Structures of *Saccharomyces cerevisiae N*-myristoyltransferase with Bound MyristoylCoA and Peptide Provide Insights about Substrate Recognition and Catalysis†

Thalia A. Farazi,‡ Gabriel Waksman,§ and Jeffrey I. Gordon*,‡

Departments of Molecular Biology and Pharmacology, and Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received January 22, 2001; Revised Manuscript Received March 23, 2001

ABSTRACT: MyristoylCoA:protein *N*-myristoyltransferase (Nmt) attaches myristate to the N-terminal Gly residue of proteins involved in a variety of signal transduction cascades, and other critical cellular functions. To gain insight about the structural basis of substrate recognition and catalysis, we determined the structures of a binary complex of *Saccharomyces cerevisiae* Nmt1p with myristoylCoA to 2.2 Å resolution and of a ternary complex of Nmt1p with a nonhydrolyzable myristoylCoA analogue [*S*-(2-oxo)pentadecylCoA] and an octapeptide substrate (GLYASKLA) to 2.5 Å resolution. The binary complex reveals how myristoylCoA alters the conformation of the enzyme to promote binding of both myristoylCoA and peptide and identifies the backbone amides of F170 and L171 as an oxyanion hole which polarizes the reactive thioester carbonyl. The ternary complex structure reveals details of the enzyme's peptide binding specificity and illuminates its mechanism of acyl transfer. The N-terminal Gly ammonium is positioned in close proximity to the C-terminal carboxylate of the protein, where it is poised to undergo the required deprotonation to an amine. In this conformation, the nucleophile is 6.3 Å away from the thioester carbonyl. A catalytic mechanism is proposed whereby, once deprotonation is initiated, the N-terminal Gly amine can approximate the thioester carbonyl by rotating along Ψ. This motion is facilitated by a H-bond network and leads to reaction between the glycine nitrogen nucleophile and the carbonyl. Loss of CoA from the tetrahedral intermediate may be facilitated by intramolecular H-bonding of the sulfur to the adenylamine of CoA. This affords a compact leaving group and lends a role for the observed bends in the CoA structure. The absolute requirement for Gly at the N-terminus of substrates is explained by the requirement for flexible rotation of its amine.

MyristoylCoA:protein *N*-myristoyltransferase (Nmt)[1] catalyzes the co-translational covalent linkage of myristate, a 14 carbon saturated fatty acid, to the N-terminal glycine of protein substrates. These substrates play important roles in a variety of cellular processes and include proteins involved in signal transduction cascades (kinases, kinase substrates, and phosphatases) and vesicular and protein trafficking. The myristoyl group typically helps regulate interactions between an N-myristoylprotein and cellular membranes or protein partners (*1*).

Nmt is an attractive therapeutic target. Genetic studies have established that the enzyme is essential for survival of the two organisms that most commonly cause systemic fungal infections in immunocompromised humans (*Candida albicans* and *Cryptococcus neoformans*) (*2*, *3*). The divergent peptide substrate specificities of fungal and human Nmts have been exploited to design species-selective peptidomimetic inhibitors that are fungicidal (*4*). In addition, a number of enveloped and nonenveloped viruses, including HIV-1, depend on host cell Nmt-catalyzed N-myristoylation of viral proteins for assembly and/or infection.

Understanding the structural basis for this enzyme's protein substrate specificity and catalytic mechanism should not only help guide development of useful new therapeutic compounds, but will also help provide insights about structure−activity relationships in other members of the large superfamily of GCN5-related *N*-acetyltransferases (GNAT) (*5*).

Saccharmyces cerevisiae Nmt1p is the most thoroughly studied of the 20 orthologous *N*-myristoyltransferases that can be currently found in GenBank. This monomeric 455 residue enzyme has an ordered bi−bi reaction mechanism: myristoylCoA binds first followed by peptide; after catalysis, CoA is released followed by myristoylpeptide. Pre-steady-state kinetic studies have disclosed that a step after chemical transformation limits the overall reaction rate (*6*). Nmt1p is highly selective for myristoylCoA in vitro and in vivo (*7*). Approximately 1% of yeast proteins appear to be substrates for the enzyme (*8*). Known yeast N-myristoylproteins have an absolutely conserved Gly at their N-terminus (position 1). Ser and Lys typically reside at positions 5 and 6, respectively.

† This work was supported by a grant from the National Institutes of Health (AI38200).

* To whom correspondence should be addressed. Phone: (314) 362-7243. Fax: (314) 362-7047. E-mail: jgordon@molecool.wustl.edu.

‡ Department of Molecular Biology and Pharmacology.

§ Department of Biochemistry and Molecular Biophysic. E-mail: waksman@biochem.wustl.edu.

[1] Abbreviations: Nmt, myristoylCoA:protein *N*-myristoyltransferase; GNAT, GCN5-related *N*-acetyltransferase; PEG, poly(ethylene glycol).

10.1021/bi0101401 CCC: $20.00
Published on Web 05/04/2001

Nmt1p has been crystallized as a ternary complex containing two bound substrate analogues: *S*-(2−oxo)pentadecyl-CoA, which contains an additional methylene interposed between the fatty acid carbonyl and the CoA sulfur, and SC-58272, a competitive dipeptide inhibitor that retains the critical Ser and Lys recognition elements (*9*). This ternary complex structure was refined to 2.9 Å resolution (*9*). The Nmt1p fold consists of a large saddle-shaped β-sheet flanked on both sides by α-helices. The fold has a pseudo 2-fold symmetry, with each half topologically equivalent to the monomer structure of GNAT superfamily members (*1*, *5*). The N-terminal half of Nmt1p forms the *S*-(2−oxo)-pentadecylCoA binding site while the C-terminal half forms most of the peptidomimetic binding site.

A comparison of the 2.45 Å resolution structure of *C. albicans* apo-Nmt (*10*) with the 2.9 Å structure of the ternary *S. cerevisiae* Nmt1p complex suggested that the enzyme undergoes several conformational changes upon ligand binding. However, because these two structures represented the "end-states" in the pathway leading to ternary complex formation, the structure of the intermediate Nmt1p:myristoylCoA complex could only be inferred. The ternary complex structure suggested that like other acyl transfer reactions, the Nmt reaction proceeds via direct nucleophilic addition−elimination. Catalytic transfer of myristate from CoA to the peptide substrate Gly1 through this mechanism requires polarization of the thioester carbonyl to make the carbon amenable to nucleophilic attack, stabilization of transition states and reaction intermediates, and deprotonation of the glycine ammonium to generate a nucleophilic amine. The ternary complex structure indicated that polarization and transition state stabilization are likely accomplished by positioning the carbonyl in an oxyanion hole. Since the SC-58272 dipeptide inhibitor did not contain a N-terminal Gly, the mechanism of deprotonation of the glycine ammonium could not be directly assessed, nor was the structural basis of the enzyme's absolute requirement for this residue ascertained. The dipeptide inhibitor also precluded full analysis of the structural basis for Nmt's selectivity for its nascent protein substrates. We have addressed these issues by determining the structures of a binary Nmt1p:myristoyl-CoA complex and a ternary complex with a natural octapeptide substrate plus *S*-(2−oxo)pentadecylCoA.

