Shotgun Cloning and Characterization of the Thymidylate Synthase-Encoding Gene from *Mycobacterium bovis* BCG

Sohkichi Matsumoto¹, Hideharu Yukitake¹, Naoya Ohara¹, Tohru Dairi², Hiroji Kanbara³, and Takeshi Yamada*¹

¹Nagasaki University School of Dentistry, Nagasaki, Nagasaki 852, Japan, ²Biotechnology Research Center, Toyama Prefectural University, Kosugi, Toyama 939-03, Japan, and ³Department of Protozoology, Institute of Tropical Medicine, Nagasaki, Nagasaki, 852, Japan

Received July 14, 1997; in revised form, October 1, 1997. Accepted October 6, 1997

Abstract: The shotgun cloning of a *Mycobacterium bovis* BCG (BCG) genome into pBluescript SK (+) successfully yielded a 0.9 kbp fragment, confirming the ability of *Escherichia coli* thyA mutant MH2702 to grow in a thymine-depleted medium. This DNA fragment contained a gene homologous to the thymidylate synthase (TS)-encoding genes (thyA) of other organisms. An inverted repeat sequence and open reading frame (ORF) were observed at the upstream region of the thyA. A computer analysis revealed that the protein encoded by this ORF possessed a structure unique for a DNA binding protein.

Key words: Mycobacterial enzyme, thyA, Gene expression

Mycobacterial diseases remain one of the major public health problems in the world. *Mycobacterium* can remain in a dormant, non-replicating state in human cells and can be reactivated to cause disease under certain environmental or immunological conditions (35). Over one billion people are infected with dormant *Mycobacterium*. It is important to understand the growth regulation of *Mycobacterium* because organisms in the dormant stages are more resistant to standard anti-mycobacterial drugs than when in the growth stage. The regulation of the DNA synthesis is a key role in the growth of organisms. To investigate it, gene encoding DNA polymerase I (17), topoisomerase (36), DNA gyrase (20), and the replication origins (29) were cloned from *Mycobacterium* and analyzed. TS is an essential enzyme which catalyzes the reductive methylation of dUMP to TMP. It is the only enzyme in the folate metabolism which catalyzes one-carbon transfer reaction to oxidize CH₂H₄ folate, and plays a key role in DNA synthesis. It has been reported that TS plays a critical role in relation to the cell cycle (23, 27). In this study, we cloned a TS gene from *M. bovis* BCG by shotgun cloning using an *E. coli* thyA mutant. This is the first study of TS from *Mycobacterium*.

Materials and Methods

Bacterial strains and culture. *M. bovis* BCG (Tokyo strain) was grown in Sauton medium at 37 C. *E. coli* strain XL1-Blue, which was used as a host for plasmids pBluescript SK or KS (pBSSK or pBSKS) (Stratagene Cloning System, Calif., U.S.A.), pGEX4T-3 (Phar- mocia Biotech, Tokyo), and their derivatives were grown in 2xYT broth (30). *E. coli* strain MH2702 (thyA114, relA1, thi-1; a gift from Mary Berlyn, University of Yale) was used to test plasmids for TS activity by genetic complementation.

Shotgun cloning of *M. bovis* thyA. To isolate thyA, a genomic library was constructed from *M. bovis* BCG chromosomal DNA using pBSK (+). The genomic DNA of *M. bovis* BCG was fragmented by repeated passage through a syringe needle until major parts of the DNA were between 0.5 to 5 kbp and ligated into *Eco* RI digested pBSK(+) through *Eco* RI linker DNA. The ligated DNAs were introduced to MH2702. The growth rate of MH2702 harboring each recombinant plasmid in M9 medium (30) containing 1 µg/ml of thiamine and 50

*Address correspondence to Dr. Takeshi Yamada, Nagasaki University School of Dentistry, Sakamoto 1-7-1, Nagasaki, Nagasaki 852, Japan. E-mail: f0338@cc.nagasaki-u.ac.jp

Abbreviations: BCG, *Mycobacterium bovis* BCG; GST, glutathione S-transferase; nt, nucleotide; ORF, open reading frame; pBSKS, pBluescript KS; pBSSK, pBluescript SK; RBS, ribosome binding sequence; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.
μg/ml ampicillin was determined at an optical density of 600 nm. One transformant designated MM3869 (MH2702 harboring plasmid pB3869) was obtained during this complementation selection and was chosen for further study.

