Solving the Conundrum: Widespread Proteins Annotated for Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine

Nicholas O. Schneider

Lambros J. Tassoulas

Danyun Zeng

Amanda J. Laseke

Nicholas J. Reiter

See next page for additional authors

Authors

Nicholas O. Schneider, Lambros J. Tassoulas, Danyun Zeng, Amanda J. Laseke, Nicholas J. Reiter, Lawrence P. Wackett, and Martin St. Maurice

Chemistry Faculty Research and Publications/College of Arts and Sciences

Solving the Conundrum: Widespread Proteins Annotated for Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine

Nicholas O. Schneider
Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin
Lambros J. Tassoulas
BioTechnology Institute, University of Minnesota, St. Paul, Minnesota
Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, St. Paul, Minnesota
Danyun Zeng
Department of Chemistry, Marquette University, Milwaukee, Wisconsin
Amanda J. Laseke
Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin
Nicholas Reiter

Department of Chemistry, Marquette University, Milwaukee, Wisconsin
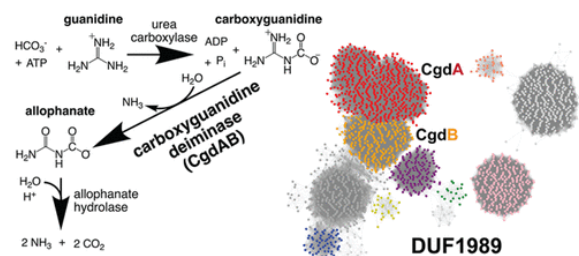
Lawrence P. Wackett

BioTechnology Institute, University of Minnesota, St. Paul, Minnesota
Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, St. Paul, Minnesota

Martin St. Maurice

Department of Biology, Marquette University, Milwaukee, Wisconsin

Abstract

Free guanidine is increasingly recognized as a relevant molecule in biological systems. Recently, it was reported that urea carboxylase acts preferentially on guanidine, and consequently, it was considered to participate directly in guanidine biodegradation. Urea carboxylase combines with allophanate hydrolase to comprise the activity of urea amidolyase, an enzyme predominantly found in bacteria and fungi that catalyzes the carboxylation and subsequent hydrolysis of urea to ammonia and carbon dioxide. Here, we demonstrate that urea carboxylase and allophanate hydrolase from *Pseudomonas syringae* are insufficient to catalyze the decomposition of guanidine. Rather, guanidine is decomposed to ammonia through the combined activities of urea carboxylase, allophanate hydrolase, and two additional proteins of the DUF1989 protein family, expansively annotated as urea carboxylase-associated family proteins. These proteins comprise the subunits of a heterodimeric carboxyguanidine deiminase (CgdAB), which hydrolyzes carboxyguanidine to *N*-carboxyurea (allophanate). The genes encoding CgdAB colocalize with genes encoding urea carboxylase and allophanate hydrolase. However, 25% of urea carboxylase genes, including all fungal urea amidolyases, do not colocalize with *cgdAB*. This subset of urea carboxylases correlates with a notable Asp to Asn mutation in the carboxyltransferase active site. Consistent with this observation, we demonstrate that fungal urea amidolyase retains a strong substrate preference for urea. The combined activities of urea carboxylase, carboxyguanidine deiminase and allophanate hydrolase represent a newly recognized pathway for the biodegradation of guanidine. These findings reinforce the relevance of guanidine as a biological metabolite and reveal a broadly distributed group of enzymes that act on guanidine in bacteria.

A subset of fungi and bacteria use the enzyme urea amidolyase (UAL) as an alternative to urease in catalyzing urea decomposition.(2,3) UAL is a multifunctional, biotin-dependent enzyme that catalyzes the ATP-dependent carboxylation of urea to allophanate (*N*-carboxyurea) and the subsequent hydrolysis of allophanate into ammonia and CO_2.(5) The complete activity of UAL is composed of three distinct reactions divided over two enzymatic components: urea carboxylase (UC; E.C. 6.3.4.6) and allophanate hydrolase (AH; E.C. 3.5.1.54). UC first catalyzes the carboxylation of a covalently tethered biotin cofactor from HCO_3^- in the biotin carboxylase domain, with the concomitant cleavage of ATP, and then catalyzes a carboxyl group transfer from carboxybiotin to urea in the carboxyltransferase domain, generating allophanate. Allophanate subsequently diffuses to the active site of AH, where it is hydrolyzed to ammonia and CO_2. The enzyme activities for UC and AH are encoded

on separate but proximally related genes in bacteria, while, in most fungi, the two genes are fused to encode a multifunctional UAL, with the N-terminal AH connected by a short linker to the C-terminal UC.(6–8)

UC and AH display a close evolutionary and functional association. It has been suggested that UC and AH coevolved in bacteria and, following horizontal gene transfer, subsequently fused into a single UAL gene in fungi.(3) Interestingly, in many bacteria, there are several additional genes of unknown function located in proximity to the genes encoding UC and AH, arranged in the context of an operon. This operon includes a likely ABC transporter and two proteins of unknown function, designated "urea amidolyase associated proteins (UAAP)".(4) UAAP 1 and 2 are both members of a Pfam Domain of Unknown Function known as DUF1989.

Curiously, the enzymatic function of UAL is redundant with the enzyme urease and there are many bacteria that contain both.(3,9,10) This redundancy raises some interesting questions. Why would the biotin cofactor-dependent, ATP-consuming activity of UAL be required (or evolutionarily retained) when the activity of the broadly distributed urease accomplishes, in a single thermodynamically favorable reaction, the identical task? The decomposition of urea by urease does not require the input of energy, nor must it overcome any challenges of intermediate channeling between multiple, distinct active sites. It has been postulated that, for a subset of fungi, replacing urease with UAL serves to free the organism from the costly demands of regulating nickel.(2) However, urease is broadly distributed in the bacteria,(11) and bacteria require transition metals like nickel and cobalt that are otherwise dispensable in the higher fungi.(12) Thus, for the many bacteria that encode UC and AH, transition metal regulation does not offer a satisfying explanation for the UC-AH/urease redundancy. An alternative explanation emerged recently, when it was determined that bacterial UC and AH are typically under the control of a guanidine riboswitch and that UC displays a strong substrate preference for guanidine over urea.(13) This observation led to the proposal that the primary function of UC and AH in bacteria is not to act on urea but rather to participate in the sequential decomposition of guanidine to NH_3 and CO_2. This hypothesis offered an explanation for the broad distribution and functional redundancy of UC enzymes and raised the potential that guanidine serves as a relevant nitrogen source for bacteria.

Here, using enzymes from *Pseudomonas syringae*, we test the hypothesis that UC and AH act on guanidine and demonstrate that these two enzymes alone are insufficient to catalyze the decomposition of guanidine. We clarify that the proteins annotated as urea amidolyase associated proteins, UAAP1 and UAAP2, are subunits of a heteromeric carboxyguanidine deiminase (CgdAB), and we demonstrate that the combined activities of UC, AH, and CgdAB serve to decompose guanidine to NH_3. The CgdAB enzyme represents a newly recognized and broadly distributed class of deiminase. The genes encoding CgdAB fall within a large, putative metalloenzyme superfamily and are highly correlated with the genes for UC and AH. Interestingly, the substrate preference for guanidine is not universal among all UC enzymes. We show that guanidine is a very poor substrate for the fungal UAL enzymes from *Saccharomyces cerevisiae* and *Candida albicans*, suggesting an evolutionary divergence in UC that is governed by substrate preference.

Materials and Methods

General

All reagents, unless mentioned otherwise, were purchased from Millipore-Sigma. Guanidine hydrochloride (electrophoresis grade) from FisherChemical was obtained from Fisher Scientific. NMR spectra were acquired on a Varian VNMRS 600 MHz spectrometer, equipped with a *z*-axis pulsed field gradient triple resonance (^1H, ^{13}C, ^{15}N) cryogenic probe. The molecular weights used for calculating k_{cat} in these studies were predicted using the recombinant protein sequence, yielding the following values: *P. syringae* UC (*Ps*UC) = 135 781 g mol^{-1}; *S. cerevisiae* UAL (*Sc*UAL) = 208 476 g mol^{-1}; *C. albicans* UAL (*Ca*UAL) = 207 094 g mol^{-1}.

Enzyme Purification

UC and AH from *P. syringae* pv tomato str. DC3000 (strain ATCC BAA-871D-5), both encoding N-terminal recombinant (His)$_8$-tags, were overexpressed in BL21(DE3) *E. coli* cells as previously described.(4) For *P. syringae* AH (*Ps*AH) expression, *E. coli* BL21(DE3) cells were transformed with a pET28a-(His)$_8$-TEV vector harboring the *Ps*AH gene. For *Ps*UC expression, *E. coli* BL21(DE3) cells were cotransformed with a pET28a-(His)$_8$-TEV vector harboring *Ps*UC and vector pCY216 encoding *E. coli* biotin protein ligase (BirA).(14) For the expression of *Ps*UC in BL21(DE3) cells, a 250 mL overnight culture in LB media was used to inoculate 8 × 1 L of M9 minimal media containing 50 µg/mL kanamycin and 30 µg/mL chloramphenicol. For the expression of *Ps*AH in BL21(DE3), a 250 mL overnight culture in LB media was used to inoculate 8 × 1 L of M9 minimal media containing 25 µg/mL kanamycin. All cultures were grown at 37 °C to an OD$_{600}$ of 0.8–1.0, at which point the cultures were chilled on ice for ~15 min. For *Ps*UC coexpression, the cultures were induced with 20 mM arabinose and 1 mM isopropyl 1-thio-β-d-galacto-pyranoside (IPTG). For *Ps*AH expression, the cultures were induced with 1 mM IPTG. The cultures were incubated at 16 °C with shaking for ~16 h prior to harvesting the cells by centrifugation.

