11-30-2018

A Gcn5-Related N-Acetyltransferase (GNAT) Capable of Acetylating Polymyxin B and Colistin Antibiotics in Vitro

Mateusz P. Czub
University of Virginia

Brian Zhang
San Francisco State University

M. Paul Chiarelli
Loyola University Chicago

Karolina A. Majorek
University of Virginia

Layton Joe
San Francisco State University

See next page for additional authors

Recommended Citation

Czub, Mateusz P.; Zhang, Brian; Chiarelli, M. Paul; Majorek, Karolina A.; Joe, Layton; Prebski, Pryemyslaw J.; Revilla, Alina; Wu, Weiming; Becker, Daniel P.; Minor, Wladek; and Kuhn, Misty L.. A Gcn5-Related N-Acetyltransferase (GNAT) Capable of Acetylating Polymyxin B and Colistin Antibiotics in Vitro. Biochemistry, 57, 51: 7011-7020, 2018. Retrieved from Loyola eCommons, Chemistry: Faculty Publications and Other Works, http://dx.doi.org/10.1021/acs.biochem.8b00946

A Gcn5-related N-acetyltransferase (GNAT) capable of acetylating polymyxin B and colistin antibiotics *in vitro*

Mateusz P. Czub[#1,2], Brian Zhang[#3], M. Paul Chiarelli[4], Karolina A. Majorek[1,2], Layton Joe[3], Przemyslaw J. Porebski[1,2], Alina Revilla[3], Weiming Wu[3], Daniel P. Becker[4], Wladek Minor[1,2], and Misty L. Kuhn[*3]

[1]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, USA

[2]Center for Structural Genomics of Infectious Diseases (CSGID), University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA

[3]Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, CA 94132, USA

[4]Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, IL 60660, USA

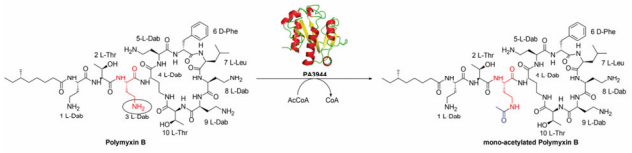
[*]To whom correspondence may be addressed: Department of Chemistry and Biochemistry, San Francisco State University, 1600 Holloway Ave., San Francisco, CA 94132. Tel.: 415-405-2112; E-mail: mkuhn@sfsu.edu

[#]Both authors contributed equally

Keywords: Gcn5-related N-acetyltransferase, GNAT, polymyxin B, colistin, polymyxin E, Pseudomonas aeruginosa, antibiotic acetylation

Abbreviations: Gcn5-related *N*-acetyltransferase (GNAT), 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), Terrific Broth (TB), isopropyl β-D-1-thiogalactopyranoside (IPTG), beta-mercaptoethanol (BME), acetyl-coenzyme A (AcCoA), L-2,4-diaminobutyric acid (Dab), L-2,3-diaminopropionic acid (Dap), 2-nitro-5-thiobenzoate (TNB^{2-}), trifluoracetic acid (TFA), methanol (MeOH)

TOC graphic:



Abstract
Deeper exploration of uncharacterized Gcn5-related *N*-acetyltransferases has the potential to expand our knowledge of the types of molecules that can be acylated by this important superfamily of enzymes and may offer new opportunities for biotechnological applications. While determining native or biologically relevant *in vivo* functions of uncharacterized proteins is ideal, their alternative or promiscuous *in vitro* capabilities provide insight into key active site interactions. Additionally, this knowledge can be exploited to selectively modify complex molecules and reduce byproducts when synthetic routes become challenging. During our exploration of uncharacterized Gcn5-related *N*-acetyltransferases from *Pseudomonas aeruginosa,* we identified such an example. We found that the PA3944 enzyme acetylates both polymyxin B and colistin on a single diaminobutyric acid residue closest to the macrocyclic ring of the antimicrobial peptide and determined the PA3944 crystal structure. This finding is important for several reasons: 1) to our knowledge this is the first report of enzymatic acylation of polymyxins and thus reveals a new type of substrate that this enzyme family can use, 2) the enzymatic acetylation offers a controlled method for antibiotic modification compared to classical promiscuous chemical methods, and 3) the site of acetylation would reduce the overall positive charge of the molecule, which is important for reducing nephrotoxic effects and may be a salvage strategy for this important class of antibiotics. While the physiological substrate for this enzyme remains unknown, our structural and functional characterization of PA3944 offers insight into its unique non-canonical substrate specificity.

Introduction

Gcn5-related *N*-acetyltransferases (GNATs) are a fascinating family of enzymes that acylate a variety of substrates ranging from small molecules to macromolecules. Members of this family are well known for their roles in aminoglycoside antibiotic resistance, histone modification, protein acetylation, xenobiotic metabolism, and other cellular processes [reviewed in[1,2]]. They accomplish this by acylating a primary amine of an acceptor molecule with a donor molecule like acetyl-coenzyme A (AcCoA). While *N*-acylation is most common, *O*-[3,4] and *S*-acylation[5] can also occur. The structures and functions of GNATs from all domains of life have been extensively studied; however, a significantly large number of them remain uncharacterized. On average, most organisms have approximately 20 GNATs encoded within their genomes; however, this number varies by organism and can be much lower or much greater.[2] For instance, *Campylobacter jejuni* has 4 genes that encode GNATs, while *Streptomyces coelicolor* has 72.[2] Thus, not all GNATs are conserved across organisms, which implies that the presence of certain GNATs may be tailored to the individual needs of the organism.

To gain a deeper understanding of the diversity of GNAT functions within a single bacterium, we have selected *Pseudomonas aeruginosa* PAO1 as a model system to further study GNATs of unknown function. The rationale for this selection is the following. First, this bacterium is found in a variety of environments, including in water,[6] in soil near the roots of plants,[7] on medical devices such as catheters,[8] and in the lungs of cystic fibrosis patients.[9] It is a highly antibiotic-resistant bacterium that causes severe nosocomial infections, especially in immunocompromised individuals.[10] Thus, knowledge gained about GNATs from this organism has important implications in agriculture and human disease as it is one of the ESKAPE[11] pathogens. Second, it is genetically tractable, i.e. it can be manipulated to explore the effects of deleting or overexpressing a gene in the organism in the laboratory, and an entire curated database (Pseudomonas Genome DB; http://www.pseudomonas.com/[12]) is maintained to house information related to its study. Finally, *P. aeruginosa* contains 36 GNATs in its genome (**Supplemental Table 1**) and the majority of them have been annotated as hypothetical proteins or as proteins of unknown function. Nine of them have been structurally characterized (**Supplemental Table 1**), but the depth of functional information is even less.

Due to the diversity of functions of GNATs already identified to date, we expect that exploring uncharacterized GNAT proteins will expand the repertoire of substrates utilized by members of this family of proteins. Additional benefits to studying GNATs of unknown function include improving their functional annotation across diverse genomes and identifying new targeted reactions that can be utilized for enzyme-mediated chemical synthesis or biotechnological applications. In an effort to expand our knowledge of the functional space of GNATs, we previously designed a broad-substrate screening assay to identify potential lead compounds to further explore the substrate specificity and catalytic activity of uncharacterized GNATs.[13] While this screen has its limitations, it does provide a starting point for further GNAT functional characterization.

Here, we focused on the structural and kinetic characterization of the PA3944 enzyme. We previously screened this enzyme against our panel of potential substrates and found it could acetylate polymyxin B and E (colistin) antimicrobial peptide antibiotics as well as the dipeptide aspartame[13]. Since the enzymatic activity toward polymyxins is novel, we concentrated our structural and functional characterization of PA3944 on its ability to acetylate polymyxins. These antibiotics have five potential sites for *N*-acetylation; therefore, we sought to further explore whether the enzyme was specific for one site or if it was promiscuous in its acetylation activity. While we do not believe the native function of this enzyme is to modify polymyxin antibiotics, we do think we can uncover clues about PA3944 substrate specificity by studying how this enzyme modifies these compounds. Since this is the first description of a polymyxin antibiotic modifying enzyme, we also determined its crystal structure. Collectively, our study provides structural and functional information about this enzyme, which will be useful for: 1) identifying other potential polymyxin-modifying GNAT enzymes and 2) engineering GNATs to selectively modify these antibiotics to salvage them in the clinic or be used for other biotechnological applications.

Methods

Clone

The *PA3944* gene (UniProtKB Q9HX72) from *Pseudomonas aeruginosa* was subcloned into the p11 vector as described previously.[14] It contains an ampicillin resistance cassette and gene expression is under control of the T7 promoter. Protein produced with this system contains an N-terminal methionine residue followed by a three amino acid spacer (GSS), a hexahistidine tag, another four amino acid spacer (SSGR), a tobacco etch virus (TEV) protease cleavage site (ENLYFQ/G), and a histidine spacer residue (i.e. MGSSHHHHHHSSGRENLYFQGH) prior to the first methionine of the PA3944 protein sequence.

Protein expression and purification for kinetics and mass spectrometry assays

Cell growth and protein expression—The clone containing the *PA3944* gene was transformed into *E. coli* BL21(DE3) competent cells and was stored as a glycerol stock at -80°C until ready to use. A 5 ml starter culture of lysogeny broth (LB) media with 100 μg/ml ampicillin was then inoculated and grown overnight at 37°C with shaking at 200 rpm on a benchtop shaker. The next day, 2.5 ml of starter culture was added to 250 ml of Terrific Broth (TB) with 100 μg/ml ampicillin in a 2 L glass Erlenmeyer flask and shaken at 200 rpm at 37°C until the OD_{600nm} reached 0.6-0.8. Cells were placed on ice until cool and then 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression at RT with shaking at 150 rpm overnight. Cells were harvested by centrifugation at 2200 x g for 30 min at 4°C, and the pellet was resuspended in 37.5 ml of cold lysis buffer (10 mM Tris-HCl pH 8.3, 500 mM NaCl, 5 mM imidazole, 5% glycerol, and 5 mM beta-mercaptoethanol (BME)), sonicated on ice for 5 min, and then stored at -80°C until ready to purify.

Protein purification and tag cleavage—The lysed cells were thawed and centrifuged at 25,000 x g for 45 min in a Sorvall SS-34 rotor at 4°C. The crude extract (supernatant) was

loaded onto a 1 ml HisTrap FF nickel-affinity column (GE Healthcare) that had been equilibrated with buffer A (10 mM Tris-HCl pH 8.3, 500 mM NaCl, and 5 mM BME) on an ÄKTA Start FPLC (GE Healthcare). Afterwards, the column was washed with 5 CVs of 5% buffer B (10 mM Tris-HCl pH 8.3, 500 mM NaCl, 5 mM BME, and 500 mM imidazole), and protein was eluted with 100% buffer B over 5 CVs. To remove imidazole, we performed a buffer exchange of the eluted protein into buffer A using a PD-10 gravity flow column (GE-Healthcare). Finally, we concentrated the protein using a 10K MWCO Sartorius VivaSpin concentrator. To remove the polyhistidine tag, 5 mg of protein was incubated with TEV protease (previously purified in our laboratory using the same protocol but stored in 10 mM Tris-HCl pH 8.3, 500 mM NaCl, 5 mM BME, 2 mM EDTA, 5 mM DTT, and 50% glycerol) in a 20:1 ratio at 4°C overnight in cleavage buffer (50 mM Tris pH 8.3, 1 mM DTT, 300 mM NaCl, 5% glycerol) in a 15 ml 10K MWCO Slide-a-Lyzer Mini-dialysis device (Thermo Fisher Scientific). Cleavage buffer was exchanged once after 2 hrs of dialysis. (We chose incubation for cleavage at 4°C because at higher temperatures all of the protein precipitated). The next morning, the dialysate was centrifuged at 12,000 x g for 10 min at 4°C and the supernatant was then loaded onto the affinity column previously equilibrated with buffer A. The cleaved protein was eluted with a gradient of 0-30% buffer B exchanged into buffer without BME (10 mM Tris-HCl pH 8.3 and 500 mM NaCl) because the BME interferes with the colorimetric enzyme kinetic assay. Protein was stored in aliquots at -80°C. Under these storage conditions, protein was stable for at least 1 year and no change in enzymatic activity was observed.

Protein crystallization and size-exclusion chromatography

Detailed methods for protein crystallization and size-exclusion chromatography can be found in Supplemental Information.

Data collection and structure determination

Diffraction data were collected at 100 K at the LS-CAT 21 ID-G (wavelength 0.97856 A, 12670 eV; PDB ID 6edd) and SBC-CAT 19-ID (0.97926 A, 12661 eV; PDB ID 6edv) beamlines at the Advanced Photon Source at Argonne National Laboratory. Data were processed with HKL-2000[15] and structure solution and model building were performed with HKL-3000[16] coupled with MOLREP,[17] BUCCANEER,[18] and Fitmunk.[19] Models were further refined with REFMAC5[20] and rebuilt with COOT.[21] The quality of the models were assessed with Molprobity[22] and wwPDB Validation Service.[23] Molecular replacement was using PDB ID: 3fbu (UniProt ID: Q81Q99) as the search model. All structures have been deposited into the Protein Data Bank (PDB) with the following accession codes: 6edv and 6edd (**Table 1**) and diffraction images have been deposited into the Integrated Resource for Reproducibility in Macromolecular Crystallography[24] (proteindiffraction.org) with the following identifiers, respectively: doi.org/10.18430/m36edd and doi.org/10.18430/m36edv.

Table 1. Data collection and refinement statistics for PA3944 structures.

