Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2017 October 17.

Published in final edited form as:

Biochemistry. 2017 July 25; 56(29): 3818–3825. doi:10.1021/acs.biochem.7b00506.

Biochemical Investigation of Rv3404c from *Mycobacterium tuberculosis*‡

Murray M. Dunsirn[§], **James B. Thoden**[§], **Michel Gilbert**[¶], and **Hazel M. Holden**[§,*]

[§]Department of Biochemistry, University of Wisconsin, Madison, WI 53706

[¶]National Research Council Canada, Human Health Therapeutics, Ottawa, Ontario, K1A 0R6, Canada

Abstract

The causative agent of tuberculosis, *Mycobacterium tuberculosis*, is a bacterium with a complex cell wall and a complicated life cycle. The genome of *M. tuberculosis* contains well over 4000 genes thought to encode proteins. One of these codes for a putative enzyme referred to as Rv3404c, which has attracted research attention as a potential virulence factor for over 12 years. Here we demonstrate that Rv3404c functions as a sugar *N*-formyltransferase that converts dTDP-4-amino-4,6-dideoxyglucose into dTDP-4-formamido-4,6-dideoxyglucose using N^{10}-formyltetrahydrofolate as the carbon source. Kinetic analyses demonstrate that Rv3404c displays a significant catalytic efficiency of 1.1×10^4 $M^{-1}s^{-1}$. In addition, we report the X-ray structure of a ternary complex of Rv3404c solved in the presence of N^5-formyltetrahydrofolate and dTDP-4-amino-4,6-dideoxyglucose. The final model of Rv3404c was refined to an overall *R*-factor of 16.8% at 1.6 Å resolution. The results described herein are especially intriguing given that there have been no published reports of *N*-formylated sugars associated with *M. tuberculosis*. The data thus provide a new avenue of research into this fascinating, yet deadly organism that apparently has been associated with human infection since ancient times.

TOC image

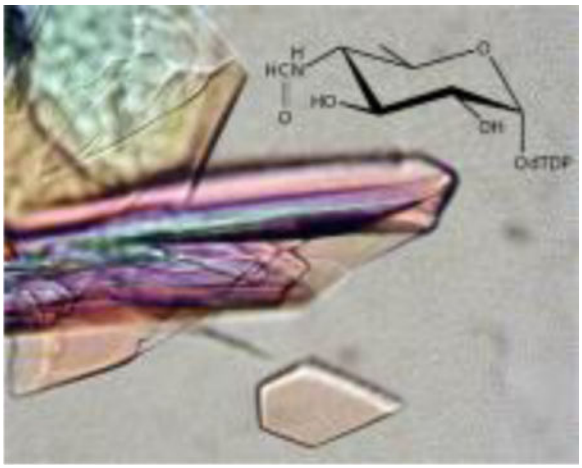
‡This research was supported in part by NIH grant GM115921 (to H. M. H.).

*To whom correspondence should be addressed. Hazel_Holden@biochem.wisc.edu, FAX: 608-262-1319, PHONE: 608-262-4988.

X-ray coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, N. J. (accession number 5VYQ).

The authors have no competing financial interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Introduction

According to the World Health Organization, tuberculosis is one of the top ten causes of death worldwide, and in 2015 alone, it accounted for more fatalities than HIV and malaria combined. It is a highly contagious and airborne disease caused by *Mycobacterium tuberculosis*, a rod-shaped organism demonstrating Ziehl-Neelsen staining properties due to the presence of mycolic acids on its cell surface.[1] Interestingly, it is thought that *M. tuberculosis* and humans have cohabited since the time humans emerged as a species.[2] Treatment for tuberculosis typically requires extensive use of isoniazid, rifampicin, ethambutol, and pyrazinamide for two months followed by an additional four months of isoniazid and rifampicin.[3] Of particular concern is the appearance of multi-drug resistant and extensively drug-resistant bacterial strains, referred to as MDR-TB and XDR-TB, respectively.[4] Treatment is also complicated by the fact those infected with *M. tuberculosis* may be asymptomatic. Indeed, only 5 to 15% of those with latent infections will eventually develop active tuberculosis in their lifetimes.[5] As a consequence, these individuals function as “repositories” for the organism, which is especially concerning given that approximately one third of the world’s population has a latent *M. tuberculosis* infection.

Distinct from both Gram negative and Gram positive bacteria, *M. tuberculosis* has a unique core cell wall structure composed of an inner peptidoglycan layer, a middle highly branched arabinogalactan zone, and an outer covering composed of mycolic acids, which are oriented perpendicular to the plane of the membrane.[6,7] Due to the role of the cell wall in both bacterial survival and pathogenesis, those enzymes involved in its biosynthesis have been and continue to be targets for drug design.

In 2014, X-ray coordinates for an uncharacterized protein, Rv3404c from *M. tuberculosis* H37Rv, were deposited into the Protein Data Bank (accession codes 4PZU and 4Q12). Whereas this protein was suggested to be a sugar *N*-formyltransferase, there were no biochemical data presented to validate this assertion. The annotation was intriguing given that there have been no reports in the literature regarding the existence of *N*-formylated sugars in *M. tuberculosis*. Thus far, research has shown that these types of unusual formylated sugars are only found on the lipopolysaccharides of pathogenic Gram negative

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

bacteria such as *Campylobacter jejuni, Francisella tularensis*, *Providencia alcalifaciens, Salmonella enterica*, and *Brucella melitensis*.[8–15] Whether these unusual sugars are important for virulence in these organisms is presently unknown. Notably, however, the loss of activity of a sugar *N*-formyltransferase in *Brucella abortus* results in a bacterial strain with attenuated pathogenicity.[13]