*Ps*AH and *Ps*UC were purified using sequential Ni^{2+}-affinity and ion-exchange chromatography. Prior to sonication, the cell paste equivalent to 8 L of liquid culture was resuspended in 150 mL of Buffer A [15 mM Tris-HCl, pH 7.8; 5 mM imidizole, 200 mM NaCl, 0.5 mM EGTA, 6 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 µM pepstatin, and 5 µM epoxysuccinyl-l-leucylamido (4-guanida) butane (E-64)]. The cells were sonicated for 8 min while maintaining the temperature below 10 °C. The sonicate was centrifuged for 30 min at 48 000*g* and at 4 °C. The resulting supernatant was loaded onto a 5 mL Ni^{2+}-affinity column at a flow rate of 2 mL/min and washed with Buffer A containing 10 mM imidazole. A gradient elution of 10–250 mM imidazole was used to elute the protein from the column. The elution was dialyzed overnight against Buffer B (20 mM triethanolamine, 50 mM NaCl, 1 mM EGTA, 2 mM 1,4-dithiothreitol (DTT), pH 8.0) at 4 °C. The dialyzed elution was loaded on Q-Sepharose fast flow resin at a flow rate of 2 mL/min and eluted from the column using a gradient of 50–1000 mM NaCl. The eluted protein was dialyzed twice against Buffer C (20 mM Bis-Tris propane, 15 mM NaCl, 10 mM MgCl$_2$, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), pH 7.2) at 4 °C for 4 h. *Ps*UC was concentrated using a 100 kDa molecular weight centrifugal filter, and *Ps*AH was concentrated using a 30 kDa molecular weight centrifugal filter. The concentrated protein was drop-frozen in liquid nitrogen and stored at −80 °C.

Carboxyguanidine deiminase subunits α and β from *P. syringae* (CgdA and CgdB; renamed from urea amidolyase associated protein; urea carboxylase-associated family protein [*P. syringae* group]; accession numbers WP_005764729.1 and WP_005764727.1, respectively) were expressed and purified as previously described.(4) CgdB with an N-terminal calmodulin (CaM) binding domain expressed from vector pKLD66nCBP and N-terminally (His)$_8$-tagged CgdA expressed from a modified pET28a-(His)$_8$-TEV vector were coexpressed in *E. coli* BL21Star(DE3). Proteins were produced using an 8 L batch culture in M9 minimal medium (containing 200 µg/mL ampicillin and 50 µg/mL kanamycin). After reaching an OD$_{600}$ of 1.0 at 37 °C, the culture was induced for 16 h at 16 °C with 1 mM IPTG prior to harvesting the cells by centrifugation. The cell paste from the equivalent of 4 L of culture was resuspended in 150 mL lysis buffer (20 mM Tris Base, pH 7.8; 5 mM imidazole; 200 mM NaCl; 0.5 mM EGTA; 5 µM E-64; 1 µM pepstatin A; and 1 mM PMSF). The cells were sonicated for 8 min while maintaining the temperature below 10 °C. The sonicate was centrifuged for 30 min at 48 000 g and 4 °C. The supernatant was loaded on a 5 mL Ni^{2+}-affinity column at a flow rate of 2 mL/min. The column was subsequently washed with 100 mL of lysis buffer and 100 mL of nickel wash buffer (20 mM Tris, pH 7.8; 10 mM imidazole; 200 mM NaCl; 0.5 mM EGTA). The protein was eluted with a gradient of 20–250 mM imidazole in nickel wash buffer. Fractions containing the protein were pooled and dialyzed 3 times for at least 4 h at 4 °C against storage buffer (20 mM Bis-Tris Propane, pH 7.8; 50 mM NaCl; 1 mM TCEP). After dialysis, the protein was loaded on a 10 mL column of CaM resin at a flow rate of 2 mL per minute, at 4 °C. The column was washed with CaM wash buffer

(20 mM Tris Base, pH 7.8; 200 mM NaCl; and 2 mM CaCl$_2$). Following the wash, the protein was eluted off the column with 50 mL CaM elution buffer (20 mM Tris Base, pH 7.8; 200 mM NaCl; and 2 mM EGTA). Fractions containing protein were pooled and dialyzed against storage buffer two times for at least 4 h. The resulting purified CgdAB protein was concentrated using a 30 kDa molecular weight centrifugal filter (final concentrations ranging from 2–10 mg/mL), drop frozen in liquid nitrogen and stored at −80 °C.

Enzyme Assays

ATP cleavage was measured using a standard enzyme coupled assay that employed the activities of pyruvate kinase and lactate dehydrogenase, as previously described.(4) The enzyme assay was performed in a solution containing 50 mM HEPES (pH 7.3), 50 mM NaCl, 8 mM MgSO$_4$, 8 mM NaHCO$_3$, 0–12 mM urea or 0–10 mM guanidine, 1.5 mM phosphoenolpyruvate, 0.15 mM NADH, 100 µM ATP, 5 U mL^{-1} of pyruvate kinase and 12.5 U mL^{-1} of lactate dehydrogenase at a temperature of 25 °C. To initiate the reaction, *Ps*UC or UAL was added to a final assay concentration of 25–50 or 36 µg mL^{-1}, respectively. Kinetic data were collected either using a Shimadzu UV-1800 spectrophotometer in 1 mL quartz cuvettes at a final assay volume of 1 mL or in 96-well plates using a SpectraMax i3x Multi-Mode plate reader at a final assay volume of 200 µL.

The production of ammonia was measured by a standard coupled assay employing the activity of glutamate dehydrogenase (GDH). These assays were performed in a solution of 50 mM HEPES (pH 7.3), 50 mM NaCl, 8 mM MgSO$_4$, 8 mM NaHCO$_3$, 50 mM urea or 10 mM guanidine, 20 mM α-ketoglutarate, 0.15 mM NADH, 100 µM ATP, 25 µg mL^{-1}*Ps*UC, 100 µg mL^{-1}*Ps*AH, and 20 U mL^{-1} GDH at room temperature. Where mentioned, *Ps*CgdAB was added at a final concentration of 25 µg mL^{-1}. The assay was measured using a SpectraMax i3x Multi-Mode Plate Reader at a final assay volume of 200 µL. To reduce the contribution from contaminating ammonia introduced from the urea and α-ketoglutarate stocks, all assay components, with the exception of UC, AH, and CgdAB, were mixed together to eliminate excess ammonia. Once the change in absorbance became undetectable (∼15–20 min; ΔAbs ∼0.1 AU), the appropriate combination of *Ps*UC, *Ps*AH, and/or *Ps*CgdAB was added to initiate the assay.

The apparent inhibition of *Ps*UC by guanidine was tested using the GDH coupled enzyme assay as described above, but with final concentrations of guanidine fixed at 0, 0.075, 0.15, and 0.3 mM and measured against urea concentrations of 0, 1.5, 3, 8, 12, and 15 mM.

The biotin-dependence of the ATP cleavage in the presence of urea and guanidine was evaluated in the presence of egg white avidin. *Ps*UC was incubated at a concentration of 25 µg mL^{-1} with a 10-fold molar excess of avidin. The same concentration of *Ps*UC was incubated with assay buffer containing 50 mM HEPES (pH 7.3), 50 mM NaCl, and 8 mM MgSO$_4$ for the negative control. Following a 1 h incubation at room temperature, the ATP cleavage activity was measured as described above, in the presence of 0 and 50 mM urea, and 10 mM guanidine. Kinetic data were collected in 96-well plates using a SpectraMax i3x Multi-Mode Plate Reader at a final assay volume of 200 µL.

NMR-Based Assays

^{13}C NMR experiments with NaH^{13}CO$_3$ were performed at 6 mM MgATP to drive the reaction forward over multiple hours of NMR data acquisition. Assays included 20 mM ^{13}C-sodium bicarbonate, 50 mM HEPES (pH 7.3), 50 mM NaCl, 8 mM MgSO$_4$, 5% D$_2$O, and either 10 mM guanidine or 50 mM urea. The reaction was initiated with the addition of *Ps*UC, to a final concentration of 25 µg mL^{-1}. The *t* = 0 time point was collected for a sample that included the addition of assay buffer in place of both *Ps*UC and ATP. To inactivate the enzymatic reaction, EDTA was prepared in assay buffer and added to the reaction at a final concentration of 10 mM.

^{13}C NMR 1D experiments with ^{13}C-guanidine and ^{13}C-urea were performed in the presence of 6 mM ATP, 5% D$_2$O, 50 mM HEPES (pH 7.3), 50 mM NaCl, 8 mM MgSO$_4$, 20 mM NaHCO$_3$, 250 µg/mL CgdAB, and either 10 mM ^{13}C-

guanidine or ^{13}C-urea. The reaction was initiated by the addition of *Ps*UC to a final concentration of 100 µg/mL. These same experiments were also performed in the presence of 100 µg/mL urease or ~725 µg/mL *Ps*AH.