MATERIALS AND METHODS

Crystallization. The N-terminal 33 residues of Nmt1p were disordered in the previous ternary complex structure (*9*). Kinetic studies indicate that removal of the N-terminal 35 residues of Nmt1p has no detectable effect on enzyme activity in vitro or in vivo (*11*). Therefore, $Met^{1} \rightarrow Lys^{33}$ were deleted in pBB131 (*9*) to produce pBB501. Thioesterase-free nmtΔM1-K33p (referred to below as Nmt1p) was produced in *Escherichia coli* and purified to apparent homogeneity using protocols described earlier (*9*). The enzyme was concentrated to 50−200 mg/mL in 50 mM *N*-[2-hydroxyethyl]piperazine-*N*′-[2-ethanesulfonic acid] (pH 7.4), 1 mM EDTA and 1 mM DTT. To obtain crystals of the binary Nmt1p:myristoylCoA complex, 1.5 mol of myristoylCoA were added to each mole of the purified apoenzyme. The protein concentration was adjusted to 25 mg/mL, and crystals were grown at 20 °C in hanging drops over a solution of 0.1 M ammonium acetate, 0.1 M sodium cacodylate (pH 6.4), and 20% poly(ethylene glycol) (PEG) 4000. Sizable crystals (0.6 × 0.6 × 0.6 mm) appeared within a few days. Crystals were cryoprotected by increasing the PEG concentration to 27% and adding glycerol to a final concentration of 5%. For the ternary complex, 1.5 M equiv of *S*-(2-oxo)pentadecylCoA and GLYASKLA (from the N-terminus of *S. cerevisiae* Arf2p) were added, and the protein concentration was adjusted to 25 mg/mL. Crystals were grown at 20 °C in hanging drops over 10 mM $NiCl_2$, 0.1 M sodium cacodylate (pH 6.2), and 18% PEG 2000 monomethyl ether. Crystals (0.2 × 0.2 × 0.2 mm) appeared within a few days and were cryo-protected by adding glycerol (final concentration = 25%).

Crystals of the binary complex diffracted to 2.2 Å, were in space group $P222_1$ with unit cell dimensions $a = 75$ Å, $b = 97$ Å, and $c = 142$ Å, and contained 2 molecules in the asymmetric unit. Crystals of the ternary complex diffracted to 2.5 Å, were in space group $C222_1$ with unit cell dimensions similar to those of the binary complex, and contained one molecule in the asymmetric unit. All data were processed and reduced using DENZO and SCALEPACK (*12*).

Structure Determination. Structures were solved using molecular replacement as implemented by the program AMoRe (*13*). The coordinates of the protein part of the Nmt1p:*S*-(2−oxo)pentadecylCoA:*SC-58272* complex described earlier (*9*) were used as a search model for the Nmt1p:*S*-(2−oxo)pentadecylCoA:*GLYASKLA* ternary complex. The coordinates of the protein part of the Nmt1p:*S*-(2−oxo)pentadecylCoA:*GLYASKLA* ternary complex described in this report were used as a search model for the binary complex. Nonambiguous MR solutions, with peak heights well above background, were found for both complexes [*c* value = 47; *R*-factor (after rigid body refinement) = 40 for the binary complex (2 molecules); *c* value = 48; *R*-factor (after rigid body refinement) = 43 for the ternary complex (1 molecule)]. For the binary complex, an electron density simulated annealing omit map (myristoylCoA omitted) was calculated using $F_{obs} - F_{calc}$ coefficients, and the resulting model phases yielded clear interpretable electron density for myristoylCoA. For the ternary complex, a similar map yielded excellent electron density for *S*-(2−oxo)-pentadecylCoA but patchy density for the peptide. As the refinement proceeded (see below), improved model phases resulted in improved definition of the peptide's electron density.

Refinement. The two models were refined against the data sets shown in Table 1 using conjugate gradient minimization and simulated annealing in torsion angle or Cartesian space (program CNS) (*14*). CNS parameter and topology files for *S*-(2−oxo)pentadecylCoA and myristoylCoA were obtained from the hetero-compound information center at the Uppsala software factory. The free *R*-factor was monitored as an indicator of model improvement. After conjugate gradient minimization, *B*-factor refinement was performed (program CNS). For the ternary complex, a global anisotropic *B*-factor scale correction was used. The average *B*-factors for the binary and ternary complex were 29 and 55, respectively, which is similar to the slope of the Wilson plot for both complexes (29 and 61, respectively). Individual *B*-factor refinement of the ternary complex model resulted in higher *B*-factors for peptide's atoms than for protein residues that

Table 1: Data Collection and Refinement Statistics

Nmt1p complex	binary	ternary
data collection		
resolution	30−2.2 Å	30−2.5 Å
total/unique reflections	341253/53112	225123/18335
completeness (%)[a]	91.8	94.7
R_{sym} (%)[b]	4.8	5.4
refinement		
resolution (Å)	30−2.2	30−2.5
no. of reflections[c]	46 470	16 703
total no. of atoms	7263	3652
R-factor/R_{free} (%)[d]	23.6/26.9	24.2/28.8
rmsd[e]		
bonds (Å)	0.007	0.008
angles (deg)	1.4	1.6
B values (Å^2)	1.1/1.5	1.2/1.6

[a] Completeness for $I/\sigma(I) > 1.0$. [b] $R_{sym} = \sum|I - \langle I\rangle|/\sum I$, where I = observed intensity, and $\langle I\rangle$ = average intensity from multiple observations of symmetry related reflections. [c] Numbers reflect the "working set" of reflections at $F/\sigma(F) > 2.0$. [d] R_{free} was calculated on the basis of 4.7 and 4.6% of the total number of reflections randomly omitted from the refinement for the binary and ternary complex, respectively. [e] Root-mean-square (rms) deviation from ideal bond lengths and angles and rms deviation in *B*-factors of bonded atoms (main chain/side chains).

form the peptide binding site. Hence, the occupancy for the peptide was adjusted to 0.7, resulting in peptide *B*-factors that were similar in magnitude as those of the protein. After bulk solvent correction, the *R*-factor, and free *R*-factor are 23.6 and 26.9%, respectively, for the binary complex (30−2.2 Å), and 24.2 and 28.8% for the ternary complex (30−2.5 Å), with good stereochemistry (Table 1).

The current model for the binary complex contains two Nmt1p molecules in the asymmetric unit (each comprising 422 residues spanning Ala34 to Leu455), two myristoylCoA molecules, and 272 well-ordered waters. Electron density was poor for 14 side chains located on the surface of the protein: these were built as alanines. All residues are in the allowed regions of the Ramachandran plot. The current model for the ternary complex contains the same 422 residues, *S*-(2−oxo)pentadecylCoA, GLYASKLA (residue 7 built as an alanine), 139 well-ordered waters, and two Ni atoms. Electron density was poor for 14 side chains at the surface of the protein, which were built as alanines. All residues in the ternary complex are in the allowed regions of the Ramachandran plot. The coordinates of the binary and ternary complexes have been deposited in the Protein Data Bank (accession codes 1IIC and 1IID, respectively).

