Structure of the *Escherichia coli* GlmU Pyrophosphorylase and Acetyltransferase Active Sites[†,‡]

Laurence R. Olsen and Steven L. Roderick*

Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

Received October 30, 2000; Revised Manuscript Received December 4, 2000

ABSTRACT: *N*-Acetylglucosamine-1-PO_4 uridyltransferase (GlmU) is a trimeric bifunctional enzyme that catalyzes the last two sequential reactions in the de novo biosynthetic pathway for UDP-GlcNAc. The X-ray crystal structure of *Escherichia coli* GlmU in complex with UDP-GlcNAc and CoA has been determined to 2.1 Å resolution and reveals a two-domain architecture that is responsible for these two reactions. The C-terminal domain is responsible for the CoA-dependent acetylation of Glc-1-PO_4 to GlcNAc-1-PO_4 and displays the longest left-handed parallel β-helix observed to date. The acetyltransferase active site defined by the binding site for CoA makes use of residues from all three subunits and is positioned beneath an open cavity large enough to accommodate the Glc-1-PO_4 acetyl acceptor. The N-terminal domain catalyzes uridyl transfer from UTP to GlcNAc-1-PO_4 to form the final products UDP-GlcNAc and pyrophosphate. This domain is composed of a central seven-stranded β-sheet surrounded by six α-helices in a Rossmann fold-like topology. A Co^{2+} ion binds to just one of the two independent pyrophosphorylase active sites present in the crystals studied here, each of which nonetheless binds UDP-GlcNAc. The conformational changes of the enzyme and sugar nucleotide that accompany metal binding may provide a window into the structural dynamics that accompany catalysis.

The intermediary metabolite UDP-GlcNAc[1] is of central importance to all organisms. Detailed studies of the four enzymatic steps leading to the production of UDP-GlcNAc from fructose 6-phosphate were described only recently for bacteria (*2*, *3*), and still more recently for eukaryotic organisms (*4*−*6*). UDP-GlcNAc is an essential precursor for many biomolecules, including lipopolysaccharide, peptidoglycan, enterobacterial common antigen, and teichoic acid in bacteria as well as GPI anchors and chitin in eukaryotes (*7*−*10*).

The *Escherichia coli glmU* gene encodes the protein *N*-acetylglucosamine-1-phosphate pyrophosphorylase (*2*). This enzyme is a bifunctional homotrimer with 456 residues per subunit that catalyzes the final two biosynthetic reactions leading to UDP-GlcNAc

$$\text{GlcN-1-PO}_4 + \text{acetyl-CoA} \rightarrow \text{GlcNAc-1-PO}_4 + \text{CoA}$$

$$\text{GlcNAc-1-PO}_4 + \text{UTP} \rightarrow \text{UDP-GlcNAc} + \text{PP}_i$$

The two active sites are segregated into domains that are thought to act independently (*3*, *11*). The acetyltransferase activity is confined to the C-terminal domain and precedes uridyl transfer (*11*). The uridyltransferase active site is contained in the N-terminal domain.

The sequence of the C-terminal domain responsible for the CoA-dependent acetylation of GlcN-1-PO_4 is homologous to the sequences of a number of bacterial and plant acetyltransferases possessing a hexapeptide repeat sequence motif (*3*). This motif is characterized by imperfect, tandem repeated copies of the six-residue sequence theme [LIV]-[GAED]-X_2-[STAV]-X (*12*, *13*). The hexapeptide repeat sequence directs folding of a structural domain termed a left-handed parallel β-helix (LβH) (*14*). Structural data are available for a number of such hexapeptide enzymes, including UDP-*N*-acetylglucosamine acyltransferase (*14*), a thermophilic carbonic anhydrase (*15*), tetrahydrodipicolinate *N*-succinyltransferase (*16*), and a xenobiotic acetyltransferase (*17*).

The second reaction catalyzed by GlmU involves uridyl transfer from UTP to GlcNAc-1-PO_4 which produces the final pyrophosphate and UDP-GlcNAc reaction products. The N-terminal pyrophosphorylase domain is homologous with a variety of nucleotide diphosphate sugar pyrophosphorylases from bacteria as well as higher organisms which include a conserved L-X_2-G-X-G-T-X-M-X_4-P-K sequence (*6*). The pyrophosphorylase activity of these homologues is stimulated

[†] This work was supported by the National Institutes of Health (Grant AI42154).

[‡] The X-ray coordinates have been deposited in the RCSB Protein Data Bank (entry 1hv9) and will be released upon publication.

* To whom correspondence should be addressed: Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461. Phone: (718) 430-2784. Fax: (718) 430-8565. E-mail: roderick@aecom.yu.edu.

[1] Abbreviations: CoA, coenzyme A; DapD, tetrahydrodipicolinate *N*-succinyltransferase; DTT, dithiothreitol; GlcNAc, *N*-acetylglucosamine; GlmU, *N*-acetylglucosamine-1-phosphate uridyltransferase; IPTG, isopropyl thiogalactopyranoside; LpxA, UDP-*N*-acetylglucosamine acyltransferase; LβH, left-handed parallel β-helix; MES, 2-(*N*-morpholino)ethanesulfonic acid; MIR, multiple isomorphous replacement; NCS, noncrystallographic symmetry; PCR, polymerase chain reaction; PIP, di-μ-iodobis(ethylenediamine)diplatinum(II) nitrate; rms, root-mean-square; UDP-GlcNAc, uridine 5′-diphospho-*N*-acetylglucosamine; PaXAT, *Pseudomonas aeruginosa* xenobiotic acetyltransferase; σ, root-mean-square units.

10.1021/bi002503n CCC: $20.00
Published on Web 01/26/2001

Table 1: Data Measurement Statistics[a]

data set	no. of observed reflections	no. of unique reflections	resolution (Å)	completeness	$\langle I/\sigma(I)\rangle$	R_{merge}	R_{iso}
native I	495967	91622	2.1	0.98	7.9	0.136	–
native II	190141	34521	2.8	0.98	6.7	0.109	0.058
CH_3HgCl I	25798	8686	5.0	0.82	13.4	0.080	0.067
CH_3HgCl II	98919	31049	2.8	0.90	8.8	0.084	0.062
cis-$Pt(NH_3)_2Cl_2$	42363	7829	5.0	0.74	18.6	0.066	0.051
PIP	90506	30457	2.8	0.86	6.8	0.072	0.173

[a] The data sets are the two native data sets (native I and native II) and the data sets for the three heavy atom derivatives: CH_3HgCl I (0.2 mM, 17 h soak), CH_3HgCl II (0.2 mM, 16 h soak), *cis*-$Pt(NH_3)Cl_2$ (1.3 mM, 2 h soak), and PIP (0.34 mM, 5.5 day soak). The total number of observed reflections, the number of unique reflections, the fractional completeness of the data set, the ratio of individual intensities to their standard deviations, and the merging and isomorphous residuals are given. $R_{merge} = \Sigma|I_i - \langle I\rangle|/\Sigma|I_i|$ within one data set, and $R_{iso} = \Sigma|F_{Nat} - F_{Der|}|/\Sigma|F_{Nat}|$ between native and derivative data sets. For R_{iso} calculations, the heavy atom derivative data sets are compared to the native II data set. The value of R_{iso} for the native II data set is a comparison to that for the native I data set.

by various divalent cations, including Mg^{2+}, Co^{2+}, and Mn^{2+} (*3*, *4*).

