RESEARCH LETTER

# A novel insertion sequence, IS*1642*, of *Mycobacterium avium*, which forms long direct repeats of variable length

Zhenyu Piao, Keigo Shibayama, Shigetarou Mori, Jun-ichi Wachino & Yoshichika Arakawa

Department of Bacterial Pathogenesis and Infection Control, National Institute of Infectious Diseases, Tokyo, Japan

**Correspondence:** Keigo Shibayama, Department of Bacterial Pathogenesis and Infection control, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi-Murayama, Tokyo 208-0011, Japan. Tel.: +81 42 561 0771; fax: +81 42 561 7173; e-mail: keigo@nih.go.jp

Received 1 September 2008; accepted 19 November 2008.
First published online 22 December 2008.

DOI:10.1111/j.1574-6968.2008.01459.x

Editor: Roger Buxton

**Keywords**
insertion sequence; direct repeat; inverted repeat; transposase.

## Abstract

A new insertion sequence (IS), IS*1642*, was identified in a *Mycobacterium avium* strain isolated from a human patient. IS*1642* had a size of 1642 bp and contained a single ORF encoding a probable transposase of 503 amino acid residues homologous (79% identity) to that of IS*1549* found in *Mycobacterium smegmatis*. The IS*1642* included imperfect inverted repeats (5′-cctgacttttatca-3′, 5′-tgataaaagtcggg-3′) on its ends, and was flanked by direct repeats of variable length ranging from 5 to 161 bp. It was suggested that the IS*1642* was widely distributed in many *M. avium* strains of human patients, and the Southern blot profile of IS*1642* was very diverse among the strains examined. The transposition event of IS*1642* was observed by *in vitro* repeated passages, showing that the IS*1642* is actually a transposable element. In light of these characteristics, IS*1642* could be a new useful marker when genotyping with high discrimination is required.

## Introduction

The insertion sequence (IS) on a bacterial genome facilitates gene rearrangements, which could contribute to the evolution of the organism (Mahillon & Chandler, 1998). ISs are widely distributed in most microorganisms, including the *Mycobacterium* species. For example, genome sequence analysis revealed that the *Mycobacterium avium* ssp. *paratuberculosis* strain K-10 contained 19 kinds of ISs, with 58 total copies in the genome (Li *et al.*, 2005). The genome of *Mycobacterium tuberculosis* H37Rv contained 56 loci with homology to ISs (Gordon *et al.*, 1999).

Bacterial ISs usually contain one or several ORFs, which encode enzymes such as transposases that catalyze the movement within the genome. ISs typically contain short terminal inverted repeat sequences (IRs), ranging from 10 to 40 bp. Upon insertion into the host genome, ISs are flanked on either side by short directly repeated sequences (DRs). The length of a direct repeat, which is usually a fixed characteristic of each IS, generally ranges between 2 and 14 bp (Mahillon & Chandler, 1998).

ISs are used as markers in restriction fragment length polymorphism studies for species typing and for molecular epidemiological purposes [e.g. IS*6110* in *M. tuberculosis* (Otal *et al.*, 1991; Small & van Embden, 1994) and IS*1245* in *M. avium* (Guerrero *et al.*, 1995; Pestel-caron & Arbeit, 1998; Ritacco *et al.*, 1998; van Soolingen *et al.*, 1998; Motiwala *et al.*, 2006)]. When isolates from different sources have few distinguishing phenotypic characteristics, the use of ISs as probes could enable powerful discrimination. On the other hand, because of their mobility, IS could be potential useful markers for identifying substrains or tracking the genetic drift (Hernandez Perez *et al.*, 1994; Laurent *et al.*, 2002).

In this study, we identified a novel IS, designated IS*1642*, which was flanked by unusually long, variable-length direct repeats, in *M. avium* clinical isolates. Here, we present molecular and genetic characterizations of this IS.

## Materials and methods

### Bacterial strains and growth conditions

*Mycobacterium avium* clinical strains isolated from human patients were kindly provided by Dr K. Ogawa of NHO Higashi Nagoya National Hospital and Dr Matsumoto of the

Osaka Prefectural Medical Center for Respiratory and Allergic Diseases, Japan. These strains were identified as the *Mycobacterium avium/intracellulare* complex by the *Mycobacterium* identification kit (Kyokuto Pharmaceutical Industrial Co. Ltd), and further identified *as M. avium* by PCR (Nishimori *et al.*, 1995). *Mycobacterium avium* ssp. *paratuberculosis* K-10 (BAA968), two *M. avium* strains (25291 and 15769), and two *M. intracellulare* strains (13950 and 25225) were obtained from the American Type Culture Collection (ATCC). Five *M. intracellulare* strains, *Mycobacterium marinum*, *Mycobacterium szulgai*, *Mycobacterium simiae*, *Mycobacterium fortuitum*, and *Mycobacterium abscessus*, and *Mycobacterium bovis* BCG Japanese strain were from our laboratory. The laboratory strain *M. tuberculosis* H37Rv was also included for analysis. These strains were cultured on Middlebrook 7H10 supplemented with 10% (v/v) OADC enrichment (BD) at 37 °C.

### DNA manipulations

Mycobacterial genomic DNA was isolated as described (Pelicic *et al.*, 1997), with a minor modification. Bacterial cells were harvested and suspended in 1 mL of acetone. The cells were pelleted by centrifugation (10 min at 5000 *g*). The pellet was resuspended in 500 μL solution I (25% sucrose/50 mM Tris · HCl, pH 8.0/50 mM EDTA/500 mg $mL^{-1}$ lysozyme) and incubated overnight at 37 °C. Then 500 μL of solution II (100 mM, Tris · HCl, pH 8.0/1% SDS/400 μg $mL^{-1}$ proteinase K) was added, and the samples were incubated for 5 h at 55 °C. Genomic DNA was extracted from the lysate using the bacteria genomicPrep Kit (GE Healthcare).

### PCR and nucleotide sequencing

The primers used in this study are listed in Table 1. PCRs were performed with Phusion high-fidelity DNA polymerase (New England Biolabs). Thermal cycling conditions comprised preincubation at 98 °C for 30 s, followed by 30 cycles at 98 °C for 10 s, at 68 or 55 °C for 30 s, and at 72 °C for 2 min, and a final extension at 72 °C for 10 min. For examination of direct repeats, genome DNA was digested by PvuII or NotI, and self-ligated. The direct repeat region was amplified by PCR with the primer set IS1642-DR1 and IS1642-DR2 (Table 1), which match the ends of IS*1642* and face the outwards. Two amplified products were obtained, and the nucleotide sequences were determined by sequencing. For sequencing, PCR products were purified using the GFX PCR purification kit (GE Healthcare). Sequencing was carried out using the Big Dye DNA sequencing kit (Applied Biosystems) and the ABI Prism 3130XL Genetic Analyzer. Nucleotide sequences and deduced amino acid sequences were analyzed by GENETYX-MAC software, version 14.0.1 (Genetyx Co., Tokyo, Japan).

**Table 1.** List of primers used in this study

| Primer | Sequence (5′–3′) |
|---|---|
| IS1642-1F | TTGTGTAGGGCTGTGACCTG |
| IS1642-1R | ACGTAGGCTGTGGATGTTTG |
| IS1642-2F | TGACCTGTGTCTTCGGTTG |
| IS1642-2R | TGTGCTTGCCGGCTTGGATAG |
| IS1642-3F | TCATCGAACCGACCAGCAAG |
| IS1642-3R | CACCACATCAGGTAAAAACG |
| IS1642-DR1 | CTACCGTACCGTCACTATCC |
| IS1642-DR2 | TCTTCACCGTGCGCACATAG |
| PIS-2 | CGGGATCCCCGCCATTTCACCTGAAACC |
| PIS-3 | CCCAAGCTTGCGAACAAAAAATCGACGCC |

### Southern blot analysis

Genome DNA was digested with restriction endonuclease PvuII (New England Biolabs), which had no recognized sites in the IS*1642* sequence. The digested DNA samples were electrophoresed on a 0.8% agarose gel. DNA fragments were transferred onto a nylon membrane and hybridized with a digoxigenin-labelled probe prepared using the PCR DIG Probe Synthesis Kit (Roche) with primer sets IS1642-3F and IS1642-3R designed to amplify a 515-bp portion of IS*1642*. Hybridized bands were visualized by chemiluminescence detection (Labeling and Detection Starter Kit II, Roche).

### Promoter activity analysis

The entire IS*1642* region was amplified by PCR with the primer set PIS-2 and PIS-3 (Table 1), and cloned upstream of a promoterless green fluorescent protein (GFP) gene on a cloning vector pVV16 (a kind gift from Dr Vissa). The plasmids were transformed into *Mycobacterium smegmatis* $mc^2$ 155, and the expression of these genes was assessed by measuring the fluorescence of GFP with a fluorescence plate reader (Perkin Elmer, Wallac 1420 ARVO MX).

### Nucleotide sequence accession numbers

AB453386 is the GenBank accession number for the nucleotide sequence of the entire region of insertion element IS*1642*. The GenBank accession number for the nucleotide sequence of the region containing insertion element IS*1642* and the 161-bp direct repeats is AB453387.