Rv3404c has attracted attention as a potential virulence factor for over 12 years because it is either differentially expressed under various conditions and/or because of its degree of conservation amongst strains from the *M. tuberculosis* complex, a genetically related group of *Mycobacterium* species that can cause human and/or animal tuberculosis.[16,17] Importantly, inactivation of Rv3404c gives rise to increased survival times for severe combined immune deficiency (SCID) mice infected with *M. tuberculosis*.[18,19]

Here we describe a biochemical and X-ray structural analysis of Rv3404c. Our data demonstrate that it catalyzes the conversion of dTDP-4-amino-4,6-dideoxyglucose (dTDP-Qui4N) to dTDP-4-formamido-4,6-dideoxyglucose (dTDP-Qui4NFo) as indicated in Scheme 1. We also demonstrate that it utilizes N^{10}-formyltetrahydrofolate (N^{10}-formyl-THF) as the carbon source (Scheme 1). The research presented herein thus provides a molecular framework for further cellular and molecular biological investigations into the physiological role of this enzyme in *M. tuberculosis*.

Materials and Methods

Protein Expression and Purification

The gene encoding Rv3404c from *M. tuberculosis* was synthesized by Integrated DNA Technologies using codons optimized for protein expression in *Escherichia coli*. The gene was placed into a modified pET28b vector (Novagen), which leads to a protein with an N-terminal polyhistidine tag as previously described.[20] The pET28t-*rv3404c* plasmid was utilized to transform Rosetta2(DE3) *E. coli* cells (Novagen). The cultures were grown in lysogeny broth supplemented with kanamycin and chloramphenicol (both at 50 mg/L concentration) at 37 °C with shaking until an optical density of 0.8 was reached at 600 nm. The flasks were cooled in an ice bath, and the cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside and allowed to express protein at 16 °C for 24 h.

The cells were harvested by centrifugation and frozen as pellets in liquid nitrogen. The pellets were subsequently disrupted by sonication on ice in a lysis buffer composed of 50 mM sodium phosphate, 20 mM imidazole, 10% glycerol, and 300 mM sodium chloride (pH 8.0). The lysate was cleared by centrifugation, and Rv3404c was purified at 4 °C utilizing nickel nitrilotriacetic acid resin (Qiagen) according to the manufacturer's instructions. All buffers were adjusted to pH 8.0 and contained 50 mM sodium phosphate, 300 mM sodium chloride, and imidazole concentrations of 20 mM for the wash buffer and 300 mM for the elution buffer. The protein was collected and dialyzed against 10 mM Tris-HCl (pH 8.0) and 200 mM NaCl and concentrated to ~24 mg/mL based on a calculated extinction coefficient of 1.09 $(mg/mL)^{-1}cm^{-1}$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Determination of Substrate Preference

Test reactions using Rv3404c (1 mg/mL), 2 mM N^{10}-formyl-THF, 50 mM HEPPS (pH 8.0), and 1 mM of various nucleotide-linked sugars were allowed to incubate for 12 h at 37 °C. For each reaction, the enzyme was removed via filtration through a 10 kDa cutoff filter, and the mixture diluted 10× with water and examined by HLPC. Specifically, a 0.0 – 1.0 M gradient of ammonium acetate (pH 4.0) was utilized with a 1 mL Resource-Q column. Of the various nucleotide-linked sugars tested, as described in Results, the only reaction demonstrating loss of the starting sugar was that when dTDP-QuiN4 served as the substrate. The product of the Rv3404c/dTDP-QuiN4 reaction was evaluated by electrospray ionization mass spectrometry in the negative ion mode.

Kinetic Analyses using dTDP-Qui4N as the Substrate

Kinetic parameters for Rv3404c were determined via a discontinuous assay using an ÄKTA Purifier HPLC system as described previously.[8] The 1.5 mL reaction mixtures contained 5 mM N^{10}-formyl-THF, 50 mM HEPPS (pH 8.0), 0.28 μM Rv3404c, and dTDP-Qui4N concentrations ranging from 0.01 mM – 2.0 mM. A plot of initial velocity versus concentration was analyzed using PRISM (GraphPad Software, Inc.) and fitted to the equation $v_o=(V_{max}[S])/(K_M+[S])$

Crystallization

Large crystals of Rv3404c were grown in hanging drop experiments from 20% poly(ethylene glycol) 3350, 200 mM KCl, and 100 mM HEPPS (pH 8.0) and in the presence of dTDP-Qui4N and N^5-formyl-THF. They belonged to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 47.5 Å, b = 73.0 Å, and c = 173.3 Å. The asymmetric unit contained one dimer. For X-ray data collection, the crystals were transferred to a cryo-protectant solution composed of 26% poly(ethylene glycol) 3350, 200 mM NaCl, 200 mM KCl, 200 mM LiCl, 5 mM dTDP-Qui4N, 5 mM N^5-formyl-THF, and 13% ethylene glycol.

X-ray Data Collection, Processing and Structural Analysis

An X-ray data set was collected in house and was processed with SAINT and scaled with SADABS (Bruker AXS Inc.). Relevant X-ray data collection statistics are listed in Table 1. The structure was solved via molecular replacement with PHASER[21] and using PDB entry 4PZU as a search probe. Iterative cycles of model-building with COOT[22,23] and refinement with REFMAC[24] were employed to produce the final X-ray model. Refinement statistics are provided in Table 2.