All NMR experiments were performed at 293 K on a Varian VNMRS 600 spectrometer equipped with a cryogenic probe. For each reaction, a series of ^{13}C 1D spectra were acquired continuously on the same sample, with each time point representing the average spectrum during the designated time frame. An averaged spectrum was obtained in time windows of 1 or 3 h and data was collected for 12 to 21 h after the reaction was initiated with *Ps*UC. All spectra were processed and analyzed in SpinWorks 4 with the identical window function, baseline correction, and processing parameters. The assigned chemical shift resonances follow a Lorenztian distribution pattern and the calculated relative resonance intensities were measured using an identical vertical scale.

Bioinformatics

CgdA and CgdB sequences were mined from the RefSeq genome database by first generating a sequence similarity network (SSN) using the EFI-EST tool to perform pairwise BLAST comparisons on 10,000 related sequences.(15) Cytoscape was used to visualize the clustering in the SSN and identify the clusters containing CgdA and CgdB sequences.(16) This process resulted in a total of 2400 sequences from the NCBI RefSeq database and ~7000 sequences when combined with the EMBL-EBI database. To annotate gene contexts properly, Hidden Markov Models (HMM) were built for CgdA, CgdB as well as frequently co-occurring TetR protein and *ykkC* riboswitch using the software HMMER v 3.1b2.(17) A GitHub repository containing the HMMs used in the study can be accessed with the link at https://github.com/ltassoulas/CgdAB. Additional HMMs from the TIGRFAMs database (J. Craig Venter Institute) were used to annotate UC (TIGR02712), AH (TIGR02713) and urea carboxylase associated transport proteins (TIGR0327, TIGR03428).(18) These models were then used with the tool RODEO (Rapid ORF Description & Evaluation Online) to analyze the gene contexts of 2400 Cgd RefSeq sequences within a ± 8 gene window.(19)

Results

The subtype 1 *ykkC* motif RNA serves as a riboswitch to sense and respond to guanidine.(13) This class of riboswitches, renamed as guanidine-I riboswitches, controls the expression of genes proposed to be involved in degrading or exporting guanidine from bacteria in order to reduce guanidine toxicity. Interestingly, guanidine-I riboswitches are often observed to be associated with UC encoding genes, and moreover, UC from *Oleomonas sagaranensis* displayed an ~40-fold higher catalytic efficiency for ATP cleavage in the presence of guanidine as compared to urea.(13) Consequently, it was suggested, though not directly tested, that UC (renamed "guanidine carboxylase") and the associated AH act directly in degrading guanidine to ammonia and CO_2.(13) Given our prior characterization of several UAL enzyme systems from both fungi and bacteria,(4,20) we sought to reproduce and further elaborate on this initial functional characterization to test the hypothesis of guanidine decomposition in bacteria and more fully describe the relevant metabolic pathway.

Characterization of UC and AH Activities with Guanidine

We assessed the relative rate of ATP cleavage for UC from *P. syringae* (*Ps*UC) in the presence and absence of both urea and guanidine (Figure 1) and observed results similar to those reported for *O. sagaranensis* UC.(7,13) The catalytic efficiency of *Ps*UC for ATP cleavage in the presence of urea is $(2.8 \pm 0.6) \times 10^2$ M^{-1} s^{-1}, compared with $(1.1 \pm 0.5) \times 10^4$ M^{-1} s^{-1} in the presence of guanidine, corresponding to an ~40-fold difference. Notably, the K_M is reduced from a value of 4.9 ± 0.7 mM for urea to a value of 0.21 ± 0.09 mM for guanidine, suggesting that guanidine is the more likely physiological substrate. The kinetic constants determined for ATP cleavage, both in the presence of urea and in the presence of guanidine, are similar to previous reports.(4,7,13)
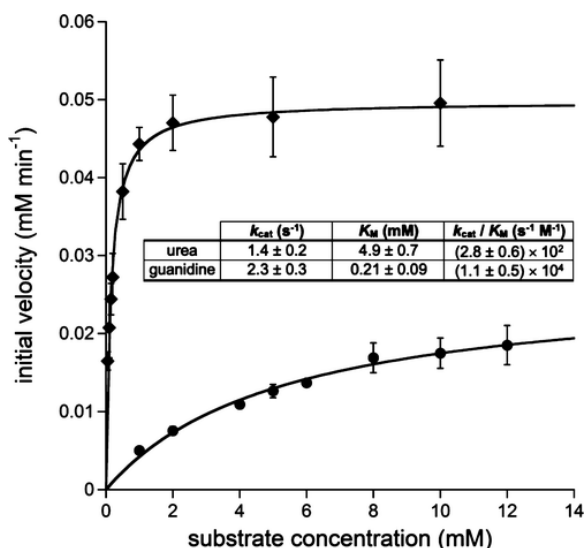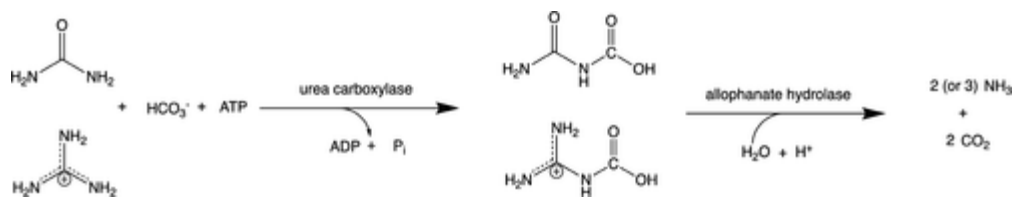
Figure 1. Kinetics of ATP cleavage by *Ps*UC in the presence of urea and guanidine. A representative plot of the initial velocity for ATP cleavage as a function of substrate concentration (guanidine, closed diamonds; urea, closed circles) in the presence of 50 μg mL^{-1}*Ps*UC. The data were fit by nonlinear regression to the standard Michaelis–Menten equation. The inset table presents the average values and standard deviations for the kinetic constants determined from three independent trials. For each trial, three independent initial velocity measurements were recorded for each substrate concentration.

We next sought to determine whether the combined activities of *Ps*UC and AH from *P. syringae* (*Ps*AH) could catalyze the complete decomposition of guanidine to ammonia and CO_2, as hypothesized by Nelson et al.(13) (Scheme 1). With urea as a substrate, ammonia formation was detected using a well-established GDH assay, as previously described.(4) However, in the presence of guanidine, no detectable ammonia was produced by the combined activities of *Ps*UC and *Ps*AH. This result indicated one of two possibilities: either carboxyguanidine is not a product of the reaction catalyzed by UC, or AH is unable to directly deaminate carboxyguanidine.

Scheme 1. Enzymatic Decomposition of Urea by UC and AH and a Previously Proposed Analogous Route for the Decomposition of Guanidine

Evidence for Carboxyguanidine as the Product of UC

Notably, the ATP cleavage assays measure only the first half-reaction of the UC reaction. While the ATP cleavage assay revealed a greatly enhanced catalytic efficiency for ATP cleavage in the presence of guanidine as compared to urea, it does not directly demonstrate the formation of the proposed product, carboxyguanidine. To directly observe carboxyguanidine formation by *Ps*UC, we attempted to monitor the formation of ^{13}C-carboxyguanidine from NaH^{13}CO$_3$ by ^{13}C NMR. This method is not appropriate for obtaining or comparing kinetic data but, instead, offers a direct, qualitative measure of product formation. When incubating *Ps*UC in the presence of ATP, NaH^{13}CO$_3$, and urea, a signal corresponding to the carboxylated product, ^{13}C-allophanate, appeared and gradually increased over 6 h (Figure 2A). Upon inactivating the Mg^{2+}-dependent biotin carboxylase activity of *Ps*UC with EDTA at 6 h, the signal intensity gradually began to diminish, consistent with the

decomposition of allophanate to urea and $H^{13}CO_3^-$ at neutral pH.[21] By comparison, when incubating PsUC with ATP, $NaH^{13}CO_3$, and guanidine, no signal corresponding to ^{13}C-carboxyguanidine was observed in the NMR spectra (Figure 2B). Despite the enhanced catalytic efficiency of ATP cleavage in the presence of guanidine, the absence of a carboxyguanidine signal suggests that either PsUC does not produce carboxyguanidine or the carboxyguanidine produced is insufficiently stable to be detected by this NMR-based assay.
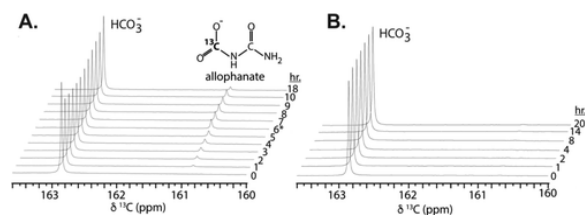

Figure 2. ^{13}C NMR spectroscopic analysis of PsUC incubated with ATP,^{13}C-HCO_3^-, and (A) urea or (B) guanidine. (A) The sample containing 6 mM MgATP, 20 mM ^{13}C-sodium bicarbonate, and 50 mM urea was measured in a 1D ^{13}C NMR experiment (0 h time point), and the reaction was initiated by the addition of PsUC to a final concentration of 25 μg/mL. After 6 h of the reaction incubation, EDTA was added as a quenching agent to 10 mM, denoted by an asterisk. The presence of EDTA terminates the enzymatic reaction, as previously demonstrated for pyruvate carboxylase.[1] The signal corresponding to the carboxy carbon atom of allophanate is indicated at 160.9 ppm. (B) The sample containing 6 mM MgATP, 20 mM ^{13}C-sodium bicarbonate, and 10 mM guanidine was measured in a 1D ^{13}C NMR experiment (0 h time point), and the reaction was initiated by the addition of PsUC to a final concentration of 25 μg/mL. All spectra were calibrated and scaled to the 1 h reaction time point.