## Results

### Identification of new IS, IS*1642*

A new IS was identified in an *M. avium* strain isolated from a human patient. The IS was found 673 bp upstream of the start codon of the gene corresponding to the MAP0076 gene of the *M. avium* ssp. *paratuberculosis* K-10 genome. This

sequence contained a nucleotide sequence of 1642 bp, with a GC content of 64%, which approximates those of the mycobacterial genomes (62–70%) (Wayne & Kubica, 1986). The sequence contained a single ORF coding for a protein of 503 amino acids and 14-bp imperfect inverted repeats (5′-cctgacttttatca-3′, 5′-tgataaaagtcggg-3′) at its ends. The complete nucleotide sequence of the IS and the deduced amino acid sequence is shown in Fig. 1a. Database searches revealed that this IS was 79% identical to that of IS*1549* of *M. smegmatis* (accession number AF006614) (Plikaytis *et al.*, 1998) at the amino acid sequence level. Hence, we assigned IS*1642* to the newly identified IS.

IS*1549* is an insertion element that was reported to be distantly related to the IS*4* family (Plikaytis *et al.*, 1998). Transposases of the IS*4* family typically contain conserved regions, N3 and C1, but IS*1549* lacks the N3 region (Plikaytis *et al.*, 1998). Analysis of alignment indicated that IS*1642* also contained the C1 region, but not the N3 region. The C1 signature sequence of the IS*4* family is Y-($X_2$)-R-($X_3$)-E-($X_6$)-K (Rezsöhazy *et al.*, 1993), and the corresponding IS*1642* sequence was Y-($X_2$)-L-($X_3$)-E-($X_6$)-K. It was reported that IS*1549* possesses unique N2 and N3 regions. These sequences were found in IS*1642* as well (Fig. 1a). While the N2 signature sequence was D-($X_2$)-T-(X)-YFE-($X_{10}$)-G-(X)-SK (Alexander *et al.*, 2003), the corresponding IS*1642* sequence was D-($X_2$)-T-(X)-HFE-($X_{10}$)-G-(X)-SK. The N3 signature sequence was AD-(X)-G-($X_5$)-N (Alexander *et al.*, 2003), and the corresponding IS*1642* sequence was identical.

## Identification of direct repeats

IS*1642* identified upstream of the MAP0076 gene was flanked by direct repeats of 161 bp, which was a duplication of the nucleotide sequence of the region upstream of the MAP0076 gene. Compared with the usual direct repeats of general ISs, this direct repeat is extremely long. Because Southern blot analysis indicated that the *M. avium* strain contained multiple copies of IS*1642* (Fig. 2, lane 1), we sought to determine the length of the direct repeats of IS*1642* of other copies. Two direct repeats were successfully identified by the method described in Materials and methods. The length of the direct repeats was 5 and 59 bp,

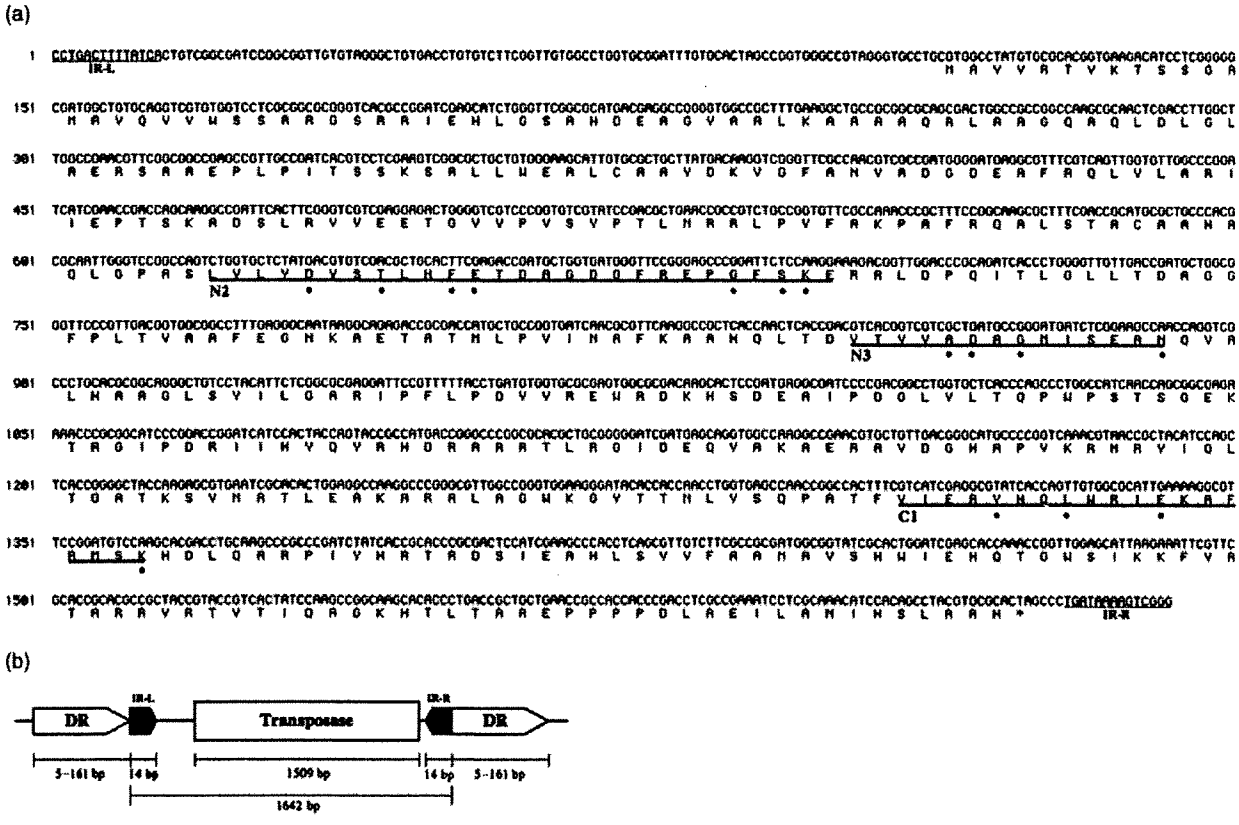


**Fig. 1.** Nucleotide sequence and structure of IS*1642*. (a) Nucleotide sequence of IS*1642* and deduced amino acid sequence of ORF. Inverted repeats, designated IR-R and IR-L, are underlined. Conserved regions, N2, N3, and C1, are underlined in bold. Dots below amino acids indicate the conserved signature sequences of these regions. (b) Schematic representation of the IS*1642* and the DR. The IR and the DR, and the beginning and end of the ORF encoding a putative transposase are shown.

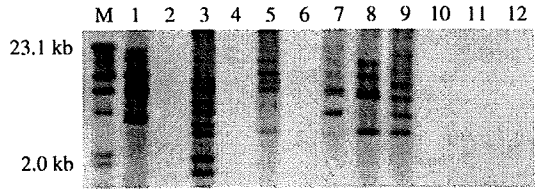


**Fig. 2.** Southern blot profile of IS*1642* of *Mycobacterium avium* strains. M, molecular size marker, lambda HindIII. lanes 1–9, human *M. avium* isolates; lane 10, *M. avium* ssp. *paratuberculosis* K-10; lane 11, *M. avium* ssp. *avium* ATCC25291; lane 12, *M. avium* ssp. *avium* ATCC15769. Numbers on the left indicate the sizes of the DNA marker.

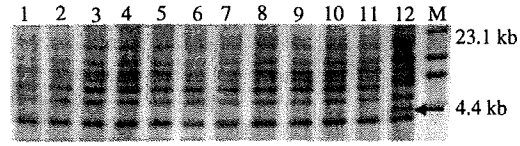


**Fig. 3.** Change in the Southern blot profile of IS*1642* of the human *Mycobacterium avium* isolate by *in vitro* repeated passages. lane 1, original *M. avium* cells; lanes 2–11, *M. avium* cells subcultured once from 10 individual colonies from the original culture plate; lane 12, *M. avium* cells passaged 10 times from the original culture plate. M, molecular size marker, lambda HindIII. The additional band that appeared after the repeated passage is indicated by the arrow. Numbers on the right indicate the sizes of the DNA marker.

respectively. The IS with the 5 bp direct repeat was found to be within the coding region of MAP2026 gene. The sequence flanking the 59 bp direct repeat sequence did not match with any gene in the database. We sought to find a conserved sequence at the insertion sites, but sequence specificity was not obvious among these three target sites.

### Distribution of IS*1642*

The distribution of IS*1642* among *M. avium* strains was examined by Southern blot analysis with the following strains: *M. avium* ssp. *paratuberculosis* K-10, *M. avium* ssp. *avium* ATCC25291, *M. avium* ssp. *avium* ATCC15769, and nine *M. avium* strains, isolated from different patients at NHO Higashi Nagoya National Hospital. The DNA bands were found in only some of the *M. avium* strains isolated from human patients, but not in *M. avium* ssp. *paratuberculosis* K-10, *M. avium* ssp. *avium* ATCC25291, and *M. avium* ssp. *avium* ATCC15769 (Fig. 2). The Southern blot profile indicated that the band patterns of IS*1642* were very diverse and the copies were multiplied. *Mycobacterium intracellulare* ATCC13950, *M. intracellulare* ATCC25225, five *M. intracellulare* clinical isolates, and seven other mycobacterial species were negative for IS*1642* on the Southern analysis profile (data not shown). The prevalence of IS*1642* among *M. avium* strains was further examined with eight strains isolated from human patients at the Osaka Prefectural Medical Center for Respiratory and Allergic Diseases by PCR with the three sets of specific primers (primers IS1642-1F and IS1642-1R, IS1642-2F and IS1642-2R, IS1642-3F, and IS1642-3R) (Table 1). All strains were positive for IS*1642* (data not shown), suggesting that IS*1642* is widely distributed among *M. avium* clinical strains.