Results and Discussion

Substrate Preference

Thus far, five nucleotide-linked amino sugars have been identified as substrates for the sugar *N*-formyltransferases: dTDP-3-amino-3,6-dideoxyglucose (dTDP-Qui3N), dTDP-3-amino-3,6-dideoxygalactose, dTDP-4-amino-4,6-dideoxyglucose (dTDP-Qui4N), UDP-4-amino-4-deoxyarabinose, and GDP-4-amino-4,6-dideoxymannose. Whereas these

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

nucleotide-linked sugars are not commercially available, they have been enzymatically synthesized in our laboratory.[8–12,14,25] From this list, the only nucleotide-linked sugar that served as a substrate for Rv3404c was dTDP-Qui4N. Specifically, the mass spectrum for dTDP-Qui4N shows a peak having a m/z = 545.9. Upon incubation of dTDP-Qui4N and N^{10}-formyl-THF with Rv3404c, a peak was observed in the mass spectrum having a m/z = 574.0. This is the mass that would be expected for the formylated product, dTDP-Qui4NFo.

Kinetic Analysis

To observe the formation of dTDP-Qui4NFo, the reaction was allowed to incubate for 12 h at 37 °C as described in Materials and Methods. In order to determine whether the enzymatic rate of Rv3404c is physiologically relevant, a kinetic analysis was subsequently conducted by measuring initial velocities versus dTDP-Qui4N concentrations in 2 min assays. Initial test assays demonstrated that the reactions were linear for at least 5 min. As can be seen in Figure 1, Rv3404c displays normal Michaelis-Menten kinetics. From these data, the K_M was determined to be 0.08 ± 0.01 mM, and the k_{cat} was calculated to be 0.9 ± 0.1 s^{-1}. The overall catalytic efficiency or k_{cat}/K_M for Rv3404c is 1.1 × 10^4 ± 800 $M^{-1}s^{-1}$. The magnitude of the catalytic efficiency is similar to that observed for other sugar *N*-formyltransferases and is highly suggestive that dTDP-Qui4N is the natural substrate for Rv3404c.

Structure of the Ternary Complex

The structure of Rv3404c in complex with N^5-formyltetrahydrofolate (N^5-formyl-THF) and dTDP-Qui4N was determined to 1.6 Å resolution and refined to an overall *R*-factor of 16.8%. For crystallization purposes, N^5-formyl-THF was utilized rather than N^{10}-formyl-THF because it is a stable analogue. It is not catalytically competent, however.

Shown in Figure 2a is a ribbon representation of the Rv3404c dimer, which displays overall dimensions of ~57 Å × 59 Å × 104 Å. The dimer is rather elongated such that the N-termini of the subunits are separated by ~100 Å. Each monomer folds into two distinct regions. The larger N-terminal domain is defined by Met 1 to Arg 183, whereas the C-terminal lobe is formed by Arg 184 to Pro 232. The N-terminal domain is characterized by a six-stranded mixed β-sheet flanked on either side by two α-helices. The C-terminal region initiates with a short β-strand (Arg 190 to Thr 192) and an α-helix (Phe 193 to Ala 202). This short β-strand interacts with the symmetry-related β-strand in the second subunit thereby contributing to the subunitsubunit interface. There are two additional β-strands (Ala 211 to Val 214 and Lys 220 to Pro 229) that are connected by a Type I turn (Asp 215 to Gly 218). The second β-strand of this hairpin motif is also involved in formation of the subunitsubunit interface. The active sites of the dimer are separated by ~50 Å.

The electron densities in Subunit 1 corresponding to dTDP-Qui4N and N^5-formyl-THF are displayed in Figure 2b. Only electron density for dTDP was observed in Subunit 2. The electron density for the pteridine ring of the cofactor was well-ordered but that for the remainder of the ligand disordered. The electron density observed for the dTDP-Qui4N ligand, however, was unambiguous. The ribose of the dTDP-Qui4N ligand adopts the C3′-*endo* pucker whereas the pyranosyl group of the substrate shows the $^4C_1$ chair conformation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A close-up stereo view of the Rv3404c active site is presented in Figure 2c. The thymine ring of the dTDP-linked sugar substrate is wedged between two aromatic side chains, Phe 93 from the N-terminal domain and Phe 208 from the C-terminal region. In addition, the thymine ring forms hydrogen-bonding interactions with the side chain of Asn 210 from the C-terminal domain. Key electrostatic interactions occur between the pyrophosphoryl group of the dTDP-Qui4N substrate and the side chains of Asn 10, His 63, Lys 65, and Tyr 139. There are five waters that lie within 3.2 Å of the pyrophosphoryl moiety. The carbonyl oxygen of Gly 91 and the backbone amide nitrogen of Phe 93 participate in hydrogen bonding interactions with the C-3′ and C-2′ hydroxyl groups of the hexose, respectively. There are three residues strictly conserved amongst the *N*-formyltransferases that play key roles in catalysis.[15] In Rv3404c these correspond to Asn 80, His 82, and Asp 117. The conserved histidine is thought to act as the catalytic base that abstracts a proton from the amino group of the sugar. This is required for the eventual transfer of the formyl group from the cofactor to the amino group of the dTDP-sugar substrate. A peak of strong electron density was observed lying within 2.0 Å of the C-4′ amino nitrogen of the hexose. On the basis of the observed electron density and in consideration of both the cryo-protectant conditions used for X-ray data collection and the coordination geometry surrounding the peak, it was modeled as a lithium ion (Figure 2c). The close distance between this cation and the C-4′ amino nitrogen of the hexose argues that the amino group is unprotonated when bound in the active site.

The initial model of the apo form of Rv3404c was deposited into the Protein Data Bank under the accession number 4PZU. The α-carbons for this apo form and the ternary complex model presented here superimpose with a root-mean-square deviation of 0.5 Å for 216 target atoms. Overall, the models are decidedly similar except for the position of the loop defined by His 63 to Pro 69, which connects the third β-strand to the third α-helix of the subunit. This difference, highlighted in Figure 3, results from the binding of N^5-formyl-THF, which causes the loop to splay outwards from the active site. The largest movement occurs at Lys 65 where the positions of the α-carbons in the two models differ by 4.0 Å.

