Published in final edited form as:

Biochemistry. 2017 July 18; 56(28): 3657–3668. doi:10.1021/acs.biochem.7b00494.

Biochemical Characterization of WbkC, an *N*-formyltransferase from *Brucella melitensis*

Alexander S. Riegert¹, Daniel P. Chantigian^{1,3}, James B. Thoden¹, Peter A. Tipton², and Hazel M. Holden^{1,*}

¹Department of Biochemistry, University of Wisconsin, Madison, WI, 53706

²Department of Biochemistry, University of Missouri, Columbia, MO, 65211

Abstract

It has become increasingly apparent within the last several years that unusual *N*-formylated sugars are often found on the O-antigens of such Gram negative pathogenic organisms as *Francisella tularensis*, *Campylobacter jejuni*, and *Providencia alcalifaciens*, amongst others. Indeed, in some species of *Brucella*, for example, the O-antigen contains 1,2-linked 4-formamido-4,6-dideoxy- α -D-mannosyl groups. These sugars, often referred to as *N*-formylperosamine, are synthesized in pathways initiating with GDP-mannose. One of the enzymes required for the production of *N*-formylperosamine, namely WbkC, was first identified in 2000 and was suggested to function as an *N*-formyltransferase. Its biochemical activity was never experimentally verified, however. Here we describe a combined structural and functional investigation of WbkC from *Brucella melitensis*. Four high resolution X-ray structures of WbkC were determined in various complexes to address its active site architecture. Unexpectedly, the quaternary structure of WbkC was shown to be different from that previously observed for other sugar *N*-formyltransferases. Additionally, the structures revealed a second binding site for a GDP molecule distinct from that required for GDP-perosamine positioning. In keeping with this additional binding site, kinetic data with the wild type enzyme revealed complex patterns. Removal of GDP binding by mutating Phe 142 to an alanine residue resulted in an enzyme variant displaying normal Michaelis-Menten kinetics. These data suggest that this nucleotide binding pocket plays a role in enzyme regulation. Finally, by using an alternative substrate, we demonstrate that WbkC can be utilized to produce a trideoxysugar not found in nature.

Graphical abstract

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

³Present address: Department of Rehabilitation Medicine, University of Minnesota, Minneapolis, MN, 55455

X-ray coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, N. J. (accession nos. 5VYR, 5VYS, 5VYT, and 5VYU).

The authors have no competing financial interests.

Introduction

Brucellosis is a zoonotic infection caused by bacteria belonging to the genus *Brucella*.¹ These Gram negative organisms are facultative intracellular pathogens, and as a consequence infections with them can lead to such chronic health issues as arthritis and endocarditis, as well as systemic involvement of various organs including the heart and brain.¹ Whereas *Brucella melitensis*, *Brucella suis*, *Brucella abortus*, and *Brucella canis* mainly affect sheep, pigs, cattle, and dogs, respectively, humans can become infected through the direct contact with sick animals or the consumption of unpasteurized dairy products or tainted foods. Those most at risk include veterinarians, zoo technicians, and farmers. Acute brucellosis is characterized by such symptoms as general weakness, muscle and joint pain, headaches, fever, and gastrointestinal distress, amongst others. Worldwide *B. melitensis* is the most prevalent and virulent of the species with more than 500,000 human cases reported each year.² The Centers for Disease Control and Prevention classify *B. melitensis*, *B. suis*, and *B. abortus* as select agents that have the potential to be utilized as bioterrorism agents. Interestingly, brucellosis is thought to be an ancient disease with reports, for example, of Egyptian bones that date back to 750 BC showing evidence of infection.³

The major virulence factor for any of the *Brucella* species appears to be the lipopolysaccharide or LPS, a complex glycoconjugate that extends farthest away from the bacterium outer cell wall.¹ Conceptually, the LPS can be envisioned in terms of three distinct components: the lipid A, the core oligosaccharide, and the O-antigen chain. The O-antigen of smooth-type *Brucella* is highly unusual in that it contains repeated 1,2-linked 4-formamido-4,6-dideoxy- α -D-mannosyl groups forming an unbranched homopolymer chain of up to 100 units (Scheme 1).^{4,5}

The pathway for the biosynthesis of 4-formamido-4,6-dideoxy-D-mannose, hereafter referred to as *N*-formylperosamine, was proposed in 2000 and is outlined in Scheme 2.⁶

Note that *N*-formylperosamine is synthesized as a GDP-linked derivative before it is appended to the O-antigen chain. Given our recent research interest in *N*-formyltransferases

involved in unusual sugar biosynthesis,^{7–12} we were especially intrigued by the enzyme WbkC, which catalyzes the last step in the formation of *N*-formylperosamine, namely the transfer of a formyl group from *N*¹⁰-formyltetrahydrofolate to GDP-4-amino-4,6-dideoxymannose (abbreviated as GDP-perosamine). For reader clarity, the structure of *N*¹⁰-formyltetrahydrofolate is provided in Scheme 3. In addition, the structure of *N*⁵-formyltetrahydrofolate, a stable cofactor analog used in this investigation, is also shown. Importantly, loss of activity of WbkC results in a bacterial strain with attenuated pathogenicity, and thus the three-dimensional structure of WbkC could serve as a scaffold for structure-based drug design.¹³

Here we report four X-ray structures of WbkC that were determined to address the active site architecture of this fascinating sugar *N*-formyltransferase. In addition, a detailed kinetic analysis was conducted to assess the enzymatic activity of WbkC. Finally, we provide data demonstrating that WbkC not only functions on GDP-perosamine, but also on the alternative substrate GDP-3-deoxyperosamine, thereby allowing for the *in vitro* production of an unusual sugar not normally encountered in nature.