### Mobility of IS*1642*

Because the Southern blot profile of IS*1642* was very different among strains, we explored whether or not the IS*1642* exhibits frequent mobility by examining changes in the Southern blot pattern during repeated passages *in vitro*. One *M. avium* clinical strain was streaked on a 7H10 agar plate, and passaged 10 times. A single colony was used for inoculation at each passage. After the passages, genome DNA was extracted. For reference, 10 colonies were individually subcultured once from the initial culture plate and genome DNA samples were extracted. The genome DNA samples were subjected to Southern blot analysis. While the original cells and the cells of 10 colonies with a single subculture showed identical patterns, the passaged cells contained an additional band (Fig. 3), indicating that transposition of the IS occurred during the *in vitro* repeated culture.

## Discussion

In this study, we discovered a new IS, IS*1642*, in *M. avium* clinical strains isolated from human patients. Our results showed that multiple copies of IS*1642* are present in several *M. avium* strains. The deduced amino acid sequence of the ORF of IS*1642* was highly homologous to the transposase of *M. smegmatis* IS*1549* (Plikaytis *et al.*, 1998). IS*1549* is an insertion element whose transposases exhibited homology to that of IS*1623* and IS*1634*, and it was suggested that these transposases represent an emerging group in the IS*4* family because they exhibit a characteristic lack of the typical conserved N3 region of IS*4* family and yet possess unique N2 and N3 motifs (Vilei *et al.*, 1999; Alexander *et al.*, 2003). A notable feature of IS*1549* and IS*1634* is that they are flanked by unusual long direct repeats that may vary in length (Plikaytis *et al.*, 1998; Vilei *et al.*, 1999). The lengths of the direct repeats of these ISs range up to 500 bp, in contrast to most usual ISs flanked by short direct repeats of 2–14 bp. IS*1642* found in this study also exhibited these characteristics. IS*1642* contained the C1 region and unique N2 and N3 motifs, but not the N3 region of the IS*4* family, suggesting that IS*1642* belongs to this new group. IS*1642* was flanked by direct repeats of variable lengths ranging up to 161 bp. Considering that IS*1642* was homologous to the Iss, which form variable length direct repeats, it would be likely that the direct repeat sequences found in this study were actually created by IS*1642*. The actual range of lengths

of direct repeats may be larger as reported in other ISs. Because the homology level of the entire region of the amino acid sequence was relatively low among ISs of this group (e.g. 38% identity between IS*1549* and IS*1634*), the conserved structures of the amino acid sequence might play an important role in the formation of the characteristic long direct repeats of variable lengths. There was no conserved sequence at the insertion sites among the three target sites, suggesting that the insertion events by IS*1642* take place randomly on the genome.

To date, the advantage of formation of long, variable-length direct repeats is not well elucidated. While insertion with a long target duplication may decrease the likelihood of destroying essential genes at the target sites on the host genome, it would be expected that an insert flanked by long direct repeats could be easily removed by homologous recombination between the repeats. Nobusato *et al.* (2000) reported insertions with long target duplications in the restriction and modification enzyme genes of *Helicobacter pylori* strains. They considered that the long duplication may control the copy number of the genes, thus keeping expression of the genes at an appropriate level. Further studies are required to elucidate the significance of the long, variable-length direct repeats.

Our results suggested that IS*1642* is widely distributed among *M. avium* clinical strains. In addition, the experiment of repeated passage suggested that IS*1642* is indeed capable of frequent transposition within the genome. This is consistent with the observation that the Southern blot profile was very divergent among the strains tested. Considering this polymorphism, it would be rather inappropriate to use the Southern blot pattern of IS*1642* as a genetic typing tool for classification of different clinical strains. Alternatively, genotyping by IS*1642* could be useful to confirm the clonality of strains because it would enable high-precision discrimination.

IS elements reportedly often carry an outward-directed promoter sequence, bringing about a constitutive expression of downstream genes at insertion sites (Safi *et al.*, 2004; Soto *et al.*, 2004). IS*1549* was found to show promoter activity (Plikaytis *et al.*, 1998). Because IS*1642* was homologous with IS*1549*, we examined IS*1642* for promoter activity. However, the expression of GFP was at an undetectable level. IS*1642* would have no or very low outward-directed promoter activity, or the promoter sequence could be formed as a hybrid of the IS sequence and genome sequence at insertion sites, thus turning on the expression of otherwise silent genes on the genome, as reported by Szeverényi *et al.* (1996). It might also be possible that promoters of IS1642 could not be recognized by *M. smegmatis*.

IS*1642* would be involved in the gene rearrangements on the genome, thus contributing to evolution of the organism, like many other ISs. Although *M. avium* is an opportunistic pathogen, treatment is often very difficult once the infection is established, despite long-term administration of antibiotics. It may be possible that IS*1642*, which has a relatively high mobility, serves to facilitate establishment of chronic infection by adapting the phenotype including the pathogenicity and drug resistance of the organism for conditions at the focus of the infection *in vivo*.

## Acknowledgements

We thank Dr K. Ogawa and Dr T. Matsumoto for providing *M. avium* clinical isolates. Collection of *M. avium* strains was supported by grant H18-Shinkou-Ippan-011, and analysis of the IS was supported by grant H19-Shinkou-Ippan-006, from the Ministry of Health, Labour and Welfare of Japan.

## References

Alexander DC, Jones JRW & Liu J (2003) A rifampin-hypersensitive mutant reveals differences between strains of *Mycobacterium smegmatis* and presence of a novel transposon, IS1623. *Antimicrob Agents Ch* **47**: 3208–3213.

Gordon SV, Heym B, Parkhill J, Barrell B & Cole ST (1999) New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **145**: 881–892.

Guerrero C, Bernasconi C, Burki D, Bodmer T & Telenti A (1995) A novel insertion element from *Mycobacterium avium*, IS*1245*, IS a specific target for analysis of strain relatedness. *J Clin Microbiol* **33**: 304–307.

Hernandez Perez M, Fomukong NG, Hellyer T, Brown IN & Dale JW (1994) Characterization of IS*1110*, a highly mobile genetic element from *Mycobacterium avium*. *Mol Microbiol* **12**: 717–724.

Laurent JP, Faske S & Cangelosi GA (2002) Characterization of IS*999*, an unstable genetic element in *Mycobacterium avium*. *Gene* **294**: 249–257.

Li L, Bannantine JP, Zhang Q, Amonsin A, May BJ, Alt D, Banerji N, Kanjilal S & Kapur V (2005) The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *P Natl Acad Sci USA* **102**: 12344–12349.

Mahillon J & Chandler M (1998) Insertion sequences. *Microbiol Mol Biol R* **62**: 725–774.

Motiwala AS, Li L, Kapur V & Sreevatsan S (2006) Current understanding of the genetic diversity of *Mycobacterium avium* subsp. *paratuberculosis*. *Microbes Infect* **8**: 1406–1418.

Nishimori K, Eguchi M, Nakaoka Y, Onodera Y, Ito T & Tanaka K (1995) Distribution of IS*901* in strains of *Mycobacterium avium* complex from Swine by using IS*901*-detecting primers that discriminate between *M. avium* and *Mycobacterium intracellulare*. *J Clin Microbiol* **33**: 2102–2106.

Nobusato A, Uchiyama I, Ohashi S & Kobayashi I (2000) Insertion with long target duplication: a mechanism for gene

mobility suggested from comparison of two related bacterial genomes. *Gene* **259**: 99–108.

Otal I, Martin C, Vincent-Levy-Frebault V, Thierry D & Gicquel B (1991) Restriction fragment length polymorphism analysis using IS*6110* as an epidemiological marker in tuberculosis. *J Clin Microbiol* **29**: 1252–1254.

Pelicic V, Jackson M, Reyrat JM, Jacobs Jr WR, Gicquel B & Guilhot C (1997) Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*. *P Natl Acad Sci USA* **94**: 10955–10960.

Pestel-Caron M & Arbeit RD (1998) Characterization of IS1245 for strain typing of *Mycobacterium avium*. *J Clin Microbiol* **36**: 1859–1863.

Plikaytis BB, Crawford JT & Shinnick TM (1998) IS*1549* from *Mycobacterium smegmatis* forms long direct repeats upon insertion. *J Bacteriol* **180**: 1037–1043.

Rezsöhazy R, Hallet B, Delcour J & Mahillon J (1993) The IS*4* family of insertion sequences: evidence for a conserved transposase motif. *Mol Microbiol* **9**: 1283–1295.

Ritacco V, Kremer K, van der Laan T, Pijnenburg JEM, de Haas PEW & van Soolingen D (1998) Use of IS901 and IS1245 in RFLP typing of *Mycobacterium avium* complex: relatedness among serovar reference strains, human and animal isolates. *Int J Tuberc Lung D* **2**: 242–251.

Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R & Howard ST (2004) IS*6110* functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol Microbiol* **52**: 999–1012.

Small PM & van Embden JDA (1994) Molecular epidemiology of tuberculosis. *Tuberculosis: Pathogenesis, Protection, and Control* (Bloom BR, ed), pp. 569–582. American Society for Microbiology, Washington, DC.

Soto CY, Menéndez MC, Pérez E, Samper S, Gómez AB, García MJ & Martín C (2004) IS*6110* mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. *J Clin Microbiol* **42**: 212–219.

Szeverényi I, Hodel A, Arber W & Olasz F (1996) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. *Gene* **174**: 103–110.

van Soolingen D, Bauer J, Ritacco V *et al.* (1998) IS*1245* restriction fragment length polymorphism typing of *Mycobacterium avium* isolates: proposal for standardization. *J Clin Microbiol* **36**: 3051–3054.

Vilei EM, Nicolet J & Frey J (1999) IS*1634*, a novel insertion element creating long, variable-length direct repeats which is specific for *Mycoplasma mycoides* subsp. *mycoides* small-colony type. *J Bacteriol* **181**: 1319–1323.

Wayne LG & Kubica GP (1986) Genus *Mycobacterium* Lehmann and Neumann. *Bergey's Manual of Systematic Bacteriology, Vol. 2* (Krieg NR & Holt JG, eds), pp. 1436–1457. The William & Wilkins Co., Baltimore, MD.

Protein Expression and Purification 69 (2010) 99–105

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

# Purification and molecular characterization of a novel diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate phosphorylase from *Mycobacterium tuberculosis* H37Rv

Shigetarou Mori, Keigo Shibayama *, Jun-ichi Wachino, Yoshichika Arakawa

*Department of Bacterial Pathogenesis and Infection Control, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashimurayama-shi, Tokyo 208-0011, Japan*

ARTICLE INFO

*Article history:*
Received 17 June 2009
and in revised form 16 September 2009
Available online 22 September 2009

*Keywords:*
*Mycobacterium tuberculosis*
Expression
Nucleotide
Phosphorylase
Histidine triad motif

ABSTRACT

In this study, Rv2613c, a protein that is encoded by the open reading frame *Rv2613c* in *Mycobacterium tuberculosis* H37Rv, was expressed, purified, and characterized for the first time. The amino acid sequence of Rv2613c contained a histidine triad (HIT) motif consisting of H-phi-H-phi-H-phi-phi, where phi is a hydrophobic amino acid. This motif has been reported to be the characteristic feature of several diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate (Ap4A) hydrolases that catalyze Ap4A to adenosine 5′-triphosphate (ATP) and adenosine monophosphate (AMP) or 2 adenosine 5′-diphosphate (ADP). However, enzymatic activity analyses for Rv2613c revealed that Ap4A was converted to ATP and ADP, but not AMP, indicating that Rv2613c has Ap4A phosphorylase activity rather than Ap4A hydrolase activity. The Ap4A phosphorylase activity has been reported for proteins containing a characteristic H-X-H-X-Q-phi-phi motif. However, no such motif was found in Rv2613c. In addition, the amino acid sequence of Rv2613c was significantly shorter compared to other proteins with Ap4A phosphorylase activity, indicating that the primary structure of Rv2613c differs from those of previously reported Ap4A phosphorylases. Kinetic analysis revealed that the $K_m$ values for Ap4A and phosphate were 0.10 and 0.94 mM, respectively. Some enzymatic properties of Rv2613c, such as optimum pH and temperature, and bivalent metal ion requirement, were similar to those of previously reported yeast Ap4A phosphorylases. Unlike yeast Ap4A phosphorylases, Rv2613c did not catalyze the reverse phosphorolysis reaction. Taken together, it is suggested that Rv2613c is a unique protein, which has Ap4A phosphorylase activity with an HIT motif.



## Introduction

*Mycobacterium tuberculosis* causes tuberculosis (TB),[1] a serious bacterial infection. Every year, 1.8 million people die from TB and 9.3 million people are newly infected. Further, multidrug-resistant TB has posed a serious problem in recent years [1]. Therefore, there is an urgent need for countermeasures against TB and detailed studies on *M. tuberculosis*. This study was designed to elucidate the function of proteins with unidentified activity to discover potential drug targets. In this study, we focused on the Rv2613c protein encoded by the open reading frame *Rv2613c* of *M. tuberculosis* H37Rv because the *Rv2613c* gene has been shown to be one of its essential genes by mutagenesis study [2]. However, the functions of this gene are yet to be specified.

The importance of this gene is also supported by the fact that the neighboring genes in the same operon of the *Rv2613c* gene encode important proteins. The operon containing the *Rv2613c* gene comprises 6 genes (*Rv2614c–Rv2609c*) [3]. It has been shown that 3 genes (*Rv2612c*, *Rv2611c*, and *Rv2610c*) in this operon participate in cell wall biosynthesis: the *Rv2612c* gene encodes phosphatidylinositol synthase [4], *Rv2611c* encodes a protein with similarities to bacterial acyltransferases [5], and *Rv2610c* encodes α-mannosyltransferase [6]. Further, the *Rv2614c* gene is predicted to code for threonyl-tRNA synthetase, as suggested by an amino acid sequence homology search [3].

We found that the amino acid sequence of the *Rv2613c* gene contains a histidine triad (HIT) motif consisting of H-phi-H-phi-H-phi-phi, where phi is a hydrophobic amino acid (Fig. 1). It was indicated that the proteins containing the HIT motif possess hydrolase activity as reported in adenosine 5′-monophosphoramide (AMP–$NH_2$) hydrolases [7] and diadenosine polyphosphate hydrolases [8,9]. Therefore, it was predicted that Rv2613c also functions as a hydrolase. In contrast, in this study, we found that Rv2613c unexpectedly shows phosphorylase activity (Eq. (1)), rather than hydrolase activity (Eqs. (2) and (3)), against diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate (Ap4A). Here, we report a detailed characterization of Rv2613c.

* Corresponding author. Fax: +81 42 561 7173.
*E-mail address:* keigo@nih.go.jp (K. Shibayama).

[1] *Abbreviations used:* TB, tuberculosis; HIT, histidine triad; ATP, adenosine 5′-triphosphate; AMP, adenosine monophosphate; ADP, adenosine 5′-diphosphate; HPLC, high-performance liquid chromatography; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

