Crystal Structure and Catalytic Mechanism of PglD from *Campylobacter jejuni* 

Nelson B. Olivier and Barbara Imperiali  
From the Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The carbohydrate 2,4-diactetamido-2,4,6-trideoxy-α-D-glucopyranosylamine (BacAc2) is found in a variety of eubacterial pathogens. In *Campylobacter jejuni*, PglD acetylates the C4 amino group on UDP-2-acetamido-4-amino-2,4,6-trideoxy-α-D-glucopyranosylamine (UDP-4-amino-sugar) to form UDP-BacAc2. Sequence analysis predicts PglD to be a member of the left-handed β helix family of enzymes. However, poor sequence homology between PglD and left-handed β helix enzymes with existing structural data precludes unambiguous identification of the active site. The co-crystal structures of PglD in the presence of citrate, acetyl coenzyme A, or the UDP-4-amino-sugar were solved. The biological assembly is a trimer with one active site formed between two protomers. Residues lining the active site were identified, and results from functional assays on alanine mutants suggest His-125 is critical for catalysis, whereas His-15 and His-134 are involved in substrate binding. These results are discussed in the context of implications for proteins homologous to PglD in other pathogens.

N-Linked glycosylation involves the covalent attachment of a carbohydrate moiety to a protein at the amide nitrogen of an asparagine side chain in the consensus sequence Asn-Xaa-Ser/Thr (1). Although the existence of archaeal glycoproteins was described more than 30 years ago (2), N-linked glycosylation was only recently discovered in *Campylobacter jejuni* (3). This eubacterium is a Gram-negative pathogen known to be the leading cause of gastroenteritis in developed countries and has been identified as the most frequently occurring infection predictor of cholera. The phenotypes for impaired N-linked glycosylation in *C. jejuni* are a reduction in natural transformability (7), reduced interaction with epithelial cells in vitro (3), and reduced colonization in animals (3, 8). In *C. jejuni* the first sugar of the heptasaccharide that is N-linked to proteins is BacAc2 (9), a diacetylated form of N-acetylglucosamine (10). BacAc2 is initially synthesized as a uridine diphosphate (UDP) derivative. Gene products of the *C. jejuni* protein glycosylation (pgl) locus form the N-linked heptasaccharide biosynthetic pathway (3), and knockouts of any enzyme involved in the biosynthesis of UDP-BacAc2 disrupts formation of the heptasaccharide (11). The biosynthesis of UDP-BacAc2 occurs by the sequential modification of UDP-N-acetylglucosamine (GlcNAc) at the C6 and C4 positions (Fig. 1). Initially, PglF, a membrane-bound protein and member of the short-chain dehydrogenase family of enzymes, conducts an NAD⁺-dependent dehydration at C6 to form UDP-2-acetamido-2,6-dideoxy-α-D-xylo-hexulose (12). The dehydration step results in formation of a ketone at C4, where PglE conducts a pyridoxal-5'-phosphate-dependent transamination using L-glutamate as the source of the transferred amine to form the UDP-4-amino-sugar (12). PglD, the focus of this study, then transfers an acetyl group from acetyl coenzyme A (AcCoA) to the C4 amine to form UDP-BacAc2 (13).

Bioinformatic analysis shows that PglD contains a series of imperfect tandem repeats collectively known as a hexapeptide repeat motif (14, 15). The pattern of the repeated unit generally conforms to the sequence (LIV)₁, (GAED)₂, X₃, X₄, (STAV)₃, and X₅. The first crystal structure solved of a protein containing this signature sequence was that of UDP-N-acetylglucosamine acetyltransferase (16). The tertiary structure formed by residues in the hexapeptide repeat was shown to be a left-handed β helix. LBH enzymes that have been characterized biochemically are known to acylate substrates such as UDP-GlcNAc (17), galactosides (18), and antibiotics (19, 20). The proposed mechanism of catalysis for LBH enzymes has been summarized in a review by Field and Naismith (21). Briefly, the substrate is activated by abstraction of a proton from either a hydroxyl group or a protonated primary amine by a side-chain functional group, usually the imidazole of histidine. The activated substrate then conducts a nucleophilic attack on the acetyl group of AcCoA, forming a tetrahedral intermediate which is followed by the subsequent release of the deacetylated coenzyme and product.