Materials and Methods

Cloning, Expression, and Purification of Wild Type WbkC

The gene encoding WbkC from *B. melitensis* was synthesized by DNA2.0 for optimized *Escherichia coli* codon usage. It was provided in a pJ411 expression plasmid that ultimately leads to a protein with an N-terminal poly-histidine tag possessing a TEV protease cleavage site. This plasmid was utilized to transform Rosetta(DE3) *E. coli* cells (Novagen). The cells harboring the pJ411-*wbkC* plasmid were grown with shaking at 37 °C in lysogeny broth supplemented with kanamycin (50 mg/L) and chloramphenicol (50 mg/L). When the optical density reached 0.8 at 600 nm, the flasks were cooled in an ice water bath. Following the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside, the cells were allowed to express protein at 16 °C for 18 h.

WbkC was purified by standard procedures using Ni-nitrilotriacetic acid resin. Following purification, TEV protease was added in a 1:30 molar ratio, and the reaction dialyzed at 4 °C against 50 mM sodium phosphate, 300 mM sodium chloride, and 20 mM imidazole (pH 8.0) for 36 h. Uncleaved protein and the TEV protease were removed by passage over Ni-nitrilotriacetic acid resin. The tag-free protein was dialyzed against 10 mM Tris (pH 8.0) and 200 mM NaCl and concentrated to approximately 11 mg/mL based on the calculated extinction coefficient of 1.11 (mg/mL)⁻¹ cm⁻¹.

To express selenomethionine-labeled WbkC, the *E. coli* Rosetta(DE3) cells containing the pJ411-*wbkC* plasmid were grown in M9 minimal media at 37 °C supplemented with kanamycin (50 mg/L) and chloramphenicol (50 mg/L) to an optical density of 0.9 at 600 nm. 100 mg/L each of L-lysine, L-threonine, and L-phenylalanine, and 50 mg/L each of L-leucine, L-isoleucine, L-valine, and L-selenomethionine were added to the cultures. After 20 min of additional growth, the cells were cooled on an ice water bath for 20 min. Following the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside, the cells were allowed to

express protein at 16 °C for 18 h. The protein was purified and concentrated as described for the wild type enzyme.

Initial X-ray Structural Analysis

Crystallization conditions were surveyed by the hanging drop method of vapor diffusion using a sparse matrix screen developed in the Holden laboratory. Experiments were conducted at both room temperature and at 4 °C with either the apoprotein or with the protein in the presence of GDP and *N*^δ-formyltetrahydrofolate (*N*^δ-formyl-THF). Crystals of the protein/GDP/*N*^δ-formyl-THF complex grew at room temperature after three to four weeks from 12 – 15% poly(ethylene glycol) 8000, 2% 2-methyl-2,4-pentanediol, 5 mM GDP, 5 mM *N*^δ-formyl-THF, and 100 mM Homo-PIPES (pH 5.0). The crystals belonged to the space group *P*₄₃₂₁₂ with unit cell dimensions of *a* = *b* = 66.5 Å and *c* = 235.9 Å. The asymmetric unit contained one dimer. Prior to X-ray data collection, the crystals were transferred to solutions composed of 25% poly(ethylene glycol) 8000, 2% 2-methyl-2,4-pentanediol, 350 mM NaCl, 5 mM GDP, 5 mM *N*^δ-formyl-THF, and 100 mM Homo-PIPES (pH 5.0). Crystals of the selenium-labeled protein were similarly prepared.

X-ray data to 2.3 Å resolution from the selenium-labeled protein crystals were collected at the Structural Biology Center beamline 19-BM at a peak wavelength of 0.9794 Å (Advanced Photon Source). The X-ray data were processed and scaled with HKL3000.¹⁴ The structure was solved via single wavelength anomalous dispersion phasing. Twelve of the eighteen possible selenium atoms were located using the software package SOLVE.¹⁵ The resulting electron density map was further improved by twofold averaging and solvent flattening with RESOLVE and allowed for an initial model of WbkC to be built.¹⁶ X-ray data to 1.9 Å resolution from a crystal of the wild type protein were subsequently collected in-house at 100 K using a Bruker AXS Platinum 135 CCD detector controlled with the Proteum software suite (Bruker AXS Inc.). The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics and operated at 50 kV and 90 mA. These X-ray data were processed with SAINT (Bruker AXS Inc.) and internally scaled with SADABS (Bruker AXS Inc.). Using this dataset, the initial WbkC model was subjected to alternating cycles of manual model building with COOT and refinement with REFMAC.^{17,18}

Site-Directed Mutagenesis and Subsequent Structural Analyses

The first structure solved of WbkC employed crystals grown at pH 5.0. There were many breaks in the electron density map corresponding to surface loops. In an effort to grow better crystals at a higher pH value, the following surface mutations, based on observed crystalline contacts, were constructed: E68A, D78A, R94A, E68A/D78A, E68A/R94A, D78A/R94A, and E68A/D78A/R94A. These site-directed mutant proteins, as well as all others in this study, were prepared via the Stratagene QuikChange method.

Only the D78A mutant protein produced crystals that could be grown at a higher pH value. Specifically, the crystals were obtained from hanging drop experiments at room temperature against solutions composed of 18 – 22% poly(ethylene glycol) 5000, 5 mM GDP, 5 mM *N*^δ-formyl-THF, and 100 mM MES (pH 6.0). Prior to X-ray data collection, the crystals were

transferred to solutions composed of 29% poly(ethylene glycol) 5000, 300 mM NaCl, 15% ethylene glycol, 5 mM GDP, 5 mM *N*⁵-formyl-THF, and 100 mM MES (pH 6.0). The crystals belonged to the space group *P*₄₃₂₁₂ with unit cell dimensions of *a* = *b* = 65.9 Å and *c* = 233.1 Å. The asymmetric unit contained one dimer. The structure was solved by molecular replacement with the software package PHASER and using as the search model the WbkC structure determined from crystals grown at pH 5.0.¹⁹ This new model was refined to 1.7 Å resolution by alternating cycles of manual manipulation with COOT and refinement with REFMAC.