Recently, the X-ray crystal structure of a truncated form of GlmU composed of 331 of 456 residues confirmed earlier biochemical studies regarding the organization of its catalytic domains (*1*). The structure of the pyrophosphorylase domain revealed its interactions with UDP-GlcNAc in the absence of divalent cations, and the structure of the inactive truncated acetyltransferase active site demonstrated that a portion of the C-terminal domain was folded as a LβH.

We report here the crystal structure of full-length *E. coli* GlmU determined to 2.1 Å resolution. This structure provides a view of the intact acetyltransferase active site with bound CoA as well as two conformationally distinct forms of the pyrophosphorylase active site with bound UDP-GlcNAc, corresponding to metal-bound and metal-free forms of the enzyme.

EXPERIMENTAL PROCEDURES

Overexpression, Preparation, and Crystallization of E. coli GlmU. The *E. coli glmU* gene was cloned by PCR from genomic DNA, inserted into a pET3a expression vector, and purified to homogeneity using anion exchange and gel filtration chromatography. The subunit mass of the purified enzyme was determined to be 49 193 ± 5 Da by electrospray injection mass spectrometry, in agreement with the predicted subunit mass of 49 190 Da for the conceptually translated protein. Crystals were grown by the hanging drop vapor diffusion method at ambient temperature. Drops consisted of 4 μL of 8.3 mg/mL protein, 14 mM UDP-GlcNAc, 28 mM $MgCl_2$, and 19 mM CoA mixed with 4 μL of reservoir solutions containing 1.65 M ammonium sulfate, 50 mM MES (pH 5.4−6.4), and 2−10 mM $CoCl_2$. Crystals were visible in 48 h and grew to dimensions of 0.3 mm × 0.3 mm × 0.4 mm. These crystals belong to space group *R*32 ($a = b =$ 104.5 Å and $c = 648.2$ Å) and diffract X-rays to at least 2.1 Å resolution.

X-ray Data Collection and Primary Phasing. Two native X-ray diffraction data sets were measured. Native data set I was collected to 2.1 Å resolution at Molecular Structure Corp. (The Woodlands, TX) on an R-Axis IV image plate detector using Cu Kα radiation from a Rigaku RU-H3R X-ray generator equipped with Osmic Blue optics and operating at 50 kV and 100 mA. Native data set II and all of the heavy atom derivative data sets were collected with a Siemens X-1000 multiwire detector using Cu Kα X-rays produced by a Rigaku RU-200 rotating anode generator operating at 50 kV and 80 mA. All data sets were measured at ambient temperature using a single crystal, except for the CH_3HgCl derivative data set, which required two crystals. Data from the multiwire detector were processed using XDS (*18*) and reduced using SCALEPACK (*19*). The image plate data were processed using DENZO and SCALEPACK (Table 1).

Experimental phases were determined by the method of MIR using native data set II and the *cis*-$Pt(NH_3)_2Cl_2$, CH_3-HgCl, and PIP derivative data sets (Table 1). Heavy atom binding sites were identified by inspection of difference Patterson and cross-phased Fourier maps. Phases were calculated using SHARP (*20*) to 2.8 Å resolution.

Model Building and Atomic Parameter Refinement. The MIR electron density map was solvent flattened with SOLOMON (*21*) and averaged about a noncrystallographic 2-fold axis relating the two subunits. The resulting map and the original MIR map were used to build an initial atomic model using the graphics program O (*22*). Upon examination of the model, it became clear that two different NCS operators were necessary to relate different portions of the polypeptide chain as a consequence of different hinge bending angles between the N- and C-terminal domains of the two independent subunits contained in these crystals.

The initial atomic model was refined using 2.8 Å resolution data with X-PLOR (*23*), employing grouped NCS restraints. Parallel refinements were initially carried out using diffraction data from both native data sets. Because these test refinements yielded similar residuals, all further work was carried out using native set I for which higher-resolution data had been measured. Five percent of the data was reserved for the calculation of R_{free} (*24*). Rigid-body refinement of the model followed by simulated annealing at 3000 K and application of a bulk solvent correction yielded a model with a crystallographic *R*-factor of 29.2% and an R_{free} of 33.2% for all data to 2.8 Å resolution. Iterative rounds of model building and least-squares refinement followed by resolution extension and gradual removal of NCS restraints resulted in a final model at 2.1 Å resolution with a crystallographic *R*-factor of 21.1% ($R_{free} = 27.7\%$) (Table 2).

In the course of building the pyrophosphorylase domain atomic model, it became apparent that the conformations of the two pyrophosphorylase domains were not identical and that just one active site contained a metal ion. This ion was initially modeled as Mg^{2+} on the basis of its concentration

Table 2: Structure Refinement Statistics[a]

resolution (Å)	2.1
no. of protein atoms	6762
no. of substrate atoms	174
no. of waters	259
no. of cobalts	5
rms deviations from ideality	
bond lengths (Å)	0.008
bond angles (deg)	1.3
dihedral angles (deg)	27.9
temperature factors (Å^2)	
mean (all atoms)	21.6
mean (CoA and UDP-GlcNAc)	28.2
R-factor (%)	21.1 (27.7)
R_{free} (%)	24.8 (32.2)

[a] *R*-factor (%) $= \Sigma|F_o - F_c|/\Sigma|F_o| \times 100$ for all available data, but excluding data reserved for the calculation of R_{free}. R_{free} (%) $= \Sigma|F_o - F_c|/\Sigma|F_o| \times 100$ for a 5% subset of X-ray diffraction data omitted from refinement calculations. Values in parentheses refer to the corresponding statistic calculated for data in the highest-resolution bin (2.2−2.1 Å).

in both crystallization solutions (14 mM) and the synthetic storage liquor used to harvest crystals (28 mM) and the fact that Mg^{2+} is thought to be the preferred physiological activator for the pyrophosphorylase and the preference of Mg^{2+} for octahedral coordination geometry. However, subsequent refinement of the model resulted in unusually low thermal factors for this ion, suggesting that Mg^{2+} was not sufficiently electron rich to account for this region of electron density. The only apparent alternative to Mg^{2+} was Co^{2+}, which was included as an additive in crystallization experiments (2−10 mM) and also in synthetic storage liquors (3 mM).