1046-5928/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2009.09.010

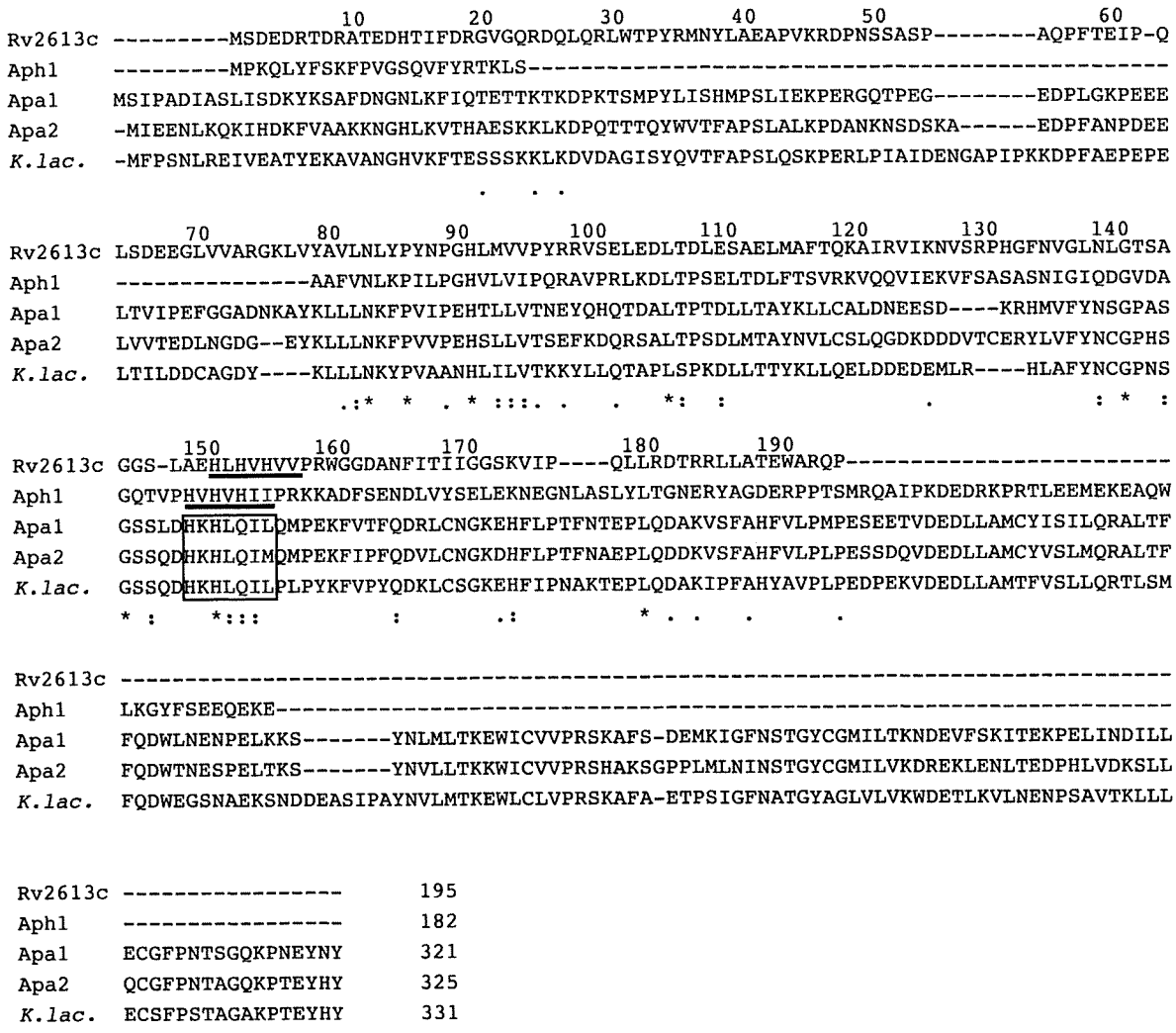


**Fig. 1.** The amino acid sequence of Rv2613c and sequence alignments between Rv2613c, *Schizosaccharomyces pombe* Ap4A hydrolase (Aph1), *Saccharomyces cerevisiae* Ap4A phosphorylase I and II (Apa1 and Apa2), and *Kluyveromyces lactis* Ap4A phosphorylase II. The HIT motif is underlined, and the H-X-H-X-Q-phi-phi motif is framed. Identical residues are denoted by an asterisk (*); strongly conserved residues, by a colon (:); and weakly conserved residues, by a period (.). *K. lac.*, *K. lactis* Ap4A phosphorylase II. The numbers along the top refer to the position of an amino acid in the sequence of Rv2613c, and the numbers in the last alignment indicate the lengths of each of the amino acid sequences. The SWISS-PROT or TrEMBL accession numbers are as follows: Rv2613c, O06201; Aph1, P49776; Apa1, P16550; Apa2, P22108; *K. lac.*, P49348.

$$\text{Ap4A} + \text{phosphate} \rightarrow \text{adenosine 5}'\text{-triphosphate (ATP)} + \text{adenosine 5}' - \text{diphosphate (ADP)} \quad (1)$$

$$\text{Ap4A} \rightarrow \text{ATP} + \text{adenosine monophosphate (AMP)} \quad (2)$$

$$\text{Ap4A} \rightarrow 2\text{ADP} \quad (3)$$

## Materials and methods

### *Cloning of the Rv2613c gene from M. tuberculosis H37Rv genomic DNA*

Genomic DNA from *M. tuberculosis* H37Rv (NHJ1633) was isolated as previously described [10]. On the basis of its nucleotide sequence [3], the 588-bp *Rv2613c* gene was amplified from the genomic DNA by polymerase chain reaction (PCR) (Takara PCR Thermal Cycler Dice mini, Takara Bio Inc., Shiga, Japan). PCR was performed in a reaction mixture (50 μl) containing 1 U of Phusion high-fidelity DNA Polymerase (New England BioLabs, Ipswich, MA), 250 ng genomic DNA, 25 pmol of NdeI primer (5′-GCCATATGAGTG ACGAGGACCG-3′), 25 pmol of HindIII primer (5′-GCAAGCTTTGGT TGCCGAG-3′), 10 nmol of deoxyribonucleotide triphosphates, 3% dimethyl sulfoxide, and reaction buffer that was supplied with the DNA polymerase. The cycle conditions for PCR were as follows: 25 cycles at 98 °C for 10 s, 70 °C for 30 s, and 72 °C for 60 s. The amplified gene was confirmed by sequencing with an Applied Biosystems 3130*xl* Genetic Analyzer (Applied Biosystems, Carlsbad, CA). The PCR product was inserted into the NdeI/HindIII site of an expression vector, pCold I (Takara), and the plasmid was named pMS2613c.

### *Protein expression and purification*

*Escherichia coli* BL21(DE3)pLysS (Novagen, Madison, WI) was transformed with pMS2613c and cultured aerobically for 7 h at 37 °C in 200 ml of Luria–Bertani medium containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol until the absorbance at 600 nm ($A_{600}$) was 1.0. The culture was then transferred into 2 l of the same medium containing the same antibiotics and cultured under aerobic conditions at 37 °C. When $A_{600}$ reached 0.5, isopropyl-β-D-thiogalactopyranoside was added at a final concentration of 0.5 mM, and the culture was allowed to progress under aerobic conditions at 15 °C for 24 h. The bacterial cells were collected by centrifugation at 8000g for 10 min and resuspended in Buffer A [20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, and 40 mM imidazole]. The cells were disrupted by sonication on

ice using a UP50H sonicator (Hielscher Ultrasonics, Teltow, Germany). Unbroken cells and debris were removed by centrifugation at 20,000g for 30 min, after which the clear supernatant was collected and used as the cell extract. The cell extract was injected onto a HisTrap HP Column (1.6 × 2.5 cm) (GE Healthcare Bio-Sciences, Buckinghamshire, UK) that was equilibrated with Buffer A, and the protein was eluted with a linear gradient of imidazole (40–500 mM; 150 ml). The recombinant Rv2613c with His-tag was eluted with 0.15–0.25 M imidazole. The protein was concentrated by an Amicon Ultra-15 (Millipore, Billerica, MA) at 4 °C and loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (1.6 × 60 cm) (GE Healthcare Bio-Sciences) equilibrated with Buffer B [20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, and 250 mM imidazole]. The sample was separated with Buffer B into 1.2 ml fractions. Fraction Nos. 24–30 were combined and dialyzed against Buffer C [20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid sodium salt (HEPES–Na; pH 7.6) and 0.5 mM dithiothreitol] at room temperature. This sample was used as the purified preparation of Rv2613c.

### *Assays*

Enzyme activity was determined by quantifying the substrate amount with high-performance liquid chromatography (HPLC) (Shimadzu, Kyoto, Japan) using previously described methods with modifications [11]. The reaction mixture (100 µl) consisted of 50 mM HEPES–Na (pH 7.6), 1 mM $MnCl_2$, 100 µg/ml bovine serum albumin, 0.1 mM Ap4A, 5 mM phosphate, and purified Rv2613c. This mixture was incubated at 37 °C for 10 min. The reaction was stopped by heating at 95 °C for 4 min, followed by centrifugation of the reaction solution at 12,000g for 3 min. The clear supernatant was analyzed using a 4.6 × 125 mm Partisphere 5 µm SAX HPLC column (Whatman, Kent, UK), which was equilibrated with pure water. The adsorbed nucleotides were eluted using the following gradient generated by mixing pure water with the elution buffer [1.3 M $(NH_4)_2HPO_4$ (pH 4.8) with $H_3PO_4$]: 0–5 min, 0% elution buffer; 5–55 min, 0–50% elution buffer at 1.0 ml/min, with detection at 254 nm.

One unit (U) of enzyme activity was defined as the degradation of 1.0 µmol Ap4A in 1 min at 37 °C. The effect of bivalent metal ions on enzymatic activity was determined using reaction mixtures in which 1.0 mM $MnCl_2$ was replaced by 2.0 mM each of $MnCl_2$, $CoCl_2$, $CaCl_2$, and $MgCl_2$. The optimum pH for enzymatic activity was assayed using sodium cacodylate trihydrate (pH 6.0 and 6.4), HEPES–Na (pH 7.0 and 7.4), Tris–HCl (pH 8.4), 3-(cyclohexylamino)-2-hydroxy-1-propanesulfonic acid (pH 9.0 and 9.4), and 3-(cyclohexylamino)-1-propanesulfonic acid (pH 10.0, 10.4, and 11.0). The $K_m$ and $V_{max}$ values were calculated by using the Prism software (GraphPad Software, La Jolla, CA). These kinetic values are given as means ± standard errors of three independent determinations.

### *Other analytical methods*

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) was performed as described previously [12]. The proteins in the gel were visualized using Bio-Safe Coomassie Stain (Bio-Rad, Hercules, CA). The molecular mass of purified Rv2613c was estimated by gel filtration chromatography with a Superdex 200 10/300 GL column (1.0 × 30 cm) (GE Healthcare Bio-Sciences) and a Gel Filtration Calibration kit (GE Healthcare Bio-Sciences), as recommended by the manufacturer. Protein concentration was determined using a BCA Protein Assay kit (Pierce, Rockford, IL). Sequence alignment was performed using the CLUSTAL W 1.83 software [13].