As will be described in Results and Discussion, two additional site-directed mutant proteins were constructed, namely the C47A and the F142A variants. Crystals of the C47A variant grew at room temperature from 19 – 25% poly(ethylene glycol) 5000, 5 mM GDP-perosamine, 5 mM *N*⁵-formyl-THF, and 100 mM MOPS (pH 7.0). The required GDP-perosamine was prepared as previously described.²⁰ Prior to X-ray data collection, the crystals were transferred to solutions composed of 30% poly(ethylene glycol) 5000, 300 mM NaCl, 15% ethylene glycol, 5 mM GDP-perosamine, 5 mM *N*⁵-formyl-THF, and 100 mM MOPS (pH 7.0). The crystals belonged to the space group *P*₄₃₂₁₂ with unit cell dimensions of *a* = *b* = 65.9 Å and *c* = 234.7 Å and one dimer in the asymmetric unit. The structure was solved by molecular replacement and refined to 2.2 Å resolution.

Crystals of the F142A mutant variant were grown at 4 °C from 13 – 15% poly(ethylene glycol) 5000, 2% glycerol, 5 mM GDP-perosamine, 5 mM *N*⁵-formyl-THF, and 100 mM MES (pH 6.0). They belonged to the space group *P*₂₁₂₁₂ with unit cell dimensions of *a* = 67.3 Å, *b* = 68.0 Å, and *c* = 236.0 Å. The asymmetric unit contained two dimers. Prior to X-ray data collection, the crystals were transferred to solutions composed of 24% poly(ethylene glycol) 5000, 15% glycerol, 300 mM NaCl, 5 mM GDP-perosamine, 5 mM *N*⁵-formyl-THF, and 100 mM MES (pH 6.0). The structure was solved by molecular replacement and refined to 2.2 Å.

To determine the structure of the F142A protein/*N*¹⁰-formyl-THF/GDP-perosamine ternary complex, the F142A crystals were soaked in a solution composed of 19% poly(ethylene glycol) 5000, 5 mM *N*¹⁰-formyl-THF, 5 mM GDP-perosamine, and 100 mM MES (pH 6.0). The solution was exchanged four times to ensure removal of *N*⁵-formyl-THF. Soaked crystals were subsequently transferred to solutions composed of 30% poly(ethylene glycol) 5000, 2% isopropanol, 15% glycerol, 300 mM NaCl, 5 mM GDP-perosamine, 5 mM *N*¹⁰-formyl-THF, and 100 mM MES (pH 6.0). The crystals belonged to the space group *P*₄₃₂₁₂ with unit cell dimensions of *a* = *b* = 66.8 Å and *c* = 235.9 Å. The asymmetric unit contained one dimer. The structure was solved by molecular replacement, and the model was refined to 2.2 Å resolution. The *N*¹⁰-formyl-THF cofactor was prepared as previously described.⁷ Relevant X-ray data collection and model refinement statistics for all of the structures are provided in Tables 1 and 2, respectively.

Kinetic Analyses

In general, the kinetic constants for WbkC with GDP-perosamine and *N*¹⁰-formyl-THF were determined via a discontinuous assay using an ÄKTA HPLC system. The reaction rates were measured by calculating the amount of formylated product (GDP-*N*-formylperosamine)

produced on the basis of the peak area at 253 nm. The area was correlated to concentration via a calibration curve created with standard samples that had been treated in the same manner as the reaction aliquots.

Specifically, the kinetic parameters were measured at room temperature using 1 mL reaction mixtures containing 5 mM *N*¹⁰-formyl-THF, 50 mM HEPPS (pH 8.0), 0.17 μM enzyme, and substrate concentrations ranging from 0.025 mM to 10.0 mM GDP-perosamine. Four 250 μL aliquots were taken at time points of 0, 30, 60, and 90 s. The reactions were quenched by the addition of 6 μL of 6 M HCl. Afterwards, 200 μL of carbon tetrachloride was added, the samples were vigorously mixed and spun at 14000 × *g* for 2 min, and 200 μL of the aqueous phase was taken for subsequent analysis via HPLC. The samples were diluted with 2 mL of water and loaded onto a 1 mL ResQ column. The products were quantified after elution with a 10-column volume gradient from 0 to 400 mM LiCl (pH 4.0, HCl). The D78A variant, which is referred in this paper as wild type for the sake of clarity, displayed similar kinetic properties/parameters (unpublished data).

The kinetic constants for D78A/F142A variant with GDP-perosamine and *N*¹⁰-formyl-THF were determined in an analogous manner. The enzyme concentration utilized was 0.52 μM and substrate concentrations ranged from 0.02 mM to 4.0 mM GDP-perosamine. Kinetic parameters were as follows: $K_m = 0.46 (\pm 0.06)$ mM, $k_{cat} = 0.54 (\pm 0.04)$ s⁻¹, k_{cat}/K_m 1200 (± 180) M⁻¹s⁻¹.