In an attempt to confirm the identity of the metal ion, native I data set was reprocessed without merging of the Bijvoet pairs to retain the observed differences in diffraction intensity due to anomalous scattering. Cobalt has a relatively strong anomalous signal at the Cu Kα wavelength used in this study ($\Delta f'' = 3.61$ e$^-$), while the anomalous signal due to magnesium is negligible at this wavelength ($\Delta f'' = 0.18$ e$^-$). These reprocessed data were used to calculate three anomalous difference Fourier maps using either the original MIR phase set or phase sets from the working model with or without inclusion of the active site Mg^{2+} ion. The expectation was that a significant peak on these maps at the metal ion position would be observed only if it were a strong anomalous scatterer at this wavelength. Irrespective of the phase set that was used, the strongest observed peak in these anomalous difference Fourier maps (approximately 12σ) was indeed centered at a position corresponding to the coordinating metal ion, identifying Co^{2+} as the true occupant. Refinement of the model with Mg^{2+} replaced with Co^{2+} yielded thermal factors for this Co^{2+} ion which were more similar to those of surrounding atoms, additionally supporting this identification of the bound metal ion. The ability of Co^{2+} to activate pyrophosphorylases containing the characteristic pyrophosphorylase sequence fingerprint suggests that Co^{2+} may bind in a manner similar to that of Mg^{2+} (*4*).

RESULTS AND DISCUSSION

Atomic Model. The atomic model of *E. coli* GlmU contains residues Asn3−Pro452 and Asn4−Pro452 for the two crystallographically independent subunits. The model also includes two copies of UDP-GlcNAc and CoA, 259 water molecules, and five Co^{2+} ions. This model has good geometry, with 89% of the residues in the most favored portion of the Ramachandran plot and the remainder in additionally allowed regions as determined by the program PROCHECK (*25*).

Domain Organization, Overall Polypeptide Chain Fold, and Crystal Packing. The asymmetric unit of these crystals contains two structurally nonidentical subunits of GlmU. Each of these two subunits is situated near a crystallographic 3-fold axis which generates two distinct trimeric copies of the enzyme, each formed from just one of these subunit types. The trimeric quaternary structure of GlmU is thought to be the native structure of the enzyme on the basis of gel filtration chromatography, which indicates a mass of approximately 150 kDa (data not shown), and the sequence similarity of its C-terminal acetyltransferase domain to other trimeric hexapeptide acyltransferases.

The overall polypeptide chain fold is similar to that reported by Brown et al. for the truncated enzyme (*1*), with the trimeric subunit interface formed by three copies of the C-terminal LβH domain (residues Phe260−Ala437) arranged parallel to a crystallographic 3-fold axis. This domain is connected to the N-terminal pyrophosphorylase domain (residues Asn3−Arg229) by a long 21-residue connector helix (residues Leu230−Ala250) oriented nearly perpendicular to the 3-fold axis of the enzyme. The result is a trimeric molecule with a mushroom-like appearance consisting of a left-handed β-helical stalk connected to a globular pyrophosphorylase domain cap (Figures 1 and 2).

The C-terminal domains of the two crystallographically independent trimers pack base-to-base to form a hexameric aggregate with approximate 32-point symmetry. Additional crystal packing interactions involve the N-terminal domains interacting with one another. These interactions lead to small but significant hinge bending rotations of these domains with respect to their C-terminal domains. These differences appear to prevent the conformational change in one subunit that is required for Co^{2+} binding. Because of this, one crystallographically independent subunit contains Co^{2+} at its uridyltransferase active site and the other does not. The distance between the uridyltransferase active site Co^{2+} and the closest CoA thiol group is 50 Å, in agreement with the proposal that the two active sites of GlmU operate independently and without substrate channeling (*3*, *11*).

Structure of the Pyrophosphorylase Domain. The overall structure of the N-terminal pyrophosphorylase domain is that of a seven-stranded mixed β-sheet surrounded by six α-helices and a two-stranded β-sheet sitting atop the larger mixed sheet (Figure 2).[2] The active site pocket is bounded by two lobes. The first lobe (residues Asn3−Val111 and His216−Asn227) includes strands β1−β4 which interact primarily with the nucleotide moiety of UDP-GlcNAc. The second lobe (residues Glu112−Val215) includes strands β5−β7, which interact primarily with the GlcNAc portion of the

[2] The following secondary structure nomenclature is used for the pyrophosphorylase domain of *E. coli* GlmU (*1*). According to this nomenclature, the residue ranges for β-strands are 6−12 (β1), 50−55 (β2), 72−75 (β3), 98−103 (β4), 126−132 (β5), 142−145 (β5a), 148−153 (β5b), 165−176 (β6), and 212−215 (β7). The residue ranges for α-helices are 34−45 (α1), 59−65 (α2), 82−87 (α3), 111−118 (α4), 177−185 (α5), and 201−208 (α6).

