Published in final edited form as:

J Struct Biol. 2018 May ; 202(2): 175–181. doi:10.1016/j.jsb.2018.01.006.

Structure of the *Bacillus anthracis* dTDP-L-rhamnose biosynthetic pathway enzyme: dTDP- α -D-glucose 4,6-dehydratase, RfbB

Trevor Gokey^a, Andrei S. Halavaty^{b,c}, George Minasov^{b,c}, Wayne F. Anderson^{b,c}, and Misty L. Kuhn^{a,*}

^aDepartment of Chemistry and Biochemistry, San Francisco State University, USA

^bDepartment of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, USA

^cCenter for Structural Genomics of Infectious Diseases (CSGID), USA

Abstract

Many bacteria require L-rhamnose as a key cell wall component. This sugar is transferred to the cell wall using an activated donor dTDP-L-rhamnose, which is produced by the dTDP-L-rhamnose biosynthetic pathway. We determined the crystal structure of the second enzyme of this pathway dTDP- α -D-glucose 4,6-dehydratase (RfbB) from *Bacillus anthracis*. Interestingly, RfbB only crystallized in the presence of the third enzyme of the pathway RfbC; however, RfbC was not present in the crystal. Our work represents the first complete structural characterization of the four proteins of this pathway in a single Gram-positive bacterium.

Keywords

cell wall polysaccharide producing enzymes; co-crystallization of metabolic enzymes; dehydratase; protein crystallization; therapeutic drug target

Background on L-rhamnose and enzymes of the dTDP-L-rhamnose biosynthetic pathway

Numerous bacteria contain L-rhamnose as a structural component of their cell envelopes (Mistou et al., 2016). This sugar is critical in many bacteria because it is incorporated into cell wall polysaccharides in Gram-positive bacteria (Mistou et al., 2016) and lipopolysaccharides in many Gram-negative bacteria (Jiang et al., 1991; Macpherson et al., 1994; Marolda & Valvano, 1995; Rochetta et al., 1999). The activated donor dTDP-L-

*To whom correspondence should be addressed: Dr. Misty L. Kuhn, San Francisco State University, 1600 Holloway Ave., San Francisco, CA 94132, mkuhn@sfsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

rhamnose is used to transfer L-rhamnose to components of the cell envelope and is synthesized using a series of four enzymes that comprise the dTDP-L-rhamnose biosynthetic pathway (Giraud & Naismuth, 2000). Genes that encode these enzymes are annotated inconsistently across organisms, as *rml*, *rfb*, or *rff* A, B, C, and D. The effect of preventing incorporation of L-rhamnose into bacterial cell walls varies by species, but it causes severe growth and cell division abnormalities in Gram-positive bacteria like *Streptococcus pyogenes* and *Streptococcus mutans* (Beek et al., 2015). Since this sugar is not found in eukaryotes, targeting enzymes from biosynthetic pathways that produce bacterial cell wall components is one avenue for therapeutic drug development toward bacterial pathogens (Adibekian et al., 2011; Maki & Renkonen, 2004; Mistou et al., 2016).