Analysis of the Product using GDP-3-deoxyperosamine as a Substrate

1 mL reactions composed of 1 mg/mL enzyme, 1 mM GDP-perosamine (or 1 mM GDP-3-deoxyperosamine), 2 mM *N*¹⁰-formyl-THF, and 50 mM HEPPS (pH 8.0) were allowed to incubate at 37 °C overnight. The enzymes were removed via filtration through a 10 kDa membrane. The reaction mixtures were diluted tenfold with water, and the reaction products resolved via anion exchange chromatography (1 mL ResQ, 0 – 1 M ammonium acetate (pH 4.0), 15 column volume gradient). Electrospray ionization mass spectral analyses of the substrates and products were performed at the University of Wisconsin Biotechnology Mass Spectroscopy Facility. The GDP-3-deoxyperosamine substrate was prepared as previously described.²⁰

Results and Discussion

Initial X-ray Model of WbkC

The structure of the apo-form of WbkC was first determined to 1.9 Å resolution using crystals grown at pH 5.0. Due to the lack of bound ligands, however, there were numerous regions in the electron density map that were disordered. In addition, experience has shown that the low pH employed for crystallization probably would hinder the subsequent investigation of ligand:protein interactions, or even more importantly result in artefactual ligand binding. On the basis of this preliminary model, and in light of crystal packing considerations, various surface mutations were constructed via site-directed mutagenesis. Ultimately, the surface mutation, D78A, allowed for crystals of WbkC to be grown at pH 6.0, which yielded an electron density map with fewer breaks. Thus, the following

discussion refers to the D78A mutant variant, which for this investigation will be considered “wild type.” Note that the position of the mutation is well over 20 Å from the active site pocket.

The crystals utilized in the X-ray analysis of WbkC at pH 6.0 belonged to the space group $P4_32_12$ and contained a dimer in the asymmetric unit. In keeping with the packing of the asymmetric unit, gel filtrations experiments demonstrated that WbkC migrates as a dimer (data not shown). A ribbon representation of the dimer, which has overall dimension of $\sim 43 \text{ Å} \times 87 \text{ Å} \times 63 \text{ Å}$, is displayed in Figure 1a. The model was solved to a nominal resolution of 1.7 Å and refined an overall R -factor of 18.7%. The total buried surface area upon dimerization is $\sim 3000 \text{ Å}^2$ and is formed by the first, second, and six α -helices of each subunit defined by Leu 16 to Gly 27, Cys 47 to Leu 55, and Ala 161 to Asp 184, respectively.

The electron density map clearly revealed the presence of a disulfide bridge formed between Cys 47 from each subunit. Although WbkC was crystallized in the presence of GDP and N^5 -formyl-THF, only fragments of these ligands could be modeled into the observed electron densities. Shown in Figure 1b is the electron density corresponding to these fragments as observed in Subunit 2. Surprisingly, in addition to the GDP and N^5 -formyl-THF fragments, a third peak of electron density corresponding to guanine was observed binding in Subunit 2 (Figure 1b). Most likely it was a contaminant of the GDP solutions utilized in the investigation. As discussed below, the position of the guanine marks where the GDP-sugar substrate ultimately binds. Electron densities for the guanine ring as well as for the N^5 -formyl-THF fragment were also observed in Subunit 1, but there was no density for the GDP fragment.

WbkC contains 259 amino acids.²¹ The polypeptide chain for Subunit 1 extends from Ala 2 to Ser 252 with a break between Leu 235 and Tyr 242, whereas the polypeptide chain for Subunit 2 is continuous from Ala 2 to Ser 246. The α -carbons for the two subunits of the dimer superimpose with a root-mean-square deviation of 0.3 Å. As such, the following discussion will refer only to Subunit 2 given that the polypeptide chain was continuous from the N- to C-terminus.

A stereo view of the ribbon representation for Subunit 2 is presented in Figure 1c. The overall architecture of the subunit is dominated by a seven-stranded mixed β -sheet that is flanked on one side by two α -helices and on the other by three α -helices, one of which is quite long (Ala 161 – Asp 184). In addition to these secondary structural elements, there are two additional α -helices and a β -hairpin containing a Type I' turn. The observed break in Subunit 1 occurs in the β -hairpin motif. Two *cis* prolines are present at positions 112 and 229. The *cis* proline located at position 112 is part of a conserved signature sequence of HxSLLPKxxG, first identified in an *N*-formyltransferase (ArnA) from *E. coli*.²² In WbkC, the sequence is HPSLLPAYRG with these residues adopting a random coil conformation that connects β -strand five to α -helix five. The second *cis* proline at 229 is not strictly conserved amongst the sugar *N*-formyltransferases. It is located in the random coil defined by Phe 226 to Ala 233, which connects α -helix seven to β -strand eight. This random coil

region is proline-rich with a sequence of FFPPFPPA, and the conformation is such that it projects the side chains of Phe 227 and Phe 230 over the guanine moiety (Figure 1c).

Kinetic Analyses

As is evident in Figure 2a, the steady-state kinetic behavior of WbkC is complex. Substrate inhibition by GDP-perosamine is evident above 250 μM ; the velocity does not decrease to zero at high concentrations, indicating that the substrate inhibition is partial. The experimentally determined initial velocities were fitted to Equation 1:²³

$$v = \frac{V_{\max} \left(1 + \frac{V'_{\max} \cdot A^n}{V_{\max} \cdot K_i} \right)}{1 + \frac{K_m}{A^n} + \frac{A^n}{K_i}} \quad (1)$$

This equation describes partial substrate inhibition caused by the binding of a second substrate molecule. In the present case, this represents binding of GDP-perosamine to the WbkC•N¹⁰-formyl-THF•GDP-perosamine ternary complex, and K_i is its dissociation constant. When there is no cooperativity between the two subunits, n has a value of 1. The model presumes that the inhibitory complex is not inactive, but causes the reaction to follow an alternate pathway to product formation, whose maximal velocity is described by $V'_{\max}$.

