doi:10.1016/j.jmb.2011.04.059 J. Mol. Biol. (2011) 410, 93–104

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structural Insights into the Novel Diadenosine 5′,5‴-P^1,P^4-Tetraphosphate Phosphorylase from *Mycobacterium tuberculosis* H37Rv

Shigetarou Mori*, Keigo Shibayama, Jun-Ichi Wachino and Yoshichika Arakawa

Department of Bacteriology II, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashi-Murayama-shi, Tokyo 208-0011, Japan

Received 25 October 2010; received in revised form 22 April 2011; accepted 23 April 2011
Available online 4 May 2011

Edited by M. Guss

Keywords:
X-ray crystallography; mycobacteria; nucleotide; mutation; kinetics

Rv2613c is a diadenosine 5′,5‴-P^1,P^4-tetraphosphate (Ap_4A) phosphorylase from *Mycobacterium tuberculosis* H37Rv. Sequence analysis suggests that Rv2613c belongs to the histidine triad (HIT) motif superfamily, which includes HIT family diadenosine polyphosphate (Ap_nA) hydrolases and Ap_4A phosphorylases. However, the amino acid sequence of Rv2613c is more similar to that of HIT family Ap_nA hydrolases than to that of typical Ap_4A phosphorylases. Here, we report the crystal structure of Rv2613c, which is the first structure of a protein with Ap_nA phosphorylase activity, and characterized the structural basis of its catalytic activity. Our results showed that the structure of Rv2613c is similar to those of other HIT superfamily proteins. However, Asn139, Gly146, and Ser147 in the active site of Rv2613c replace the corresponding Gln, Gln, and Thr residues that are normally found in HIT family Ap_nA hydrolases. Furthermore, analyses of Rv2613c mutants revealed that Asn139, Gly146, and Ser147 are important active-site residues and that Asn139 has a critical role in catalysis. The position of Gly146 might influence the phosphorylase activity. In addition, the tetrameric structure of Rv2613c and the presence of Trp160 might be essential for the formation of the Ap_4A binding site. These structural insights into Rv2613c may facilitate the development of novel structure-based inhibitors for treating tuberculosis.

Introduction

Mycobacterium tuberculosis is the causative agent of tuberculosis (TB). Every year, approximately 1.7 million people die from TB worldwide, and more than 8.9 million people are newly infected with *M. tuberculosis*.[1] Furthermore, multidrug-resistant TB and extensively drug-resistant TB have become serious problems recently.[2] As a result, novel anti-TB drugs are needed urgently. To facilitate the structure-based design of new anti-TB drugs, our group has investigated the structure–function relationships of *M. tuberculosis* proteins such as the NAD kinase–NAD complex[3] and Rv2613c, which is a novel diadenosine 5′,5‴-P^1,P^4-tetraphosphate (Ap_4A) phosphorylase (EC 2.7.7.53).[4] The *Rv2613c* gene was previously shown to be an essential gene in *M. tuberculosis* H37Rv,[5] and Rv2613c was recognized as a target for new anti-TB drugs by *in silico* analysis.[6] Therefore, we have determined the

**Corresponding author*. E-mail address: mshige@nig.go.jp.

Abbreviations used: Ap_4A, diadenosine 5′,5‴-P^1,P^4-tetraphosphate; HIT, histidine triad; Ap_nA, diadenosine polyphosphate; TB, tuberculosis; SAD, single-wavelength anomalous dispersion; PDB, Protein Data Bank; AMW, adenosine monotungstate; IB2, P^1,P^2-methylene-P^3-thio-diadenosine triphosphate; SeMet, selenomethionine.

0022-2836/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.

three-dimensional structure of Rv2613c and elucidated its structure–function relationships.

Since the amino acid sequence of Rv2613c has a histidine triad (HIT) motif (His-φ-His-φ-His-φ-φ, where φ represents a hydrophobic amino acid), Rv2613c is thought to belong to the HIT superfamily. This superfamily also includes nucleotide hydrolases and transferases such as adenosine monophosphate (AMP) lysine hydrolases, HIT family diadenosine polyphosphate (Ap_nA) hydrolases, galactose-1-phosphate uridylyltransferases, and Ap_4A phosphorylases.[7] The amino acid sequence of Rv2613c has both similarities with and differences from the sequences of other Ap_4A phosphorylases in the HIT superfamily. For example, *Kluyveromyces lactis* Ap_4A phosphorylase (Swiss-Prot ID: APA2_KLULA) and *Saccharomyces cerevisiae* Ap_4A phosphorylases 1 and 2 (Swiss-Prot IDs: APA1_YEAST and APA2_YEAST) have a His-X-His-X-Gln motif instead of the HIT motif.[8,9] In addition, unlike these phosphorylases, which are approximately 330 amino acids long,[8,9] Rv2613c is only 195 amino acids long; therefore, its length is similar to that of *Schizosaccharomyces pombe* Ap_nA hydrolase (Swiss-Prot ID: APH1_SCHPO), which is 182 amino acids long.[10] Furthermore, the amino acid sequence of Rv2613c is more similar to that of HIT family Ap_nA hydrolases than typical Ap_4A phosphorylases. For instance, the amino acid sequence of Rv2613c is 28.8% identical and 61.9% similar to *Sc. pombe* Ap_nA hydrolase, whereas it is 16.2% identical and 60.7% similar to *S. cerevisiae* Ap_4A phosphorylase 1. Collectively, these observations suggest that the amino acid sequence of Rv2613c is more related to HIT family Ap_nA hydrolases than to typical Ap_4A phosphorylases.

Furthermore, we recently showed that Rv2613c converts Ap_4A into ATP and adenosine 5′-diphosphate (ADP) in the presence of inorganic phosphate (Eq. (1)), so it is a phosphorylase rather than a hydrolase.[4] In contrast, HIT family Ap_nA hydrolases convert Ap_nA into Ap_{n-1} and AMP (Eq. (2)):[7]

$$Ap_4A + \text{inorganic phosphate} \rightarrow ATP + ADP \quad (1)$$

$$Ap_nA \rightarrow Ap_{n-1} + AMP \quad (2)$$

Previous crystallographic studies of HIT superfamily proteins have revealed that the HIT motif is part of the substrate binding site and that the conserved central His residue in the HIT motif is important for catalytic activity.[11,12] In addition, a Ser residue, which is upstream of the HIT motif, is important for the catalytic activity of AMP-lysine hydrolase[13] and galactose-1-phosphate uridylyltransferase.[14] However, the three-dimensional structure and catalytic residues of an Ap_4A phosphorylase are not known. Therefore, we aimed to determine the crystal structure of Rv2613c and to characterize and compare the structural basis of its catalytic activity with those of HIT family Ap_nA hydrolases and typical Ap_4A phosphorylases.

Previously, we also showed that the multimerization of Rv2613c differs from those of HIT family Ap_nA hydrolases and typical Ap_4A phosphorylases.[4] Specifically, Rv2613c is homotetrameric in solution,[4] whereas HIT family Ap_nA hydrolases and typical Ap_4A phosphorylases are homodimeric[15] and monomeric,[16,17] respectively. Therefore, we also investigated the importance of multimerization for the catalytic activity of Rv2613c.

Recently, we reported the crystallization and preliminary X-ray analysis of Rv2613c.[18] Here, we describe the 1.89-Å-resolution crystal structure of Rv2613c and provide insights into the structural basis of its catalytic activity. To our knowledge, this is the first report of the crystal structure of a protein with Ap_nA phosphorylase activity.