The structure of Rv3404c described here is similar to that previously reported for VioF from *P. alcalifaciens* O30.[10] These two enzymes display amino acid sequence identities and similarities of 36% and 59%, respectively. In keeping with the apparent amino acid sequence homology between the two proteins, the α-carbons for the two enzymes correspond with a root-mean-square deviation of 0.9 Å. Shown in Figure 4 is a superposition of the active sites for these two *N*-formyltransferases. The only significant differences are the positions of the pyranosyl groups in the active site pockets and the conformations of the Lys 65/Lys 77 side chains. As can be seen, in Rv3404c, the amino group of the pyranosyl moiety is appropriately positioned to interact with His 82, which is the presumed catalytic base on the basis of that observed for other sugar *N*-formyltransferases.[15] In VioF, the amino group is located at ~7 Å from His 94, suggesting that the dTDP-sugar bound in a nonproductive conformation due to the crystallization conditions. This same nonproductive conformation was also observed in the structural investigation of WbtJ from *F. tularensis*.[9] As such, the structure of Rv3404c presented here represents a more meaningful description of the Michaelis complex for those *N*-formyltransferases that act upon dTDP-Qui4N substrates.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The three-dimensional structures of *N*-formyltransferases that function on dTDP-Qui3N rather than dTDP-Qui4N have been reported within recent years.[8,11,12] One of these enzymes is WlaRD from *C. jejuni*.[8] A superposition of the active sites for Rv3404c and WlaRD with bound dTDP-sugars is presented in Figure 5. Strikingly, there is little amino acid sequence identity between Rv3404c and WlaRD with the exception of the conserved asparagine/histidine/aspartate triad that is observed in all *N*-formyltransferases studied to date. The most significant changes in the identities of the amino acids lining the active sites of Rv3404c and WlaRD include: Asn 10/Lys 9, Lys 65/Asp 79, His 63/Glu 77, Pro 83/Phe 97, and Trp 92/Val 106. As a result, the locations of the α- and β-phosphoryl groups of the dTDP-sugars in the active sites are significantly different. Specifically, the α- and β-phosphoryl moieties of dTDP-Qui4N and dTDP-Qui3N are shifted by 1.6 Å and 3.7 Å, respectively. Remarkably, as can be seen in Figure 5, the amino groups attached to either C-4′ or C-3′ are located in nearly identical positions in the active sites. Indeed, in Rv3404c, the distance between the C-4′ amino group and His 82 is 3.5 Å whereas in WlaRD the distance between the C-3′ amino group and His 94 is 3.9 Å. It is clear from this investigation, as was first suggested in the structural analysis of VioF from *P. alcalifaciens* O30, that substrate specificity amongst the *N*-formyltransferases is determined by interactions between the proteins and the pyrophosphoryl moieties of the substrates. These differing interactions allow the *N*-formyltransferases to direct the sugar amino groups into the correct positions for catalysis.

Unusual dideoxysugars such as Qui4N and Qui4NFo have been observed on the O-antigens of various Gram negative bacteria.[26] Strikingly, however, there has only been one report, published in 1998, that describes the observation of a quinovose-based sugar in a *Mycobacterium* species. Specifically, the polysaccharidic backbone of the 6-*O*-methylglucose lipopolysaccharide in *Mycobacterium bovis* BCG strain Pasteur has been shown to contain Qui2N.[27] From the investigation reported here, it is now clear that Rv3404c functions as a sugar *N*-formyltransferase using dTDP-Qui4N as its substrate. That the *M. tuberculosis* genome contains a gene encoding this enzyme is, indeed, puzzling given that there are no reports in the literature regarding the observation of *N*-formylated sugars in this organism. The fact that the gene encoding Rv3404c is highly conserved amongst strains from the *M. tuberculosis* complex suggests that the protein plays a key role in the virulence of the organism. The hydrophobicity of the *M. tuberculosis* cell wall has major impacts on its virulence, pathogenicity, aerosol transmission, and intrinsic resistance to many antibiotics.[28,29] Thus the formylation of Qui4N might be important for the effective transport or function of this residue in the context of the *M. tuberculosis* hydrophobic cell wall.

The locations of *N*-formylated sugars in *M. tuberculosis* and the life cycle stage at which they are produced remain questions that can only be addressed by further physiological studies. The data presented herein will undoubtedly spark an entirely new avenue of tuberculosis research. If, indeed, these sugars are important for cell wall maintenance or intracellular persistence, for example, then the three-dimensional model of Rv3404c described here will provide an important scaffold for structure-based drug design. Moreover, the absence of these sugars in humans further supports the targeting of their biosynthetic pathways in the continued search for new anti-tuberculosis therapeutics.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Abbreviations

dTDP	thymidine diphosphate
HEPPS	*N*-2-hydroxyethylpiperazine-*N*′-3-propanesulfonic acid
HPLC	high-performance liquid chromatography
N^5-formyl-THF	N^5-formyltetrahydrofolate
N^{10}-formyl-THF	N^{10}-formyltetrahydrofolate
Qui2N	2-amino-2,6-dideoxyglucose
Qui3N	3-amino-3,6-dideoxyglucose
Qui3NFo	3-formamido-3,6-dideoxyglucose
Qui4N	4-amino-4,6-dideoxyglucose
Qui4NFo	4-formamido-4,6-dideoxyglucose
Tris	*tris*- (hydroxymethyl)aminomethane