There are too many parameters in Equation 1 for the data to be fitted unequivocally, however. To simplify the analysis, we analyzed the data obtained at very low substrate concentrations independently. The velocities obtained when GDP-perosamine was varied between 50 μM and 260 μM fit very well to the Michaelis-Menten equation and yielded values of 4.56 ± 0.10 $\mu\text{M/s}$ for $V_{\max}$ and 184 ± 8 μM for K_m (Figure 2b). We assumed a value of 0.8 $\mu\text{M/s}$ for $V'_{\max}$, based on inspection of the experimental data, and fitted the dataset to Equation 1, also setting $V_{\max}$ as a fixed parameter with a value of 4.6 $\mu\text{M/sec}$. A reasonable fit to the data was obtained, with values of 140 ± 20 μM for K_m and 310 ± 40 μM for K_i . The close agreement between the values determined for K_m from fitting the limited dataset to the Michaelis-Menten equation and from fitting the entire dataset to Equation 1 is noteworthy. However, systematic deviations between the experimental data and the fit are evident. We explored whether a model in which cooperativity between the two subunits was allowed, *i.e.*, the concentration terms in Equation 1 were raised to an exponent. Figure 2a shows that the data are fitted very well by a model that incorporates modest cooperativity, described by a Hill coefficient of 1.50 ± 0.05 . From this analysis, values of 50 ± 5 μM and 190 ± 10 μM were determined for K_m and K_i , respectively. Further evidence for cooperativity was provided by the kinetic analysis of the F142A mutant protein as described below.

Substrate inhibition usually arises from the binding of two molecules of the same substrate. The present structural analysis provides a strong suggestion that the second binding site is in the vicinity of F142, which is at one end of the active site cleft. It is possible that binding of a second GDP-perosamine interferes with product release, leading to a reduced $V_{\max}$.

Additional Structures of WbkC

Whereas the initial structural analysis of WbkC allowed for its overall tertiary and quaternary structure to be described, details regarding its active site architecture were missing given that only fragments of GDP and *N*^δ-formyl-THF were observed in the electron density map. There was also concern that the disulfide bridge formed between the two subunits of the dimer might have affected ligand binding. As such, the site-directed mutant variant, C47S, was subsequently constructed. Crystals of this mutant protein, grown in the presence of GDP and *N*^δ-formyl-THF, also belonged to the space *P*₄₃₂₁² with the asymmetric unit containing a dimer. The model was refined to 2.2 Å resolution with an overall *R*-factor of 18.9%. The α-carbons between the wild type and the C47S mutant variant superimpose with a root-mean-square deviation of 0.2 Å, suggesting no major conformational changes occurred upon formation of the disulfide bridge during the crystallization process. In this structure, however, GDP was observed binding to Subunit 2 in the same region where the guanine ring was located in the original model. In addition, fragments of *N*^δ-formyl-THF and GDP were also observed. Only GDP and a fragment of *N*^δ-formyl-THF were observed binding to Subunit 1.

A close-up stereo view of the ligand binding pocket is presented in Figure 3. The guanine ring of the GDP moiety is surrounded on one side by Leu 203 and on the other by Phe 230. It is also hydrogen bonded to a water molecule and the carbonyl oxygen of Met 225. The side chains of Arg 45 and Arg 88, as well as two water molecules and the backbone amide group of Asn 118, serve to anchor the pyrophosphoryl group into the active site cleft. With respect to the GDP fragment, the side chains of Phe 142 and Asp 143 play key roles in ligand positioning. Given the close approach of the side chain of Asp 143 to the GDP fragment, it can be concluded that it is protonated. This may be a function of the pH at which the crystals were obtained.

Our previous investigations on sugar *N*-formyltransferases have demonstrated that the *p*-aminobenzoic acid moiety of the *N*^δ-formyl-THF cofactor occupies the region where the GDP fragment binds. In light of this result, and in keeping with the unusual kinetics displayed by the wild type enzyme, it was postulated that this additional binding site perhaps functions in a regulatory role. In order to address this hypothesis, the site-directed mutant variant, F142A, was subsequently constructed, and the protein was crystallized in the presence of GDP and *N*^δ-formyl-THF. Crystals of the F142A mutant protein belonged to the space group *P*₂₁₂₁₂₁ with two dimers in the asymmetric unit. The model for the F142A variant was refined to 2.2 Å resolution with an overall *R*-factor of 20.5%. The electron densities for the polypeptide chains of Subunits 2, 3, and 4 were continuous whereas that for Subunit 1 had a break between Ile 237 to Tyr 242. The four subunits in the asymmetric unit superimpose with root-mean-square deviations of between 0.3 to 0.6 Å. In all cases, the GDP fragments were missing, and in their positions were electron densities corresponding to the 2-amino-4-oxo-5-formyl-6-methylpteridine and *p*-aminobenzoic acid groups of the *N*^δ-formyl-THF ligands. The F142A mutant variant did not exhibit substrate inhibition, but did display kinetic behavior expected for an enzyme with modest cooperativity (Figure 4). The best fit of the experimental data yielded values of $259 \pm 5 \mu\text{M/s}$ for V_{max} and $400 \pm 20 \mu\text{M}$

for K_m . It is noteworthy that the Hill coefficient determined in this analysis was 1.4 ± 0.1 , closely matching the value determined for the wild type enzyme.

Structure of the Ternary Complex of WbkC

In an attempt to produce a ternary complex of WbkC in the presence of N^{10} -formyl-THF and GDP-perosamine, the F142A crystals were soaked in solutions containing 10 mM concentrations of both ligands. Clear density was observed in both subunits for the N^{10} -formyl-THF cofactors. This represents only the second time in which the catalytically competent cofactor has been seen binding to an *N*-formyltransferase. The first observation was during our structural investigation of a sugar *N*-formyltransferase from *Campylobacter jejuni*.⁷ In Subunit 2 only GDP could be modeled into the electron density whereas in Subunit 1 there was weak density for the perosamine group. A close-up view of the active site with N^{10} -formyl-THF and GDP-perosamine (Subunit 1) is presented in Figure 5. There are three residues, Asn 105, His 107, and Asp 143 that are strictly conserved amongst the *N*-formyltransferases. Mutation of any of these residues, for example in the case of the *C. jejuni* *N*-formyltransferase, results in complete loss of enzymatic activity.⁷ The N^{10} -formyl-THF cofactor is anchored in place by hydrogen bonds to backbone amide and carbonyl groups and two water molecules. With respect to the GDP-perosamine ligand, the guanine ring is surrounded by Leu 203 and Phe 230 and lies within 3.2 Å to the carbonyl oxygen of Met 225 and two waters. Four additional water molecules interact with the substrate. The guanidinium groups of Arg 45 and Arg 88 interact with the α -phosphoryl group of the substrate. The electron density for the perosamine ring was weak but it appears that there are no hydrogen bonding interactions between it and the protein. Importantly the distance between the pyranosyl C-4' amino group and the carbonyl carbon of the N^{10} -formyl-THF cofactor is 4.2 Å, and it is nearly 7 Å from His 107, which is thought to be the catalytic base. Most likely the GDP- sugar bound in a nonproductive conformation perhaps due to the lower pH at which the crystals were grown.

