Int J Clin Exp Med 2014;7(12):5385-5393
www.ijcem.com /ISSN:1940-5901/IJCEM0002982

Original Article

Polymorphisms of FtsK/SpoIIIE protein in *Mycobacterium tuberculosis* complex strains may affect both protein function and host immune reaction

Yi Jiang[1,2*], Haican Liu[1,2*], Yan Qiu[1,2*], Guilian Li[1,2], Xiangfeng Dou[3], Kanglin Wan[1,2]

[1]*State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, P. R. China;* [2]*Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310003, China;* [3]*Beijing Center for Diseases Prevention and Control, Beijing 100013, China.* [*]*Equal contributors.*

Received October 5, 2014; Accepted December 8, 2014; Epub December 15, 2014; Published December 30, 2014

Abstract: FtsK/SpoIIIE family ATPases are conserved throughout bacteria and are involved in the translocation of DNA and proteins through membrane-spanning pores. Through comparison of 26 whole-genome-sequenced *M. tuberculosis* complex (MTBC) strains downloaded from NCBI website, we found that FtsK/SpoIIIE protein presented polymorphisms. One *M. bovis* strain and three BCG strains even showed the two T cell epitopes missing. Then we chose 159 clinical *M. tuberculosis* isolates from China, amplified gene encoding FtsK/SpoIIIE protein (Rv3871) and compared the sequences. The results showed that there are polymorphisms existed in FtsK/SpoIIIE protein among MTBC, which may affect both protein function and host immune reaction. In addition, position 1497 could be used as a good phylogenetic marker for Beijing strains.

Keywords: *Mycobacterium tuberculosis*, FtsK/SpoIIIE protein, polymorphisms

Introduction

The majority of protein export systems contain associated ATPases that supply energy for functions including assembly of the secretory apparatus and substrate translocation. FtsK/SpoIIIE family ATPases are conserved throughout bacteria and are involved in the translocation of DNA and proteins through membrane-spanning pores [1]. This translocation is important for cell division, sporulation, DNA conjugation, and other essential cell processes [1, 2]. Rv3871 is FtsK/SpoIIIE family ATPases in *Mycobacterium tuberculosis* [3]. It is a cytosolic ATPase that has been shown to bind to the C-terminal of CFP10 in the ESAT6: CFP10 complex and escort it to Rv3870, the membrane-bound component of the Esx-I system protein, and thereby allow its secretion [4]. The Esx-I system has several other substrates that are co-secreted and are mutually dependent upon each other for secretion, such that the inhibition of secretion of any of these substrates can affect the secretion of the rest of the substrates [5]. Through functional and comparative genomic studies, it is now known that the secretion of ESAT6 and CFP10 is crucial for stimulating host immunogenicity [6] while imparting a fully virulent phenotype to *M. tuberculosis* [7-9].

Through comparison of 26 whole-genome-sequenced *M. tuberculosis* complex (MTBC) strains downloaded from NCBI website, we found that FtsK/SpoIIIE protein presented polymorphisms. One *M. bovis* strain and three BCG strains even showed the two T cell epitopes missing. All these give us a hint that polymorphisms of the FtsK/SpoIIIE protein in the MTBC may be the reason for changes in the antigen produced, which may cause alterations of related functions and host immune reaction.

Materials and methods

Ethics statement

The study obtained approval from the Ethics Committee of National Institute for Com-

Table 1. 26 MTBC strains whose data were obtained from NCBI website

ID No.	Strain name
01	*Mycobacterium tuberculosis* H37Rv
02	*Mycobacterium tuberculosis* CDC1551
03	*Mycobacterium bovis* AF2122/97
04	*Mycobacterium bovis* BCG str. Pasteur 1173P2
05	*Mycobacterium tuberculosis* H37Ra
06	*Mycobacterium tuberculosis* F11
07	*Mycobacterium bovis* BCG str. Tokyo 172
08	*Mycobacterium tuberculosis* KZN 1435
09	*Mycobacterium tuberculosis* KZN 4207
10	*Mycobacterium bovis* BCG str. Mexico
11	*Mycobacterium tuberculosis* UT205
12	*Mycobacterium tuberculosis* RGTB327
13	*Mycobacterium tuberculosis* CCDC5180
14	*Mycobacterium tuberculosis* CCDC5079
15	*Mycobacterium tuberculosis* CTRI-2
16	*Mycobacterium tuberculosis* RGTB423
17	*Mycobacterium tuberculosis* KZN 605
18	*Mycobacterium tuberculosis* 7199-99
19	*Mycobacterium bovis* BCG str. Korea
20	*Mycobacterium tuberculosis* str. Erdman = ATCC 35801
21	*Mycobacterium tuberculosis* str. Beijing/NITR203
22	*Mycobacterium tuberculosis* str. Haarlem/NITR202
23	*Mycobacterium tuberculosis* CAS/NITR204
24	*Mycobacterium tuberculosis* EAI5/NITR206
25	*Mycobacterium tuberculosis* EAI5
26	*Mycobacterium tuberculosis* str. Haarlem

municable Disease Control and Prevention, Chinese Center for Disease Control and Prevention. The patients with TB included in the present research protocol were given a Subject information sheet and they all gave written informed consent to participate in the study.