References

1. Madigan, MT., Martinko, JM. Biology of Microorganisms. 11. Pearson; Prentice Hall, Upper Saddle River, New Jersey: 2006.
2. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 2005; 1:e5. [PubMed: 16201017]
3. Cole ST. Inhibiting *Mycobacterium tuberculosis* within and without. Philos Trans R Soc Lond B Biol Sci. 2016; 371doi: 10.1098/rstb.2015.0506
4. Brouqui P, Quenard F, Drancourt M. Old antibiotics for emerging multidrug-resistant/extensively drug-resistant tuberculosis (MDR/XDR-TB). Int J Antimicrob Agents. 2017; 49:554–557. [PubMed: 28336312]
5. Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent *Mycobacterium tuberculosis* infection. N Engl J Med. 2015; 372:2127–2135. [PubMed: 26017823]
6. Brennan PJ. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis (Edinb). 2003; 83:91–97. [PubMed: 12758196]
7. Abrahams KA, Besra GS. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. Parasitology. 2016:1–18.
8. Thoden JB, Goneau MF, Gilbert M, Holden HM. Structure of a sugar *N*-formyltransferase from *Campylobacter jejuni*. Biochemistry. 2013; 52:6114–6126. [PubMed: 23898784]
9. Zimmer AL, Thoden JB, Holden HM. Three-dimensional structure of a sugar *N*-formyltransferase from *Francisella tularensis*. Protein Sci. 2014; 23:273–283. [PubMed: 24347283]
10. Genthe NA, Thoden JB, Benning MM, Holden HM. Molecular structure of an *N*-formyltransferase from *Providencia alcalifaciens* O30. Protein Sci. 2015; 24:976–986. [PubMed: 25752909]
11. Woodford CR, Thoden JB, Holden HM. New role for the ankyrin repeat revealed by a study of the *N*-formyltransferase from *Providencia alcalifaciens*. Biochemistry. 2015; 54:631–638. [PubMed: 25574689]
12. Woodford CR, Thoden JB, Holden HM. Molecular architecture of an *N*-formyltransferase from *Salmonella enterica* O60. J Struct Biol. 2017; doi: 10.1016/j.jsb.2017.03.002

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

13. Lacerda TL, Cardoso PG, Augusto de Almeida L, Camargo IL, Afonso DA, Trant CC, Macedo GC, Campos E, Cravero SL, Salcedo SP, Gorvel JP, Oliveira SC. Inactivation of formyltransferase (wbkC) gene generates a *Brucella abortus* rough strain that is attenuated in macrophages and in mice. Vaccine. 2010; 28:5627–5634. [PubMed: 20580469]
14. Riegert AS, Chantigian DP, Thoden JB, Tipton PA, Holden HM. Biochemical characterization of WbkC, an *N*-formyltransferase from *Brucella melitensis*. Biochemistry. 2017; doi: 10.1021/acs/biochem.7b00494
15. Holden HM, Thoden JB, Gilbert M. Enzymes required for the biosynthesis of *N*-formylated sugars. Curr Opin Struct Biol. 2016; 41:1–9. [PubMed: 27209114]
16. Betts JC, McLaren A, Lennon MG, Kelly FM, Lukey PT, Blakemore SJ, Duncan K. Signature gene expression profiles discriminate between isoniazid-, thiolactomycin-, and triclosan-treated *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2003; 47:2903–2913. [PubMed: 12936993]
17. Dubnau E, Chan J, Mohan VP, Smith I. Responses of *Mycobacterium tuberculosis* to growth in the mouse lung. Infect Immun. 2005; 73:3754–3757. [PubMed: 15908407]
18. McAdam RA, Quan S, Smith DA, Bardarov S, Betts JC, Cook FC, Hooker EU, Lewis AP, Woollard P, Everett MJ, Lukey PT, Bancroft GJ, Jacobs WR Jr, Duncan K. Characterization of a *Mycobacterium tuberculosis* H37Rv transposon library reveals insertions in 351 ORFs and mutants with altered virulence. Microbiology. 2002; 148:2975–2986. [PubMed: 12368431]
19. Forrellad MA, Klepp LI, Gioffre A, Sabio y Garcia J, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F. Virulence factors of the *Mycobacterium tuberculosis* complex. Virulence. 2013; 4:3–66. [PubMed: 23076359]
20. Thoden JB, Holden HM. The molecular architecture of human *N*-acetylgalactosamine kinase. J Biol Chem. 2005; 280:32784–32791. [PubMed: 16006554]
21. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007; 40:658–674. [PubMed: 19461840]
22. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
23. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
24. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997; 53:240–255. [PubMed: 15299926]
25. Genthe NA, Thoden JB, Holden HM. Structure of the *Escherichia coli* ArnA *N*-formyltransferase domain in complex with N(5) -formyltetrahydrofolate and UDP-Ara4N. Protein Sci. 2016; 25:1555–1562. [PubMed: 27171345]
26. Knirel, YA., Valvano, MA. Bacterial Lipopolysaccharides. Springer-Verlag; New York: 2011.
27. Tuffal G, Albigot R, Riviere M, Puzo G. Newly found 2-*N*-acetyl-2,6-dideoxy-beta-glucopyranose containing methyl glucose polysaccharides in *M.bovis* BCG: revised structure of the mycobacterial methyl glucose lipopolysaccharides. Glycobiology. 1998; 8:675–684. [PubMed: 9621108]
28. Jarlier V, Nikaido H. Mycobacterial cell wall: structure and role in natural resistance to antibiotics. FEMS microbiology letters. 1994; 123:11–18. [PubMed: 7988876]
29. Jankute M, Nataraj V, Lee OY, Wu HHT, Ridell M, Garton NJ, Barer MR, Minnikin DE, Bhatt A, Besra GS. The role of hydrophobicity in tuberculosis evolution and pathogenicity. Sci Rep. 2017; 7:1315. [PubMed: 28465507]
30. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. J Mol Biol. 1993; 231:1049–1067. [PubMed: 8515464]
31. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA: 2002. The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