These results suggest a discrepancy: the catalytic efficiency for ATP cleavage is ~40 fold higher in the presence of guanidine compared to urea, but ammonia is not produced from guanidine by the combined action of PsUC and PsAH, nor is the proposed product, carboxyguanidine, observed by ^{13}C NMR. A possible explanation for these results is that guanidine simply serves to accelerate the rate of abortive ATP cleavage in the biotin carboxylase domain of UC, without participating directly as a carboxyl group acceptor. There are several documented examples of allosteric effectors and cosubstrates altering the rate of abortive ATP cleavage in the related biotin-dependent enzyme, pyruvate carboxylase.[22,23] To determine whether the enhanced catalytic efficiency of ATP cleavage results from guanidine acting as a carboxyl-group acceptor in the full biotin-dependent reaction, we assessed the biotin dependence by performing the ATP cleavage reaction in the presence of avidin. This approach has often been used in UC and other biotin-dependent enzymes.[7,24−26] If the enhanced rate of ATP cleavage requires that guanidine accept a carboxyl group from biotin in the overall reaction, the presence of avidin would be expected to significantly reduce the rate of ATP cleavage in the presence of guanidine. Indeed, the rate of ATP cleavage in the presence of both urea and guanidine was almost completely eliminated in the presence of avidin (Figure S1). This result strongly suggests that guanidine serves as a carboxyl group acceptor substrate in the overall biotin-dependent reaction catalyzed by UC.

There is presently only a single available crystal structure of a UC enzyme, determined from the UC domain of *Kluyvermyces lactis* UAL.[27,28] While the precise orientation of urea reported in this structure is subject to interpretation, the structure clearly indicates the potential for either urea or guanidine to be accommodated in the carboxyltransferase domain active site. It has been suggested that the more positively charged guanidinium ion is likely to be better accommodated than urea by the conserved Asp residue (Asp 1584 in *K. lactis*UAL) in the active site[13] (Figure S2). To confirm that guanidine can bind in the carboxyltransferase domain active site, we assessed the potential for guanidine to act as an apparent competitive inhibitor of urea in the conversion of urea to NH_3 and CO_2 by the combined actions of PsUC and PsAH. Indeed, guanidine acts as a competitive inhibitor of urea in this assay (Figure 3), with an estimated value of $K_i^{apparent} = 80$ μM, consistent with the measured K_M for guanidine of 0.21 mM (Figure 1). Taken together, our data confirm that guanidine can bind in the CT domain

active site of UC and that it likely acts to accept a carboxyl group from the biotin cofactor. The inability to observe carboxyguanidine by ^{13}C NMR, therefore, results from the instability of the carboxyguanidine product on the time scale of the NMR experiment. The high instability of carboxyguanidine is supported by the complete absence of its description in the peer-reviewed literature (searching SciFinder containing >10^8 chemical substances), despite guanidine having been known since 1861(29) and with 1 833 429 guanidinium compounds listed in SciFinder at the time of this writing. Nevertheless, the inability of *Ps*UC and *Ps*AH to directly convert guanidine into NH_3 and CO_2 suggests that, if carboxyguanidine is produced by *Ps*UC, an additional factor(s) is necessary for the enzyme-catalyzed decomposition of carboxyguanidine. Given that carboxyguanidine has the potential to rapidly decompose back to guanidine and HCO_3^- in neutral solution, this additional factor is expected to have enzymatic activity.
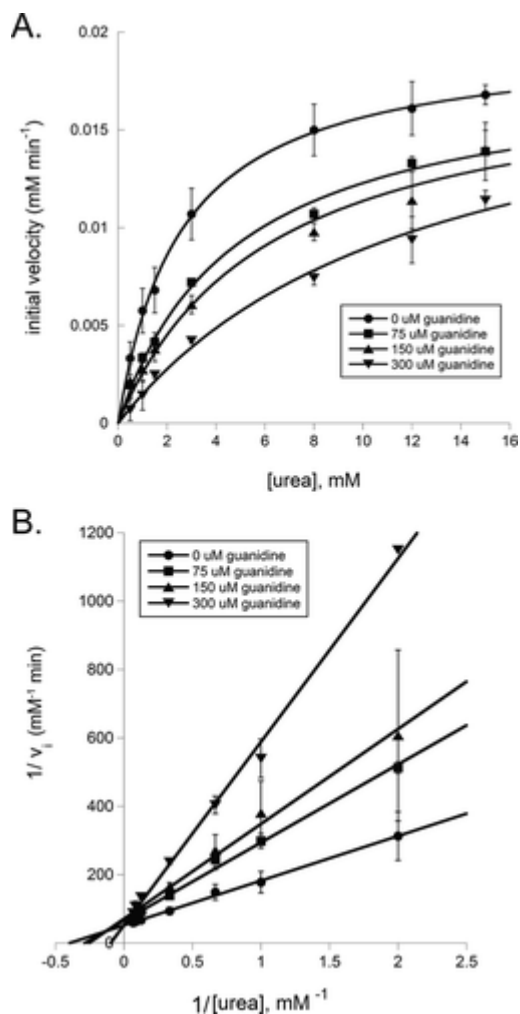

Figure 3. Guanidine is mutually exclusive with urea in *Ps*UC. (A) NH_3 production from the decomposition of urea through a 1:1 molar ratio of *Ps*UC and *Ps*AH, measured using the GDH-coupled assay. Initial velocity rates were measured in the presence of four fixed guanidine concentrations (0 µM, circles; 75 µM, squares; 150 µM, triangles; 300 µM, inverse triangles). The data were fit by nonlinear regression to the standard Michaelis–Menten equation, where $K_M^{app} = K_M(1 + [I]/K_I)$. A replot of K_M^{app} vs [I] revealed a K_I^{app} = 79 µM. (B) Double reciprocal plot of the initial velocity data, revealing a competitive inhibition profile that is consistent with mutually exclusive binding of guanidine and urea to *Ps*UC.

UAAP 1 and 2 Are Subunits of a Heterodimeric Carboxyguanidine Deiminase

We have previously reported that, in bacteria, the genes encoding two proteins of unknown function, annotated as UAAP1 and UAAP2, are often located in close proximity with the genes encoding UC and AH

(Figure 4A).(4) More recently, it has been suggested that these urea amidolyase associated proteins are among the gene products controlled by the guanidine-I riboswitch and that they may contribute to guanidine degradation.(13) The recombinant UAAP1 and UAAP2 proteins were soluble only when they were coexpressed together and they purified from *E. coli* as a heteromer (Figure S3). We have previously demonstrated that UAAP1/2 from *P. syringae* (*Ps*UAAP1/2; NCBI WP_005764729.1 and WP_005764727.1, respectively) does not contribute to the degradation of urea by *Ps*UC and *Ps*AH.(4) We now sought to evaluate whether these proteins of unknown function contribute to the degradation of guanidine, either directly or through the combined activities of UC and AH. We assessed the production of ammonia from urea and guanidine in the presence of various combinations of *Ps*UC, *Ps*AH, and *Ps*UAAP1/2. *Ps*UAAP1/2 is unable to directly act on guanidine: neither *Ps*UAAP1/2 alone nor in combination with *Ps*AH resulted in any guanidine decomposition to NH_3 (data not shown). As previously reported,(4) the combined activities of *Ps*UC and *Ps*AH were sufficient to degrade urea to NH_3 in the presence of ATP and HCO_3^-, and *Ps*UAAP1/2 had no effect on the rate of ammonia production from urea (Figure 4B). However, when *Ps*UC, *Ps*AH, and *Ps*UAAP1/2 were incubated with guanidine, ATP, and HCO_3^-, a significant rate of NH_3 production was detected. Ammonia was also produced when *Ps*UC and *Ps*UAAP1/2 were incubated with these same substrates in the absence of *Ps*AH, at approximately one-third of the rate in the presence of *Ps*AH (Figure 4B). We conclude that UAAP1/2 acts on the carboxyguanidine produced by *Ps*UC to generate ammonia. Two possible mechanisms can be envisioned for this reaction. The decomposition of carboxyguanidine can proceed through an amidinohydrolase-like mechanism (e.g., arginase) to produce urea and carbamate or, alternatively, it can proceed through an iminohydrolase-like mechanism (e.g., arginine deiminase) to produce allophanate and ammonia (Scheme 2).(30) Either mechanism is consistent with the observations summarized in Figure 4B.
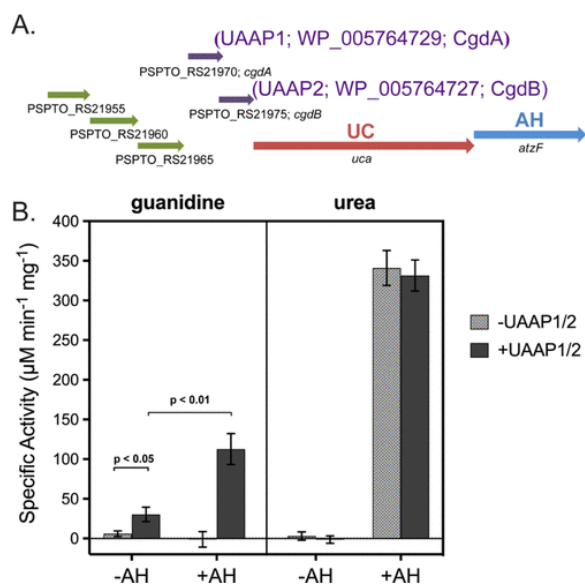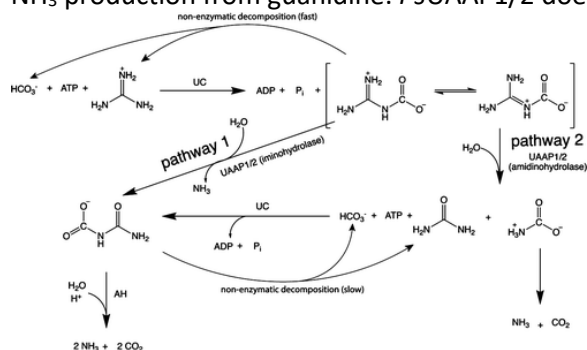