*This work was supported, in whole or in part, by National Institutes of Health Grant Post-doctoral Fellowship F32GM075415 (NIGMS; to N. B. O.) and Grant GM39334 (to B. I.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S3.  
2 The atomic coordinates and structure factors (codes 3BSS, 3BSY, and 3BSW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).  
3 To whom correspondence should be addressed: MIT, 77 Massachusetts Ave., 18-590, Cambridge, MA 02090. Fax: 617-452-2419; E-mail: imper@mit.edu.
Sequence alignments of PglD and other LβH proteins with existing crystal structures share ~25% sequence similarity that is primarily localized to the LβH domain (data not shown). Sequences that align more favorably with the full-length PglD are proteins from a variety of pathogenic and non-pathogenic organisms (supplemental Fig. S1). Noteworthy among these are homologs found in the Neisseria bacterial species. PglB, a bifunctional transmembrane protein, has been implicated in the biosynthesis of a diacetyl-trideoxyhexose found O-linked to pilin from Neisseria gonorrhoeae and Neisseria meningitides (22, 23). Another PglD homolog is NeuD from Mannheimia hemolytica and Streptococcus agalactiae. In S. agalactiae this protein is known to O-acetylate sialic acid and is required for capsular polysaccharide sialylation (24).

There have been several recent advances toward determining the three-dimensional structure of PglD. A crystal structure of PglD in the apo state has been solved by the Protein Structure Initiative. This structure has been made publicly available from the Protein Data Bank (www.rcsb.org) under the identifier 2NPO (Fig. 2). Although analysis of the structure by the authors is pending publication, it can be seen that the crystal structure is unique and non-redundant, thus providing a valuable representation of other sequences. More recently, two other crystal structures of PglD were solved, one in the presence of citrate (3BFP) and the other in the presence of coenzyme A (2VHE) (25). Although lacking the critical acetyl group on coenzyme A, the authors combined structural analysis with functional assays on site-directed mutants to identify several residues lining the active site as important for catalysis. Moreover, computational and molecular modeling efforts were incorporated into the study resulting in a proposal for the mechanism of catalysis and a mode for binding of the sugar substrate to protein. Although previously modeled by computational methods, a crystal structure of the native sugar substrate would provide physical evidence for describing the mode of substrate binding. Furthermore, structural data that present the active site in dissimilar chemical environments may aid in understanding the function of specific residues during catalysis.

In this report we describe the results of biophysical and biochemical studies designed to further elucidate the catalytic mechanism of PglD. The co-crystal structures of PglD in the presence of citrate, AcCoA, or the UDP-4-amino-sugar have been solved. Each structure shows a chemical environment in the active site that is distinct from that in the previously published crystal structures. Using sedimentation velocity AUC, we also show that PglD self-associates as a homotrimer in solution. Comparison of the structures reveals that the extreme C-terminal portion of the protein undergoes a coenzyme-dependent cis-trans
amide bond isomerization between Val-190 and Pro-191, resulting in an interchange of coils between protomers in the biological assembly. Combining the results from structure and function experiments, we propose a detailed mechanism of catalysis for PglD in the formation of UDP-BacAc₂ and discuss some of the implications for homologs in other pathogenic organisms.

**EXPERIMENTAL PROCEDURES**

**Molecular Biology**—The pglD gene was amplified from genomic DNA (ATCC 700819, designation NCTC 11168) as described elsewhere (13). The amplicon encoding the full-length protein was engineered with the restriction sites NcoI and XhoI, then subcloned into the pETGQ vector (26). This vector was used to express constructs with a thrombin-cleavable octahistidine tag at the N terminus. Site-directed mutagenesis was accomplished using the QuikChange site-directed mutagenesis protocol from Stratagene.

**Protein Expression and Purification**—Heterologous expression was accomplished using the *Escherichia coli* BL-21(DE3) strain (Stratagene). Cells were transformed with pETGQ-construct plasmids and grown to an A₆₀₀ of ~0.6 absorbance units at 37 °C in Luria-Bertani broth; the cultures were cooled to ~16 °C and then induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside. Incorporation of selenomethionine was accomplished following a protocol described elsewhere (27). Twenty hours after induction the cells were harvested by centrifugation and resuspended in ice-cold buffer composed of 50 mM HEPES, 10 mM imidazole, 150 mM NaCl, pH 7.1, at 1/20 the original culture volume. Maintaining a working temperature of 4 °C, the cells were lysed by sonication, and the lysate was cleared by centrifugation in a Type 45 Ti rotor (Beckman/Coulter) at 35,000 rpm. The construct was bound to nickel-nitrioltriacetic acid (Qiagen) in batch using 1 ml of resin loaded onto an SP-Sepharose cation exchange column (GE Healthcare), and the protein eluted with a linear NaCl gradient. Protein expression and purification were confirmed by SDS-PAGE.