Results

Structure determination

Previously, we showed that the Rv2613c crystal belongs to the *C*2 space group and has unit cell parameters of $a = 101.5$ Å, $b = 63.6$ Å, $c = 79.1$ Å, and $\beta = 110.9°$.[18] In addition, we reported that Rv2613c exists as a homotetramer of 25-kDa subunits in solution[4] and that there are two subunits per asymmetric unit, which are characterized by a V_m of 2.41 $Å^3$ Da^{-1} and a solvent content of 49.1%.[18] Here, we solved the crystal structure of Rv2613c at 1.89 Å resolution by using single-wavelength anomalous dispersion (SAD). The final *R*-factor and R_{free} were 17.6% and 20.1%, respectively (Table 1). The final model of Rv2613c consisted of 325 residues, 233 water molecules, 3 phosphate ions, and 2 tetraethylene glycols in the asymmetric unit, which included subunits A and B (Rv2613c-A/-B) (Fig. 1a). Electron density was present for all residues in Rv2613c-A, except for the His tag and for residues 1–13, 36–54, and 171–175. Similarly, electron density was present for all residues in Rv2613c-B, except for the His tag and for residues 1–23. In addition, residues 196–198 in Rv2613c-B were derived from the expression vector. The average *B*-factors for all atoms of Rv2613c-A, Rv2613c-B, water molecules, phosphate ions, and tetraethylene glycols were 32.7, 26.5, 36.3, 36.6, and 43.8 $Å^2$, respectively. Ramachandran plot analysis showed that the φ/ψ angle pairs of most residues (322/325; 99.1%) were in the favored regions, whereas the pairs of the remaining residues (3/325; 0.9%) were in the allowed regions (Table 1).

Table 1. Data collection and structure refinement statistics

	Native	SeMet
Data collection		
Wavelength (Å)	1.00000	0.97901
Resolution range (Å)	50.0–1.89 (1.93–1.89)[a]	50.0–2.50 (2.54–2.50)
Space group	*C*2	*C*2
Cell dimensions *a*, *b*, *c* (Å)	101.5, 63.6, 79.1	101.3, 63.5, 79.5
Cell dimension β (°)	110.9	110.6
Unique reflections	37,314 (1826)	16,529 (808)
Average redundancy	7.4 (6.4)	7.6 (7.3)
Completeness (%)	99.8 (98.3)	99.9 (99.9)
Mean $I/\sigma(I)$	42.2 (4.4)	33.6 (6.9)
R_{merge}[b]	6.1 (38.0)	8.6 (33.4)
Refinement		
Resolution range	39.4–1.89 (1.94–1.89)	
Number of reflections	36,096 (2327)	
R-factor (%)[c]	17.6 (19.5)	
R_{free} (%)[d]	20.1 (26.6)	
RMSD from ideality		
Bonds (Å)	0.005	
Angles (°)	1.025	
Ramachandran plot		
Favored region (%)	99.1	
Allowed region (%)	0.9	
Outlier region (%)	0	

[a] Data for the highest-resolution shells are given in parentheses.
[b] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is its average.
[c] *R*-factor $= \sum_{hkl}|F_{obs} - k|F_{calc}| / \sum_{hkl}|F_{obs}|$, where *k* is the scaling factor.
[d] This value is based on 5% of the data excluded from refinement at random.

Protein architecture and conformation of prosthetic groups

The overall structure of Rv2613c consisted of a small N-terminal domain (residues 1–39) and a large C-terminal domain (residues 40–195). In Rv2613c-A, the N-terminal domain consisted of two β-strands, whereas in Rv2613c-B, the N-terminal domain consisted of one α-helix and one β-strand (Fig. 1a and b). The C-terminal domain formed an α/β structure in both Rv2613c-A and Rv2613c-B, and was involved in the dimer interaction between Rv2613c-A and Rv2613c-B in the asymmetric unit (Fig. 1a and b). The main structural feature of this dimer was an antiparallel β-sheet with a normal left-handed twist (Fig. 1a). This β-sheet consisted of 10 β-strands, which included strands C, D, E, G, and F from Rv2613c-A, and strands M, N, L, K, and J from Rv2613c-B (Fig. 1a and b). Two long α-helices (α4 in Rv2613c-A and α12 in Rv2613c-B) also were involved in the dimer interaction and were located inside this β-sheet (Fig. 1a and b).

The quaternary structure of Rv2613c was a tetramer consisting of a dimer of dimers (Rv2613c-A/-B + Rv2613c-A′/-B′) (Fig. 1c), in agreement with our previous observation that Rv2613 tetramerizes in solution.[4] The N-terminal domain was involved in intersubunit contacts (Fig. 1b and c). Specifically, at the A–B′ contact, a mixed β-sheet and a parallel β-sheet consisted of three β-strands (A and H from Rv2613c-A and I′ from Rv2613c-B′) and two β-strands (B from Rv2613c-A and O′ from Rv2613c-B′). Similarly, at the B–A′ contact, a mixed β-sheet and a parallel β-sheet also were composed of three β-strands (I from Rv2613c-B, and A′ and H′ from Rv2613c-A′) and two β-strands (O from Rv2613c-B and B′ from Rv2613c-A′) (Fig. 1b and c). In addition, three α-helices (α6 in Rv2613c-A, and α8 and α15 in Rv2613c-B) were involved in this tetramerization (Fig. 1c).

In Rv2613c-A, one phosphate ion and one tetraethylene glycol were bound to the C-terminal domain, while two phosphate ions and one tetraethylene glycol were bound to the C-terminal domain in Rv2613c-B (Fig. 1a). One phosphate ion was bound to the same binding site in each subunit, while the other phosphate ion in Rv2613c-B was bound to the C-terminal domain (Fig. 1a). In addition, the binding site and conformation of tetraethylene glycol molecule in Rv2613c-A were different from those of the tetraethylene glycol molecule in Rv2613c-B (Fig. 1a).

Structural comparisons

We searched for similar protein structures. As a result, the structure of Rv2613c is similar to those of other proteins in the HIT superfamily such as *Homo sapiens* fragile HIT protein [Fhit; Protein Data Bank (PDB) IDs: 6FIT and 1FHI; *Z*-score = 16.1 and 15.8, respectively], which is a HIT family Ap_nA hydrolase;[11,12] Zn-bound HIT family protein from *Mycobacterium smegmatis* (MSMEG5028; PDB ID: 3O0M; *Z*-score = 16.1); and ADP-glucose phosphorylase from *Arabidopsis thaliana* (Ath; PDB ID: 1Z84; *Z*-score = 11.5).[19] For example, superimposition of the structures of Rv2613c, Fhit, and MSMEG5028 revealed that the structure of Rv2613c-A/-B is similar to those of the Fhit and MSMEG5028 dimers (Fig. 2a). In particular, the 10-stranded antiparallel β-sheet and inner long α-helices of the Rv2613c dimer were in good agreement with those of the Fhit and MSMEG5028 dimers (Fig. 2a). In contrast, neither Fhit nor MSMEG5028 was structurally similar to the N-terminal domain of Rv2613c (Fig. 2a).

In addition, since this is the first report of the crystal structure of a protein with Ap_nA phosphorylase activity, we compared the structure of Rv2613c with that of Ath in place of a typical Ap_4A phosphorylase. Ath has a His-X-His-X-Gln motif, which is characteristic of typical Ap_4A phosphorylases. In addition, the amino acid sequence of Ath is 36% identical and 56% similar to *S. cerevisiae* Ap_4A phosphorylase 1. Therefore, we considered the