## Results

### *Expression and purification of recombinant Rv2613c*

Recombinant Rv2613c was purified to homogeneity by two-step column chromatography (Fig. 2A, Table 1). The purity of the

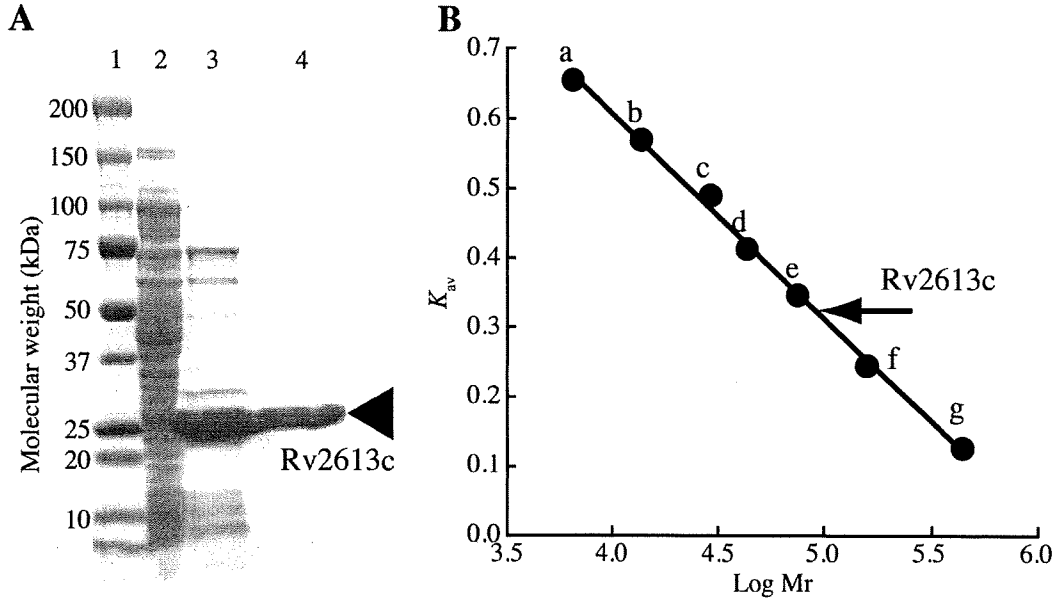


**Fig. 2.** Molecular mass of Rv2613c. (A) Purification of recombinant Rv2613c. SDS–PAGE gel stained with Coomassie blue. Lane 1, Precision Plus Protein Standards (Bio-Rad); lane 2, cell extract of MS2613c; lane 3, elution fraction obtained during the first column chromatography; lane 4, purified Rv2613c; the arrow indicates the position of Rv2613c. (B) Estimation of the molecular mass of Rv2613c. Purified Rv2613c was loaded onto a Superdex 200 column. The point at which Rv2613c was eluted is indicated by an arrow. The protein standards used were: *a*, aprotinin (6.5 kDa); *b*, ribonuclease A (13.7 kDa); *c*, carbonic anhydrase (29 kDa); *d*, ovalbumin (43 kDa); *e*, conalbumin (75 kDa); *f*, aldolase (158 kDa); and *g*, ferritin (440 kDa). $K_{av}$ values were calculated using the following equation: $K_{av} = (V_e - V_o)/(V_c - V_o)$, where $V_o$ is the column void volume, $V_e$ is the elution volume, and $V_c$ is the geometric column volume. The $V_o$ value used was the $V_e$ of Blue Dextran 2000. $M_r$, molecular weight.

**Table 1**
Purification of recombinant Rv2613c from *M. tuberculosis* H37Rv.

| Step | Total protein (mg) | Total activity (unit)[a] | Yield (%) | Specific activity (unit/mg) | Purification (fold) |
|---|---|---|---|---|---|
| Cell extract | 296.0 | 891 | 100 | 3.01 | 1.0 |
| HisTrap | 20.0 | 202 | 23 | 10.08 | 3.3 |
| Sephacryl S-200 | 15.6 | 169 | 19 | 10.84 | 3.6 |

[a] Phosphorylase activity of Rv2613c in the presence of 0.1 mM Ap4A.

preparation of recombinant Rv2613c was greater than 95% as judged by the results of N-terminal amino acid sequence analysis and mass spectrometry. N-terminal amino acid sequence analysis revealed that the N-terminal methionine of the purified Rv2613c was deleted. The molecular weight was determined to be 24683.45 Da by time-of-flight mass spectrometry, consistent with that calculated from the amino acid sequence containing His-tag. Analysis of gel filtration chromatography indicated that the molecular weight of the protein is approximately 100 kDa (Fig. 2B), indicating that Rv2613c formed a homotetramer of 25 kDa subunits in solution.

### *Ap4A phosphorylase activity of Rv2613c*

The amino acid sequence of Rv2613c contained an HIT motif (Fig. 1). This motif has been reported to be a characteristic feature of several Ap4A hydrolases that convert Ap4A to ATP and AMP or 2 ADP. However, Ap4A was unexpectedly converted to ATP and ADP, but not AMP, in the presence of 5 mM phosphate added to 50 mM HEPES (pH 7.6), 1 mM $MnCl_2$, and 0.1 mM Ap4A (Fig. 3B). This activity increased in a phosphate concentration-dependent manner (Fig. 4A). When other dinucleoside polyphosphate [Np($n$)N'] substrates were used, such as 5',5'''-$P^1$,$P^3$-triphosphate (Ap3A), diadenosine 5',5'''-$P^1$,$P^5$-pentaphosphate (Ap5A), diadenosine 5',5'''-$P^1$,$P^6$-hexaphosphate (Ap6A), diguanosine 5',5'''-$P^1$,$P^4$-tetraphosphate (Gp4G), and diguanosine 5',5'''-$P^1$,$P^5$-pentaphosphate (Gp5G), the major products were Np($n$ – 1) and N'DP, but not N'MP (Table 2). When AMP–$NH_2$ was used as a substrate, hydrolyzed products were not detected, including AMP (Table 2). These results indicate that the catalytic function of Rv2613c resembles a phosphorylase rather than a hydrolase. Thus, Rv2613c appears to be an Ap4A phosphorylase.

### *Substrate specificity and kinetic parameters*

Rv2613c activity was further examined using various substrates listed in Table 2. Ap5A, $P^1$-(5'-adenosyl)$P^4$-(5'-guanosyl) tetraphosphate (Ap4G), $P^1$-(5'-adenosyl)$P^5$-(5'-guanosyl) pentaphosphate (Ap5G), Gp4G and Gp5G, as well as Ap4A, were efficiently phosphorylated by Rv2613c (Table 2). Ap3A and Ap6A were less efficiently phosphorylated, and substrates with a single nucleoside were barely phosphorylated (Table 2). This suggested that the optimum substrates for Rv2613c were dinucleoside polyphosphates containing four or five phosphate residues.

The $K_m$ and $V_{max}$ values for Ap4A and phosphate were determined by evaluating the rates using varying concentrations of each substrate (Fig. 4). The $K_m$ values for Ap4A and phosphate were 0.10 ± 0.001 and 0.94 ± 0.062 mM, respectively, and the $V_{max}$ values for Ap4A and phosphate were 21.87 ± 0.765 and 26.96 ± 0.687 U/mg, respectively. It has been reported that yeast Ap4A phosphorylase catalyzed the reverse phosphorolysis reaction, which resulted in the production of Ap4A and phosphate from ATP and ADP [14]. However, Rv2613c did not exhibit this activity (Fig. 3C).

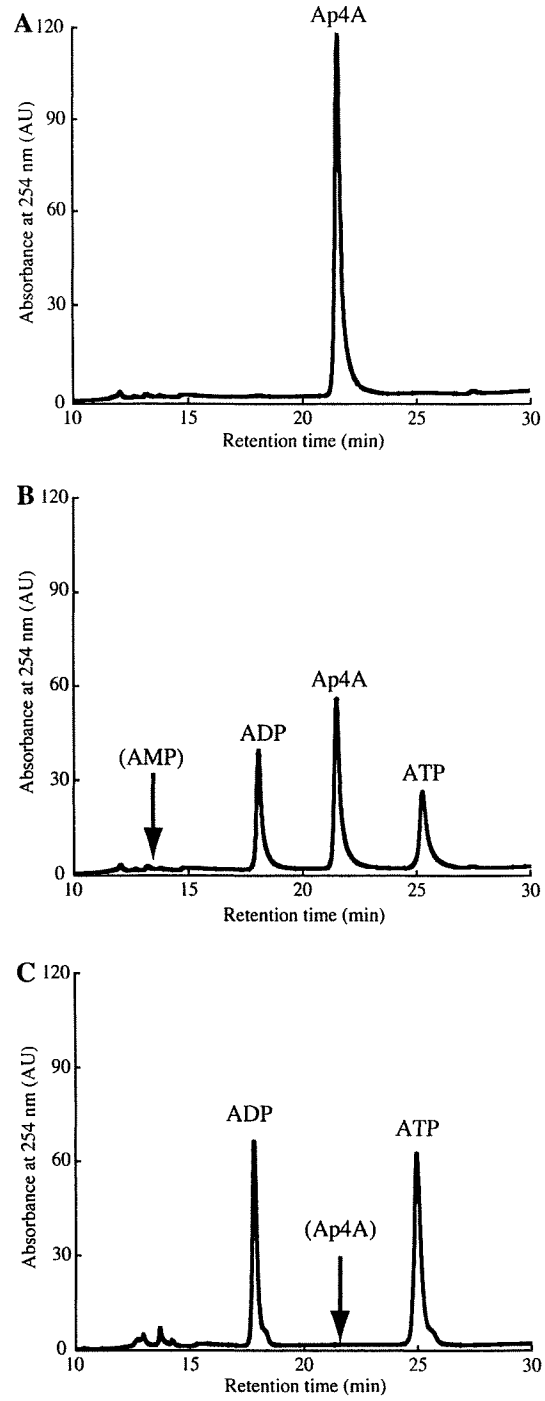


**Fig. 3.** HPLC analyses of products and substrates, which were performed in the absence (A) or presence (B) of purified Rv2613c (0.2 μg) using the conditions described in "Materials and methods". (C) Instead of Ap4A and phosphate, ATP and ADP were used as substrates in the presence of purified Rv2613c (10 μg) at 37 °C for 2 h. The Ap4A, ADP, and ATP peaks are indicated. The retention times for AMP (B) and Ap4A (C) are indicated by arrows.