RESULTS AND DISCUSSION

Nmt1p:myristoylCoA Binary Complex. The overall structure of the *S. cerevisiae* Nmt1p:myristoylCoA binary complex is similar to that of the *C. albicans* apo-enzyme (Figure 1A). There are two notable exceptions. MyristoylCoA binding is associated with an ordering of residues 34−55, forming a 3_{10} helix, A′, that functions as an important part of the myristoylCoA binding site. MyristoylCoA binding also influences the conformation of a loop connecting helix A with strand b (the Ab loop). A conformational change in this loop helps accommodate both of the enzyme's substrates (see below).

Figure 1, panels B and C, show the myristoylCoA binding site, and summarize the contacts between this substrate and Nmt1p. The four bends in bound myristoylCoA give it an overall conformation of a question mark. H38, K39, F40, and W41 of the 3_{10} A′ helix, as well as its positive macroscopic dipole, contact the 3′ phosphate of CoA. The CoA pyrophosphate H-bonds with amide nitrogens and side chain atoms of residues in the eC loop and αC helix, generating the first bend of myristoylCoA. N8p and C6p of pantetheine are positioned within 4 Å of E105 and Y103 in the flexible Ab loop. These interactions, together with a water-mediated H-bond between O5p of pantetheine and the N6a amine of adenine, produce a second bend in myristoylCoA that directs C5p through S1p of pantetheine toward the active site. The N6a amine also is within H-bonding distance (3.4 Å) to the CoA sulfur. The thioester carbonyl of myristoylCoA is H-bonded to the backbone amides of F170 and L171. These two residues, positioned in a bulge in βe (Figure 1B), form the oxyanion hole that polarizes the thioester carbonyl. A third bend, located at this carbonyl, directs the proximal portion of the acyl chain around adenine toward αC. A fourth bend, located at C6m of myristate, is produced by the positioning of the side chains of W41 (3_{10} A′ helix), I168 (βe), I187 (αC), and Y204 (βf). This fourth bend directs C7−C14 of myristate into a deep interior pocket formed, in part, by the W41 indole. The floor of the acyl chain binding pocket is formed by side chains from V194, A202, and F425, and main chain atoms from T191, W200 and H201.

The conformation of myristoylCoA in the binary complex is very similar to the conformation of *S*-(2−oxo)pentadecylCoA in the ternary complex described in this paper (Figure 1D) and in our previously reported ternary complex with substrate analogue inhibitors (*9*). The carbonyl group and the sulfur link of myristoylCoA and *S*-(2−oxo)pentadecylCoA are comparably positioned. The alignment of these groups causes the CO−CH_2−S methylene to pucker laterally, allowing the carbonyl to be positioned within the oxyanion hole formed by F170 and L171. The ω-terminal methyl of myristoylCoA is positioned within 0.8 Å of its location in bound *S*-(2−oxo)pentadecylCoA (Figure 1D). These findings are consistent with earlier isothermal titration calorimetric (ITC) studies that showed that myristoylCoA and *S*-(2−oxo)-pentadecylCoA bind to apo-Nmt1p with similar thermodynamic signatures (*15*).

As mentioned above, the bent conformation of CoA places its sulfur within 3.4 Å of its N6a amine. This feature may allow formation of an intramolecular H-bond that stabilizes the CoA thiol formed during the nucleophilic addition−elimination reaction, and facilitates its release from the enzyme as a compact globular product (see below).

Our conclusion, derived from a comparison of the Nmt1p:myristoylCoA binary complex structure with the *C. albicans* apo-Nmt structure, that myristoylCoA binding induces formation of a N-terminal 3_{10} helix A′, is consistent with the results of a recent stopped flow fluorescence study (*6*). This study disclosed that myristoylCoA binding to apo-Nmt1p occurs in a two-step process with a fast phase and a slow phase. The fast phase is most likely diffusion controlled. The slow phase is independent of myristoylCoA concentration, and most likely represents the rate of a conformational change. Formation of the N-terminal 3_{10} helix should be reported by its residue W41 which makes contact with the 3′ phosphate of CoA and the acyl chain of myristate.