Strains and DNA preparation

Firstly, we used FtsK/SpoIIIE protein sequences of 26 MTBC strains (**Table 1**) from NCBI genome website to compare and analyze. These 26 strains included one *M. africanum* strain, one *M. bovis* strain and 20 *M. tuberculosis* strains. We can see that FtsK/SpoIIIE protein presented polymorphisms among 26 MTBC strains (**Figure 1**). Four strains presented a same nonsynonymous mutation of P31A. *Mycobacterium tuberculosis* F11 had one nonsynonymous mutation of P288S. One *M. bovis* strain and three BCG strains even showed the two T cell epitopes missing. All these give us a hint that polymorphisms of the FtsK/SpoIIIE protein in the MTBC may be the reason for changes in the antigen produced, which may cause alterations of related functions and host immune reaction.

Then we chose 159 clinical *M. tuberculosis* isolates to clarify this hypothesis. 159 strains were selected from 2346 MTBC strains isolated in Beijing Municipality and 12 provinces and autonomous regions (See **Table 2**), China, which were genotyped by Spoligotyping in a previous study [10]. Strains belonging to all major and rare spoligotypes in China were included. Considering the predominance of the Beijing family strains in China, we chose about half of the Beijing family strains (81 strains) and half non-Beijing family strains (78 strains). We randomly selected the 81 Beijing family strains from 1738 Beijing strains among 2346 strains. The remaining 78 strains were selected from 608 non-Beijing family isolates. Furthermore, we attempted to include strains representing different spoligotypes that were isolated from different places. **Table 2** shows the numbers of strains used in this study that were obtained from different provinces in China. The spoligotype patterns of 159 stains were showed in **Table 3**.

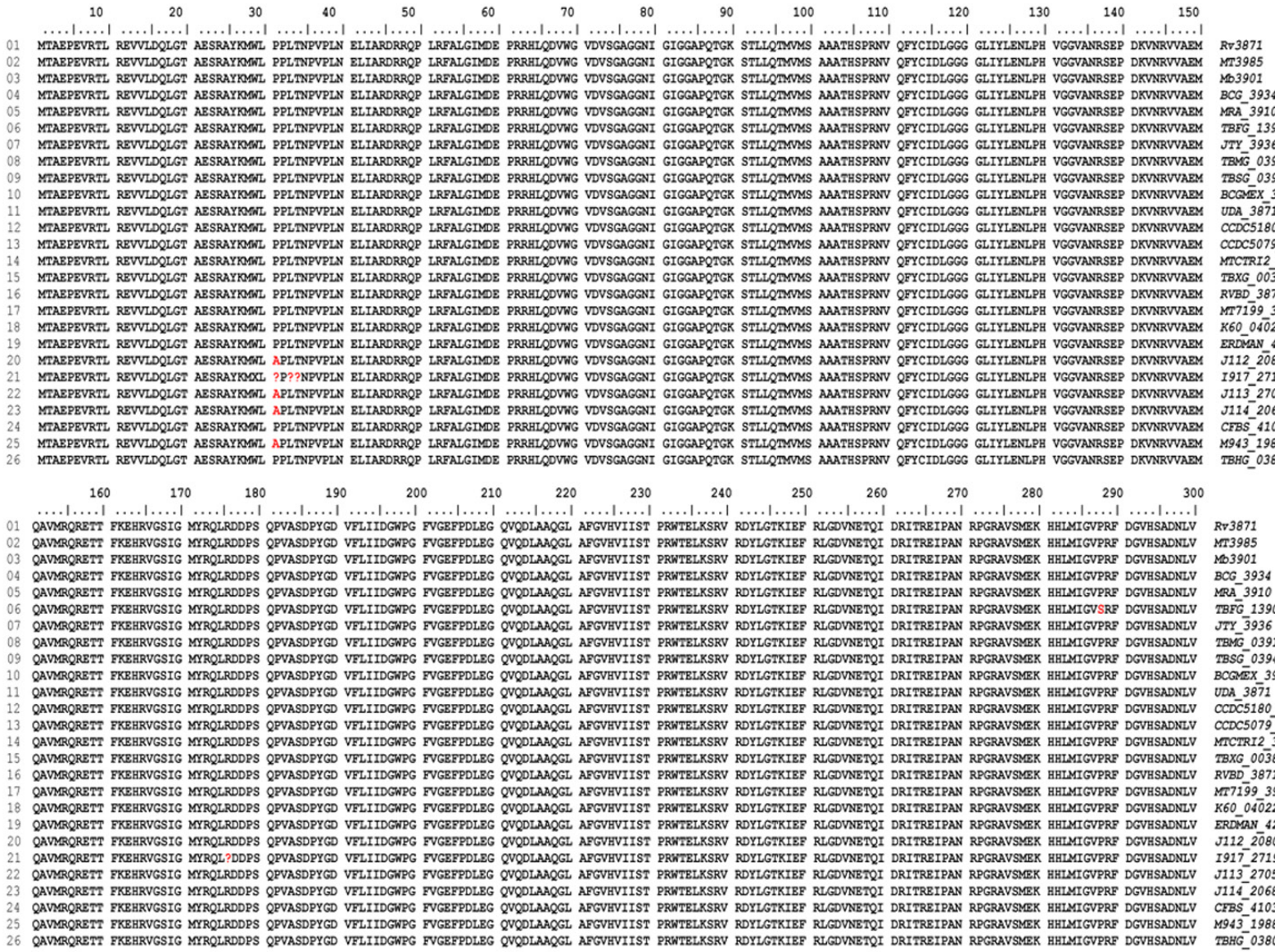
The strains were cultured using a standard Löwenstein-Jensen medium method, heat inactivated and then used directly in polymerase chain reactions (PCRs).

