THE PROTEIN SOCIETY

Investigation of a sugar *N*-formyltransferase from the plant pathogen *Pantoea ananatis*

Daniel L. Hofmeister, James B. Thoden, and Hazel M. Holden*

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

Received 12 December 2018; Accepted 14 January 2019
DOI: 10.1002/pro.3577
Published online 21 January 2019 proteinscience.org

Abstract: ***Pantoea ananatis*** **is a Gram-negative bacterium first recognized in 1928 as the causative agent of pineapple rot in the Philippines. Since then various strains of the organism have been implicated in the devastation of agriculturally important crops. Some strains, however, have been shown to function as non-pathogenic plant growth promoting organisms. To date, the factors that determine pathogenicity or lack thereof between the various strains are not well understood. All *P. ananatis* strains contain lipopolysaccharides, which differ with respect to the identities of their associated sugars. Given our research interest on the presence of the unusual sugar, 4-formamido-4,6-dideoxy-D-glucose, found on the lipopolysaccharides of *Campylobacter jejuni* and *Francisella tularensis*, we were curious as to whether other bacteria have the appropriate biosynthetic machinery to produce these unique carbohydrates. Four enzymes are typically required for their biosynthesis: a thymidylyltransferase, a 4,6-dehydratase, an aminotransferase, and an *N*-formyltransferase. Here, we report that the gene SAMN03097714_1080 from the *P. ananatis* strain NFR11 does, indeed, encode for an *N*-formyltransferase, hereafter referred to as PA1080c. Our kinetic analysis demonstrates that PA1080c displays classical Michaelis–Menten kinetics with dTDP-4-amino-4,6-dideoxy-D-glucose as the substrate and N^{10}-formyltetrahydrofolate as the carbon source. In addition, the X-ray structure of PA1080c, determined to 1.7 Å resolution, shows that the enzyme adopts the molecular architecture observed for other sugar *N*-formyltransferases. Analysis of the *P. ananatis* NFR11 genome suggests that the three other enzymes necessary for *N*-formylated sugar biosynthesis are also present. Intriguingly, those strains of *P. ananatis* that are non-pathogenic apparently do not contain these genes.**

Keywords: bacterial sugar biosynthesis; 4-formamido-4,6-dideoxy-D-glucose; *N*-formyltransferase; lipopolysaccharide; O-antigen; *Pantoea ananatis*

Introduction

Pantoea ananatis is a Gram-negative yellow-pigmented bacterium first discovered in 1928 as a source of pineapple rot in the Philippines.[1] It is a ubiquitous organism that can survive under a myriad of environmental conditions including airplane fuel, and it is now considered an emerging plant pathogen.[2,3] Infections of such crops as maize, rice, melons, and onions with *P. ananatis* have

Abbreviations: dTDP, thymidine diphosphate; dTMP, thymidine monophosphate; HEPPS, *N*-2-hydroxyethylpiperazine-*N*-3-propanesulfonic acid; HPLC, high-performance liquid chromatography; TEV, tobacco etch virus; Tris, *tris*-(hydroxymethyl) aminomethane.

X-ray coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, N. J. (Accession no. 6NBP) and were released on December 19, 2018.

Broader statement: *Pantoea ananatis* is a Gram-negative bacterium that can infect such agriculturally important crops as maize and rice. Various strains of *P. ananatis* contain lipopolysaccharides, which are complex glycoconjugates that often contain unusual dideoxysugars. The research presented here demonstrates that *P. ananatis* strain NFR11 has the required enzymes to produce an *N*-formylated sugar that may ultimately be important for the virulence of this emerging plant pathogen.

Grant sponsor: National Institutes of Health GM115921; Grant sponsor: NIH GM115921.

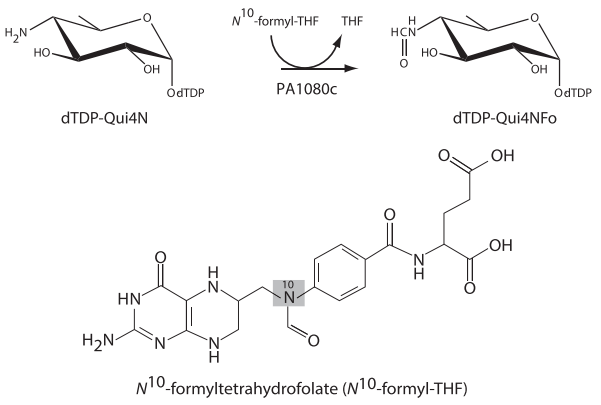
**Correspondence to*: Hazel M. Holden, Department of Biochemistry, University of Wisconsin, Madison, WI 53706. E-mail: hazel_holden@biochem.wisc.edu

led to devastating economic consequences. Indeed, the outbreak of Sweet Vidalia onion rot in Georgia in 1997 resulted in significant agricultural damage, with some farms suffering nearly 100% loss.[4] In South Africa, the bacterium has been linked to the widespread destruction of *Eucalyptus* trees.[5] Unlike typical plant pathogens, some strains of *P. ananatis* are also capable of infecting humans and insects.[3] Whereas most studies of this bacterium have focused on its role in plant disease, there are strains of *P. ananatis* that have been shown to promote crop growth.[3] The factors that allow *P. ananatis* to adapt and survive in such diverse natural habitats are unclear. Equally important, the biochemical elements required for the pathogenicity or lack thereof of specific strains are not completely understood.

As is typical for Gram-negative bacteria, *P. ananatis* contains a lipopolysaccharide or LPS, a complex glycoconjugate that extends outwards from the organism's outer cell membrane. The LPS consists of three components: the lipid A portion that links the glycoconjugate to the bacterial cell membrane, the core polysaccharide, and the O-antigen, which often contains unusual di- and trideoxysugars. It is the O-antigen that contributes to the wide species variation observed in nature.[6] Thus far, the structure of the O-antigen for one strain of *P. ananatis* (AEP17) has been determined via NMR spectroscopy and shown to contain an unusual D-galacturonamide moiety.[7]

Given our long-standing interest in unusual sugar biosynthesis, and in particular on *N*-formylated sugars, we utilized a simple bioinformatics analysis to determine whether any strains of *P. ananatis* contained the genes required for the production of such carbohydrates. Briefly, we performed a BLAST® search to discover gene sequences that were similar to those encoding enzymes previously investigated in the laboratory.[8–15] The vast majorities of "hits" were annotated as L-methionyl-tRNA *N*-formyltransferases. For those open-reading frames, whereby the complete bacterial genome sequences were known, the surrounding DNA sequences were subsequently analyzed via BLAST® to determine if the additional genes required for *N*-formylated sugar biosynthesis were also present, namely those encoding a thymidylyltransferase, a 4,6-dehydratase, and a pyridoxal 5′-phosphate-dependent aminotransferase.