B. anthracis is a Gram-positive bacterium that causes the disease Anthrax and is currently classified as a Category A Priority Pathogen by the US National Institute of Allergy and Infectious Diseases due to its ease of transmission and high mortality rate. Throughout history this disease has resulted in the death of significant numbers of humans and livestock around the world, but it is largely seen as a disease of the past for humans, except as a bioterrorist agent or potential re-emerging disease. The dTDP-L-rhamnose biosynthetic enzymes have been widely studied across a variety of bacteria, however, the full pathway of enzymes has not been structurally characterized from a single Gram-positive bacterium. Therefore, we have determined the structures of all four enzymes of the pathway from this important human pathogen (Protein Data Bank (PDB) accession codes: RfbA (4ecm; Baumgartner *et al.*, 2017), RfbB (6bi4), RfbC (3ryk; Shornikov *et al.*, 2017), and RfbD (3sc6; Law *et al.*, 2017)), and will describe the structure of the second enzyme dTDP- α -D-glucose 4,6-dehydratase (RfbB) here. RfbB produces dTDP-4-keto-6-deoxyglucose from dTDP- α -D-glucose using an NAD⁺ coenzyme via a series of oxidation, dehydration, and reduction steps. First, NAD⁺ oxidizes the glucosyl C4' of dTDP-glucose. Second, water is eliminated from the glucosyl C5' and C6' of the newly formed intermediate. Finally, NADH reduces this intermediate at the glucosyl C6' position to form the product dTDP-4-keto-6-deoxyglucose (Allard et al., 2002; Gabriel & Lindquist, 1968; Gross et al., 2000; Hegeman et al., 2001). The third enzyme of the dTDP-L-rhamnose pathway (RfbC) can then use this product to proceed with biosynthesis of dTDP-L-rhamnose. Alternatively, dTDP-4-keto-6-deoxyglucose can be used by enzymes of other metabolic pathways since this sugar is a precursor to nearly every known deoxy-sugar biosynthetic pathway (Liu & Thorson, 1994), including polyketide biosynthetic pathways.

Protein expression and purification

The RfbB and RfbC genes from *Bacillus anthracis* str. Ames were cloned into the pMCSG19c ampicillin resistant vector using previously described procedures (Kwon & Peterson, 2014). The genes in this vector are expressed with an N-terminal maltose binding protein (MBP) tag followed by a tobacco vein mottling virus (TVMV) protease cleavage site, a polyhistidine tag, tobacco etch virus (TEV) protease cleavage site, and the protein of interest. The TVMV protease is constitutively expressed by an integrated gene on the plasmid, which removes the MBP tag prior to purification with Ni-NTA affinity chromatography. Each plasmid was separately transformed into kanamycin resistant *E. coli* BL21(DE3) Magic cells (Dieckman *et al.*, 2002) for protein expression. The RfbB protein

was grown in 1.5 L of M9 medium in the presence of SeMet as described previously (Makowska-Grzyska et al., 2012) and the RfbC protein was grown and expressed separately in 1.5 L of Terrific Broth. Cells were grown to an OD_{600nm} of 0.6–0.8 at 310.15 K and then cultures were cooled on ice. Protein expression was induced using 0.5 mM IPTG at room temperature overnight with shaking. The next day cells from both cultures (RfbB SeMet and RfbC native) were harvested by centrifugation, resuspended in 100 mL of lysis buffer (10 mM Tris HCl pH 8.3, 500 mM NaCl, 10% glycerol, 0.01% Igepal CA630, and 5 mM beta-mercaptoethanol (BME)), and sonicated together. Both RfbB and RfbC proteins were purified at the same time on the same 5 mL Ni-NTA affinity column and concentrated using previously described procedures (Kuhn et al., 2013). The polyhistidine tags were not removed from the proteins. When we previously expressed and purified the proteins separately, the average yield of RfbB and RfbC was approximately 45 mg/L and 25 mg/L, respectively (Shornikov *et al.*, 2017). Therefore, the relative ratio of RfbB:RfbC proteins in the mixture used for crystallization was estimated as approximately 2:1. As a result, the purified protein used for crystallization was not necessarily a stable complex of RfbB and RfbC in equal molar ratios.

Crystallization

RfbB and RfbC catalyze two sequential steps of the dTDP-L-rhamnose biosynthetic pathway and in some metabolic pathways sequential genes can produce multi-protein complexes (Marsh *et al.*, 2013). Since we were unable to obtain crystals of the RfbB protein even in the presence of ligands, we screened RfbB for crystals in the presence of RfbC because it is the next enzyme in the biosynthetic pathway and may help stabilize RfbB for crystallization (See results and discussion for full explanation of how we arrived at this approach). The total RfbB and RfbC protein concentration used for crystallization was 7.5 mg/mL in 10 mM Tris HCl pH 8.3, 500⁺. One μ L of protein was mM NaCl, 5 mM BME, 10% glycerol, 5 mM MgCl₂, and 1 mM NAD added to 1 μ L of reservoir solution in a 96-well microplate made for sitting drop vapor diffusion. Crystals of the RfbB protein grew in the condition containing 0.1 M Tris HCl pH 8.0, 1.56 M ammonium sulfate, and 8.5% (w/v) PEG3350.