Primers

As the gene sequence of Rv3871 reached to 1775 bp, we split it into two parts and designed two pairs of primers to amplify the gene sequence. The primers (from the 5' to 3' end) used in this study were designed with DNASTAR software according to H37Rv genome sequence and were showed in **Table 4**.

PCR

The PCR were performed in a total volume of 20 μL. The PCR mix contained 10 μL PCR buf-


10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
01 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM Rv3871
02 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM MT3985
03 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM Mb3901
04 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM BCG_3934
05 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM MRA_3910
06 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM TBFG_13906
07 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM JTY_3936
08 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM TBMG_03919
09 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM TBSG_03943
10 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM BCGMEX_3933
11 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM UDA_3871
12 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM CCDC5180_3546
13 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM CCDC5079_3601
14 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM MTCTRI2_3950
15 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM TBXG_003890
16 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM RVBD_3871
17 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM MT7199_3940
18 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM K60_040220
19 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM ERDMAN_4245
20 MTAEPEVRTL REVVLDQLGT AESRAYKMWL APLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM J112_20800
21 MTAEPEVRTL REVVLDQLGT AESRAYKMXL ?P??NPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM I917_27190
22 MTAEPEVRTL REVVLDQLGT AESRAYKMWL APLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM J113_27050
23 MTAEPEVRTL REVVLDQLGT AESRAYKMWL APLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM J114_20680
24 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM CFBS_4103
25 MTAEPEVRTL REVVLDQLGT AESRAYKMWL APLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM M943_19885
26 MTAEPEVRTL REVVLDQLGT AESRAYKMWL PPLTNPVPLN ELIARDRRQP LRFALGIMDE PRRHLQDVWG VDVSGAGGNI GIGGAPQTGK STLLQTMVMS AAATHSPRNV QFYCIDLGGG GLIYLENLPH VGGVANRSEP DKVNRVVAEM TBHG_03809
160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
01 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV Rv3871
02 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV MT3985
03 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV Mb3901
04 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV BCG_3934
05 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV MRA_3910
06 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVSRF DGVHSADNLV TBFG_13906
07 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV JTY_3936
08 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV TBMG_03919
09 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV TBSG_03943
10 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV BCGMEX_3933
11 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV UDA_3871
12 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV CCDC5180_3546
13 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV CCDC5079_3601
14 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV MTCTRI2_3950
15 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV TBXG_003890
16 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV RVBD_3871
17 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV MT7199_3940
18 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV K60_040220
19 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV ERDMAN_4245
20 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV J112_20800
21 QAVMRQRETT FKEHRVGSIG MYRQL?DDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV I917_27190
22 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV J113_27050
23 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV J114_20680
24 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV CFBS_4103
25 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV M943_19885
26 QAVMRQRETT FKEHRVGSIG MYRQLRDDPS QPVASDPYGD VFLIIDGWPG FVGEFPDLEG QVQDLAAQGL AFGVHVIIST PRWTELKSRV RDYLGTKIEF RLGDVNETQI DRITREIPAN RPGRAVSMEK HHLMIGVPRF DGVHSADNLV TBHG_03809