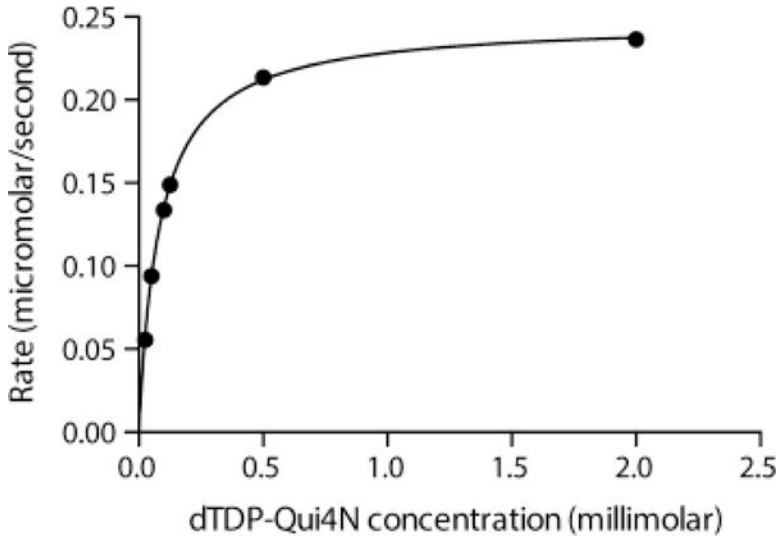


Figure. 1.

Steady state kinetic analysis of Rv3404c. Shown is a plot of the Rv3404c reaction rate versus dTDP-Qui4N concentration. The enzyme demonstrates normal Michaelis-Menten kinetics. In presenting the data as we do, we are adhering to standard conventions in enzymology. Measuring velocities over a wide range of substrate concentrations allows us to obtain data that define both k_{cat} and k_{cat}/K_M well, which is not accomplished by measuring replicates at fewer different concentrations. The graph shown allows for a qualitative appreciation of the quality of the data; the quantitative goodness-of-fit to the Michaelis-Menten equation is given by the standard errors as described in Results and Discussion.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

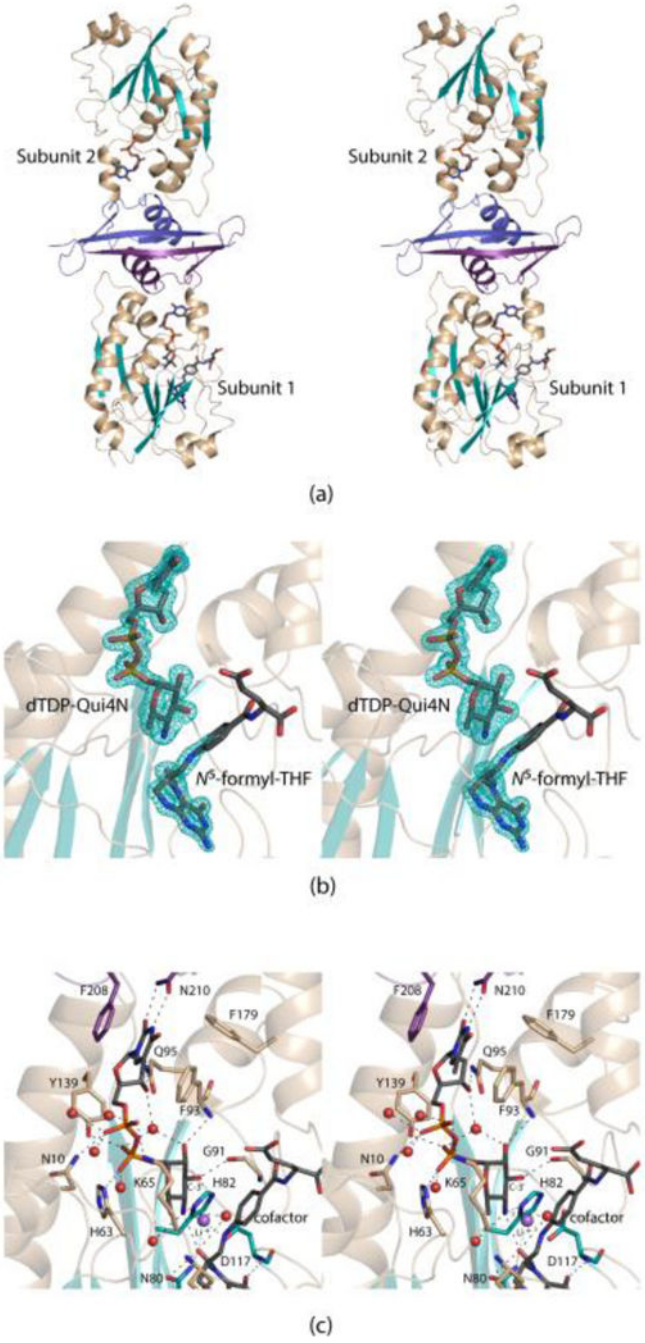


Figure. 2.