Data collection and processing

Crystals of RfbB were soaked in a cryo-protectant solution, which was made by mixing in a 1:1 ratio 3.6 M ammonium sulfate and 50% sucrose, and flash-cooled in liquid nitrogen. Data were collected on the 21ID-G beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source at Argonne National Laboratory. Data were indexed, scaled, and integrated with HKL-2000 (Otwinowski & Minor, 1997). The data collection and processing statistics are presented in Table 1.

Structure solution and refinement

The RfbB structure was determined using Phaser (McCoy et al., 2007) from the CCP4 package (Winn et al., 2011). The structure with PDB code 1r66 (51% sequence identity to RfbB; Allard et al., 2004) was used as a search model for molecular replacement and the

initial model was rebuilt with ARP/wARP (Morris et al., 2003), manually corrected in Coot (Emsley & Cowtan 2004; Emsley et al., 2010), and refined with REFMAC v5.7 (Murshudov et al., 2011). At this point NAD^+ , sucrose, sulfate and a nickel ion (coordination number 4, which is more frequent for nickel; the nickel was likely from purification since we used nickel affinity chromatography) were fitted in the well defined electron density maps and the model was refined for several cycles using Refmac and corrected in Coot. During the final stages of refinement, we applied the translation-libration-screw (TLS) group correction. The TLS groups were defined using the TLSMD server (Painter & Merritt, 2006; <http://skuld.bmsc.washington.edu/~tlsmd/>). The quality of the final model was assessed using the PDB validation server (<http://deposit.pdb.org/validate/>) and MolProbity (Chen et al., 2010; Davis et al., 2007; <http://molprobity.biochem.duke.edu/>). Structure refinement statistics and the model quality are shown in Table 1. Chains A and C are missing residues 86–94, 273, 274, and 322; Chains B and D are missing residues 86–94, 272–274, and 322. Coordinates and supporting experimental data for the RfbB structure have been deposited into the Protein Data Bank under the accession code 6bi4.

Approach for obtaining crystals of RfbB protein from *Bacillus anthracis*

We used several traditional strategies to attempt to crystallize the *B. anthracis* RfbB protein, including screening it for crystallization in the presence and absence of NAD^+ and/or dTDP- α -D-glucose substrates. Additionally, we cloned the RfbB protein into different vectors (e.g. ones that produced proteins with either an N-terminal or C-terminal polyhistidine tag, or N-terminal MBP tag), and screened the proteins produced against a panel of different crystallization screens. The construct in the pMCSG7 vector that contains an N-terminal polyhistidine tag did not produce soluble protein, while constructs with a C-terminal polyhistidine tag (pMCSG28 vector) or N-terminal MBP tag (pMCSG19c vector) did. The RfbB protein produced from the pMCSG28 vector failed crystallization trials, and the RfbB protein produced from the pMCSG19c vector was used for further crystallization trials as described here. Since we were unable to obtain single crystals of RfbB using standard methods and we were able to produce soluble protein of all other proteins of the dTDP-L-rhamnose biosynthetic pathway (RfbA, RfbC, and RfbD) from the same organism, we utilized an alternative strategy of screening different combinations of *B. anthracis* Rfb proteins for crystallization in hopes that we would obtain crystals of RfbB.