A perusal of the *P. ananatis* strain NFR11 genome highlighted such a protein of interest that was encoded by SAMN03097714_1080 and originally annotated as an L-methionyl-tRNA *N*-formyltransferase. This particular strain of *P. ananatis* was isolated in the United States from switchgrass.[16] Here, we report the three-dimensional structure of the protein encoded by SAMN03097714_1080 and hereafter referred to as PA1080c. On the basis of both structural and biochemical data, we suggest that PA1080c is, in fact, a sugar *N*-formyltransferase that converts dTDP-4-amino-4,6-dideoxy-D-glucose (dTDP-Qui4N) to dTDP-4-formamido-4,6-dideoxy-D-glucose (dTDP-Qui4NFo) and employs N^{10}-formyltetrahydrofolate (N^{10}-formyl-THF) as the carbon source (Scheme 1). Our analysis provides new molecular details into this ubiquitous yet unconventional plant pathogen.



Scheme 1. Reaction catalyzed by PA1080c and its required cofactor.

Results and Discussion

Kinetic properties

As described in the Materials and Methods section, the gene encoding PA1080c was chemically synthesized and utilized to produce recombinant protein. Given its amino acid sequence similarity and homology to Rv3404c, a sugar *N*-formyltransferase from *Mycobacterium tuberculosis* (39% and 48%, respectively), we anticipated that PA1080c would function on dTDP-Qui4N and would employ N^{10}-formyl-THF as the carbon source.[15] To test this hypothesis, a simple activity assay was conducted overnight with a mixture of PA1080c, dTDP-Qui4N, and N^{10}-formyl-THF. The structures of the substrate and cofactor are provided in Scheme 1. Shown in Figure 1 are HPLC traces of the reaction mixture at time zero (blue) and time 12 h (red). Peaks 1, 2, and 4 correspond to dTDP-Qui4N, dTMP (a contaminant), and N^{10}-formyl-THF/THF, respectively. After 12 h a new peak (Peak 3) was observed with a retention volume of 13 mL. The ligand corresponding to this peak was purified, and its identity confirmed to be dTDP-Qui4NFo via electrospray ionization mass spectrometry in the negative ion mode (Fig. 2).

A detailed kinetic analysis was subsequently conducted via a discontinuous HPLC assay. A plot of dTDP-Qui4N concentration versus reaction rate is presented in Figure 3. PA1080c demonstrates classical Michaelis–Menten kinetics. The K_M for dTDP-Qui4N is 0.29 ± 0.05 m*M* and the k_{cat} is 0.84 ± 0.11 s^{-1}. The overall catalytic efficiency or k_{cat}/K_M for PA1080c is 2.9×10^3 (± 400)

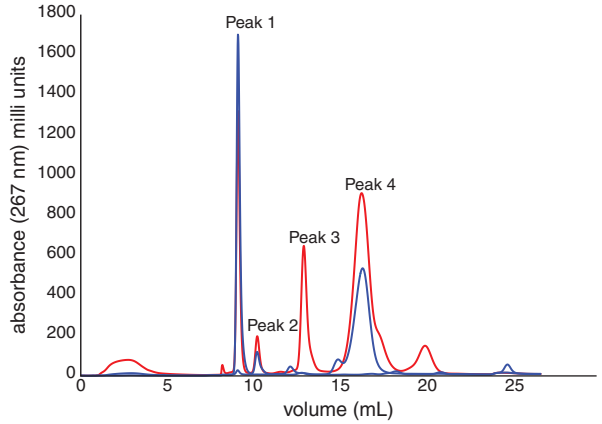


Figure 1. HPLC activity assay of PA1080c. To test the activity of PA1080c, a mixture of the enzyme with dTDP-Qui4N and N^{10}-formyl-THF was incubated overnight. Shown are the HPLC traces of the reaction mixture at time zero (blue) and time 12 h (red). Peaks 1, 2, and 4 correspond to dTDP-Qui4N, dTMP (a contaminant), and N^{10}-formyl-THF/THF, respectively. After 12 h a new peak (Peak 3) was observed with a retention volume of 13 mL.

M^{-1} s^{-1}, which is comparable to that observed for other sugar *N*-formyltransferases.[8–10,13–15,17]

Overall three-dimensional structure of PA1080c

The crystals of PA1080c utilized in this investigation were grown from poly(ethylene glycol) 3350 at pH 8.0 and in the presence of 5 m*M* N^5-formyl-THF (a stable analog of N^{10}-formyl-THF) and 5 m*M* dTDP-Qui4N. They belonged to the space group $P3_121$ with one polypeptide chain in the asymmetric unit. The model was refined to an overall *R*-factor of 18.8% at 1.7 Å resolution. Shown in Figure 4(A) is the observed electron density corresponding to the bound ligands. As can be seen, the electron density for the dTDP-sugar is unambiguous whereas that for the cofactor is weak due to low occupancy. Experiments to achieve higher occupancy for the cofactor were unsuccessful. These included soaking the crystals for days in either 5 m*M* N^5-formyl-THF or N^{10}-formyl-THF.

A ribbon representation of the monomer is presented in Figure 4(B). The structure is dominated by a seven stranded mixed β-sheet flanked on either side by α-helices. There is an additional β-hairpin motif