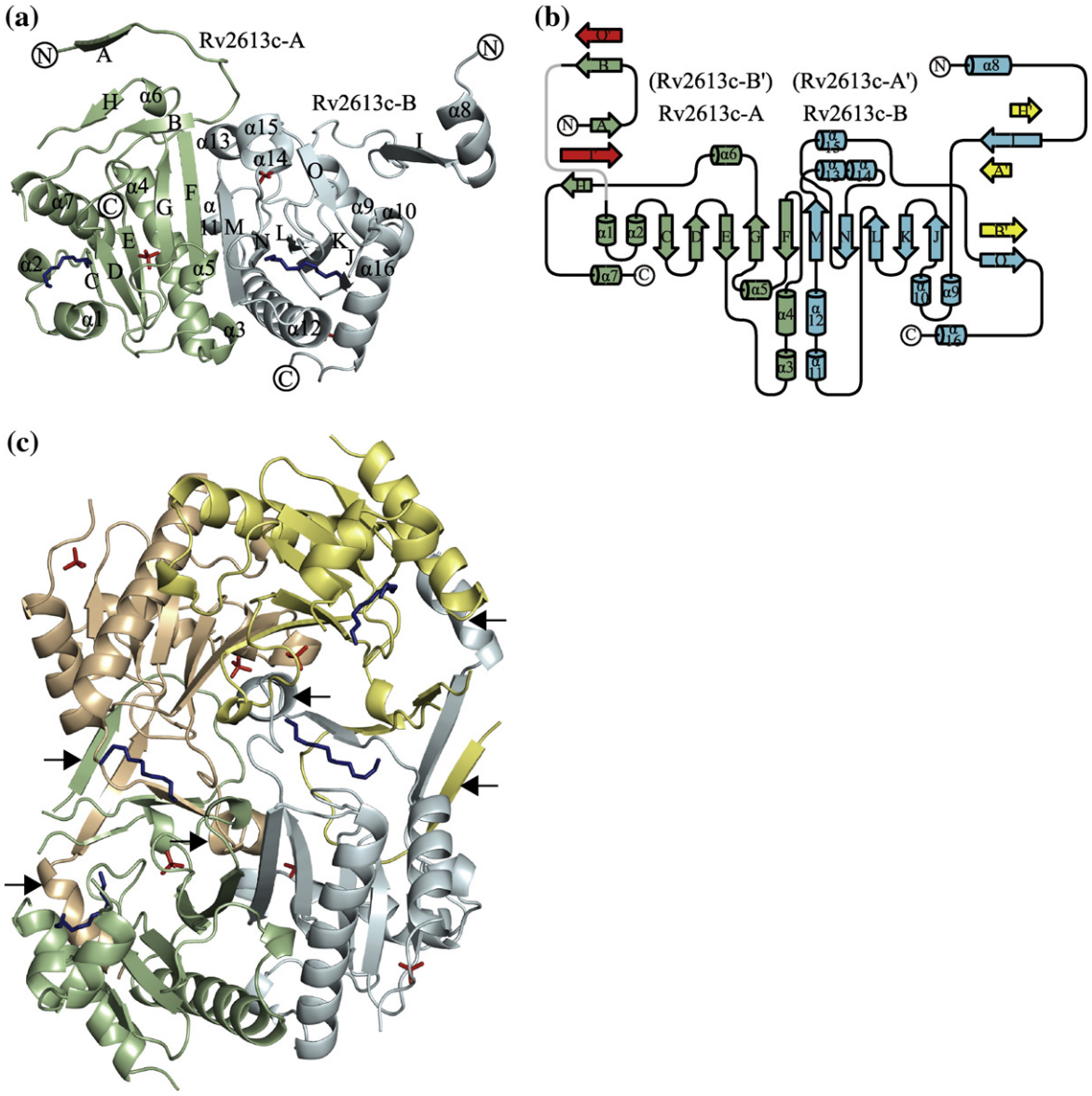

Fig. 1. Structure of Rv2613c. (a) Ribbon diagram of Rv2613c subunit A (green) and subunit B (cyan) (Rv2613c-A/-B) in the asymmetric unit. The N-terminus and the C-terminus are indicated by circled letters. α-Helices are numbered, and β-strands are lettered in order from the N-terminus of Rv2613c-A to the C-terminus of Rv2613c-B. Phosphate ions and tetraethylene glycols are modeled as red and blue sticks, respectively. (b) Secondary structure of the Rv2613c dimer shown in (a). Rv2613c-A and Rv2613c-B are indicated in green and cyan, respectively. The N-terminus, C-terminus, α-helices, and β-strands are labeled as in (a). In Rv2613c-A, the loops, which were not modeled due to poor electron density, are shown as gray lines. The three β-strands from Rv2613c-A′ and the two β-strands from Rv2613c-B′ are shown in yellow and red, respectively. (c) Ribbon diagram of the quaternary structure of Rv2613c. The asymmetric unit contains two subunits (Rv2613c-A/-B or Rv2613c-A′/-B′). Rv2613c-A, Rv2613c-B, Rv2613c-A′, and Rv2613c-B′ are shown in green, cyan, yellow, and light brown, respectively. Phosphate ions and tetraethylene glycols are represented as in (a). The arrows show A–B′ and B–A′ contacts.

structure of Ath an appropriate comparison structure. Superimposition of the structures of Rv2613c and Ath showed that the structure of Rv2613c-A/-B is similar to that of the Ath monomer (Fig. 2b). In particular, the 10-stranded antiparallel β-sheet and inner long α-helices of the Rv2613c dimer were in good agreement with that of the Ath monomer (Fig. 2b). In addition, Ath exhibited structural similarity to the N-terminal domain of Rv2613c (Fig. 2b). Furthermore, the structure of the Rv2613c tetramer was similar to that of the Ath dimer. Moreover, the interaction between Rv2613c-A/-B and Rv2613c-A′/-B′ also was similar to the interaction in the Ath dimer. Finally, unlike the crystal

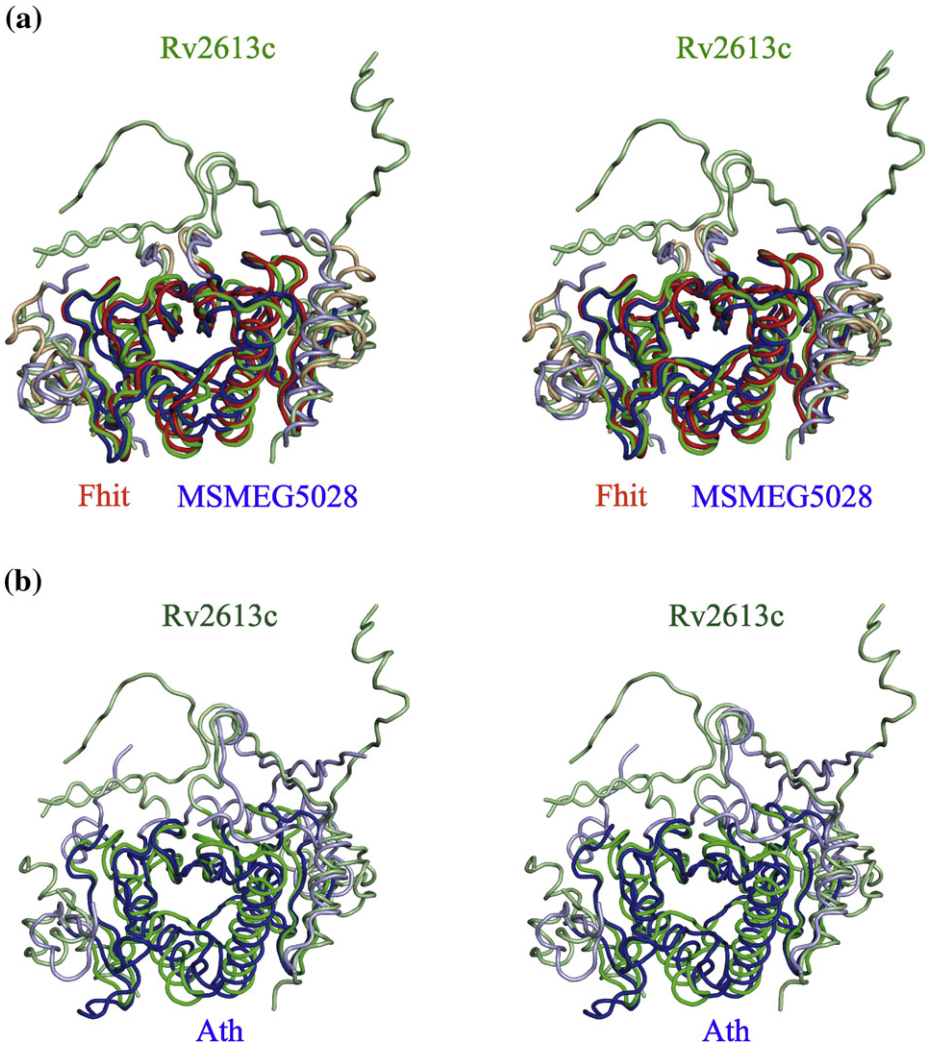

Fig. 2. Stereo representation of structural comparisons between Rv2613c and representative proteins in the HIT superfamily. (a) Superimposition of the structure of Rv2613c-A/-B on those of *H. sapiens* Fhit (PDB ID: 6FIT) and the Zn-bound HIT family protein from *M. smegmatis* (MSMEG5028; PDB ID: 3O0M) dimers. Rv2613c, Fhit, and MSMEG5028 are shown in green, red, and blue, respectively. Regions that are structurally similar are illustrated with darker colors compared to those that are divergent. (b) Superimposition of the structure of Rv2613c-A/-B on that of the ADP-glucose phosphorylase from *A. thaliana* (Ath; PDB ID: 1Z84) monomer. Rv2613c and Ath are shown in green and blue, respectively. Regions that are structurally similar are depicted with darker colors compared to those that are divergent. The structurally similar region of the Ath monomer is probably an internally duplicated subdomain.

structures of MSMEG5028 and Ath, which include a zinc ion, the crystal structures of Rv2613c and Fhit did not include any metal ions.