**Sedimentation Velocity AUC**—Experiments were conducted in an Optima XL-I ultracentrifuge (Beckman/Coulter) using an An60 Ti four-hole rotor at the Boston Biomedical Research Institute (Watertown, MA). Each experiment was conducted with the temperature in the centrifugation chamber at 37 °C and a rotor speed of 50,000 rpm. The centrifuge was retrofitted with a turbo diffusion pump, circumventing contamination of the optics with oil from the conventional diffusion pump when conducting experiments at temperatures above 25 °C. Data were acquired with the interference optics system using sapphire windows. Each cell assembly was composed of 12-mm double-sector Epon centerpieces with interference slit window holders (Biomolecular Interaction Technologies Center). Samples were dialyzed for 24 h in the gel-filtration running buffer before the experiment. Three sample cells were loaded, with each having a different concentration of PglD (70 ± 1, 25 ± 1, and 6 ± 1 μM) and analyzed in the centrifuge simultaneously. Data were analyzed with the software package SEDANAL (28), and the program SEDNTERP was used to estimate the partial specific volume of the protein and density of the solvent.

**Crystallization and Data Collection**—Protein solutions with UDP-4-amino-sugar or AcCoA were made such that the final concentration of the added substrate was 5 mM and were then incubated on ice for 1 h. The UDP-4-amino-sugar was enzymatically synthesized in vitro using the method previously described (13). The protein solution was concentrated and diluted three times with the filtrate using Amicon 10,000 M₇ₐ cut-off concentrators. Before setting up trays with sitting or hanging drops, the protein was concentrated to 10 mg/ml. The crystallization drops were formed by mixing 1.5 μl of protein solution with 1.5 μl of reservoir solution. All crystals except for the SeMet derivative were grown with a reservoir solution of 20% polyethylene glycol 1000, 100 mM phosphate-citrate, 200 mM Li₂SO₄, pH 4.2. The SeMet derivative crystal grew with a reservoir solution containing 1.0 M sodium citrate and 100 mM imidazole, pH 8.0. Crystals were cryoprotected in a reservoir solution supplemented with 20% glycerol and 5 mM substrate as necessary. Intensity data were collected at 110 K on beamline X6A (National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY) as summarized in Table 1. All data sets were indexed, integrated, and scaled using the HKL2000 software suite of programs, and the scaled intensities were converted to structure factors using the program TRUNCATE (29).

**Structure Determination and Refinement**—The citrate-bound structure was solved by the method of single isomorphous replacement with anomalous scattering. Heavy atom sites in the substructure were identified using SHELXD against data collected at the selenium peak wavelength and truncated to 2.5 Å. Three of five possible selenium sites for a single molecule of PglD in the asymmetric unit were located. The correlation coefficients for all/weak reflections were 17.2/11.1, and the Patterson figure of merit was 18.3. Structure factors from the native data were merged with initial phases using CAD, phase extension to 1.77 Å, and density modification was carried out using SHELXE. Values for contrast, connectivity, mean mapCC, and pseudo-free CC were 1.1, 96, 94, and 80%, respectively. The initial model was built with ARP/wARP (30), fitting 190 of 198 residues in the construct sequence using the automated tracing function. The structures of UDP-4-amino-sugar-bound and AcCoA-bound protein were solved by molecular replacement using the software programs PHASER and MOLREP, respectively. In each case the search model was that of the citrate-bound structure, omitting the citrate molecule.
Table 1

| Data collection | UDP-4-amino | AcCoA | Citrate | SeMet derivative |
|-----------------|--------------|-------|---------|-----------------|
| Space group     | P4, 32       | P2, 2, 2 | P6       | P6              |
| Unit cell dimen. | 162, 162     | 62, 92, 125 | 86, 86, 65 | 86, 86, 65      |
| Resolution (Å)  | 30–2.30      | 50–1.80 | 30–1.77 | 30–2.2         |
| Ramachandran plot | 11.0 (60.7) | 6.3 (36.1) | 6.2 (57.1) | 13.6 (61.8)     |
| Completeness (%) | 100 (100)    | 99.9 (99.9) | 99.7 (99.6) | 100 (100)       |
| Redundancy (%)  | 30.2 (30.8)  | 8.0 (7.5)  | 5.7 (5.7)  | 5.6 (5.6)       |
| Rmerge (%)      | 12.0         | 12.0     | 12.0     | 12.0            |

Refinement

Resolution (Å) 30–2.30 30–1.80 30–1.77 30–2.2
Unique reflections 31,265 56,913 25,746
Rmerge (%): 17.8/19.1 18.4/22.0 18.3/19.9
Completeness (%): 1,687 5,006 1,397
Protein 1,447 4,226 1,397
Water 38 153 13
Overall 202 627 206

B-factors (Å²)

Overall 28.6 19.6 25.9
Protein 27.0 17.0 22.9
Organic 23.7 25.7 21.4
Water 41.4 36.3 46.4
r.m.s.d.