### *Effects of bivalent metal ions, pH, and temperature*

Rv2613c required bivalent metal ions such as $Mn^{2+}$, $Co^{2+}$, $Ca^{2+}$, and $Mg^{2+}$ to exert its Ap4A phosphorylase activity similar to yeast Ap4A phosphorylase (Table 3) [15,16]. The most efficient activity occurred in the presence of $Mn^{2+}$. Rv2613c activity was observed over a broad pH range (7.4–10.0), with optimal activity at pH 8.0 (Fig. 5A). The optimal temperature for Rv2613c activity was 30 °C

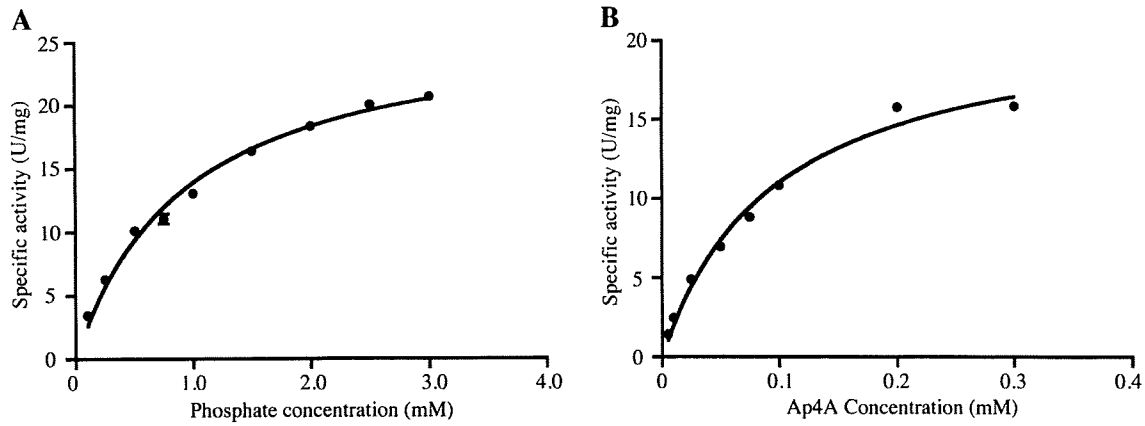


**Fig. 4.** The effect of substrate concentration on Ap4A phosphorylase activity. Varying concentrations of phosphate (A) or Ap4A (B) were added to reaction mixture as described in Materials and methods. Error bars indicate the standard error for three experiments.

**Table 2**
Nucleotide substrate utilization by Rv2613c.

| Nucleotides | Relative activity (%)[a] | Major products |
|---|---|---|
| Ap3A | 19 | ADP |
| Ap4A | 100 | ATP, ADP |
| Ap5A | 106 | p4A, ADP |
| Ap6A | 25 | p5A (probably)[b], ADP |
| Ap4G | 100 | ATP, GDP |
| Ap5G | 76 | p4A, GDP |
| Gp4G | 107 | GTP, GDP |
| Gp5G | 106 | p4G, GDP |
| Adenosine tetraphosphate (p4A) | <1[c] | ADP |
| ATP | <1 | ADP |
| ADP-ribose | <1 | ADP |
| ADP-glucose | <1 | ADP |
| Guanosine tetraphosphate (p4G) | <1 | GDP |
| Guanosine 5′-triphosphate (GTP) | <1 | GDP |
| Guanosine 5′-diphoshate glucose (GDP-glucose) | <1 | GDP |
| GDP-mannose | <1 | GDP |
| Adenylyl imidodiphosphate | <1 | ADP |
| Guanylyl imidodiphosphate | <1 | GDP |
| AMP-$NH_2$ | ND | – |

ND, not detected.
[a] The relative activity in the presence of 0.1 mM Ap4A was taken as 100%.
[b] The detected peak was considered to be adenosine pentaphosphate (p5A), as p5A was not available.
[c] Relative activity was below 1%.

**Table 3**
Effect of bivalent metal ions.

| Bivalent metal ions | Relative activity (%)[a] |
|---|---|
| None | ND |
| $Mn^{2+}$ | 100 |
| $Co^{2+}$ | 68 |
| $Ca^{2+}$ | 64 |
| $Mg^{2+}$ | 34 |

ND, not detected.
[a] Activity in the presence of 2.0 mM $MnCl_2$ was considered to be 100%.

(Fig. 5B), and the activity completely disappeared after treatment at 65 °C for 10 min. The optimum pH and temperature for Rv2613c activity were similar to those required for yeast Ap4A phosphorylase activity [15].

## Discussion

This study revealed that Rv2613c has phosphorylase activity rather than hydrolase activity against Ap4A (Fig. 3B) and other nucleotides (Table 2). Previously, Ap4A phosphorylase was reported primarily in yeast [14–17] and its activity was found in green algae [18] and cyanobacteria [19]. The physiological functions of Ap4A phosphorylase in vivo are unclear. Nevertheless, it was reported that the regulation of intracellular Ap4A concentra-

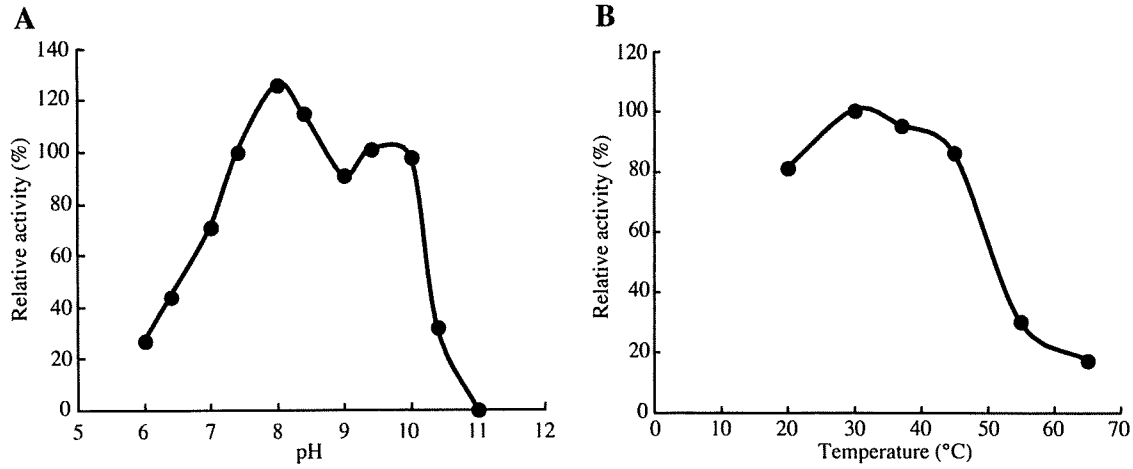


**Fig. 5.** Properties of Rv2613c. pH (A) and temperature (B) dependence of Rv2613c activity. (A) Activity in the presence of HEPES–Na (pH 7.4) was considered to be 100% and (B) the activity at 37 °C was considered to be 100%.

tion is indispensable in bacteria. In *E. coli*, the increment in the amount of Ap4A affects motility and cell metabolism [20]. In *Salmonella enterica*, appropriate regulation of Ap4A concentration is essential for intracellular invasion [21]. From these results, it is conceivable that the control of intracellular Ap4A concentration similarly plays a significant role in the survival and virulence of *M. tuberculosis*, and the inhibition of Rv2613c activity could affect survival and virulence.

It was previously shown that Ap4A phosphorylases have an H-X-H-X-Q-phi-phi motif in place of a typical HIT motif (Fig. 1). Ap4A phosphorylase was considered to be a member of the HIT protein superfamily, which contains a short motif related to the HIT motif [22–24]. However, Rv2613c has the HIT motif, but not the H-X-H-X-Q-phi-phi motif. In addition, unlike yeast Ap4A phosphorylases, which are composed of approximately 330 amino acids, Rv2613c is composed of 195 amino acids and is similar to yeast Ap4A hydrolase (Fig. 1). These observations indicate that Rv2613c is a unique Ap4A phosphorylase with a primary structure homologous to that of Ap4A hydrolase rather than typical Ap4A phosphorylases. Furthermore, mammals, including human, have no primary structures that are homologous to that of yeast Ap4A phosphorylase or Rv2613c. Ap4A phosphorylase activity has not been reported in mammals. Therefore, Rv2613c is a potential target for the development of new anti-TB drugs that act specifically against *M. tuberculosis*.