Production of a Novel GDP-linked Sugar

Given the lack of observed protein interactions with the C-3' hydroxyl group of the GDP-perosamine substrate, we next tested whether WbkC could function on GDP-3-deoxyperosamine in our continued attempts to provide a “molecular toolbox” for the enzymatic synthesis of new di- and trideoxysugars. Shown in Figure 6a is the mass spectrum of the WbkC natural substrate, GDP-4-amino-4,6-dideoxymannose. The peak at 587.0916 corresponds to the substrate. In Figure 6b, the mass spectrum is shown following incubation of WbkC with GDP-4-amino-4,6-dideoxymannose. The peak at 615.0864 corresponds to the formylated product (addition of the expected 28 amu). The mass spectrum for GDP-3-deoxyperosamine should demonstrate a peak at 571, due to the loss of the hydroxyl group, and indeed that is the case for our starting substrate as displayed in Figure 6c. Upon incubation of WbkC with GDP-3-deoxyperosamine, a peak is observed at 599.0928 (Figure 6d), which corresponds to the addition of a formyl group.

Comparison with Other Sugar *N*-formyltransferases

The first structural investigation of a sugar *N*-formyltransferase, namely ArnA from *E. coli*, was reported back in 2005.^{22,24,25} This enzyme does not, however, produce a formylated

sugar ultimately found on the O-antigen; rather formylation of the sugar is thought to be an obligatory step in the modification of lipid A with 4-amino-4-deoxy-L-arabinose.²⁶ There was a hiatus of investigations on sugar *N*-formyltransferases until 2013 when the structure of WlaRD from *C. jejuni* was reported.^{7,12} This enzyme catalyzes the formylation of either dTDP-3-amino-3,6-dideoxyglucose (dTDP-Qui3N) or dTDP-3-amino-3,6-dideoxygalactose (dTDP-Fuc3N). Its natural substrate *in vivo* is presently unknown, although the reported kinetic data demonstrate that it has a higher catalytic efficiency with dTDP-Qui3N as a substrate ($1.1 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ versus $5.6 \times 10^1 \text{ M}^{-1}\text{s}^{-1}$). A ribbon representation of WlaRD is displayed in Figure 7a. Following this investigation, the three-dimensional structures of two additional *N*-formyltransferases from *Pseudomonas alcalifaciens* O40 and *Salmonella enterica* O60 were reported.^{9,27} These enzymes, referred to as QdtF and FdtF, function specifically on dTDP-Qui3N and dTDP-Fuc3N, respectively. Surprisingly, these two enzymes contained, in addition to the canonical *N*-formyltransferase folds, ankyrin repeats at the C-terminal regions that housed additional dTDP-sugar binding pockets. It is thought that the ankyrin repeats play an allosteric role. A ribbon representation of QdtF is presented in Figure 7b. As can be seen, the ankyrin repeats splay away from the main body of the dimer. With the exception of these ankyrin repeats, however, the three-dimensional architectures of the *N*-formyltransferases that function on dTDP-linked 3-amino sugars are similar.

There are additional structures of *N*-formyltransferase now known, namely WbtJ and VioF from *Francisella tularensis* and *P. alcalifaciens* O30, respectively.^{8,10} These enzymes function on dTDP-4-amino-4,6-dideoxyglucose (dTDP-Qui4N). Whereas their overall active site architectures are similar to WlaRD, QdtF, and FdtF, their quaternary structures are remarkably different as can be seen in the ribbon representation for VioF in Figure 7c.

Given that the substrate for WbkC, the focus of this investigation, is a GDP-linked 4-amino sugar, it might have been anticipated that it would adopt a similar quaternary structure to that observed for WbtJ and VioF. As can be seen by comparing Figure 1a and Figure 7c, the quaternary structures of WbkC, WbtJ, and VioF are unmistakably different. Indeed, the WbkC structure reported here represents yet another example of how sugar *N*-formyltransferases form functional dimers. Shown in Figure 8 is a superposition of the ribbon representations for WlaRD, VioF, and WbkC. The N-terminal domains of each protein display the same three-dimensional architectures. The differences in quaternary structures arise primarily from the divergence of the polypeptide chains in the C-terminal regions. The α -carbons for WbkC and VioF superimpose with a root-mean-square deviation of 1.9 Å for 205 target pairs whereas those for WbkC and WlaRD correspond with a root-mean-square deviation of 1.5 Å for 208 target pairs. Note that the positions of the cofactors and the three catalytic residues (Asn 105, His 107, and Asp 143 in WbkC) in all these enzymes are decidedly similar. Both Asn 105 and His 107 are positioned on the fifth β -strand of the subunit, and Asp 143 is located in the loop connecting the sixth and seventh β -strands.

Summary

In conclusion, the research described here on WbkC provides new molecular insight into the intriguing sugar *N*-formyltransferases. The quaternary structure of WbkC was completely unexpected as was the additional binding pocket for a GDP-based ligand. The complex kinetic patterns displayed by WbkC suggest a degree of cooperativity between the two subunits, which has not been previously observed for other family members. Finally, we show that WbkC can function on GDP-3-deoxypentosamine as well as on GDP-pentosamine, thereby providing an enzymatic method for producing a novel trideoxysugar.