Bond lengths (Å) 0.013 0.010 0.010
Bond angles (°) 1.383 1.620 1.189
Ramachandran plot, %

Overall 85.1/14.9/0.0 86.2/13.6/0.2 89.0/11.0/0.0
Protein 86.2/13.6/0.2 89.0/11.0/0.0 85.1/14.9/0.0
Organic 86.2/13.6/0.2 89.0/11.0/0.0 85.1/14.9/0.0
Overall 85.1/14.9/0.0 86.2/13.6/0.2 89.0/11.0/0.0
PDB code 3BSS 3BSY 3BSW

---

Kinetic Data Measurement and Analysis—Histidine mutant reactions were conducted in size exclusion chromatography buffer supplemented with 6 μg of bovine serum albumin to serve as a carrier protein and 2 mM AcCoA while varying the concentration of the UDP-4-amino-sugar in a volume of 30 μl. Protein quantities per reaction per mutant construct are provided in Table 2. pH-dependent reactions contained 50–150 pg of enzyme and 6 μg of bovine serum albumin. Reactions in the range of pH 6.5–9.0 were conducted in duplicate with a three-component buffer system consisting of 20 mM MES (pKₐ 6.2), 20 mM HEPES (pKₐ 7.6), 20 mM BICINE (pKₐ 8.3), and 150 mM NaCl. All reaction mixtures were incubated for 20 min at 37 °C, boiled for 2 min, and then filtered through a 10,000 M₉ cut-off membrane and analyzed by capillary electrophoresis. The initial reaction rates were fitted to the Michaelis-Menten equation for one substrate using the program SigmaPlot Version 9.0. The kinetic parameters were determined by fitting the initial reaction rates to the Michaelis-Menten equation for one substrate using the program SigmaPlot Version 9.0. Results are given as average values ± standard deviations.

RESULTS

Self-association in Solution—Proteins belonging to the 1βH superfamily of enzymes are generally expected to form trimers.
FIGURE 3. Co-crystal structures. Shown are protein-substrate interactions limited to those discussed in the text. A, PgLID with acetyl coenzyme-A, 3BSY. B, PgLID with the UDP-4-amino-sugar, 3BSS. The boxed area indicates the region depicted in the stereo graphic; blue mesh represents simulated annealing $F_o - F_c$ omit maps contoured to $3 \sigma$; phase calculations excluded coordinates of the respective substrates.
Structure and Catalytic Mechanism of PglD

(16). To investigate the possibility for self-association by PglD, we analyzed a purified protein sample using sedimentation velocity AUC at 37 °C in the absence of exogenous coenzyme or sugar substrate. The sedimentation coefficient of PglD was found to be 5.80 ± 0.02 Svedberg units and showed no concentration dependence to the sedimentation coefficient over a protein concentration range of 6–70 μM (supplemental Fig. S2A). Sample concentrations were calculated using the F-statistics function in SEDANAL (28); the S.D. of the fit was 0.006 fringes using the Levenberg-Marquardt fitting method. A single species model fit using a 95% confidence interval resulted in a molecular mass of 66 ± 3 kDa (supplemental Fig. S2B). The molecular weight of the construct analyzed was 21.4 kDa, calculated by sequence and verified by MALDI-mass spectrometry (data not shown), which suggests that PglD associates as a homotrimer in solution at 37 °C.

Citrate-bound Structure—The co-crystal structure of PglD in the presence of citrate was solved by single isomorphous replacement with anomalous scattering using native data combined with data collected from an isomorphous selenomethionine derivative protein crystal. Statistics for data collection and structure refinement are presented in Table 1. The citrate-bound and 2NPO structures were both solved in the hexagonal space group P63. A single protomer of PglD may be defined as a homotrimer where the carbonyl oxygens are parallel to each other, and are buried from the AUC experiments, we designate the homotrimer as the biologically relevant arrangement.

Acetyl Coenzyme A Binding Site—The AcCoA-bound structure was solved in the orthorhombic space-group P212121 (3BSY, Fig. 3A). Three molecules of AcCoA in association with three molecules of PglD comprise the asymmetric unit. The homotrimeric assembly is centered on a pseudosymmetric 3-fold axis. In the AcCoA-bound structure, each coenzyme molecule was observed at the cleft of two protomers. Within the homotrimeric assembly, one of the coenzyme molecules interacts with a third protomer. The sequence of hexapeptide repeats remains largely unbroken and, unlike several other LβH proteins that use AcCoA as the coenzyme (1XAT (38), 1HM8 (42), 1KHR (43), 1KQA (41), 1MR9 (44), 1T3D (45), 2IU8 (46)), PglD does not present a coil that extends across the interprotomer cleft. Residues from PglD that form key hydrogen bonds and van der Waals contacts with the coenzyme as well as residues primarily responsible for catalysis are located within the LβH domain. Asn-118-a and His-134-a donate hydrogen bonds to the carbonyl oxygen of the thioester. The pantetheine moiety of the coenzyme is fixed in the cleft of the binding pocket; thus, the thioester carbonyl carbon remains exposed to solvent. Ser-136-a is also in the vicinity of the active site, where the β-hydroxyl group hydrogen bonds with the backbone carbonyl of Leu-153-a. The pantetheine moiety is in the extended conformation where the carbonyl oxygens are parallel to each other, point perpendicular to the homotrimer 3-fold axis, and are buried within the interprotomer cleft. The main-chain amides from Ile-155-a and Gly-173-a hydrogen-bound with carbonyl oxygens of the pantetheine moiety. The carbonyl oxygen of Gly-173-a hydrogen-bonds with the C6 amine on the coenzyme nucleotide. The nature of the contacts between the coenzyme and protomer-b is primarily hydrophobic.