PDB ID **Diffraction images DOI**	6edd	6edv
Data collection		
Beamline	LS-CAT 21 ID-G	SBC-CAT 19 ID
Resolution (Å)	1.55 (1.55-1.58)	1.35 (1.35-1.37)
Space group	*P*1	$P2_12_12_1$
Wavelength (Å)	0.97856	0.97926
a, b, c (Å)	36.45, 44.17, 60.12	36.44, 44.03, 111.64
α, β, γ (°)	81.88, 73.32, 89.94	90, 90, 90
Completeness (%)	96.2 (93.5)	100.0 (99.7)
Observed reflections	121370	558894
Unique reflections	49684	40378
<I> / <σ I>	21.3 (1.9)	46.4 (2.2)
CC1/2 last shell	0.77	0.68
Redundancy	2.4 (2.1)	13.8 (4.6)
R_{merge}	0.053 (0.397)	0.122 (0.778)
Wilson B factor (Å^2)	15.7	13.9
Refinement		
R_{work} / R_{free}	0.166 / 0.194	0.133 / 0.162
Bond lengths rmsd (Å)	0.008	0.013
Bond angles rmsd (Å)	1.4	1.6
Mean B value (Å^2)	20	19
Number of protein atoms	3125	1513

Number of water atoms	476	155
Number of ligands / ions atoms	130	57
Clashscore	2.5	0.98
Clashscore percentile (100)	99	99
Rotamer outliers (%)	0.00	0.00
Ramachandran outliers (%)	0.00	0.00
Residues with bad bonds / angles (%)	0.00 / 0.05	0.00 / 0.00
MolProbity score	1.03	0.79

*Values in parentheses refer to the highest resolution shell.

Enzyme kinetics assays for colorimetric method and mass spectrometry

Steady-state enzyme kinetic assays—All kinetic assays were performed in triplicate using the previously described discontinuous colorimetric method[13] with minor modifications. Briefly, all reactions were performed in 50 mM Tris-HCl pH 8.0 except for pH studies where the buffer was 50 mM Bis-Tris-propane at pH 6.5, 7.0, 8.0, 9.0, or 9.5 in a total volume of 50 µl for 10 min at 37°C and were initiated with 10 µl of enzyme. Reactions were terminated using 50 µl of 100 mM Tris-HCl pH 8.0 and 6 M guanidine HCl to denature the protein and then 200 µl of Ellman's reagent in buffer (0.2 mM 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), 100 mM Tris-HCl pH 8.0, and 1 mM EDTA) was added to each sample and incubated at RT prior to measuring the absorbance at 415nm with a fixed wavelength filter on a BioTek ELx808 microplate reader. Each molecule of CoA produced during the reaction reacts with one molecule of (DTNB), which forms a 2-nitro-5-thiobenzoate (TNB^{2-}) anion under basic conditions that absorbs at 415 nm. CoA standards (2.5 and 5 nmol) were used to determine the response factor for TNB^{2-} at 415 nm, which was used to convert TNB^{2-} absorbance to the number of nmols of CoA product produced during the enzymatic reaction. One unit of enzymatic activity is defined as 1 µmol of TNB^{2-} produced per min. Kinetic parameters were determined by fitting data to the Michaelis-Menten equation using Origin 2017 software. Details of the concentrations of donor (AcCoA) and acceptor substrates used for each assay are described below.

The PA3944 enzyme was screened against a panel of the following compounds in Tris-HCl pH 8.0 buffer to determine its substrate specificity and promiscuity: L-glutamate, L-aspartate, taurine, carnosine, beta-alanine, aspartame, glycine, L-alanine, L-2,3-diaminopropionic acid (DAP), colistin, polymyxin B, L-2,4-diaminobutyric acid (Dab), L-ornithine, and L-lysine. Each reaction was performed in the presence of 0.5 mM AcCoA, 5 mM compound and 0.91 µM enzyme. Substrate saturation curves were produced using

0.64 μM of tagged or cleaved PA3944 enzyme. The tagged enzyme was used for substrate screening assays and pH studies, while the cleaved protein was used for substrate saturation curves where kinetic parameters were determined in **Table 2**. For polymyxin B, colistin, and Dab substrate saturation curves AcCoA concentration was held constant at 0.5 mM while acceptor substrate was varied from 0-10 mM for polymyxin B and colistin, and 0-20 mM for Dab. For AcCoA substrate saturation curves, polymyxin B or colistin was held constant at 5 mM while AcCoA was varied from 0-2 mM.

Table 2. Comparison of kinetic parameters for tagged and cleaved PA3944 enzyme.

PA3944 Enzyme	Substrate	K_M (mM)			k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
Tagged	Polymyxin B (constant [AcCoA])[a]	2.57	±	0.23	0.653	2.54×10^2
	Colistin (constant [AcCoA])[a]	2.28	±	0.17	0.672	2.95×10^2
	DAB (constant [AcCoA])[a,b]	8.60	±	1.59	0.264	30.7×10^1
	AcCoA (constant [Polymyxin B])[c]	0.118	±	0.028	0.602	5.10×10^3
	AcCoA (constant [Colistin])[c]	0.105	±	0.017	0.471	4.49×10^3
Cleaved	Polymyxin B (constant [AcCoA])[a]	4.00	±	0.32	0.923	3.65×10^2
	Colistin (constant [AcCoA])[a]	2.50	±	0.13	0.894	3.58×10^2
	DAB (constant [AcCoA])[a,b]	10.0	±	1.2	0.391	3.91×10^1
	AcCoA (constant [Polymyxin B])[c]	0.137	±	0.020	0.610	4.45×10^3
	AcCoA (constant [Colistin])[c]	0.118	±	0.021	0.456	3.86×10^3

[a]The concentration of AcCoA was held constant at 0.5 mM while the acceptor substrate(s) were varied.
[b]At DAB concentrations higher than 10 mM substrate inhibition was observed; therefore, kinetic parameters were calculated based on data collected from 0-10 mM DAB.
[c]The concentration of polymyxin B or colistin was held constant at 5 mM while the concentration of AcCoA was varied.

Enzymatic assays for mass spectrometric analysis—Enzymatic assays were performed for 1 hr at 37°C in triplicate in a 100 μl reaction volume in the presence of 50 mM Tris-HCl pH 8.0, 1 mM AcCoA, 1 mM of either polymyxin B or colistin, and 0.64 μM of tagged PA3944 enzyme. Control reactions were performed in absence of enzyme. Reactions were terminated by adding 50% methanol to the reaction mixture and then frozen and stored at -80°C. Reaction products were analyzed by mass spectrometry as described below.

Mass spectrometry
Reaction mixtures were extracted by solid phase extraction prior to mass spectrometry analysis. They were then combined with 450 μl of trimethylamine acetate pH 8.0, loaded onto a Waters Corp HLB Oasis solid phase extraction cartridge (6 ml, 200 mg) and washed with 5 ml of 100 mM triethyl acetate pH 8.0 and 5 ml of distilled water. The peptides were

eluted with 5 ml of a 70:30 acetonitrile:water solution and evaporated to near dryness in a Speedvac evaporator. The residue was dissolved in a 75% methanol:water solution for direct infusion into the mass spectrometer. Molecular weight and product ion mass spectra were acquired with a ThermoFinnigan Advantage (San Jose, CA) LCQ ion trap mass spectrometer using electrospray ionization (ESI) in the positive ion mode. The tandem mass spectra energy parameter was set to 50-60% of its maximum value for fragmentation of all the singly-protonated peptides in this study. Mass spectra of all the peptide ions were acquired by direct infusion of the 3:1 methanol:water mixture. The capillary temperature was 200°C and the spray voltage was 4.5 kV. Molecular weight and product ion spectra were acquired from each reaction mixture and are shown in Supplemental Information.

Results and Discussion

PA3944 genomic context and protein properties

The *PA3944* gene from *P. aeruginosa* PAO1 lies within a five-gene operon (PA3941-PA3945) that encodes four hypothetical proteins and one predicted acyl-CoA thioesterase II (tesB). Two of these hypothetical proteins (PA3944 and PA3945) are predicted to have a GNAT domain, but their function is unknown and to our knowledge they have not been previously studied. The organization of this operon is conserved in *P. aeruginosa* strains PAO1 (an opportunistic pathogen and widely studied laboratory strain), DK2 (isolated from Danish Cystic Fibrosis patients), M18 (found in the sweet melon rhizosphere and produces anti-fungal compounds), and LESB58 (a hypervirulent Cystic Fibrosis isolate from Liverpool). The *PA3944* gene encodes a 21.9 kDa protein with a pI of 5.32 and is found in both pathogenic and nonpathogenic strains of Pseudomonas in 289 genera.[25]

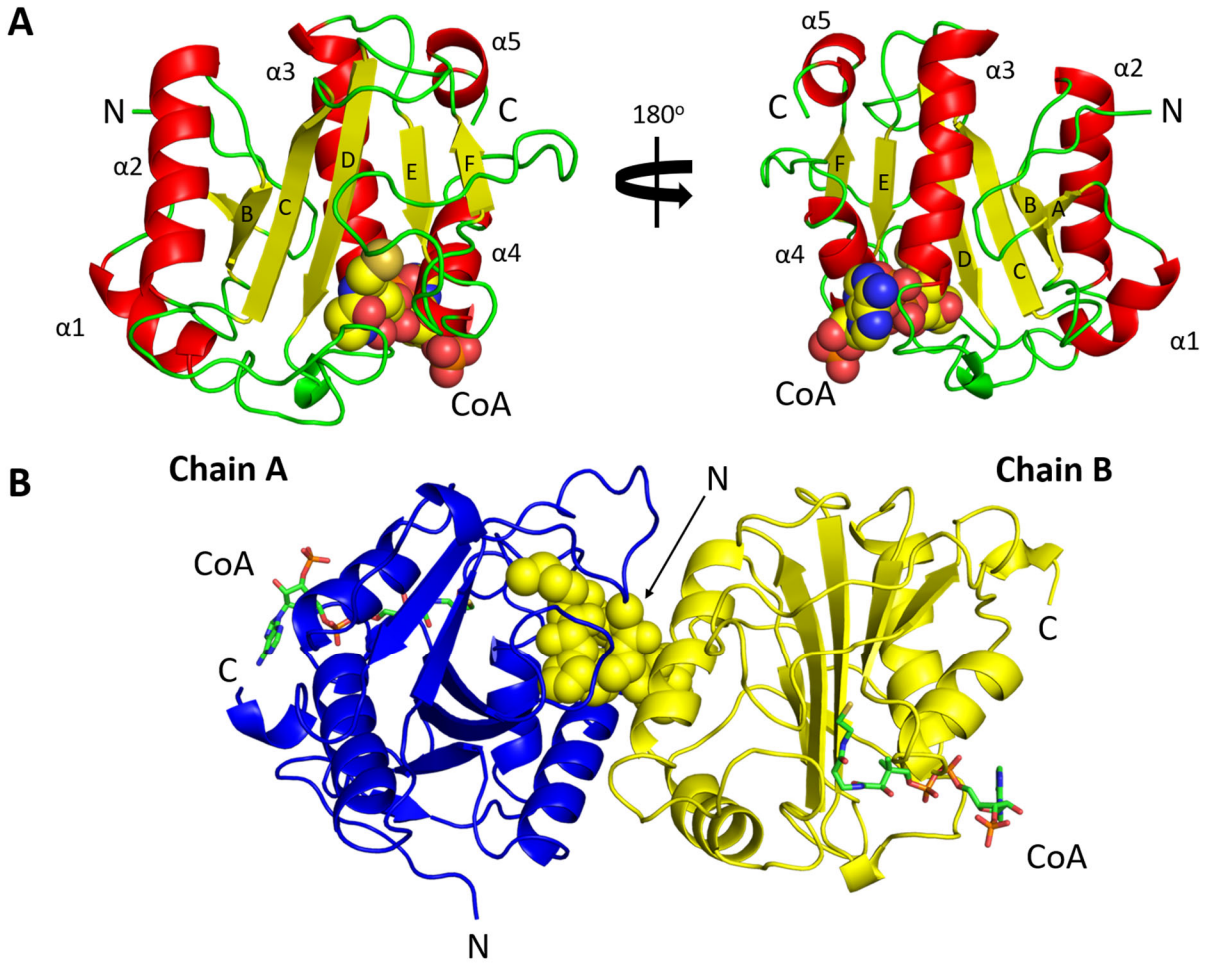
3D crystal structure of PA3944

We previously screened the PA3944 protein against a panel of potential substrates and found the enzyme exhibited the highest activity toward aspartame, polymyxin B and colistin (polymyxin E).[13] The rationale for including these compounds in the screening assay was that they were relatively inexpensive peptides with several potential sites for *N*-acylation. Due to the unusual substrate specificity of PA3944 exhibited toward polymyxin B and colistin, we sought to determine its crystal structure to learn more about how these antibiotics become acetylated. To begin, we screened this protein for crystals in the presence and absence of AcCoA/CoA and a variety of substrates identified from our previous broad-substrate screening assay.[13] Two structures of PA3944 with CoA were determined in different space groups: $P2_12_12_1$ with a monomer in the asymmetric unit, (PDB: 6edv) and $P1$ with two monomers in the asymmetric unit (PDB: 6edd) (**Table 1**).

The PA3944 protein structure consists of five α-helices that surround an antiparallel β-sheet formed by six β-strands (**Figure 1**), which is a typical signature of GNATs[26]. The canonical V-like splay is formed between β-strands B and C, and the topology of the PA3944 structure is similar to that of the PA4794 GNAT we studied previously.[27] CoA was modeled into the electron density found between helices α3, α4 and β-strands D, F and stabilized by interactions with the following residues: Trp105, Leu107, Gly113, Arg114, Gly115, Arg118, Thr141, Asn145, Ser148 and Arg154. This CoA binding pocket is also

conserved across the *P. aeruginosa* PA3270, PA4794 and PA2578 GNATs (**Supplemental Table 1**).