**DNA sequencing and computer analysis.** DNA fragments were sequenced using a Taq Dye Primer Cycle Sequencing Kit and an automated 370A DNA sequencer system (Applied Biosystems, Inc., Calif., U.S.A.). A sequence homology search was carried out using the DDBJ (Shizuoka, Japan) database with the FASTA program from the Genetics Computer Group (University of Wisconsin, Wisc., U.S.A.).

**Southern hybridization.** Each chromosomal DNA (500 ng) of *M. bovis* BCG and *Mycobacterium tuberculosis* H37Rv were digested with *Sal I* or *Bam HI* and fractionated by agarose gel electrophoresis using 0.8% gel. The fraction was then transferred onto a nylon membrane (Gene Screen Plus, NEN Research Products, Boston, Mass., U.S.A.) and hybridization was carried out as described previously (22).

**Expression of TS as a protein fused with Schistosoma japonicum glutathione S-transferase (GST) and antibody preparation.** To prepare the GST and TS hybrid products (GST-BTS), primer A, 5'-CCCgaattcCCATGCTCGAAACGGGTACGCCCAAA, and primer B, 5'-CCCCctc gagTCATACCGCGACTGGAGCTTTGATCGCC, (the small letters indicate the added Eco RI and Xho I sites) were synthesized. The DNA products amplified from pB3869 DNA by polymerase chain reaction with primers A and B were digested with *Eco RI* and *Xho I*. They were then inserted into the same sites of pGEX4T 3, and designated pGEXthyA. The expression of GST-BTS was carried out according to the manufacturer’s instructions. The solubilized cell lysates containing GST-BTS were then fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The gel was stained with Coomassie Brilliant Blue-R. The band containing GST-BTS was cut out, homogenated, and administered to A/j mice. The antiserum against the GST-BTS fusion protein thus obtained was used.

**Construction of pB3869KS.** pB3869 was digested with both *Bam HI* and *Hind III*. The 0.9 kbp DNA fragment containing the thyA gene was inserted into the same site of pBSKS (+), and designated pB3869KS.

**Western blot analysis.** The cells were lysed in SDS gel-loading buffer and then boiled for 5 min. Each sample, approximately 20 μg protein, was fractionated by SDS-PAGE (12.5% gel). After that, the proteins in the gel were blotted onto a 0.45 μm nitrocellulose membrane. After blocking with 3% bovine serum albumin, the membrane was probed with diluted anti-GST-BTS antiserum, diluted (1:100) by overnight incubation at 4 °C. The following procedure was carried out as described previously (22).

**Results**

**Shotgun Cloning of thyA Using the thyA Mutant E. coli Strain MH2702 and Determination of the Primary Structure**

We demonstrated that pB3869 confers the growth of MH2702 in media lacking thymine (Fig. 1). The sequence analysis showed that pB3869 had 868 by of the inserted DNA (nucleotide (nt) 390-1258 in Fig. 2) and contained a 759 by ORE. The homology search showed that the amino acids sequence of the ORF had strong homology with those of the TS of various organisms.