Our rationale for this approach of combining different proteins within the same pathway was as follows: if proteins within a metabolic pathway have been shown to form multi-protein complexes (Marsh *et al.*, 2013) and it is theorized that proteins in the cell are in close proximity with each other (Sreer 1987), then the presence of the other Rfb proteins of the pathway may stabilize the RfbB protein for crystallization. We used a similar strategy to obtain crystals for structure determination of the *B. anthracis* RfbC protein (PDB ID: 3ryk; Shornikov *et al.*, 2017). Our approach for crystallization of RfbB differed from that of RfbC because here we produced the RfbB protein with selenomethionine (SeMet) and all of the other Rfb proteins (RfbA, RfbC, and RfbD) natively (in absence of SeMet; see Shornikov *et al.*, 2017 for more information). We then screened different combinations of Rfb proteins (e.g. RfbB and RfbA; RfbB and RfbC; RfbB and RfbD; RfbA, RfbB, and RfbC; RfbB, RfbC, and RfbD; RfbA, RfbB, and RfbD; RfbA, RfbB, RfbC, and RfbD) in different

proportions and the presence and absence of substrates. This enabled us to streamline crystal screening of RfbB specifically during data collection since we could detect signal from SeMet if RfbB was present. The combination of proteins that produced RfbB crystals for data collection and structure solution was RfbB and RfbC in the presence of NAD⁺. Although crystals were SeMet derivatives, anomalous signal was not strong enough to solve the structure by Single Anomalous Dispersion (SAD) techniques. Therefore, we solved the structure by Molecular Replacement. It appeared that only the RfbB protein crystallized since we did not observe electron density for the RfbC protein. One possible reason for the absence of a complex of RfbB and RfbC is that the complex may be more soluble and we likely had supersaturation for RfbB alone. The RfbB protein used in each of these different crystallization trials was from the same batch of protein; therefore, we concluded the presence of RfbC aided the crystallization of RfbB. However, since each combination of proteins was only screened once, it is possible some other factor besides RfbC created favorable conditions for RfbB to crystallize.

Structure of RfbB

The RfbB structure was determined in the *I*422 space group with 4 chains in the asymmetric unit at 2.9 Å resolution (PDB ID: 6bi4). RfbB is a homodimeric α/β protein (Figure 1) with two distinct domains: the N-terminal coenzyme-binding domain and the C-terminal nucleotide-binding domain. A four-helix bundle creates the dimeric interface of RfbB (Figure 1C), which is dictated predominately through hydrophobic interactions. Two unresolved regions in the RfbB protein structure were present: residues 86 through 94 in the coenzyme-binding domain and residues 272, 273, 274 and 322 of the nucleotide-binding domain (Figure 1, Figure 2). The residues that are unresolved vary by chain (see *Structure solution and refinement* section for details). All four chains have NAD⁺ bound in the coenzyme-binding domain and sulfate ions bound in the nucleotide-binding domain similar to where the pyrophosphate moiety of dTDP- α -D-glucose binds in homologous structures. Two nickel ions and two sucrose molecules were observed at the dimer interface, which allowed for close interaction between the individual dimers to enhance crystal packing, but are not physiologically relevant and are not critical for catalysis due to their location in the structure. A series of functionally conserved active site residues were observed in the RfbB structure, including Y150, D126, and E127 (Figure 2), which are important for abstracting and donating protons during the course of the reaction.

Structural comparison of RfbB homologs

RfbB has two flexible regions that are disordered in our *B. anthracis* RfbB structure (Figure 3A). In homologous structures, the 273–276 region of RfbB exhibits significant reorganization from a loop to an alpha-helix upon binding dTDP- α -D-glucose substrate, which acts as a lid over the nucleotide-binding region of the protein (Figure 3A). A comparison of the 6bi4 RfbB structure with those of other homologs shows the two drastic open (Allard et al., 2001; RmlB, PDB ID: 1g1a) and closed (DesVI, PDB ID: 1r6d) conformations of this loop (Figure 3B and C, respectively). In the closed conformation of the DesVI 1r6d structure, this loop contains R274 that interacts with both α - and β -phosphates of dTDP- α -D-glucose and the conserved D90 residue (Figure 3C). On the