Catalytically important residues

The binding site of the phosphate ion in Rv2613c (Fig. 3a) corresponded to that of the tungstate ion of adenosine monotungstate (AMW) in Fhit (Fig. 3b) and that of the α-phosphate of AMP in Ath (Fig. 3c). Since HIT superfamily proteins, including Rv2613c, act on the α-phosphate of ribonucleotides,[4,7] it is likely that the phosphate ion binding site of Rv2613c is its active site. In this site, Asn139, Ser147, His153, and His155 were bound to the phosphate ion, and Gly146 was positioned near the phosphate ion (Fig. 3a). Thus, these five residues are probably catalytically important for the enzyme activity of Rv2613c.

Furthermore, the multiple sequence alignment of Rv2613c and related HIT superfamily proteins provides additional evidence for the importance of these residues. For example, His153 and His155 were in the HIT motif, whereas Asn139, Gly146, and Ser147 were upstream of this motif (Fig. 4). In addition, the multiple sequence alignment of the region containing these five residues in the

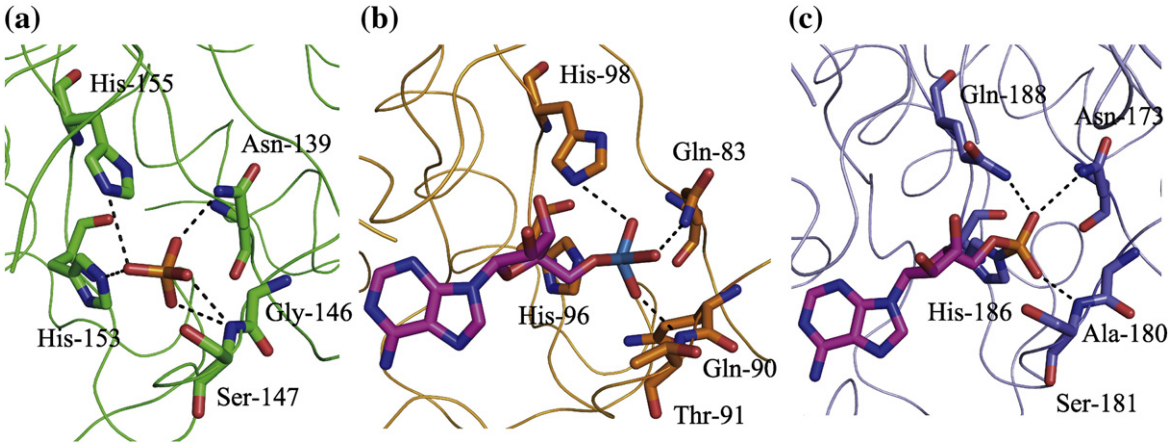

Fig. 3. Comparison of the active sites of Rv2613c, Fhit, and Ath. (a) Active site of Rv2613c. Asn139, Gly146, Ser147, His153, His155, and the phosphate ion are represented by sticks. Carbon, nitrogen, oxygen, and phosphorus atoms are shown in green, blue, red, and orange, respectively. (b) Active site of Fhit (PDB ID: 6FIT). Gln83, Gln90, Thr91, His96, His98, and AMW are represented by sticks. The carbon atoms in the amino acids and adenosine are shown in orange and purple, respectively. Nitrogen, oxygen, and tungsten atoms are shown in blue, red, and cyan, respectively. (c) Active site of Ath (PDB ID: 1Z84). Asn173, Ala180, Ser181, His186, Gln188, and AMP are represented by sticks. The carbon atoms in the amino acids and adenosine are shown in light blue and purple, respectively. Nitrogen, oxygen, and phosphorus atoms are shown in dark blue, red, and orange, respectively.

phosphate binding site of Rv2613c and the corresponding residues in the AMW binding site of Fhit showed that His153 and His155 in Rv2613c correspond to His96 and His98 in Fhit, respectively (Figs. 3a and b and 4). Similarly, Asn139, Gly146, and Ser147 in Rv2613c are equivalent to Gln83, Gln90, and Thr91 in Fhit, respectively (Figs. 3a and b and 4). In addition, the Gln, Gln, and Thr residues, which corresponded to Gln83, Gln90, and Thr91, respectively, in Fhit were conserved among HIT family Ap$_n$A hydrolases (Fig. 4). On the other hand, the multiple sequence alignment showed that Asn139, Ser147, and His153 in the phosphate binding site of Rv2613c corresponded to Asn173, Ser181, and His186 in the AMP binding site of Ath, respectively (Figs. 3a and c and 4). Similarly, Gly146 and His155 in Rv2613c are equivalent to Ala180 and Gln188 in Ath, respectively (Figs. 3a and c and 4). In addition, the Asn and Ser residues, which corresponded to Asn139 and Ser147, respectively, in Rv2613c were conserved among Ap$_4$A phosphorylases and ADP-glucose phosphorylase (Fig. 4).

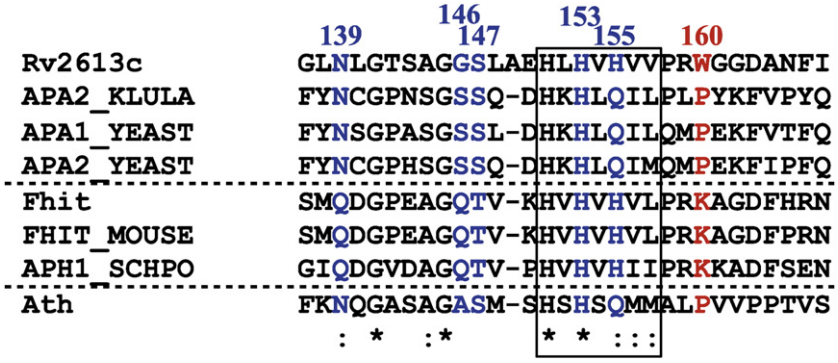

Fig. 4. Multiple sequence alignment of the active-site region of Rv2613c and representative Ap$_4$A phosphorylases (APA2_KLULA, APA1_YEAST, and APA2_YEAST) and HIT family Ap$_n$A hydrolases [Fhit; FHIT_MOUSE (*Mus musculus*); and APH1_SCHPO], and ADP-glucose phosphorylase (Ath). The Swiss-Prot (UniProtKB) IDs of these proteins are as follows: Rv2613c (O06201), APA2_KLULA (P49348), APA1_YEAST (P16550), APA2_YEAST (P22108), Fhit (P49789), FHIT_MOUSE (O89106), APH1_SCHPO (P49776), and Ath (Q9FK51). The position of residues in Rv2613c is shown above the alignment. The residues that correspond to Asn139, Gly146, Ser147, His153, and His155 in Rv2613c are highlighted in blue, while those that correspond to Trp160 in Rv2613c are highlighted in red. The HIT and His-X-His-X-Gln motifs are boxed. Identical and highly conserved residues are denoted by an asterisk (*) or a colon (:), respectively. The multiple sequence alignment was generated by CLUSTAL W 1.83.