To obtain the upstream region of the TS gene, we cloned pM4329 containing a region (nt 1–845 in Fig. 2) upstream of thyA from the genomic library using a DNA probe of the TS gene (nt 509–1048). The inserted DNA fragment in pM4329 was sequenced. The DNA sequence determined in this study is shown in Fig. 2. A possible ribosome binding sequence (RBS), GAGGA, that exhibits strong homology to the DNA sequence complementary to the 3' end of the 16S rRNA of *M. bovis* BCG (31) was found 14 nt upstream of the initiation codon (GTG). The promoter-like sequence which was identical to the −10 consensus sequence of pro-

![Fig. 1. The genetic complementation.](image-url)
moters from *M. tuberculosis* was observed (1). In the 5’ upstream region of thyA, an ORF was observed. This ORF consisted of 165 bp, starting with a GTG start codon and ending with a TAG stop codon. A possible RBS (AGGA) and a promoter-like sequence (TAGCTT) were also observed. Prediction of a secondary structure of this protein by Chou and Fasman’s method suggests that this ORF may encode a protein with a helix-turn-helix structure, which is a DNA binding protein motif (32).

The amino acid sequence deduced from the *M. bovis* TS gene was compared with those of the TS from several organisms containing *E. coli* (2), *Lactobacillus casei* (21), *T4* phage (8), mouse (25) and human (31), and the TS domain from *Plasmodium falciparum* (from His-323 to Ala-608) (7) and *Leishmania major* (from His-232 to Val-520) (5). The results are shown in Fig. 3. They indicate that the TS of *M. bovis* BCG has strong homology with those of other organisms. Sixty-seven percent identity with *E. coli* TS, 59% identity with *L. casei* TS, 46% identity with *P. falciparum* TS domain, 49% identity with *L. major* TS domain, 54% identity with mouse TS, 54% identity with human TS, and 48% identity with *T4* phage TS was observed. Amino acid sequences in the regions (Thr-36 to Lys-43) and (Met-131 to His-137), which correspond to consensus sequences for the folate binding site and dUMP binding site, respectively, were observed (15, 16). The essential amino acids for the enzymatic activity (9, 10) were also conserved in the TS of *M. bovis* BCG, such as Arg-11, Arg-212, Cys-136, and Tyr-154.

### Genomic Analysis

A Southern blot analysis was carried out (Fig. 4). The *Sal I* digested fragments with a length of 1.4 kbp and the *Bam HI* digested fragments with a length of 18 kbp gave clear hybridization bands (lanes 1, 2). The same hybridization patterns were also obtained from the genome from *M. tuberculosis H37Rv* (lanes 3, 4). *M. bovis* BCG is an attenuated derivative of *Mycobacterium bovis*, the genome of which is closely related to that of *M. tuberculosis* (13). These results show that both *M. bovis* BCG and *M. tuberculosis* have a single copy of thyA and that its location on their genomes may be identical.

### The Expression of Recombinant TS in *E. coli*

The thyA gene was expressed as a protein fused with GST. It was analyzed by SDS-PAGE (Fig. 5). The over-expressed band was seen at the expected size (54 kDa) in lane 3.

### Analysis of TS Expression by Western Blotting

The expression of recombinant TS in MH2702 was analyzed by Western blotting with the antibody against GST-BTS (Fig. 6). A protein band at 28 kDa, immunoreacting to the antibody, was observed only in the cell lysate of the MH2702 harboring pB3869 (lane 3). MH2702 harboring only the vector plasmid (lane 2) or the vector plasmid containing thyA in the opposite direction (lane 4) did not express TS, and also could not grow in media lacking thymine (data not shown). These
Fig. 3. The amino acid sequences were compared with those of TS or TS domains from several organisms containing E. coli (2), L. casei (21), T4 phage (8), P. falciparum (His-323 to Ala-608) (7), L. major (His-232 to Val-520) (5), mouse (25) and human (33). Amino acid sequences at positions which correspond to a consensus sequences for the folate binding site (Tyr-36 to Lys-43) and dUMP binding site (Met-131 to His-137) are underlined. The asterisks indicate identical amino acids in all proteins. The essential amino acids for the enzymatic activity of TS are boxed. The alignments were carried out using the malign program from the Genetics Computer Group (University of Wisconsin) through the DDBJ (Shizuoka, Japan).
results indicate that transcription occurred from the lac promoter of the vector and translation of TS was from the naive GTG initiation codon directed by the mycobacteria.