In the apo state, the last 10 residues of the C terminus form a coil that interacts with the active-site partner. In the presence of coenzyme the direction of the coil was altered such that residues of the coil now interact with residues of the cognate protomer (Fig. 4). The turn in the backbone occurs between Val-190 and Pro-191, converting the peptide bond from a trans to a cis conformation. This isomerization provides an unobstructed path for the nitrogenous base of the coenzyme to favorably interact with protein residues. Superimposing Cα carbons from the LβH domain (Ile-79—Gly-185) of the UDP-4-amino-sugar and AcCoA-bound structures results in a r.m.s.d. of 0.16 Å. From this comparison we observed that the main-chain carbonyl oxygen of Gly-189 would create a steric clash with the adenosine of the coenzyme if the gate did not open (Fig. 4).
UDP-4-amino-sugar Binding Site—The UDP-4-amino-sugar-bound crystal structure of PglD (3BSS, Fig. 3B) was solved in the cubic space group P4_32 by molecular replacement using the citrate-bound model as the search ensemble (3BSW, coordinates of the citrate molecule excluded). The asymmetric unit contains one molecule of the UDP-4-amino-sugar bound to one molecule of PglD. Also unique to this structure, the Matthews coefficient for the asymmetric unit is 8.1 Å³/Da; therefore, ~85% of the unit cell volume is occupied by solvent.

In the apo state the N-terminal domain contains three parallel β strands and three α helices that form a β-α-β-α-β-α motif (Fig. 2). Helix α2 is formed by residues Met-40—Thr-45, inclusive. In the presence of the sugar substrate, however, these residues form a coil (Fig. 5). Unraveling of the helix allows for side and main chain atoms of the protein to make direct contact with the sugar substrate. Protein residues that interact with the UDP-sugar substrate reside primarily in the N terminus and belong to a single protomer with two exceptions (Fig. 3B). On the adjacent protomer forming the coenzyme-binding site is His-125, which interacts with the pyranose C4-amine. The second is Asn-162, which donates a hydrogen bond to the carboxyl oxygen of the pyranose C2-acetyl group. Two conserved aspartic acid residues (Asp-35 and -36) accept hydrogen bonds from the ribosyl 3'-hydroxyl group and the uridine imide, respectively. The uridine ring is stabilized in the binding pocket by alignment with the equatorial face of Phe-37 and the hydrophobic pocket formed by Tyr-10 and isoleucine side chains 55, 60, and 64.

In the 2NPO structure the side chain of Lys-38 was modeled, but the residues Ala-12 and Ser-13 were not. In the presence of the UDP-sugar substrate, the electron density was of sufficient quality to fit the previously unmodeled residues. Helix α2 is clearly dissolved, and the resulting coil appears stabilized in part by the formation of hydrogen bonds between protein and the UDP-sugar substrate (Fig. 5). The side chain of Ser-13 plays a substantial role in the binding pocket by forming hydrogen bonds with the substrate α-phosphate and Nε of Lys-38. Also, in the presence of substrate the side chain of His-15 is tucked into the binding pocket, and the main chain amides of Gly-14 and His-15 form hydrogen bonds with the sugar substrate β-phosphate.

Alanine Mutants of Active Site Residues—LβH enzymes involved in acyl transfer often have a histidine residue impli-
cated in the catalytic mechanism. The side chain imidazole is purported to act as a general base by abstracting a proton from the substrate, thereby activating the substrate for nucleophilic attack on the acyl group. We identified three histidine residues within the region of the active site that could potentially play key roles in catalysis (His-15, His-125, and His-134) (Fig. 6A).

Alanine mutants for each histidine were prepared, and apparent kinetic parameters were determined using capillary electrophoresis to monitor the reaction progress; the results are summarized in Table 2. The H125A mutation results in a decrease in catalytic efficiency ($k_{cat}/K_m$) by nearly 4 orders of magnitude, suggesting that the native histidine is responsible for activation of the substrate. H15A results in a mutant with a higher $K_m$ and an approximate 3-fold decrease in the turnover number ($k_{cat}$). These results suggest that His-15 is important for binding of the UDP-4-amino-sugar and may play a role in catalysis. Mutant H134A retains the wild type $K_m$ but suffers a 30-fold decrease in catalytic efficiency, suggesting that the native residue is important for catalysis.