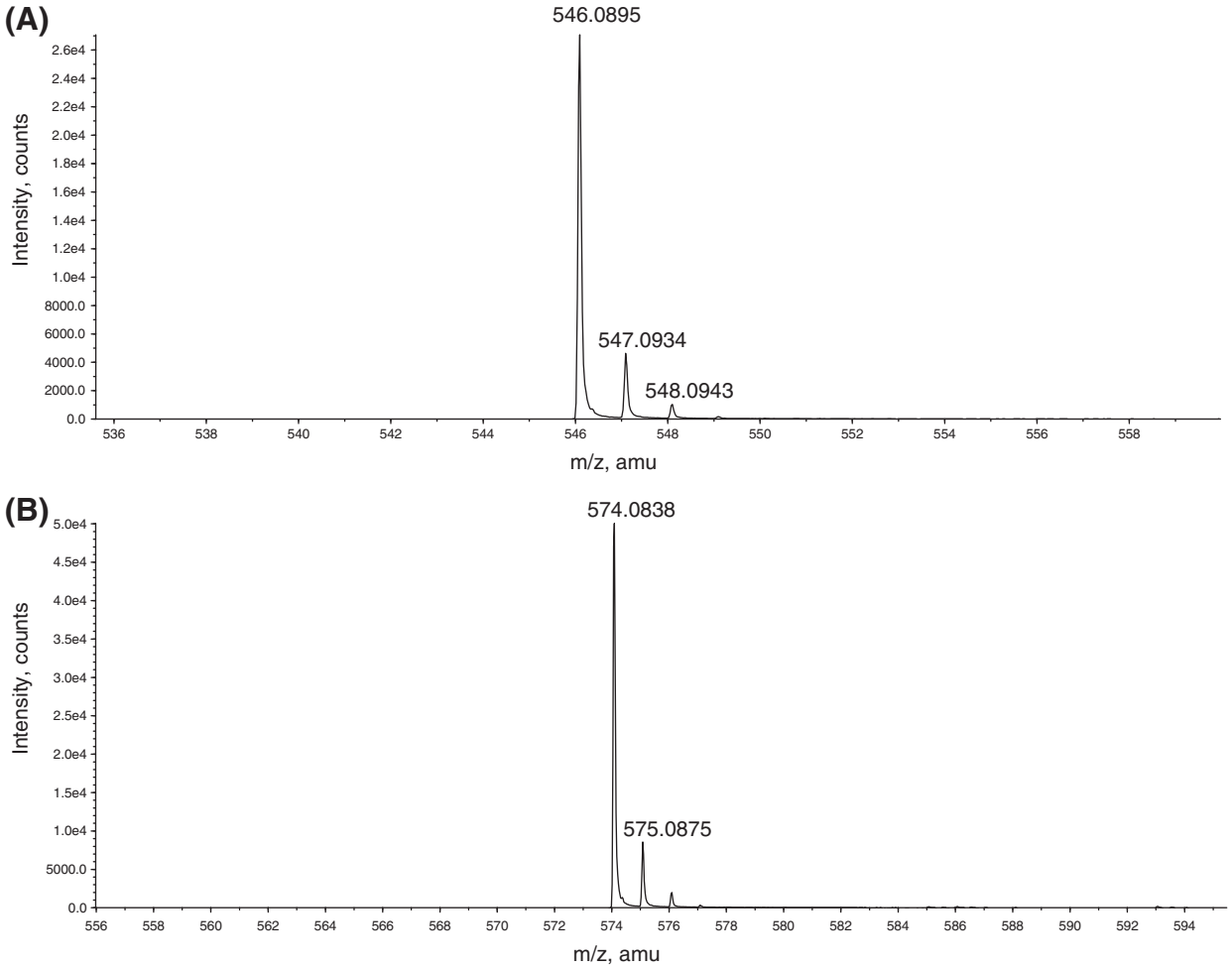


Figure 2. Electrospray ionization mass spectrometry before and after assay incubation. Shown in (A) is the mass spectrum for dTDP-Qui4N, which corresponds to Peak 1 in Figure 1. The mass spectrum for the ligand associated with Peak 3 (Fig. 1) that formed after incubation is presented in (B), and it corresponds to dTDP-Qui4NFo.

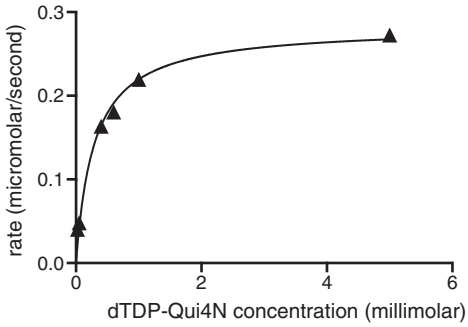


Figure 3. Initial velocity kinetic data for PA1080c. The concentration of N^{10}-formyl-THF was held constant at 5 m*M*, whereas the dTDP-Qui4N concentrations ranged from 0.01 m*M* to 5.0 m*M*. In presenting the data as we do, we are adhering to standard conventions in enzymology. Measuring velocities 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 the Materials and Methods section.

delineated by Ala 240 to the C-terminus. The first 12 residues at the N-terminus were disordered. In addition, there was a break between Ala 143 and Asp 146. This region connects β-strands six and seven [Fig. 4(B)].

Examination of the crystalline packing arrangement shows that PA1080c forms the dimeric quaternary structure observed for other sugar *N*-formyltransferases that function on dTDP-Qui4N.[9,11,15] The total buried surface area for the dimer, displayed as a ribbon representation in Figure 4(C), is ~3600 Å^2.

Active site architecture

A close-up view of the PA1080c active site is displayed in Figure 5. The thymine ring of dTDP-Qui4N, surrounded by the aromatic side chains of Tyr 122 and Phe 237, is hydrogen bonded into the active site by Asn 239 and three water molecules. The imidazole ring of His 234 abuts one side of the dTDP-Qui4N ribosyl moiety, which is further anchored to the protein via a hydrogen bond between its 3-hydroxyl group and the side chain of Gln 124. The side chains of His 92, Lys 94, and Tyr 168 interact with the pyrophosphoryl group of the dTDP-Qui4N ligand. Finally, the pyranosyl group of dTDP-Qui4N lies within hydrogen bonding distance to the backbone carbonyl of Gly 120, the backbone amide of Tyr 122, and four water molecules.

There are three amino acid residues that appear to be strictly conserved amongst the *N*-formyltransferases.[19] In PA1080c, these correspond to Asn 109, His 111, and Asp 146 (Fig. 5). It is thought that the conserved histidine serves as a catalytic base to remove a proton from the sugar amino group as it attacks the carbonyl carbon of N^{10}-formyl-THF. In keeping with this hypothesis, $N^{\delta 1}$ of His 111 lies within 3.9 Å of the pyranosyl C-4′ amino group. In other sugar *N*-formyltransferases, the side chains of the conserved aspartate and asparagine lie within 3 Å of one another. As can be seen in Figure 5, however, Asp 146 is positioned outside of the active site (it is preceded by the break in the electron density between Ala 143 and Asp 146). This movement is most likely due to the low occupancy of the N^5-formyl-THF ligand, as discussed in more detail below.

Comparison with M. tuberculosis *Rv3404c*

The first model of a sugar *N*-formyltransferase that functions on dTDP-Qui4N was described in 2014.[9] Unfortunately, all attempts to prepare a ternary complex of the enzyme from *Francisella tularensis* with dTDP-Qui4N and a tetrahydrofolate-based cofactor were unsuccessful. Subsequently, the structure of a similar enzyme from *Providencia alcalifaciens* O30 was reported in a ternary complex but the conserved histidine was located at ~7 Å from the sugar amino group, suggesting that the dTDP-Qui4N ligand had bound in a non-productive conformation due to the crystallization conditions.[11] In 2017, the structure of Rv3404c from *M. tuberculosis* was determined, and the ternary complex model demonstrated that the conserved histidine was appropriately positioned to interact with the sugar amino group.[15]

A superposition of the ribbon representations for Rv3404c and PA1080c is presented in Figure 6(A). The alpha carbons for the two proteins superimpose with a root-mean-square deviation of 1.5 Å. The dTDP-sugar ligands adopt virtually identical conformations within the active site regions of these enzymes. The PA1080c protein is 26 amino acid residues longer than the Rv3404c enzyme, and this difference primarily results in a larger surface loop defined by Lys 39 to Phe 46 and a longer N-terminal region [Fig. 6(A)].