Several enzymatic properties of Rv2613c, such as optimum pH and temperature, and bivalent metal ion requirement, were similar to those of previously reported yeast Ap4A phosphorylases (Fig. 5, Table 3) [15,16]. These enzymatic properties of Rv2613c were similar to those of *E. coli* and *Homo sapiens* Ap4A hydrolases as well [25,26]. Under pH 6.8–7.0, which is a physiological condition for *M. tuberculosis* growth, Rv2613c actually showed Ap4A phosphorylase activity (Fig. 5A), indicating that it actually functions as Ap4A phosphorylase in *M. tuberculosis* cells. While yeast Ap4A phosphorylase activity catalyzes the reverse reaction by which Ap4A was synthesized from ATP and ADP at pH 7.0 [27], Rv2613c did not show this activity (Fig. 3C), suggesting that other enzymes are responsible for the synthesis of Ap4A in *M. tuberculosis*. It was reported that aminoacyl-tRNA synthetases are involved in the production of Ap4A [28,29]. It is possible that a neighboring gene, the *Rv2614c* gene, which is predicted to encode threonyl-tRNA synthetase, is involved in the synthesis of Ap4A in *M. tuberculosis*.

Previous analysis of crystal structures revealed that the nucleotide-binding mode of the HIT protein superfamily was conserved and that the HIT motif constituted the part of substrate binding site [30]. As Rv2613c possesses HIT motif, it would be likely that Rv2613c also has a similar structure of binding mode of nucleotides. Another study showed that the His residue at the C-terminal end of the HIT motif plays a critical role in the hydrolysis process of the reaction by an amino acid substitution experiment [31]. As shown in this study, Rv2613c contains HIT motif but exhibited phosphorylase activity rather than hydrolase activity. These observations suggest that other amino acid residues common with Ap4A phosphorylase, such as Leu(82), Val(95), Asn(139), and Ser(147), may participate in the process of addition of phosphate that takes place after the cleavage process of substrates. Crystal structure analysis for Rv2613c is currently underway in order to determine its conformation with the aim of developing novel anti-TB drugs.

## Acknowledgments

The expression and purification of Rv2613c were carried out using a grant (H19-Shinkou-Ippan-006) from the Ministry of Health, Labour and Welfare of Japan. Molecular characterization of Rv2613c was carried out using a Grant-in-Aid for Young Scientists ((B) 20780066) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

[1] World Health Organization, WHO report 2009: Global Tuberculosis Control, Epidemiology, Strategy, Financing, WHO, Geneva, 2009.
[2] C.M. Sassetti, D.H. Boyd, E.J. Rubin, Genes required for mycobacterial growth defined by high density mutagenesis, Mol. Microbiol. 48 (2003) 77–84.
[3] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murph, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[4] M. Jackson, D.C. Crick, P.J. Brennan, Phosphatidylinositol is an essential phospholipid of mycobacteria, J. Biol. Chem. 275 (2000) 30092–30099.
[5] J. Kordulakova, M. Gilleron, G. Puzo, P.J. Brennan, B. Gicquel, K. Mikusova, M. Jackson, Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species, J. Biol. Chem. 278 (2003) 36285–36295.
[6] M.E. Guerin, J. Kordulakova, F. Schaeffer, Z. Svetlikova, A. Buschiazzo, D. Giganti, B. Gicquel, K. Mikusova, M. Jackson, P.M. Alzari, Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from mycobacteria, J. Biol. Chem. 282 (2007) 20705–20714.
[7] P. Bieganowski, P.N. Garrison, S.C. Hodawadekar, G. Faye, L.D. Barnes, C. Brenner, Adenosine monophosphoramidase activity of Hint and Hnt1 supports function of Kin28, Ccl1, and Tfb3, J. Biol. Chem. 277 (2002) 10852–10860.
[8] Y. Huang, P.N. Garrison, L.D. Barnes, Cloning of the *Schizosaccharomyces pombe* gene encoding diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate (Ap4A) asymmetrical hydrolase: sequence similarity with histidine triad (HIT) protein family, Biochem. J. 312 (1995) 925–932.
[9] L.D. Barnes, P.N. Garrison, Z. Siprashvili, A. Guranowski, A.K. Robinson, S.W. Ingram, C.M. Croce, M. Ohta, K. Huebner, Fhit, a putative tumor suppressor in humans, is a dinucleoside 5′,5‴-P$^1$,P$^3$-triphosphate hydrolase, Biochemistry 35 (1996) 11529–11535.
[10] V. Pelicic, M. Jackson, J.M. Reyrat, W.R. Jacobs Jr., B. Gicquel, C. Guilhot, Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*, Proc. Natl. Acad. Sci. USA 94 (1997) 10955–10960.
[11] N.R. Leslie, A.G. McLennan, S.T. Safrany, Cloning and characterisation of hAps1 and hAps2, human diadenosine polyphosphate-metabolising Nudix hydrolases, BMC Biochem. 3 (2002) 20.
[12] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[13] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[14] P. Plateau, M. Formant, J.M. Schmitter, J.M. Buhler, S. Blanquet, Isolation, characterization, and inactivation of the *APA1* gene encoding Yeast diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate phosphorylase, J. Bacteriol. 171 (1989) 6437–6445.
[15] A. Guranowski, S. Blanquet, Phosphorolytic cleavage of diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate. Properties of homogeneous diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate alpha, beta-phosphorylase from *Saccharomyces cerevisiae*, J. Biol. Chem. 260 (1985) 3542–3547.
[16] P. Plateau, M. Fromant, J.M. Schmitter, S. Blanquet, Catabolism of bis(5′-nucleosidyl) tetraphosphates in *Saccharomyces cerevisiae*, J. Bacteriol. 172 (1990) 6892–6899.
[17] W. Mulder, I.H. Scholten, H. Roon, L.A. Grivell, Isolation and characterisation of the linked genes APA2 and QCR7, coding for Ap4A phosphorylase II and the 14 kDa subunit VII of the mitochondrial bc1-complex in the yeast *Kluyveromyces lactis*, Biochim. Biophys. Acta 1219 (1994) 719–723.
[18] A.G. McLennan, E. Mayers, S. Hankin, N.M. Thorne, M. Prescott, R. Powls, The green alga *Scenedesmus obliquus* contains both diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate (asymmetrical) pyrophosphohydrolase and phosphorylase activities, Biochem. J. 300 (1994) 183–189.
[19] A.G. Mclennan, E. Mayers, D.G. Adams, Anabaena flos-aquae and other cyanobacteria possess diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate (Ap4A) phosphorylase activity, Biochem. J. 320 (1996) 795–800.
[20] S.B. Farr, D.N. Arnosti, M.J. Chamberlin, B.N. Ames, An apaH mutation causes AppppA to accumulate and affects motility and catabolite repression in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 86 (1989) 5010–5014.
[21] T.M. Ismail, C.A. Hart, A.G. McLennan, Regulation of dinucleoside polyphosphate pools by the YgdP and ApaH hydrolases is essential for the ability of *Salmonella enterica* serovar typhimurium to invade cultured mammalian cells, J. Biol. Chem. 278 (2003) 32602–32607.
[22] B. Séraphin, The HIT protein family: a new family of proteins present in prokaryotes, yeast and mammals, DNA Sequence 3 (1992) 177–179.
[23] C. Brenner, P. Bieganowski, H.C. Pace, K. Huebner, The histidine triad superfamily of nucleotide-binding proteins, J. Cell. Physiol. 181 (1999) 179–187.

[24] C. Brenner, Hint, Fhit, and GalT: function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases, Biochemistry 41 (2002) 9003–9014.
[25] A. Guranowski, H. Jakubowski, E. Holler, Catabolism of diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate in procaryotes. Purification and properties of diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate (symmetrical) pyrophosphohydrolase from *Escherichia coli* K12, J. Biol. Chem. 258 (1983) 14784–14789.
[26] D. Lazewska, E. Starzynska, A. Guranowski, Human placental (*Asymmetrical*) diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate hydrolase: purification to homogeneity and some properties, Protein Expr. Purif. 4 (1993) 45–51.
[27] A. Brevet, H. Coste, M. Fromant, P. Plateau, S. Blanquet, Yeast diadenosine 5′,5‴-P$^1$,P$^4$-tetraphosphate alpha, beta-phosphorylase behaves as a dinucleoside tetraphosphate synthetase, Biochemistry 26 (1987) 4763–4768.
[28] P.C. Zamecnik, M.L. Stephenson, C.M. Janeway, K. Randerath, Enzymatic synthesis of diadenosine tetraphosphate and diadenosine triphosphate with a purified lysyl-sRNA synthetase, Biochem. Biophys. Res. Commun. 6 (1966) 91–97.
[29] O. Goerlich, R. Foeckler, E. Holler, Mechanism of synthesis of adenosine(5′)tetraphospho(5′)adenosine (AppppA) by aminoacyl-tRNA synthetases, Eur. J. Biochem. 126 (1982) 135–142.
[30] C. Brenner, P. Garrison, J. Gilmour, D. Peisach, D. Ringe, G.A. Petsko, J.M. Lowenstein, Crystal structures of HINT demonstrate that histidine triad proteins are GalT-related nucleotide-binding proteins, Nat. Struct. Biol. 4 (1997) 231–238.
[31] H.C. Pace, P.N. Garrison, A.K. Robinson, L.D. Barnes, A. Draganescu, A. Rösler, G.M. Blackburn, Z. Siprashvili, C.M. Croce, K. Huebner, C. Brenner, Genetic, biochemical, and crystallographic characterization of Fhit–substrate complexes as the active signaling form of Fhit, Proc. Natl. Acad. Sci. USA 95 (1998) 5484–5489.