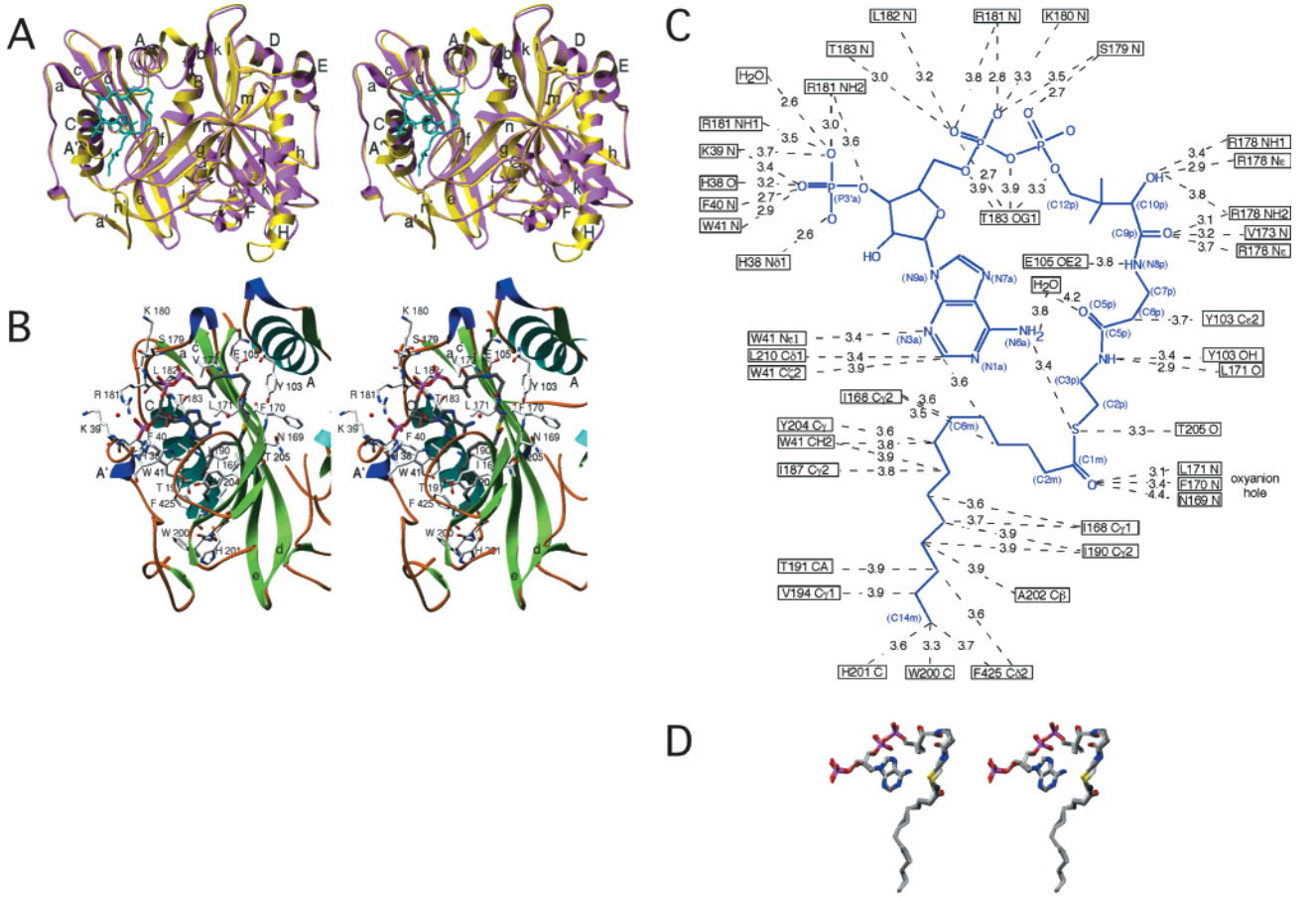

FIGURE 1: Structure of a binary Nmt1p:myristoylCoA complex. (A) Ribbon diagram comparing the binary *S. cerevisiae* Nmt1p:myristoylCoA complex (yellow) with the 2.45 Å structure of *C. albicans* apo-Nmt (magenta). Note the two major structural differences: ordering of residues 34−55 as a 3_{10} helix (A′) in the binary complex, and the conformation of the Ab loop. Nomenclature for helices and strands are taken from ref 9. (B) MyristoylCoA binding site. Secondary structure code: α helices, aqua; 3_{10} helices, deep blue; *β* strands, green; loops, amber. Atom color code: oxygen, red; phosphorus, pink; carbon, silver; nitrogen, purple; sulfur, yellow. The carbon atoms of the ligand are dark gray. (C) Summary of contacts between myristoylCoA and Nmt1p. Contact distances are in angstroms. Atoms in adenosine and phosphates are denoted with an "a" suffix, while atoms in pantetheine and myristate are denoted with a "p" and "m". (D) Superimposition of the structures of myristoylCoA from the binary complex (back) and *S*-(2-oxo)pentadecylCoA from the Nmt1p:*S*-(2-oxo)pentadecylCoA:GLYASKLA ternary complex (front). Atom color code as in panel B.

Comparison of the *C. albicans* apo-enzyme and *S. cerevisiae* binary complex reveals a second conformational change involving the Ab loop. In the *C. albicans* apo-enzyme, E109 to D112 of the Ab loop (equivalent to E105 to D108 in *S. cerevisiae* Nmt1p) obstruct the peptide binding site as well as the myristoylCoA binding site: the side chain of E109 is positioned where the C6p of myristoylCoA is located in the binary *S. cerevisiae* complex, and the Cα atoms of E109 and D110 are positioned where the Leu +2 residue of the GLYASKLA substrate is positioned in the ternary complex (see below).

The Ab loop in the binary complex shows conformational heterogeneity. The two molecules in the asymmetric unit (designated A and B) have different Ab loop conformations. Both conformations have higher *B*-factors (40−62 for A; 57−74 for B) compared to the rest of the protein (average = 29). In molecule A, the Ab loop has swung away by ∼8 Å and is directed out to solvent, and the peptide binding site appears more open than in the apo-enzyme structure (Figures 1A). Molecule B has similar conformation to that observed in the Nmt1p:*S*-(2−oxo)pentadecylCoA:*GLYASKLA* ternary complex structure. These findings help explain why apo-Nmt is *un*able to bind peptide prior to binding of myristoylCoA (*15*) (also see discussion below of the role of myristoylCoA in forming part of the peptide binding site).

Ternary Nmt1p:S-(2-oxo)pentadecylCoA:GLYASKLA Complex. Our previously reported ternary complex structure contained a competitive dipeptide inhibitor (SC-58272; K_i = 43 nM) derived from GLYASKLS (N-terminus of the *S. cerevisiae* N-myristoylprotein, Arf2p). Although this inhibitor retained the critical Ser5 and Lys6 recognition elements, the N-terminal GLYA was replaced with a 2-methylimidizole and rigidified carbon linker (*p*-[2-methyl-1-imidazol-1-yl)-butyl]phenylacetyl), while the C-terminal Leu-Ser was represented by cyclohexylethyl (*9*). Thus, the peptidomimetic could not provide information about the positioning of the Gly ammonium/amine in the active site of Nmt1p, nor could it provide information about the features of the enzyme that define its preferences for residues 1−4 or 6−8 in this or other substrates. Therefore, we crystallized a ternary complex of Nmt1p with bound *S*-(2−oxo)pentadecylCoA and GLYASKLA. Although *S*-(2−oxo)pentadecylCoA was required to generate a stable ternary complex, we were able to model myristoylCoA from the binary complex. The results provide, for the first time, the configuration of all components in Nmt's active site.