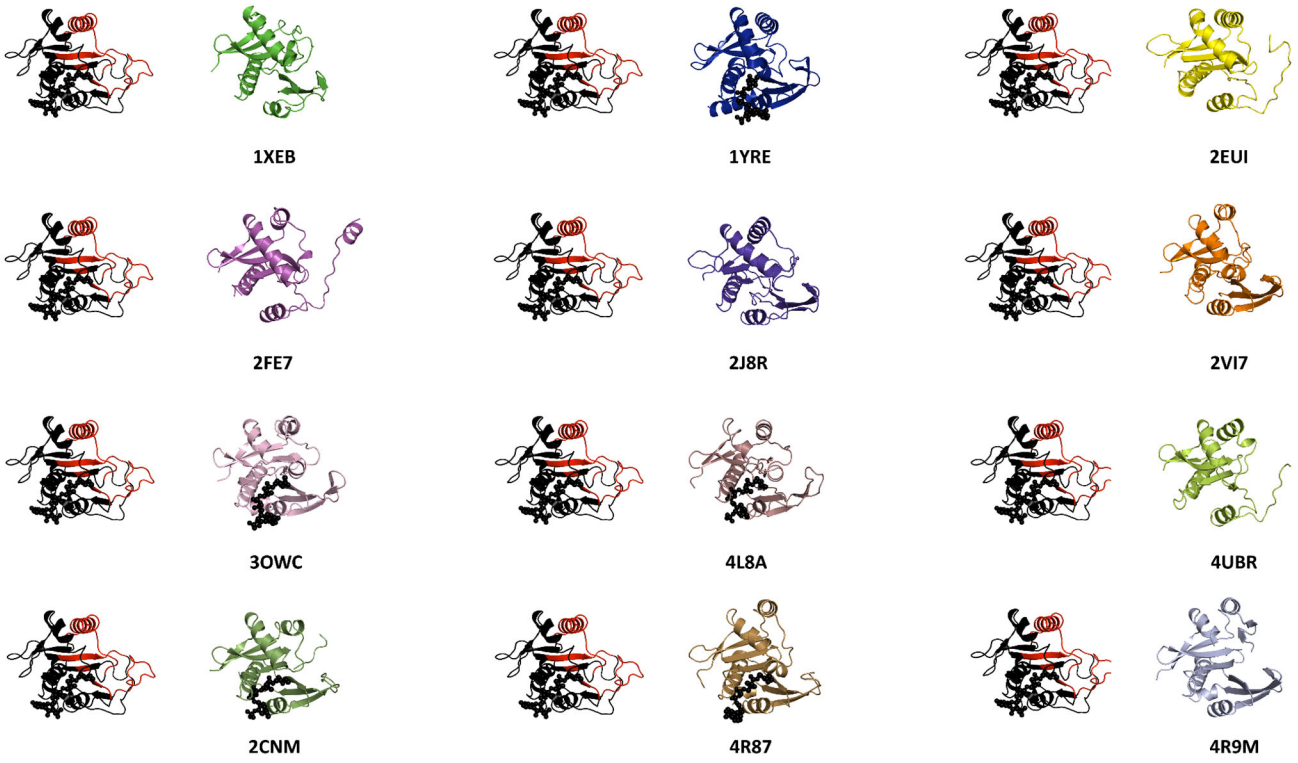
Figure 1. Crystal structures of PA3944 in complex with CoA. **(A)** Monomer assigned to orthorombic space group (PDB ID: 6edv). α helices (red) are labeled from α1 to α5, β strands (yellow) from A to F and loops are shown in green. CoA is shown as spheres. **(B)** The N-terminus (spheres) of one monomer (PDB ID: 6edd, chain B, yellow) is bound to the active site of the other monomer (PDB ID: 6edd, chain B, blue). CoA is shown as green sticks. The N- and C-termini are labeled as N and C in both panels.



We performed a structural similarity search of PA3944 against the PDB using VAST[28] and found the highest structural similarity toward two uncharacterized proteins from *Bordetella pertussis* (PDB ID: 3juw, 24.2% sequence identity) and *Bacillus halodurans* (PDB ID: 2qml, 10.8 % sequence identity). The structure in the PDB that exhibited the highest

sequence identity is an uncharacterized acetyltransferase from *Agrobacterium tumefaciens* (PDB ID: 2fsr, 32.2% sequence identity). We then used DALI[29] to compare the PA3944 structure to all other known *P. aeruginosa* GNATs (**Figure 2**) and found the highest structural similarity between PA3944 and the uncharacterized PA3270 protein (PDB ID: 1yre, 22.9% sequence identity). To determine whether we could predict the possible function of PA3944 from sequence or structure, we broadened our search to include GNATs that had been previously characterized from other organisms as described in Supplemental Methods. Our comparison of the amino acid sequences of the acceptor substrate binding sites of these proteins did not show any sequence similarity to PA3944; therefore, we were unable to predict a possible native function for this enzyme based on previously characterized GNATs. Moreover, we observed that the topology of the GNAT structures were highly variable and the most similar proteins adopted different conformations of some of their secondary structures (**Figure 2**). Thus, it was not feasible to infer function from structure for the PA3944 enzyme.

Figure 2. Structural comparison of PA3944 with other GNATs from *Pseudomonas aeruginosa* and other organisms with known functions. The structure of PA3944 is shown in red and black on the left of the structure being compared. The putative acceptor substrate binding region of PA3944 is shown in red and the CoA ligand is shown as black spheres. The corresponding PDB ID is shown and gene names that correspond to the PDB IDs are as follows: 1xeb (PA0115), 2fe7 (PA0478), 3owc (PA2578), 2cnm (RimI from *Salmonella typhimurium*), 1yre (PA3270), 2j8r (PA4866), 4l8a (PA4794), 4r87 (SpeG from *Vibrio cholerae*), 2eui (PA4026), 2vi7 (PA1377), 4ubr (PA4534), and 4r9m (SpeG from *Escherichia coli*).


1XEB
1YRE
2EUI
2FE7
2J8R
2VI7
3OWC
4L8A
4UBR
2CNM
4R87
4R9M

The only difference between our two PA3944 structures ($P2_12_12_1$ PDB ID: 6edv, and $P1$ PDB ID: 6edd) was the presence of a portion of the N-terminus in the active site of the adjacent monomer of the 6edd structure (**Figure 1**). Here, the histidine residue that remains after tag cleavage and the first four residues of the protein (HMNAN) are observed in the electron density and the δ-2 nitrogen of Asn2 from one chain forms a H-bond with the backbone oxygen of Ser46 on a mobile loop of the second chain. In this conformation, the ε-carbon atom of Met1 of the N-terminus bound in the active site of the opposite monomer is 5.4 Å from the sulfhydryl group of CoA. Residues Met41-Leu48 on the loop between helices α1 and α2 interact with the N-terminus and block the entrance to the active site. The N-terminus of the 6edd structure was not modeled due to lack of electron density for the first eight residues.

Since the PA3944 protein appeared to adopt two different oligomeric states (monomer and apparent dimer) in the different crystals, we performed size-exclusion chromatography to determine its oligomeric state in solution. We found the protein eluted as a monomer regardless of pH (see Methods for more details; **Supplementary Figure 1**). This result is consistent with PISA predictions[30] which did not show any significant interactions between monomers of the $P1$ (6edd) structure except those connected with binding a small portion of the N-terminus. Therefore, we conclude that the PA3944 protein is a monomer in solution and the observed binding of the N-terminus in the $P1$ (6edd) structure is a result of crystal packing but may indicate that the protein is capable of binding a peptide or peptide-like molecule. This is further supported by the fact that the closest sequence and structural homologs with known function acetylate proteins (ribosomal protein alanine acetyltransferase from *Salmonella typhimurium,* Uniprot ID Q8ZJW4_SALTY), or cyclic peptides (FsC-acetyl coenzyme A N(2)-transacetylase from *Neosartorya fumigate,* Uniprot ID SIDG_ASPFU) (**Figure 3**).

Figure 3. Pairwise sequence comparison between PA3944 and closest GNAT family proteins with known functions. Sequences are colored according to pairwise similarity from red (highly similar regions) to white (intermediately similar) to blue (highly dissimilar). Similarity was calculated as BLOSUM62 scores averaged over three residues. The CoA binding site in PA3944 is colored in orange, while its acceptor substrate binding site is in orange. The sequences are labeled with Uniprot IDs.

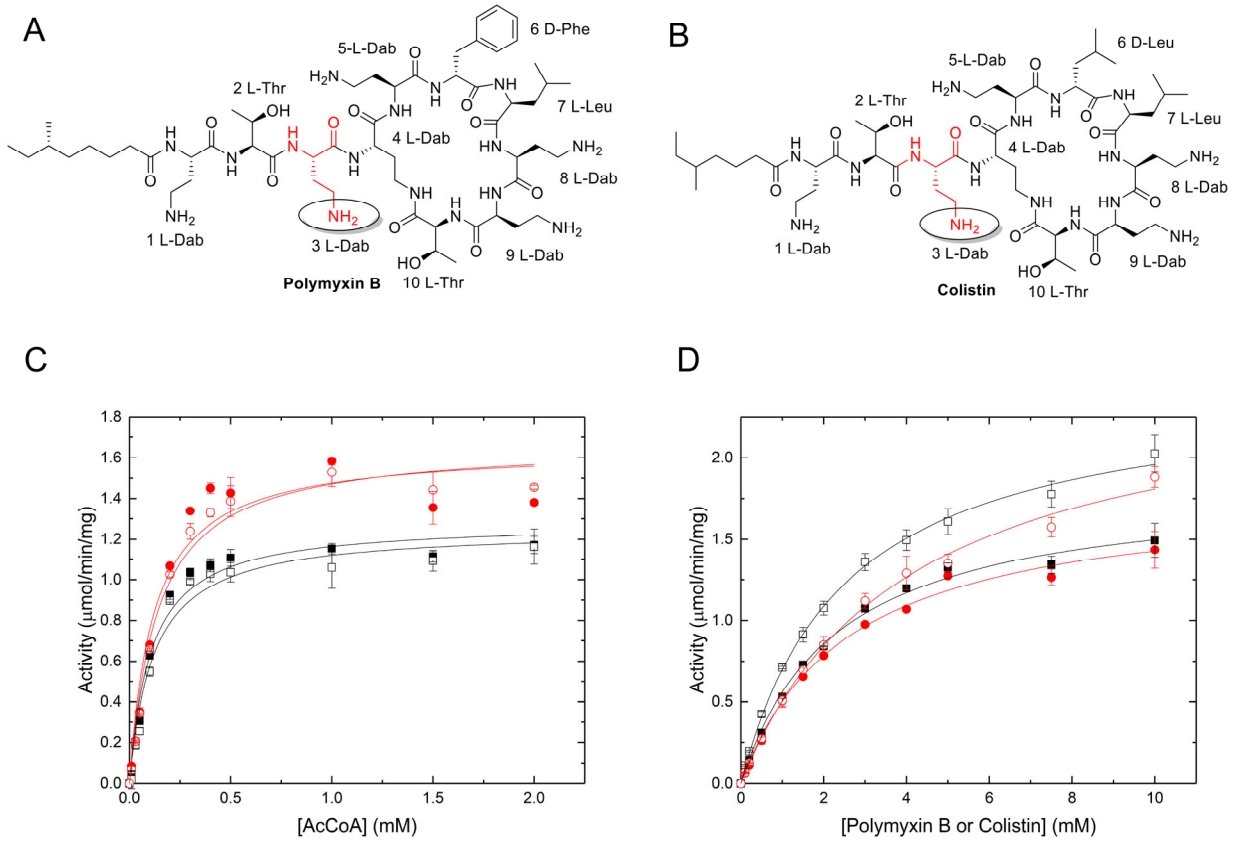
PA3944 acetylates polymyxin B and colistin on a single diaminobutyric acid residue

Polymyxins are polycationic macrocyclic antimicrobial peptides that have been isolated from strains of the bacterium *Bacillus polymyxa*, and they disrupt the structure of the bacterial cell membrane by interacting with its phospholipids. The most prevalent polymyxins include polymyxin B (B1 and B2) and colistin (polymyxin E1 and E2). These two families of polymyxins differ in their D-amino acid composition and in the fatty acids that are attached to the N-terminus of the peptide. Both families contain identical L-amino acid residues; however, polymyxin B also contains a D-phenylalanine residue whereas colistin has a D-leucine residue in its place. The fatty acid composition varies for each type of polymyxin, where polymyxin B1 and colistin E1 have a 6-methyloctanoyl group, whereas polymyxin B2 and colistin E2 bear a 6-methylheptanoyl residue [reviewed in[31]].

Since polymyxins have five potential sites for *N*-acetylation on diaminobutryic acid (Dab) residues within their structures (**Figure 4**), it was unclear whether the PA3944 enzyme acetylated a single Dab or multiple Dabs on each antibiotic. Our initial hypothesis was that the enzyme might acetylate Dab5 since it is located near the D-Phe/D-Leu of polymyxin B/colistin and most closely resembles the structure of aspartame, but we could not rule out the possibility that other Dabs might be modified. To determine the site(s) of acetylation on polymyxin B and colistin, we performed the enzymatic acetylation reaction and analyzed the products of the reaction using MS/MS. Product ion analyses of colistin and polymyxin B indicate that predominant acetylation product of PA3944 is a singly acetylated species and occurs only once on residue 3—the primary amine of the Dab residue proximal to the macrocycle in both antibiotics.

Figure 4. Structures of polymyxins and substrate saturation curves for PA3944. The structures of polymyxin B and colistin are shown in panels A and B, respectively. The regioselective acetylation by PA3944 occurs on the Dab residue highlighted in red and circled. AcCoA substrate saturation curves are shown in panel C at constant concentration of polymyxin B (red circles) and colistin (black squares). The tagged PA3944 protein is represented as filled circles or squares and the cleaved protein is represented by open circles or squares. Polymyxin B and colistin substrate saturation curves are shown in panel D at constant concentration of AcCoA. Coloring for

polymyxin B, colistin, tagged, and cleaved protein is the same as panel C. Error bars represent the standard deviation of triplicate reactions. See Methods for more details regarding assay conditions.



Mass spectrometry analysis of the enzymatic reaction containing polymyxin B yielded a singly acetylated polymyxin B1 $(M+H)^+$ ion at m/z 1246.3 that was selected for product ion analysis. Smaller abundances of singly-acetylated polymyxin ions containing sodium were observed at m/z 1267.8 and m/z 1289.7 as well. The acetylated polymyxin B2 formed ions in much smaller abundances (**Supplemental Figure S2**). No unmodified polymyxin B1 or B2 were observed in this mass spectrum. The product ion spectrum of the singly acetylated polymyxin B1 (m/z 1246.3; **Supplemental Figure S3**) shows an ion at m/z 904.8 formed by cleavage of the side chain resulting in the loss of a neutral fragment containing the (unmodified) primary amine group that is farthest away from the cyclic peptide ring. The product ion observed at m/z 762.4 is the cyclic peptide fragment formed by loss of the side chain (analogous to the m/z 728.3 in the product ion spectrum of the singly-acetylated colistin on Dab3 in **Supplemental Scheme S1**). Full scan MS analysis

of the control reaction mixture containing polymyxin without PA3944 enzyme showed no acetylated products.