Table 2. Kinetic values of Rv2613c and derivatives

Enzyme	Activity (U mg^{-1})[a]	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m (M^{-1} s^{-1})
Wild type	8.69	0.10 ± 0.001	8.48 ± 0.313	84.82
N139A	ND[b]	—	—	—
N139Q	ND	—	—	—
G146S	2.63	0.08 ± 0.014	1.99 ± 0.115	24.88
G146Q	1.19	0.08 ± 0.019	1.38 ± 0.108	17.25
S147A	0.96	0.07 ± 0.015	0.48 ± 0.036	6.90
S147T	1.92	0.11 ± 0.018	1.49 ± 0.130	13.62
N139Q/S147T	ND	—	—	—
W160A	ND	—	—	—

[a] Specific activity determined in the presence of 1.0 mM Ap_4A.
[b] ND, not detected (i.e., enzyme activity was not detected using up to 100 μg of purified enzyme for up to 2 h).

To test the hypothesis that Asn139, Ser147, and Gly146 are catalytically important residues in Rv2613c, we examined the catalytic activities of seven Rv2613c mutants, namely N139A, N139Q, G146S, G146Q, S147A, S147T, and N139Q/S147T. Three mutants (N139A, N139Q, and N139Q/S147T) did not exhibit any detectable catalytic activity, while the activity of the other four mutants (G146S, G146Q, S147A, and S147T) was lower than that of wild-type Rv2613c (Table 2). In addition, the K_m value of the G146S, G146Q, S147A, and S147T mutants was similar to that of wild-type Rv2613c; however, the K_{cat} and K_{cat}/K_m values of these four mutants were markedly lower than those of wild-type Rv2613c (Table 2).

We also determined the importance of Gly146 on the production of AMP during catalysis. Wild-type Rv2613c and G146S produced very little AMP (1.2 and 1.5 nmol from 100 nmol of Ap_4A, respectively; Fig. 5a and b), whereas G146Q produced more AMP (11.2 nmol from 100 nmol of Ap_4A; Fig. 5c).

Relationship between tetramerization and catalytic activity

To elucidate the relationship between the tetramerization and the catalytic activity of Rv2613c, we predicted the binding site of Ap_4A and its mode of binding in Rv2613c by superimposing the α-phosphate of Ap_4A on the phosphate ion in the putative active site of Rv2613c-A. As a result, one adenosine binding site corresponded to the tetraethylene glycol binding site in Rv2613c-A, while the other adenosine binding site corresponded to the tetraethylene glycol binding site in Rv2613c-B′ (Fig. 6a and b). Furthermore, the putative position of adenine, which was bound to Rv2613c-B′, could make π–π stacking interactions with Trp160 in this subunit (Fig. 6b). Therefore, we predicted that the tetramerization of Rv2613c (i.e., A–B′ and B–A′ contacts) is involved in the formation of the Ap_4A binding site and that Trp160 might participate in

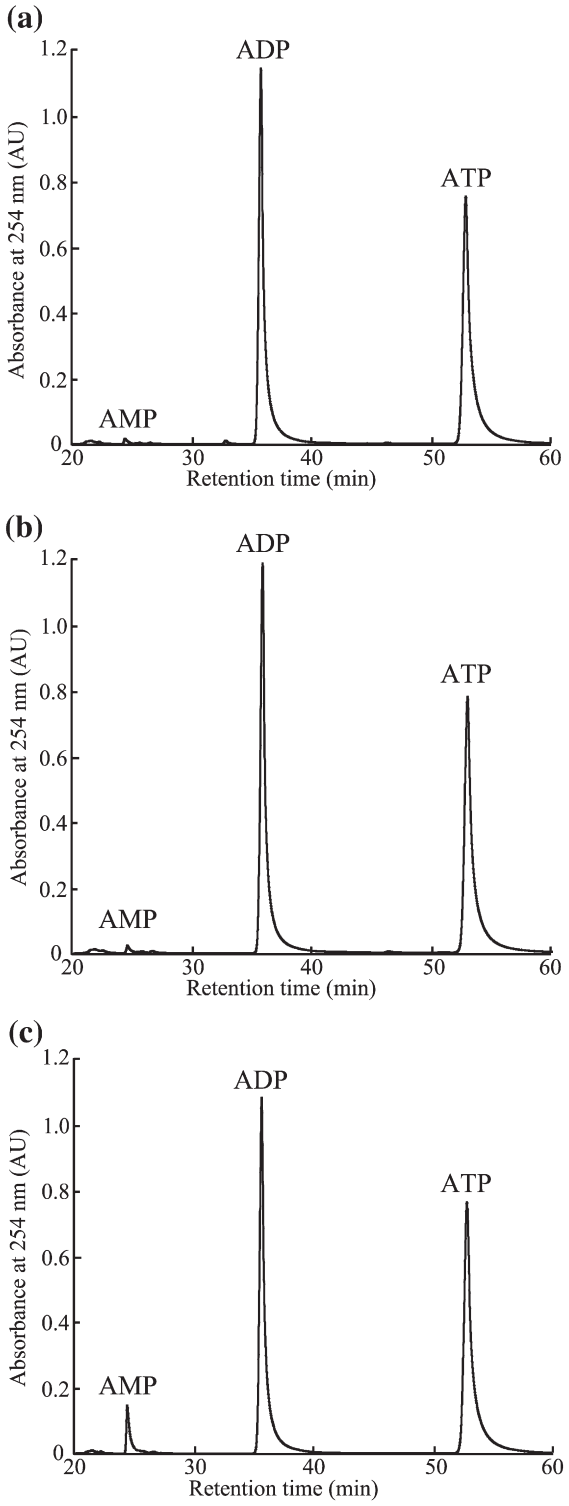

Fig. 5. High-performance liquid chromatography analyses of products formed by the complete degradation of 100 nmol of Ap_4A with purified Rv2613c (a), G146S (b), or G146Q (c). The reaction conditions are described in Materials and Methods.

the binding of Ap_4A. The latter hypothesis is supported by the observation that the W160A mutant of Rv2613c did not have any detectable catalytic activity (Table 2). In addition, the multiple sequence alignment in Fig. 4 showed that Trp160 in Rv2613c corresponds to a Pro residue in typical Ap_4A phosphorylases and ADP-glucose phosphorylase, and to a Lys residue in HIT family Ap_nA hydrolases. Finally, in the putative Ap_4A binding site of Rv2613c-A, the side