Overall structure of Rv3404c. A ribbon representation of the dimer is presented in stereo in (a). The N-terminal domain of each subunit is highlighted in wheat and teal whereas the C-terminal domains are colored in purple and blue for Subunits 1 and 2, respectively. The electron densities corresponding to dTDP-Qui4N and N^5-formyl-THF in Subunit 1 are shown in stereo in (b). The electron density map was calculated with (F_O-F_C) coefficients and contoured at 3σ. The ligands were not included in the X-ray coordinate file used to calculate the omit map, and thus there is no model bias. A close-up stereo view of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Rv3404c active site is presented in (c). Ordered water molecules are depicted as red spheres. The position of the Li^+ ion is indicated by the purple sphere. Possible hydrogen bonding interactions within 3.2 Å are indicated by the dashed lines. Whereas most of the amino acid residues involved in ligand binding are contributed by the N-terminal domain, the side chains of Phe 208 and Asn 210 that interact with the thymine ring are located in the C-terminal region. This figure, and Figs. 3, 4, and 5 were prepared with the software package PyMOL.[31]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

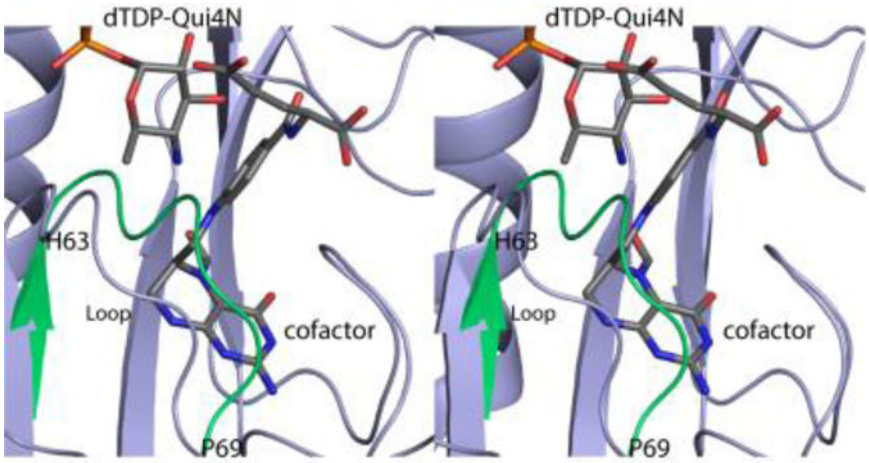


Figure 3.

Comparison of the apo and Rv3404c ternary complex models (in stereo). The only significant difference between the Rv3404c model with bound ligands and the apo form is the position of the loop defined by His 63 to Pro 69. The conformation of the loop observed in the apo structure is highlighted in green.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

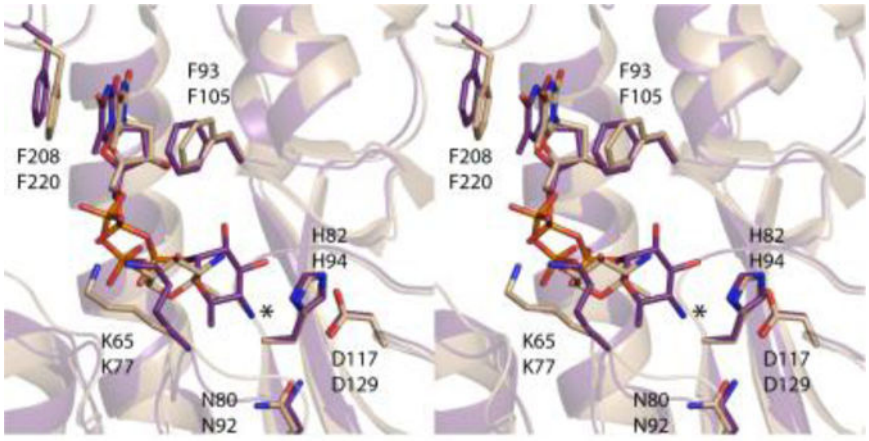


Figure 4.

Comparison of the active sites for Rv3404c and VioF from *P. alcalifaciens* O30. Both enzymes utilize dTDP-Qui4N as substrates. Shown in stereo is a superposition of their active sites with Rv3404c displayed in purple bonds and VioF highlighted in wheat bonds. The top and bottom labels refer to residues in Rv3404c and VioF, respectively. The positions of the strictly conserved asparagine/histidine/aspartate triads are nearly identical in the two enzymes. Note that whereas both Rv3404c and VioF contain a lysine residue in a similar position (Lys 65/Lys 77), the orientations of the side chains are different due to the manner in which the dTDP-Qui4N ligands bind to the two proteins. The observed variations in the binding of the dTDP-sugar substrates are most likely due to crystallization artifacts. Indeed, the orientation of the ligand in VioF is not compatible for catalysis which requires close proximity of the sugar amino group to the conserved histidine residue, His 94. The asterisk indicates the position of the amino group of dTDP-Qui4N when bound to Rv3404c.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