Acknowledgments

A portion of the research described in this paper was performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source (U. S. Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357). We gratefully acknowledge Dr. Norma E. C. Duke for assistance during the X-ray data collection at Argonne. The authors thank an anonymous reviewer for bringing reference 30 to their attention. This research was supported in part by NIH grant GM115921 (to H. M. H.).

Abbreviations

GDP

guanosine diphosphate

GDP-pentosamine

GDP-4-amino-4,6-dideoxy-D-mannose

GDP-*N*-formylpentosamine

GDP-4-formamido-4,6-dideoxy-D-mannose

HEPPS

N-2-hydroxyethylpiperazine-*N'*-3-propanesulfonic acid

Homo-PIPES

homopiperazine-1,4-bis(2-ethanesulfonic acid)

HPLC

high-performance liquid chromatography

MES

2-(*N*-morpholino)ethanesulfonic acid

MOPS

3-(*N*-morpholino)propanesulfonic acid

***N*-formylpentosamine**

4-formamido-4,6-dideoxy-D-mannose

***N*⁵-formyl-THF**

*N*⁵-formyltetrahydrofolate

***N*¹⁰-formyl-THF**

*N*¹⁰-formyltetrahydrofolate

TEV

tobacco etch virus

Tris

tris-(hydroxymethyl)aminomethane

References

1. Young EJ. An overview of human brucellosis. *Clin Infect Dis*. 1995; 21:283–289. [PubMed: 8562733]
2. Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: a re-emerging zoonosis. *Vet Microbiol*. 2010; 140:392–398. [PubMed: 19604656]
3. Pappas G, Papadimitriou P. Challenges in *Brucella* bacteraemia. *Int J Antimicrob Agents*. 2007; 30(Suppl 1):S29–31. [PubMed: 17706402]
4. Cloeckaert A, Grayon M, Verger JM, Letesson JJ, Godfroid F. Conservation of seven genes involved in the biosynthesis of the lipopolysaccharide O-side chain in *Brucella* spp. *Res Microbiol*. 2000; 151:209–216. [PubMed: 10865948]
5. Cardoso PG, Macedo GC, Azevedo V, Oliveira SC. *Brucella* spp noncanonical LPS: structure, biosynthesis, and interaction with host immune system. *Microb Cell Fact*. 2006; 5:13. [PubMed: 16556309]
6. Godfroid F, Cloeckaert A, Taminiau B, Danese I, Tibor A, de Bolle X, Mertens P, Letesson JJ. Genetic organisation of the lipopolysaccharide O-antigen biosynthesis region of *Brucella melitensis* 16M (wbk). *Res Microbiol*. 2000; 151:655–668. [PubMed: 11081580]
7. Thoden JB, Goneau MF, Gilbert M, Holden HM. Structure of a sugar *N*-formyltransferase from *Campylobacter jejuni*. *Biochemistry*. 2013; 52:6114–6126. [PubMed: 23898784]
8. 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]
9. 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]
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. 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]
12. 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]
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. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*. 2006; 62:859–866. [PubMed: 16855301]
15. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr*. 1999; 55(Pt 4):849–861. [PubMed: 10089316]
16. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr*. 2000; 56(Pt 8):965–972. [PubMed: 10944333]
17. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]

18. 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]
19. 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]
20. Cook PD, Holden HM. GDP-perosamine synthase: structural analysis and production of a novel trideoxysugar. *Biochemistry*. 2008; 47:2833–2840. [PubMed: 18247575]
21. Godfroid F, Taminiau B, Danese I, Denoel P, Tibor A, Weynants V, Cloeckaert A, Godfroid J, Letesson JJ. Identification of the perosamine synthetase gene of *Brucella melitensis* 16M and involvement of lipopolysaccharide O side chain in *Brucella* survival in mice and in macrophages. *Infect Immun*. 1998; 66:5485–5493. [PubMed: 9784561]
22. Gatzeva-Topalova PZ, May AP, Sousa MC. Crystal structure and mechanism of the *Escherichia coli* ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry*. 2005; 44:5328–5338. [PubMed: 15807526]
23. Fromm, HJ. Initial Rate Enzyme Kinetics. Springer-Verlag; New York: 1975.
24. Gatzeva-Topalova PZ, May AP, Sousa MC. Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance. *Structure*. 2005; 13:929–942. [PubMed: 15939024]
25. Williams GJ, Breazeale SD, Raetz CR, Naismith JH. Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis. *J Biol Chem*. 2005; 280:23000–23008. [PubMed: 15809294]
26. Breazeale SD, Ribeiro AA, McClerren AL, Raetz CR. A formyltransferase required for polymyxin resistance in *Escherichia coli* and the modification of lipid A with 4-amino-4-deoxy-L-arabinose. Identification and function of UDP-4-deoxy-4-formamido-L-arabinose. *J Biol Chem*. 2005; 280:14154–14167. [PubMed: 15695810]
27. Woodford CR, Thoden JB, Holden HM. Molecular architecture of an *N*-formyltransferase from *Salmonella enterica* O60. *J Struct Biol*. 2017; available online 2 March 2017. doi: 10.1016/j.jsb.2017.03.002
28. 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]
29. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA: 2002.
30. Hensley P, Yang YR, Schachman HK. On the detection of homotropic effects in enzymes of low co-operativity. Application to modified aspartate transcarbamoylase. *J Mol Biol*. 1981; 152:131–152. [PubMed: 7040676]

Figure 1.