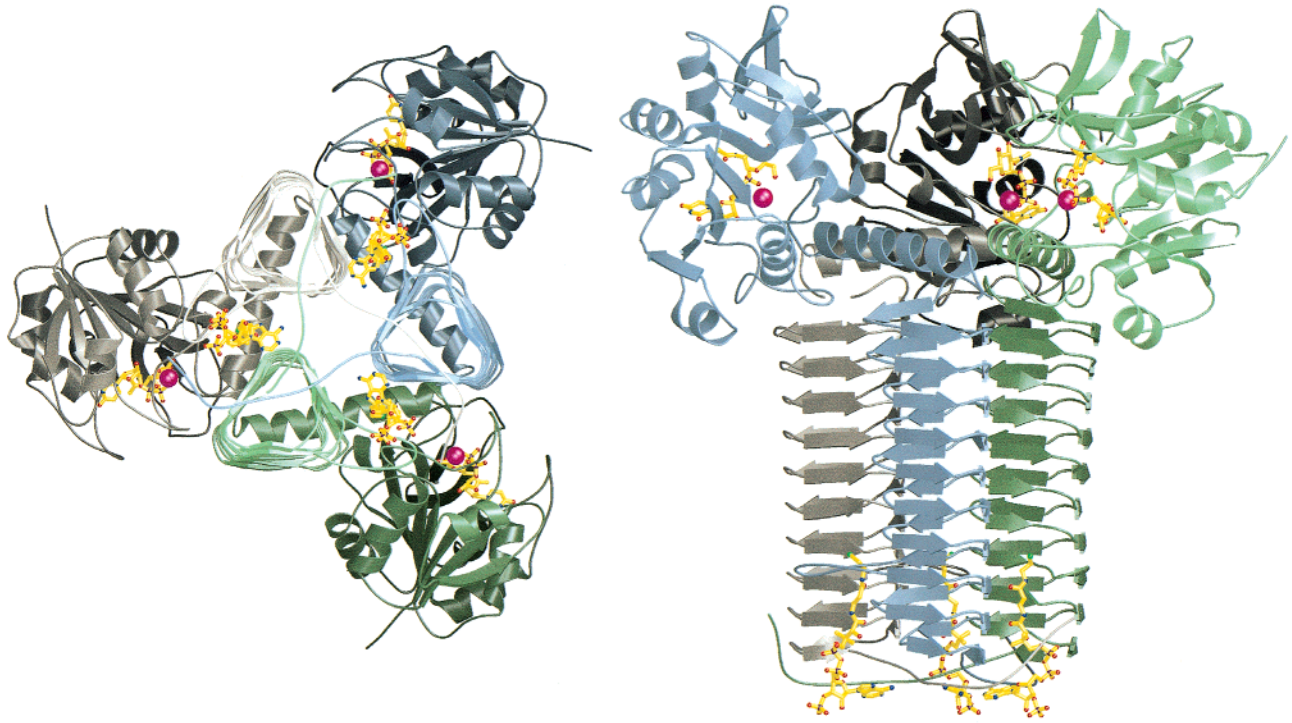

FIGURE 1: Ribbon diagrams of the polypeptide chain conformation of trimeric GlmU (left) viewed parallel to the molecular 3-fold axis and from the C-terminal end of the LβH and (right) viewed perpendicular to the molecular 3-fold axis. CoA and UDP-GlcNAc as well as the active site Co^{2+} ion are depicted. This figure was prepared with MOLSCRIPT (*30*) and Raster3D (*31*).

sugar nucleotide. Strands β5a and β5b form a sheet which encloses one end of the active site pocket.

Pyrophosphorylase Active Site. Superposition of the two pyrophosphorylase domains guided by the C_α positions of the central β-sheet indicates that a conformational change has occurred on metal binding (Figure 3). Although this conformational change leaves most of the central β-sheet unchanged, large portions of the first lobe of the domain (residues Asn3−Val111 and His216−Asn227) move nearly 3 Å toward the bound UDP-GlcNAc in the metal-containing active site. Such a movement may be prevented in the metal-free active site since the first lobe of this domain forms crystal contacts with the static second lobe of the metal-containing pyrophosphorylase domain. A particularly mobile polypeptide loop of the first lobe inserted between β1 and α1 (residues Ala13−Lys33) contains much of the L-X_2-G-X-G-T-X-M-X_4-P-K conserved pyrophosphorylase sequence fingerprint in the residue range of Leu11−Lys25. The overall change in the conformation of the sugar nucleotide involves a 1 Å translation of the uridine nucleoside portion of UDP-GlcNAc that interacts with the first domain lobe, but little change in the location or conformation of the GlcNAc moiety that interacts primarily with the static second lobe.

The coordination of the Co^{2+} ion is created by a bidentate interaction utilizing a nonbridge oxygen from each of the α- and β-phosphates of UDP-GlcNAc (metal ion distances of 2.3 and 2.3 Å), the side chain oxygen atoms of two conserved residues, Asp105 (2.1 Å) and Asn227 (2.2 Å), and two well-ordered water molecules (2.4 and 2.5 Å) (Figure 4). These groups complete a geometrically precise octahedral coordination sphere. There are several changes in the pattern of interactions to the sugar nucleotide as a result of Co^{2+} binding to the pyrophosphorylase active site. Asp105 moves 2.4 Å to coordinate the metal ion and no longer forms a hydrogen bond with the ribose 3′-hydroxyl group, which instead interacts with the peptide nitrogen of Gly104 and the peptide carbonyl oxygen of Leu11. The phenolic hydroxyl of Tyr103 moves 2.7 Å on metal binding and donates a hydrogen bond to the acetyl carbonyl oxygen of GlcNAc, no longer forming a hydrogen bond with O5 (ring oxygen) of GlcNAc, which instead interacts with a water molecule. Three of the oxygen atoms in the metal ligand sphere (an α-phosphate nonbridge oxygen and a single side chain oxygen atom from both Asp105 and Asn227) interact with the side chain ammonium group of Lys25, a conserved residue of the pyrophosphorylase sequence fingerprint (residues Leu11−Lys25). The side chain of this lysine residue is in contact with the side chain of the conserved Met19. Two of the metal ligand oxygen atoms are provided by water. These two water molecules, as well as one additional water molecule, hydrogen bond to the primary 6-hydroxyl of the GlcNAc moiety. Metal binding imposes direct constraints on the structure of UDP-GlcNAc involving the torsion angles of the phosphate groups. In the Co^{2+}-free active site, the nonbridging oxygens take on a staggered conformation, whereas in the cobalt-containing active site, they are eclipsed without short contacts between nonbridge oxygen atoms as is required by the bidentate interaction with the metal ion.

It is interesting that the repositioning of UDP-GlcNAc and the mobile polypeptide loop upon Co^{2+} binding results in a structure consistent with work involving mutagenesis of Gly14 to alanine (*1*). The K_m values for GlcNAc-1-PO_4 and UTP for this mutant are nearly unchanged relative to that of the native enzyme, despite modeling work with the structure of the truncated and metal-free enzyme which suggested that