Analysis of the enzymatic reaction containing colistin showed the singly acetylated colistin E2 $(M+H)^+$ ion at m/z 1197.8 and the singly acetylated $(M+Na)^+$ E1 and E2 ions at m/z 1233.9 and m/z 1219.9, respectively (**Supplemental Figure S4**). A protonated molecular ion from unmodified colistin E2 was observed at m/z 1155.8. Full scan MS analysis of the control reaction mixture containing colistin without PA3944 enzyme showed no acetylated products. The m/z 1197.8 ion, corresponding to a singly acetylated colistin E2, was subjected to product ion analysis to determine the site of colistin acetylation (**Supplemental Figure S5**). The cleavage of the colistin side chain produces an acylium ion at m/z 997.4 (containing the acetyl group) formed by the loss of a fragment containing the unmodified primary amine, closest to chain terminus, analogous to the m/z 904.8 observed in the product ion analysis of polymyxin. The formation of a product ion at m/z 728.3 corresponds to the unacetylated cyclic peptide fragment and demonstrates that the acetyl modification is on the primary amine closest to the cyclic peptide (Dab3) (**Supplemental Scheme S2**).

These results strongly suggest that the enzymatic reaction of PA3944 toward both polymyxin B and colistin is specific for Dab3, the Dab residue proximal to the cyclic peptide (**Figure 4**), as we did not observe multiple acetylation events on either of these peptides under our described assay conditions. The colistin E2 isoform was the predominant parent species as well as the predominant acetylated species observed in the MS assays, while the relative abundance of polymyxins B1 and B2 were similar. Since both E1/B1 and E2/B2 isoforms of the antibiotics were found to be acetylated, the differences in fatty acid and D-amino acid composition does not appear to majorly affect enzymatic acetylation of these substrates. Thus, the selectivity with respect to the location of acetylation on both polymyxins was consistent. This enzymatic specificity is in contrast to the nonselective chemical methods of acetylating polymyxin antibiotics, such as acetyl chloride or acetic anhydride, which are very reactive and nonselective.

PA3944 exhibits similar catalytic efficiencies toward both polymyxin acceptor substrates

We kinetically characterized the PA3944 enzyme toward polymyxin B and colistin using acetyl coenzyme A (AcCoA) as the acyl donor. Since polymyxin B and colistin structures are quite similar (**Figure 4**), we expected the PA3944 enzyme would exhibit comparable kinetic parameters toward both acceptor substrates. We found no major difference in the catalytic efficiency of the enzyme toward polymyxin B or colistin when comparing tagged or cleaved enzyme kinetic parameters (**Table 2 and Figure 4**). This indicates the differences in fatty acid at the N-terminus and D-amino acid within the peptide sequence of polymyxin B and colistin do not significantly alter PA3944 activity. Additionally, no significant difference in catalytic efficiency was seen when AcCoA was varied and concentrations of polymyxins were held constant regardless of whether the affinity tag was present or removed. Thus, the presence of the affinity tag does not significantly alter PA3944 enzymatic activity.

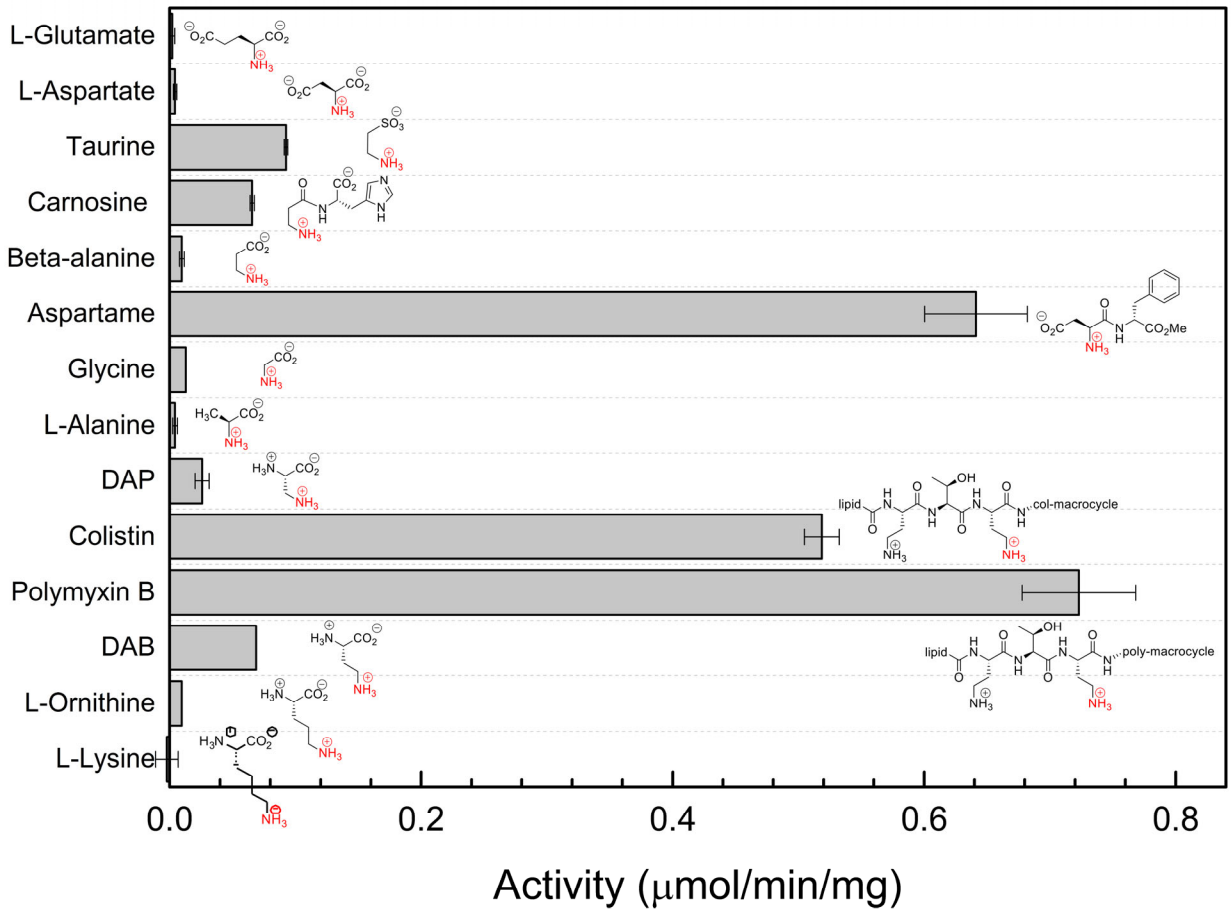
PA3944 is not an efficient diaminobutyric acid acetyltransferase

Since PA3944 acetylated Dab3 on polymyxins, we wondered whether it would acetylate free Dab amino acids more efficiently than those within the antibiotics. Dab acetyltransferases, which are GNATs and part of the ectoine biosynthetic pathway, have been studied from the halophilic bacteria *Methylomicrobium alcaliphilum, Methylophaga thalassica, Methylophaga alcalica,* and *Halomonas elongata*.[32-34] We found the apparent affinity of PA3944 for Dab was 1-2 orders of magnitude lower (8.6 or 10 mM; tagged vs cleaved, **Table 2**) than that reported for *M. alcaliphilum*, *M. thalassica*, and *M. alcalica* (0.460, 0.365, and 0.375 mM, respectively[33,34]), and its catalytic efficiency was one order of magnitude lower toward Dab compared to polymyxin B and colistin (**Table 2**). Therefore, we concluded that under our described reaction conditions the PA3944 enzyme is an inefficient Dab acetyltransferase and likely has an alternative native function.

PA3944 expanded substrate screening assays

To further explore substrate specificity of the PA3944 enzyme, we performed an expanded screening assay with compounds denoted in **Figure 5**. Similar to our previous results, the enzyme exhibited the highest activity toward aspartame, polymyxin B and colistin, but it is not obvious why polymyxins and aspartame would be acetylated to a similar extent since their structures are quite different. The previous substrate screen[13] did not include a complete series of compounds to evaluate which structural components of these molecules are important for PA3944 recognition. Therefore, we selected compounds that were commercially available to determine structural elements necessary for acetylation of aspartame and polymyxins (**Figure 5**). First, we focused on molecules that represented structural components of aspartame. We found L-glutamate, L-aspartate, glycine, and alanine are very poor substrates for the PA3944 enzyme. Limited activity was observed for substrates bearing an aminoethyl group (taurine, carnosine, and diaminobutyric acid); however, beta-alanine, which also contains an aminoethyl group, was a poor substrate. This indicates that there are structural elements in aspartame that are more favorable for acetylation than the simpler amine containing substrates we assayed (**Figure 5**). We are currently in the process of synthesizing aspartame analogs to learn more about the structural elements of aspartame that are recognized by PA3944.

Figure 5. Substrate screening of PA3944. Average activity of PA3944 from triplicate reactions toward each substrate is shown and more details of assay conditions are found in Methods. Structures of each substrate are overlaid onto the bar graph with the site of acetylation highlighted in red. Only partial structures of polymyxin B and colistin are shown. Full structures of these compounds can be found in Figure 4. Error bars represent the standard deviation of triplicate reactions.



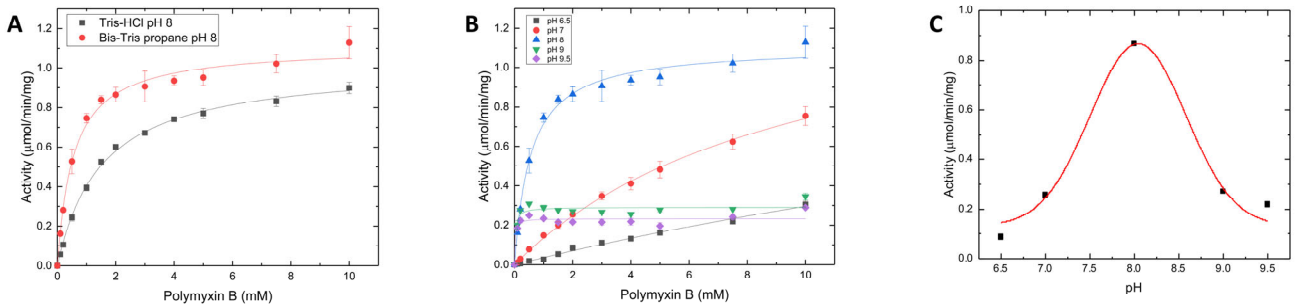
Next, we turned our attention to PA3944 activity toward polymyxin B and colistin. Both polymyxins contain diaminobutyric acid residues; therefore, we investigated the importance of chain length between the alpha carbon and primary amine by comparing activity toward free amino acids glycine, L-alanine, 2,3-diaminopropionic acid (Dap), 2,4-diaminobutyric acid (Dab), L-ornithine, and L-lysine (**Figure 5**). We found the optimal distance between the alpha carbon and the primary amine was the presence of two methylenes, like that of Dab. Therefore, we performed substrate saturation curves for Dab and found that the catalytic efficiency was one order of magnitude lower for the single free Dab residue than when it is incorporated into the cyclic peptide of polymyxin B or colistin (**Table 2**). This indicates that differences in charges of substrates or additional interactions between the enzyme and polymyxin substrates enhance activity and are important for substrate specificity.

Polymyxin acetylation is enzymatic and optimal activity occurs at pH 8

GNATs use a general acid/base catalytic mechanism to acetylate their acceptor substrates. This occurs when the enzyme deprotonates a primary amine of an acceptor substrate using either an acidic active site residue or water molecule, and then a residue like tyrosine is used to reprotonate CoA after acetyltransfer. As the pH of the buffer used for enzymatic

assays increases, there is a higher likelihood that the acceptor substrate is already deprotonated and would cloud whether or not the observed reaction is indeed enzymatic. Therefore, to ensure that the polymyxin acetylation reaction we observed was enzymatic, we performed a series of kinetic assays in different buffers and at different pHs. First, we measured the kinetic activity of PA3944 toward polymyxin B in the presence of Tris-HCl pH 8.0 or Bis-Tris-propane pH 8.0 buffer and compared the kinetic parameters (**Figure 6A**). We found the catalytic efficiency of the enzyme in Tris-HCl pH 8.0 and Bis-Tris-propane pH 8.0 was $2.47x10^2$ $M^{-1}s^{-1}$ and 7.33×10^2 $M^{-1}s^{-1}$, respectively, indicating the catalytic efficiency of the enzyme in Bis-Tris-propane buffer is nearly 3-fold higher than in Tris-HCl. Next, we performed substrate saturation curves toward polymyxin B in Bis-Tris-propane buffer at different pHs (**Figure 6B**). The activity of the enzyme varied with pH and did not continue to increase linearly at higher pH, which would be expected if non-enzymatic acetylation were occurring. We found the enzymatic activity toward polymyxin B at different pHs exhibited a gaussian curve (**Figure 6C**) with a pH optimum at 8.0. This type of curve matches what would be expected for a general acid/base catalytic mechanism where two residues are likely participating in the reaction. Additional plots of log V_{max} as a function of pH, log K_m as a function of pH, and log k_{cat}/K_m as a function of pH are presented in **Supplemental Figure S6**. Note that substrate saturation curves at pH 6.5 and 7.0 (**Figure 6B**) did not reach saturation, so the apparent V_{max} and K_m values used to produce the curves in **Supplemental Figure S6** may affect data interpretation. Overall, our results provide evidence that the polymyxin acetylation reaction is indeed enzymatic. Studies to identify these critical residues for enzymatic activity are currently underway in our laboratory.

Figure 6. PA3944 activity toward polymyxin B in different buffers and pHs. A) Comparison of polymyxin B substrate saturation curves at constant concentration of AcCoA (0.5mM) in Tris-HCl and Bis-Tris propane buffers at pH 8.0. Data were fit using non-linear regression as described in materials and methods. B) Polymyxin B substrate saturation curves as a function of pH in Bis-Tris propane buffers from pH 6.5-9.5. Data were fit using non-linear regression as described in materials and methods. C) PA3944 activity toward 2 mM polymyxin B and 0.5 mM AcCoA in Bis-Tris propane buffers at pH 6.5, 7.0, 8.0, 9.0, and 9.5. Data were fit to a gaussian curve. See materials and methods for further details.