**pH Dependence of Kinetic Parameters for Substrate Acetylation**—The kinetic parameters for the acetylation of the UDP-4-amino-sugar by PglD were measured over the range of pH 6.5–9.0. The $k_{cat}/K_m$ versus pH profile for the reaction is shown in Fig. 7. The profile shows a break in the slope near neutral pH, approaches a limiting value near pH 8.5, and continues a decline at pH 9.0. Fitting the data to Equation 1 yields $pK_a$ 7.6 values of 7.5 ± 0.1 and 9.5 ± 0.4 for two groups, where

### Table 2

**Kinetic parameters for native and point mutants of PglD**

| Construct | Protein | $K_m$ | $k_{cat}$ | $k_{cat}/K_m$ | $V_{max}$ |
|-----------|---------|-------|-----------|--------------|-----------|
| Native    | 2.5 × 10⁻⁶ | 1.0 ± 0.1 | 314       | 314          | 1.47 × 10⁴ |
| H15A      | 5.0 × 10⁻⁷ | 2.5 ± 0.3 | 138       | 55           | 6.46 × 10³ |
| H125A     | 6.5 × 10⁻⁴ | 1.0 ± 0.2 | 9.1 × 10⁻² | 9.1 × 10⁻²   | 4.25      |
| H134A     | 2.0 × 10⁻⁶ | 0.8 ± 0.2 | 10        | 13           | 4.71 × 10² |

**FIGURE 7.** pH dependence of log ($k_{cat}/K_m$) for the PglD-catalyzed acetylation of UDP-4-amino-sugar in the presence of 2 mM AcCoA. Data were fit to Equation 1, resulting in $pK_a$ 7.6 = 7.5 ± 0.1 and $pK_a$ 9.5 ± 0.4.

**FIGURE 8.** Proposed mechanism of catalysis by PglD. Proton abstraction, Glu-126 serves to increase the basicity of His-125. Asn-118 and His-134 hydrogen-bond with the thioester carbonyl oxygen resulting in the thioester acetate being exposed to solvent. His-125 abstracts a proton from the carbohydrate C4 amine to produce a lone pair of electrons that perform a nucleophilic attack on the thioester acetate. Tetrahedral intermediate, the electron rich tetrahedral intermediate is stabilized by polar side chains near the active site. His-125 is returned to the pre-catalytic state by interacting with Glu-124. Expel and Regenerate, delocalization of electrons in the tetrahedral intermediate results in the release of the de-acetylated cofactor and release of a proton from the C4 aminoacetyl group of UDP-BacAc$_2$. Glu-124 is returned to the pre-catalytic state by protonating the thiolate.
the first must be deprotonated and the second protonated for the enzyme-catalyzed reaction to occur. These values are consistent with the deprotonation of a histidine side chain (His-125) and primary amine in the protonated state (UDP-4-amino-sugar). Data for reactions outside this range of pH were not included in the calculations due to extremely high error values.

DISCUSSION

BacAc2 is an essential component of the N-linked heptasaccharide in C. jejuni (11). Because formation of the oligosaccharide may affect the pathogenicity of the bacterium (3, 8), it is imperative to gain insight into the enzymatic reactions in which BacAc2 is biosynthesized and utilized. Shown in Fig. 8 is the proposed mechanism for UDP-BacAc2 biosynthesis by PglD. The finding that His-125 serves as the key residue for catalysis was consistent with the role played by key histidines in other LβH enzymes (38, 40, 41, 45). Another notable theme that recurs in LβH enzymes is the existence of a negative dipole interacting with the key catalytic histidine (38, 41, 43, 47). This component of the dyad serves to increase the basicity of the histidine, thus enhancing the ability to act as a general base by abstracting a proton from the substrate to be acylated. PglD follows a similar chemical logic; the basicity of the His-125 imidazole moiety is increased by the carboxylate of Glu-126 interacting with Nε1 (Fig. 6A). In the citrate-bound model, the side chain carboxylate of Glu-124 points toward the coenzyme-binding pocket, clearly positioned to hydrogen bond with the side chain carboxylate of Glu-124 points toward the coenzyme-deprotonated the Nβ atom may be protonated by the side chain of Glu-124 once it has abstracted a proton from the substrate to facilitate the nucleophilic attack on the acetate carbonyl group of the coenzyme. The tetrahedral intermediate then breaks down from the C4 amine to make ethanolamine shifts the pKa of 10.7 in water at 25°C. Adding a hydroxyl group on the β carbon to make ethanolamine shifts the pKa to 9.4 (48). The UDP-4-amino-sugar has a hydroxyl group at the C3 position that is also β to the C4 primary amine, suggesting that the pKa 7.6 of the primary amine on the PglD substrate is similar to ethanolamine. Therefore, His-125 activates the protonated C4 amine via proton abstraction, thus facilitating the nucleophilic attack on the acetate carbonyl group of the coenzyme. The second proton on the C4 amine would be lost to solvent during the breakdown of the tetrahedral intermediate. The proton abstracted from the UDP-4-amino-sugar is removed from His-125 by Glu-124 and is subsequently transferred to the thiolate.