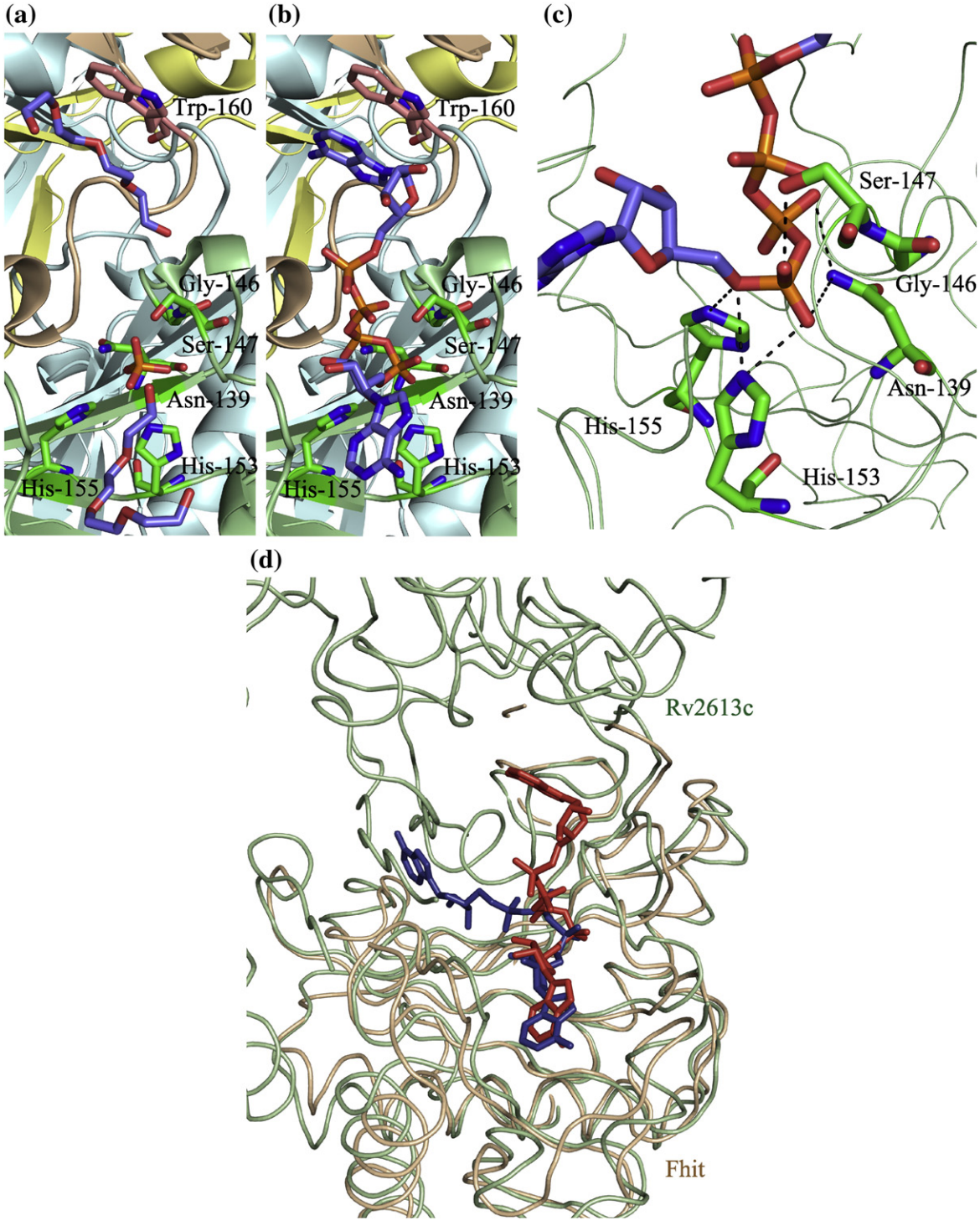

Fig. 6 (*legend on next page*)

chain of Asn139 would bind to both α-phosphoryl and β-phosphoryl oxygens of Ap_4A, and the side chains of Ser147, His153, and His155 would bind to the α-phosphoryl oxygen of Ap_4A (Fig. 6c).

A comparison of the putative Ap_4A binding site in Rv2613c with the P^1,P^2-methylene-P^3-thiodiadenosine triphosphate (IB2) binding site in Fhit[12] revealed that the putative binding site of the adenosine of Ap_4A in Rv2613c-A was the same as the adenosine binding site of IB2 in Fhit. However, the putative binding site of the other adenosine of Ap_4A in Rv2613c-B′ was different from the binding site of the other adenosine of IB2 in Fhit (Fig. 6d).

Discussion

Recently, we reported that Rv2613c is a tetrameric Ap_4A phosphorylase, but its amino acid sequence is more similar to HIT family Ap_nA hydrolases than to typical Ap_4A phosphorylases.[4] Now, we have solved the 1. 89-Å-resolution crystal structure of Rv2613c (Fig. 1a and Table 1), which is the first known structure of a protein with Ap_nA phosphorylase activity, and characterized the structural basis of its catalytic activity.

We found that the overall and active-site structures of Rv2613c were similar to those of other proteins in the HIT superfamily (Figs. 2 and 3). Specifically, the structure of Rv2613c-A/-B was in agreement with those of the Fhit dimer, MSMEG5028 dimer, and Ath monomer (Fig. 2a and b). Furthermore, a 10-stranded antiparallel β-sheet and long inner α-helices were common secondary structures in these proteins (Fig. 2a and b). Consistent with our previous results,[4] we also found that Rv2613c is tetrameric (Fig. 1c) and that its N-terminal domain is involved in intersubunit contacts (Fig. 1b and c). In contrast, Fhit and MSMEG5028 do not have a structural equivalent of the N-terminal domain of Rv2613c; therefore, they are homodimeric rather than tetrameric. In contrast, since Ath has a structural equivalent of the N-terminal domain of Rv2613c (Fig. 2b), it forms a homodimer, which corresponds to the Rv2613c tetramer. Therefore, a consensus structure of the N-terminal domain of Rv2613c might be required for the multimerization of HIT superfamily proteins.

In Rv2613c, one phosphate ion was bound to the same binding site in each subunit (Figs. 1c and 3a), which is thought to be the active site of Rv2613c. The interaction of Asn139 and Ser147 with the phosphate ion and the close proximity of Gly146 to the phosphate ion (Fig. 3a) suggested that these residues are catalytically important. We evaluated the catalytic importance of these residues by measuring the enzymatic activities of mutant Rv2613c proteins with point mutations in these residues. The lack of catalytic activity in the N139A, N139Q, and N139Q/S147T mutants (Table 2) showed that Asn139 is catalytically important. In addition, the similar K_m values and decreased K_{cat} values of the G146S, G146Q, S147A, and S147T mutants compared with those of wild-type Rv2613c (Table 2) suggested that Gly146 and Ser147 are probably important for the catalytic activity of Rv2613c rather than for Ap_4A binding. In the putative Ap_4A binding site of Rv2613c, the interactions between the side chain of Asn139 and the α-phosphoryl and β-phosphoryl oxygens of Ap_4A, as well as those between the side chain of Ser147 and the α-phosphoryl oxygen of Ap_4A (Fig. 6c), suggested that Asn139 and Ser147 are important to properly orient the α-phosphates and β-phosphates of Ap_4A for catalysis. This hypothesis is in agreement with two previous studies of HIT superfamily proteins, namely AMP-lysine hydrolase[13] and galactose-1-phosphate uridylyltransferase,[14] which showed the importance of a Ser residue in the active site. Furthermore, the catalytic importance of Asn139 and Ser147 in Rv2613c is consistent with the conservation of Asn and Ser residues at these positions in typical Ap_4A phosphorylases and ADP-glucose phosphorylase, but not in HIT family Ap_nA hydrolases, which have Gln and Thr residues in their active site (Figs. 3b and 4).

Fig. 6. Putative Ap_4A binding site in Rv2613c and comparison with the IB2 binding site in Fhit. (a) The binding sites for phosphate and tetraethylene glycol at the A–B′ contact of Rv2613c. Rv2613c subunits are colored as in Fig. 1c. Nitrogen, oxygen, and phosphorus atoms are shown in dark blue, red, and orange, respectively. The carbon atoms of tetraethylene glycol are shown in light blue. The carbon atoms of Asn139, Gly146, Ser147, His153, and His155 in Rv2613c-A are shown in green, and those of Trp160 in Rv2613c-B′ are shown in pink. (b) Putative Ap_4A binding site of Rv2613c. To create this model, we superimposed the α-phosphate of Ap_4A onto the phosphate ion in Rv2613c-A as shown in (a), and then we optimized the structure of Ap_4A by using the CNSsolve program. The Rv2613c subunits are colored as in Fig. 1c. The carbon atoms of adenosine are shown in light blue. Nitrogen, oxygen, and phosphorus atoms are colored as in (a). The carbon atoms of Asn139, Gly146, Ser147, His153, and His155 in Rv2613c-A, and those of Trp160 in Rv2613c-B′, are also colored as in (a). (c) Close-up view of the putative active sites of Rv2613c. Asn139, Gly146, Ser147, His153, His155, and Ap_4A colored as in (b). (d) Comparison of the putative binding mode of Ap_4A in Rv2613c and the binding mode of IB2 in Fhit (PDB ID: 1FHI). The backbones of the Rv2613c tetramer and the Fhit dimer are shown in green and light brown, respectively. Ap_4A and IB2 are shown in red and blue, respectively.

In addition to the catalytic importance of Gly146, the proximity of this residue to the α-phosphates and β-phosphates of Ap_4A in the putative Ap_4A binding site of Rv2613c (Fig. 6c) showed that this residue might influence the phosphorylase activity of Rv2613c. Moreover, G146Q produced about 10-fold more AMP than wild-type Rv2613c or G146S (Fig. 5). In addition, Gly146 in Rv2613c corresponds to Ser and Glu residues in typical Ap_4A phosphorylases and HIT family Ap_nA hydrolases, respectively (Fig. 4).