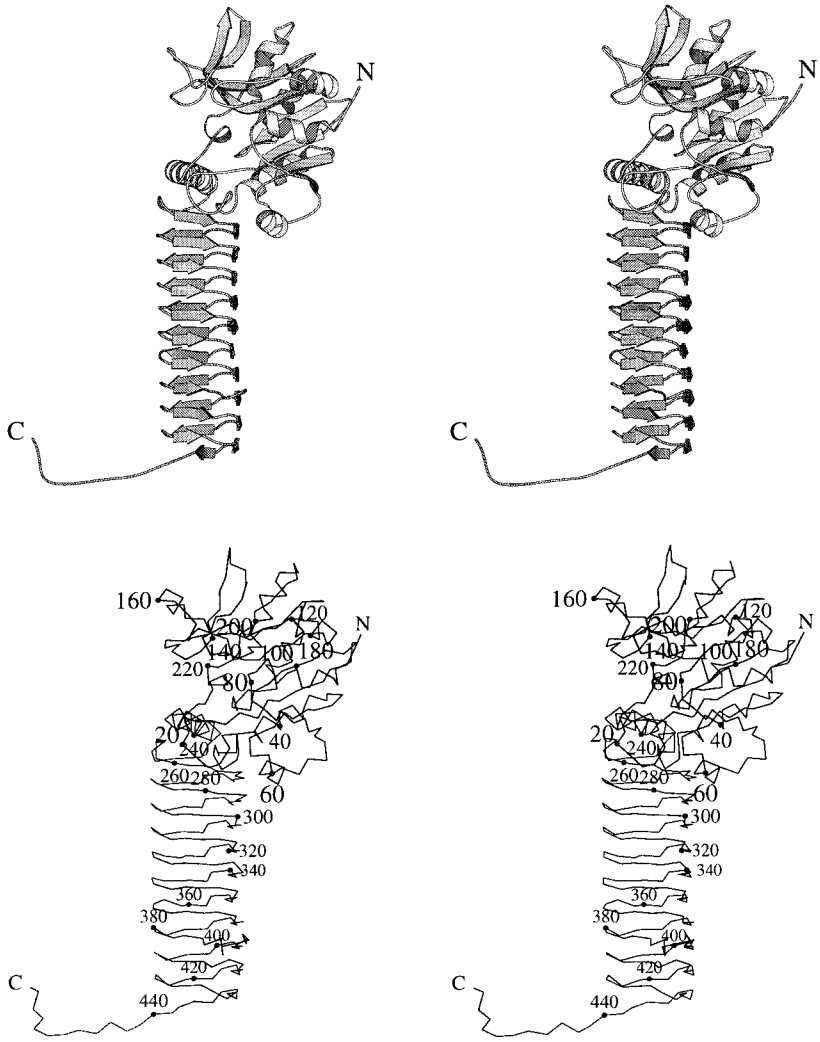

FIGURE 2: Divergent stereoview diagrams of a GlmU subunit. (Top) Ribbon diagram. (Bottom) Numbered α-carbon representation.

a steric conflict of the alanine methyl group with the ribose 3′-hydroxyl would result. For the structure of the active site containing metal reported here, no such steric conflict is implied due to a 1.6 Å shift in the C_α of this residue and a 0.9 Å shift in the position of the 3′-hydroxyl group of the ribose on metal binding.

A large cavity is bounded on one side by the two remaining unliganded nonbridge oxygen atoms of UDP-GlcNAc and on the other by the side chains of Asp193 and Asn195, two residues that have been previously proposed to interact with a metal ion (*1*). This hypothesis remains reasonable in light of the pyrophosphorylase active site conformation induced by Co^{2+} binding reported here. Indeed, this pocket has the proper proportions based on modeling experiments to accommodate both pyrophosphate and a second metal ion and appears to be the only such region proximal to the UDP-GlcNAc binding site of sufficient volume.

Structure of the Acetyltransferase Domain. The C-terminal portion of *E. coli* GlmU is dominated by the appearance of a long LβH composed of 11 complete or partial coils (residues Phe260−Ala437). This domain is the longest LβH observed to date. This structural domain corresponds to the imperfect tandem-repeated hexapeptide sequence and folds as a large coiled prism as if the polypeptide chain were wound in a left-handed spiral around the surface of an equilateral prism (Figure 2). The simplified hexapeptide repeat rule of [LIV]-[GAED]-X_2-[STAV]-X is composed of six residue types that are denoted *i*, $i + 1$, $i + 2$, $i + 3$, $i + 4$, and $i + 5$. An idealized coil of the LβH domain is composed of three such hexapeptide repeat units, corresponding to a canonical 18-residue coil length. The flat faces of the LβH are parallel β-sheets formed by stacks of short untwisted parallel β-strands. These strands participate in classical hydrogen bonding interactions with adjacent β-strands (*14*). The parallel β-strands, termed PB1, PB2, and PB3, are separated either by turns (T1, T2, and T3) or by an inserted loop at the T3 turn position (Figures 5 and 6).

The axis of each individual LβH is very nearly parallel to the 3-fold axis that relates subunits of the trimer. The association of the three LβH domains surrounding this 3-fold axis forms a trimeric core of the enzyme with approximate dimensions of 44 Å × 44 Å × 48 Å and accounts for the majority of intersubunit interactions. The prismatic coil of the LβH domain is interrupted once between residues Thr384 and Thr395 by a 10-residue inserted loop that projects from the T3 turn of coil C8 into the intersubunit space and adjacent to the thiol group of CoA. This inserted loop disobeys the hexapeptide repeat sequence motif rule and cannot be accommodated within the LβH fold as a result. The model also contains two Co^{2+} ions that bind on the 3-fold axis of each trimer. Each of these Co^{2+} ions receives ligands from

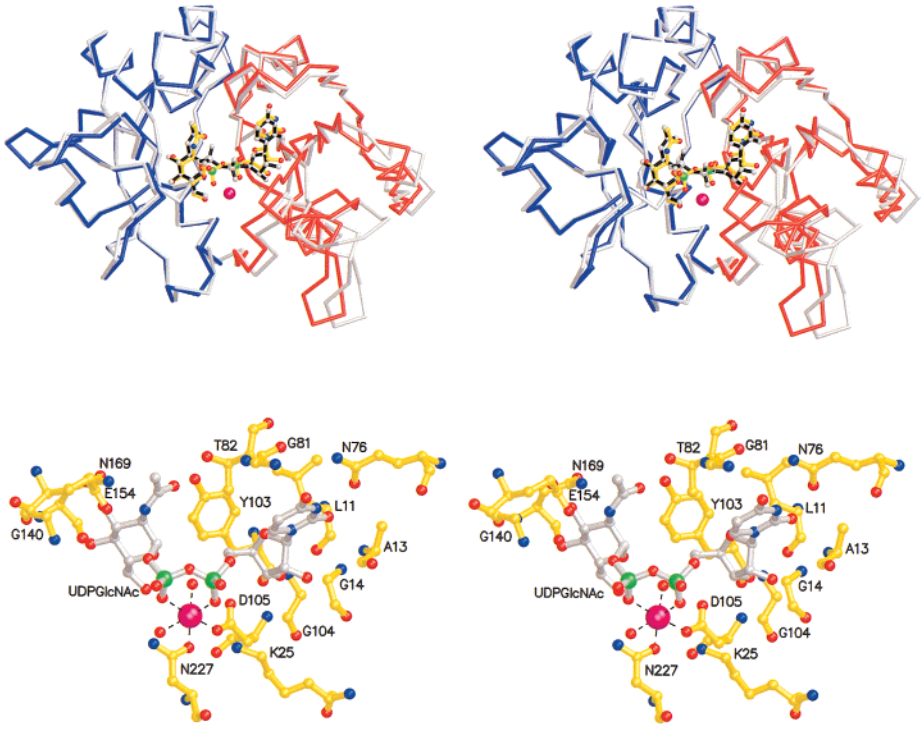

FIGURE 3: Divergent stereoview diagrams of the pyrophosphorylase domain of GlmU. (Top) Least-squares superposition of the two independent pyrophosphorylase domains. The Co^{2+}-free domain and associated UDP-GlcNAc are depicted (gray) as are the mobile first lobe (red) and the static second lobe (blue) of the Co^{2+}-bound domain. (Bottom) The active site depicting UDP-GlcNAc, Co^{2+}, and interacting residues.