Figure 4. The combination of *Ps*UC and *Ps*UAAP1/2 (CgdAB) is required for the decomposition of guanidine to ammonia. (A) Localized map of genes encoding UC, AH, and UAAP1 and UAAP2 on the genome of *P. syringae* pv tomato strain DC3000. The genes encoding UAAP1/2 (CgdAB) are annotated as *PSPTO_RS21970* and *PSPTO_RS21975* (purple), and the genes encoding UC and AH are annotated as *uca* (red) and *atzF* (blue), respectively. The gene annotations are indicated below the arrows that represent the relative location, size and direction of the genes. The genes annotated as *PSPTO_RS21955*, *PSPTO_RS21960*, and *PSPTO_RS21965* (green) encode a putative ABC-type transporter and are in suitably close proximity to comprise an operon with the other genes. Reproduced from ref (4). (B) The production of NH_3 from the decomposition of urea or guanidine by *Ps*UC was measured in the presence and absence of equimolar *Ps*UAAP1/2, and/or equimolar *Ps*AH. For each set of experiments, the presence or absence of *Ps*UAAP1/2 (CgdAB) is indicated by the lighter shade to the left (−UAAP1/2) and the darker shade to the right

(+UAAP1/2). The panel on the left presents the data for guanidine, while the panel on the right presents the data for urea. The absence or presence of PsAH in the reaction is indicated by −AH or +AH in the x-axis. In the absence of PsAH, there is a significant increase ($p < 0.05$) in NH_3 production from guanidine in the presence of PsUAAP1/2. The addition of PsAH to this system results in an additional significant increase ($p < 0.01$) in NH_3 production from guanidine. PsUAAP1/2 does not alter the rate of NH_3 production from urea.

Scheme 2. Two Possible Pathways for the Complete Enzymatic Decomposition of Guanidine

To distinguish between an iminohydrolase mechanism (Scheme 2; pathway 1) and an amidinohydrolase mechanism (Scheme 2; pathway 2), we performed assays to evaluate product formation by ^{13}C NMR. Incubating ^{13}C-urea with ATP and HCO_3^- in the presence of PsUC yielded a product signal corresponding to ^{13}C-allophanate, as expected (Figure S4). We next incubated ^{13}C-guanidine with ATP and HCO_3^- in the presence of PsUC and PsUAAP1/2. Initially, a signal corresponding to ^{13}C-allophanate appeared, and this was followed, after a clear lag period, by an additional signal corresponding to ^{13}C-urea (Figure 5A,B). The appearance of the ^{13}C-allophanate peak prior to the appearance of the ^{13}C-urea peak strongly supports an iminohydrolase mechanism (Scheme 2; pathway 1). The lagging appearance of urea could arise either by the enzyme partitioning between pathways 1 and 2 or from the gradual decomposition of ^{13}C-allophanate to urea and HCO_3^-. To distinguish between these possibilities, we incubated ^{13}C-guanidine, ATP and HCO_3^- in the presence of PsUC, PsUAAP1/2 and PsAH. We expected that, if the enzyme partitions between pathways 1 and 2, a signal corresponding to ^{13}C-urea would continue to be observed, even in the presence of PsAH. However, only a signal corresponding to $H^{13}CO_3^-$ appeared under these conditions (Figure 5C,D). We conclude that the signal corresponding to urea in Figure 5A resulted from the decomposition of allophanate to urea and HCO_3^- at neutral pH.(21) As further confirmation, we next incubated ^{13}C-guanidine, ATP, and HCO_3^- in the presence of PsUC, PsUAAP1/2, and urease, which will immediately decompose any urea produced by the combined actions of PsUC and PsUAAP1/2. The signal intensity corresponding to ^{13}C-allophanate increased at the same rate as it did in the absence of urease (Figure 5EF), in further support of the iminohydrolase mechanism of Scheme 2; pathway 1. We thus functionally assign UAAP1 and UAAP2 as the α and β subunits, respectively, of a newly characterized enzyme, which we have named carboxyguanidine deiminase (Cgd). The proteins, formerly assigned the designations urea-carboxylase associated protein 1 and 2 are henceforth renamed CgdA and CgdB, and the heteromeric carboxyguanidine deiminase enzyme is designated CgdAB.
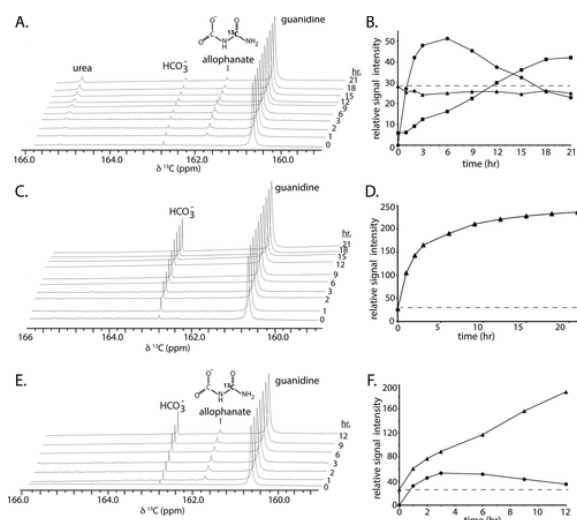
Figure 5. ^{13}C NMR spectroscopic analysis of guanidine decomposition by *Ps*UC and CgdAB. (A) 1D ^{13}C NMR data of the sample containing 6 mM ATP, 20 mM NaHCO₃, 250 µg/mL CgdAB, and 10 mM ^{13}C guanidine (0 h time point). The reaction was initiated by *Ps*UC to a final concentration of 100 µg/mL. Time points at 1, 2, 3, 6, 9, 12, 15, 18, 21 h show the production of allophanate over time and subsequent lag in the formation of urea. (B) Plot of time vs relative signal intensity for urea (squares), bicarbonate (triangles), and allophanate (black circles) chemical shifts. (C) 1D ^{13}C NMR data of the sample containing 6 mM ATP, 20 mM NaHCO₃, 250 µg/mL CgdAB, and 10 mM ^{13}C guanidine in the presence of 725 µg/mL *Ps*AH (0 h time point). The reaction was initiated by *Ps*UC to a final concentration of 100 µg/mL. Time points at 1, 2, 3, 6, 9, 12, 15, 18, 21 h show only the formation of bicarbonate over time. (D) Plot of time vs relative signal intensity for the bicarbonate (triangle) chemical shift. All spectra were calibrated and scaled to the 1 h reaction time point. (E) 1D ^{13}C NMR data of the sample containing 6 mM ATP, 20 mM NaHCO₃, 250 µg/mL CgdAB, and 10 mM ^{13}C guanidine in the presence of 100 µg/mL urease (0 h time point). The reaction was initiated by *Ps*UC to a final concentration of 100 µg/mL. Time points at 1, 2, 3, 6, 9, 12 h show the formation of allophanate over time and subsequent formation of bicarbonate. (F) Plot of time vs relative signal intensity for bicarbonate (triangles) and allophanate (closed circles) chemical shifts. The dashed line in panels (B), (D) and (F) represents the relative signal intensity for ^{13}C-bicarbonate at $t = 0$.

Carboxyguanidine Deiminase Is a Member of a Large, Uncharacterized Metalloenzyme Superfamily

The α and β subunits of Cgd from *P. syringae* (*Ps*CgdA and *Ps*CgdB) show 30% amino acid sequence identity to each other and each is homologous to members of the Pfam Domains of Unknown Function (DUF) protein family known as DUF1989 (Figure 6). At the time of this writing, there were 14 726 members of DUF1989, of which 54.5% are indicated from this study to be CgdAB enzymes. In genome annotations, they have been denoted as urea-carboxylase-associated, or UAAP1 and UAAP2, reflecting their genomic association and the absence of a demonstrated function. In all instances that we have examined, the two genes encoding these proteins are adjacent, consistent with our finding that they interact to form a monofunctional heteromeric enzyme.

