan, J.E. (2009) *Escherichia coli* unsatu-

rated fatty acid synthesis: complex transcription of the *fabA* gene and *in vivo* identification of the essential reaction catalyzed by FabB. *J Biol Chem* **284:** 29526–29535.
Feng, Y., and Cronan, J.E. (2010) Overlapping repressor binding sites result in additive regulation of *Escherichia coli* FadH by FadR and ArcA. *J Bacteriol* **192:** 4289–4299.
Feng, Y., and Cronan, J.E. (2011a) Complex binding of the FabR repressor of bacterial unsaturated fatty acid biosynthesis to its cognate promoters. *Mol Microbiol* **80:** 195–218.
Feng, Y., and Cronan, J.E. (2011b) The *Vibrio cholerae* fatty acid regulatory protein, FadR, represses transcription of *plsB*, the gene encoding the first enzyme of membrane phospholipid biosynthesis. *Mol Microbiol* **81:** 1020–1033.
Feng, Y., Xu, J., Zhang, H., Chen, Z., and Srinivas, S. (2013a) *Brucella* BioR regulator defines a complex regulatory mechanism for bacterial biotin metabolism. *J Bacteriol* **195:** 3451–3467.
Feng, Y., Zhang, H., and Cronan, J.E. (2013b) Profligate biotin synthesis in alpha-proteobacteria – a developing or degenerating regulatory system? *Mol Microbiol* **88:** 77–92.
Guo, M., Feng, H., Zhang, J., Wang, W., Wang, Y., Li, Y., *et al*. (2009) Dissecting transcription regulatory pathways through a new bacterial one-hybrid reporter system. *Genome Res* **19:** 1301–1308.
Haldimann, A., and Wanner, B.L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* **183:** 6384–6393.
Hebbeln, P., Rodionov, D.A., Alfandega, A., and Eitinger, T. (2007) Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. *Proc Natl Acad Sci USA* **104:** 2909–2914.
Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77:** 51–59.
Knowles, J.R. (1989) The mechanism of biotin-dependent enzymes. *Annu Rev Biochem* **58:** 195–221.
Lin, S., and Cronan, J.E. (2010) Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst* **7:** 1811–1821.
Lin, S., and Cronan, J.E. (2012) The BioC O-methyltransferase catalyzes methyl esterification of malonyl-acyl carrier protein, an essential step in biotin synthesis. *J Biol Chem* **287:** 37010–37020.
Lin, S., Hanson, R.E., and Cronan, J.E. (2010) Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nat Chem Biol* **6:** 682–688.
Manandhar, M., and Cronan, J.E. (2013) Proofreading of noncognate acyl adenylates by an acyl-coenzyme a ligase. *Chem Biol* **20:** 1441–1446.
Otsuka, A., and Abelson, J. (1978) The regulatory region of the biotin operon in *Escherichia coli*. *Nature* **276:** 689–694.
Parish, T., and Stoker, N.G. (2000) Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology* **146** (Part 8): 1969–1975.
Ploux, O., Soularue, P., Marquet, A., Gloeckler, R., and Lemoine, Y. (1992) Investigation of the first step of biotin biosynthesis in *Bacillus sphaericus*. Purification and characterization of the pimeloyl-CoA synthase, and uptake of pimelate. *Biochem J* **287** (Part 3): 685–690.
Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., *et al*. (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **69:** 326–356.
Reyrat, J.M., and Kahn, D. (2001) *Mycobacterium smegmatis*: an absurd model for tuberculosis? *Trends Microbiol* **9:** 472–474.
Rodionov, D.A., and Gelfand, M.S. (2006) Computational identification of BioR, a transcriptional regulator of biotin metabolism in Alphaproteobacteria, and of its binding signal. *FEMS Microbiol Lett* **255:** 102–107.
Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2002) Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res* **12:** 1507–1516.
Smeulders, M.J., Keer, J., Speight, R.A., and Williams, H.D. (1999) Adaptation of *Mycobacterium smegmatis* to stationary phase. *J Bacteriol* **181:** 270–283.
Tyagi, J.S., and Sharma, D. (2002) *Mycobacterium smegmatis* and *tuberculosis*. *Trends Microbiol* **10:** 68–69.
Yang, M., Chen, Y., Zhou, Y., Wang, L., Zhang, H., Bi, L.J., and Zhang, X.E. (2012a) MSMEG_2731, an uncharacterized nucleic acid binding protein from *Mycobacterium smegmatis*, physically interacts with RPS1. *PLoS ONE* **7:** e36666.
Yang, M., Gao, C., Cui, T., An, J., and He, Z.G. (2012b) A TetR-like regulator broadly affects the expressions of diverse genes in *Mycobacterium smegmatis*. *Nucleic Acids Res* **40:** 1009–1020.
Yang, M., Gao, C.H., Hu, J., Dong, C., and He, Z.G. (2014) Characterization of the interaction between a SirR family transcriptional factor of *Mycobacterium tuberculosis*, encoded by Rv2788, and a pair of toxin-antitoxin proteins RelJ/K, encoded by Rv3357 and Rv3358. *FEBS J* **281:** 2726–2737.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.