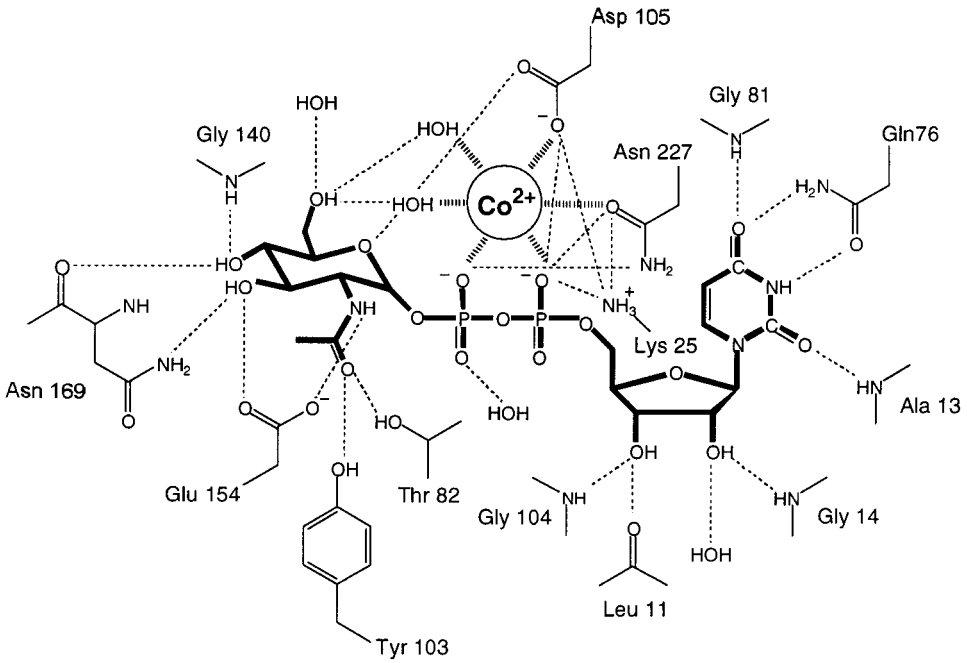

FIGURE 4: Interactions of UDP-GlcNAc at the Co^{2+}-bound pyrophosphorylase active site. Hydrophilic interactions are depicted with dashed segments. Interactions defining the octahedral coordination sphere of the metal ion are depicted as hashed segments.

three copies of the side chain carboxylate of Asp406. These ions do not appear to play any obvious catalytic role.

The frequent truncation of T3 turns marks a distinguishing feature of the LβH of GlmU. The T3 turn position is that turn from which loops are inserted in hexapeptide enzymes, providing a means by which hexapeptide acyltransferases attain structural and functional diversity in the context of the regular structure of the LβH (*26*). However, these turns are normally distinguished from the other two turn positions (T1 and T2) only at the sites of inserted loops. In contrast, five T3 turns of GlmU are shortened by one residue (coils C3, C4, C6, C7, and C10), two contain residues at $i + 2$ with flipped peptide planes that point the peptide carbonyl toward the N-terminal end of the LβH (C1 and C2), and one contains a *cis*-proline at the $i + 3$ position (C9) and one an inserted loop (C8) (Figure 6). Just one T3 turn (C5) is in

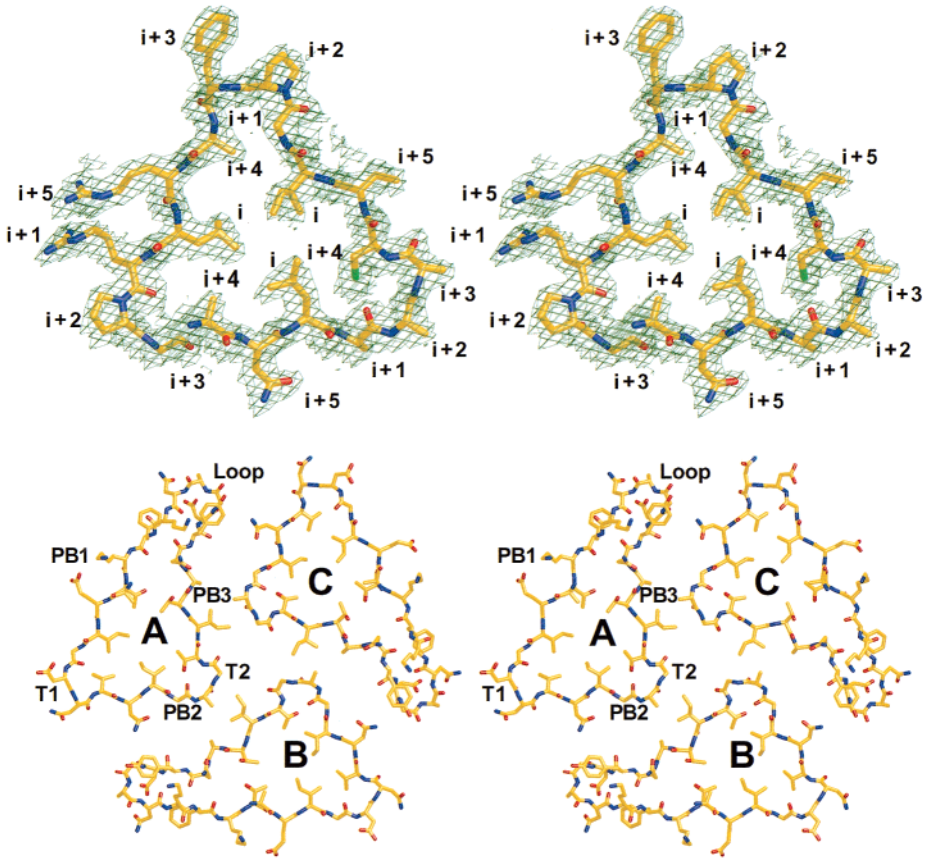

FIGURE 5: Divergent stereoview diagrams of the LβH domain. (Top) Coil C5 (residues Ala318−Gly335) with residue types (i, $i + 1$, $i + 2$, $i + 3$, $i + 4$, and $i + 5$) labeled. The density is from the final $2F_o - F_c$ map and is contoured at 1.5σ. (Bottom) The three copies of coil C8 (residues Ala370−Thr395) of the trimer viewed parallel to the molecular 3-fold axis and depicting the inserted loop region (residues Cys385−Lys394). The location of the subunits (A−C), parallel β-strands (PB1−PB3), and turn positions (T1, T2, and Loop) are indicated. For this coil, an inserted loop replaces the ordinary tight turn at this position.