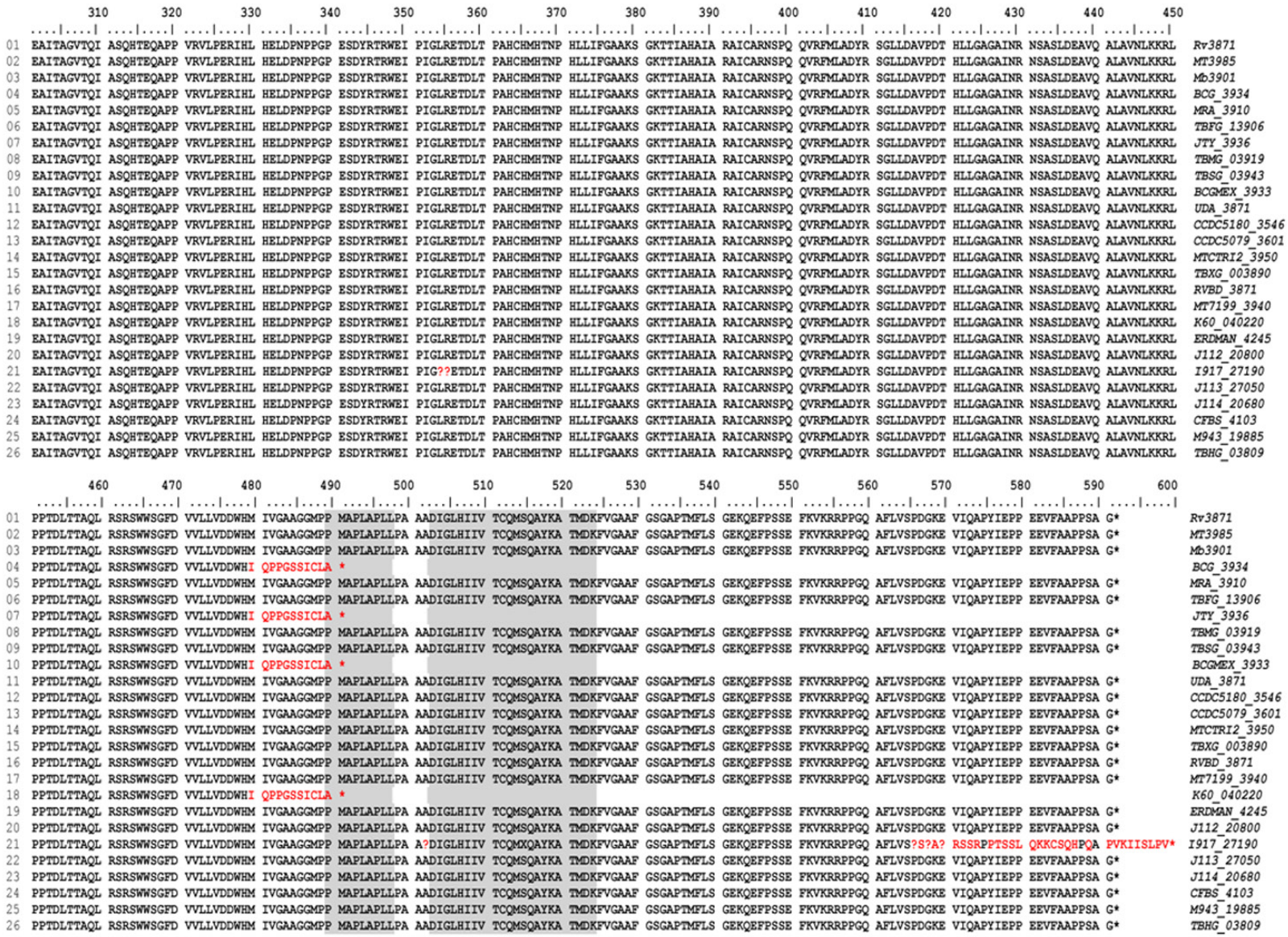


Figure 1. AA sequence alignment for antigen Rv3871 of 26 whole-genome-sequenced MTBC strains. AA changes marked in red color. Shading indicates T cell epitope area. "?" means there were undetermined base(s) in the sequence.