The *cgd* genes are found in the Kingdom Bacteria, most commonly in the Phylum Proteobacteria, but are also found in the Phyla Nitrospirae, Actinobacteria, Planctobacteria, Firmicutes, and Verrucomicrobia. The number of genes distributed across diverse bacteria and the observation that sequence divergence matches taxonomic divergence both suggest that CgdAB is an ancient member of the DUF1989 protein family.
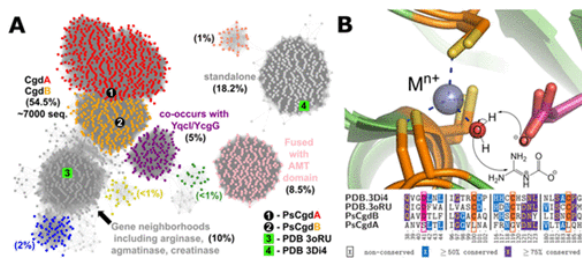
Figure 6. Carboxyguanidine deiminase is located within the DUF1989 family. (A) Sequence Similarity Network of the DUF1989 family with over 12 800 sequences were clustered using the EFI-EST using a cutoff *E*-value of 43. CgdA/B sequences comprise more than half of the recorded sequences of DUF1989 members. Labels with green squares indicate sequences with known PDB structures and the CgdA/B sequences characterized in this study are labeled with black circles. Percentages are a percent of total sequences for each labeled cluster. (B) Conserved aspartate and metal binding residues in the DUF1989 family. PDB structures 3DI4 and 3ORU of uncharacterized DUF1989 members overlaid indicate three conserved cysteine ligands binding a metal and a conserved, proximal aspartate residue that has a putative role in activating water that could hydrolyze carboxyguanidine. Multiple sequence alignment of these structures with *Ps*CgdA and *Ps*CgdB indicate these residues are conserved in *Ps*CgdB but not *Ps*CgdA. Sequence numbering is based on *Ps*CgdB.

The DUF1989 family, while not previously characterized functionally, has two members for which X-ray structures have been determined (PDB 3ORU, 3DI4) (Figure 6). The Sequence Similarity Network (SSN) indicates that these two structurally defined proteins are not Cgd proteins, which is also indicated by completely different gene neighborhoods than *cgdAB*. The deposited structures both reveal a metal coordinated by cysteine ligands and a proximal aspartate (Figure 6B). These residues are conserved in CgdB, but not in CgdA and suggest a role in activating water for substrate hydrolysis. Evaluating the gene neighborhoods of other clusters of DUF1989 sequences suggests that another 10% of DUF1989 sequences could be involved in metabolism of guanidinium containing compounds (e.g., arginine, creatine, agmatine). Other notable clusters of DUF1989 sequences have an aminomethyltransferase (AMT) fusion with a DUF1989 domain (8.5% of sequences, e.g., NCBI WP_067612148.1) and another cluster (5%) has a co-occurring YqcI/YcgG protein of unknown function, which has a suggested role in processing nonproteinogenic arginine-based natural products by nonribosomal peptide synthetases (e.g., NCBI WP_056147698.1).(31)

Evaluating the gene contexts of *cgd* sequences within an eight gene window indicates that >96% colocalize with the UC gene and more than two-thirds colocalize with an AH gene (Figure S5). Thus, this gene cluster encodes all of the necessary genes for a bacterium to decompose guanidine and urea to ammonia (Scheme 2). Two-thirds of *cgdAB* genes have a guanidine-I-riboswitch upstream while another 22% and 15% have regulatory elements corresponding to uncharacterized TetR and NikR family regulatory proteins, respectively, that may also be guanidine sensitive (ex. NCBI WP_114699518.1, WP_012288442.1). Transport genes are also co-occurring; 71% of *cgd* sequences co-occur with ABC transport cassette genes and 19% with a passive transport permease gene.

All genes encoding CgdAB proteins identified here have neighboring genes identified as UC which are likely, instead, guanidine carboxylase genes. However, not all annotated bacterial UCs are genetically linked to CgdAB proteins. In bacteria, 76% of the carboxylases are CgdAB-linked and hence are likely guanidine carboxylases, while the remaining 24% are independent of Cgd proteins. Given that finding, we examined guanidine carboxylase/urea carboxylase sequence signatures for markers that might differentiate between carboxylases evolved for preferential reactivity in carboxylating guanidine versus urea. The crystal structure of the UC domain from the fungal *K. lactis* UAL (PDB 3VA7) was used to optimally align sequences. All active site residues are conserved across fungi and bacteria with the exception of one residue: a residue corresponding to N1330 in *K. lactis* UAL is conserved in fungal UALs and in ~25% of bacterial UCs (Figure S6A). The remaining ~75% of

bacterial UCs show an aspartate residue in the same position. The dichotomy at this position correlates strongly with the presence or absence of *cgdAB* genes. Ninety-seven percent of recorded UC genes encoding an Asp at this position had *cgdAB* in the gene neighborhood, while less than one percent of UC genes encoding an Asn had nearby *cgdAB* genes (Figure S6B). This correlation suggests that the aspartate or asparagine tracks with preferential activity versus guanidine or urea, respectively.

Fungal UALs Do Not Act on Guanidine

Analysis of the gene context of UAL in fungi reveals that there are no genes encoding CgdAB within at least an eight gene window of the UAL gene. Since most known riboswitches occur only in bacteria, we anticipate that the regulation of UAL in fungi differs from many of its bacterial homologues that are controlled by the guanidine-I riboswitches and which are found in the context of an operon with *cgdA* and *cgdB*. We characterized the UAL enzymes from both *S. cerevisiae* and *C. albicans*. Notably, both of these enzymes maintain an Asn at the equivalent to position N1330 of *K. lactis* UAL. In marked contrast to the bacterial homologues, both fungal UAL enzymes (*Sc*UAL and *Ca*UAL) showed a very strong substrate preference for urea over guanidine, with a catalytic efficiency 4 orders of magnitude higher for urea compared to guanidine (Figure 7). Thus, the ability to decompose guanidine through the combined activities of UC, CgdAB and AH may be limited to bacteria.
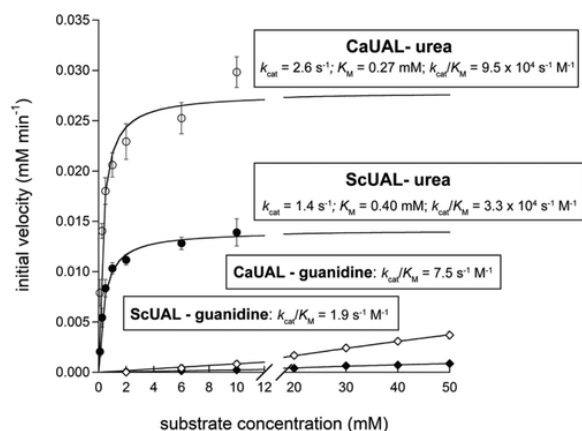

Figure 7. Kinetics of ATP cleavage by fungal UAL reveal a strong substrate preference for urea over guanidine. The initial velocity for ATP cleavage was measured as a function of substrate concentration (guanidine, diamonds; urea, circles) for *Sc*UAL (open symbols) and *Ca*UAL (closed symbols). The kinetic constants reveal that fungal *Sc*UAL and *Ca*UAL show high substrate selectivity for urea over guanidine. For ATP cleavage in the presence of urea, the kinetic constants were determined by nonlinear regression fits of the rates to the standard Michaelis–Menten equation. The k_{cat}/K_M for ATP cleavage in the presence of guanidine was estimated from the slope of the line, assuming that [guanidine] is $< K_M$. For each trial, three independent initial velocity measurements were recorded for each substrate concentration.

Discussion

In a commentary that followed the initial discovery and characterization of UC in bacteria,(7) Hausinger presciently noted the possibility that UC may have a substrate preference other than urea.(32) The recent discovery that UC in many (but not all) bacteria is under control of a guanidine-I riboswitch and the characterization of an enhanced guanidine-dependent ATP cleavage activity in UC from *O. sagaranensis* further supports the notion that UC has a substrate preference for guanidine.(13) Here, we offer strong evidence to support that the product of the reaction catalyzed by UC in the presence of guanidine, ATP, and HCO_3^- is, indeed, carboxyguanidine, and we establish the downstream pathway that contributes to the complete decomposition of guanidine to ammonia and CO_2 in bacteria. The identification and characterization of this pathway further clarifies the redundancy between urease and UC and suggests a straightforward way to predict

whether an enzyme is a urea carboxylase or a guanidine carboxylase, based on the presence or absence of genes corresponding to CgdAB and a corresponding sequence signature.

Guanidine has not been extensively studied as either a nitrogen source or a metabolite in bacteria, but it is a natural product and likely to be generated during the catabolism of natural and synthetic guanidinium compounds. In fact, *P. syringae* produces guanidine from arginine and α-ketoglutarate and this has recently been harnessed in an effort to produce guanidine by biotechnology.(33−35) A large number of guanidine-containing natural products are biosynthesized by prokaryotes and eukaryotes.(36) In addition, cyanoguanidine is applied on a ton scale as a fertilizer additive(37) and metformin, a biguanidine, is one of the most prescribed drugs globally and is found in some aquatic environments at microgram per liter levels.(38) As a consequence, it is expected that a broad subset of soil microorganisms can access and assimilate nitrogen from guanidine. While a few studies have described the microbial degradation of guanidinium ion both in surface waters and soil, these did not elucidate routes for its biodegradation.(39−42) The pathway described here, under control of a riboswitch or other regulatory elements, is expected to be a major contributor to guanidine mineralization in bacteria. Furthermore, the broad distribution of CgdAB among bacteria, spanning ∼6000 current examples, suggests that metabolic or environmentally derived guanidine has long been a relevant factor, greatly predating its occurrence from anthropogenic chemicals that have entered global soils and water.