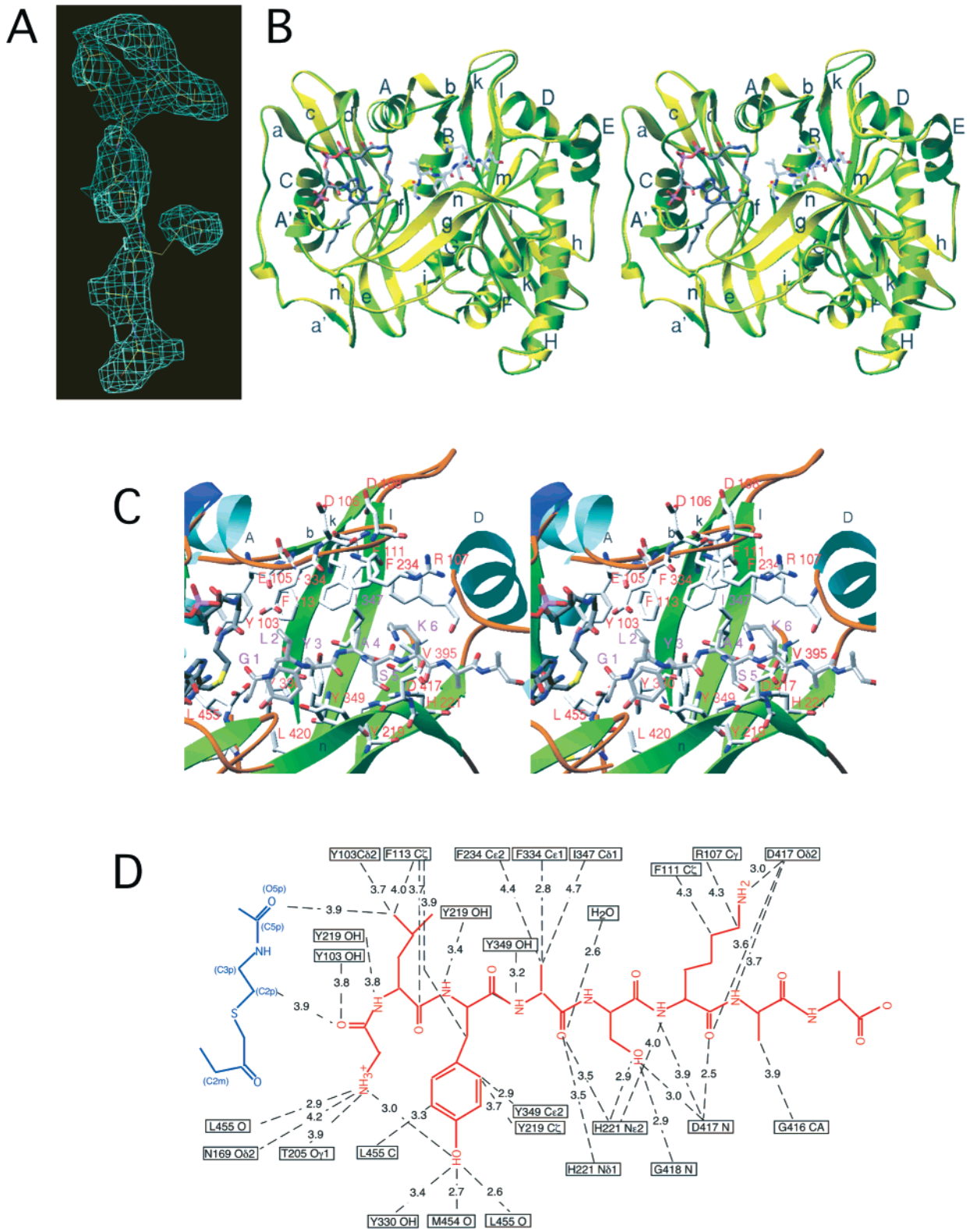

FIGURE 2: Structure of a ternary Nmt1p:*S*-(2-oxo)pentadecylCoA:GLYASKLA complex. (A) Electron density map of bound GLYASKLA. SigmaA-weighted $F_o - F_c$ omit map was generated by omitting residues within 3.5 Å radius around the octapeptide substrate. The map is contoured at 2σ. (B) Ribbon diagram of the ternary complex (green) with the binary complex (yellow). (C) GLYASKLA binding site. See Figure 1 for color codes and text for details. (D) Contacts between GLYASKLA and Nmt1p (distances in angstroms).

The overall conformation of the octapeptide is shown in Figure 2, panels A−C. Gly1 and the Leu2 side chain of GLYASKLA interact with C5p−C′—a segment of *S*-(2−oxo)pentadecylCoA that forms an integral part of the peptide binding site. This finding provides additional structural insights about why peptide does not bind to Nmt1p in the absence of myristoylCoA (*15*). The contacts between the side chain of Leu2 and C5p−C′ of *S*-(2−oxo)pentadecylCoA generate a 90° bend in the peptide backbone, turning it away from the myristoylCoA analogue and toward a peptide

binding groove (Figure 2, panels B and C). The peptide backbone from Leu 2 to Ala 4 is directed toward a large aromatic/hydrophobic patch formed by residues in the core of the C-terminal β-sheet composed of strands g, k, l, n, and j (Figure 2C). This patch is curved in such a way that at Ala 4 the path of the peptide is blocked (notably by F234), turning it toward residues in strand n and in the groove between strand n and the DE loop (Figure 2C). The bend in the peptide's backbone at Ser5 and Lys6 brings the peptide's main chain in a parallel β-sheet arrangement with strand n (Figure 2C). The configuration of the peptide in this region is stabilized by contacts with the protein detailed below, as well as by intrapeptidic H-bonds, e.g., between the carbonyl oxygen of Tyr3 and the amide nitrogen of Ser5.

Peptide: Nmt1p Interactions. (1) Features of Peptide Substrates. In vitro kinetic studies of >100 synthetic peptides derived from the N-terminal sequences of known N-myristoylproteins and purified Nmt1p (*16*) provided a set of empiric rules to search the yeast genome for putative Nmt1p substrates (*8*). Gly was absolutely required at the +1 position. Glu, Asp, Arg, Lys, His, Pro, Phe, Tyr, and Trp were not allowed at the +2 position. All possible amino acids are allowed at +3 and +4. Only Ser, Thr, Ala, Gly, Cys, or Asn were permitted at the +5. All possible residues except Pro were allowed at +6. A search of 6310 ORFs in the Saccharomyces Genome database (SGD) using PatMatch (http://genome-www.stanford.edu/Saccharomyces) yielded 71 known or putative N-myristoylproteins. Ser is most commonly encountered at position +2 (14/71), followed by Leu (13), and Asn (9). Ser (11) is also most common at +3, followed by aromatics [Phe (5), Tyr (5)], and Val (8). The +4 position (all possibilities allowed in the search) shows a preference for uncharged residues [Ser (8), Leu (7), Asn (6), and Val (6)]. Position +5 exhibits a marked preference for Ser (37/71). Despite allowing all amino acids except Pro, the +6 position shows a marked preference for Lys (16), followed by Ser (9), and Ile (8). Thus, GLYASKLA in the ternary complex provides a good representation of the N-terminal sequences in known or putative *S. cerevisiae* N-myristoylproteins. Below we describe the subsites coordinating each of the peptide's residues.

(2) Position +1. See the discussion of the catalytic mechanism in the next section.

(3) Position +2. Leu2 in GLYASKLA is positioned over the aromatic side chains of Y103 and F113 (Ab loop). It is surrounded on one side by V104, F111 (Ab loop), and F334 (βk), and on the other side by the aliphatic chain of E105 (Ab loop), and C4p−C7p of pantetheine (Figure 2, panels C and D). This environment favors binding of a hydrophobic residue at position +2. Ser may be allowed owing to the presence of at least three potential H-bonding partners: the O5p carbonyl from myristoylCoA, the main chain carbonyl oxygen of Y103, and the E105 carboxylate.

(4) Position +3. An aromatic cluster serves to accommodate Tyr3 in GLYASKLA. The Tyr interacts with Y219, Y330, and Y349 (Figure 2, panels C and D). L420 also contributes to the generally hydrophobic environment. On the basis of these observations, one would predict a marked preference for substrates with a hydrophobic/aromatic side chain at position +3. The known or putative *S. cerevisiae* N-myristoylproteins identified from the genome database search revealed that hydrophobic/aromatic residues are often present at this location, as is Ser. The Tyr-rich environment could provide H-bond interactions with a Ser hydroxyl.

(5) Position +4. Ala4 is located in a large hydrophobic/aromatic pocket containing F111, F234, F334, I347, and V395 (Figure 2, panels C and D). This pocket could accommodate a larger hydrophobic residue (Leu and Val are common at this position in yeast N-myristoylproteins).

(6) Position +5. Ser is strongly preferred at position +5 (as noted above). The Ser hydroxyl H-bonds with H221 in the Nmt1p:*S*-(2−oxo)pentadecylCoA:GLYASKLA complex, as it does in the ternary complex with bound substrate analogue inhibitors (*9*). This hydroxyl is also within H-bond distance from the backbone amides of G418 and D417 (Figure 2, panels C and D).