Table 2. No. of strains in different provinces of China

Places	No. of isolates
Anhui Province	11
Shannxi Province	14
Beijing Municipality	10
Fujian Province	25
Gansu Province	11
Guangxi Zhuang Autonomous Region	27
Sichuan Province	1
Henan Province	12
Hunan Province	7
Xizang (Tibet) Autonomous Region,	5
Xinjiang Uygur Autonomous Region	11
Jilin Province	13
Zhejiang Province	12

Table 3. No. of strains of each Spoligotype pattern

Spoligotypes	No. of strains
Beijing	81
T	10
U	25
MANU	10
Haarlem	5
EAI	1
LAM	2
S	1
CAS	4
H37Rv family	1
New	19

fer, 100 nM each primer, 200 μM each of the four dNTPs and 0.5 U DNA Taq Polymerase (Takara). An initial denaturation of 5 min at 94°C was followed by 35 cycles of denaturation at 94°C for 45 s, annealing at 62°C for 45 s and extension at 72°C for 1 min, followed by a final extension at 72°C for 10 min.

Negative controls (reagents only, no DNA) were included each time when the PCR was performed. The positive control was 500 pg DNA from *M. tuberculosis* H37Rv. The presence and size of each PCR product were determined by electrophoresis on 2% agarose gel in Tris-boric acid-EDTA buffer followed by staining with ethidium bromide.

We performed all of the PCRs at least twice to validate the reproducibility. The variants were confirmed by sequencing of the new PCR products.

Sequence

The sequences of the PCR products were determined by ABI 3730xl DNA Analyzer.

Data analysis

The sequences were first aligned by ClustalW [11] software with the *Rv3871* gene sequence from *M. tuberculosis* H37Rv genome to determine the Rv3871 region, and then this region was split out by a personalized Perl script. The sequence compare and translate were carried out by Bioedit software. The mutated protein structures were predicted by Phyre2 software online (http://www.sbg.bio.ic.ac.uk/phyre2).

Results

Mutation and deletion in gene sequences

All 159 clinical strains we chose presented relative PCR products of antigen Rv3871-1 and Rv3871-2. **Table 5** showed the mutations in the gene sequences of Rv3871. Among 159 strains, there were six nonsynonymous mutations, six synonymous mutations and one three-base deletion in *Rv3871*. Strain JL06035, GX06055, AH03040, FJ05132 and HeN05007 all had one different nonsynonymous mutation. Four Beijing strains, GX06058, GX06047, XZ06007 and GS05121, presented same synonymous mutation. GX06002, GX06046 and FJ05143 owned different one synonymous mutation. There were one nonsynonymous mutation, one synonymous mutation and one three-base deletion in strain HuN06002. Position 1479 presented higher polymorphism, as 87 strains owned an sSNPs (G-A) (**Table 7**). We counted the frequencies of the synonymous mutation of position 1479. Among the 81 isolates of Beijing phenotype, 97.5% (n=79) isolates presented A; meanwhile, among the 78 non-Beijing isolates, only seven isolates were G in the position 1479 of Rv3871.

Changes in protein level

Table 5 showed the AA change and position in antigen Rv3871 among 159 strains. For all changes, six synonymous mutations occurred, which resulted in no AA change. Six nonsynonymous mutations led to AA change of the pro-