It has been suggested that the identity of the diacetyl trideoxy hexose found O-linked to serine residues in pili from the Neisseria species may be BacAc2 (23). Sequence analysis of PglB from N. gonorrhoeae and N. meningitides predicts the N-terminal domain to be homologous to PglC from C. jejuni, the phosphoglycosyltransferase that transfers phospho-BacAc2 to undecaprenyl phosphate. The C-terminal domain of PglB is homologous to the acetyltransferase PglD from C. jejuni presented in this report (supplemental Fig. S1). The residues that directly contact the carbohydrate moiety of the UDP-4-amino-sugar in PglD are His-15, His-125, and Asn-162 (Fig. 3B). The corresponding residues in PglB of the Neisseria species are His-210, His-333, and Gln-370, respectively. Therefore, the possibility that BacAc2 is the O-linked diacetyl-trideoxy hexose on pili from Neisseria is consistent with our results.

The crystal structures and functional data presented herein provide a new map to investigate other LβH enzymes homologous to PglD. Although PglD catalyzes an acetyl group transfer to an amine, the mechanism presented may also be valid for the transfer of acyl groups in the formation of esters. Recently the enzyme NeuD from S. agalactiae was shown to be involved in the O-acetylation of N-acetylneuraminic acid (Neu5Ac), a variant of sialic acid in group B Streptococcus serotypes (24). In the course of their studies the authors modeled NeuD after the crystal structure (1KRR) of the galactose acetyltransferase from E. coli (41) (16% sequence identity in the alignment). The point mutation K123A was made in NeuD because the residue aligned with the catalytically important His-115 in galactose acetyltransferase. Using an in vivo assay, the authors reported that neuraminate O-acetylation was significantly decreased in the mutant strain. Therefore, it was surmised that Lys-123 was important for catalysis and that the mechanism of acetyl transfer was similar to that of galactose acetyltransferase. Aligning NeuD with PglD produces a more favorable fit than with galactose acetyltransferase (sequence identity of 22%, supplemental Fig. S1). Lys-123 from NeuD aligns perfectly with Lys-110 from PglD. In PglD, Lys-110 is positioned on the outer surface of the LβH and, on the basis of our structure, does not appear to directly participate in catalysis. We propose that His-138 in NeuD, which is flanked by the residues glutamate and histidine, catalyzes the O-acetylation of its substrate in a mechanism similar to what we have outlined for PglD. It is possible that the K123A mutation in NeuD disrupted the native protein fold or unfavorably altered protein-protein interactions in vivo.

We have identified 14 LβH proteins that show significant sequence homology to PglD, 11 of which have not been characterized biochemically (supplemental Fig. S1). The identified sequences represent proteins presumed to be involved in carbohydrate synthesis, amino acid modification, or drug resistance.

Structure and Catalytic Mechanism of PglD

The crystal structures and functional data presented herein provide a new map to investigate other LβH enzymes homologous to PglD. Although PglD catalyzes an acetyl group transfer to an amine, the mechanism presented may also be valid for the transfer of acyl groups in the formation of esters. Recently the enzyme NeuD from S. agalactiae was shown to be involved in the O-acetylation of N-acetylneuraminic acid (Neu5Ac), a variant of sialic acid in group B Streptococcus serotypes (24). In the course of their studies the authors modeled NeuD after the crystal structure (1KRR) of the galactose acetyltransferase from E. coli (41) (16% sequence identity in the alignment). The point mutation K123A was made in NeuD because the residue aligned with the catalytically important His-115 in galactose acetyltransferase. Using an in vivo assay, the authors reported that neuraminate O-acetylation was significantly decreased in the mutant strain. Therefore, it was surmised that Lys-123 was important for catalysis and that the mechanism of acetyl transfer was similar to that of galactose acetyltransferase. Aligning NeuD with PglD produces a more favorable fit than with galactose acetyltransferase (sequence identity of 22%, supplemental Fig. S1). Lys-123 from NeuD aligns perfectly with Lys-110 from PglD. In PglD, Lys-110 is positioned on the outer surface of the LβH and, on the basis of our structure, does not appear to directly participate in catalysis. We propose that His-138 in NeuD, which is flanked by the residues glutamate and histidine, catalyzes the O-acetylation of its substrate in a mechanism similar to what we have outlined for PglD. It is possible that the K123A mutation in NeuD disrupted the native protein fold or unfavorably altered protein-protein interactions in vivo.