The importance of H221 has been demonstrated in two ways. First, a color colony sectoring assay (described in ref *17*) was performed. A plasmid encoding NmtH221Ap (H221 to Ala), was introduced into a strain of yeast containing a null allele of the endogenous *NMT1* gene plus another episome encoding wild-type Nmt1p. nmtH221Ap was not able to support survival, i.e., there was selective pressure to retain the plasmid containing wild type *NMT1* (data not shown). Second, in vitro steady state kinetic analysis of purified nmtH221Ap revealed that, compared to Nmt1p, it had a markedly increased K_m for a Ser5-containing octapeptide substrate, and a marked increase in K_i for SC-58272 (*9*).

(7) Position +6. In the ternary complex containing the SC-58272 dipeptide inhibitor, the ϵ-amino group of the retained Lys was surrounded by a negative electrostatic field formed by the side chains of D106, D108 (Ab loop), and D417 (*9*). The density for Lys6 in GLYASKLA is interrupted (Figure 2A). Lys6 was built so that it was coordinated by D417, just as it is in SC-58272. However, D106 and D108 in the Nmt1p:*S*-(2-oxopentadecylCoA):GLYASKLA complex do not make contact with this Lys (Figure 2C). The aliphatic portion of Lys6 is surrounded by the aromatic side chains of F111 and F234 (as it is in the Nmt1p:*S*-(2−oxo)-pentadecylCoA:SC58272 ternary complex), and by the aliphatic portion of R107. F111 and F234 are rotated 90° from their position in the *C. albicans* apo-Nmt structure, allowing their benzenes to approach the aliphatic portion of Lys6.

(8) Positions +7 and +8. Leu7 and Ala8 make very few contacts with Nmt1p (Figure 2, panels C and D). As a result, they are unlikely to contribute in a significant manner to binding energetics. This conclusion is supported by the results of in vitro steady-state kinetic analysis of the effects of progressive C-terminal truncations of octapeptide substrates (*16*). The open-endedness of the peptide-binding groove in this region would allow the enzyme to accommodate a nascent substrate for co-translational modification.

Unlike *S. cerevisiae* Nmt1p, there are no reports of the results of systematic analyses of the interactions of mammalian Nmts with octapeptide substrates. An alignment of the 19 known Nmts revealed that *S. cerevisiae* Nmt1p residues involved in GLYASKLA binding are almost all absolutely conserved. F334 of Nmt1p is represented by a Ser, and I347 by an Ala, in human Nmt. Both F334 and I347 form part of a hydrophobic pocket and were involved in coordination of Ala4 in GLYASKLA. Thus, we would predict that the Ala and Ser substitutions in human Nmt

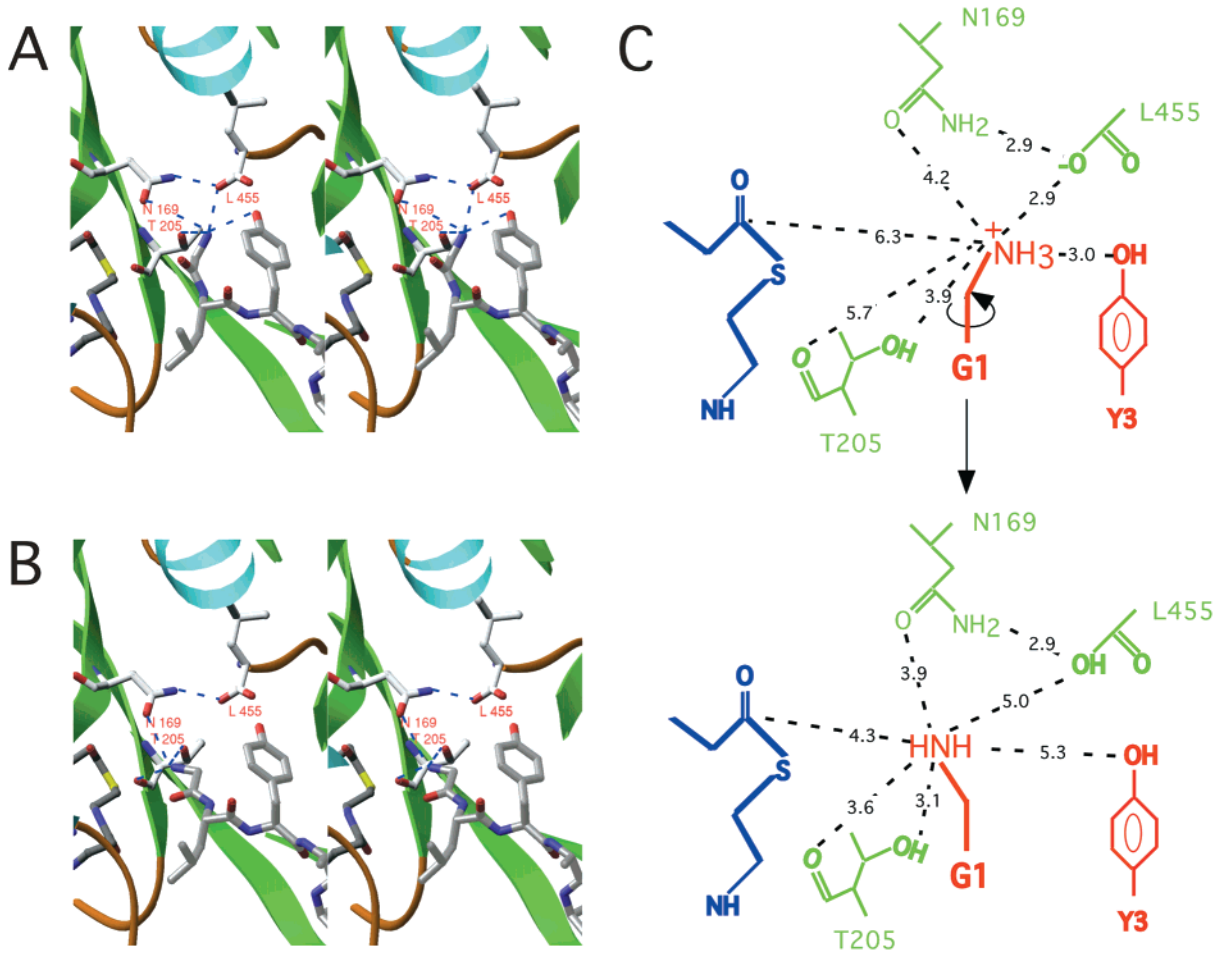

FIGURE 3: Two alternative conformers of Gly1, and the associated H-bonding networks, proposed to be involved in catalysis. (A) Gly 1 "conformer" seen in the ternary complex structure. (B) Gly1 conformer after a 180° rotation around Ψ. (C) Summary of the H-bond networks in the two conformers.

should allow it to accommodate bulkier and/or more polar residues at position +4 of its substrates.