Polymyxins are heterogeneous and may affect reproducibility

Since commercial polymyxins are heterogeneous mixtures of fermentation products isolated from bacteria and they can adhere to polystyrene, it is important to note that different preparations of solutions of these molecules may vary in composition from batch to batch [reviewed in[35]]. While this potentially presents challenges for interpreting and reproducing kinetic parameters from different laboratories, in our hands two different PA3944 enzyme preparations and polymyxin B preparations were used to collect the enzyme kinetic assay data in **Figures 4-6**. We observed differences in k_{cat} and K_m from the two preparations; however, the catalytic efficiencies of the enzymes were nearly identical. Thus, catalytic efficiencies may be the more useful parameter for comparing PA3944 enzymatic activity in the future from lab to lab.

The non-native function of PA3944 could be exploited for biotechnology applications

Due to the rise in antibiotic resistance toward most antibiotics, polymyxins are being used as a last resort to treat multi-drug resistant or extreme drug resistant bacterial infections [see reviews[31,36]]. While we have shown that PA3944 can acetylate polymyxin antibiotics *in vitro,* we do not think this is its physiological function. The rationale for our conclusion is primarily based on the fact that the apparent affinity of PA3944 for both polymyxins is in the millimolar range (**Table 2**) and reported breakpoints for polymyxin B or colistin for *P. aeruginosa* resistance to these antibiotics is in the micromolar range (greater than or equal to 4 μg/mL (~3 μM) for colistin and 8 μg/mL (~6 μM) for polymyxin B[35]). While it is unlikely that the native function of PA3944 is to modify polymyxins, its ability to singly acetylate Dab3 on the fatty acyl tail of these important antibiotics could be exploited. For instance, i*n vitro* chemical mechanisms of acylation of polymyxins are promiscuous, and having a controlled acylation method for modifying these antibiotics enables specific chemical transformations of these important molecules. Additionally, future directed evolution of the PA3944 enzyme may allow for selective acylation of other sites of these types of antibiotics, which could significantly improve synthetic strategies of these molecules. Furthermore, if the PA3944 enzyme is able to recognize other antimicrobial peptides in a similar manner, it could also be used to selectively modify structurally similar octapeptin antimicrobial peptides. While the native function of PA3944 remains unknown and warrants further investigation, the regioselectivity of this enzyme may be exploited for future biotechnological applications (e.g. designing reporters) and/or drug development.

Supporting Information

Supplemental Methods for protein expression and purification for crystallization, protein crystallization, size-exclusion chromatography, and sequence and structural comparison of homologs. Supplemental Results including mass spectrometry of acetylated polymyxin B and colistin. Supplemental Table S1, Figures S1-S6, and Schemes SC1, SC2.

Acknowledgements

We would like to acknowledge the following people for their helpful discussions and/or technical assistance: George Gassner, Laura Conrad-Miller, Teaster Baird Jr., Marc O. Anderson, Jorge C. Escalante-Semerena, Robert Hancock, Reza Falsafi, Jackson Baumgartner, David T. Tran, Brian Amsler, and Andrew S. Kuhn. Additionally, I (MLK) will be forever grateful to Gerry Wright for sage advice and guidance and critical reading

of the manuscript. Funding for this work was provided by the National Science Foundation (CHE-1708863 to MLK and CHE-1708927 to DPB), a Facilitating Research and Creative Work at San Francisco State University Collaborative Grant (to MLK). This research was also funded by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract nos. HHSN272201200026C and HHSN272201700060C and by NIH grant GM117325 (to WM). This research used resources of the Advanced PhotonSource, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The results shown in this report are derived from work performed at SBC-CAT and LS-CAT sectors. SBC-CAT is operated by UChicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

References

1. Favrot, L.; Blanchard, J. S.; Vergnolle, O. (2016) Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry 55*, 989-1002.

2. Hentchel, K. L.; Escalante-Semerena, J. C. (2015) Acylation of Biomolecules in Prokaryotes: a Widespread Strategy for the Control of Biological Function and Metabolic Stress. *Microbiol. Mol. Biol. Rev. 79*, 321-346.

3. Hegde, S. S.; Javid-Majd, F.; Blanchard, J. S. (2001) Overexpression and mechanistic analysis of chromosomally encoded aminoglycoside 2'-N-acetyltransferase (AAC(2')-Ic) from Mycobacterium tuberculosis. *J. Biol. Chem. 276*, 45876-45881.

4. Daigle, D. M.; Hughes, D. W.; Wright, G. D. (1999) Prodigious substrate specificity of AAC(6')-APH(2"), an aminoglycoside antibiotic resistance determinant in enterococci and staphylococci. *Chem. Biol. 6*, 99-110.

5. Gu, L.; Geders, T. W.; Wang, B.; Gerwick, W. H.; Hakansson, K.; Smith, J. L.; Sherman, D. H. (2007) GNAT-like strategy for polyketide chain initiation. *Science 318*, 970-974.

6. Caskey, S.; Stirling, J.; Moore, J. E.; Rendall, J. C. (2018) Occurrence of Pseudomonas aeruginosa in waters: implications for patients with cystic fibrosis (CF). *Lett. Appl. Microbiol. 66*, 537-541.

7. Green, S. K.; Schroth, M. N.; Cho, J. J.; Kominos, S. K.; Vitanza-jack, V. B. (1974) Agricultural plants and soil as a reservoir for Pseudomonas aeruginosa. *Appl. Microbiol. 28*, 987-991.

8. Mittal, R.; Aggarwal, S.; Sharma, S.; Chhibber, S.; Harjai, K. (2009) Urinary tract infections caused by Pseudomonas aeruginosa: a minireview. *J. Infect. Public Health 2*, 101-111.

9. Bhagirath, A. Y.; Li, Y.; Somayajula, D.; Dadashi, M.; Badr, S.; Duan, K. (2016) Cystic fibrosis lung environment and Pseudomonas aeruginosa infection. *BMC Pulm. Med. 16*, 174.

10. Obritsch, M. D.; Fish, D. N.; MacLaren, R.; Jung, R. (2005) Nosocomial infections due to multidrug-resistant Pseudomonas aeruginosa: epidemiology and treatment options. *Pharmacotherapy 25*, 1353-1364.

11. Rice, L. B. (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE*. J. Infect. Dis. 197, 1079-81*.

12. Winsor, G. L.; Griffiths, E. J.; Lo, R.; Dhillon, B. K.; Shay, J. A.; Brinkman, F. S. (2015) Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res. 44*, D646-D653.

13. Kuhn, M. L.; Majorek, K. A.; Minor, W.; Anderson, W. F. (2013) Broad-substrate screen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity. *Protein Sci. 22*, 222-230.

14. Zhang, R.; Skarina, T.; Katz, J.; Beasley, S.; Khachatryan, A.; Vyas, S.; Arrowsmith, C.; Clarke, S.; Edwards, A.; Joachimiak, A. (2001) Structure of Thermotoga maritima stationary phase survival protein SurE: a novel acid phosphatase. *Structure 9*, 1095-1106.

15. Otwinowski, Z.; Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276, 307-26.*

16. Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallographica Section D: Biological Crystallography 62*, 859-866.

17. Vagin, A.; Teplyakov, A. MOLREP: an automated program for molecular replacement. *Journal of applied crystallography* **1997***, 30*, 1022-1025.

18. Cowtan, K. (2008) Fitting molecular fragments into electron density. *Acta Cryst. D64*, 83-89.

19. Porebski, P. J.; Cymborowski, M.; Pasenkiewicz-Gierula, M.; Minor, W. (2016) Fitmunk: improving protein structures by accurate, automatic modeling of side-chain conformations. *Acta Cryst. D72*, 266-280.

20. Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D67*, 355-367.

21. Emsley, P.; Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst. D 60*, 2126-2132.

22. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.,3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375-83.

23. Gore, S.; García, E. S.; Hendrickx, P. M.; Gutmanas, A.; Westbrook, J. D.; Yang, H.; Feng, Z.; Baskaran, K.; Berrisford, J. M.; Hudson, B. P. (2017) Validation of structures in the Protein Data Bank. *Structure 25*, 1916-1927.

24. Grabowski, M.; Langner, K. M.; Cymborowski, M.; Porebski, P. J.; Sroka, P.; Zheng, H.; Cooper, D. R.; Zimmerman, M. D.; Elsliger, M.; Burley, S. K. (2016) A public database of macromolecular diffraction experiments. *Acta Cryst. D72*, 1181-1193.

25. Winsor, G. L.; Griffiths, E. J.; Lo, R.; Dhillon, B. K.; Shay, J. A.; Brinkman, F. S. (2015) Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res. 44*, D646-D653.

26. Vetting, M. W.; de Carvalho, Luiz Pedro S; Yu, M.; Hegde, S. S.; Magnet, S.; Roderick, S. L.; Blanchard, J. S. (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys. 433*, 212-226.

27. Majorek, K. A.; Kuhn, M. L.; Chruszcz, M.; Anderson, W. F.; Minor, W. (2013) Structural, Functional, and Inhibition Studies of a Gcn5-related N-Acetyltransferase (GNAT) Superfamily Protein PA4794. *J. Biol. Chem. 288*, 30223-30235.

28. Madej, T.; Lanczycki, C. J.; Zhang, D.; Thiessen, P. A.; Geer, R. C.; Marchler-Bauer, A.; Bryant, S. H. (2013) MMDB and VAST : tracking structural similarities between macromolecular complexes. *Nucleic Acids Res. 42*, D297-D303.

29. Holm, L.; Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res. 44*, W351-W355.

30. Krissinel, E.; Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774-797.

31. Velkov, T.; Thompson, P. E.; Nation, R. L.; Li, J. (2009) Structure− activity relationships of polymyxin antibiotics. *J. Med. Chem. 53*, 1898-1916.

32. Ono, H.; Sawada, K.; Khunajakr, N.; Tao, T.; Yamamoto, M.; Hiramoto, M.; Shinmyo, A.; Takano, M.; Murooka, Y. (1999) Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, Halomonas elongata. *J. Bacteriol. 181*, 91-99.

33. Mustakhimov, I. I.; Rozova, O. N.; Reshetnikov, A. S.; Khmelenina, V. N.; Murrell, J. C.; Trotsenko, Y. A. (2008) Characterization of the recombinant diaminobutyric acid acetyltransferase from Methylophaga thalassica and Methylophaga alcalica. *FEMS Microbiol. Lett. 283*, 91-96.

34. Reshetnikov, A.; Mustakhimov, I.; Khmelenina, V.; Trotsenko, Y. A. (2005) Cloning, purification, and characterization of diaminobutyrate acetyltransferase from the halotolerant methanotroph Methylomicrobium alcaliphilum 20Z. *Biochemistry (Moscow) 70*, 878-883.

35. Bakthavatchalam, Y. D.; Pragasam, A. K.; Biswas, I.; Veeraraghavan, B. (2018) Polymyxin susceptibility testing, interpretative breakpoints and resistance mechanisms: An update. *J. Glob. Antimicrob. Resist. 12,* 124-136.

36. Velkov, T.; Roberts, K. D.; Thompson, P. E.; Li, J. (2016) Polymyxins: a new hope in combating Gram-negative superbugs? *Future Med. Chem. 10,* 1017-25.

SUPPLEMENTAL INFORMATION

A Gcn5-related N-acetyltransferase (GNAT) capable of acetylating polymyxin B and colistin antibiotics *in vitro*

Mateusz P. Czub[#1,2], Brian Zhang[#3], M. Paul Chiarelli[4], Karolina A. Majorek[1,2], Layton Joe[3], Przemyslaw J. Porebski[1,2], Alina Revilla[3], Weiming Wu[3], Daniel P. Becker[4], Wladek Minor[1,2], and Misty L. Kuhn[*3]

[1]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, USA

[2]Center for Structural Genomics of Infectious Diseases (CSGID), University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA

[3]Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, CA 94132, USA

[4]Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, IL 60660, USA

SUPPLEMENTAL METHODS

Materials

Acetyl coenzyme A (AcCoA; #A2181) and coenzyme A (CoA; #C3019) trilithium salts, and polymyxin B (#P4932) and colistin (#C4461) sulfate salts were purchased from Sigma-Aldrich. All other reagents for biochemical assays were purchased at the highest quality available.

Protein expression and purification for crystallization

The ampicillin resistant plasmid containing the *PA3944* gene was transformed into chloramphenicol resistant *E. coli* BL21(DE3) CodonPlus-RIL cells (Agilent). A starter culture of LB with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol was inoculated and grown overnight at 37°C at 200 rpm. The following morning 10 ml of starter culture was used to inoculate 1 L of LB. Cells were grown until the OD_{600nm} reached 0.8 and then cooled to 16°C. Protein expression was induced with 1 mM IPTG and expressed overnight at 16°C with shaking. The next day, the cells were harvested by centrifugation (25 min at 4000 x g at 4°C in a Beckman Coulter Avanti J-26 XP centrifuge) and the supernatant was discarded. The pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 7.8, 2 mM imidazole, 150 mM sodium chloride) and homogenized in a glass homogenizer. One tablet of protease inhibitor (cOmplete Mini, EDTA-free Protease Inhibitor Cocktail Tablets, Roche) and a pinch of lysozyme (Sigma product # L-6876) and 2 µl of Benzonase Nuclease (Sigma product # E1014) were then added to the cell suspension, sonicated on ice, and centrifuged (45 min at 142,400 x g, and 4°C in a Beckman Coulter Optima L-80 XP ultracentrifuge). The supernatant was loaded onto a gravity flow Ni-NTA column (Qiagen) previously equilibrated with lysis buffer and the flow-through was collected and reapplied on the column three times to extend the length of interaction time between the protein and resin. The resin was then washed with wash buffer (50 mM Tris-HCl pH 7.8, 10 mM imidazole, 600 mM sodium chloride) and protein was eluted with elution buffer (50 mM Tris-HCl pH 7.8, 250 mM imidazole, 200 mM sodium chloride). To remove the polyhistidine tag, the eluted protein was combined with tobacco etch virus (TEV) protease in a 15:1 ratio and dialyzed in a 10 kD MWCO Slide-A-Lyzer dialysis cassette (Thermo Fisher Scientific) against 1 L of dialysis buffer (50 mM Tris-HCl pH 7.8, 150 mM sodium chloride, 14 mM BME) overnight at 4°C. The dialyzed protein was loaded onto the nickel-affinity column that had been equilibrated with dialysis buffer. The flow-through was collected, concentrated with an Amicon Ultra Centrifugal Filter with a 10K MWCO, and loaded onto a HiLoad 16/60 Superdex 200 gel filtration column using an AKTA FPLC (GE Healthcare) in 100 mM Tris-HCl pH 7.5 and 150 mM NaCl buffer. Polyhistidine tag cleavage was confirmed by SDS-PAGE.