		PB1				T1				PB2				T2				PB3				T3			
		i+4	i+5	i	i+1	i+2	i+3			i+4	i+5	i	i+1	i+2	i+3			i+4	i+5	i	i+1	i+2	i+3		
C1	260																	F	D	L	R	G	T		265
C2	266	L	T	H	G	R	D	V	E	I	D	T	N					V	I	I	E	G	N		283
C3	284	V	T	L	G	H	R	V	K	I	G	T	G					C	V	I	K	N			300
C4	301	S	V	I	G	D	D	C	E	I	S	P	Y					T	V	V	E	D			317
C5	318	A	N	L	A	A	A	C	T	I	G	P	F					A	R	L	R	P	G		335
C6	336	A	E	L	L	E	G	A	H	V	G	N	F					V	E	M	K	K			352
C7	353	A	R	L	G	K	G	S	K	A	G	H	L					T	Y	L	G	D			369
C8	370	A	E	I	G	D	N	V	N	I	G	A	G					T	I	T	LOOP				384
C9	395	T	I	I	G	D	D	V	F	V	G	S	D					T	N	L	V	A	P		412
C10	413	V	T	V	G	K	G	A	T	I	A	A	G					T	T	V	T	R			429
C11	430		N	V	G	E	N	A	L	A															437

FIGURE 6: Structure-based amino acid sequence alignment of the LβH domain of GlmU (residues Phe260−Ala437) identifying structurally equivalent residues in each of the 11 coils. Each horizontal line represents one complete or partial coil. The parallel β-strands that form the planar faces of the LβH are denoted PB1, PB2, and PB3, and the turns separating these strands are denoted T1, T2, and T3. The residue types within each hexapeptide unit are also indicated (i, $i + 1$, ...). The conserved hydrophobic residues at position i are boxed. Residues in the left-handed conformation (main chain torsion angle $\varphi > 0$) are depicted in bold type. The small residues at the corners of each coil at position $i + 4$ are inversely shaded. Residues Cys385−Lys394 are not included in the coiled β-helix and form the extended loop structure.

the classical LβH turn conformation similar to the turns that predominate at the T1 and T2 positions.

All of the cysteine residues of GlmU are reduced. The disulfide bond of the truncated enzyme located between Cys307 and Cys324 and within the core of the LβH domain at the $i + 4$ positions of two adjacent coils is not observed here. This disulfide bond is perhaps a consequence of the increased solvent accessibility of these two residues due to the truncation of the LβH at residue 331 or an increase in the conformational mobility of the last β-helical coil as a result of the truncation (*1*). The 1350-fold reduced activity of the Cys307 to alanine mutant reported by Mengin-Lecreulx and co-workers for the full-length enzyme remains enigmatic (*27*). The apparent loss of a hydrogen bond within the LβH from the Cys307 sulfhydryl group to the peptide carbonyl oxygen of Ala321 at the $i + 1$ position that would result from mutation of residue 307 to alanine is difficult to invoke as an explanation for this loss of activity since the chromatographic behavior and trimeric quaternary structure of this mutant protein were unaltered, and in any event, the acetyltransferase active site for the enzyme resides solely outside the LβH (see below). The distance between the sulfur atoms of Cys307 and the closest CoA thiol group of the trimer is 23 Å.

Acetyltransferase Active Site. The position of CoA bound to the C-terminal domain defines the location of the acetyltransferase active site at a cleft between the LβH domains of two adjacent subunits (Figure 1).[3] CoA is bound in a manner similar to that observed for DapD and PaXAT, two hexapeptide acyltransferases in which the pantetheine arm of the cofactor is arranged in an extended conformation and directed parallel to the 3-fold axis of the trimer (*17*, *26*).

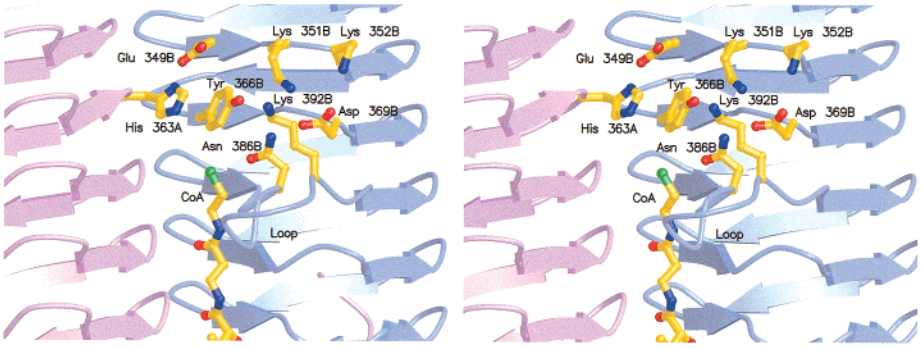

FIGURE 7: Divergent stereoview diagrams of the acetyltransferase active site depicting the thiol group of CoA and residues from the A subunit (left) and B subunit (right).

In particular, the pantetheine carbonyl groups of the cofactor bound to these enzymes receive a hydrogen bond from the backbone amide groups of $i + 2$ residues in the terminal and penultimate T2 turns of the LβH domain. The cofactor bound to GlmU makes similar interactions, with its pantetheine carbonyl groups accepting hydrogen bonds from the backbone amide groups of Ser405A and Ala423A. Distance matching between the carbonyl oxygen atoms of the pantetheine arm, separated by 5.1 Å, and the 4.9 Å distance between the parallel β-strands of adjacent LβH coils is responsible for the extended conformation of the pantetheine arm, as has been found for all hexapeptide acyltransferases studied to date. This is in contrast to the diversity of observed protein-bound CoA conformations found in the structural database (*28*).

GlmU preferentially utilizes residues from the A subunit to bind CoA in the acetyltransferase binary complex studied here. Only the hydrogen bond from the carbonyl oxygen of the inserted loop residue Tyr387B to a pantetheine amide group is contributed by the B subunit. The pyrophosphate group of the cofactor is largely solvent exposed and interacts with at least two well-ordered water molecules. The conformation of the adenine base allows its exocyclic amino group to donate a hydrogen bond to the main chain carbonyl oxygen of Arg440A. This residue also interacts directly through its side chain guanidinium group with the pantetheine hydroxyl and carbonyl groups and promotes a twist of the base plane to an orientation roughly perpendicular to the β-helical axis, in contrast to more parallel orientations previously observed. The ribose 3′-phosphate of the cofactor interacts with Lys446C in one subunit. In the other subunit, the side chain amino group of this residue is rotated out of hydrogen bonding distance, and the phosphate forms a hydrogen bond to Arg429 as a crystal contact.

A unique feature of the GlmU acetyltransferase active site described here is the involvement of a 19-residue C-terminal tail from subunit C in formation of the active site. This tail exits the LβH at residue Ile438C and extends across the bottom of the trimer and into an adjoining active site formed by the A and B subunits (Figures 1 and 2). van der Waals interactions of the side chain of Ile438C with the cofactor adenine and three hydrophilic contacts of Arg440C with the cofactor occur in the active site situated between the C and A subunits. The tail is then directed underneath the A subunit by the interactions of two glutamine residues, 443C and 445C, with peptide groups of LβH coil C11. The tail reemerges in the active site formed between the A and B subunits with an interaction of Lys446C with the cofactor ribose 3′-phosphate and contacts of the indole ring of Trp449C with a pantetheine methyl group and the side chain of Leu436A. Additional contacts of the C subunit tail with the A subunit are made by the side chain of Arg451C, which interacts with the side chain hydroxyl group of Thr420A and the backbone carbonyl oxygens of Asp400A and Gly418A.