Conformational Changes and Peptide Acquisition. Our studies now show that myristoylCoA forms part of the peptide binding site and that binding of myristoylCoA frees part of the peptide binding site by altering the conformation of the Ab loop, in agreement with the ordered bi bi reaction mechanism. In the previously reported ternary complex of Nmt1p with two bound substrate analogues, the Ab loop formed extensive contacts with *S*-(2−oxo)pentadecylCoA and the SC-58272 dipeptide inhibitor (*9*). In contrast, the Ab loop is less ordered with fewer contacts in the ternary complex containing GLYASKLA. These interactions are limited to the aliphatic portion of the side chain of E105 and the +2 Leu, plus R107 and F111 and the aliphatic portion of Lys +6 (Figure 2, panels C and D). Therefore, it is tempting to speculate that the extent of ordering of the Ab loop may correlate with the overall catalytic efficiency of different Nmt1p substrates. In this scenario, the peptidomimetic functions as a potent inhibitor in part because it forms a tight dead-end complex with extensive contacts with components of the Ab loop.

Catalytic Mechanism. Figure 3A shows the positioning of Gly1 in the active site of the ternary complex, with myristoylCoA modeled in lieu of *S*-(2−oxo)pentadecylCoA. The environment provides several H-bond interactions. The Gly1 nitrogen is 2.9 Å from the OT1 atom of L455, allowing for deprotonation. This ammonium/amine is also positioned in the proximity of the hydroxyl of Tyr3 in GLYASKLA (3.0 Å), the Oγ hydroxyl of T205 (3.9 Å; this Thr is conserved in all 20 Nmts), and the Oδ atom of N169 (4.2 Å; also conserved). The distance and angle between the Gly1 nitrogen and the thioester carbonyl is greater (6.3 Å) than would be predicted in near attack conformers (ground state conformers that resemble the transition state; see ref *18*).

If the Gly amine is allowed to rotate by ∼180° around Ψ (Figure 3B), it would be positioned so that it could readily attack the polarized carbonyl of myristoylCoA (*19*) (see earlier discussion regarding Nmt's oxyanion hole). With this rotation, the amine would no longer H-bond with the C-terminal carboxylate, but would instead form strong H-bonding interactions with N169 [Oδ (3.9 Å)] and T205 [Oγ (3.1 Å), O (3.6 Å)].

Therefore, the following mechanism involving two conformers is proposed (illustrated in Figures 3C and 4). One conformer, seen in the ternary structure, links the Gly1 ammonium/amine to the C-terminal carboxylate for deprotonation. The other, inferred, conformer forms after a rotation around Ψ is stabilized by additional and stronger H-bonds, and allows the neutral glycine nitrogen to approach the carbonyl group along the antibonding orbital axis, as is well-known for such reactions (*19*). In doing so, the main chain carbonyl of T205 may serve as an H-bond acceptor for the amine nucleophile. Indeed, as negative charge develops on

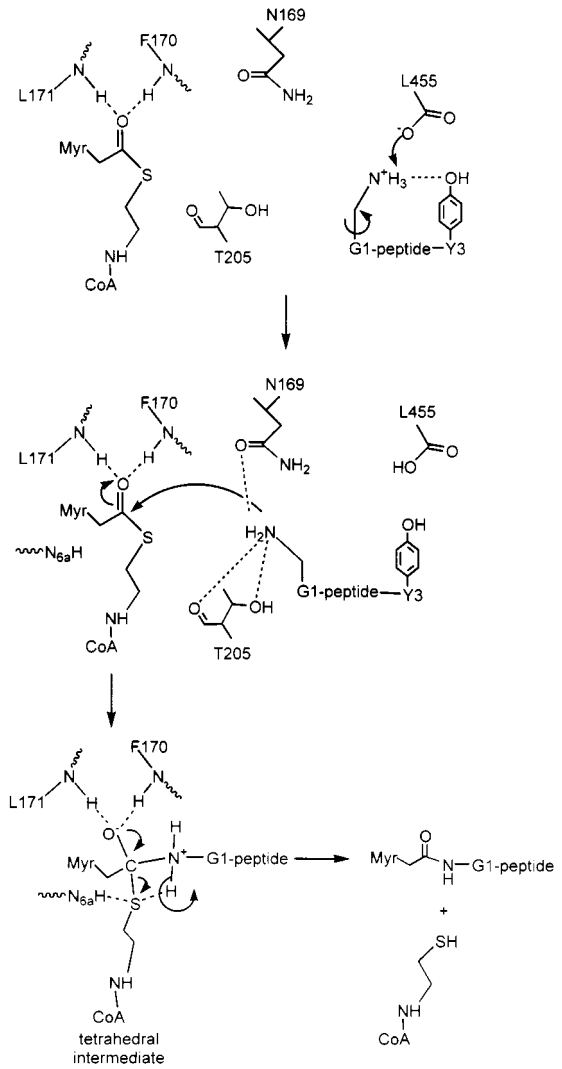

FIGURE 4: Summary of proposed reaction mechanism. See text for details.

the reactive thioester carbonyl, this H-bond may become bifurcated, thus stabilizing the developing tetrahedral intermediate (Figure 4). As the reaction proceeds, CoA is released as an anion that requires stabilization, possibly by H-bonding. It is interesting to note, that this H-bond could be provided intramolecularly by hydrogens on N6 of adenine: the ternary structure indicates that the S−N distance is 3.1 Å. The use of this intramolecular stabilization mechanism is economical and accounts for the bent conformation of CoA. An attractive feature of this mechanism is that the departing CoA is globular: its compact size would facilitate diffusion from the reactive site.

The absolute requirement for an acceptor Gly1 residue is a distinctive feature of protein N-myristoylation. Previous studies have shown that substitution of Ala for Gly at position +1 in a variety of octapeptide sequence contexts generates a competitive Nmt1p inhibitor (*16*). Although inspection of the ternary complex structure suggests that the active site of Nmt1p could accommodate Ala at this position, the required rotation of the terminal amine toward the reactive thioester carbonyl would be sterically hindered by the presence of the residue's Cβ methyl group. In addition, earlier "mutagenesis" studies of octapeptide substrates have shown that aromatic residues at position +2 are poor substrates and good inhibitors (*16*). The ternary complex structure indicates that an aromatic residue may not allow proper positioning of Gly1 so that it would be competent to perform the nucleophilic attack on the thioester carbonyl of myristoylCoA, i.e., the cluster of Y103, F111, F113, and F334 (Figure 2C) may promote tight binding of an aromatic residue limiting movement of the Gly amine. This idea is supported further by the observation that Pro substitution at position +2 also produces a poor substrate (*16*).

Other GNAT Family Members. As noted in the introductory portion of this manuscript, Nmt is a member of the GNAT superfamily. Other family members catalyze the transfer of an acetyl group from CoA to a primary amino group of various targets, including serotonin, histones, and aminoglycoside antibiotics (reviewed in ref *5*). Each of the symmetry-related halves of Nmt is topologically equivalent to the monomer core structure of GNAT superfamily members. Kinetic and structural studies of two GNATs, serotonin *N*-acetyltransferase, and histone *N*-acetyltransferase (human P/CAF and *Tetrahymena* GCN5) disclosed that they too follow an ordered reaction mechanism (*20*−*24*). The X-ray crystallographic structures of serotonin *N*-acetyltransferase (AANAT) without any bound ligands, and with a bound bisubstrate analogue inhibitor (*23*, *25*), indicate that, like Nmt, the bound acylCoA allows formation of part of the binding site for the second substrate. Comparison of the structures of the *Tetrahymena* GCN5 apo-enzyme, a binary complex of the enzyme with acetylCoA, and a ternary complex containing CoA plus a peptide derived from histone H3, show modest conformational changes upon acetylCoA binding, which might indicate a link between CoA and H3 acquisition (*24*).