An amino acid sequence alignment for the four *N*-formyltransferases that have thus far been shown to function on dTDP-Qui4N is provided in Figure 6(B). Asn 109, His 111, and Asp 146 are strictly conserved not only amongst the sugar *N*-formyltransferases but also amongst the L-methionyl tRNA *N*-formyltransferases and the glycinamide ribonucleotide transformylases.[20,21] There are three residues that are only conserved amongst the *N*-formyltransferases that function on dTDP-Qui4N: His 92, Lys 94, and Asn 239. The first two are intimately involved in positioning the pyrophosphoryl moiety of the dTDP-sugar ligand into the active site whereas the third hydrogen bonds to the substrate's thymine ring [Fig. 6(C)]. Given that these residues are not conserved in those *N*-formyltransferases that function on dTDP-Qui3N, their existence in an "open-reading frame" sequence may aid in more appropriate annotations.[8,10,13] Likewise, unique to the sugar

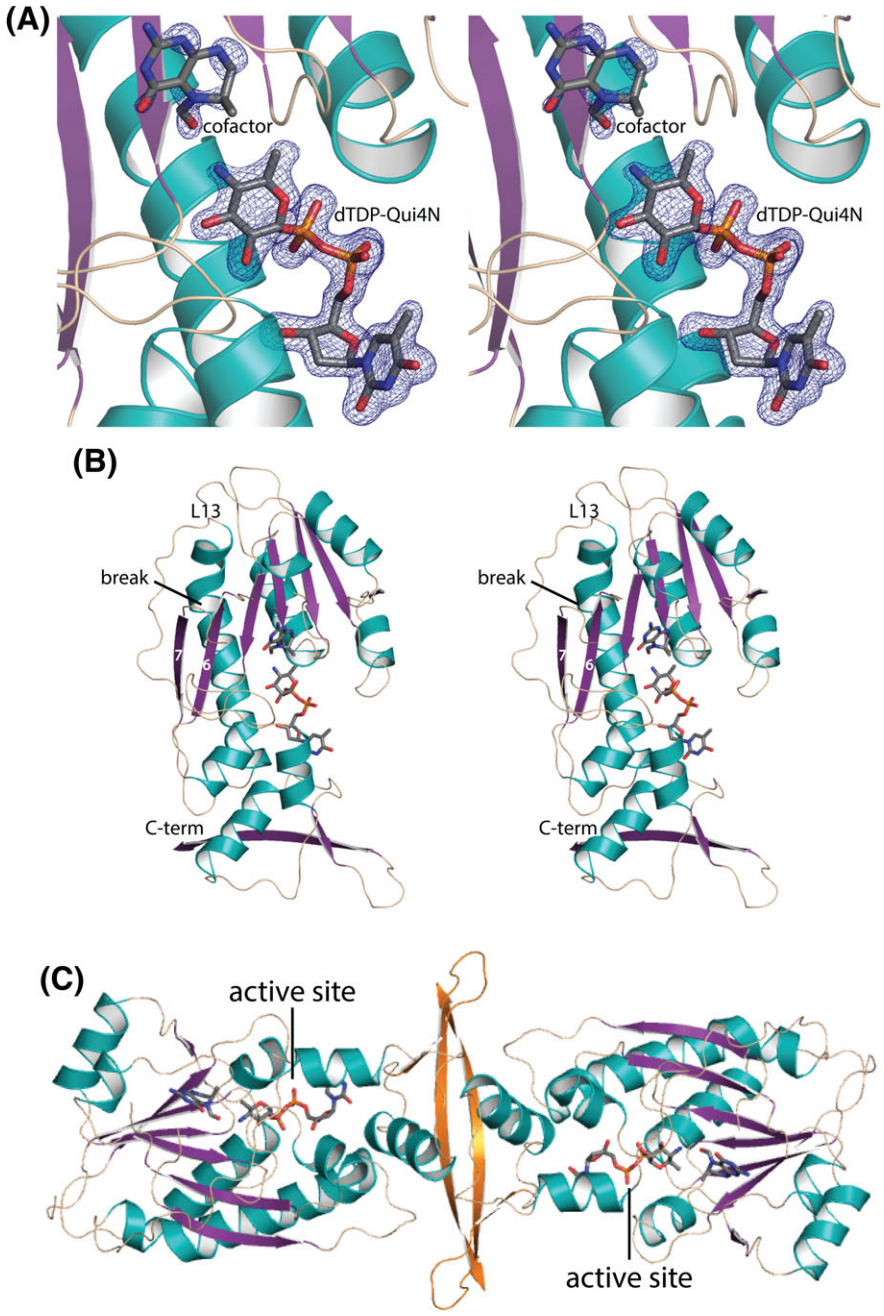


Figure 4. Overall structure of PA1080c. Shown in stereo in (A) is the observed electron density for the dTDP-sugar ligand and the tetrahydrofolate-based cofactor. The omit map was calculated with coefficients of the form F_o–F_c where F_o and F_c were the native and calculated structure factor amplitudes, respectively. The map was contoured at 3σ. The *B*-factors for the dTDP-sugar and the cofactor were 22.4 and 49.9 Å^2, respectively. A stereo ribbon representation of the monomer in the asymmetric unit is presented in (B). The α-helices and β-strands are colored in teal and purple, respectively. The quaternary structure of PA1080c is presented in (C) with the subunit:subunit interface highlighted in orange. This figure and Figures 5 and 6 were prepared with the software package PyMOL.[18]

N-formyltransferases described here is the characteristic signature sequence defined by Pro 112 to Tyr 122 (PA1080c numbering). This region, which connects β-strand 5 to α-helix 5, surrounds both the cofactor and the dTDP-sugar as can be seen in Figure 6(C). The aromatic side chain of Trp 121 plays a key role in binding the pyranosyl ring of the substrate by participating in CH/π interactions. These types of interactions were first suggested to be important determinants in carbohydrate/protein recognition well over 40 years ago and since have been investigated and extensively reviewed.[22–24]

Biochemical context

In 2012, the structure of the O-antigen from *P. alcalifaciens* O30 was shown to contain Qui4NFo, and the pathway for its biosynthesis was proposed, as outlined in Scheme 2.[17] In addition to the