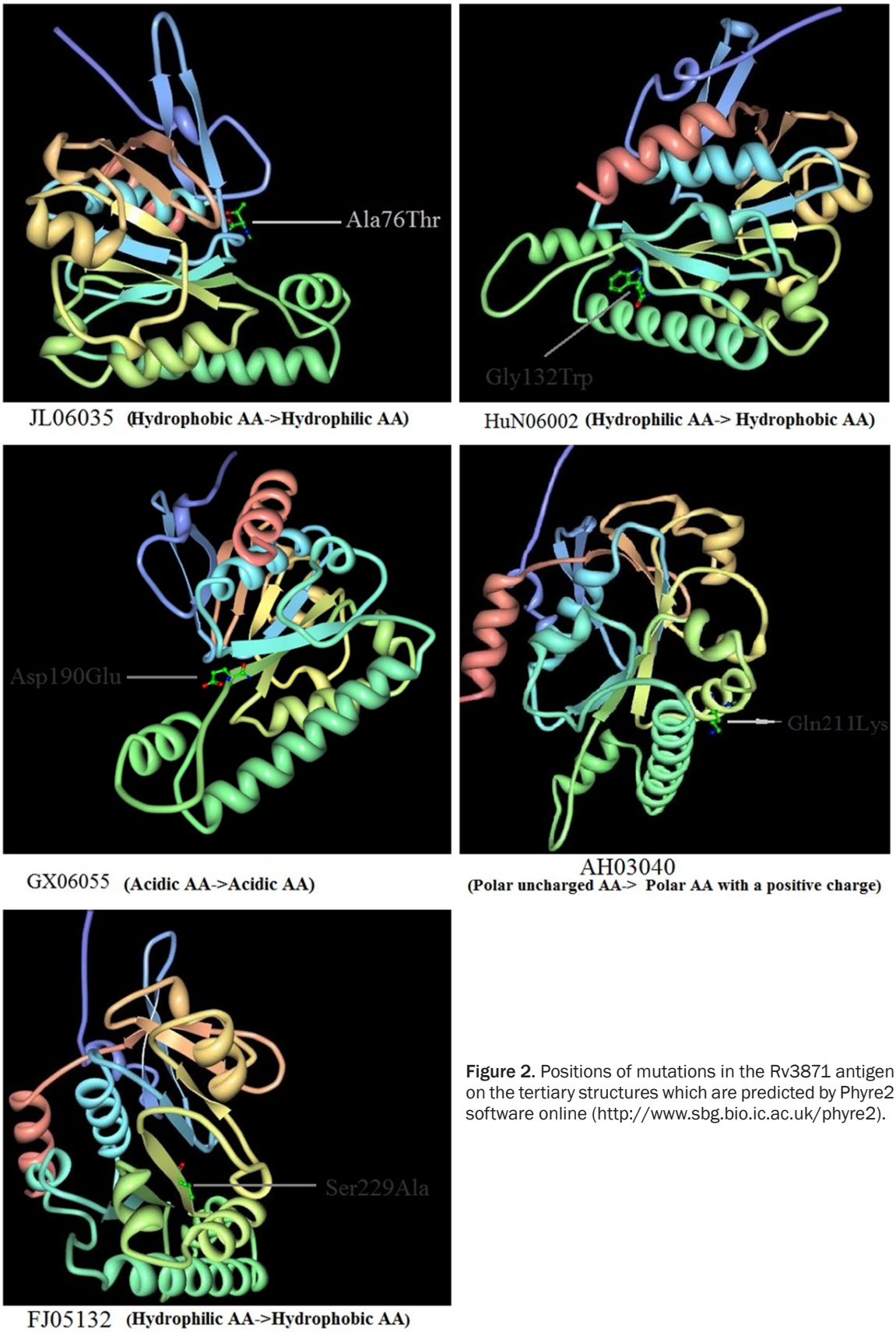


Figure 2. Positions of mutations in the Rv3871 antigen on the tertiary structures which are predicted by Phyre2 software online (http://www.sbg.bio.ic.ac.uk/phyre2).

Table 4. Primers used in the study

Gene product	Length (bp)	Primers
Rv3871-1	1507	5'-CGCGCATTCACAGGTTCAC-3'F
		5'-AACCGCTTCTTCAGGTTGAC-3'R
Rv3871-2	909	5'-GTATCCACCTGCACGAACTC-3'F
		5'-TCCCGTACACAAACCGTTCG-3'R

Table 5. Changes in antigen Rv3871 among 159 clinical strains*

Isolates	Base change	AA change	Spoligotypes
JL06035	G226A	(GCC) A76T (ACC)	MANU
HuN06002	G394T; T396G Position 397-399 deletion	(GGT) G132W (TGG) Position 133 deletion	Beijing
GX06055	C570A	(GAC) D190E (GAA)	Beijing
AH03040	C631A	(CAG) Q211K (AAG)	Beijing
FJ05132	T685G	(TCC) S229A (GCC)	T
GX06058	T891C	No change	Beijing
GX06047			
XZ06007			
GS05121			
HeN05007	A1564G	(ATG) M522V (GTG)	S
GX06002	G1557A	No change	U
GX06046	C1663A	No change	Beijing
FJ05413	C1761T	No change	U
87 strains	G1479A	No change	-#

*: Use the CDS of Rv3871 of *M. tuberculosis* H37Rv strain as the reference sequence. #: Details are showed in **Table 7**.

Table 6. AA changes of the T cell epitopes included in antigen Rv3871

IEDB_ID	Peptide sequence	AA changes
8651	DIGLHIIVTCQMSQAYKATMDK	ATG(M)-GTG(V)
144929	PMAPLAPLL	No change

tein. After we predicted protein structures by Phyre2 software online, we found that AA32-309 constitute the domain. Five nonsynonymous mutations and one three-base deletion were located on the domain region (See **Figure 2**). Also, four of five nonsynonymous mutations were all changed between amino acids with different properties, which may induce alteration of protein function. Mutation in HeN05007 was not located in the region, which resulted in no function change.

Changes in T cell epitopes

There are two human T cell epitopes in the antigen Rv3871 (**Table 6**) according to the Immune Epitopes Database (IEDB) [12]. For 26 MTBC strains from NCBI website, one *M. bovis* strain and three BCG strains showed two T cell epitopes missing. Among 159 strains, nonsynonymous mutation in HeN05007 affected one of two T cell epitopes in antigen Rv3871.

Discussion

In this study, we chose 159 clinical *M. tuberculosis* strains in China which originated from a very large geographical area and had different spoligotyping patterns; the data provided by them could therefore be representative of genetic diversity that might be present within China, at least to some extent.