The catalytic mechanism proposed for Nmt, AANAT, and histone *N*-acetyltransferase involves direct nucleophilic attack of the second substrate's primary amine on the reactive thioester carbonyl of bound acylCoA. In each case, this attack appears to be aided by polarization of the thioester carbonyl by an oxyanion hole, and a catalytic base that generates the nucleophilic amine by proton abstraction. The GNAT members use the same structural element to serve as their oxyanion hole, i.e., backbone amides in a conserved "β bulge" (*5*). However, they exploit different elements for base catalysis. As discussed above, the main chain C-terminal carboxylate appears to play this role in Nmt. On the other hand, two other members of the superfamily were found to utilize the side chain of a glutamate residue as their catalytic base. In the yeast GCN5 histone *N*-acetyltransferase, there is compelling evidence from structural and mutagenesis studies that the side chain of a glutamate in β strand 4 (E173) functions as the catalytic base, exerting its effects through an ordered water (*22*, *24*). In Esa1, the catalytic subunit of the yeast NuA4−acetylase complex that targets histone H4 for acetylation, the side chain of a glutamate (E338) located in strand 10 rather than strand 4 appears to function as the catalytic base (*26*). To date, site-directed mutagenesis studies of AANAT (ovine serotonin *N*-acetyltransferase) have not yielded a straightforward conclusion about the identity of its catalytic base (*23*).

One other interesting distinction among Nmt and other GNAT members relates to the reprotonation of the thiolate leaving group. As noted above, we propose that in Nmt N6 of the CoA adenine is involved in reprotonation. In AANAT,

this role has been tentatively assigned to Y168 (*23*).

Prospectus. The ternary structure of Nmt1p now provides molecular details of the interactions of each residue of a representative octapeptide substrate with the acetyltransferase. As such, the ternary structure represents a starting point for understanding how orthologous Nmts have diverged in their peptide substrate specificities and for designing new classes of inhibitors of this GNAT family member that may be useful therapeutically.

ACKNOWLEDGMENT

We thank Klaus Fütterer for his invaluable assistance with data collection during the beginning of this study, Jill Manchester for help with protein purification, plus George Gokel and Ying Li for valuable discussions. This work was supported by NIH grant AI38200. TF was supported in part by Medical Scientist Training Program grant GM07200.

REFERENCES

1. Bhatnagar, R. S., Ashrafi, K., Futterer, K., Waksman, G., and Gordon, J. I. (2000) in *The Enzymes: Protein Lipidation* (Tamanoi, F., and Sigman, D. S., Eds.) pp 241−290, Academic Press, Inc., San Diego, California.
2. Weinberg, R. A., McWherter, C. A., Freeman, S. K., Wood, D. C., Gordon, J. I., and Lee, S. C. (1995) *Mol. Microbiol. 16*, 241−250.
3. Lodge, J. K., Jackson-Machelski, E., Toffaletti, D. L., Perfect, J. R., and Gordon, J. I. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12008−12012.
4. Lodge, J. K., Jackson-Machelski, E., Higgins, M., McWherter, C. A., Sikorski, J. A., Devadas, B., and Gordon, J. I. (1998) *J. Biol. Chem. 273*, 12482−12491.
5. Dyda, F., Klein, D. C., and Hickman, A. B. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 81−103.
6. Farazi, T. A., Manchester, J. K., and Gordon J. I. (2000) *Biochemistry 39*, 15807−15816.
7. Johnson, D. R., Bhatnagar, R. S., Knoll, L. J., and Gordon, J. I. (1994) *Annu. Rev. Biochem. 63*, 869−914.
8. Ashrafi, K., Farazi, T. A., and Gordon, J. I. (1998) *J. Biol. Chem. 273*, 25864−25874.
9. Bhatnagar, R. S., Futterer, K., Farazi, T. A., Korolev, S., Murray, C. L., Jackson-Machelski, E., Gokel, G. W., Gordon, J. I., and Waksman, G. (1998) *Nat. Struct. Biol. 5*, 1091−1097.
10. Weston, S. A., Camble, R., Colls, J., Rosenbrock, G., Taylor, I., Egerton, M., Tucker, A. D., Tunnicliffe, A., Mistry, A., Mancia, F., de la Fortelle, E., Irwin, J., Bricogne, G., and Pauptit, R. A. (1998) *Nat. Struct. Biol. 5*, 213−221.
11. Rudnick, D. A., Johnson, R. L., and Gordon, J. I. (1992) *J. Biol. Chem. 267*, 23852−23861.
12. Otwinowski, Z. (1993) in *Proceedings of the CCP4 study weekend: data collection and processing* (Sawyers, L. I., and Bailey, S., Ed.) pp 56−62, SERC Daresbury Laboratory, Warrington, U.K.
13. Navaza, J. (1994) *Acta Crystallogr., Sect. A 50*, 157−163.
14. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
15. Bhatnagar, R. S., Schall, O. F., Jackson-Machelski, E., Sikorski, J. A., Devadas, B., Gokel, G. W., and Gordon, J. I. (1997) *Biochemistry 36*, 6700−6708.
16. Towler, D. A., Gordon, J. I., Adams, S. P., and Glaser, L. (1988) *Annu. Rev. Biochem. 57*, 69−99.
17. Zhang, L., Jackson-Machelski, E., and Gordon, J. I. (1996) *J. Biol. Chem. 271*, 33131−33140.
18. Bruice, T. C., and Benkovic, S. J. (2000) *Biochemistry 39*, 6267−6274.
19. Burgi, H. B., Dunitz, J. D., Lehn, J. M., and Wipff, G. (1974) *Tetrahedron 30*, 1563−1572.
20. De Angelis, J., Gastel, J., Klein, D. C., and Cole, P. A. (1998) *J. Biol. Chem. 273*, 3045−3050.
21. Lau, O. D., Courtney, A. D., Vassilev, A., Marzilli, L. A., Cotter, R. J., Nakatani, Y., and Cole, P. A. (2000) *J. Biol. Chem. 275*, 21953−21959.
22. Tanner, K. G., Langer, M. R., and Denu, J. M. (2000) *Biochemistry 39*, 11961−11969.
23. Hickman, A. B., Namboodiri, M. A., Klein, D. C., and Dyda, F. (1999) *Cell 97*, 361−369.
24. Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, C. D., and Marmorstein, R. (1999) *Nature 401*, 93−98.
25. Hickman, A. B., Klein, D. C., and Dyda, F. (1999) *Mol. Cell 3*, 23−32.
26. Yan, Y., Barlev, N. A., Haley, R. H., Berger, S. L., and Marmorstein, R. (2000) *Mol. Cell 6*, 1195−205.

BI0101401