Protein crystallization

The PA3944 protein was crystallized using hanging drop vapor diffusion at 16°C in 3-Well Midi Crystallization plates (Swissci) that were set using a Mosquito crystallization robot (TTP Labtech). Crystals used for data collection and structure determination were from two different preparations of protein but in the same crystallization condition (MCSG1 screen, well C11: 100 mM Tris-HCl pH 7.0, 200 mM calcium acetate monohydrate, 20%w/v PEG 3000): 1) cleaved PA3944 protein at 14 mg/ml co-crystallized in the presence of 5 mM puromycin and CoA with a ratio of protein

to mother liquor of 1:1; ethylene glycol was used as a cryoprotectant. 2) cleaved PA3944 protein at 10 mg/ml co-crystallized in the presence of 5 mM colistin and CoA with a ratio of protein to mother liquor of 3:2; no cryoprotectant was added when harvesting crystals for data collection. Crystals were harvested using Hampton CryoLoops and dried for 10-15 minutes over 1 M sodium chloride (a slow dehydration technique) and then frozen in liquid nitrogen.

Size-exclusion chromatography

The oligomeric state of the PA3944 protein was determined using size exclusion chromatography with a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare). A calibration curve was prepared from retention times and molecular weights of albumin (66.5 kDa), ovalbumin (44.3 kDa), chymotrypsin (25 kDa) or ribonuclease A (13.7 kDa) in 100 mM Tris-HCl pH 7.0 buffer. To determine whether pH affected the oligomeric state of the PA3944 protein, separate measurements were determined in the following buffers that contained 150 mM NaCl: 100 mM potassium phosphate pH 6.0; 100 mM Tris-HCl at pH 7.0 or 8.0; and 100 mM CHES at pH 9.0. We also assessed whether the presence of AcCoA affected the oligomeric state of PA3944 by analyzing the retention times of the PA3944 protein in the presence of a 2-fold excess of AcCoA: protein. 15 nmol of protein was injected onto the column at a flow rate of 1.5 ml/min for each measurement.

Sequence and structural comparison of homologs

To identify PA3944 homologs, we queried the SwissProt database using the PA3944 sequence and single iteration BLAST (E-value cutoff of 10). From these results, we selected sequences of proteins with 3D structures and experimentally verified functions or preliminary substrates. These included two polyamine acetyltransferases (Uniprot IDs ATDA_VIBCH and ATDA_ECOLI) and a fusarinine C siderophore acetyltransferase (Uniprot ID: SIDG_ASPFU)[1-3]. Two sequences of protein acetyltransferase RimI from *E. coli* and *S.typhimurium* (Uniprot IDs Q8ZJW4)SALTY, RIMI_ECO57, respectively) and three sequences of GNAT acetyltransferases from *P. aeruginosa* (Uniprot IDs: Q9HUU7, and Q9HV14) that have been found to acetylate L-methionine sulfoximine and L-methionine sulfone[4], and a C-terminal lysine residue of a peptide[5], respectively, as well as UniProt ID: Q9I0Q8, which was included in our broad substrate screening[6] but does not have a known function. The sequence alignment (**Figure 3**) was performed using the program PROMALS3D[7,8]. A structural alignment of all GNATs from *P. aeruginosa* with crystal structures was performed using secondary-structure-matching (SSM) with the program SUPERPOSE from the CCP4 suite[9]. The acceptor substrate binding site in all structures was visually identified.

SUPPLEMENTAL RESULTS

Mass spectrometry of acetylated colistin and polymyxin B

Molecular weight and product ion spectra were acquired from each reaction mixture. Product ion analyses of colistin and polymyxin indicate that the primary site of enzymatic acetylation occurs (just once) on the primary amine on the side chain closest to the cyclic peptide at the end of the chain. Molecular weight analysis of reaction mixtures containing AcCoA and

polymyxin in presence and absence of enzyme are shown in Figure S2. These spectra strongly suggest polymyxin acetylation only occurs in the presence of the PA3944 enzyme. In Figure S2A, $(M+H)^+$ and $(M+Na)^+$ ions formed from polymyxin B1 are observed at m/z 1204 and m/z 1226 and those for polymyxin B2 are observed at m/z 1190 and m/z 1211, respectively. No acetylated polymyxin ions are observed in the spectrum of the control reaction in absence of enzyme. Figure S2B shows monoacetylated polymyxin B1 at m/z 1246 $(M+H)^+$, m/z 1268 $(M+Na)^+$, and m/z 1290 $(M-H+2Na)^+$, respectively. No multiply acetylated polymyxin molecules are observed.

Product ion spectra of the $(M+H)^+$ ion at m/z 1246 formed from the monoacetylated polymyxin is shown in Figure S3. An ion observed at m/z 904.8 is formed by cleavage on the side chain resulting in the loss of a neutral fragment containing the (unmodified) primary amine that is farthest away from the cyclic peptide ring. The m/z 904.8 fragment ion contains the acetyl group. The product ion observed at m/z 762.4 is the cyclic peptide fragment formed by loss of the side chain analogous to the m/z 728.3 in the product ion spectrum of the singly-acetylated polymyxin B in Figure S3 (see also Scheme S1). If polymyxin B was acetylated on the primary amine closest to the end of the fatty acyl chain (Dab1) we would expect to see an ion at m/z 863. This ion is not observed, so we conclude that the primary amine closest to the peptide ring in polymyxin is the main site of acetylation.

Molecular weight analysis of reaction mixtures containing AcCoA and colistin in the presence and absence of enzyme are shown in Figure S4. Spectra of the control reaction show $(M+H)^+$ and $(M+Na)^+$ of unacetylated colistin E2 (m/z 1156 and m/z 1178) and colistin E1 molecules (m/z 1170 and m/z 1192), respectively (Figure S4A). Ions formed from monoacetylated colistin E2 are observed in Figure S4B at m/z 1198 $(M+H)^+$ and m/z 1220 $(M+Na)^+$. A smaller abundance of the $(M+Na)^+$ formed from monoacetylated colistin E1 is observed at m/z 1234. These spectra strongly suggest the acetylation of colistin only occurs in the presence of the PA3944 enzyme. No m/z values consistent with the $(M+H)^+$ and $(M+Na)^+$ ions formed from multiply acetylated colistin are observed. We do not know the identities of the additional ions observed at m/z 1313, 1327, and 1449 but each appear to become acetylated in the presence of the enzyme due to the observation of ions at m/z 1356, 1370, and 1491. Since colistin sulfate is supplied as a mixture of molecules, we cannot conclusively state the identity of these ions; however, it is clear that the apparently acetylated molecules are only observed in the presence of the PA3944 enzyme.

The protonated molecule ion at m/z 1198 was subjected to product ion analysis (Figure S5) to locate the position of acetyl modification on the colistin molecule (Scheme S2). The product ion spectra suggest that the primary position of acetylation is on the primary amine on the side chain closest to the peptide ring. The observation of an acylium ion formed by the cleavage of colistin produces an ion at m/z 997.4 (containing the acetyl group) that is formed by the loss of a fragment containing the unmodified primary amine farthest away from the cyclic peptide ring. The formation of a product ion at m/z 728.3, corresponding to the unacetylated cyclic peptide fragment, strongly suggests that acetylation occurs on the primary amine closest to the cyclic peptide (Scheme S2). Modification of the amine closest to the end of the peptide chain would have yielded an acylium ion at m/z 955 formed by the loss of a fragment containing the acetyl group and this ion is not observed in any significant abundance. Therefore, we conclude the primary amine closest to the peptide ring is the principal site of modification.

SUPPLEMENTAL TABLES, FIGURES, and SCHEMES

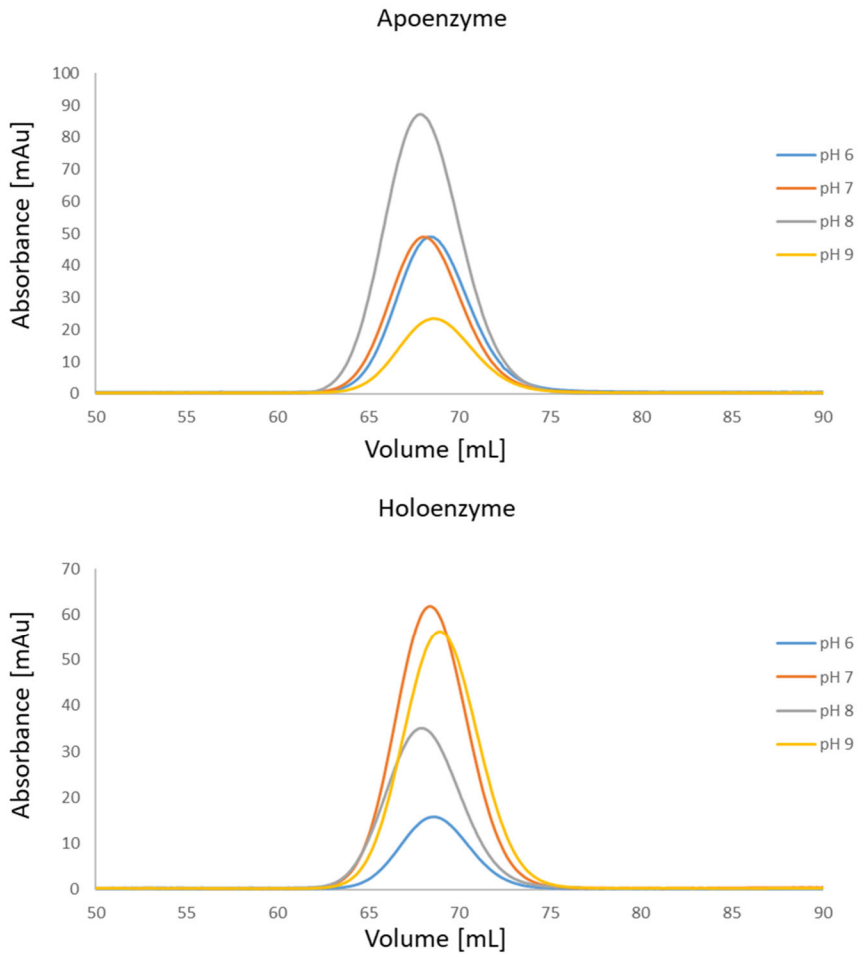
Table S1. Summary of *Pseudomonas aeruginosa* GNAT genes, proteins, and structures.

Pseudomonas gene	UniProt ID	PDB IDs of crystal structures
PA0115	Q9I717	1xeb
PA0249	Q9I6P0	
PA0478	Q9I640	2fe7
PA0483	Q9I635	
PA0711	Q9I5L7	
PA1062	Q9I4R2	
PA1377	Q9I3W7	2vi7
PA1428	Q9I3R6	
PA1472	Q9I3P0	
PA1749	Q9I2Y8	
PA1885	Q9I2L2	
PA1928	Q9I2H6	
PA1943	Q9I2G1	
PA2271	Q9I1K2	
PA2578	Q9I0Q8	3owc
PA2631	Q9I0K7	
PA3127	Q9HZ97	
PA3248	Q9HYZ2	
PA3270	Q9HYX1	1yre
PA3368	Q9HYN3	
PA3460	Q9HYE6	
PA3944	Q9HX72	
PA3945	Q9HX71	
PA4026	Q9HX01	2eui
PA4114	Q9HWR6	
PA4166	Q9HWL5	
PA4166	Q9HWL5	
PA4534	Q9HVP3	4ubr, 5ib0
PA4678	Q9HVB7	
PA4794	Q9HV14	3kkw, 4kot, 4oae, 4l8a, 4oad, 4kub, 4klw, 4kow, 4klv, 4kua, 4kov, 4l89, 4kos, 4kor, 4kou, 4koy, 3pgp, 4m3s, 4kox
PA4863	Q9HUV0	
PA4866	Q9HUU7	1yvo, 2j8m, 2j8n, 2j8r, 2bl1
PA5204	P22567	
PA5432	Q9HTD4	
PA5433	Q9HTD3	
PA5475	Q9HT95	

* Based on predictions from InterPro, each protein listed is either annotated as a GNAT or contains a GNAT domain.

Figure S1. Size exclusion chromatograms of PA3944 protein with different buffers in the absence (apoenzyme) and presence of AcCoA (haloenzyme) (A), calculated molecular weights (B), and calibration curve (C). See Methods for more specific details.