Rv3871, one of gene products encoded by RD1, is a cytosolic ATPase that has been shown to bind to the C-terminal of CFP10 in the ESAT6: CFP10 complex and escort it to Rv3870, the membrane-bound component of the EsxI system protein, and thereby allow its secretion [5]. It showed that the two T cell epitopes were missing in BCG and *M. bovis* strains, which at some level explains that BCG vaccine and *M. bovis* strains are less likely to induce host immune reaction than *M. tb* strains. Also, one

Table 7. Nucleotide in position 1497 of Rv3871 among 159 strains

Strains	Spoligo-typing	Base	Strains	Spoligo typing	Base	Strains	Spoligo-typing	Base	Strains	Spoligo-typing	Base
AH03002	Beijing	A	FJ05349	T	G	GX06044	Beijing	A	JL06007	Beijing	A
AH03009	T	G	FJ05357	New	G	GX06045	Beijing	A	JL06009	Beijing	A
AH03019	Beijing	A	FJ05395	T	G	GX06046	Beijing	A	JL06015	Beijing	A
AH03026	Beijing	A	FJ05406	EAI	G	GX06047	Beijing	A	JL06016	Beijing	A
AH03029	Beijing	A	FJ05413	U	G	GX06048	Beijing	A	JL06017	Beijing	A
AH03031	Beijing	G	FJ05425	New	G	GX06052	Beijing	A	JL06018	Beijing	A
AH03032	Beijing	A	FJ05484	U	G	GX06055	Beijing	A	JL06020	Beijing	A
AH03035	Beijing	A	FJ05490	U	G	GX06058	Beijing	A	JL06023	Beijing	A
AH03037	Beijing	A	FJ05554	U	G	GX06059	Beijing	A	JL06035	MANU	A
AH03040	Beijing	A	FJ06068	MANU	A	GX06066	U	G	JL06151	MANU	G
AH04125	New	G	FJ06025	T	G	GX06087	New	G	JL06183	New	G
ShanX05092	New	G	FJ06038	Haarlem	G	GX06088	LAM	G	ZJ06001	Haarlem	G
ShanX05094	Beijing	A	FJ06051	New	G	GX06097	New	G	ZJ06003	Beijing	A
ShanX05096	Beijing	A	FJ06057	New	G	GX06112	U	G	ZJ06006	Beijing	A
ShanX05098	Beijing	A	FJ07031	LAM	G	GX06117	U	G	ZJ06007	Beijing	A
ShanX05106	Beijing	A	FJ07033	H37Rv family	G	GX06129	MANU	G	ZJ06009	Beijing	A
ShanX05105	Beijing	A	FJ07042	Haarlem	G	GX06130	U	G	ZJ06010	Beijing	A
ShanX05111	Beijing	A	FJ07049	MANU	G	HeN05007	S	G	ZJ06015	Beijing	A
ShanX05115	Beijing	G	FJ07061	MANU	G	HeN05015	U	A	ZJ06018	Beijing	A
ShanX05124	U	G	GS05040	T	G	HuN06002	Beijing	A	ZJ06020	Beijing	A
ShanX05177	New	G	GS05112	Beijing	A	HuN06004	MANU	G	ZJ06027	Beijing	A
ShanX05178	U	G	GS05113	Beijing	A	HuN06009	Beijing	A	ZJ06040	Beijing	A
ShanX05260	U	G	GS05114	Beijing	A	HuN06022	Beijing	A	ZJ06098	U	G
ShanX05290	U	G	GS05116	Beijing	A	HuN06026	T	G	HeN06022	Beijing	A
ShanX05296	MANU	A	GS05121	Beijing	A	HuN06099	U	A	HeN06023	Beijing	A
BJ05012	Beijing	A	GS05125	Beijing	A	HuN06101	New	G	HeN06024	Beijing	A
BJ05013	Beijing	A	GS05127	Beijing	A	XZ06003	CAS	G	HeN06035	Beijing	A
BJ05015	Beijing	A	GS05129	Beijing	A	XZ06107	New	G	HeN06039	Beijing	A
BJ05016	Beijing	A	GS05130	Beijing	A	XJ06013	New	G	HeN06040	Beijing	A
BJ05022	Beijing	A	GS05138	MANU	G	XJ06018	CAS	G	HeN06041	Beijing	A
BJ05024	Beijing	A	GX06002	U	G	XJ06025	U	G	HeN06042	Beijing	A
BJ05025	Beijing	A	GX06027	U	G	XJ06060	New	G	HeN06043	Beijing	A
BJ05028	Beijing	A	GX06037	Beijing	G	XJ06087	U	A	HeN06045	Beijing	A
BJ05029	Beijing	A	GX06040	U	G	XJ06106	U	A	XZ06002	Beijing	A
BJ05030	Beijing	A	GX06137	T	G	XJ06112	MANU	A	XZ06006	Beijing	G
FJ05009	New	G	GX06160	New	G	XJ06116	U	G	XZ06007	Beijing	A
FJ05063	T	G	GX06187	New	G	XJ06153	CAS	G			
FJ05086	Haarlem	A	GX06203	U	G	XJ06183	U	G			
FJ05132	T	G	GX06204	New	G	XJ06188	CAS	G			
FJ05159	T	A	SC06055	New	G	JL06003	Beijing	A			
FJ05199	Haarlem	G	GX06043	U	G	JL06006	Beijing	A			