opposite side, a mobile loop forms an alpha-helix when dTDP- α -D-glucose is bound, bringing the H88 residue contained within this loop of the 1r6d structure close enough to interact with the phosphates of the substrate (Figure 3C) and stabilize the closed conformation of the protein. The RmlB 1g1a structure of the open conformation shows the residues on these two loops are far from each other (Figure 3B), and in our 6bi4 RfbB structure they are disordered. The increased flexibility of the loop of RfbB compared to homologs from Gram-negative bacteria (e.g. 1g1a) may be due to the presence of a P96 residue, which is not available to H-bond with the conserved D88 like the backbone nitrogen of alanine does in the 1g1a structure, or lack of substrate bound in the active site. It appears Gram-positive bacteria have a conserved proline residue in this position while Gram-negative bacteria have either an alanine or valine residue. The 1oc2 structure from the Gram-positive bacterium *Streptococcus suis* has this conserved proline (Figure 2), but it has a substrate analog bound in its active site and this flexible loop adopts a helix conformation similar to the 1r6d structure.

Usefulness of structures of all enzymes in metabolic pathway from same organism

While several structures of RfbB homologs have been determined at higher resolution than our *B. anthracis* RfbB structure, this was the highest resolution data we could obtain, even after attempting several traditional and alternative crystallization strategies. In spite of its lower resolution, our structure of RfbB from *B. anthracis* is still valuable for several reasons. First, it completes the structural characterization of all of the enzymes of the dTDP-L-rhamnose pathway in a single Gram-positive bacterium. This may be useful for downstream computational studies of protein-protein interactions or be used for fitting structures into lower resolution maps of Cryo-EM experiments (Sudha *et al.*, 2014). Second, proteins within a specific organism likely form protein-protein interactions that are unique to their sequences compared to other organisms (Rekha *et al.*, 2005; Sudha *et al.*, 2014). Therefore, having structures of all of the proteins from one organism may provide new avenues for targeted drug design at protein interfaces and may provide insight as to how this pathway is regulated in the cell if these proteins indeed form a multi-protein complex.

Conclusions

Our approach for crystallizing proteins within a metabolic pathway presented here may be a strategy others may use for salvaging failed protein crystallization attempts if all proteins within a pathway are cloned and produce soluble protein. This work completes our full structural characterization of the individual enzymes of the dTDP-L-rhamnose biosynthetic pathway from a single Gram-positive bacterium, which may be useful for future studies of potential protein-protein complexes of enzymes of this pathway in *B. anthracis* or other important organisms.

Acknowledgments

This project was funded in part with Startup Funds from San Francisco State University (to MLK). Additional funding for this project included Federal funds from the National Institute of Allergy and Infectious Diseases,

National Institutes of Health (NIH), U.S. Department of Health and Human Services, under Contracts No. HHSN272200700058C and HHSN272201200026C (to WFA).