A



B

Apoenzyme				
pH	6	7	8	9
Volume [ml]	68.5	68.1	67.9	68.5
Calculated mass [kDa]	24.2	24.7	25.0	24.2
Holoenzyme				
pH	6	7	8	9
Volume [ml]	68.6	68.4	67.9	69.0
Calculated mass [kDa]	24.0	24.3	25.0	23.5

C

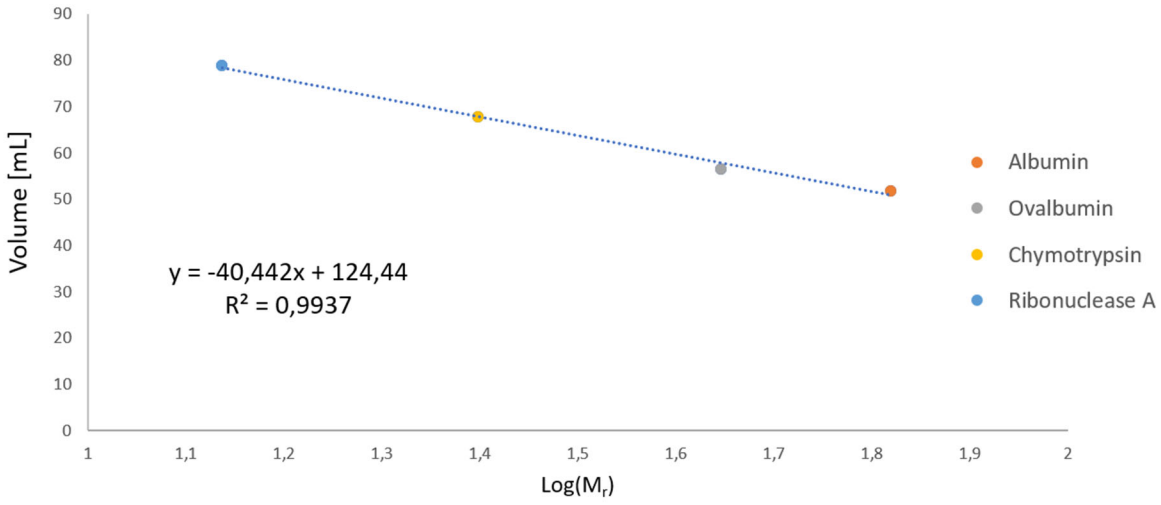

Volume [mL]
90
80
70
60
50
40
30
20
10
0
1
1,1
1,2
1,3
1,4
1,5
1,6
1,7
1,8
1,9
2
Log(M$_r$)
y = -40,442x + 124,44
R² = 0,9937
Albumin
Ovalbumin
Chymotrypsin
Ribonuclease A

Figure S2. Full scan ESI mass spectrum of a reaction mixture containing polymyxin B and AcCoA in A) absence of PA3944 enzyme and B) presence of PA3944 enzyme. The electrospray mass spectra of the polymyxin peptide mixture (Figure S2A) show protonated $(M+H)^+$ and cationized $(M+Na)^+$ molecule ions generated from form B1 at m/z 1203.9 and at m/z 1225.8, respectively. Smaller abundances of the protonated molecule ion formed from the B2 form are observed at m/z 1189.7. No acetylated products are observed without the enzyme present in the reaction mixture. When the enzyme is present in the reaction mixture (Figure S2B) only ions generated from the acetylated polymyxin are observed. Protonated molecule ions from the B2 and B1 forms modified by single acetyl groups are observed at m/z 1231.3 and m/z 1245.3. Ions containing one $(M+Na)^+$ and two sodium ions $(M+2Na)^+$ are formed from the B1 containing one acetyl group are observed at m/z 1267.8 and m/z 1289.7. No unmodified polymyxin or polymyxin molecules modified by multiple acetyl groups are observed in the Figure 2B spectra. See Methods for more specific details about sample preparation (positive ion mode).

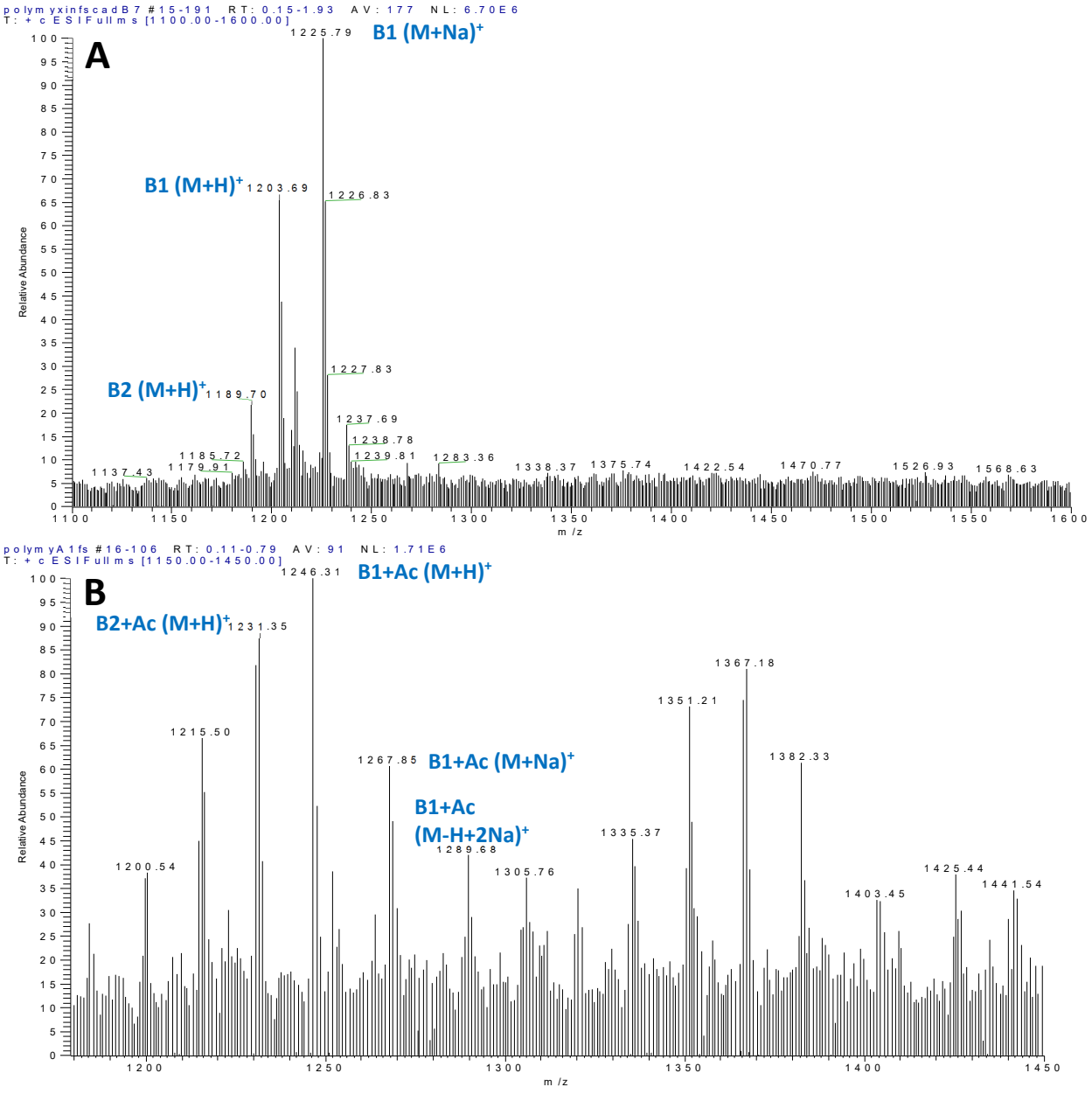


Figure S3. Product ion mass spectra of the protonated molecule ion of **A)** unmodified polymyxin B that was observed at m/z 1204 and **B)** polymyxin containing one acetyl group that was observed at m/z 1246 (positive ion mode), showing loss of water as the primary fragmentation in both cases.

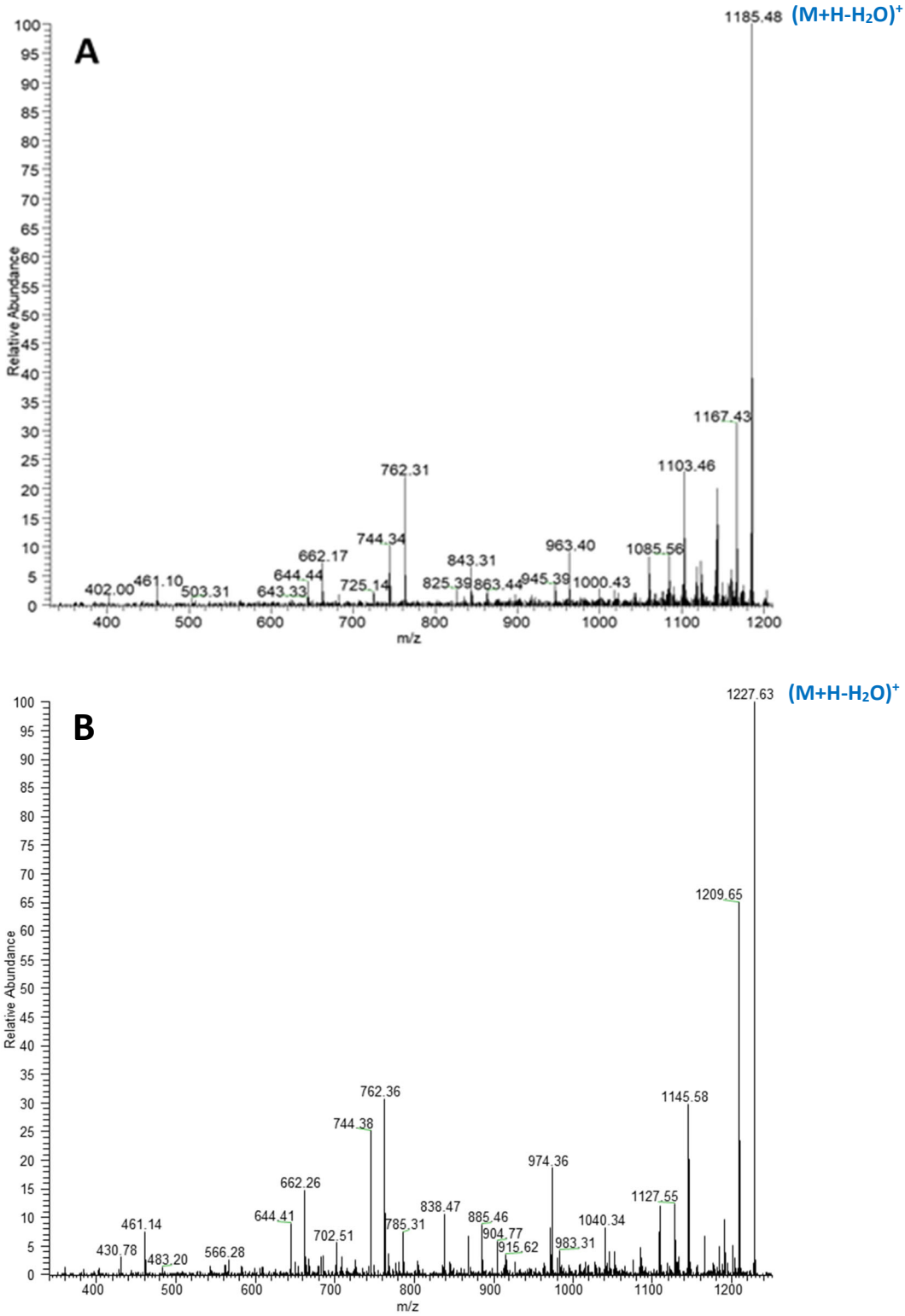


Figure S4. Full scan ESI mass spectra of a reaction mixture containing colistin A and B and AcCoA A) in the absence of PA3944 enzyme and B) in the presence of PA3944 enzyme. See Methods for more specific details about sample preparation (positive ion mode).

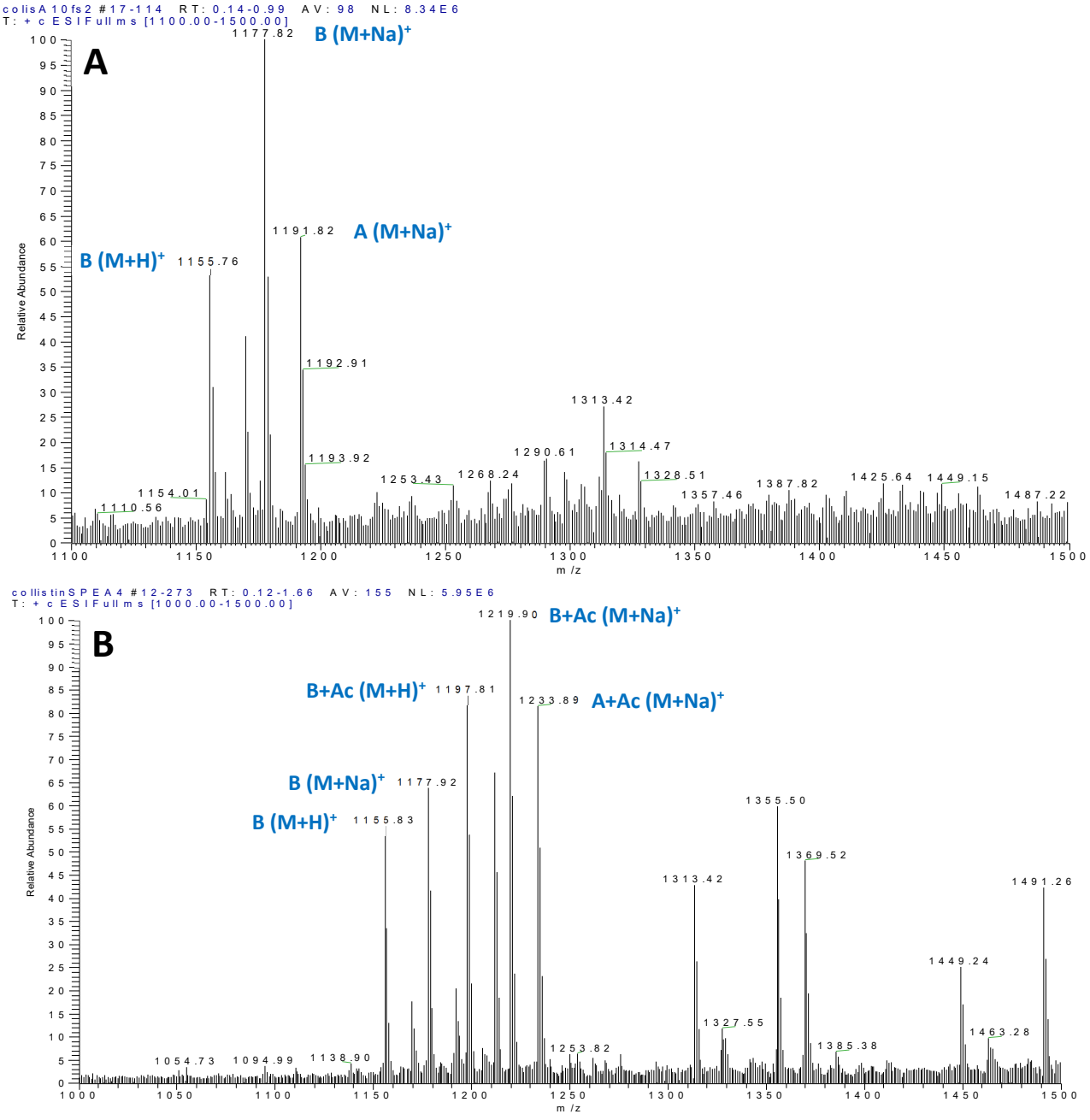


Figure S5. Product ion mass spectra of the protonated molecule ion of **A)** unmodified colistin that had been observed at m/z 1156 and **B)** of the colistin ion containing one acetyl group (positive ion mode) that had been observed at m/z 1198. Loss of water was observed as the primary fragmentation in both cases.