Discussion

TS is produced in a large amount in cancer cells and virus-infected cells, though generally it is synthesized at a low level (18). Therefore, drugs targeting TS are used for the treatment of various proliferative diseases (4). For the development of drugs even more effective than those currently used, and to better understand these drugs, TS genes have been cloned from several organisms and extensively studied (2, 12, 28). In this study, the shotgun cloning of M. bovis BCG genome into pBluescript SK (+) successfully yielded a 0.9 kbp fragment containing the TS gene. The primary structure of thyA was determined and the amino acids sequence deduced. It shows that the TS of M. bovis BCG has strong homology with those of other organisms as shown in Fig. 3. However, we found that E. coli XL1-Blue strain expressing the TS of M. bovis BCG was more resistant to 5-fluorouracil (an anti-cancer chemotherapeutic agent targeting TS) than the parental E. coli (data not shown). This suggests that each TS from various organisms may have somewhat of a different structure and therefore has different sensitivity against the drug, although the difference can also be explained by gene dosage effect. This indicates the possibility of organism-specific drug development.

In eucaryotic cells, the expression of TS was growth-rate-dependent and greatly increased at the G1/S phase.
In bacteria, thymine excess causes an imbalance of the nt pools at the replication fork and thereby impairs the fidelity of DNA replication (19). Therefore, the precise regulation of TS expression is necessary. In E. coli, the majority of thyA transcripts arise from the umpA promoter, which is located just upstream of the thyA. Therefore, TS is transcribed from the umpA promoter with translational coupling (3, 14). It was reported that TS expression in E. coli was regulated with the transcriptional terminator in the E. coli thyA structural gene (3). Such a structure was not observed in the thyA of M. bovis BCG. Interestingly, the inverted repeat sequence at the promoter-like region of thyA and the possible ORF, which is presumed to encode a DNA binding protein, were observed. This ORF did not have homology with the umpA gene. These structures of the upstream region of the thyA may regulate TS expression in M. bovis BCG.

A large amount of GST-BTS was expressed. However, the thyA mutant strain MH2702 expressing GST-BTS could not grow in the M9 medium lacking thymine (data not shown). On the other hand, functional TS was expressed under the control of lac promoter when the RBS and GTG initiation codon of thyA were utilized. It is worth noting that it is generally more useful to express a non-chimeric form by using its own RBS when purifying mycobacterial enzymes and antigens with a native configuration.

Recent gene technology has made it possible to use thyA for several applications. thyA can be substituted as a safe selectable marker for antibiotic-resistant genes. As for Mycobacterium, an improved vaccine against tuberculosis is in urgent demand, because low efficacy against tuberculosis was observed in individual vaccine trials of M. bovis BCG against tuberculosis (6, 26, 34). Genetically attenuated live strains such as a thyA mutant strain have the potential of becoming new effective vaccine strains.

Tuberculosis remains to be a significant medical problem. There have been few studies to biochemically analyze Mycobacterium enzymes although biochemical reactions by the enzymes are of particular importance with regard to pathogenicity. It has been suggested that TS has the potential to influence the pathogenicity of intracellular bacteria (11, 24). Slow growth is one of the most important characteristics of virulent Mycobacterium and may be strongly related with intracellular parasitism. The TS expression system may be one of the keys to understanding the cell growth of Mycobacterium. This study will furnish significant information for the research of mycobacteria.

We thank Mary Berlyn for kindly providing E. coli thyA mutant MH2702 and S. Matsunaga, O. Fujishita, H. Hotokezaka, H. Uemura, M. Naito, and T. Matsuo for helpful discussions.

References

SHOTGUN CLONING OF thyA FROM MYCOBACTERIUM

177: 1879–1882.