The structure of the predicted Ap_4A binding site of Rv2613c (Fig. 6b) and the lack of catalytic activity of the W160A mutant (Table 2) suggested that the tetramerization of Rv2613c (i.e., A–B′ and B–A′ contacts) and Trp160 may be essential for the formation of the Ap_4A binding site. Furthermore, the ability of Rv2613c to tetramerize, in contrast to the inability of Fhit to multimerize, might account for the difference between the putative Ap_4A binding site of Rv2613c and the IB2 binding site of Fhit (Fig. 6d).

The tetrameric structure of Rv2613c is unique among enzymes that use Ap_nA as a substrate. In addition, Trp160 in Rv2613c is not conserved in other Ap_4A phosphorylases and Ap_nA hydrolases in the HIT superfamily (Fig. 4). These differences suggest that Rv2613c may have a unique Ap_4A binding site, which may facilitate the design of specific inhibitors against Rv2613c. Ap_4A is found in both prokaryotes and eukaryotes,[20,21] and is involved in virulence and biofilm formation.[22,23] Furthermore, *Rv2613c* is an essential gene in *M. tuberculosis* H37Rv.[5] Therefore, the specific inhibitors against Rv2613c be potentially useful anti-TB drugs. As a result, we are currently screening for novel inhibitors of Rv2613c.

Materials and Methods

Selenomethionine substitution and site-directed mutagenesis of Rv2613c

The *Rv2613c* gene from *M. tuberculosis* H37Rv was cloned into the pCold I expression vector (Takara Bio, Inc., Shiga, Japan) to construct the pMS2613c plasmid, as described previously.[4] To express selenomethionine (SeMet)-substituted Rv2613c (SeMet-Rv2613c), we transformed pMS2613c into *Escherichia coli* B834(DE3) (Merck Novagen, Darmstadt, Germany). In addition, site-directed mutagenesis was used to introduce point mutations (N139A, N139Q, G146S, G146Q, S147A, S147T, N139Q/S147T, and W160A) into Rv2613 by using the QuickChange Site-Directed Mutagenesis Kit II (Agilent Technologies, Inc., Santa Clara, CA) with pMS2613c as template. The resulting plasmids were sequenced with an Applied Biosystems 3130*xl* Genetic Analyzer (Applied Biosystems, Carlsbad, CA) to confirm the mutation and were transformed into *E. coli* BL21(DE3)pLysS (Merck Novagen).

Expression and purification of Rv2613c

A single transformed colony of *E. coli* B834(DE3) was inoculated in 5 mL of Luria–Bertani medium and incubated overnight at 37 °C. Subsequently, 2 mL of the overnight culture was transformed into 200 mL of M9 minimal medium supplemented with the following: 1% glucose; 2 mM $MgSO_4$; 0.1 mM $CaCl_2$; 50 mg L^{-1} ampicillin; Kao Vitamins (Sigma-Aldrich Japan, Tokyo, Japan); 5 μg mL^{-1} each of L-Trp and L-Tyr; and 50 μg mL^{-1} each of L-Ala, L-Arg, L-Asn, L-Asp, L-Cys, L-Glu, L-Gln, L-Gly, L-His, L-Ile, L-Leu, L-Lys, L-Phe, L-Pro, L-Ser, L-Thr, L-Val, and L-SeMet. Then, this culture was grown under aerobic conditions for 7 h at 37 °C until the absorbance at 600 nm (A_{600}) had reached 1.0. Afterwards, the culture was transferred into 2 L of the same supplemented medium and cultured under the same conditions until an A_{600} of 0.4 had been reached. At that time, we added 0.5 mM isopropyl β-D-thiogalactopyranoside (Wako Pure Chemical Industries Ltd., Osaka, Japan) (final concentration) to induce protein expression and continued incubating the culture for 24 h at 15 °C. The Rv2613c point mutants were expressed in the same manner.

To purify the SeMet-Rv2613c and mutant Rv2613c proteins, we centrifuged the resulting *E. coli* cells at 8000*g* for 10 min at 4 °C, and then resuspended them in buffer A [20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, and 40 mM imidazole]. Subsequently, the cells were disrupted by sonicating them on ice with a UP50H sonicator (Hielscher Ultrasonics, Teltow, Germany). Unbroken cells and cellular debris were removed by centrifugation at 20,000*g* for 30 min at 4 °C. Then, the clarified supernatant was loaded onto a HisTrap HP column (1.6 cm × 2.5 cm; GE Healthcare Bio-Sciences, Buckinghamshire, UK) that was equilibrated with buffer A, and the protein of interest was eluted with a linear gradient of imidazole (40–250 mM). Next, the protein was loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (1.6 cm × 60 cm; GE Healthcare Bio-Sciences) that was equilibrated with buffer B [20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, and 150 mM imidazole]. The sample was separated with buffer B into 1.2-mL fractions. For SeMet-Rv2613c, fractions 24–30 were pooled and then dialyzed against buffer C [20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Hepes)-Na (pH 7.6) and 0.5 mM dithiothreitol] using Slide-A-Lyzer Dialysis Cassettes (10,000 weight cutoff; Thermo Fisher Scientific, Waltham, MA) at room temperature. For the Rv2613c mutants, fraction 25 was dialyzed against buffer C using Slide-A-Lyzer Dialysis Cassettes (10,000 weight cutoff; Thermo Fisher Scientific) at room temperature. These dialyzed proteins were used for all experiments in this study.

Crystallization and data collection

We crystallized Rv2613c and collected the diffraction data as described previously.[18] Briefly, SeMet-Rv2613c was crystallized by hanging-drop vapor diffusion. Specifically, 5 μL of protein (10 mg mL^{-1}) was mixed with 5 μL of crystallization solution [0.1 M sodium cacodylate (pH 6.7), 0.2 M lithium sulfate, and 28% polyethylene glycol 400], and then the drop was suspended on a siliconized coverslip over 0.6 mL of the crystallization solution. Prismatic colorless crystals of SeMet-Rv2613c

formed after about 4 weeks at 20 °C and grew to a maximum size of 0.5 mm. The crystals were harvested using a CryoLoop (Hampton Research, Aliso Viejo, CA) and flash cooled to 100 K in a nitrogen gas stream.

The diffraction data for SeMet-Rv2613c were collected at the AR-NW12A station of the Photon Factory (Tsukuba, Japan) at a wavelength of 0.97901 Å with an ADSC Quantum 210r detector (Area Detector Systems Corporation, Poway, CA) located 240 mm from the crystal in 2-s 1° oscillations at 100 K. The diffraction data were processed, merged, and scaled using the HKL2000 (DENZO and SCALEPACK) software package.[24]

Structure determination and model analysis

The crystal structure of Rv2613c was solved by using SAD. SAD phasing and phase refinement were performed by using the PHENIX software package.[25] We built a model of Rv2613c with the Coot program,[26] and then refined it with CCP4,[27] CNSsolve,[28] and PHENIX to 1.89 Å resolution. We validated the model by using MolProbity.[29]

Ribbon diagrams were generated by PyMOL,[30] and protein structure topology cartoons were produced by TopDraw.[31] The atomic coordinates of crystal structures were downloaded from PDB†. We searched for similar protein structures by using the Dali server‡[32] and performed pairwise structure comparisons with the DaliLite program.[33] Sequence alignments were performed by using CLUSTAL W 1.83.[34]

Enzyme activity assays

The enzyme activity of Rv2613c and its mutants was measured by using a high-performance liquid chromatography system (Shimadzu, Kyoto, Japan) to quantify the amount of remaining substrate after the reaction, as described previously,[4] with some modifications. The reaction mixture (100 μL) consisted of 50 mM Hepes-Na (pH 7.6), 1 mM $MnCl_2$, 100 μg mL^{-1} bovine serum albumin, 1 mM Ap_4A, 5 mM phosphate, and purified enzymes. This mixture was incubated at 37 °C for 30 min. The reaction was stopped by heating at 95 °C for 4 min, followed by centrifugation of the reaction solution at 12,000*g* for 3 min. The clear supernatant was analyzed using a 4.6 mm × 125 mm Partisphere 5-μm SAX HPLC column (Whatman, Kent, UK), which was equilibrated with pure water. The adsorbed nucleotides, such as AMP, ADP, ATP, and Ap_4A, were eluted with a gradient of water and elution buffer [1.3 M $(NH_4)_2HPO_4$ with H_3PO_4 (pH 4.8)] as follows: 0–10 min, 0% elution buffer; 10–110 min, 0–50% elution buffer at a flow rate of 0.5 mL min^{-1}. The eluted nucleotides were detected at 254 nm. One unit of enzyme activity was defined as the degradation of 1.0 μmol of Ap_4A in 1 min at 37 °C. K_m and K_{cat} values were calculated by using GraphPad Prism (GraphPad Software, La Jolla, CA). Results were expressed as the mean (standard error of the mean) of three independent measurements.