We have identified 14 LβH proteins that show significant sequence homology to PglD, 11 of which have not been characterized biochemically (supplemental Fig. S1). The identified sequences represent proteins presumed to be involved in carbohydrate synthesis, amino acid modification, or drug resistance.
Among these proteins are the capsular polysaccharide synthesis protein Wbp1 from *Francisella tularensis*, a highly infectious aerosolizable intracellular pathogen that causes tularemia or rabbit fever (49), sialic acid acetyltransferase NeuD from *M. homoe lytica*, an opportunistic pathogen of cattle, sheep, and other ruminants, and the PglB of the pilin glycosylation pathway from *Wolinella succinogenes*, a non-pathogenic bacterium that contains the genes of several virulence factors found in pathogenic bacteria such as *Helicobacter pylori* and *C. jejuni* (50). Investigators that study pathways leading to pathogenicity in these bacteria can, therefore, use the structures of PglD as a framework for deciphering the mechanisms of metabolite modification.

Acknowledgments—We are grateful to R. Grant from the Massachusetts Institute of Technology biology department for learned discussions and assistance with screening crystals, the staff at the National Synchrotron Light Source beamline X6A for access via the General User program, and especially J. Jackson for assistance in data collection and reduction, W. Stafford and D. Hayes from the Boston Biomedical Research Institute (Watertown, MA) for use of an ALIC with a turbo diffusion pump and assistance in data collection and analysis, M. Gawnowicz from Columbia University for MALDI-mass spectrometry data collection, R. Sasisekharan for use of the capillary electrophoresis system, E. Gouaux from the Vollum Institute (Portland, OR) for the generous gift of the pETGQ vector, and members of the Imperiali, Drennan, Stubbe, and Licht laboratories for helpful discussions.

REFERENCES

1. Weerapana, E., and Imperiali, B. (2006) Glycobiology 16, 91–101
2. Mescher, M. F., and Strominger, J. L. (1976) *J. Biol. Chem.* 251, 2005–2014
3. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., and Guerry, P. (1999) *Mol. Microbiol.* 32, 1022–1030
4. Jacobs, B. C., Rothbarth, P. H., van der Meche, F. G., Herbrink, P., Schmitz, P. I., de Klerk, M. A., and van Doorn, P. A. (1998) *Neurology* 51, 1110–1115
5. Koga, M., Gilbert, M., Li, J., Koike, S., Takahashi, M., Furukawa, K., Hirata, K., and Yuki, N. (2005) *Neurology* 64, 1605–1611
6. Rees, J. H., Soudain, S. E., Gregson, N. A., and Hughes, R. A. (1995) *Science* 268, 270–273
7. Rikov, M., Taulbee, R. W., and Dingley, K. R. (1993) *Structure* 1, 109–119
8. Beaman, T. W., Binder, D. A., Blanchard, J. S., and Roderick, S. L. (1997) *Biochemistry* 36, 489–494
9. Beaman, T. W., Sugantino, M., and Roderick, S. L. (1998) *Biochemistry* 37, 6689–6696
10. Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996) *EMBO J.* 15, 2323–2330
11. Olsen, L. R., Vetting, M. W., and Roderick, S. L. (2007) *Protein Sci.* 16, 1230–1235
12. Wang, X. G., Olsen, L. R., and Roderick, S. L. (2002) *Structure* 10, 581–588
13. Sulzenbacher, G., Gal, L., Penef, C., Fassy, F., and Bourne, Y. (2001) *J. Biol. Chem.* 276, 11844–11851
14. Sugantino, M., and Roderick, S. L. (2002) *Biochemistry* 41, 2209–2216
15. Kehoe, L. E., Snidwongse, I., Courvalin, P., Rafferty, J. B., and Murray, I. A. (2003) *J. Biol. Chem.* 278, 29963–29970
16. Pye, V. E., Tingey, A. P., Robson, R. L., and Moody, P. C. (2004) *J. Biol. Chem.* 279, 40729–40736
17. Buettow, L., Smith, T. K., Dawson, A., Fyffe, S., and Hunter, W. N. (2007) *Proc. Natl. Acad. Sci. U. S. A.* 104, 4321–4326
18. Johnson, C. M., Huang, B., Roderick, S. L., and Cook, P. F. (2004) *Biochemistry* 43, 15534–15539
19. Boudavar, S. (ed) (1996) *The Merck Index*, 12th Ed., Merck Research Laboratories, Whitehouse Station, NJ p. 636
20. Oyston, P. C., Sjostedt, A., and Tittball, R. W. (2004) *Nat. Rev. Microbiol.* 2, 967–978
21. Baar, C., Epping, M., Raddatz, G., Simon, J., Lanz, C., Klimmek, O., Nanadukumar, R., Gross, R., Rosinus, A., Keller, H., Jagtap, P., Linke, B., Meyer, F., Lederer, H., and Schuster, S. C. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, 11690–11695