References

- Adibekian A, Stallforth P, Hecht ML, Werz DB, Gagneux P, Seeberger PH. Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chem Sci*. 2011; 2:337–344.
- Allard ST, Beis K, Giraud MF, Hegeman AD, Gross JW, Wilmouth RC, Whitfield C, Graninger M, Messner P, Allen AG, Maskell DJ, Naismith JH. Toward a structural understanding of the dehydratase mechanism. *Structure*. 2002; 10:81–92. [PubMed: 11796113]
- Allard ST, Cleland WW, Holden HM. High resolution X-ray structure of dTDP-glucose 4,6-dehydratase from *Streptomyces venezuelae*. *J Biol Chem*. 2004; 279:2211–2220. [PubMed: 14570895]
- Allard ST, Giraud MF, Whitfield C, Graninger M, Messner P, Naismith JH. 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*. 2001; 307:283–295. [PubMed: 11243820]
- Baumgartner J, Lee J, Halavaty AS, Minasov G, Anderson WF, Kuhn ML. Structure of the *Bacillus anthracis* dTDP-L-rhamnose-biosynthetic enzyme glucose-1-phosphate thymidyltransferase (RfbA). *Acta Cryst*. 2017; F73:621–628.
- Beek SL, Le Breton Y, Ferenbach AT, Chapman RN, Aalten DM, Navratilova I, Boons GJ, McIver KS, van Sorge NM, Dorfmüller HC. GacA is essential for Group A *Streptococcus* and defines a new class of monomeric dTDP-4-dehydrorhamnose reductases (RmlD). *Mol Microbiol*. 2015; 98:946–962. [PubMed: 26278404]
- Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst*. 2010; D66:12–21.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007; 35:W375–W383. [PubMed: 17452350]
- Dieckman L, Gu M, Stols L, Donnelly MI, Collart FR. High throughput methods for gene cloning and expression. *Protein Expr Purif*. 2002; 25:1–7. [PubMed: 12071692]
- Emsley P, Cowtan K. Coot: model building tools for molecular graphics. *Acta Cryst*. 2004; D60:2126–2132.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst*. 2010; D66:486–501.
- Gabriel O, Lindquist LC. Biological mechanisms involved in the formation of deoxy sugars. IV Enzymatic conversion of thymidine diphosphoglucose-4T to thymidine diphospho-4-keto-6-deoxyglucose-6T. *J Biol Chem*. 1968; 243:1479–1484. [PubMed: 4869561]
- Gerrata B, Cleland WW, Frey PA. Mechanistic roles of Thr134, Tyr160, and Lys164 in the reaction catalyzed by dTDP-glucose 4,6-dehydratase. *Biochemistry*. 2001; 40:9187–9195. [PubMed: 11478886]
- Giraud MF, Naismith JH. The rhamnose pathway. *Curr Opin Struct Biol*. 2000; 10:687–696. [PubMed: 11114506]
- Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paem J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 2010; 38:W695–W699. [PubMed: 20439314]
- Gross JW, Hegeman AD, Vestling MM, Frey PA. Characterization of enzymatic processes by rapid mix-quench mass spectrometry: the case of dTDP-glucose 4,6-dehydratase. *Biochemistry*. 2000; 39:13633–13640. [PubMed: 11076501]
- Hegeman AD, Gross JW, Frey PA. Probing catalysis by *Escherichia coli* dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of functional amino acid residues at the active site. *Biochemistry*. 2001; 40:6598–6610. [PubMed: 11380254]