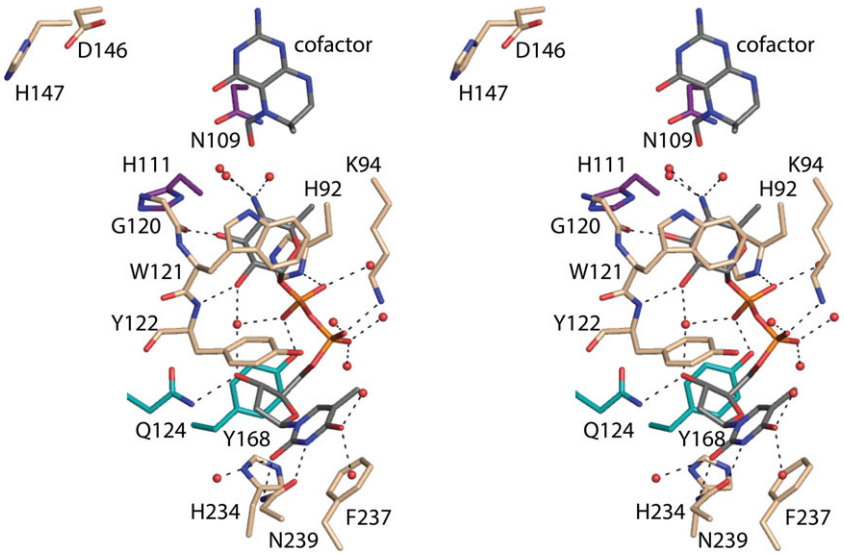


Figure 5. Stereo view of the PA1080c active site. Those residues that lie within ~3.2 Å of the dTDP-sugar and the tetrahydrofolate-based cofactor are displayed. The dashed lines indicate possible hydrogen bonding interactions. Ordered water molecules are depicted as red spheres. The coloring scheme is the same as described in Figure 4.

N-formyltransferase, three other enzymes are required: a glucose-1-phosphate thymidylyltransferase, a 4,6-dehydratase, and a pyridoxal 5′-phosphate-dependent aminotransferase. We have demonstrated in this investigation that PA1080c functions as a sugar *N*-formyltransferase. Given that the structure of the O-antigen of the *P. ananatis* strain studied here is not known, however, the question then is whether the organism has the other required enzymes to produce *N*-formylated sugars. DNA sequence analysis of the *P. ananatis* NFR11 genome, and in consideration of operon context, suggests that the bacterium has, indeed, the required enzymatic activities. Listed in Table I are the genes encoding these hypothetical proteins and their amino acid sequence similarities and homologies to known enzymes. The putative *P. ananatis* thymidylyltransferase and the 4,6-dehydratase are remarkably similar to previously studied enzymes. Whereas the presumed aminotransferase shows less amino acid sequence similarity to known protein structures, the gene encoding it lies within the same operon as that for PA1080c.

One important question that needs to be addressed is whether the presence of an *N*-formylated sugar on the O-antigen of a bacterium plays a role in virulence. Given that the existence of an *N*-formylated sugar on the O-antigen of *Pseudomonas aeruginosa* was first reported in 1985, it is somewhat surprising that there is still a paucity of data regarding the biological roles of these carbohydrates.[28] Interestingly, the loss of an *N*-formyltransferase in *Brucella melitensis*, a zoonotic Gram-negative bacterium that infects sheep and causes brucellosis in humans, results in a bacterial strain with attenuated pathogenicity.[29]

Pantoea ananatis is a fascinating plant pathogen in that certain strains cause disease whereas others promote plant growth.[3] Currently completed or draft genomes of various *P. ananatis* strains are listed in Table II. Strikingly, in those non-pathogenic organisms, the genes encoding the enzymes required for *N*-formylated sugar biosynthesis are apparently absent according to our simple BLAST® analysis. Whether this trend continues as more genomes are sequenced remains to be determined.

In summary, we have shown that PA1080c functions as an *N*-formyltransferase and that most likely at some point in the life cycle of *P. ananatis* strain NFR11 *N*-formylated sugars are produced. Additional biological investigations will be required to define the location of these sugars. Indeed, it is not clear whether they are found on the O-antigen or perhaps on exopolysaccharides. In addition, by comparing the structures and amino acid sequences of four *N*-formyltransferases that function on dTDP-Qui4N, several signature sequences have been identified that will aid in more accurate annotations of protein functions based on amino acid sequences.

Materials and Methods

Cloning, expression, and purification

The gene encoding SAMN03097714_1080 (PA1080c) from *P. ananatis* strain NFR11 was synthesized by Integrated DNA Technologies and placed into pET28T3G, a laboratory pET28b(+) vector that had been previously modified to incorporate a TEV protease cleavage recognition site after the N-terminal polyhistidine tag.[30]

The pET28T3G-*PA1080c* plasmid was utilized to transform Rosetta2(DE3) *Escherichia coli* cells

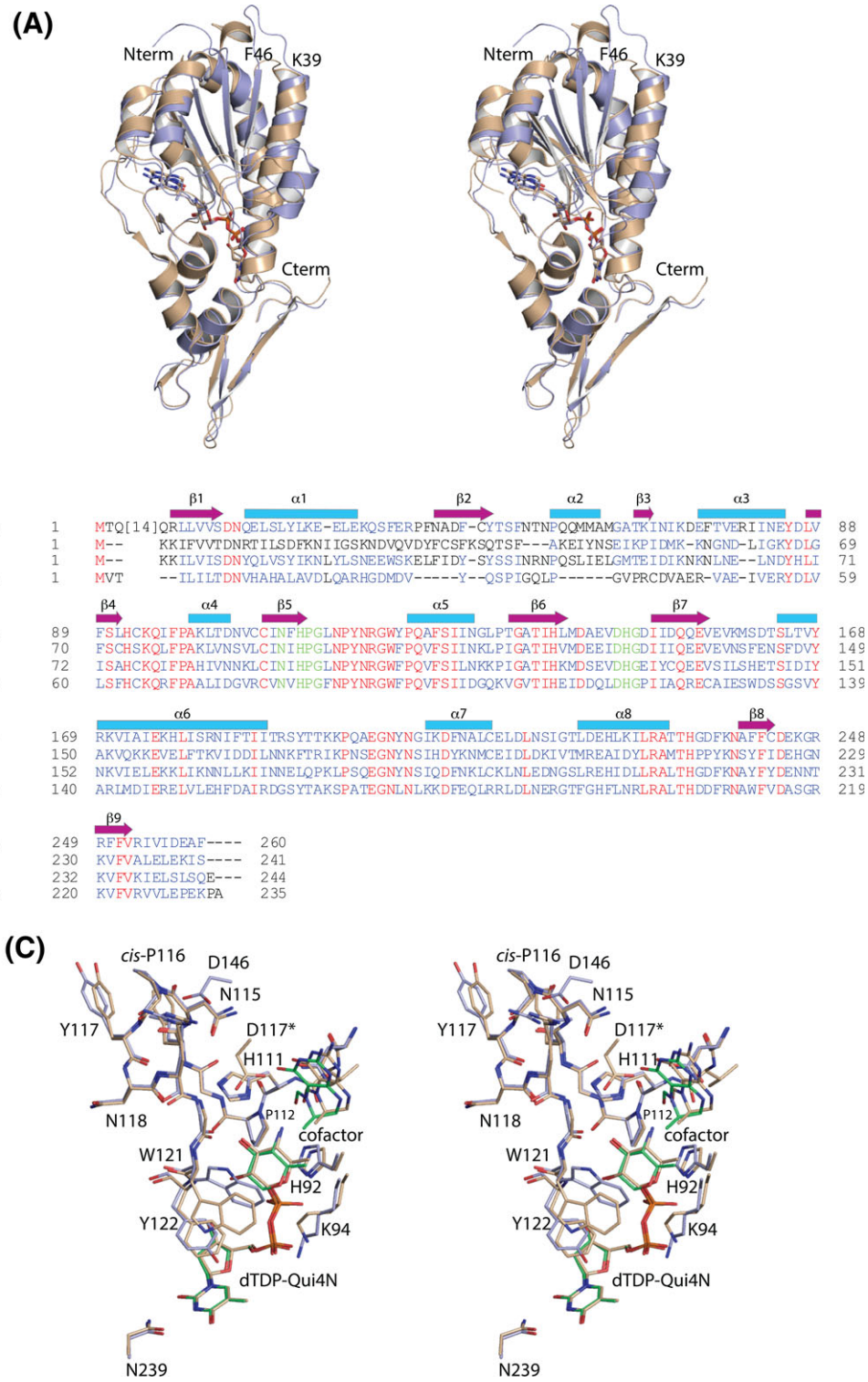
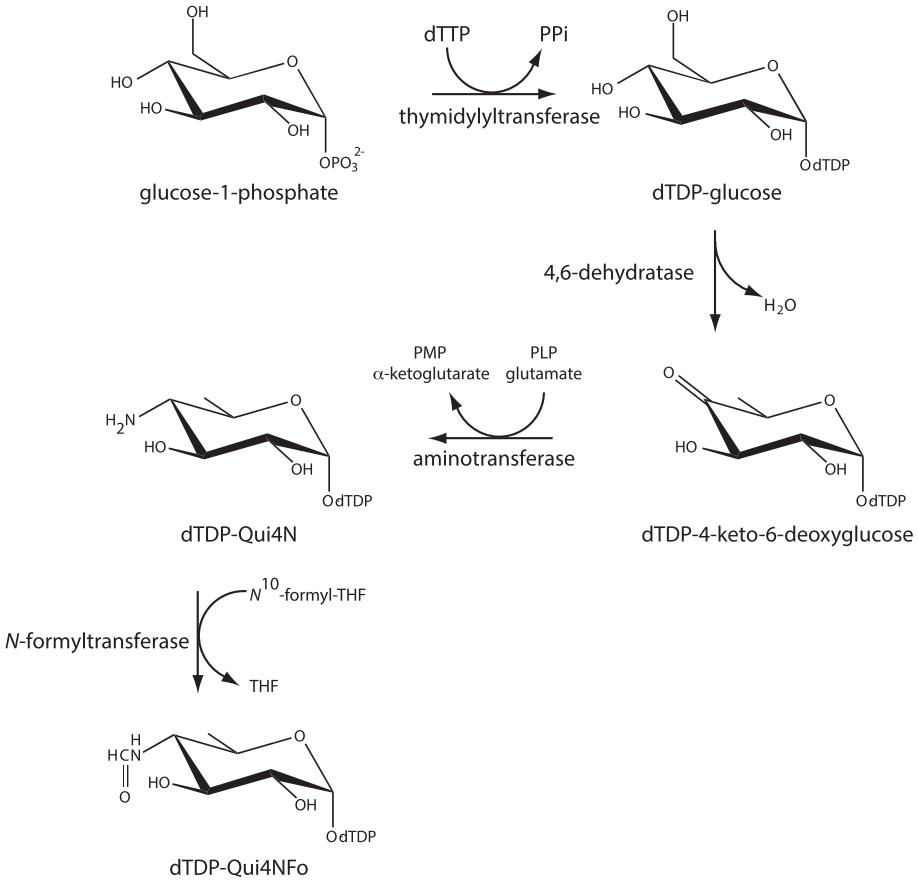


Figure 6. Comparison of PA1080c with other *N*-formyltransferases. A superposition of the ribbon drawings for PA1080c and *M. tuberculosis* Rv3404c, colored in blue and wheat, respectively, is shown in stereo in (A). An amino acid sequence alignment of known *N*-formyltransferases that function on dTDP-Qui4N is provided in (B). The positions of the β-strands and α-helices are indicated by the magenta arrows and the blue rectangles, respectively. A superposition of the active sites for PA1080c and *M. tuberculosis* Rv3404c, displayed in blue and wheat, respectively, is presented in (C). The labeled residues refer to those in PA1080c (with the exception of D117*, which corresponds to the residue found in Rv3404c).

(Novagen). The cultures were grown in lysogeny broth supplemented with kanamycin (50 mg/L) and chloramphenicol (50 mg/L) 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 protein expression was initiated with the addition of 1 m*M* isopropyl β-D-1-thiogalactopyranoside. The cells were then allowed to grow at 22°C for 18 h.



Scheme 2. Predicted pathway for the production of dTDP-Qui4NFo.

Table I. *Genes Encoding Proteins Required for dTDP-Qui4N Biosynthesis*

Gene identifier	Proposed function	Closest relative of known structure/function	Amino acid sequence identity/similarity (%)	PDB code and reference
SAMN03097714_1322	Glucose-1-phosphate thymidylyltransferase	G1P-TT or RmlA *Escherichia coli*	89/93	1H5T[25]
SAMN03097714_1324	4,6-Dehydratase	RmlB *Salmonella enterica*	82/88	1G1A[26]
SAMN03097714_1082	Aminotransferase	DesI *Streptomyces venezuelae*	34/52	2PO3[27]

The cells were harvested by centrifugation and disrupted by sonication on ice. The lysate was cleared by centrifugation, and PA1080c was purified with Prometheus™ Ni-NTA agarose (Prometheus Protein Biology Products) according to the manufacturer's instructions. The polyhistidine tag was removed by

Table II. Pantoea ananatis Strains *(Reported in Weller-Stuart et al.[3])*

Strain	Status	Interest	Source	Presence of gene encoding *N*-formyltransferase
P. ananatis AJ13355	Complete	Biotechnology	Soil	No
P. ananatis AMG521	Draft	Plant growth promoter	Rice paddy	No
P. ananatis B1-9	Draft	Plant growth promoter	Rhizosphere of onion root	No
P. ananatis CFH 7–1	Draft	Plant pathogen	Cotton ball	Yes
P. ananatis LMG 5342	Complete	clinical isolate	human wound	Yes
P. ananatis NS296	Draft	Endophyte	Rice seed	Yes
P. ananatis NS303	Draft	Endophyte	Rice seed	Yes
P. ananatis NS311	Draft	Endophyte	Rice seed	Yes
P. ananatis PA4	Draft	Plant pathogen	Maize	Yes
P. ananatis PaMB1	Draft	Plant pathogen	Maize	Yes
P. ananatis RSA47	Draft	Endophyte	Rice seed	Yes

digestion with TEV protease. The TEV protease and remaining tagged PA1080c were removed by passage over Ni-NTA agarose, and the tag-free protein was dialyzed against 10 m*M* Tris–HCl (pH 8.0) and 200 m*M* NaCl and concentrated to 19 mg/mL based on an extinction coefficient of 1.64 $(mg/mL)^{-1}$ cm^{-1}.

Activity assay

A reaction mixture containing 50 m*M* HEPPS (pH 8.0), 1 m*M* dTDP-Qui4N, 2 m*M* N^{10}-formyl-THF, and 0.5 mg/mL PA1080c was allowed to incubate at room temperature overnight. The enzyme was then removed via filtration with a 10 kDa cutoff filter, and the mixture was diluted 10× with water and examined by HPLC. A 0–1 M gradient of ammonium acetate (pH 4.0) was utilized with a 1 mL Resource-Q column. Evaluation of the reaction products was performed via electrospray ionization mass spectrometry in the negative ion mode. The required dTDP-Qui4N was prepared as previously described.[9]

Kinetic analyses

Kinetic parameters for PA1080c were determined via a discontinuous assay using an ÄKTA Purifier HPLC system. Reaction rates were determined by measuring the amount of *N*-formylated product formed on the basis of peak area as monitored at 267 nm. The concentration was determined from the peak area via a calibration curve with standard samples that had been treated in the same manner as the reaction time points. The 1.5 mL reaction mixtures contained 5 m*M* N^{10}-formyl-THF, 50 m*M* HEPPS (pH 8.5), 0.01 mg/mL PA1080c, and dTDP-Qui4N concentrations ranging from 0.01 m*M* to 5.0 m*M*. Five 250 μL aliquots were taken over 2.5 min, and the reaction aliquots quenched by the addition of 12 μL of 6 M HCl. Following the addition of 200 μL of carbon tetrachloride and vigorous mixing, the samples were spun at 14,000*g* for 2 min, and 200 μL of the aqueous phase removed for HPLC analysis. The samples were diluted with 2 mL water and loaded onto a 1 mL Resource-Q column. Products were quantified after elution with an 8-column volume gradient from 0 to 400 m*M* LiCl (pH 4.0, HCl). A plot of initial velocity versus concentration was analyzed using PRISM (GraphPad Software, Inc.) and fitted to the equation $v_o = (V_{max}[S])/(K_M + [S])$.

Table III. *X-ray Data Collection and Model Refinement Statistics*

Resolution limits (Å)	50–1.7 (1.8–1.7)[a]
Number of independent reflections	32,264 (4954)
Completeness (%)	98.3 (94.1)
Redundancy	7.6 (3.7)
Avg I/Avg σ(I)	13.1 (2.1)
R_{sym} (%)[b]	6.1 (43.3)
[c]*R*-factor (overall)%/no. reflections	18.8/32,264
R-factor (working)%/no. reflections	18.6/30,727
R-factor (free)%/no. reflections	22.6/1537
number of protein atoms	1995
number of heteroatoms	307
Average *B* values	
Protein atoms (Å^2)	27.0
Ligand (Å^2)	30.7
Solvent (Å^2)	36.4
Weighted RMS deviations from ideality	
Bond lengths (Å)	0.011
Bond angles (°)	1.6
Planar groups (Å)	0.009
Ramachandran regions (%)[d]	
Most favored	97.9
Additionally allowed	2.1
Generously allowed	0.0

[a]Statistics for the highest resolution bin.
[b]$R_{sym} = (\sum|I-\bar{I}|/\sum I)\times 100$.
[c]*R*-factor = $(\Sigma|Fo-Fc|/\Sigma|Fo|)\times 100$ where *F*o is the observed structure-factor amplitude and *F*c is the calculated structure-factor amplitude.
[d]Distribution of Ramachandran angles according to PROCHECK.[35]

Crystallization and structural analysis

Crystallization conditions for PA1080c were surveyed at both room temperature and 4°C by the hanging drop method of vapor diffusion using a laboratory-based sparse matrix screen. X-ray diffraction quality crystals of the enzyme were subsequently grown from precipitant solutions containing 13–16% poly(ethylene glycol) 3350, 5 m*M* N^5-formyl-THF, 5 m*M* dTDP-Qui4N, 200 m*M* NaCl, and 100 m*M* HEPPS (pH 8.0) at 21°C. The crystals belonged to the space group $P3_121$ with unit cell dimensions of $a = b = 75.9$ Å and $c = 87.6$ Å. The asymmetric unit contained one monomer.

For X-ray data collection, the crystals were transferred to a cryoprotectant solution containing 15% poly(ethylene glycol) 3350, 400 m*M* NaCl, 5 N^5-formyl-THF, 17% ethylene glycol, 5 m*M* dTDP-Qui4N, and 100 m*M* HEPPS (pH 8.0).

An X-ray data set was collected in-house with a Bruker AXS Platinum-135 CCD detector controlled with the PROTEUM software package (Bruker AXS). The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics. The X-ray data were processed with SAINT and scaled with SADABS (Bruker AXS). The structure was solved by molecular replacement with PHASER[31] using as a search model the coordinates of PDB entry 4YFV.[11] Iterative cycles of model building with COOT and refinement with REFMAC reduced the R_{work} and R_{free} to 18.6% and 22.6%, respectively, from 50 to 1.7-Å resolution.[32–34]

All relevant X-ray data collection and model refinement statistics are presented in Table III.

Acknowledgments

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

Conflict of Interest

The authors have no competing financial interests.

References

1. Serrano FB (1928) Bacterial fruitlet brown-rot of pineapple in the Philippines. Philippine J Sci 36:271–324.
2. Coutinho TA, Venter SN (2009) *Pantoea ananatis*: an unconventional plant pathogen. Mol Plant Pathol 10:325–335.
3. Weller-Stuart T, De Maayer P, Coutinho T (2017) *Pantoea ananatis*: genomic insights into a versatile pathogen. Mol Plant Pathol 18:1191–1198.
4. Walcott RR, Gitaitis RD, Castro AC, Sanders FHJ, Diaz-Perez JC (2002) Natural infestation of onion seed by *Pantoea ananatis*, causal agent of center rot. Plant Dis 86:106–111.
5. Coutinho TA, Preisig O, Mergaert J, Cnockaert MC, Riedel K-H, Swings J, Wingfield MJ (2002) Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. Plant Dis 86:20–25.
6. Madigan MT, Martinko JM. *Biology of microorganisms*. Upper Saddle River, NJ: Prentice Hall, Pearson, 2006.
7. Contreras Sanchez-Matamoros R, Gil Serrano AM, Tejero-Mateo P, Ollero J, Megias Saavedra E, Rodriguez-Carvajal MA (2013) Structure of the O-antigen of the lipopolysaccharide isolated from *Pantoea ananatis* AEP17, a rhizobacterium associated with rice. Carbohydr Res 369: 25–30.
8. Thoden JB, Goneau MF, Gilbert M, Holden HM (2013) Structure of a sugar *N*-formyltransferase from *Campylobacter jejuni*. Biochemistry 52:6114–6126.
9. Zimmer AL, Thoden JB, Holden HM (2014) Three-dimensional structure of a sugar *N*-formyltransferase from *Francisella tularensis*. Protein Sci 23:273–283.
10. Woodford CR, Thoden JB, Holden HM (2015) New role for the ankyrin repeat revealed by a study of the *N*-formyltransferase from *Providencia alcalifaciens*. Biochemistry 54:631–638.
11. Genthe NA, Thoden JB, Benning MM, Holden HM (2015) Molecular structure of an *N*-formyltransferase from *Providencia alcalifaciens* O30. Protein Sci 24:976–986.
12. Genthe NA, Thoden JB, Holden HM (2016) Structure of the *Escherichia coli* ArnA *N*-formyltransferase domain in complex with N(5)-formyltetrahydrofolate and UDP-Ara4N. Protein Sci 25:1555–1562.
13. Woodford CR, Thoden JB, Holden HM (2017) Molecular architecture of an *N*-formyltransferase from *Salmonella enterica* O60. J Struct Biol 200:267–278.
14. Riegert AS, Chantigian DP, Thoden JB, Tipton PA, Holden HM (2017) Biochemical characterization of WbkC, an *N*-formyltransferase from *Brucella melitensis*. Biochemistry 56:3657–3668.
15. Dunsirn MM, Thoden JB, Gilbert M, Holden HM (2017) Biochemical investigation of Rv3404c from *Mycobacterium tuberculosis*. Biochemistry 56:3818–3825.
16. De Maayer P, Aliyu H, Vikram S, Blom J, Duffy B, Cowan DA, Smits THM, Venter SN, Coutinho TA (2017) Phylogenomic, pan-genomic, pathogenomic and evolutionary genomic insights into the agronomically relevant enterobacteria *Pantoea ananatis* and *Pantoea stewartii*. Front Microbiol 8:1–14.
17. Liu B, Chen M, Perepelov AV, Liu J, Ovchinnikova OG, Zhou D, Feng L, Rozalski A, Knirel YA, Wang L (2012) Genetic analysis of the O-antigen of *Providencia alcalifaciens* O30 and biochemical characterization of a formyltransferase involved in the synthesis of a Qui4N derivative. Glycobiology 22:1236–1244.
18. DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.
19. Holden HM, Thoden JB, Gilbert M (2016) Enzymes required for the biosynthesis of *N*-formylated sugars. Curr Opin Struct Biol 41:1–9.
20. Newton DT, Mangroo D (1999) Mapping the active site of the *Haemophilus influenzae* methionyl-tRNA formyltransferase: residues important for catalysis and tRNA binding. Biochem J 339:63–69.
21. Shim JH, Benkovic SJ (1999) Catalytic mechanism of *Escherichia coli* glycinamide ribonucleotide transformylase probed by site-directed mutagenesis and pH-dependent studies. Biochemistry 35:10024–10031.
22. Glickson JD, Phillips DC, Rupley JA (1971) Proton magnetic resonance study of indole NH resonances of lysozyme-assignment, deuterium exchange kinetics, and inhibitor binding. J Am Chem Soc 93:4031–4035.
23. Quiocho FA (1986) Carbohydrate-binding proteins-tertiary structures and protein–sugar interactions. Annu Rev Biochem 55:287–315.
24. Spiwok V (2017) CH/pi interactions in carbohydrate recognition. Molecules 22:9–11.
25. Zuccotti S, Zanardi D, Rosano C, Sturla L, Tonetti M, Bolognesi M (2001) Kinetic and crystallographic analyses support a sequential-ordered bi bi catalytic mechanism for *Escherichia coli* glucose-1-phosphate thymidylyltransferase. J Mol Biol 313:831–843.
26. Allard ST, Giraud MF, Whitfield C, Graninger M, Messner P, Naismith JH (2001) The crystal structure of dTDP-D-glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar *Typhimurium*, the second enzyme in the dTDP-L-rhamnose pathway. J Mol Biol 307:283–295.
27. Burgie ES, Holden HM (2007) Molecular architecture of DesI: a key enzyme in the biosynthesis of desosamine. Biochemistry 46:8999–9006.
28. Knirel YA, Vinogradov EV, Shashkov AS, Dmitriev BA, Kochetkov NK, Stanislavsky ES, Mashilova GM (1985) Somatic antigens of *Pseudomonas aeruginosa*. The structure of the O-specific polysaccharide chains of lipopolysaccharides of *P. aeruginosa* serogroup O4 (Lanyi) and related serotype O6 (Habs) and immunotype 1 (Fisher). Eur J Biochem 150:541–550.
29. 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 (2010) Inactivation of formyltransferase (wbkC) gene generates a *Brucella abortus* rough strain that is attenuated in macrophages and in mice. Vaccine 28:5627–5634.
30. Dow GT, Gilbert M, Thoden JB, Holden HM (2017) Structural investigation on WlaRG from *Campylobacter jejuni*: a sugar aminotransferase. Protein Sci 26:586–599.
31. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.
32. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D60:2126–2132.
33. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Cryst D66: 486–501.
34. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst D53:240–255.
35. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. J Mol Biol 231:1049–1067.