† www.rcsb.org
‡ ekhidna.biocenter.helsinki.fi/dali_server

Accession codes

The structure factor and atomic coordinates for Rv2613c have been deposited in the PDB under accession code 3ANO.

Acknowledgements

We thank the beamline staff at the AR-NW12A station of the Photon Factory for their assistance with data collection. Synchrotron radiation experiments at this facility were performed as outlined in proposal nos. 2008G610 and 2010G520. This study was funded by grants from the Ministry of Health, Labor, and Welfare of Japan (grant H22-Shinkou-Ippan-007) and by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (grant 20780066).

References

1. Korenromp, E. L., Bierrenbach, A. L., Williams, B. G. & Dye, C. (2009). The measurement and estimation of tuberculosis mortality. *Int. J. Tuberc. Lung Dis.* **13**, 283–303.
2. Chiang, C. Y. & Yew, W. W. (2009). Multidrug-resistant and extensively drug-resistant tuberculosis. *Int. J. Tuberc. Lung Dis.* **13**, 304–311.
3. Mori, S., Yamasaki, M., Maruyama, Y., Momma, K., Kawai, S., Hashimoto, W. *et al.* (2005). NAD-binding mode and the significance of intersubunit contact revealed by the crystal structure of *Mycobacterium tuberculosis* NAD kinase–NAD complex. *Biochem. Biophys. Res. Commun.* **327**, 500–508.
4. Mori, S., Shibayama, K., Wachino, J. & Arakawa, Y. (2010). Purification and molecular characterization of a novel diadenosine 5′,5‴-P^1,P^4-tetraphosphate phosphorylase from *Mycobacterium tuberculosis* H37Rv. *Protein Expression Purif.* **69**, 99–105.
5. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**, 77–84.
6. Raman, K., Yeturu, K. & Chandra, N. (2008). targetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst. Biol.* **2**, 109.
7. Brenner, C. (2002). Hint, Fhit, and GalT: function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases. *Biochemistry*, **41**, 9003–9014.
8. Mulder, W., Scholten, I. H., van Roon, H. & Grivell, L. A. (1994). Isolation and characterization of the linked genes APA2 and QCR7, coding for Ap_4A phosphorylase II and the 14 kDa subunit VII of the mitochondrial bc_1-complex in the yeast *Kluyveromyces lactis*. *Biochim. Biophys. Acta*, **1219**, 719–723.

9. Plateau, P., Fromant, M., Schmitter, J. M. & Blanquet, S. (1990). Catabolism of bis(5′-nucleosidyl) tetraphosphates in *Saccharomyces cerevisiae*. *J. Bacteriol.* **172**, 6892–6899.
10. Huang, Y., Garrison, P. N. & Barnes, L. D. (1995). Cloning of the *Schizosaccharomyces pombe* gene encoding diadenosine 5′,5‴-P^1,P^4-tetraphosphate (Ap_4A) asymmetrical hydrolase: sequence similarity with the histidine triad (HIT) protein family. *Biochem. J.* **312**, 925–932.
11. Lima, C. D., Klein, M. G. & Hendrickson, W. A. (1997). Structure-based analysis of catalysis and substrate definition in the HIT protein family. *Science*, **278**, 286–290.
12. Pace, H. C., Garrison, P. N., Robinson, A. K., Barnes, L. D., Draganescu, A., Rosler, A. *et al.* (1998). Genetic, biochemical, and crystallographic characterization of Fhit–substrate complexes as the active signaling form of Fhit. *Proc. Natl Acad. Sci. USA*, **95**, 5484–5489.
13. Krakowiak, A., Pace, H. C., Blackburn, G. M., Adams, M., Mekhalfia, A., Kaczmarek, R. *et al.* (2004). Biochemical, crystallographic, and mutagenic characterization of Hint, the AMP-lysine hydrolase, with novel substrates and inhibitors. *J. Biol. Chem.* **279**, 18711–18716.
14. Geeganage, S., Ling, V. W. & Frey, P. A. (2000). Roles of two conserved amino acid residues in the active site of galactose-1-phosphate uridylyltransferase: an essential serine and a nonessential cysteine. *Biochemistry*, **39**, 5397–5404.
15. Barnes, L. D., Garrison, P. N., Siprashvili, Z., Guranowski, A., Robinson, A. K., Ingram, S. W. *et al.* (1996). Fhit, a putative tumor suppressor in humans, is a dinucleoside 5′,5‴-P^1,P^3-triphosphate hydrolase. *Biochemistry*, **35**, 11529–11535.
16. Guranowski, A. & Blanquet, S. (1985). 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**, 3542–3547.
17. McLennan, A. G., Mayers, E., Hankin, S., Thorne, N. M., Prescott, M. & Powls, R. (1994). The green alga *Scenedesmus obliquus* contains both diadenosine 5′,5‴-P^1,P^4-tetraphosphate (asymmetrical) pyrophosphohydrolase and phosphorylase activities. *Biochem. J.* **300**, 183–189.
18. Mori, S., Shibayama, K., Wachino, J. & Arakawa, Y. (2010). Crystallization and preliminary X-ray analysis of the diadenosine 5′,5‴-P^1,P^4-tetraphosphate phosphorylase from *Mycobacterium tuberculosis* H37Rv. *Acta Crystallogr. Sect. F*, **66**, 279–281.
19. Mccoy, J. G., Arabshahi, A., Bitto, E., Bingman, C. A., Ruzicka, F. J., Frey, P. A. & Phillips, G. N. (2006). Structure and mechanism of an ADP-glucose phosphorylase from *Arabidopsis thaliana*. *Biochemistry*, **45**, 3154–3162.
20. McLennan, A. G., Barnes, L. D., Blackburn, G. M., Brenner, C., Guranowski, A., Miller, A. D. *et al.* (2001). Recent progress in the study of the intracellular functions of diadenosine polyphosphates. *Drug Dev. Res.* **52**, 249–259.
21. Hoyle, C. H. V., Hilderman, R. H., Pintor, J. J., Schluter, H. & King, B. F. (2001). Diadenosine polyphosphates as extracellular signal molecules. *Drug Dev. Res.* **52**, 260–273.
22. Ismail, T. M., Hart, C. A. & McLennan, A. G. (2003). 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**, 32602–32607.
23. Monds, R. D., Newell, P. D., Wagner, J. C., Schwartzman, J. A., Lu, W., Rabinowitz, J. D. & O'Toole, G. A. (2010). Diadenosine tetraphosphate (Ap_4A) metabolism impacts biofilm formation by *Pseudomonas fluorescens* via modulation of c-di-GMP-dependent pathways. *J. Bacteriol.* **192**, 3011–3023.
24. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
25. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D*, **66**, 213–221.
26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. Sect. D*, **66**, 486–501.
27. Collaborative Computational Project, Number 4. (1994). The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
28. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.
29. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J. *et al.* (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D*, **66**, 12–21.
30. The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.
31. Bond, C. S. (2003). TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*, **19**, 311–312.
32. Holm, L. & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549.
33. Holm, L. & Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics*, **16**, 566–567.
34. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). 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**, 4673–4680.