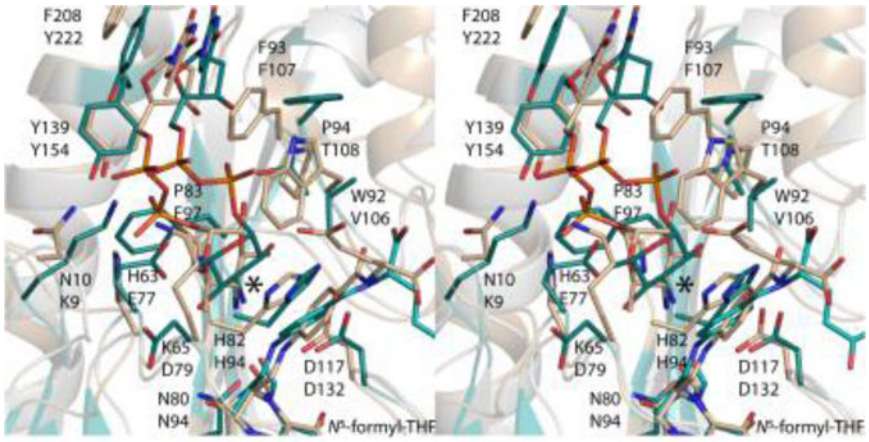
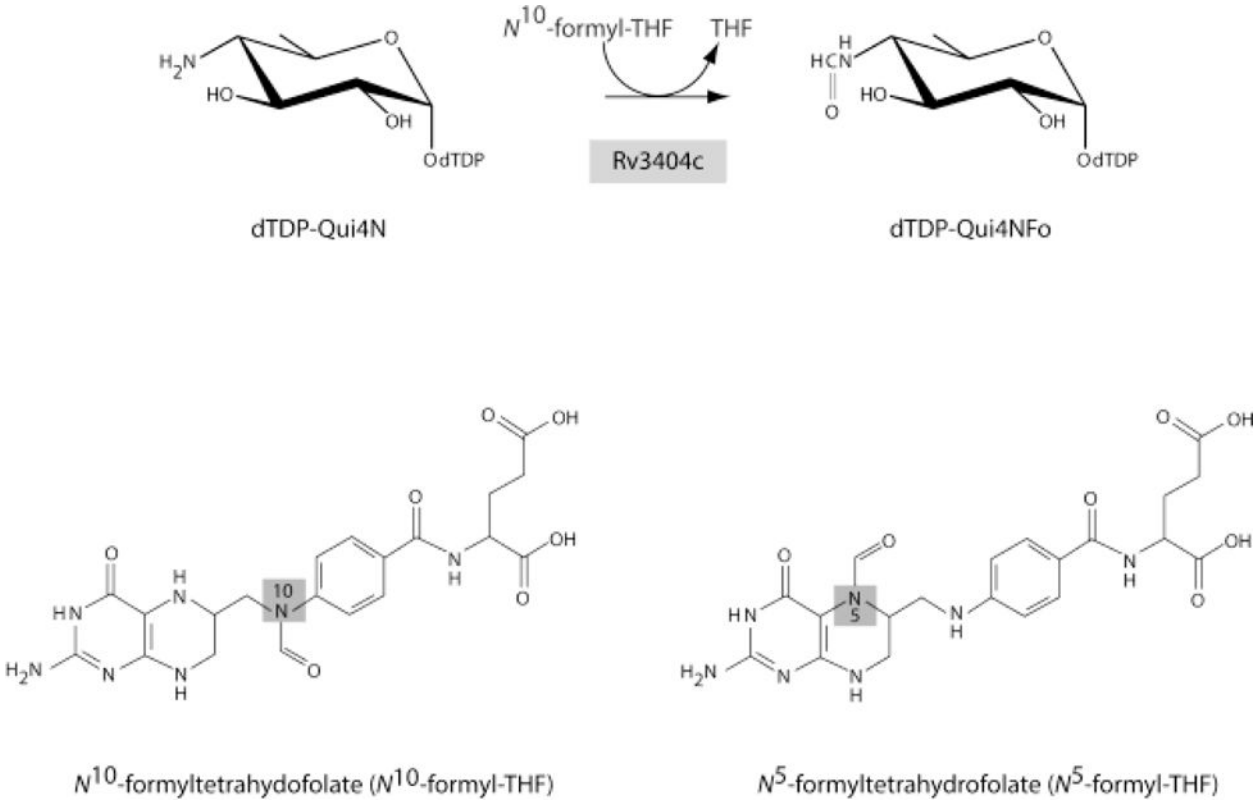


Figure 5.

Comparison of the active sites for Rv3404c and WlaRD from *C. jejuni*. The focus of this investigation, Rv3404c, functions on dTDP-Qui4N whereas the *N*-formyltransferase referred to as WlaRD acts upon dTDP-Qui3N. Shown in stereo is a superposition of the active sites for these enzymes. Side chains from Rv3404c are displayed in wheat. Those amino acid residues belonging to WlaRD are displayed in teal. The top and bottom amino acid labels refer to Rv3404c and WlaRD, respectively. The large asterisk indicates the positions of the amino groups located on the pyranosyl rings of each ligand.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Scheme 1.
Reaction catalyzed by Rv3404c and cofactors employed in this investigation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

X-ray Data Collection Statistics

resolution limits (Å)	50 – 1.6 (1.7 – 1.6)[b]
space group	$P2_12_12_1$
unit cell	
a	47.5
b	73.0
c	173.3
number of independent reflections	79772 (12792)
completeness (%)	99.1 (96.8)
redundancy	6.4 (3.2)
avg I/avg σ(I)	12.4 (2.5)
R_{sym} (%)[a]	7.8 (35.1)

[a] $R_{sym} = \left(\sum \left|I - \bar{I}\right| / \sum I\right) \times 100$.

[b] Statistics for the highest resolution bin.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Refinement Statistics

resolution limits (Å)	50 – 1.6
[a]*R*-factor (overall)%/no. reflections	16.8/79772
R-factor (working)%/no. reflections	16.7/75782
R-factor (free)%/no. reflections	20.0/3990
number of protein atoms	3768
number of heteroatoms	817
average B values	
protein atoms (Å^2)	16.4
ligand (Å^2)	20.5
solvent (Å^2)	28.9
weighted RMS deviations from ideality	
bond lengths (Å)	0.011
bond angles (°)	1.9
planar groups (Å)	0.008
Ramachandran regions (%)[b]	
most favored	98.3
additionally allowed	1.7
generously allowed	0.0

[a]*R*-factor = $(\Sigma|F_O - F_C|/\Sigma|F_O|) \times 100$ where F_O is the observed structure-factor amplitude and F_C. is the calculated structure-factor amplitude.

[b]Distribution of Ramachandran angles according to PROCHECK.[30]