Despite guanidine first being described in 1861, carboxyguanidine has never been described in the literature, consistent with an inability to synthesize and maintain it. It is, therefore, unsurprising that we were unable to detect an NMR signature for this product. Without the subsequent enzymatic deimination of carboxyguanidine to the more stable allophanate, this product is expected to quickly revert back to the initial substrates of UC: guanidine and HCO_3^-. From a biological perspective, this pathway risks a high rate of abortive ATP cleavage unless carboxyguanidine is acted upon quickly, before it decomposes back to the starting substrates. This suggests UC and CgdAB are candidates for substrate channeling, but our prior studies found no evidence for a permanent physical interaction between these proteins.(4) In the absence of a permanent interaction, the success of this metabolic pathway requires either a transient channeling interaction between UC and CgdAB and/or a highly efficient CgdAB enzyme to protect the carboxyguanidine intermediate from degradation.

We have demonstrated that CgdAB is an iminohydrolase (Figure 5; Scheme 2, pathway 1), rather than an amidinohydrolase. Arginine deiminase, along with other arginine-degrading deiminases, uses a Cys catalytic triad (Cys-His-Glu/Asp) to catalyze a double displacement mechanism through an *S*-alkylthiouronium covalent intermediate.(43) Arginase, meanwhile, acts as an amidinohydrolase that employs a dinuclear metal center to activate water for nucleophilic attack.(44) While the mechanistic details for CgdAB remain unexplored, there is strong evidence suggesting that CgdAB is a metalloenzyme distinct from both arginase and arginine deiminase. The two deposited structures of Cgd homologues of unknown function in the PDB (3DI4 and 3ORU) show a trio of conserved Cys-residues chelating a central Zn^{2+}, and these Cys residues are conserved in CgdB (Figure 6). While further investigation is warranted, metal ion assisted nucleophilic attack of water by CgdAB may be similar to that of cytidine deaminase and adenosine deaminase, which both use a Zn^{2+}-bound water for nucleophilic attack and a conserved aspartate or glutamate as a general acid/base catalyst.(45) Despite the potential similarities in mechanism, there is no apparent homology between Cgd and either adenosine deaminase or cytidine deaminase. Cytidine deaminase, arginase, and related enzymes that catalyze hydrolytic deamination via activation of water have catalytic efficiencies (and catalytic proficiencies) several orders of magnitude higher than the deiminases that catalyze hydrolysis through a double displacement mechanism.(30,46) The requirement for a high catalytic efficiency when acting on the unstable carboxyguanidine substrate may offer an additional rationale for exploiting metal-ion assisted catalysis in this deimination reaction.

Recent studies have confirmed the presence of guanidine-based riboswitches and transporters in bacteria.(13,47−50) Following on these important initial discoveries, our work serves to define the components

of the metabolic pathway responsible for guanidine decomposition in bacteria. The combined activities of UC, CgdAB and AH in the decomposition of guanidine correlate with their tight gene clustering in the majority of bacteria for which the UC gene is present. Our analysis reveals that ~75% of all genes for UC cluster with *cgdAB*. Consequently, these UC enzymes are likely to have a strong substrate preference for guanidine, as was observed here for *P. syringae* UC and previously for *O. sagaranensis* UC.(13) This subset of enzymes should appropriately be renamed as guanidine carboxylases, in accordance with prior recommendations.(13) However, we caution against universally reassigning all currently annotated UC enzymes to guanidine carboxylase. Approximately 25% of genes encoding UC, including all of the fungal UAL enzymes, do not cluster with *cgdAB* and display an almost perfect correlation with a single Asp to Asn amino acid change in the carboxyltransferase active site, exemplified by Asn 1330 of *K. lactis* UAL (Figure S6). Consistent with this analysis, the fungal UAL enzymes evaluated in the current study have a very strong substrate preference for urea over guanidine and, therefore, are appropriately annotated as UALs (Figure 7). The notion that UC has diverged among bacteria and fungi has previously been recognized.(3) While the substrate preference for guanidine and the pathway for guanidine decomposition explains some of the apparent functional redundancy between urease and UC, there remains a significant number of seemingly bona fide UCs for which the functional redundancy with urease remains unresolved. The work presented here reinforces the functional relevance of both guanidine carboxylase and urea carboxylase and highlights a newly recognized pathway for the decomposition of guanidine for nitrogen assimilation in bacteria. Furthermore, the broad distribution of carboxyguanidine deiminase among bacteria, and the absence of the corresponding genes in fungi, adds to a growing body of evidence implicating guanidine as an important but selective environmental metabolite and nitrogen source.

References

Wyatt, B. N., Arnold, L. A., and St. Maurice, M. (**2018**) A high-throughput screening assay for pyruvate carboxylase. *Anal. Biochem. 550*, 90– 98, DOI: 10.1016/j.ab.2018.04.012

Navarathna, D. H., Harris, S. D., Roberts, D. D., and Nickerson, K. W. (**2010**) Evolutionary aspects of urea utilization by fungi. *FEMS Yeast Res. 10*, 209– 213, DOI: 10.1111/j.1567-1364.2009.00602.x

Strope, P. K., Nickerson, K. W., Harris, S. D., and Moriyama, E. N. (**2011**) Molecular evolution of urea amidolyase and urea carboxylase in fungi. *BMC Evol. Biol. 11*, 80, DOI: 10.1186/1471-2148-11-80

Lin, Y., Boese, C. J., and St. Maurice, M. (**2016**) The urea carboxylase and allophanate hydrolase activities of urea amidolyase are functionally independent. *Protein Sci. 25*, 1812– 1824, DOI: 10.1002/pro.2990

Whitney, P. A. and Cooper, T. G. (**1972**) Urea carboxylase and allophanate hydrolase: two components of a multienzyme complex in Saccharomyces cerevisiae. *Biochem. Biophys. Res. Commun. 49*, 45– 51, DOI: 10.1016/0006-291X(72)90007-1

Cooper, T. G., Lam, C., and Turoscy, V. (**1980**) Structural analysis of the dur loci in *S. cerevisiae*: two domains of a single multifunctional gene. *Genetics 94*, 555– 580

Kanamori, T., Kanou, N., Atomi, H., and Imanaka, T. (**2004**) Enzymatic characterization of a prokaryotic urea carboxylase. *J. Bacteriol. 186*, 2532– 2539, DOI: 10.1128/JB.186.9.2532-2539.2004

Kanamori, T., Kanou, N., Kusakabe, S., Atomi, H., and Imanaka, T. (**2005**) Allophanate hydrolase of *Oleomonas sagaranensis* involved in an ATP-dependent degradation pathway specific to urea. *FEMS Microbiol. Lett. 245*, 61– 65, DOI: 10.1016/j.femsle.2005.02.023

Selengut, J. D., Rusch, D. B., and Haft, D. H. (**2010**) Sites Inferred by Metabolic Background Assertion Labeling (SIMBAL): adapting the Partial Phylogenetic Profiling algorithm to scan sequences for signatures that predict protein function. *BMC Bioinf. 11*, 52, DOI: 10.1186/1471-2105-11-52

Minami, T., Anda, M., Mitsui, H., Sugawara, M., Kaneko, T., Sato, S., Ikeda, S., Okubo, T., Tsurumaru, H., and Minamisawa, K. (**2016**) Metagenomic Analysis Revealed Methylamine and Ureide Utilization of Soybean-Associated Methylobacterium. *Microbes Environ. 31*, 268– 278, DOI: 10.1264/jsme2.ME16035

Oshiki, M., Araki, M., Hirakata, Y., Hatamoto, M., Yamaguchi, T., and Araki, N. (**2018**) Ureolytic Prokaryotes in Soil: Community Abundance and Diversity. *Microbes Environ. 33*, 230– 233, DOI: 10.1264/jsme2.ME17188

Zhang, Y., Rodionov, D. A., Gelfand, M. S., and Gladyshev, V. N. (**2009**) Comparative genomic analyses of nickel, colbalt, and vitamin B12 utilization. *BMC Genomics 10*, 78, DOI: 10.1186/1471-2164-10-78

Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., and Breaker, R. R. (**2017**) Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Mol. Cell 65*, 220– 230, DOI: 10.1016/j.molcel.2016.11.019

Chapman-Smith, A., Turner, D. L., Cronan, J. E., Jr., Morris, T. W., and Wallace, J. C. (**1994**) Expression, biotinylation and purification of a biotin-domain peptide from the biotin carboxy carrier protein of *Escherichia coli* acetyl-CoA carboxylase. *Biochem. J. 302* (Pt 3), 881– 887, DOI: 10.1042/bj3020881

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., and Whalen, K. L. (**2015**) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta, Proteins Proteomics 1854*, 1019– 1037, DOI: 10.1016/j.bbapap.2015.04.015

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (**2003**) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res. 13*, 2498– 2504, DOI: 10.1101/gr.1239303

Eddy, S. R. (**2011**) Accelerated Profile HMM Searches. *PLoS Comput. Biol. 7*, e1002195 DOI: 10.1371/journal.pcbi.1002195

Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., and White, O. (**2001**) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res. 29*, 41– 43, DOI: 10.1093/nar/29.1.41

Tietz, J. I., Schwalen, C. J., Patel, P. S., Maxson, T., Blair, P. M., Tai, H. C., Zakai, U. I., and Mitchell, D. A. (**2017**) A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol. 13*, 470– 478, DOI: 10.1038/nchembio.2319

Lin, Y. and St. Maurice, M. (**2013**) The structure of allophanate hydrolase from *Granulibacter bethesdensis* provides insights into substrate specificity in the amidase signature family. *Biochemistry 52*, 690– 700, DOI: 10.1021/bi301242m

Cheng, G., Shapir, N., Sadowsky, M. J., and Wackett, L. P. (**2005**) Allophanate hydrolase, not urease, functions in bacterial cyanuric acid metabolism. *Appl. Environ. Microbiol. 71*, 4437– 4445, DOI: 10.1128/AEM.71.8.4437-4445.2005

Legge, G. B., Branson, J. P., and Attwood, P. V. (**1996**) Effects of acetyl CoA on the pre-steady-state kinetics of the biotin carboxylation reaction of pyruvate carboxylase. *Biochemistry 35*, 3849– 3856, DOI: 10.1021/bi952797q

Branson, J. P. and Attwood, P. V. (**2000**) Effects of Mg(2+) on the pre-steady-state kinetics of the biotin carboxylation reaction of pyruvate carboxylase. *Biochemistry 39*, 7480– 7491, DOI: 10.1021/bi992825v

Castric, P. A. and Levenberg, B. (**1976**) Urea amidolyase of *Candida utilis*. Characterization of the urea cleavage reactions. *Biochim. Biophys. Acta 438*, 574– 583, DOI: 10.1016/0005-2744(76)90273-4

Attwood, P. V. and Graneri, B. D. (**1992**) Bicarbonate-dependent ATP cleavage catalysed by pyruvate carboxylase in the absence of pyruvate. *Biochem. J. 287* (Pt 3), 1011– 1017, DOI: 10.1042/bj2871011

Trumble, G. E., Smith, M. A., and Winder, W. W. (**1991**) Evidence of a biotin dependent acetyl-coenzyme A carboxylase in rat muscle. *Life Sci. 49*, 39– 43, DOI: 10.1016/0024-3205(91)90577-X

Fan, C., Chou, C. Y., Tong, L., and Xiang, S. (**2012**) Crystal structure of urea carboxylase provides insights into the carboxyltransfer reaction. *J. Biol. Chem. 287*, 9389– 9398, DOI: 10.1074/jbc.M111.319475

Zhao, J., Zhu, L., Fan, C., Wu, Y., and Xiang, S. (**2018**) Structure and function of urea amidolyase. *Biosci. Rep. 38*, BSR20171617, DOI: 10.1042/BSR20171617

Strecker, A. (**1861**) Untersuchungen über die chemischen Beziehungen zwischen Guanin, Xanthin, Theobromin, Caffeïn und Kreatinin [Studies on the chemical relationships between guanine, xanthine, theobromine, caffeine and creatinine]. *Liebigs Ann. Chem. 118*, 151– 177, DOI: 10.1002/jlac.18611180203

Lewis, C. A., Jr. and Wolfenden, R. (**2014**) The nonenzymatic decomposition of guanidines and amidines. *J. Am. Chem. Soc. 136*, 130– 136, DOI: 10.1021/ja411927k

Romo, A. J., Shiraishi, T., Ikeuchi, H., Lin, G. M., Geng, Y., Lee, Y. H., Liem, P. H., Ma, T., Ogasawara, Y., Shin-Ya, K., Nishiyama, M., Kuzuyama, T., and Liu, H. W. (**2019**) The Amipurimycin and Miharamycin Biosynthetic Gene Clusters: Unraveling the Origins of 2-Aminopurinyl Peptidyl Nucleoside Antibiotics. *J. Am. Chem. Soc. 141*, 14152– 14159, DOI: 10.1021/jacs.9b03021

Hausinger, R. P. (**2004**) Metabolic Versatility of Prokaryotes for Urea Decomposition. *J. Bacteriol. 186*, 2520– 2522, DOI: 10.1128/JB.186.9.2520-2522.2004

Fukuda, H., Ogawa, T., Tazaki, M., Nagahama, K., Fujii, T., Tanase, S., and Morino, Y. (**1992**) Two reactions are simultaneously catalyzed by a single enzyme: the arginine-dependent simultaneous formation of two products, ethylene and succinate, from 2-oxoglutarate by an enzyme from *Pseudomonas syringae*. *Biochem. Biophys. Res. Commun. 188*, 483– 489, DOI: 10.1016/0006-291X(92)91081-Z

Wang, B., Dong, T., Myrlie, A., Gu, L. P., Zhu, H. L., Xiong, W., Maness, P., Zhou, R. B., and Yu, J. P. (**2019**) Photosynthetic production of the nitrogen-rich compound guanidine. *Green Chem. 21*, 2928– 2937, DOI: 10.1039/C9GC01003C

Martinez, S., Fellner, M., Herr, C. Q., Ritchie, A., Hu, J., and Hausinger, R. P. (**2017**) Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist. *J. Am. Chem. Soc. 139*, 11980– 11988, DOI: 10.1021/jacs.7b06186

Berlinck, R. G. S., Bertonha, A. F., Takaki, M., and Rodriguez, J. P. G. (**2017**) The chemistry and biology of guanidine natural products. *Nat. Prod. Rep. 34*, 1264– 1301, DOI: 10.1039/C7NP00037E

Yang, M., Fang, Y., Sun, D., and Shi, Y. (**2016**) Efficiency of two nitrification inhibitors (Dicyandiamide and 3, 4-dimethypyrazole phosphate) on soil nitrogen transformations and plant productivity: a meta-analysis. *Sci. Rep. 6*, 22075, DOI: 10.1038/srep22075

Tao, Y., Chen, B., Zhang, B. H., Zhu, Z. J., and Cai, Q. (**2018**) Occurrence, Impact, Analysis and Treatment of Metformin and Guanylurea in Coastal Aquatic Environments of Canada, USA and Europe. *Adv. Mar. Biol. 81*, 23– 58, DOI: 10.1016/bs.amb.2018.09.005

Mitchell, W. R. (**1987**) Microbial-Degradation of Guanidinium Ion. *Chemosphere 16*, 1071– 1086, DOI: 10.1016/0045-6535(87)90044-0

Mitchell, W. R. (**1987**) Biodegradation of Guanidinium Ion in Aerobic Soil Samples. *Bull. Environ. Contam. Toxicol. 39*, 974– 981, DOI: 10.1007/BF01689587

Kumar, P. and Brumme, R. (**1999**) Mineralization of guanidine derivatives in soils. *Nutr. Cycling Agroecosyst. 53*, 133– 138, DOI: 10.1023/A:1009735003250

Ebisuno, T. and Takimoto, M. (**1981**) Examination of biodegradability of guanyl compounds. *Eisei Kagaku 27*, 156– 162, DOI: 10.1248/jhs1956.27.156

Galkin, A., Lu, X., Dunaway-Mariano, D., and Herzberg, O. (**2005**) Crystal structures representing the Michaelis complex and the thiouronium reaction intermediate of *Pseudomonas aeruginosa* arginine deiminase. *J. Biol. Chem. 280*, 34080– 34087, DOI: 10.1074/jbc.M505471200

Di Costanzo, L., Sabio, G., Mora, A., Rodriguez, P. C., Ochoa, A. C., Centeno, F., and Christianson, D. W. (**2005**) Crystal structure of human arginase I at 1.29-A resolution and exploration of inhibition in the immune response. *Proc. Natl. Acad. Sci. U. S. A. 102*, 13058– 13063, DOI: 10.1073/pnas.0504027102

Betts, L., Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (**1994**) Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol. 235*, 635– 656, DOI: 10.1006/jmbi.1994.1018

Snider, M. J., Gaunitz, S., Ridgway, C., Short, S. A., and Wolfenden, R. (**2000**) Temperature effects on the catalytic efficiency, rate enhancement, and transition state affinity of cytidine deaminase, and the thermodynamic consequences for catalysis of removing a substrate "anchor". *Biochemistry 39*, 9746– 9753, DOI: 10.1021/bi000914y

Sherlock, M. E. and Breaker, R. R. (**2017**) Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. *Biochemistry 56*, 359– 363, DOI: 10.1021/acs.biochem.6b01271

Sherlock, M. E., Malkowski, S. N., and Breaker, R. R. (**2017**) Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry 56*, 352– 358, DOI: 10.1021/acs.biochem.6b01270

Higgins, D. A., Gladden, J. M., Kimbrel, J. A., Simmons, B. A., Singer, S. W., and Thelen, M. P. (**2019**) Guanidine Riboswitch-Regulated Efflux Transporters Protect Bacteria against Ionic Liquid Toxicity. *J. Bacteriol. 201*, e00069-19 DOI: 10.1128/JB.00069-19

Kermani, A. A., Macdonald, C. B., Gundepudi, R., and Stockbridge, R. B. (**2018**) Guanidinium export is the primal function of SMR family transporters. *Proc. Natl. Acad. Sci. U. S. A. 115*, 3060– 3065, DOI: 10.1073/pnas.1719187115