Overall structure of WbkC. Shown in (a) is a ribbon representation of the WbkC dimer. The position of the disulfide bridge that formed during the crystallization process is indicated in a space-filling representation. A stereo view of the observed electron densities corresponding to the bound ligands in Subunit 2 is displayed in (b). The density shown is from an omit map contoured at $\sim 2\sigma$. The map was calculated with coefficients of the form $(F_o - F_c)$, where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude. Note that the map was calculated before the ligands had been

included in the X-ray coordinate file, and thus there is no model bias. A stereo view of the WbkC subunit is presented in (c). This figure and Figures 3, 5, 7, and 8 were prepared with PyMOL.²⁹

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(a)

(b)

Figure 2.

Initial velocity kinetic data for the wild type WbkC reaction. Shown in (a) is the complete dataset fitted to Equation 1, as described in the text. The dotted purple line illustrates the fit to the model with no cooperativity ($n = 1$). The solid red line shows the best fit, which yielded a value of 1.5 for n . Shown in (b) is the limited dataset that was fitted to the Michaelis-Menten equation to estimate a value for $V_{\max}$.

Figure 3.

Close-up stereo view of the active site with bound ligands. The bound ligands are displayed in gray bonds. Distances within 3.2 Å are indicated by the dashed lines. Water molecules are represented by the red spheres.

Figure 4.

Initial velocity experiments with the F142A mutant variant. The experimental points were fitted to the modified Michaelis-Menten equation in which each concentration term was raised to an exponent, n . Inset: Data replotted according to Hensley *et al.*,³⁰ to emphasize the cooperativity. The line is drawn to aid the viewer.

Figure 5. Close-up stereo view of the ternary complex of WbkC. The bound GDP-perosamine and N^{10} -formyl-THF ligands are displayed in gray bonds. Possible hydrogen bonding interactions are indicated by the dashed lines. Water molecules are displayed as red spheres.

Figure 6.

Mass spectra. (a) The peak at $m/z = 587$ corresponds to GDP-perosamine. (b) The peak at $m/z = 615$ corresponds to the product of the WbkC reaction, namely GDP-*N*-formylperosamine. (c) The peak at $m/z = 571$ corresponds to the GDP-3-deoxyperosamine. (d) The peak at $m/z = 599$ corresponds to the formylated version of GDP-3-deoxyperosamine.

Figure 7. Comparison of quaternary structures. All the sugar *N*-formyltransferases whose structures have been determined thus far function as dimers, but the manner in which the subunits come together is decidedly different. Shown in (a), (b), and (c) are the ribbon representations for WlaRD, QdtF, and VioF, respectively.

Figure 8. Superposition of the ribbon drawings for the WbkC, WlaRD, and VioF subunits. The ribbon representations for the WbkC, WlaRD, and VioF subunits are colored in light blue, light green, and light purple, respectively.

Schema 1.

Schema 2.

Schema 3.

Table 1

X-ray Data Collection Statistics.

	WbkC D78A mutation	WbkC D78A/C47S mutations	WbkC D78A/F142A mutations	WbkC D78A/ F142A mutations (soaked in GDP- perosamine and <i>N</i> ¹⁰ -formyl-THF)
resolution limits	50–1.7 (1.8–1.7) ^b	50–2.2 (2.3–2.2)	50–2.2 (2.3–2.2)	50–2.2 (2.2–2.3)
number of independent reflections	57143 (8645)	27237 (3263)	54309 (5278)	27823 (3311)
completeness (%)	98.9 (97.0)	99.6 (98.4)	96.5 (80.6)	98.4 (96.3)
redundancy	4.2 (2.6)	14.1 (8.1)	6.3 (2.2)	8.0 (3.6)
avg <i>I</i> /avg σ (<i>I</i>)	11.8 (2.2)	12.3 (4.0)	7.1 (1.8)	10.0 (2.6)
<i>R</i> _{sym} (%) ^a	6.0 (30.8)	9.6 (40.5)	9.3 (41.2)	9.1 (40.1)

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

^b Statistics for the highest resolution bin.

Table 2

Refinement Statistics.

	WbkC D78A mutations	WbkC D78A/C47S mutations	WbkC D78A/F142A mutations	WbkC D78A/F142A mutations (soaked in GDP-perosamine and N ¹⁰ -formyl-THF)
space group	<i>P</i> ₄ ₃ ₂ ₁ ²	<i>P</i> ₄ ₃ ₂ ₁ ²	<i>P</i> ₂ ₁ ₂ ₁ ² ₁	<i>P</i> ₄ ₃ ₂ ₁ ²
unit cell dimensions (Å) <i>a</i> , <i>b</i> , <i>c</i>	65.9, 65.9, 233.1	65.9, 65.9, 234.7	67.3, 68.0, 236.0	66.8, 66.8, 235.9
resolution limits (Å)	50 – 1.7	50 – 2.2	50 – 2.2	50 – 2.2
^a <i>R</i> -factor (overall)%/no. reflections	18.7/57143	18.9/27237	20.5/54309	20.9/27823
<i>R</i> -factor (working)%/no. reflections	18.5/54336	18.5/25866	20.2/51622	20.6/26412
<i>R</i> -factor (free)%/no. reflections	23.4/2807	25.9/1371	26.8/2687	26.4/1411
number of protein atoms	3929	3924	7958	3850
number of heteroatoms	452	355	500	289
average B values				
protein atoms (Å ²)	21.5	28.7	29.0	21.7
ligands (Å ²)	29.4	41.0	42.7	33.8
solvent (Å ²)	28.7	29.1	23.3	18.7
weighted RMS deviations from ideality				
bond lengths (Å)	0.011	0.013	0.013	0.013
bond angles (°)	1.7	1.6	1.7	1.6
general planes (°)	0.009	0.007	0.007	0.007
Ramachandran regions (%)^b				
most favored	92.7	91.6	91.5	91.4
additionally allowed	7.1	8.2	8.4	8.1
generously allowed		0.2	0.1	0.5
disallowed	0.2	–	–	–

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

^bDistribution of the Ramachandran angles according to PROCHECK.²⁸