of two T cell epitopes in the protein changed due to a nonsynonymous mutation, which results in alteration of immune reaction between pathogen and host.

In our study, seven strains presented polymorphisms in Rv3871. After we predicted the mutated protein structures by Phyre2 software, we found that five strains with nonsynonymous mutations were located in structure domain and four of them changed to amino acids with different property, which may result in alteration of protein function.

In the gene sequence of Rv3871, position 1479 presented higher polymorphisms, as 87 strains owned an sSNPs (G-A). As 97.5% Beijing strains showed A in the position, it could be used as a good phylogenetic marker for Beijing strains.

In conclusion, there are polymorphisms existed in FtsK/SpoIIIE protein among MTBC, which may affect both protein function and host immune reaction. In addition, position 1497 could be used as a good phylogenetic marker for Beijing strains.

Acknowledgements

We thank the staffs of the respective institutes in Beijing municipality, the 13 provinces and autonomous regions in China for their excellent contribution to this study, specially for the help of Lishui Zhang (Fujian), Yunhong Tan (Hunan), Xiujun Yang (Jilin), Chongxiang Tong (Gansu), Feiying Liu (Guangxi), Yingcheng Qi (Xinjiang), Qing Wang (Anhui), Xiaohui Cao and Ping Zhao (Beijing), Haitao Li (Henan), JunYang (Sichuan), Xuanmin Zhang (Shannxi), Li Shi (Xizang), Qing Wang (Anhui) and Xiaomeng Wang (Zhejiang). This work was funded by the project 81401647 of Natural Science Foundation of China, 2013ZX10003006 and 2013ZX10003002-001of Chinese National Key Program of Mega Infectious Diseases.

Disclosure of conflict of interest

None.

Address correspondence to: Kanglin Wan, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, P.O. Box 5, Changping, Beijing 102206, People's Republic of China. Tel: 0086 10 58900776; Fax: 0086 10 58900776; E-mail: wankanglin@icdc.cn

References

[1] Iyer LM, Makarova KS, Koonin EV and Aravind L. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 2004; 32: 5260-5279.

[2] Ramsdell TL, Huppert LA, Sysoeva TA, Fortune SM and Burton BM. Linked Domain Architectures Allow for Specialization of Function in the FtsK/SpoIIIE ATPases of ESX Secretion Systems. J Mol Biol 2014; [Epub ahead of print].

[3] McLaughlin B, Chon JS, MacGurn JA, Carlsson F, Cheng TL, Cox JS and Brown EJ. A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. PLoS Pathog 2007; 3: e105.

[4] Champion PA, Stanley SA, Champion MM, Brown EJ and Cox JS. C-terminal signal sequence promotes virulence factor secretion in Mycobacterium tuberculosis. Science 2006; 313: 1632-1636.

[5] Champion PA, Champion MM, Manzanillo P and Cox JS. ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol Microbiol 2009; 73: 950-962.

[6] Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C and Cole ST. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med 2003; 9: 533-539.

[7] Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S and Sherman DR. Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis. Mol Microbiol 2004; 51: 359-370.

[8] Stanley SA, Raghavan S, Hwang WW and Cox JS. Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proc Natl Acad Sci U S A 2003; 100: 13001-13006.

[9] Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, Eisenberg D, Russell RG, Derrick SC, Collins FM, Morris SL, King CH and Jacobs WR Jr. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A 2003; 100: 12420-12425.

[10] Dong H, Liu Z, Lv B, Zhang Y, Liu J, Zhao X and Wan K. Spoligotypes of Mycobacterium tuberculosis from different Provinces of China. J Clin Microbiol 2010; 48: 4102-4106.

[11] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23: 2947-2948.

[12] Ernst JD, Lewinsohn DM, Behar S, Blythe M, Schlesinger LS, Kornfeld H and Sette A. Meeting Report: NIH Workshop on the Tuberculosis Immune Epitope Database. Tuberculosis (Edinb) 2008; 88: 366-370.