- Jiang XM, Neal B, Santiago F, Lee SJ, Romana LK, Reeves PR. Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Mol Microbiol.* 1991; 5:695–713. [PubMed: 1710759]
- Kuhn ML, Majorek KA, Minor W, Anderson WF. Broad-substrate screen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity. *Protein Sci.* 2013; 22:222–230. [PubMed: 23184347]
- Kwon K, Peterson SN. High-throughput cloning for biophysical applications. *Method Microbiol.* 2014; 1140:61–74.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993; 26:283–291.
- Law A, Stergioulis A, Halavaty AS, Minasov G, Anderson WF, Kuhn ML. Structure of the *Bacillus anthracis* dTDP-L-rhamnose biosynthetic enzyme dTDP-4-dehydroxymannose reductases (RfbD). *Acta Cryst.* 2017; F73:644–650.
- Liu HW, Thorson JS. Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu Rev Microbiol.* 1994; 48:223–256. [PubMed: 7826006]
- Macpherson DF, Manning PA, Morona R. Characterization of the dTDP-rhamnose biosynthetic genes encoded in the rfb locus of *Shigella flexneri*. *Mol Microbiol.* 1994; 11:281–292. [PubMed: 8170390]
- Mäki M, Renkonen R. Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology.* 2004; 14:1R–15R. [PubMed: 14514713]
- Makowska-Grzyska M, Kim Y, Wu R, Wilton R, Gollapalli DR, Wang XK, Zhang R, Jedrzejczak R, Mack JC, Maltseva N, Mulligan R, Binkowski TA, Gornicki P, Kuhn ML, Anderson WF, Hedstrom L, Joachimiak A. *Bacillus anthracis* inosine 5'-monophosphate dehydrogenase in action: the first bacterial series of structures of phosphate ion-, substrate-, and product-bound complexes. *Biochemistry.* 2012; 51:6148–6163. [PubMed: 22788966]
- Marolda CL, Valvano MA. Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) rfb gene cluster: identification of functional homologs of rfbB and rfbA in the rff cluster and correct location of the rffE gene. *J Bacteriol.* 1995; 177:5539–5546. [PubMed: 7559340]
- Marsh JA, Hernández H, Hall Z, Ahnert SE, Perica T, Robinson CV. Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell.* 2013; 153:461–470. [PubMed: 23582331]
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
- Melo A, Elliott WH, Glaser L. The mechanism of 6-deoxyhexose synthesis. I Intramolecular hydrogen transfer catalyzed by deoxythymidine diphosphate D-glucose oxidoreductase. *J Biol Chem.* 1968; 243:1467–1474. [PubMed: 4869560]
- Mistou MY, Sutcliffe IC, van Sorge NM. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol Rev.* 2016; 40:464–479. [PubMed: 26975195]
- Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. *Methods in enzymology.* 2003; 374:229–244. [PubMed: 14696376]
- Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* 2011; D67:355–367.
- Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997; 276:307–326.
- Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr.* 2006; 39:109–111.
- Rekha N, Machado SM, Narayanan C, Krupa A, Srinivasan N. Interaction interfaces of protein domains are not topologically equivalent across families within superfamilies: Implications for metabolic and signaling pathways. *Proteins.* 2005; 58:339–353. [PubMed: 15562516]
- Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 2014; 42:W320–W324. [PubMed: 24753421]

- Rocchetta HL, Burrows LL, Lam JS. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. *Microbiol Mol Biol Rev*. 1999; 63:523–553. [PubMed: 10477307]
- Shornikov A, Tran H, Macias J, Halavaty AS, Minasov G, Anderson WF, Kuhn ML. Structure of the *Bacillus anthracis* dTDP-L-rhamnose-biosynthetic enzyme dTDP-4-dehydrorhamnose 3,5-epimerase (RfbC). *Acta Cryst F*. 2017:664–671.
- Srere PA. Complexes of sequential metabolic enzymes. *Annu Rev Biochem*. 1987; 56:89–124. [PubMed: 2441660]
- Stivala A, Wybrow M, Wirth A, Whisstock J, Stuckey P. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics*. 2011; 27:3315–3316. [PubMed: 21994221]
- Sudha G, Nussinov R, Srinivasan N. An overview of recent advances in structural bioinformatics of protein-protein interactions and a guide to their principles. *Prog Biophys Mol Biol*. 2014; 116:141–150. [PubMed: 25077409]
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Cryst*. 2011; D67:235–242.

Figure 1. Overall structure of *Bacillus anthracis* RfbB protein

A) Topology diagram of RfbB showing α -helices as cylinders and β -strands as arrows. β -sheets are colored in cyan, violet, or blue blocks. The C- and N-termini are colored as red and blue squares, respectively. The figure was created using Pro-origami (Stivala *et al.*, 2011) and refined using Inkscape (<https://inkscape.org/>). (B) Structure of RfbB monomer with NAD⁺ coenzyme in yellow sticks. Missing segments are shown as dashed lines, which connect β 12 to β 13 and β 4 to α 4. The C-terminus is shown as a red sphere and the N-terminus is not visible in this orientation. (C) RfbB dimer. The α 4 and α 6 helices form the dimer interface and interact primarily through hydrophobic interactions. The N- and C-termini are indicated by blue and red spheres, respectively. Figures of protein structures were created using VMD.