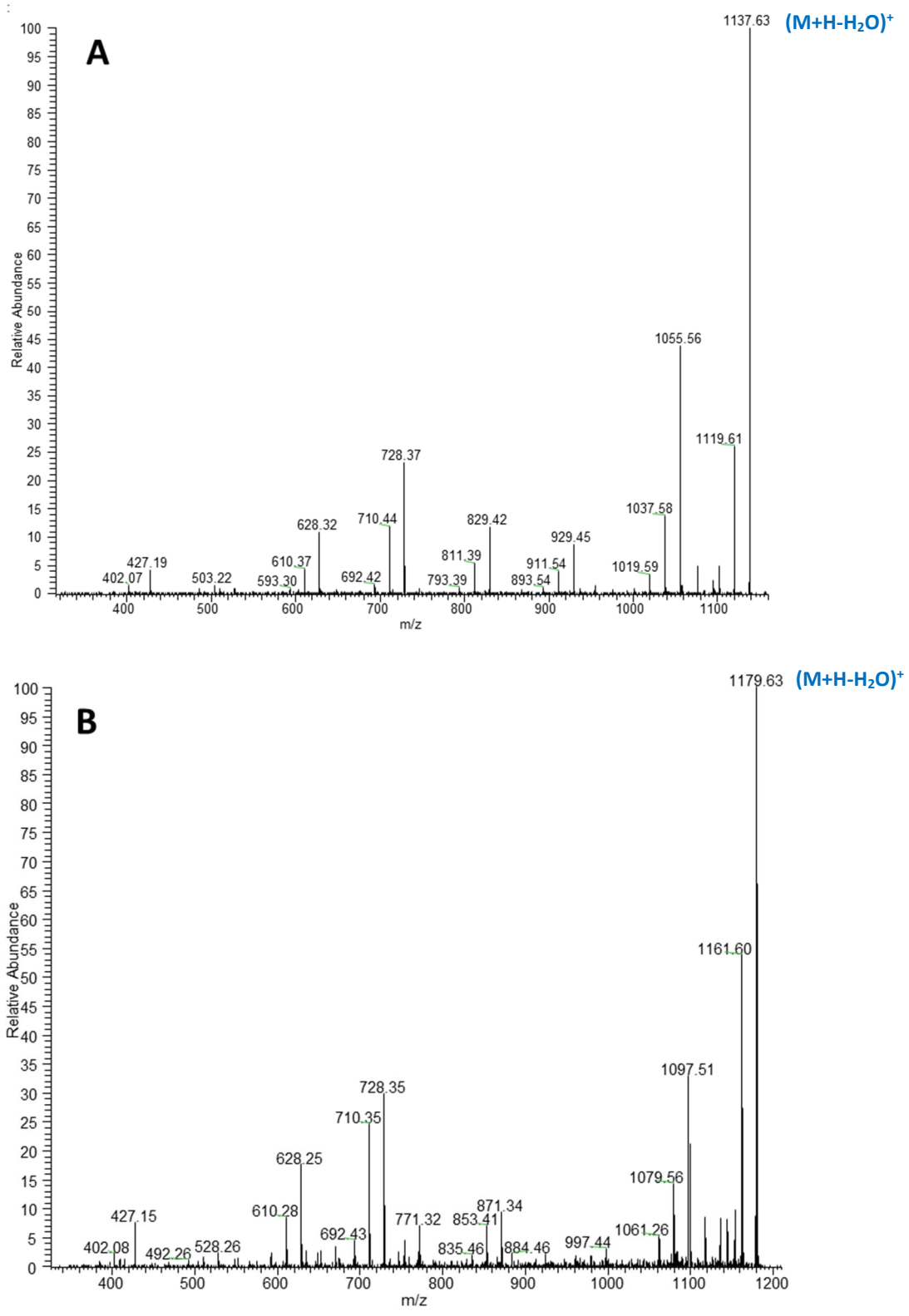
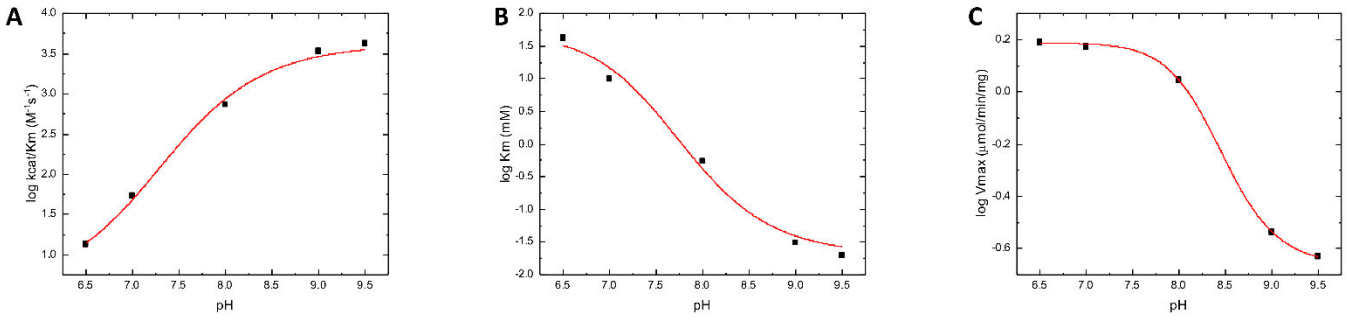
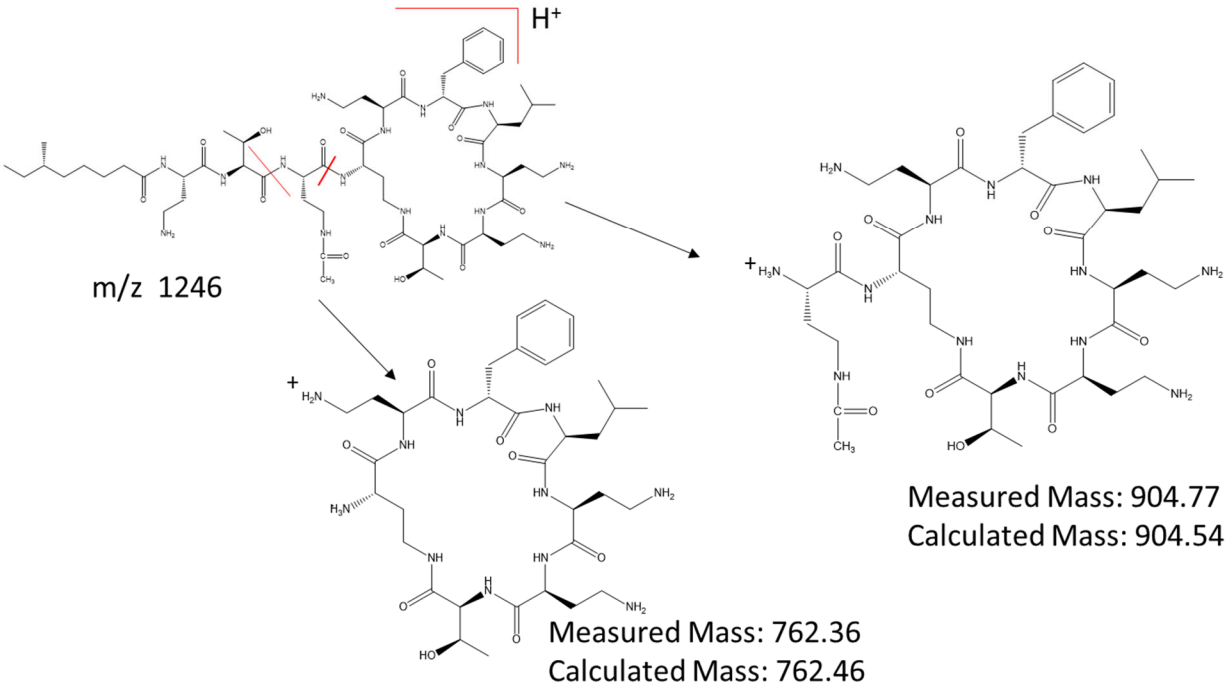


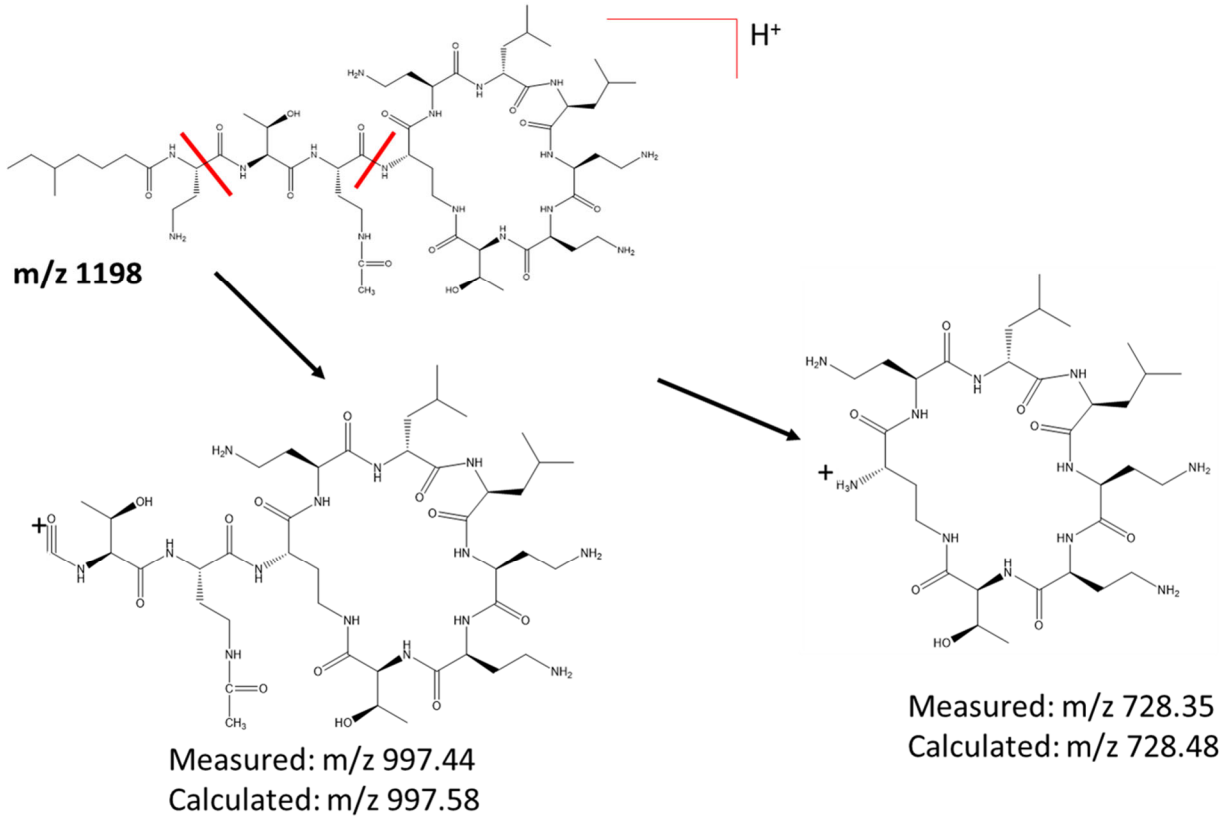
Figure S6. PA3944 pH profiles with polymyxin B as acceptor substrate. A) log k_{cat}/K_m ($M^{-1}s^{-1}$) as a function of pH, B) log K_m (mM) as a function of pH, and C) log V_{max} (µmol/min/mg) as a function of pH. Kinetic parameters used for the plots were calculated from non-linear regression fittings of the substrate saturation curves in Figure 6B to the Hill equation as described in materials and methods.



Scheme S1. Product ion formation indicative of the position of acetyl group attachment in the enzyme catalyzed acetylation of polymyxin B.



Scheme S2. Product ion formation process indicative of the position of acetyl group attachment in the enzyme catalyzed colistin acetylation.



References

1. Filippova, E. V.; Kuhn, M. L.; Osipiuk, J.; Kiryukhina, O.; Joachimiak, A.; Ballicora, M. A.; Anderson, W. F. (2015) A Novel Polyamine Allosteric Site of SpeG from Vibrio cholerae Is Revealed by Its Dodecameric Structure. *J. Mol. Biol. 427*, 1316-1334.

2. Blatzer, M.; Schrettl, M. F.; Sarg B FAU - Lindner, Herbert,H.; FAU, L. H.; Pfaller, K. F.; Haas, H. (2011) SidL, an Aspergillus fumigatus transacetylase involved in biosynthesis of the siderophores ferricrocin and hydroxyferricrocin. *Appl. Environ. Microbiol. 77,* 4959-66.

3. Fukuchi, J.; Kashiwagi, K. F.; Takio, K. F.; Igarashi, K. (1994) Properties and structure of spermidine acetyltransferase in Escherichia coli. *J. Biol. Chem. 269,* 22581-5.

4. Davies, A. M.; Tata, R.; Beavil, R. L.; Sutton, B. J.; Brown, P. R. (2007) L-Methionine sulfoximine, but not phosphinothricin, is a substrate for an acetyltransferase (gene PA4866) from Pseudomonas aeruginosa: structural and functional studies. *Biochemistry 46*, 1829-1839.

5. Majorek, K. A.; Kuhn, M. L.; Chruszcz, M.; Anderson, W. F.; Minor, W. (2013) Structural, functional, and inhibition studies of a Gcn5-related N-acetyltransferase (GNAT) superfamily protein PA4794: a new C-terminal lysine protein acetyltransferase from pseudomonas aeruginosa. *J. Biol. Chem. 288*, 30223-30235.

6. Kuhn, M. L.; Majorek, K. A.; Minor, W.; Anderson, W. F. (2013) Broad-substrate screen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity. *Protein Sci. 22*, 222-230.

7. Pei, J.; Tang M FAU - Grishin, Nick,V.; Grishin, N. V. (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res. 36,* W30-4.

8. Pei, J.; FAU, K. B.; Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res. 36,* 2295-300.

9. Krissinel, E.; Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst. D60,* 2256-68.