The use of a conformationally mobile C-terminal tail to form a portion of an acyltransferase active site is not unprecedented in hexapeptide enzymes, as the active site of DapD is formed in just such a manner on binding substrates (*16*, *26*). In this case, the 18 C-terminal residues from the A subunit are disordered in the apoenzyme but form part of the active site in a ternary complex with CoA and the *N*-acyl acceptor, 2-amino-6-oxopimelate. However, GlmU is unique in its use of three subunits to form its acyltransferase active site and in the interactions of its 19-residue C-terminal tail with the cofactors of two distinct acetyltransferase active sites. The atomic model of GlmU described here lacks four C-terminal residues which are not visible in electron density maps for either subunit. Whether these disordered residues, VKKK, form part of the active site is unknown, although the participation of the extreme C-terminal sequence is consistent with the observation that a construct of *E. coli* GlmU with a C-terminal histidine-based affinity tag is 12-fold less active than the native enzyme (*27*). The use of a C-terminal tail and a possible conformational change to form the active sites of the *N*-acyltransferases DapD and GlmU is in contrast to the active site of PaXAT, a hexapeptide *O*-acetyltransferase which undergoes little if any conformational rearrangement on binding substrates in either binary or ternary complexes and makes no use of its C-terminal sequence to form the active site (*17*).

The binding site for CoA in the GlmU acetyltransferase active site binary complex is located underneath a pocket with dimensions that appear to be sufficient to accommodate both the acetyl group of acetyl-CoA and the GlcN-1-PO_4 substrate, with one side of this pocket formed by the inserted loop from the B subunit LβH (Figure 7). Several positively charged residues, including Arg333B, Lys351B, Lys352B, and Lys392B, near this pocket could serve to bind the

[3] The subunits of the trimeric enzyme are denoted A, B, and C. When viewed with the N-terminal end of the β-helix directed upward and looking into the active site, the A subunit forms the left side of the active site interface, the B subunit is on the right, and the third subunit is designated C. Residues specifically referred to a particular subunit are denoted with a lettered suffix.

phosphate group of GlcN-1-PO_4. By analogy with the structures of DapD and PaXAT, the inserted loop region (residues 385−394) might be expected to form part of the active site binding pocket in a ternary complex, although only Tyr387B participates in contacts with the cofactor in the binary complex described here.

At the top of the proposed substrate acceptor binding pocket, His363A adopts an unusual eclipsed side chain conformation ($\chi_1 = -2°$). Such a high-energy conformation may point to the functional importance of this residue as unusual side chain conformations in proteins are frequently associated with ligand and substrate binding sites (*29*). The interaction of the side chain ND1 group of His363A with Glu349B and the stacking interaction of its imidazole ring with the phenolic ring of Tyr366B orient the NE2 group toward the CoA thiol (distance of 6 Å). Both of these interactions could serve to increase the pK_a of His363A and aid its catalytic function.

ACKNOWLEDGMENT

We thank Ms. Yu Tian and Dr. John G. Arnez for help in measuring the X-ray diffraction data.

REFERENCES

1. Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C., and Bourne, Y. (1999) *EMBO J. 18*, 4096−4107.
2. Mengin-Lecreulx, D., and van Heijenoort, J. (1993) *J. Bacteriol. 175*, 6150−6157.
3. Mengin-Lecreulx, D., and van Heijenoort, J. (1994) *J. Bacteriol. 176*, 5788−5795.
4. Szumilo, T., Zeng, Y., Pastuszak, I., Drake, R., Szumilo, H., and Elbein, A. D. (1996) *J. Biol. Chem. 271*, 13147−13154.
5. Mio, T., Yamada-Okabe, T., Arisawa, M., and Yamada-Okabe, H. (1999) *J. Biol. Chem. 274*, 424−429.
6. Mio, T., Yabe, T., Arisawa, M., and Yamada-Okabe, H. (1998) *J. Biol. Chem. 273*, 14392−14397.
7. Anderson, M. S., and Raetz, C. R. H. (1987) *J. Biol. Chem. 262*, 5159−5169.
8. Varon, D., Boylan, S. A., Okamoto, K., and Price, C. W. (1993) *J. Bacteriol. 175*, 3964−3971.
9. Tiede, A., Bastisch, I., Schubert, P., Orlean, P., and Schmidt, R. E. (1999) *Biol. Chem. Hoppe-Seyler 380*, 503−523.
10. Cabib, E. (1991) in *Encyclopedia of Plant Physiology* (Pirson, A., and Zimmerman, M., Eds.) pp 395−415, Springer-Verlag, New York.
11. Gehring, A. M., Lees, W. J., Mindiola, D. J., Walsh, C. T., and Brown, E. D. (1996) *Biochemistry 35*, 579−585.
12. Vaara, M. (1992) *FEMS Microbiol. Lett. 97*, 249−254.
13. Dicker, I. B., and Seetharam, S. (1992) *Mol. Microbiol. 6*, 817−823.
14. Raetz, C. R. H., and Roderick, S. L. (1995) *Science 270*, 997−1000.
15. Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996) *EMBO J. 15*, 2323−2330.
16. Beaman, T. W., Binder, D. A., Blanchard, J. S., and Roderick, S. L. (1997) *Biochemistry 36*, 489−494.
17. Beaman, T. W., Sugantino, M., and Roderick, S. L. (1998) *Biochemistry 37*, 6689−6696.
18. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916−924.
19. Otwinoski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
20. de La Fortelle, E., and Bricogne, G. (1996) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 472−494, Academic Press, New York.
21. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. D52*, 30−42.
22. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
23. Brunger, A. T., Krukowski, A., and Erickson, J. W. (1990) *Acta Crystallogr. A46*, 585−593.
24. Brunger, A. T. (1992) *Nature 355*, 472−475.
25. Laskowski, R. A., MacArthur, M. W., Moss, S. D., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
26. Beaman, T. W., Blanchard, J. S., and Roderick, S. L. (1998) *Biochemistry 37*, 10363−10369.
27. Pompeo, F., van Heijenhoort, J., and Mengin-Lecreulx, D. (1998) *J. Bacteriol. 180*, 4799−4803.
28. Engel, C., and Wierenga, R. (1996) *Curr. Opin. Struct. Biol. 6*, 790−797.
29. Heringa, J., and Argos, P. (1999) *Proteins 37*, 44−55.
30. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
31. Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D50*, 869−873.

BI002503N