Figure 2. Sequences of RfbB and homologs

Sequence alignment of RfbB compared to homologs from two selected Gram-positive (*Streptomyces venezuelae* and *Streptococcus suis*) and Gram-negative (*Salmonella enterica* and *Escherichia coli*) bacteria. PDB IDs are indicated to the left of the sequences. 6bi4 corresponds to RfbB from *B. anthracis*, 1r6d is DesIV from *S. venezuelae*, 1oc2 is RmlB from *S. suis*, 1g1a is RmlB from *S. enterica*, and 1bxx is RmlB from *E. coli*. Strictly conserved residues are highlighted in red with white letters. Residues with similar physicochemical properties are surrounded with blue lined boxes. The alignment was created using Clustal Omega (Goujon *et al.*, 2010) and the figure was produced using EsPrnt 3.0 (Robert & Gouet, 2014).

Figure 3. Comparison of mobile loop region of RfbB and homologs

A) Two unresolved mobile loops of the RfbB protein compared to structural homologs. The RfbB monomers are shown as gray surfaces, the unresolved segments of the protein as dashed lines, and NAD⁺ is shown as gray sticks. The *S. venezuelae* (PDB ID 1r6d) structure is shown as cyan ribbons and represents the closed conformation of the protein when dTDP- α -D-glucose (cyan sticks) and NAD⁺ substrates are bound. The *S. enterica* (PDB ID 1g1a) structure is shown as violet ribbons and represents the open conformation of the protein when dTDP- α -D-glucose is not present. B) Comparison of the active site and mobile loop residues between RfbB in gray and the *S. enterica* homolog in the open conformation in violet. NAD⁺ is shown as gray sticks. Missing regions of the RfbB structure are indicated by dashed lines. C) Comparison of the active site and mobile loop residues between RfbB and the *S. venezuelae* homolog with dTDP- α -D-glucose in cyan and NAD⁺ in gray. Missing

regions of the RfbB structure are indicated by dashed lines. Figures of protein structures were created using VMD.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Data collection, processing, and structure refinement statistics, and the model quality of RfbB.

Data collection	
Space group	<i>I</i> 422
<i>a</i> , <i>b</i> , <i>c</i> (Å)	165.56, 165.56, 292.12
α , β , γ (°)	90.00, 90.00, 90.00
Resolution range (Å)	30.00–2.90 (2.95–2.90)
No. of measured reflections	422,805
No. of unique reflections	44,858 (2,208)
Completeness (%)	100.0 (99.9)
Redundancy	9.4 (9.7)
$\langle I/\sigma(I) \rangle$	13.0 (2.9)
R_{merge}	0.15 (0.77)
Overall B factor from Wilson plot (Å ²)	70.3
Refinement	
Resolution range (Å)	29.95–2.91 (2.98–2.91)
Completeness (%)	99.7 (99.6)
No. of reflections, working set	42,594 (3,072)
No. of reflections, test set	2,264 (160)
R_{cryst}	0.167 (0.269)
R_{free}	0.204 (0.328)
No. of non-H-atoms	
Protein (No. of molecules)	10,037 (4)
Ion	2 (NI); 70 (SO4)
Ligand	176 (NAD); 46 (SUC)
Water	76
R.m.s. deviations	
Bonds (Å)	0.010
Angles (°)	1.667
Average B factors (Å ²)	
Protein	66.5
Ion	116.8 (NI); 112.3 (SO4)
Ligand	60.6 (NAD); 111.6 (SUC)
Water	59.6
TLS bodies	4 for each of the four chains
Ramachandran plot $\ddagger$	
Favored regions (%)	97.0
Allowed regions (%)	3.0
Outliers (%)	0.0

[†]Statistics are based on MolProbity (available on the web (<http://molprobity.biochem.duke.edu>), Davis *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35: W375-W383).

Values in parentheses are for highest-resolution shell.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript