Elucidating the biodegradation pathway for the pharmaceutical metformin in wastewater: Implications in human and wastewater microbiomes

---

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Lambros J. Tassoulas

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOCHEMISTRY, BIOPHYSICS AND MOLECULAR BIOLOGY

Professor Lawrence P. Wackett, Advisor

December, 2023

Acknowledgements

Firstly, I would like to express my gratitude to my advisor, Prof. Larry Wackett, from whom I have gained knowledge and experience in multiple areas of science. Dr. Wackett's vast knowledge, guidance has helped shape both this dissertation and myself as an aspiring researcher. I can recall many instances where Dr. Wackett's wisdom would make the difference between success and failure in achieving certain goals.

I would like to thank, very much, Dr. Mikael Elias for his advice and company throughout the years. I really appreciate that I went through this endeavor and I need to say that without certain encouragement from Dr. Elias, several X-ray crystallography experiments in this dissertation would not have been accomplished.

I want to acknowledge my thesis committee members Dr. Elias, Dr. Michael Freeman and Dr. Carl Rosen for overseeing my progress as a graduate student and always demonstrating great enthusiasm and curiosity.

Special thanks go to Betsy Martinez-Vaz, Jack Richman, Tony Dodge, Serina Robinson, Kelly Aukema, Sudipta Shaw, Dimitri Pérusse and to the rest of my colleagues in the Gortner laboratory who are a joy to work with and for the fruitful discussions we have had, as well as their company.

I would like to dedicate this thesis to my father, Yannis Tassoulas, for being a big influence in my life which I am grateful for.

Lambros Tassoulas, Minneapolis, December 11, 2023

The work in this dissertation was supported by grants from the National Institutes of Health (NIH), National Science Foundation (NSF) and from the University of Minnesota UMII-MnDRIVE and Health Foods Healthy Lives Institute (HFHL).

Abstract

Metformin (1,1-dimethylbiguanide) is a wonder drug and a pervasive pollutant which is taken by type-II diabetes patients and more recently for improving obesity and cancer treatment outcomes. The drug's direct mode of action is currently unknown but is thought to be dependent on microbial reactions to the drug. Large therapeutic doses (1-2 grams daily) and the global use of metformin result in over 100 million kilograms of the drug entering aquatic ecosystems each year. The biodegradation of metformin was known in wastewater treatment plants to be transformed by microbes to guanylurea but, prior to this work, no enzymes involved in the biodegradation had been identified. As a result of this work, the metformin biodegradation pathway has been completely elucidated starting with metformin hydrolysis to form dimethylamine and guanylurea by metformin hydrolase (MfmAB). The guanylurea hydrolase (GuuH) was discovered, which degrades guanylurea into guanidine, ammonia and carbon dioxide. Lastly, the enzyme CgdAB was discovered that acts as a carboxyguanidine deiminase bacteria use to assimilate nitrogen from guanidine. A secondary pathway of metformin biodegradation was also identified by discovering a biguanidase (BguH) that degrades biguanide and 1-methylbiguanide to form guanylurea, suggesting that metformin can also be demethylated by microbes. These enzymes were characterized by kinetics, X-ray crystallography or computational modelling, and bioinformatics. While the metabolism of metformin is now known in wastewater ecosystems, metabolism in the human gut has not been established but may contribute to the potency or therapeutic effect of the drug. Testing enzymes that are found in human gut microbes, homologous to the metformin hydrolase, did not show activity on metformin. However, a subset of these gut enzymes, from Gammaproteobacteria that hydrolyze agmatine, was potently inhibited by metformin and the natural product galegine, from which the drug was derived from. Agmatine is a known effector of human host metabolism and has been reported to augment metformin's therapeutic effects for type-II diabetes. This gut-derived inhibition mechanism gives new insights on metformin's action in the gut and may lead to significant discoveries in improving metformin therapy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Metformin is the first-line therapeutic drug for type-II diabetes and a pervasive pollutant of global waters. This exceptionally small molecule has treated hundreds of millions people since its discovery in 1957 and, yet interestingly, its direct mechanism of action is still unknown[1]. In addition, cohort studies have identified that metformin-treated diabetes patients show healthier outcomes than non-metformin treated diabetes patients in obesity, aging and cancer incidence and the future use of this drug could be extended beyond type-II diabetes[2]. The large daily dose of 1-2 grams, when ingested by patients, is not metabolized in the liver, and is excreted in the urine, entering wastewater. This amounts to more than 100 million kilos of metformin that is polluting the environment every year and is pervasive as caffeine, a lifestyle molecule, in global waters.

The mechanism of action for metformin has been under intense scrutiny for the past couple of decades, yet a direct mode of action is still unclear. Recent studies point to metformin's interaction with human gut microbes for exerting therapeutic effects[3–5]. A study involving intravenous administration of metformin to type-II diabetes patients, that mainly bypasses the human gut, showed little to no efficacy[6].

The large metagenome sequencing effort of the gut microbiome in recent decades has identified consistent trends in taxonomic and gene distributions that correlate with metformin treatment which include increases in short-chain fatty acid (SCFA) producing and mucin-degrading microbes [3,7,8]. These well-established trends still have not been connected to the therapeutic mechanism or side effects of the drug and remains unclear if they are linked. Growth and biophysical studies of microbes with metformin have observed negative effects in growth, chemotaxis and motility for *E. coli* but at suprapharmacological levels of metformin (>20 mM) [9]. The estimated concentration of metformin in the gut is between 1-10 mM and in the portal vein that enters the liver the concentration of metformin has been measured to be much less between 10-40 μM [5,10,11].

An important note of metformin therapy is that the drug is part of a class of molecules called biguanides which include phenformin, buformin and the parent compound galegine that is a natural product from the plant *Galega officinalis*[1]. Phenformin and buformin are more potent drugs than metformin but were eventually discontinued as metformin shows a superior safety profile. Fatal side effects of these drugs are rare but is mainly caused by lactic acidosis[12]. A possible mechanism of this side effect has been investigated in mitochondria of human cells which finds biguanides to inhibit mitochondrial respiratory complex I and IV proteins[13]. This has been postulated to be a therapeutic mechanism where biguanides cause inefficient coupling of respiration although this relies on the idea that the drugs can accumulate by many orders of magnitude in liver cells which evidence has not been conclusive[5,7]. Other drug targets have been proposed for biguanide drugs but a key problem is some of these proposed mechanisms are not shared by all of the biguanide drugs [13–15].

Human gut microbes are important in regulating human health and they are also known to modulate drug potency and side effects by gut drug metabolism in several cases[16–18]. While metformin therapy is safe and effective there is a sizeable amount (~20%) of type-II diabetes patients that are non-responsive to the drug and around 30% of patients suffer from gastrointestinal symptoms[19,20]. Gut drug metabolism could activate, deactivate or toxify the drug substance and, since there is variability of gut microbe composition from patient to patient, this may cause variable efficacy and incidence of side effects for the drug. Studies that track the disposition of metformin show complete recovery of metformin after intravenous administration but incomplete recovery (~80%) after oral dosage, possibly due to partial, gut drug metabolism[11,21]. This possibly may be due to partial, gut-drug metabolism although a meta-study concluded that metformin metabolism in humans is still equivocal [10]. There is evidence for metformin metabolism to guanylurea in rats [22].

Microbial metabolism of metformin is known in wastewater treatment plants (WWTPs) with some reports finding significant transformation (>80%)[23–25]. The transformation is known to occur due to microbes in the activated sludge of WWTPs with guanylurea found to be the main transformation product[23]. Prior studies have isolated metformin-degrading *Aminobacter* strains from activated sludge where metformin could be a sole carbon or nitrogen source and guanylurea as a fermentation product[26]. Other treatment methods have proven to be ineffective as metformin has poor affinity with activated carbon and chlorination of metformin creates N-chloro species that have been shown to be toxic to human cells[27–29]. Thus, biodegradation of metformin by activated sludge microbes is seen as a viable strategy for its removal, although the enzymes involved in metformin metabolism had previously not been identified.

1.1 Motivation

If the enzymes that encode metformin metabolism in wastewater were known, the enzymes could be used to augment wastewater treatment and in addition, be used as markers to identify gut drug metabolism of metformin in patients. Identifying genes and enzymes that mediate metformin metabolism could give insight on what metabolites are generated and the microbes involved which may reveal more about metformin's action in the gut and possibly how its potency is modulated from patient to patient. Garnering more insight on metformin's action in the gut could identify potential improvements for the therapeutic regimen like probiotics and supplements that could lower the therapeutic dose of metformin of 1-2 grams daily by several fold. This could limit unnecessary off-target effects of the drug that may be responsible for side effects.

While it is found that certain wastewater treatment plants can remove >80% of metformin entering the plants, it may still be necessary to augment this process if wastewater needs to be reused in a timely manner when freshwater sources are scarce. Knowledge of the exact enzymes that degrade metformin could be used in engineering applications to improve current removal rates.

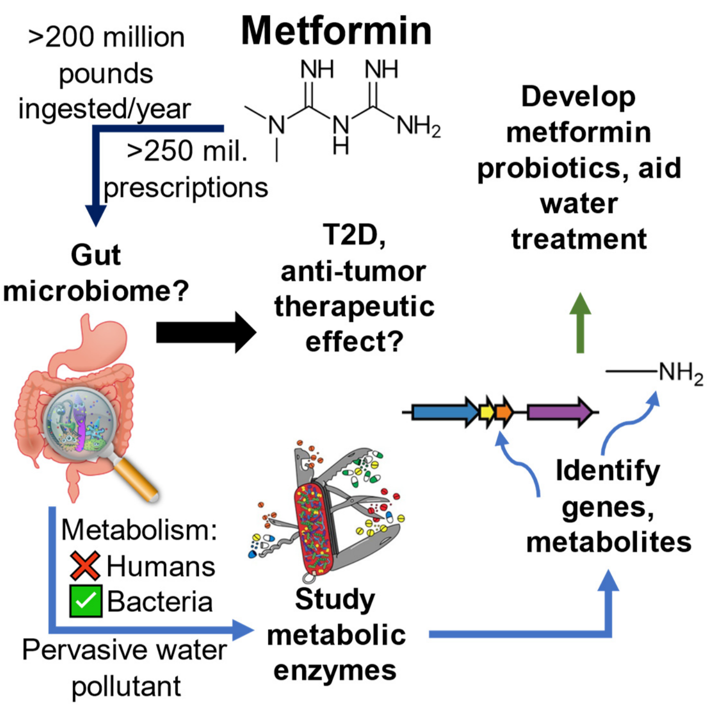


Figure 1.1: Garnering new insights into metformin action

1.2 Dissertation contributions

This dissertation aims to provide insight into metformin metabolism in nature and its action in the human gut: the enzymes that are involved, their mechanisms and the scope of the metabolism in the wastewater and human microbiomes. The genes that encode enzymes used by microbes to completely mineralize metformin into ammonium and carbon dioxide were discovered in this work and were characterized by kinetics, structural analysis and bioinformatics. These gene signatures can now be used to identify metformin metabolic capability of human or wastewater microbiomes which can be important in water treatment and drug potency. The identification of the enzyme initiating metformin metabolism that evolved from the ureohydrolase protein superfamily, also led to the discovery that metformin can bind and potently inhibit ureohydrolase proteins in gut microbiota. This last finding may be very important in revealing the direct mechanism of action for metformin in the gut and may lead to significant discoveries in improving metformin therapy.

1.3 Dissertation outline

The remainder of this dissertation is organized as follows.

- Chapter 2 goes into a guanidine mineralization pathway and the discovery of a carboxyguanidine deiminase (CgdAB) using bioinformatics and structure modelling.
- Chapter 3 describes the discovery of a guanylurea hydrolase (GuuH) from a *Pseudomonas mendocina* strain GU using enzyme kinetics and bioinformatics.
- Chapter 4 unveils the discovery of a Ni^{2+}-dependent metformin hydrolase, MfmAB, characterized by enzyme kinetics and crystallography.
- Chapter 5 presents the discovery of a biguanide hydrolase BguH in *Pseudomonas mendocina* sp. MET-2 detailing enzyme kinetics.
- Chapter 6 reveals a potential inhibition mechanism by metformin on human gut microbes of type-II diabetes patients taking metformin.
- Chapter 7 delivers concluding remarks and potential future directions that could be taken which builds on the findings from this dissertation.

Chapter 2

Solving the conundrum: widespread proteins annotated for urea metabolism in bacteria are carboxyguanidine deiminases mediating nitrogen assimilation from guanidine

Adapted with permission from N.O. Schneider, L.J. Tassoulas, D. Zeng, A.J. Laseke, N.J. Reiter, L.P. Wackett and M. St. Maurice. Solving the Conundrum: Widespread Proteins Annotated for Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine. *Biochemistry*, 59:3258−3270, 2020.

2.1 Introduction

A subset of fungi and bacteria use the enzyme urea amidolyase (UAL) as an alternative to urease in catalyzing urea decomposition (Figure 2.1)[30,31]. 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 carbon dioxide[32]. 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 (AtzF; E.C. 3.5.1.54). UC first catalyzes the carboxylation of a covalently tethered biotin cofactor from bicarbonate 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 AtzF, where it is hydrolyzed to ammonia and CO_2. The enzyme activities for UC and AtzF 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 AtzF connected by a short linker to the C-terminal UC[33–35]. UC and AtzF display a close evolutionary and functional association. It has been suggested that UC and AtzF coevolved in bacteria and, following horizontal gene transfer, subsequently fused into a single UAL gene in fungi[31].

Table 2.1: Substrate specificity of urea carboxylases (UC) and urea amidolyases (UAL)

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	Reference
P. syringae UC	urea	1.4	4.9	2.8×10^2	Schneider *et al*[36]
	guanidine	2.3	0.21	1.1×10^4	
O. sagaranensis UC	urea	5.2	5.2	1.1×10^3	Nelson *et al*[37]
	guanidine	7.4	0.19	4.0×10^4	
K. lactis UAL	urea	14.0	0.26	5.5×10^5	Zhao *et al*[38]
C. albicans UAL	urea	2.6	0.27	9.5×10^4	Schneider *et al*[36]
	guanidine	n.d	n.d	7.5	
S. *cerevisiae* UAL	urea	1.4	0.40	3.3×10^4	Schneider *et al*[36]
	guanidine	n.d	n.d	1.9	

n.d – not determined; *P. syringae – Pseudomonas syringae*; *O. sagaranensis – Oleomonas sagaranensis*; *K. lactis – Kluyveromyces lactis*; *C. albicans – Candida albicans*; *S. cerevisiae – Saccharomyces cerevisiae*

Interestingly, in many bacteria, there are several additional genes of unknown function located in proximity to the genes encoding UC and AtzF, 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)"[39]. 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[31,40,41]. 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[30]. However, urease is broadly distributed in the bacteria[42], and bacteria require transition metals like nickel and cobalt that are otherwise dispensable in the higher fungi[43]. Thus, for the many bacteria that encode UC and AtzF, transition metal regulation does not offer a satisfying explanation for the UC-AtzF/urease redundancy.

An alternative explanation emerged recently, when it was determined that bacterial UC and AtzF are typically under the control of a guanidine riboswitch and that UC displays a strong substrate preference for guanidine over urea (Table 2.1)[37]. This observation led to the proposal that the primary function of UC and AtzF 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.

In Schneider *et al*, using enzymes from *Pseudomonas syringae*, the hypothesis was that UC and AtzF act on guanidine and demonstrate that these two enzymes alone are insufficient to catalyze the decomposition of guanidine[36]. It then became clear that the proteins annotated as urea amidolyase associated proteins, UAAP1 and UAAP2, are subunits of a heteromeric carboxyguanidine deiminase (CgdAB), and it was demonstrated that the combined activities of UC, AtzF, and CgdAB serve to decompose guanidine to NH_3 (Figure 2.1)[36]. The foundational experiments in Schneider *et al* were conceptualized based on a bioinformatic analysis of the CgdAB proteins detailed here and is incorporated in the publication. In addition, structural and phylogenetic analysis was done for urea carboxylase and urea amidolyase to propose a structural determinant for their respective substrate specificities on guanidine and urea.

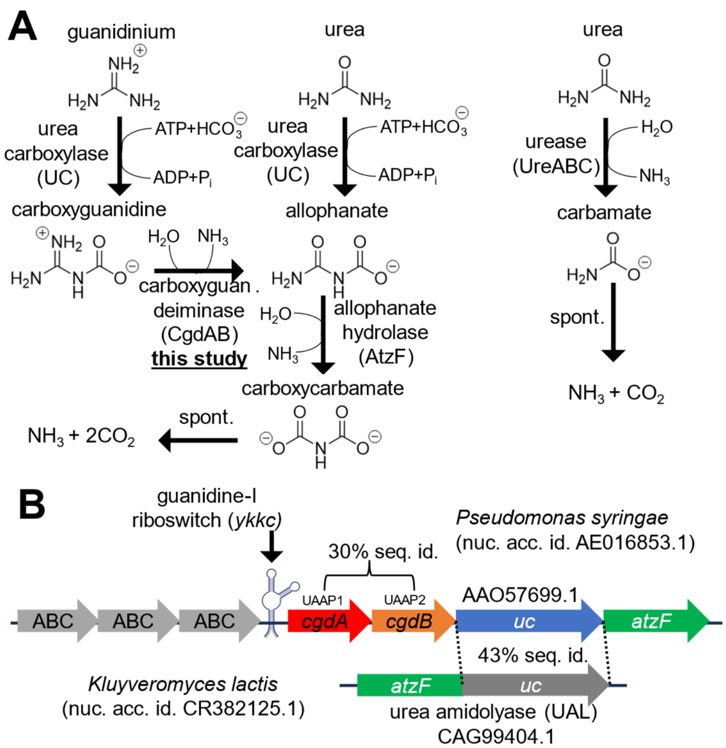


Figure 2.1: Guanidine and urea biodegradation in *P. syringae* and *K. lactis*

(A) Biodegradation pathways for guanidine and urea. Guanidine is first carboxylated by carboxybiotin bound to urea carboxylase (UC) in a ATP-dependent reaction to form carboxyguanidine. This is then hydrolyzed by carboxyguanidine deiminase (CgdAB), or frequently annotated as urea amidolyase associated proteins 1 and 2 (UAAP1/2), to form allophanate which is then broken down by allophanate hydrolase (AtzF). Urea can be broken down via UC and AtzF or urease to be completely mineralized. (B) Genomic contexts of guanidine and urea degrading operons in *P. syringae* and *K. lactis*, respectively. The guanidine degrading genes *uca*, *cgdAB* and *atzF* are co-located on the genome of *P. syringae* along with a set of ABC transporters. The expression of the genes is controlled by a guanidine-I riboswitch just upstream of the *cgdA* gene. In *Pseudomonas syringae* and most bacteria, UC and AtzF are encoded as separate genes while in fungi, these two enzymes are fused together and described as a urea amidolyase (UAL). The UC domain of UAL from *K. lactis* (KlUAL) shares 43% amino acid sequence identity with the *P. syringae* UC (PsUC).

2.2 Methods

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[44]. Cytoscape was used to visualize the clustering in the SSN and identify the clusters containing CgdA and CgdB sequences[45]. 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 cooccurring TetR protein and *ykkC* riboswitch using the software HMMER v 3.1b2 [46]. 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), AtzF (TIGR02713) and urea carboxylase associated transport proteins (TIGR0327, TIGR03428)[47]. These models were then used with the tool RODEO (Rapid ORF Description & Evaluation Online) to analyze the gene contexts of 2400 CgdB RefSeq sequences within a ± 8 gene window[48].

2.3 Results

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

The α and β subunits of Cgd from P. syringae (PsCgdA and PsCgdB) 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 2.1). As of August 2020, 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 amidolyase-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 *cgdA* and *cgdB* 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.

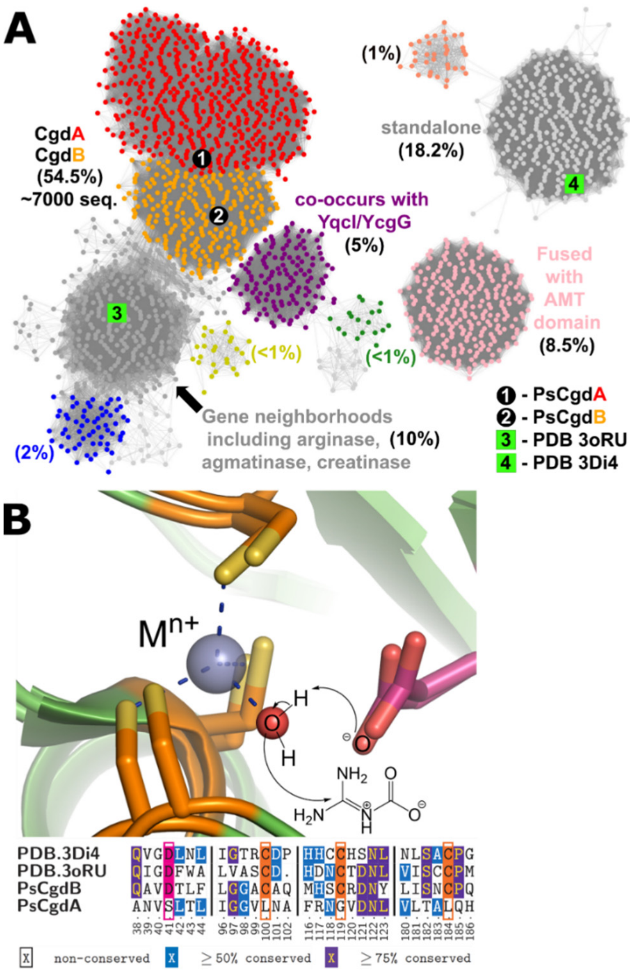


Figure 2.2: Carboxyguanidine deiminase within the DUF1989 Family:

(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 structure and the CgdA/B sequences characterized in this study are labelled with black circles. Percentages are a percent of total sequences for each labelled 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 PsCgdA and PsCgdB indicate these residues are conserved in PsCgdB but not PsCgdA. Sequence numbering is based on PsCgdB.

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. The DUF1989 family, while not previously characterized functionally, has two members for which X-ray structures have been determined (PDB 3ORU, 3DI4) (Figure 2.2). 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 *cgdA* and *cgdB* 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. The DUF1989 family, while not previously characterized functionally, has two members for which X-ray structures have been determined (PDB 3ORU, 3DI4) (Figure 2.2). 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 2.2B). 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)[49].

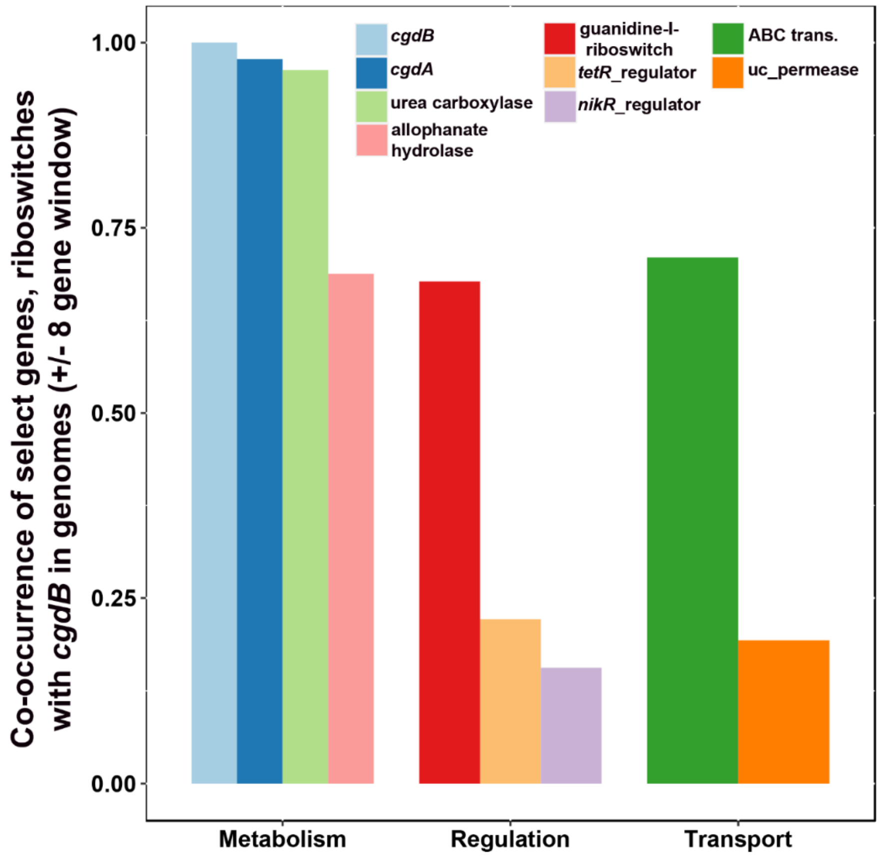
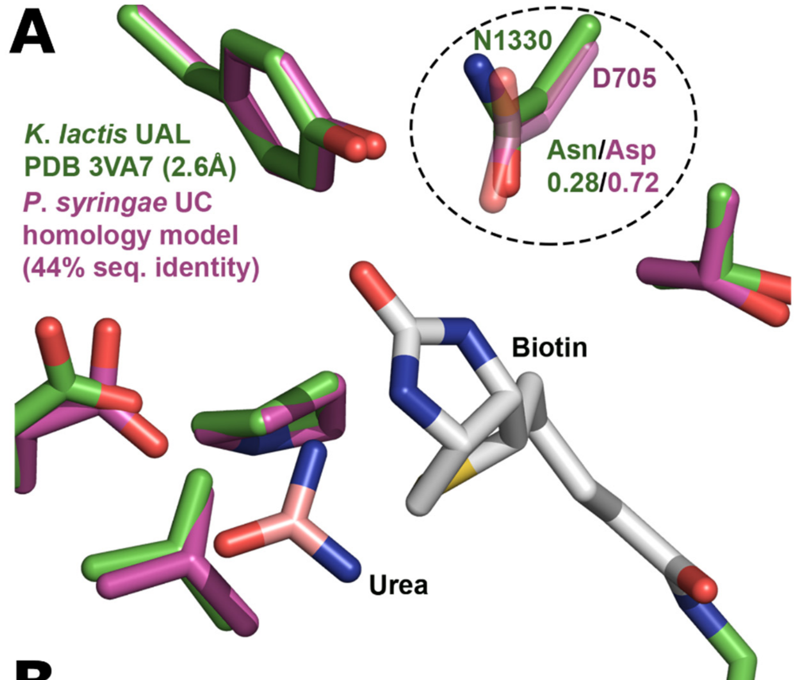


Figure 2.3: Co-occurrences with CgdB in Genomes: Gene contexts within a ±8 gene window around 1200 CgdB sequences from RefSeq genomes were counted for the genes and the guanidine-I-riboswitch above using HMMs. Since ABC transporters are multi component, the count was scaled down to the average number of ABC transporter genes per CgdB observed (n=3).

Evaluating the gene contexts of *cgdB* sequences within an eight gene window indicates that >96% colocalize with the UC gene and more than two-thirds colocalize with an AtzF gene (Figure 2.3). Thus, this gene cluster encodes all of the necessary genes for a bacterium to decompose guanidine and urea to ammonia (Figure 2.1A). 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 *cgdB* 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 primary shell, active site residues are conserved across fungal UALs and bacterials UCs 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 2.4A). 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 aspartate residue at this position had *cgdAB* in the gene neighborhood, while less than one percent of UC genes encoding an asparagine residue had nearby *cgdAB* genes (Figure 2.4B). This correlation suggests that the aspartate or asparagine tracks with preferential activity versus guanidine or urea, respectively.



B

Variant	% Co-occurrence of UC genes with *cgdB* in genomes
UC Total (Asp + Asn)	70% (1160 of 1661 seq.)
Asp variant (e.g. PsUC D705)	97% (1156 of 1188 seq.)
Asn variant (e.g. *K.lactis* UAL N1330)	0.8% (4 of 473 seq.)

Figure 2.4: Potential Structural Determinant of Substrate Specificity in urea carboxylase: (A) Active site overlay of fungal UAL from Kluyveromyces lactis (PDB 3VA7) and a homology model of PsUC. Active site residues are highly conserved across fungal and bacterial UCs except for one residue position that is predominantly an aspartate in bacteria or an asparagine residue in fungi and some bacteria. (B) Strong correlation between N1330/D705 in UC and Co-occurrence of CgdB sequences in genomes. UC sequences from the RefSeq database were separated into two groups, one with the N1330 variant and the other with the D705 variant. The gene contexts within a ±8 gene window were evaluated around the two groups of UCs to look for co-occurrence of CgdB. As a result, UCs with the N1330 variant had very little co-occurrence with CGDs in genomes (<1%) while UCs with the D705 variant are almost ubiquitous with CGDs (>97%).

2.4 Discussion

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 (Figure 2.2).

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. 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[50]. Arginase, meanwhile, acts as an amidinohydrolase that employs a dinuclear metal center to activate water for nucleophilic attack[51]. 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 2.2). 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[52]. Despite the potential similarities in mechanism, there is no apparent structural homology between CgdAB and either adenosine deaminase or cytidine deaminase.

The combined activities of UC, CgdAB and AtzF 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 ~70% 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 for *P. syringae* UC and for *O. sagaranensis* UC (Table 2.1)[36–38]. This subset of enzymes should appropriately be renamed as guanidine carboxylases, in accordance with prior recommendations[37]. 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 2.4). X-ray crystallography of *P. syringae* UC would be beneficial in seeing if there are significant differences in the secondary structure its active site compared to the *K. lactis* UAL crystal structure, that may dictate substrate specificity.

Chapter 3

Filling in the gaps in metformin biodegradation: a new enzyme and a metabolic pathway for guanylurea

Reprinted with permission from L.J. Tassoulas, A. Robinson, B. Martinez-Vaz, K.G. Aukema and L.P. Wackett. Filling in the Gaps in Metformin Biodegradation: a New Enzyme and a Metabolic Pathway for Guanylurea. *Applied and Environmental Microbiology*, 87(11): e03003-20, 2021.

3.1 Introduction

Guanylurea (carbamylguanidine) is one of the most widespread water contaminants, originating and accumulating from metformin, a first-line globally important drug[53,54]. Metformin is widely prescribed as a treatment for type 2 diabetes and has even heightened significance in light of its observed anti-inflammatory, anticancer, and antiaging effects[2,25,55,56]. Most recently, metformin has been associated with reduced mortality in COVID-19 patients suffering from preexisting conditions such as diabetes and obesity[57]. Those studies also showed that human drug-metabolizing enzymes do not transform metformin, the drug is largely excreted unchanged, and microbes in wastewater treatment plants and aquatic environments transform it to guanylurea.

Guanylurea has been widely labeled as a recalcitrant “dead-end” product of metformin in numerous wastewater treatment systems[58,59]. Guanylurea forms, and has been detected in, wastewater treatment plant effluent and coastal waters around the globe[23,25,60,61]. Metformin and guanylurea have been reported in surface waters at levels of micrograms per liter, which is extremely high for a pharmaceutical contaminant[23,62].

Metformin was detected in European surface waters at levels of up to 3 mg/L, with guanylurea concentrations exceeding those of metformin by an order of magnitude or more[63–65]. It is also relevant that metformin and guanylurea are charged molecules that bind poorly to granulated activated carbon, the worldwide standard treatment for removing pharmaceuticals, which are largely hydrophobic molecules[23,66]. In that context, biodegradation is even more crucial for their removal. The levels of these compounds entering municipal wastewater treatment are likely to increase due to metformin's broadening pharmacological efficacy, its widespread use, projected increases in the rates of obesity and diabetes, and the emerging new uses for the compound[54,65]. There is also evidence that metformin and guanylurea may impact some aquatic species[63,67,68]. In light of these different impacts, guanylurea is considered in many countries to be one of the most prominent emerging water pollutants[69].

While microbial metabolism of metformin is considered to be the major source of guanylurea, it can also derive from other anthropogenic and natural sources. Guanylurea may also accumulate from the microbial degradation of cyanoguanidine (dicyanodiamide), a common fertilizer additive[70,71]. Derivatives of guanylurea are utilized in the manufacture of flame retardants as well as propellants for energetic compounds and munitions[72,73]. Natural products containing the guanylurea moiety have been identified in the red alga *Gymnogongrus flabelliformis*. These compounds include the novel amino acid gigartinine and a guanylureido acid named gongrine[74–76], which are proposed to play a role in nitrogen metabolism[77] and may contribute to faster spring growth for certain marine plants[78].

Given the prevalent use of metformin and frequent detection of guanylurea in aquatic environments, research on the microbial degradation of these compounds is essential to develop biotechnological applications for bioremediation and wastewater treatment. Several studies have recently investigated the biodegradation of metformin and guanylurea by microbial communities isolated from activated sludge[58,61]. In one study not involving metformin, the disappearance of guanylurea was shown to be accompanied by the appearance of guanidine as measured by high-performance liquid chromatography (HPLC) analyses, suggesting guanidine as a potential transformation product[72]. These studies examined the microbial breakdown of guanylurea but did not define metabolic pathways or the genes and enzymes mediating breakdown. The present study describes the biodegradation of guanylurea by a *Pseudomonas mendocina* bacterium isolated, by enrichment, from a wastewater treatment plant. This bacterium utilized guanylurea as a nitrogen source for growth. Genome sequencing and bioinformatic analyses led to the identification of genes involved in the biodegradation of ureide and guanidinium compounds. A new member of the isochorismatase-like hydrolase protein family was shown to catalyze the conversion of guanylurea to guanidine, and the enzyme was characterized physically and kinetically. A mineralization pathway to ammonia and carbon dioxide was demonstrated. Insights obtained here may now be used with wastewater metagenome data to predict the intrinsic capacity for guanylurea biodegradation.

3.2 Methods

Enrichment cultures and isolation of guanylurea-degrading bacteria

The bacterial *Pseudomonas mendocina* strain GU was isolated from a sample of activated sludge collected at the metropolitan wastewater treatment plant in Saint Paul, MN. Isolation was achieved by enrichment culture with citrate-acetate medium and 1 g of sludge per 50 mL of minimal medium as the inoculum. The minimal medium contained the following per liter of deionized water: 5.45 g potassium phosphate diabasic, 0.2 g magnesium sulfate heptahydrate, 0.1 g sodium chloride, 1.5 g sodium acetate, and 4.7 g sodium citrate[79]. Guanylurea (1 mM) was then added as the sole nitrogen source. Enrichments and isolates were grown at 30°C in a shaking incubator at 200 to 225 rpm.

Cultures were transferred in 10-fold dilutions into fresh medium every 7 days. Individual isolates were obtained by plating 10-fold serial dilutions of the enrichments on selective guanylurea plates, transferred to Luria-Bertani (LB) plates, and then isolated by streaking on LB and minimal medium plus guanylurea plates until pure.

Growth studies

Growth studies were conducted in triplicates in citrate-acetate medium containing 1 mM each of their respective nitrogen source: guanylurea, biuret, guanidine, urea, or metformin. Nitrogen-free citrate-acetate medium served as a negative control. Citrate-acetate medium containing 6 mM ammonium chloride (NH_4Cl) served as a positive control. Cell growth was monitored spectrophotometrically at 600 nm initially at 12-hr to 24-hr intervals. Cultures with short lag phases were monitored every 4 to 6 hrs. DNA extraction and PCR testing for taxonomic identification. Genomic DNA was purified utilizing a Qiagen DNeasy blood and tissue kit (Qiagen, Valencia, CA) and Promega Wizard genomic DNA purification kit (Promega, Madison, WI). PCR amplification for taxonomic identification of bacterial strains was performed with universal 16S primers, 530F (59-GTGCCAGCMGCCGCGG-39) and 1492R (59-GGTTACCTTGTTACGACTT-39), using Q5 high-fidelity polymerase (New England BioLabs, Ipswich, MA). Amplification of the 1.4-kb DNA 16S rRNA fragment was achieved using the following conditions: 98°C for 2min, and then 35 cycles consisting of 95°C for 1 min, 60°C for 30 sec, and 72°C for 1 min; a final extension of the PCR product was performed at 72°C for 5 min. All PCRs were carried out using a concentration of 0.5 mM for each primer. Amplification products were purified using the QIAquick DNA extraction kit and sequenced (Functional Biosciences, WI, USA) to determine the identity of the microbial isolate. HPLC analysis. HPLC testing was conducted with a Waters IC-Pak anion HC column, 4.6 by 150 mm. Isocratic mobile phase consisted of a 5 mM KPO_4 buffer at pH 8.0, and the flow rate was 0.5 mL/min.

Eluents were monitored at 194 nm. Standards were in 125 mM KPO_4 buffer, pH 8.0. Spent medium was analyzed by HPLC after cells were grown in citrate-acetate medium containing 1 mM guanylurea for 24 hrs. Cells were centrifuged at 3,000×g for 15 min, and supernatant was then collected and filter sterilized through a 0.2 µm filter. Controls included sterilized N-free medium and 1 mM guanylurea solution. Enzyme was prepared at a concentration of 1mg/mL in 125 mM KPO_4 buffer, pH 8.0, with 1 mM guanylurea. Samples were incubated either overnight at room temperature or for 30 min and boiled for 2 min to inactivate enzyme.

Genome sequencing and analysis

Total genomic DNA from microbial isolates was sequenced using a Roche GS 454 FLX system and standard LR 70 chemistry. Illumina Nextera XT library preparation and sequencing (on a MiSeq with V3 chemistry and 300-bp paired-end reads) services were provided by the University of Minnesota Biomedical Genomics Center (BMGC, Saint Paul, MN, USA). Adaptors and low-quality bases were trimmed from raw reads with Trimmomatic v 0.36[80]. De novo assembly was performed using SPAdes v 3.13.0[81]. Initial genome annotation was performed with Prokka v 1.12[82]. Sequence similarity networks were made using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) retrieving 10,000 sequences related to GuuH[44]. The network was visualized in Cytoscape, and the clusters of guanylurea hydrolase, biuret hydrolase, and triuret hydrolase were identified as presented in Figure 3.5[45].

Protein expression and chromatography

The putative GuuH gene (NCBI accession no. MBF8163004.1) was codon optimized for E. coli and synthesized by Integrated DNA Technologies (Coralville, IA). The gene was cloned into a pET28b vector with Gibson assembly (New England BioLabs) using NdeI and HindIII restriction sites with an N-terminal 6xHis tag and transformed into BL21-DE3 *E. coli* cells (New England BioLabs). Site-directed mutants of the biuret hydrolase from Herbaspirillum sp. BH-1 (NCBI accession no. PLY61274.1) and *guuH* were made using the Q5 site-directed mutagenesis kit from New England BioLabs. The *guuH* gene was expressed by growing cells in lysogeny broth (LB) medium with 50 mg/mL kanamycin at 37°C and 200 rpm to an OD_{600} of 0.6 in a shake flask. The culture was cooled to 14°C and induced with 1 mM isopropyl b-D-1-thiogalactopyranoside (IPTG) and, with the same agitation, incubated for 20 hrs. Induced cells were harvested at 4,000×g and resuspended in lysis buffer (20 mM sodium phosphate, 200 mM NaCl, pH 7.4). Cells were lysed with three passes in a French press at 10,000 psi. The cells were centrifuged for 90 min at 19,000×g, and the supernatant was passed through 0.2 µm filter prior to loading into a GE Life Sciences AKTA fast liquid protein chromatography (FPLC) system for injection onto a GE Life Sciences HisTrap HP 5-ml column. After washing to limit nonspecific protein binding, GuuH was eluted with a linear gradient from 0.1 M to 0.25 M imidazole in lysis buffer over 10 column volumes with flow rate set at 2 mL/min, and fractions were collected. To determine the oligomeric state of GuuH, gel filtration was performed using a GE Healthcare HiLoad 16/600 Superdex 200-pg column. The column was equilibrated with 50 mM Tris, 200 mM sodium chloride at pH 8, and the sample was injected onto the column and washed with 1 column volume at 1 mL/min flow rate, with GuuH eluted as a homotetramer at ~96 kDa.

Enzyme assays

Enzyme activity was measured by detection of ammonia release using the Berthelot reaction as previously described [83]. Experiments were performed at room temperature in 125 mM sodium phosphate dibasic buffer at pH 8. Specific activity at 25°C was determined at the pH optimum for the enzyme using the appropriate buffer (acetate, pH 5.5 to 6.5; phosphate, pH 7.0 to 8.0; borate, pH 8.5 to 10.5) with 1 mM guanylurea. CgdAB from *Pseudomonas syringae*, a heterodimeric enzyme (NCBI accession no. WP_005764729.1 and WP_005764727.1, respectively), and allophanate hydrolase from *Pseudomonas* sp. strain ADP (NCBI accession no. WP_011117193.1) were employed in coupled enzyme assays to determine if carboxyguanidine is formed during GuuH hydrolysis of guanylurea[36]. The CgdAB protein was a gift from Martin St. Maurice of Marquette University, and it was shown to be active in a coupled assay with guanidine carboxylase with guanidine as the substrate. Enzyme kinetic assays were performed in a 125 mM sodium phosphate dibasic buffer solution, pH 8, with 1 mM guanylurea as the substrate. To initiate the reaction, GuuH was added to a final assay concentration of 0.27 mg/mL. CgdAB was added to a final assay concentration of 15.6 mg/mL and AtzF protein to a final concentration of 7.38 mg/mL. Ammonia release was measured utilizing the Berthelot assay[83]. The assay was measured in a Beckman Coulter DU 640 UV-visible (UV-Vis) spectrophotometer.

A coupled enzyme assay to determine stoichiometries of the guanylurea and ammonium liberated utilized GuuH, AtzF, and CgdAB proteins. The reactions were performed overnight in a solution containing 50 mM guanylurea in 125 mM sodium phosphate, pH 8. Enzyme concentrations were 60 mg/mL GuuH, 15.6 mg/mL CgdAB, and 7.38 mg/mL AtzF. Reaction mixtures were incubated at 20°C overnight, and stoichiometries were determined by ammonia release utilizing the Berthelot reaction as described previously.

Data availability

Genome sequences for *Pseudomonas mendocina* strain GU are available in GenBank at BioProject accession PRJNA675777 or BioSample accession SAMN16722328.

3.3 Results

Enrichment and isolation of a pure culture growing on guanylurea

Activated sludge from the Saint Paul, MN, municipal wastewater treatment plant was used as a source of bacteria for enrichment on guanylurea as the sole nitrogen source. After the first five transfers, guanylurea utilization was indicated by a significant increase in optical density in liquid culture and appearance of individual colonies on agar plates containing LB or guanylurea minimal medium. Subsequent transfers and plating of individual colonies led to the isolation of a pure culture with guanylurea-degrading capability as demonstrated by HPLC. No guanylurea was detected in the spent growth medium of the cultures after 24 hrs of incubation. Control media which were not inoculated did not exhibit significant guanylurea disappearance (less than 5%). The 16S rRNA sequence analysis identified the bacterium as a *Pseudomonas mendocina*, and it was designated here as strain GU. *Pseudomonas mendocina* strain GU was able to utilize guanidine, agmatine, and urea as sole sources of nitrogen for growth, using citrate and acetate as carbon sources (Supp. Table A.1). The doubling time with guanylurea was 44 min. Doubling times with other nitrogen-rich compounds analogous to guanylurea ranged from 66 to 92 min: agmatine (66 min), guanidine (81 min), and urea (92 min). No growth was observed when biuret, metformin, or cyanoguanidine was used as the sole nitrogen source.

Genome sequencing and analysis of P. mendocina strain GU

The strain was subjected to Illumina sequencing with 36-fold coverage. The *P. mendocina* strain GU genome showed a high degree of gene sequence relatedness (Supp. Table A.2) and synteny (Supp. Fig. A.1) with *P. mendocina* ymp, which was isolated from Yucca Mountain hazardous waste repository [84]. Properties of the two genomes are compared in Table 3.1. Overall, the two genomes shared 98% average nucleotide sequence identity, and 77.6% of the predicted proteins in the genome of *P. mendocina* strain GU shared >95% sequence identity to proteins present in the genome of strain ymp (Table S1). The major difference was the presence of an additional 665 coding sequences in strain GU (12.4% of the genome). Taken together, only 10% of the shared proteins were less than 95% identical.

Table 3.1: Comparison of general genome properties of Pseudomonas mendocina ymp and Pseudomonas mendocina strain GU[a]

General sequencing and genome property	*Pseudomonas mendocina* ymp	*Pseudomonas mendocina* strain GU
Genome coverage (fold)	14	36
Size (Mb)	5.0	5.7
G+C content (%)	64.7	64.3
No. of protein-coding genes	4,704	5,378
Coding regions (% genome)	90.8	91.5
Avg ORF[b] size (bp)	980	975

[a] *P. mendocina* ymp was sequenced by the Joint Genome Institute and strain GU was sequenced as part of the present study.

[b] ORF, open reading frame.

Identification and purification of a guanylurea-degrading enzyme

Guanylurea degradation genes or enzymes had not previously been identified. In light of this, the genome sequence of *P. mendocina* strain GU was mined for the presence of genes that might encode enzyme classes that act on the guanidine or urea functional groups that, in combination, compose guanylurea. Genes encoding enzymes annotated to be active with the guanidinium compounds arginine and agmatine were identified but, based on the flanking genes, were thought not to be relevant to guanylurea. The organism has genes encoding an active urease, but urease was previously tested for activity with guanylurea and shown to be inactive[85]. A gene encoding a protein annotated from BLAST comparisons as “cysteine hydrolase” and “nicotinamidase” was analyzed in more detail. The translated amino acid sequence showed 48% sequence identity to a biuret hydrolase (BiuH) from *Rhizobium leguminosarum* that had been characterized functionally[86]. We noted that biuret is structurally analogous to guanylurea, having only an oxygen replacing one of the guanylurea guanidinium nitrogen atoms (Figures 3.1A and 3.1B). In this context, a synthetic gene encoding the biuret hydrolase homolog (NCBI accession no. MBF8163004.1) was obtained and expressed in *Escherichia coli*, and the His-tagged enzyme was purified in one step to homogeneity via a Ni-NTA affinity column. Incubation of the purified enzyme with guanylurea produced 1 mol of ammonia per mol of guanylurea, identifying this enzyme to be capable of degrading guanylurea. No redox cofactors or oxygen was required, indicating that the enzyme was a guanylurea hydrolase.

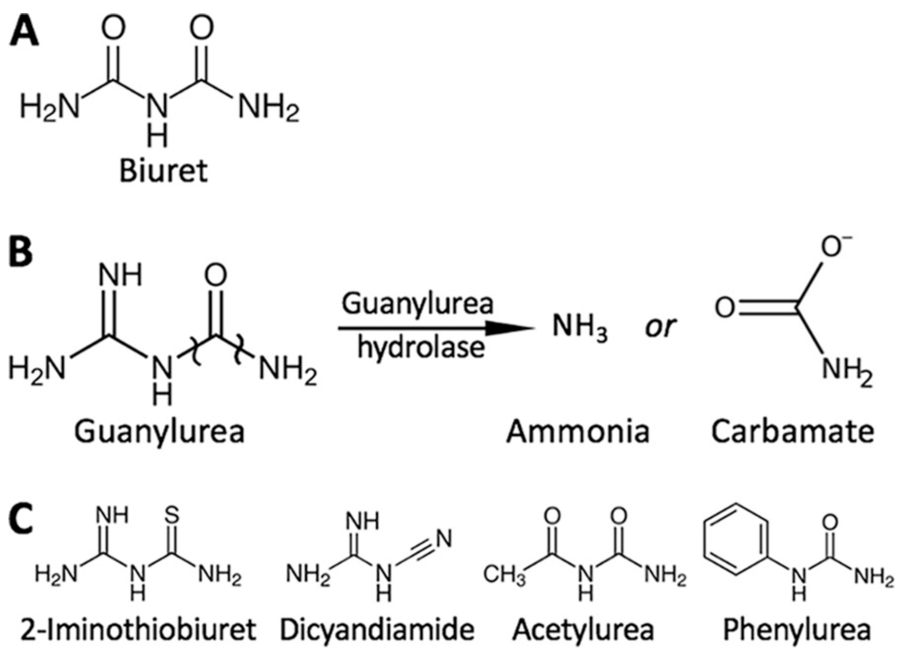


Figure 3.1: Compounds and reaction in Chapter 3

Characterization of the guanylurea hydrolase enzyme

It was important to examine the kinetics and substrate specificity of the enzyme described above to determine if this enzyme was competent to support significant guanylurea degradation and if that was the major native activity. The purified enzyme was active over a pH range of 5.5 to 8.5, with highest specific activity at pH 8.0. At the pH optimum, the specific activity of purified guanylurea hydrolase with saturating guanylurea was 13 µmol per min per mg GuuH (Table 3.2). The k_{cat}/K_M was higher (3-fold) than for the average enzyme (~1×10^5 $M^{-1}s^{-1}$) with its cognate substrate, as compiled from a study of thousands of enzymes by Davidi and coworkers[87]. The enzyme showed minimal activity on biuret, with 0.06% the specific activity measured on guanylurea. Moreover, no activity (0.01 µmol per min per mg) was detected with acetylurea, dicyandiamide, phenylurea, nitroguanidine, and 2-imino-thiobiuret (Figure 3.1C). These data indicate that the major biological function of the enzyme studied here is to release ammonia from guanylurea, allowing *P. mendocina* strain GU to grow on guanylurea.

Table 3.2: Biochemical and molecular properties of guanylurea hydrolase

Property	Experimental determination
Subunit mol wt	24.8 kDa
Subunit structure	α_4
Calculated pI	5.5
pH optimum	8.0
Sp act	13 µmol/min/mg
k_{cat}	5.2 s^{-1}
K_m	16 µM
k_{cat}/K_m	3.3×10^5 $M^{-1}s^{-1}$

Elucidating the guanylurea hydrolase reaction product

In addition to ammonia, guanidine was stably produced in incubations with guanylurea hydrolase and guanylurea. This could arise from hydrolytic cleavage of the terminal urea C-N bond or the subterminal N-C bond (Figure 3.1B). The former cleavage reaction would produce ammonia and carboxyguanidine, which readily decarboxylates, forming stable guanidine. The subterminal bond cleavage would produce guanidine directly, along with carbamate. Carbamate at neutral pH in water has a half-life of 70 ms[88], rapidly converting to carbon dioxide and ammonia. So, both reactions produce ammonia and guanidine rapidly from unstable intermediates. However, it is now possible to "observe" carboxyguanidine because an enzyme that rapidly converts it to allophanate and ammonia was recently discovered[36]. In that previous study, guanidine carboxylase produced highly unstable carboxyguanidine that was rapidly hydrolyzed into a more stable compound, allophanate, via carboxyguanidine deiminase (CgdAB). Allophanate hydrolase was used to convert allophanate to ammonia and carbon dioxide. Here, an analogous experiment was performed by adding the enzymes CgdAB and allophanate hydrolase to GuuH reactions. If carboxyguanidine was formed, the additional enzymes would release an additional three equivalents of ammonia. If guanidine and carbamate were formed, no additional ammonia would be released. None of the enzymes have activity with guanidine.

In two separate experiments, carried out as described in the Materials and Methods, no increase in ammonia was observed in kinetic (short-term incubations) or stoichiometric (long-term incubations) experiments, indicating that guanidine is produced directly. In these experiments, an up to 40-fold excess of coupling enzymes were added, which would have released additional ammonia from carboxyguanidine, if that had been produced by GuuH.

Gene regions of the P. mendocina strain GU genome related to guanylurea mineralization

With the observation that guanidine and carbamate are the likely products of the GuuH enzyme, and carbamate that decays spontaneously liberates one ammonia molecule, further ammonia release would require enzymes implicated in guanidine degradation. Guanidine metabolism has only recently been elucidated, and special identifying features of the genes were reported[36]. That information was used to aid in bioinformatic analysis of predicted protein-encoding regions in the *P. mendocina* strain GU genome (Figure 3.2).

First, the guanylurea hydrolase gene (*guuH*), identified in this study, did not appear to be contiguous to genes encoding enzymes involved in related metabolism that we could discern (Figure 3.2A). The only identifiable surrounding genes were ABC transporters of unknown function. A gene region likely to be involved in guanidine metabolism was localized ~400 genes distant (Figure 3.2B). The large gene clearly encodes an enzyme including the biotin-binding domain and other regions diagnostic for carboxylases[89]. The carboxylase is further identified as a guanidine-metabolizing enzyme by a specific active region demarcated in a previous study[36]. Guanidine carboxylases contain an aspartic acid residue in the active site at a position typically occupied by asparagine in fungal and bacterial urea carboxylases[36], and the protein here contained that diagnostic aspartate. Even stronger evidence is provided by cooccurrence of the *cgdAB* genes next to our annotated guanidine carboxylase (Figure 3.2B). The *cgdAB* genes are present in ~7,000 bacterial genomes, and 96% of the time, they colocalize with a guanidine carboxylase[36]. The CgdA and CgdB proteins were shown previously to form a complex and to transform the product of guanidine carboxylase, carboxyguanidine, to allophanate. Lastly, guanidine metabolism gene expression was shown to be regulated by guanidine riboswitches[37], and these are found directly upstream of the enzyme-encoding genes just described (Figure 3.2B). Another gene region (genes 300 to 302) in the *P. mendocina* strain GU genome encoded an allophanate hydrolase, an adjacent urea carboxylase, and a regulatory protein, respectively (Figure 3.2C). The carboxylase protein encoded by gene 301 contained an asparagine instead of aspartate at its active site, a feature consistent with a preference for urea as the substrate undergoing carboxylation[36]. Urea carboxylation produces allophanate, and allophanate hydrolase can degrade that to ammonia and carbon dioxide. In the presence of guanylurea, the five genes highlighted in red in Figure 3.2 encode four proteins (CgdA and CgdB function as one protein) that constitute a complete pathway for guanylurea that can liberate all four nitrogen atoms as ammonia (Figure 3.2D).

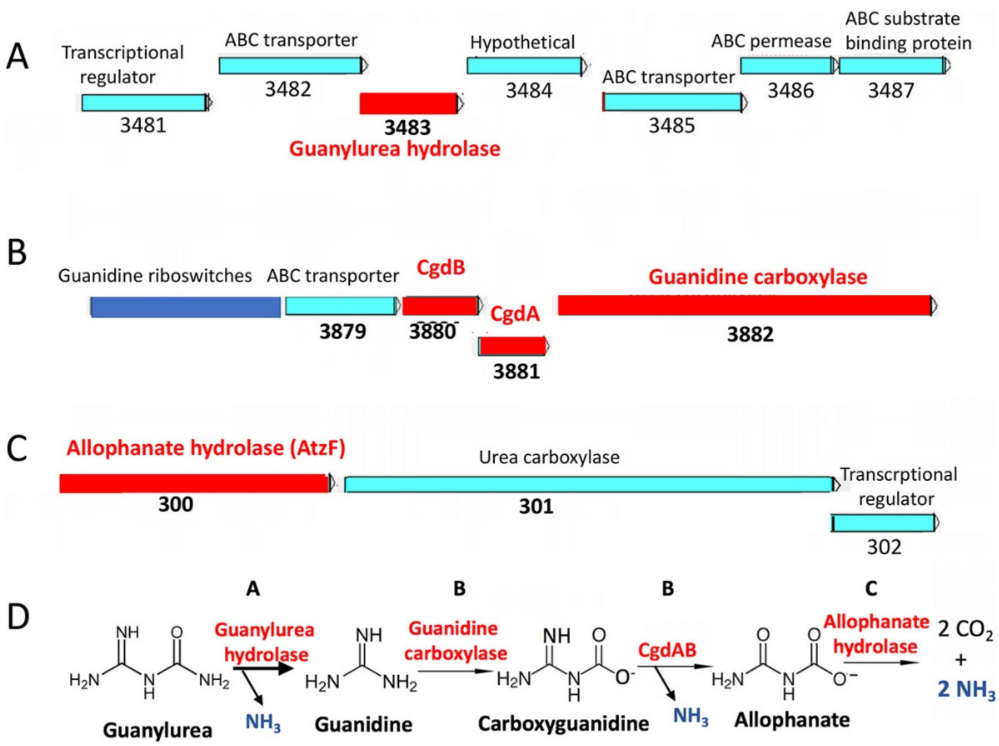


Figure 3.2: Gene regions identified in P. mendocina strain GU and biodegradative pathway

(A) Genes 3481 to 3487; (B) genes 3879 to 3882; (C) genes 300 to 302. (D) Metabolic pathway via enzyme reactions and NH_3 stoichiometry.

In addition to the genomics and bioinformatic inferences described above, there is direct experimental data consistent with the pathway shown in Figure 3.2D. First, *P. mendocina* strain GU grows readily on guanidine as a sole nitrogen source. Only one metabolic pathway for guanidine is currently known and that proceeds as shown in Figure 3.2, via guanidine carboxylase, CgdAB, and allophanate hydrolase[36], for which genes are present in *P. mendocina* strain GU. To further test the pathway shown in Figure 3.2D experimentally, we carried out ammonia stoichiometry experiments. If all of the relevant genes shown in Figure 3.2 are expressed, it is predicted that 4 equivalents of ammonia would be produced for each guanylurea molecule consumed. To test this hypothesis, parallel cultures of *P. mendocina* strain GU were grown with limiting amounts of nitrogen, and then cell densities were determined when growth ceased. As shown in Figure 3.3, the growth on 1 mol of guanylurea, containing four nitrogen atom equivalents, was the same as with 4 mol of ammonium ion. For example, separate growth cultures containing 1.92 mM ammonium chloride or 0.48 mM guanylurea each grew to an optical density at 600 nm (OD_{600}) of 0.8. Both contained the same number of nitrogen atoms, equivalent to 1.92 mM ammonia. These data are consistent with guanylurea being completely mineralized to release all the nitrogen atoms contained within the compound. One can draw out a chemically plausible pathway in which guanylurea undergoes a direct deamination reaction to produce biuret, but biuret was negative as a growth substrate and metabolism producing biuret would only yield one nitrogen equivalent, not four as was observed.

Guanylurea hydrolase is a new member of the isochorismatase hydrolase-like protein family

A query of the Pfam database with the GuuH amino acid sequence gave a match to the isochorismatase hydrolase-like (IHL) protein family (PF00857.20) with an E-value of 7.1E-242. GuuH consists of 231 amino acids and matched extensively over amino acids 26 to 217 with biuret hydrolase (BiuH) and triuret hydrolase (TrtA) (Figure 3.4). IHL proteins are sometimes denoted as cysteine hydrolases, and GuuH contains a cysteine, C171, that aligns with a cysteine in BiuH and TrtA and found in a highly conserved region (Figure 3.4A).

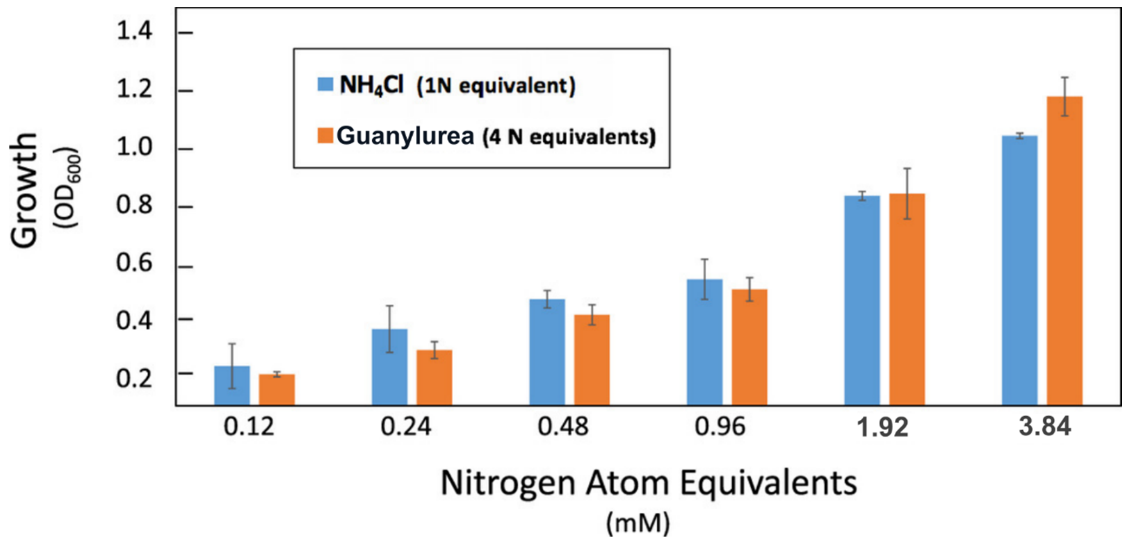


Figure 3.3: Nitrogen growth comparison by setting molar amounts of nitrogen in ammonium chloride equivalent to guanylurea

X-ray structures have been determined for biuret[90] and triuret hydrolases[91], and those studies have revealed the cysteine aligning with C171 in GuuH acts as a nucleophile to catalyze initial bond cleavage. The resulting acyl enzyme intermediate further undergoes hydrolysis to complete the reaction. Those structural and mechanistic studies have also revealed that the biuret and triuret hydrolase active sites both contain a D-K-C catalytic triad, of which all the residues reside in GuuH in alignment and in highly conserved regions of the protein overall (Figure 3.4A). One notable difference in the sequences was with a glutamine residue that helps bind substrate in biuret and triuret hydrolases and that was replaced with a glutamate in GuuH, as illustrated in Figure 3.4B. X-ray structures for biuret and triuret hydrolases reveal that the glutamine residue hydrogen bonds to the substrate amide group distal to the one reacting at the active site cysteine[90,91]. The change to an aligning glutamate, Glu221, in GuuH is logical, chemically, as the negatively charged glutamate would be expected to bind electrostatically to the positively charged guanidinium group, thus aligning the guanylurea substrate in a similar manner as biuret and triuret hydrolases align their substrates. This suggests that the glutamine-to-glutamate change is key for differentiating GuuH enzymes from biuret and triuret hydrolases.

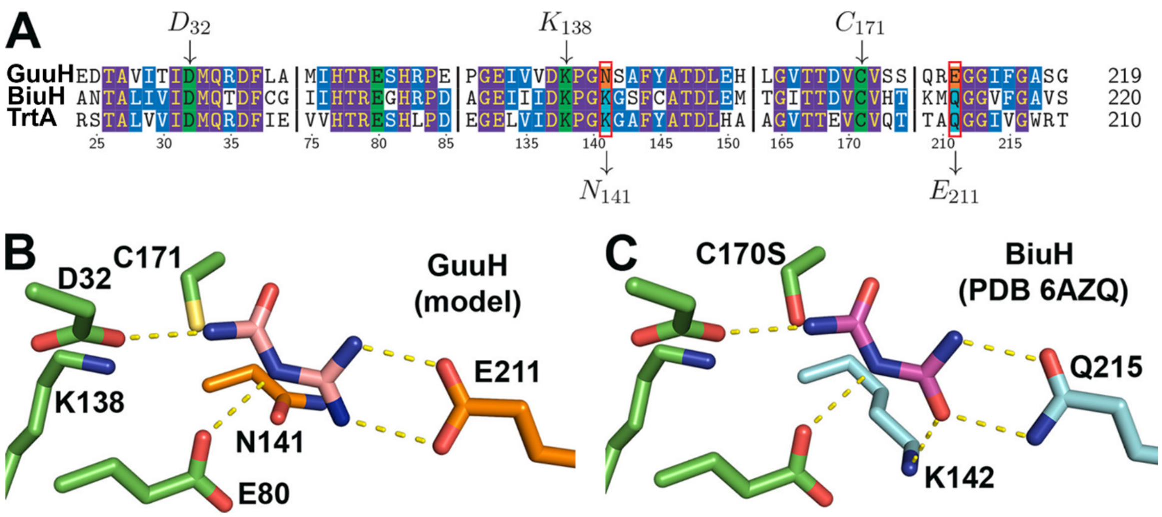


Figure 3.4: Sequence alignment and structural comparisons of guanylurea hydrolase (GuuH) with biuret hydrolase (BiuH) and triuret hydrolase (TrtA)

(A) Sequence alignment was made using NCBI COBALT alignment tool with the BiuH (PLY61274.1) and TrtA (PLY61272.1) sequences from Herbaspirillum BH-1. (B and C) Key amino acid positions differentiating guanylurea (E211) and biuret (Q215) hydrolases, derived from PDB structure 6AZQ. Homology model of GuuH made with BiuH as a template (48% sequence identity). The C170S variant of BiuH is catalytically dead, which allows for cocrystals with biuret.

To test this hypothesis, we mutated the glutamine residue in the biuret hydrolase from Herbaspirillum sp. strain BH-1 (Gln212) that corresponds to Glu221 of GuuH and replaced the residue with a glutamate residue. The resulting enzyme, BiuH Q212E, showed nearly 2 orders of magnitude diminished activity with biuret compared to that of the wild type, and guanylurea hydrolysis became the dominating function of the enzyme, with 40% higher activity with guanylurea than with biuret (Supp. Table A.3). The wild-type biuret hydrolase showed no detectable activity with guanylurea (Supp. Table A.3). This result demonstrated that a single point mutation can convert a biuret hydrolase to an enzyme significantly active with guanylurea. Triuret hydrolases have a larger active site, and the glutamine residue is significantly further away from the catalytic cysteine than in GuuH and biuret hydrolases[91]. These different properties also allow discrimination between the closely related GuuH enzymes, biuret hydrolases, and triuret hydrolases (Figure 3.5).

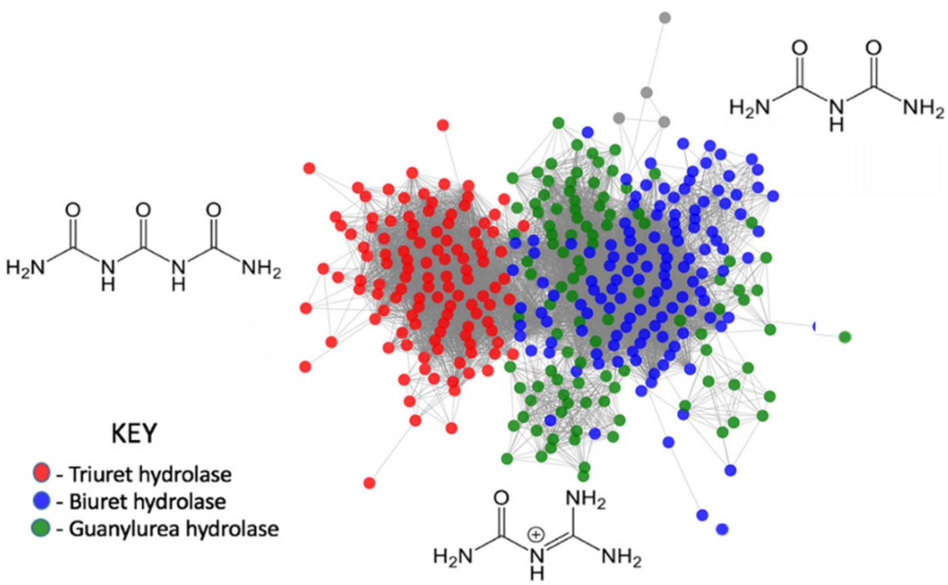


Figure 3.5: Sequence similarity network (SSN) of 3,700 sequences of IHL proteins determined or predicted to degrade triuret, biuret, or guanylurea

There are >1,900 sequences ranging in identity from 45% to 100%. SSN built using EFI-EST with a cutoff of E265 and visualized with Cytoscape.

The defining signatures that identified biuret, triuret, and guanylurea hydrolases (Figure 3.4) were used with sequence similarity network (SSN) analysis to look broadly at the interrelationships of the three homologous enzymes with closely aligning sequences pulled from GenBank. The SSN analysis showed the GuuH enzymes to be more closely related to the biuret hydrolases, consistent with the similar sizes of the substrates and their ready interconversion (Figure 3.5). The observation that there were several separated lobes of GuuH sequences in the SSN may suggest that different sequence clusters of GuuH enzymes arose independently from biuret hydrolases, consistent with our observation of a single point mutation changing activities. Guanylurea may largely derive from metformin and the fertilizer additive cyanoguanidine, both which are only recently in the environment, perhaps suggesting a recent evolutionary origin of at least some GuuHs.

2.4 Discussion

There is increasing focus on the occurrence of human pharmaceuticals in surface and wastewaters, and metformin is often at or near the top of the list with respect to the number of source detections and levels[92]. Metformin has a moderate half-life of 5 to 10 days in river simulation studies, with variation related to microbial diversity[93]. However, because of metformin's extremely high usage, with 70 million prescriptions annually and gram quantities per daily dose, the compound is continuously introduced into surface waters and wastewater treatment plants. For many anthropogenic chemicals in the environment, levels of stable microbial degradation products often exceed those of the parent compound, and this is almost invariably the case with metformin[94].

Guanylurea has long been identified as a relatively persistent and high-level transformation product of metformin[23,61]. The persistent high levels of guanylurea have led to it being labeled as a "dead-end" product of metformin. While there have been reports of guanylurea biodegradation[23,25,61], this report identifies specific genes, enzymes, and a metabolic pathway mediating its metabolism. In the present study, we isolated a bacterium from a wastewater treatment plant that degrades guanylurea to meet its nitrogen needs, sequenced the genome, identified and purified a new guanylurea-hydrolyzing enzyme, elucidated a metabolic pathway for guanylurea mineralization, and identified the evolutionary origins and connectivity of the guanylurea hydrolase to homologous enzymes in the IHL protein family, a part of the cysteine hydrolase superfamily.

As of 9 November 2020, in recorded genomes from NCBI and EMBL databases, GuuH is encoded in less than 0.5% of genomes. The potential rare occurrence of GuuH in the environment and the gene regulation in guanylurea metabolism may play a role in the accumulation of guanylurea in certain environments. Guanidine metabolic genes encoding guanidine carboxylase and carboxyguanidine deiminase can be turned on by a guanidine-sensing riboswitch, which is present in *P. mendocina* strain GU, but there has been no insight into a suppression mechanism of these genes in the presence of guanidine that could be coordinated by broader nitrogen or carbon metabolism, for example[37]. The entire regulation of guanylurea metabolic genes, which are encoded in three separate regions of the genome, is complex and will be examined in detail in a future study. The results presented here have utility in at least two distinct areas. The first is in environmental studies, specifically, where the fates of metformin, cyanoguanidine[95], and guanylurea are being examined. The results here present a roadmap for identifying specific bacteria, genes, and metabolic enzymes that would be present in a particular environment of interest. For example, wastewater metagenomic data[96] may now be culled to look for the presence of these genetic capacities. The second outgrowth of the data would be in medical studies. Metformin prescription rates are expected to rise given its efficacy for treating type 2 diabetes, obesity, and now, perhaps, COVID-19[57]. Metformin is not known to be metabolized by human drug-metabolizing enzymes, but it is known to impact the bacteria in human intestines[97]. Indeed, several studies indicate that the human gut microbiome plays a significant role in modulating metformin's therapeutic effects, but the mechanisms of that modulation are currently obscure[98]. In light of the overlapping environmental and medical implications, the metabolism of metformin, which largely appears to occur via guanylurea, has heightened importance. It is intriguing to consider that new GuuH enzymes may be arising in recent evolutionary time to handle new anthropogenic chemical inputs from the expanding usage of metformin. Biuret and triuret hydrolases appear to be ancient enzymes that have diverged with the taxonomy of the bacteria that harbor them[91,99]. Triuret hydrolases have diverged significantly from biuret hydrolases, a phenomenon clearly indicated by the visual separation on the sequence similarity network shown in Figure 3.5. In contrast, the enzymes identified by sequence signatures as GuuH are interspersed and emerging from several regions of the biuret hydrolase enzyme cluster, as shown in Figure 3.5. One explanation for this would be that GuuH activities have arisen multiple times from biuret hydrolase precursors. Given that a single mutation can impart significant guanylurea hydrolase activity upon a biuret hydrolase, it is plausible that recent evolution is occurring in different global waters as metformin input increases.

There are other precedents for recent evolution of biodegradative enzymes in response to new anthropogenic environmental inputs. For example, trans-3-chloroacrylate dehalogenase from *Pseudomonas cichorii* 170 that degrades the nematicide 1,3-dichloropropene was thought to have arisen in recent evolutionary times from 4-oxalocrotonate tautomerase[100,101]. A similar observation has been made with anthropogenic s-triazine compounds, such as the herbicide atrazine, in which atrazine chlorohydrolases arose independently from divergent members of the amidohydrolase protein superfamily[102,103]. Moreover, the fate of anthropogenic chemicals and their metabolites in the environment has been observed to change in recent times as the result of microbial enzyme evolution[104]. Given that GuuH enzymes can readily arise from a biuret hydrolase via simple mutation(s), and the increasing environmental prevalence of compounds giving rise to guanylurea, we expect that guanylurea will increasingly lose its designation as a “dead-end” metabolite.

Chapter 4

Di-nickel enzyme evolved to metabolize the pharmaceutical metformin: Impacts for wastewater and human microbiomes

Chapter prepared for submission to Proceedings of the National Academy of Sciences

4.1 Introduction

Metformin is the first-line therapeutic drug for type-II diabetes and one of the most prescribed drugs in the world with over 250 million daily prescriptions. The large daily dose of 1-2 grams, when ingested by patients, is not metabolized in the liver, and is excreted in the urine and enters wastewater. This amounts to more than 100 million kg of metformin entering the environment each year, and is nearly as pervasive as caffeine in global waters [59,105]. In addition, metformin has been identified, more recently, to have anti-obesity, anti-aging and anti-tumor properties and the future use of this drug could be extended beyond type-II diabetes [2]. The mechanism of action for metformin has been under intense scrutiny for the past couple of decades but a direct mode of action is still unclear. Recent studies point to metformin's interaction with human gut microbes for exerting therapeutic effects [3–5,7]. A study involving intravenous administration of metformin to type-II diabetes patients, that mainly bypasses the human gut, showed little to no efficacy [6].

Human gut microbes are important in regulating human health and they are also known to modulate drug potency and side effects by gut drug metabolism in several cases [16–18]. Studies that track the disposition of metformin show complete recovery of metformin after intravenous administration but incomplete recovery (~80%) after oral dosage [11,21]. This possibly may be due to partial, gut-drug metabolism although a meta-study concluded that metformin metabolism in humans is still equivocal [10]. There is evidence for metformin metabolism to guanylurea in rats [22]. Identifying genes and enzymes that mediate metformin metabolism could give insight on what metabolites are possibly generated and the genes involved in the gut.

Microbial metabolism of metformin is known in wastewater treatment plants (WWTPs) with some reports finding significant transformation (>80%) [23–25]. The transformation is known to occur due to microbes in the activated sludge of WWTPs with guanylurea found to be the main transformation product [23,68,106]. Chronic exposure of environmentally relevant concentrations of metformin and guanylurea is known to be toxic to zebrafish, affecting reproduction, larvae survival and neurobehavior [107,108]. Other water treatment methods to remove metformin have proven to be ineffective as metformin has poor affinity with activated carbon and chlorination of metformin creates N-chloro species that have been shown to be toxic to human cells [27–29]. Thus, biodegradation of metformin by activated sludge microbes is seen as a viable strategy for its removal, although the enzyme initiating its metabolism had previously not been studied.

In recent years, several microbes have been isolated from activated sludge that can utilize metformin as a carbon or nitrogen source for growth and their genomes have been sequenced [26,109–112]. *Aminobacter* isolates have been shown to break down metformin, utilizing the dimethylamine fragment and exporting guanylurea via a drug transporter [109,110]. We reported on the isolation of a *Pseudomonas* species that could utilize guanylurea as the sole nitrogen source and discovered the guanlyurea hydrolase (GuuH) that deaminates guanylurea to form guanidine [113]. The same species also possessed the genes to break down guanidine completely via guanidine carboxylase, carboxyguanidine deiminase (CgdAB) and allophanate hydrolase [36]. More recently, we reported on another *Pseudomonas* species isolate that could mineralize metformin completely and revealed, by comparative genomics with metformin-degrading *Aminobacter* genomes, a small set of identical genes that were shared [109]. In this small set of genes, two distinctively encoded proteins homologous to arginase and agmatinase, substrates that distantly resemble the structure of metformin.

Here we report that these two genes, *mfmA* and *mfmB*, encode a metal-dependent metformin hydrolase (MfmAB) and characterize its kinetics, substrate specificity, X-ray structure and find the genes to be widespread globally in wastewater treatment plants and perhaps in human gut microbiomes.

4.2 Methods

Cloning, expression and purification of MfmAB

The metforminase genes, *mfmA* and *mfm*B from *Pseudomonas mendocina* sp. MET-2 (NCBI accessions WP_254300333.1 and WP_254300332.1, respectively) were codon-optimized and cloned into *Escherichia coli* BL21 DE3 cells (New England Biolabs) using a pETDuet vector derivative with kanamycin resistance. The *mfmA* gene was cloned with a C-terminal Tobacco Etch Virus (TEV) protease cleavage site followed by a 6x His-tag and inserted, by Gibson assembly, into the first multiple cloning site (MCS1) using the NcoI and HindIII restriction sites. The *mfmB* gene was inserted into the second multiple cloning site (MCS2) using the NdeI and XhoI restriction sites. Site-directed mutants were made using the Q5 Site Directed Mutagenesis Kit from New England Biolabs. The *mfmAB* genes were expressed by growing cells in terrific broth (TB) medium supplemented with 0.5 mM $NiSO_4$ and 50 μg/mL kanamycin at 37°C and 200 rpm to an OD_{600} of 0.6 in a shake flask. The culture was cooled to 16°C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and, with the same agitation, incubated for 20 hours. Cell pellets were harvested by centrifugation at 1,500 x g for 20 min and then resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 10 mM beta-mercapatoethanol pH 7.4). The cells were lysed using a French Press with three passes at 10,000 psi and the lysate then clarified by centrifugation at 20,000 x g for one hour. Metforminase was purified from the lysate by using fast protein liquid chromatography (FPLC) and immobilized metal affinity chromatography (IMAC). Using a GE-AKTA FPLC and a GE HisTrap 5 mL column, MfmAB was purified after running an imidazole gradient from 50 mM to 500 mM and fractions collected. The expression yield for MfmAB was 16 mg per liter culture. Pooled fractions from the FPLC were buffer exchanged into storage buffer (20 mM HEPES-NaOH, 200 mM NaCl pH 8) using a 15-mL Amicon 10 kDa centrifugal filter. The His-tag of purified MfmAB was cleaved by adding 1 mg TEV protease in a dilute protein solution between 5-10 mg/mL MfmAB and the cleavage reaction was placed on a rotator at 4°C overnight. The reaction was concentrated to 2 mL and cleaved MfmAB was purified by size exclusion chromatography using the AKTA FPLC and a GE Healthcare HiLoad 16/600 Superdex 200 pg column. The column was equilibrated with storage buffer, the sample was injected onto the column, and washed with 1 column volume at 1 mL per min flow rate. MfmAB eluted as a heterotrimer with an apparent molecular weight of ~120 kDa and the preparation was then subsequently used in crystallization experiments (data not shown). For metal reconstitution experiments, MfmAB enzyme was stripped of metal by incubating the enzyme with 1 mM 1,10-phenanthroline, 2.5 mM EDTA in storage buffer for 1 hour before buffer exchanging the stripped enzyme into storage buffer using size exclusion chromatography as described before.

Enzyme activation and kinetics

Rates of substrate hydrolysis by MfmAB were determined by a spectrophotometric, coupled-enzyme assay or by an HPLC method. Prior to enzyme kinetic assays, MfmAB was pre-incubated with 1 mM $NiCl_2$ in storage buffer on ice for two hours to allow for complete activation of the enzyme. The coupled-enzyme assay included coupling of enzymes guanylurea hydrolase (GuuH) from *Pseudomonas mendocina* sp. MET-1 and bovine liver l-Glutamic dehydrogenase (GDH) with MfmAB to measure NADH oxidation that was proportional to guanylurea release from MfmAB (Supp. Fig. A.6). Guanylurea hydrolysis, via GuuH, generates guanidinium and ammonia which the latter can be used in reductive amination of 2-oxoglutarate to form l-glutamate, causing NADH oxidation that can be measured by absorbance at 340 nm. For reactions, a 10X coupled enzyme assay master mix was prepared in 50mM HEPES-NaOH pH 8 which had the following components and final concentrations: 0.3 mM NADH disodium salt (Sigma), 5 mM 2-oxoglutarate sodium salt (Aldrich), 0.8 mM adenosine diphosphate sodium salt (Sigma), 2.5 U/mL GDH from lyophilized powder (Sigma) and 0.3 mg/mL purified GuuH. Methods for expression and purification of GuuH are detailed by Martinez-Vaz et al [109]. The master mix was then diluted with buffer and purified MfmAB enzyme into wells of 96-well flat-bottom microplates and the reaction was initiated by adding substrate to make a total sample volume of 200 μL. The reactions were monitored, continuously, by absorbance at 340 nm using an Agilent BioTek Synergy HTX microplate reader and initial rates were recorded. Rates of NADH oxidation were calculated using the molar extinction coefficient for NADH, at 340 nm, of 6220 $M^{-1}cm^{-1}$ and the pathlength of the assay volume (200 μL) in the microplates, 0.56 cm. MfmAB enzyme concentrations used in the assay were between 0.01 μg/mL and 12 μg/mL to measure substrate hydrolysis with the rate being linearly dependent to enzyme concentration in this range (Supp. Fig. A.6). Negative controls for the assays included no-enzyme and enzyme with no substrate that were used to measure background oxidation of NADH in the assay over time. Glycerol and Tris were found to inhibit MfmAB and were not used in enzyme kinetic assays. Guanylurea release was determined to be stoichiometric to NADH oxidation and this was validated by an HPLC method that could separate and quantify metformin and guanylurea (Supp. Fig. A.6).

A reversed-phase HPLC method to separate guanylurea from metformin was adapted from Lin et al. and detailed by Martinez-Vaz et al. which, in brief, used a C18 column and an isocratic mobile phase of 75:25 (v/v) acetonitrile:10 mM potassium phosphate buffer pH 6.6 [109,114]. Kinetics were determined by incubating MfmAB with metformin in buffer and aliquots were quenched with 75% (v/v) acetonitrile to take fixed time point measurements of the reaction. Quenched samples were injected onto the HPLC and using standard curves, the concentrations of metformin and guanylurea were determined, based on peak area at their respective λ_{max}, 234 nm and 220 nm, respectively. Rates of metformin consumption and guanylurea formation were determined using this HPLC method and were similar to the coupled-enzyme assay (Supp. Fig. A.6). One unit of activity (U) was defined as one micromole substrate per minute at the enzyme's pH optimum at 25^{o}C.

Substrates and NMR spectroscopy

Metformin hydrochloride (Cayman Chemical), guanylurea phosphate (TCI Chemicals), dimethylamine hydrochloride (Acros), buformin hydrochloride (Enamine), phenformin hydrochloride (Cayman Chemical), biguanide hydrochloride (Synthonix), 1-methylbiguanide sulfate (Alfa), 1,1-dimethylguanidine hydrochloride (Aldrich), 1-methylguanidine hydrochloride (Aldrich), L-arginine hydrochloride (Acros), agmatine sulfate (Fluka), 4-guanidinobutyric acid (Fluka) and creatine hydrate (Sigma) were obtained with high purity (>97%). ^{1}H-NMR experiments were conducted in water with 20% D_2O using the Varian Unity Inova 400 MHz NMR system and VnmrQ 2.2 software. Sodium 3-(trimethylsilyl)-propionate-2,2,3,3-D4 salt (Cambridge Isotope Laboratories) was used as a reference standard.

Crystallization of MfmAB and structure elucidation

Initial crystallization conditions were found from the sparse matrix screen PACT Premier HT (Molecular Dimensions) using sitting drops containing 14 mg/mL MfmAB and reservoir solutions. Crystals grew from the following condition, 0.2 M $NaNO_3$, 0.1 M bis-tris propane, 20% (w/v) PEG3350. Optimization of MfmAB crystals was done by vapor diffusion in 24-well hanging drop crystallization plates. Crystal growth was sensitive to changes to the relative humidity and to prevent condensation, the air was purged from the wells with compressed nitrogen prior to sealing each well. Crystals grew in a range of conditions at 18°C with 0.2 M $NaNO_3$ between 12-16% (w/v) PEG 3350 and 0.1 M bis-tris propane pH 7.5-8.5 in drops of 1 μL of protein (5-10 mg/mL) with 1 or 2 μL of precipitant. Crystals appeared after one day and were harvested by looping them into cryoprotectant (mother solution containing 25% (v/v) ethylene glycol) and frozen in liquid nitrogen.

Diffraction data was collected using the Advanced Photon Source (Argonne, Illinois, USA) with various beamlines (Supp. Table A.6). Data was processed using XDS (Build January 26, 2018) and molecular replacement, refinement was done using Molrep and Refmac within CCP4 (Version 7.0) and Coot (v0.8.9)[115–119]. For molecular replacement, two separate models from the AlphaFold database were used (both 93% seq. id.), A0A2S0XPN7 and A0A316GGX0, to model MfmA and MfmB, respectively[120]. The molecular replacement of the MfmAB complex proceeded by first placing monomers of MfmB into the asymmetric unit followed by deleting the monomers that poorly fit the electron density map upon inspection. The resulting solution was used as a fixed model to place monomers of MfmA, by molecular replacement, which was again inspected and, after a few more iterations, produced the final model that contained multiple copies of the MfmAB complex.

Computational modelling and Bioinformatics

Docking metformin into the active site of MfmA was done using AutoDock Vina (Version 1.2.5) and ligand restraints were obtained from the ZINC20 database (ZINC12859773, https://zinc.docking.org/)[121,122]. The protein receptor was prepared using AutoDockTools4 (Version 4.2.6) with polar hydrogens added, using default charges for standard residues and the partial charges for nickel ions were set to 0.660 according to Sindhikara *et al*[123,124]. The docking was done using the AutoDock4 forcefield with default parameters except the exhaustiveness was set to 100 and the ligand guanidinium torsions were set to rotatable. The top 10 binding modes shared an AutoDock score of −5.3 and the third-best binding mode is shown in Figures 3D and 3E. Homologous MfmA and MfmB sequences were mined from the NCBI and EMBL databases by first generating a sequence similarity network (SSN) using the EFI-EST tool that performed pairwise BLAST comparisons on 10,000 related sequences[44]. Cytoscape was used to visualize the clustering in the SSN and identify the clusters containing MfmA and MfmB sequences[45].

Metal content analysis of MfmAB using ICP-OES

A sample of 10 mg of MfmAB was buffer exchanged into 20 mM HEPES-NaOH pH 8 using a centrifugal filter and digested overnight with 5% (v/v) nitric acid (trace metals grade, Sigma). The digested sample was heated to 95°C for one hour and then the sample was clarified by centrifugation and the supernatant was collected and sent for analysis. A blank sample was prepared by using the flow-through from the centrifugal filter after buffer exchange and following the same procedure. Samples were sent to the Research Analytical Laboratory at the University of Minnesota for analysis using an iCap 7600 Duo ICP-OES Analyzer for inductively coupled plasma-optical emission spectroscopy (ICP-OES).

4.3 Results

Discovery of a metformin hydrolase, MfmAB, that hydrolyzes metformin to guanylurea and dimethylamine

The metformin hydrolase genes, *mfmA* and *mfmB*, tested in this study came from *Pseudomonas mendocina* sp. MET-2, isolated from activated sludge of the Metropolitan Wastewater Treatment Plant in Saint Paul, Minnesota, United States [109]. Metformin hydrolase, MfmAB, is a complex of MfmA and MfmB proteins that are homologous to each other (34% sequence identity, seq. id.) and related to proteins from the ureohydrolase superfamily (Figure 4.1). The MfmAB genes appear, putatively, in an operon with putative nickel delivery proteins (HypB, HypA), a putative nickel importer (UreJ), and a putative metformin transporter (CodB) (Figure 4.1A). The MfmAB operon is also present and highly conserved (>97% amino acid seq. id.) in five other *Pseudomonas* and *Aminobacter* species that can grow on metformin and isolated from activated sludge across three continents (Supp. Fig. A.2). The *Pseudomonas* strains have been characterized to utilize metformin as a sole nitrogen source while the *Aminobacter* strains can utilize metformin as a carbon and nitrogen source [109–112]. The heterologous expression of MfmA or MfmB individually in *E. coli* showed no activity on metformin in lysates. Only when the two proteins were co-expressed was activity present, as determined by a decrease in metformin and concomitant appearance of guanylurea via HPLC (Figure 4.1B). In expression studies, MfmA does not produce soluble protein when expressed singly although individually expressed MfmB produces soluble protein (Supp. Fig. A.3A). Repeated attempts to obtain soluble MfmA protein were not successful. Thus, the strategy developed to purify the MfmAB complex was fusing MfmA to a C-terminal 6xHis-tag, allowing the copurification of MfmB with MfmA. Two bands of protein were seen on a denaturing polyacrylamide gel at ~41 kDa and 38 kDa corresponding to His-tagged MfmA and MfmB, respectively (Figure 4.1C). The protein gel band for MfmB was consistently denser than the band for MfmA, suggesting that the stoichiometry of the heterotrimer is 2:1 MfmB:MfmA. By gel filtration, the purified complex elutes with an apparent molecular weight of 179 ± 3 kDa (n=3) compared to a series of standards (Supp. Fig. A.13). This is considerably larger than the calculated mass of the heterotrimer of 117 kDa. Based on the crystallization model (*vide infra*), we conclude that the heterotrimer is the authentic structure in solution. We suspect the complex has a larger hydrodynamic volume than its size would suggest due to two axes being longer than the third. The effect of hydrodynamic volume on apparent molecular weight of proteins has been previously noted [125]. It was demonstrated that MfmAB could transform metformin to guanylurea, suggesting that dimethylamine was the second product of the reaction. Using ^{1}H NMR, the transformation of metformin with MfmAB was monitored by following the chemical shift of the methyl hydrogens on the dimethylamine moiety (Figure 4.1D). This chemical shift moved upfield, from 3.06 to 2.73 ppm, upon adding MfmAB to metformin, the latter chemical shift was identical to a dimethylamine standard (Supp. Fig. A.5).

MfmA and MfmB are both homologous to known enzymes such as arginase, agmatinase and guanidinium hydrolase of the ureohydrolase protein superfamily (InterPro IPR006035) that possess a binuclear metal binding site and catalyze the hydrolysis of guanidinium groups to form urea and an amine (Figure 4.1E) [126,127]. A multiple sequence alignment of the sequences of MfmA, MfmB with arginase and agmatinase shows that the conserved residues that chelate the binuclear metals are present in MfmA but not in MfmB (Figure 4.1E).

MfmAB is a nickel dependent metallohydrolase with exquisite substrate specificity

The pH optimum of MfmAB was sharp between pH 9-9.5 (Figure. 4.2A). The alkaline pH optimum is characteristic of enzymes in the ureohydrolase protein superfamily like arginase and agmatinase as it is proposed that the nucleophile is a metal-bound hydroxide in the hydrolysis reaction [128,129]. MfmAB was not fully active in the absence of certain divalent metals. When MfmAB was stripped of metals and then reconstituted with divalent metals, nickel showed the highest activation, with cobalt and manganese having less activity and iron showing activity comparable to the no metal control (Figure 4.2B). When MfmAB was co-incubated with nickel and copper or zinc, no guanylurea formation was detected by HPLC. Analysis of the metal content of purified MfmAB by inductively coupled plasma-optical emission spectroscopy (ICP-OES) showed sub-stoichiometric amounts of zinc, nickel and iron. As nickel was the most active metal, and putative nickel delivery and transport proteins appear co-expressed with MfmAB, nickel was chosen to activate MfmAB in all subsequent enzyme kinetic experiments (Figure 4.1A).

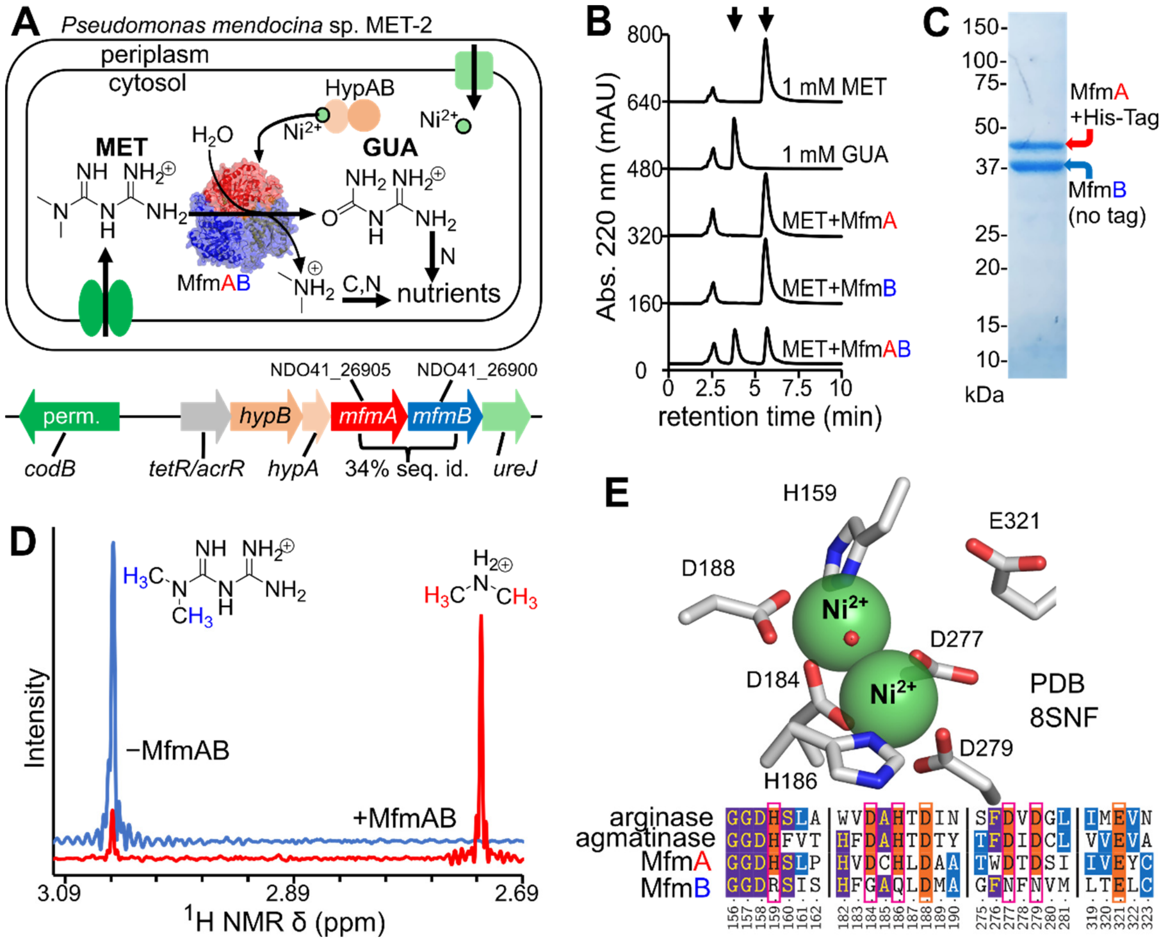


Figure 4.1: Revealing the function of metformin degrading genes in a *Pseudomonas* species.

(A) Metformin-related metabolic genes in *Pseudomonas mendocina* sp. MET-2 are in a putative operon which encode metformin hydrolase (MfmA, MfmB*)*, nickel metabolism proteins (HypB, HypA), a putative nickel importer (UreJ), and a putative metformin transporter (CodB). MfmA and MfmB are both homologs of the ureohydrolase superfamily and share 34% sequence identity to each other. The *Pseudomonas* species has additional genes to completely mineralize guanylurea as a nitrogen source and utilize dimethylamine as a source of carbon or nitrogen. NCBI Genbank identifiers for the *mfmA* and *mfmB* genes are shown above the genes in the putative operon. (B) Activity on metformin was found only in *E. coli* lysates when MfmA and MfmB were co-expressed (MfmAB) and not individually. Lysates were incubated with 1 mM metformin for one hour in 20 mM CHES pH 9 with 1 mM $NiCl_2$ before being sampled by HPLC.

Guanylurea was identified as the reaction product of MfmAB. (C) MfmB copurifies with His-tagged MfmA. Stained, denaturing, polyacrylamide gel with purified MfmAB after IMAC. A band was observed for 6xHis-tagged MfmA and a more intense band was seen for MfmB despite not being tagged. See Supp. Fig. A.3 for the full gel. (D) Identification of dimethylamine as a reaction product of MfmAB by ^{1}H NMR. NMR spectra were obtained for 50 mM metformin in 50 mM ammonium formate, 200 mM NaCl pH 8.5, 1 mM $NiCl_2$ with 20% (v/v) D_2O before (blue spectrum) and after one hour incubation with 200 μg purified MfmAB (red spectrum). The major shift for the methyl hydrogens changed upon MfmAB addition from 3.06 ppm to 2.73 ppm which was identical to the shift found for the dimethylamine standard. See Supp. Fig. A.4 for full NMR spectra. (E) MfmA, MfmB are homologous to arginase and agmatinase with binuclear divalent metal centers, yet MfmB has lost several key metal binding residues. The active site of MfmA is shown with the binuclear metal center and the several histidine and aspartate residues binding it. A multiple sequence alignment showing sequence conservation shared between human arginase I (PDB 2AEB), *E. coli* agmatinase (PDB 7LOL), MfmA and MfmB is depicted. Numbering of amino acids is based on the sequence of MfmA.

To characterize the enzyme kinetics of MfmAB, a spectrophotometric, coupled-enzyme assay included the enzymes guanylurea hydrolase (GuuH) and glutamate dehydrogenase (GDH) in admixture with MfmAB to measure NADH oxidation that was proportional to guanylurea release from MfmAB (Supp. Fig. A.6). Guanylurea is released from the MfmAB reaction with metformin and biguanide analogs and guanylurea hydrolysis, via GuuH, generates guanidinium and ammonia. The latter can be used in reductive amination of 2-oxoglutarate to form l-glutamate, supporting NADH oxidation that can be measured by a decrease in absorbance at 340 nm. MfmAB was found to be very specific to metformin, with no other substrate showing more than 0.6% of the activity observed for metformin (Figure 4.2C). Buformin (1-butylbiguanide) was the second-most active substrate while other biguanides such as phenformin (1-phenylethylbiguanide), 1-methylbiguanide and biguanide were much less reactive. Of the different guanidines tested, dimethylguanidine had similar activity to buformin, with methylguanidine having less activity (Supp. Table A.4).

.

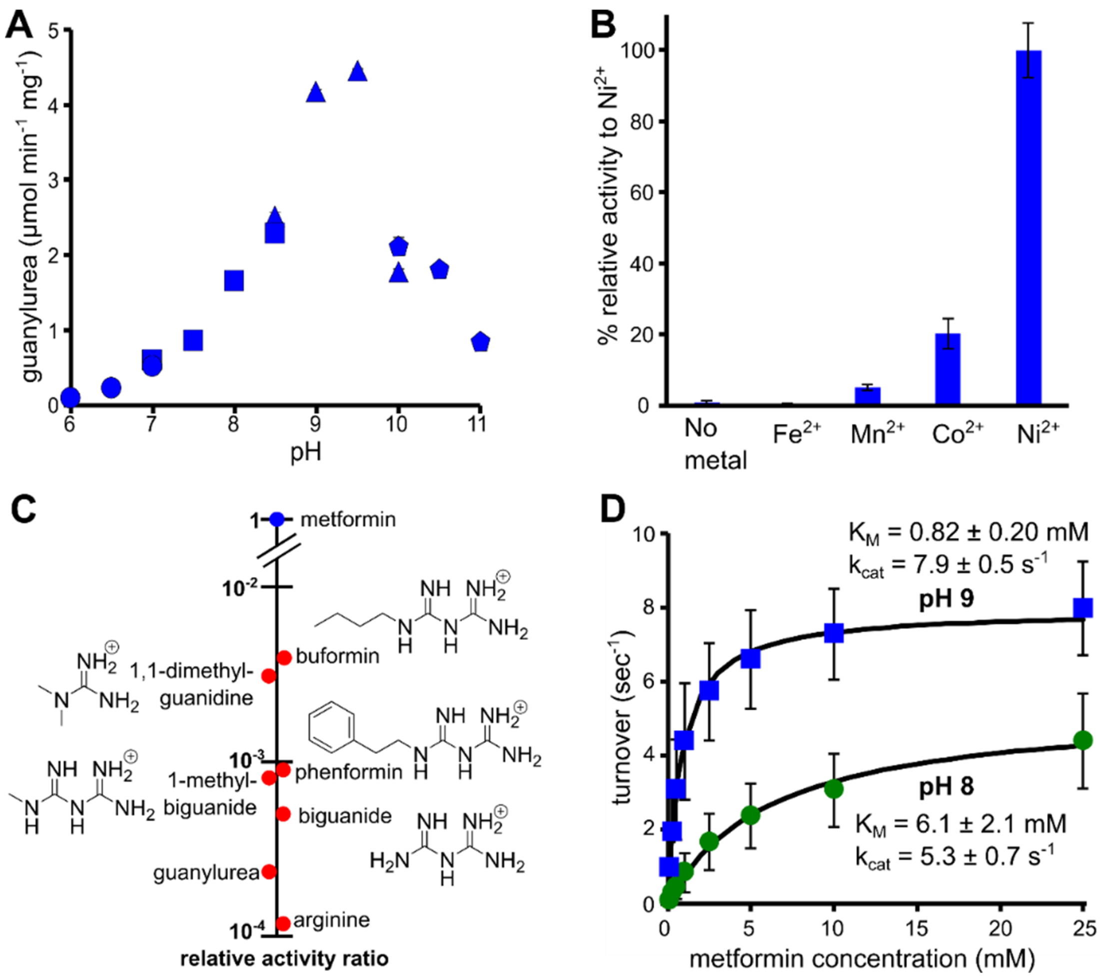


Figure 4.2: MfmAB is Ni^{2+}-dependent and shows exquisite specificity for metformin

(A) pH-Activity dependence of MfmAB. Enzyme was incubated in different pH buffers for 15 min before adding 15 mM metformin and sampling after another 15 min. The pH buffer types used, 50 mM, were PIPES (pH 6-7, shown as circles), HEPES (pH 7-8.5, squares), CHES (pH 8.5-10, triangles) and CAPS (pH 10-11, pentagons). Error bars denote one standard deviation of the mean from averaging two technical replicates. (B) Metal-activity dependence of MfmAB. Enzyme, stripped of metal, was reconstituted with or without 0.1 mM of several divalent metals and activity measured by a coupled-enzyme assay in 50 mM CHES pH 9 with 5 mM metformin. Error bars denote one standard deviation of the mean from two technical replicates. (C) Plot of specificity ratios of MfmAB activity for various substrates relative to metformin. No other substrate tested showed activity more than 1% than that of metformin. Specific activities were measured by incubating purified MfmAB with 50 mM substrate in 100 mM CHES pH 9 buffer using a coupled-enzyme assay with two technical replicates. (D) Michaelis-Menten kinetics of metformin hydrolase at pH 9 (squares) and pH 8 (circles). Activity at several metformin concentrations was measured by observing guanylurea release using a coupled-enzyme assay in either 50 mM CHES pH 9 or 50 mM HEPES pH 8 buffer. Error bars denote one standard deviation of the mean from three biological replicates. Black lines show the fit to the Michaelis-Menten equation.

Michealis-Menten kinetics were determined for the cleaved His-tag MfmAB with a K_M of 0.82 mM and a catalytic efficiency (k_{cat}/K_M) of 9.6×10^3 $M^{-1}.s^{-1}$ at the pH optimum (pH 9). At pH 8, the catalytic efficiency of MfmAB decreased to 8.7×10^2 $M^{-1}s^{-1}$ and the K_M increased to 6.1 mM (Figure 4.2D). The cleavage of the His-tag led to an activity increase of ~25% but the K_M was not affected (Supp. Table A.5). Activity of the enzyme was temperature dependent with thermal inactivation of the enzyme seen at 65°C (Supp. Fig. A.5A). The activation energy for metformin hydrolysis by MfmAB at pH 8 was estimated from an Arrhenius plot to be 52.9 kJ mol^{-1} (Supp. Fig. A.5B). This value was comparable to the activation energy for enzymatic hydrolysis of guanidine catalyzed by GdmH reported to be 56.1 kJ mol^{-1} at pH 8 [126]. Both metformin and guanidine are resonance stabilized molecules and it is estimated that GdmH accelerates guanidine hydrolysis, over the spontaneous rate, on the order of 10^{13}, similar to arginase and agmatinase of the ureohydrolase protein superfamily [126,130]

Crystal structure of MfmAB

Diffraction data for crystals of MfmAB were collected at resolution ranging from 1.85 to 2.30Å. The structure was solved by molecular replacement using, as template, separate AlphaFold models for MfmA and MfmB, respectively (Supp. Table A.6) [126]. MfmAB was found to crystallize with multiple asymmetric unit sets of parameters in space group P1 and in space group C2. The oligomeric state of MfmAB observed in the crystal structures was heterotrimeric with a stoichiometry of 2:1 MfmB:MfmA (Figure 4.3A). Both MfmA and MfmB subunits exhibit the arginase α/β fold characteristic of the ureohydrolase superfamily, with eight parallel β-sheets and alternating α-helices [131,132].

Also characteristic of this protein superfamily is the *cis*-peptide bonds present in MfmA (G156-G157) and MfmB (G135-G136). Similar to guanidine hydrolase (GdmH), MfmA and MfmB have N-terminal loops of approximately 60 and 45 amino acids, respectively, that are not commonly found in the arginase fold and lie at the inter-subunit interfaces of MfmAB (Figure 4.3B) [126]. Residues 16-24 comprise an N-terminal loop of MfmB that was highly disordered and could not be resolved from the electron density maps. When nickel was co-crystallized with MfmAB, anomalous dispersion difference maps at the K-edge for nickel identified its presence at the canonical binuclear metal binding site of MfmA subunits but not of MfmB (Figure 4.3C). The active site of MfmAB resided totally within the MfmA subunits. The MfmB subunit may occlude the entrance to the active site of MfmA but no residues of MfmB come within 10 Å of the metals bound in MfmA. No conformational differences were observed for MfmA and MfmB between the crystal structures, despite the multiple, distinct asymmetric units found.

The canonical binuclear binding site in MfmA is formed by four aspartate residues (D184, D188, D277, D279) and two histidine residues (H159, H186) (Figures 4.1E, 4.3C). The active site also includes a key glutamate residue E321 which is implicated in substrate binding for agmatinase and arginase (Figure 4.3D). The aforementioned residues correspond to critical-activity determining residues that were subject to mutagenesis in arginase and agmatinase homologs in prior studies [129,133,134]. To expand the scope of residues critical for metformin hydrolysis, select residues in the active site, that were shown to be only conserved in close MfmA homologs (>60% seq. identity), were singly substituted and specific activity was measured (Supp. Fig. A.7). These residues were Q81, D188, N200, C201 and W232. All substitutions were deleterious with the variants D188N and N200A having 0.05% and 5.6% of the activity relative to wildtype. The MfmA/D188N variant was crystallized (PDB 8SNK) and it showed occupancy for only one of the two metal binding sites in the active site (Supp. Fig. A.8).

Co-crystallization of MfmAB with metformin and other compounds was unsuccessful in achieving substrate or ligand bound structures. Computational docking of metformin into the MfmA active site was done to model the potential catalytic conformation (Figure 4.3D). A docking model positioned metformin such that the carbon atom bonded to the dimethylamine leaving group is close (2.1 Å) to the critical, attacking water molecule that is bound by the metal ions. In addition, two of the amino groups of metformin are positioned near the two metals ions at less than 2.7 Å. In this hypothetical conformation, residues N200 and E321 could provide hydrogen bonding interactions to the substrate while C201 and W232 may provide van der Waals (VDW) contacts to metformin. The MfmA active site cavity is small and metformin docks snuggly in the cavity (Figure 4.3E). Overlaying the metformin MfmA docking model with the crystal structure of human arginase I (hARG1) bound with a boronic acid analog of the substrate arginine (PDB 2AEB) showed the dimethylamine moiety of metformin aligned with the alkyl chain of the boronic acid analog (Figure 4.3F) [51]. In addition, the metformin guanidinium atoms in the docking model are coordinating the metals similarly to the crystallized boronic acid analog.

Comparing the active sites of MfmA with the homologous enzymes arginase, agmatinase and guanidinium hydrolase showed differences that relate to substrate selectivity. Firstly, the positions of the metals in MfmAB are not aligned precisely with the metals in the hARG1 structure. The metals in *E.coli* agmatinase and GdmH also show significant deviation (Figure 4.3G). Secondly, MfmAB is missing a histidine residue, that is conserved in other ureohydrolase homologs and is implicated in the catalytic mechanism [129,133]. In MfmAB, the corresponding residue is an asparagine (N200) (Figure 4.3D).

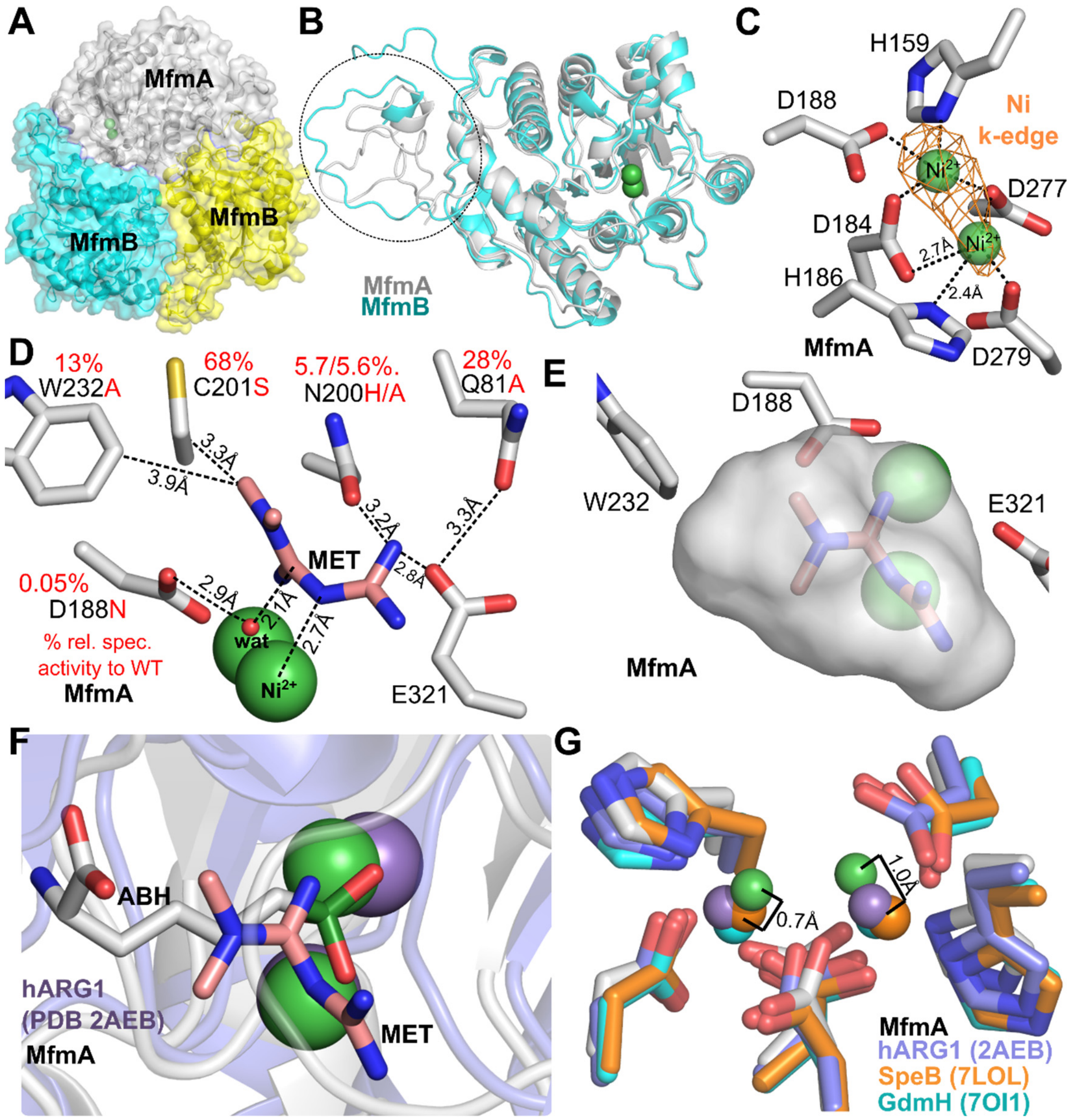


Figure 4.3: Crystal structure of the metforminase (MfmAB) complex

(A) MfmAB is a heterotrimer with one subunit of MfmA per two subunits of MfmB (PDB 8SNF). (B) Overlay of MfmA and MfmB subunits show high structural conservation apart from their N-terminal loops which provide contacts at the intersubunit interfaces of MfmAB (PDB 8SNF). (C) Active site of MfmA showing the anomalous dispersion difference maps contoured at 4σ found from diffraction data collected at the Ni edge, 8347 eV, shown in orange mesh. Data was also collected under the Ni K-edge (8250 eV), but the map (red mesh) shows no density peak at this contour level (PDB

8SNF). (D) Docking model of metformin bound to the active site of MfmA and mutagenesis. Select residues in the MfmA active site were substituted, singly, with the amino acids written in red next to the residue position and the relative specific activity to the WT stated above each residue tested. The residue substitution D188N was critical to activity which is implicated in activating the water that attacks the substrate. The relative activities reported are an average from two technical replicates with percent error of one standard deviation being less than 3%. (E) Cavity of MfmA active site with metformin docked. (F) Overlay of docked metformin in MfmA with human arginase I (hARG1) bound with a boronic acid arginine analog, ABH, (PDB 2AEB). The docking model aligns well the dimethylamine moiety with the boronic acid group. (G) Overlay of metal centers from different ureohydrolase homologs, MfmA (PDB 8SNF). hARG1 (PDB 2AEB), *E. coli* SpeB (PDB 7LOL) and GdmH (PDB 7OI1). The positioning of the metals in MfmA (green), relative to the other homologs in the superfamily may dictate substrate specificity.

4.4 Discussion

MfmAB displays exquisite substrate specificity for metformin, being able to discriminate other close substrates by several orders of magnitude (Figure 4.2C). In addition, MfmAB has a catalytic efficiency of approximately 10^4 $M^{-1}s^{-1}$, an order of magnitude less than the average (10^5 $M^{-1}s^{-1}$) for enzymes active on natural metabolites (Figure 4.2D) [135]. This is remarkable for an enzyme thought to have evolved recently; metformin has only been approved for pharmaceutical use since 1958. Prior to the introduction of metformin, the structurally-analogous biguanides buformin and phenformin were used for treating type-II diabetes but have since been discontinued in many countries due to metformin's superior safety profile [136]. However, in South America and Asia, buformin and phenformin are still being prescribed and wastewater bacteria possessing MfmAB could degrade these pharmaceuticals, albeit more slowly. Microbes able to grow on metformin have been isolated from activated sludge in China, Europe, United States [109–111]. From those genomes, we have determined the presence of high confidence homologs (>97% seq. id) of MfmA and MfmB indicating that metformin hydrolase is widespread globally in wastewater (Supp. Fig. A.2).

Evolution of the MfmAB complex

MfmB appears to be always co-located with MfmA in genomes and the role of the subunit appears to serve as a scaffold to stabilize the fold of the active MfmA subunit (Supp. Fig. A.2). Mining for related sequences of MfmA and MfmB sequences from the NCBI and EMBL databases and clustering the data using a sequence similarity network (SSN), finds only a few dozen closely related sequences (seq. id. >60%) to MfmA and MfmB in recorded databases (Supp. Fig. A.7). The cluster of MfmA sequences appears to be the closest relative of MfmB sequences, which suggests that an ancient gene duplication event of MfmA likely occurred to evolve MfmB (SI Appendix, Fig. S8). A similar example of this type of enzyme evolution is found in guanidine metabolism for the enzyme carboxyguanidine deiminase (CgdAB) [36]. The CgdAB enzyme is comprised of two homologous proteins with one chain conserving the catalytic residues (CgdB) while the other chain (CgdA) lacks the residues but is still necessary for the activity of the enzyme [36]. Examining the crystal structure of the MfmAB complex does not reveal that MfmB provides contacts for substrates bound in the MfmA active site. However, there is a highly disordered loop at the N-terminus of MfmB (residues 16-24) that may play a role in gating of substrates into the MfmA active site (Supp. Fig. A.10). In addition, it may be possible that the loop is involved to bind the HypAB proteins and assist nickel loading into the active site of MfmA.

As the proteins MfmA and MfmB only share 34% sequence identity, it does not suggest that their interaction is a recent evolutionary event in response to metformin entering the environment but the functional precursor to MfmAB is yet unknown. In a study by Chaignaud *et al*, several *Aminobacter* bacterial strains were unable to grow on metformin as the sole carbon source despite their genomes encoding a MfmA homolog that shared ~93% seq. id. with the MfmA reported here (Supp. Fig. A.11) [110]. It appears that fewer than 14 substitutions were necessary to evolve this protein to become active specifically on metformin as shown in the present study (Supp. Fig. A.11A). The substitution positions are distributed globally across the protein although substitutions on one active site loop stands out (Supp. Fig. A.11). MfmA residues 289-291 on an active site loop may provide contacts for metformin, according to the docking model, and at these positions the substitutions from MfmA, to homologs that putatively do not work on metformin, are N289S, S290T and A291S. These changes appear very subtle, but these residues are proposed to be within 3.3-3.5Å with the substrate and could provide van der Waals interactions (Supp. Fig. A.11B). Obtaining a crystal structure of one of these homologs that putatively do not work on metformin may reveal structural differences in the secondary structure of this loop or other parts of the protein that may also dictate substrate specificity. The few differences between MfmA and non-metformin degrading homologs support a recent evolution hypothesis and may be used as a signature to identify true metformin hydrolases.

The recent spread of *mfmA* and *mfmB* genes across the world appears to be facilitated by horizontal gene transfer as the genes found in metformin-degrading *Pseudomonas* or *Aminobacter* species share >97% and >99% sequence identity at the protein and nucleotide level, respectively. In metformin-degrading *Pseudomonas* strains, *mfmAB* genes are located on plasmids while in *Aminobacter* genomes, the genes are either found on the main chromosome or on a plasmid (Supp. Fig. A.2). This suggests that MfmAB first evolved from an *Aminobacter* bacterium and the genes later moved into other *Pseudomonas* and *Aminobacter* strains on plasmids which contain transposon elements and type-IV secretion systems [109].

Implications in human microbiomes

The discovery of the metformin hydrolase enzyme in wastewater may have implications in metformin's interaction with human gut microbiota of type-II diabetes patients. If MfmAB or an homologous enzyme exist in gut microbes than partial metabolism may be occurring and modulating the potency of the drug [17,18]. Another possibility is that gut enzymes, homologous to metformin hydrolase, may be inhibited by metformin which can compete with their native substrates. One homolog of MfmAB that is present in the human gut is agmatinase which acts on agmatine, an intermediate in polyamine metabolism [137]. Interestingly, metformin and agmatine have been linked in a previous mechanistic study of metformin's effect on aging using a host-microbe system with *C. elegans* as the host and *E. coli* as the gut endosymbiont [138]. *E. coli* encodes agmatinase and a deletion of this gene done in the study simulated metformin's positive effect [138]. Testing for metformin activity or inhibition of gut ureohydrolase homologs of metformin hydrolase may be important and merits further investigation.

Catalytic mechanism of MfmAB

The catalytic mechanism of metformin hydrolysis by MfmAB appears similar to other members of the ureohydrolase superfamily in that it has the critical aspartate residue (D188) that is in close proximity to the water bound by the binuclear metals (Figure 4.3D). The role of the asparate residue is to activate the water molecule and bind one of the metals [129]. Substituting this residue with asparagine (D188N) was highly deleterious to specific activity which could support its role in catalysis but it also may affect binding of one of the binuclear metals as evidenced in the crystal structure of the D188N variant (PDB 8SNK), where one of the two metals is missing (Supp. Fig. A.8). In a previous study, the corresponding mutation was done in the *E. coli* agmatinase, D153N, and it was reported that the K_M for the native substrate, agmatine, was unchanged but with a 95% reduction of the specific activity [129].

A histidine residue that is present in the active sites of arginase, agmatinase and guanidinium hydrolase, and is implicated in proton transfer to the amine leaving group on the substrate, is not present in MfmAB [128,133]. In MfmAB the corresponding residue is an asparagine (N200), in which the side chain is not ionizable and likely has a role in substrate binding and/or transition state stabilization but not able to do proton transfer (Figure 4.3D). A cysteine residue, C201, also in the active site, may be ionizable at the pH optimum of the enzyme, pH 9, but substitution of this residue with serine (C201S) did not dramatically affect the activity of the enzyme (Figure 4.3D). The proton transfer to promote the elimination of dimethylamine from metformin may be facilitated by the catalytic aspartate, D188, which may transfer the proton it abstracts from the substrate-attacking water molecule to metformin (Supp. Fig. A.12). The docking model shows that for metformin to bind in a catalytic conformation, it may require that the substrate is contorted and not planar (Figure 4.3E). This conformation could be stabilized by the active site residues (N200, E321, D188) and by the two metals binding the substrate which could serve to reduce the resonance stabilization (Figure 4.3D). The loss of planarity of amides is known to increase their lability in hydrolysis reactions and this may also apply in the metformin hydrolysis catalyzed by MfmAB [139,140].

Metal-dependency of MfmAB

The activity of metformin hydrolase is highest with divalent nickel compared to the other metals tested (Figure 4.2B). Nickel is also the preferred metal for the enzymatic hydrolysis of urea and guanidine, which are both resonance stabilized structures, like metformin, and have higher activation barriers for hydrolysis compared to substituted guanidines, arginine and agmatine, whose enzymes prefer divalent manganese [130]. The greater Lewis acidity of nickel compared to manganese, may be necessary to drive metformin and guanidine hydrolysis, by further activating the metal bound hydroxide that attacks the substrate. In addition, interactions with the substrate and geometry may also play a role in MfmAB's preference for nickel.

The delivery of nickel *in vivo*, into MfmAB, appears to be dependent on the co-expressed nickel-delivery proteins HypA and HypB. This is also likely the case for guanidinium hydrolase (GdmH) that has proteins GhaA, GhaB that are homologous to

HypA, HypB [126]. In the Irving-Williams series of divalent metals forming stable ligand complexes (ex. with proteins), only copper and zinc, which inhibit metformin hydrolase, are more competitive than nickel to form more stable complexes ($Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$) [141]. This property of nickel then necessitates that its delivery must be facilitated *in vivo* to avoid mismetallation. So the likely role of the HypAB chaperones is to selectively deliver nickel to MfmAB *in vivo* where the endogenous concentration of nickel is at low levels compared to metals such as manganese and iron [142–144].

Chapter 5

Discovery of an ultra-specific microbial biguanide hydrolase reveals an alternate metformin biodegradation strategy

5.1 Introduction

Guanylurea is the main transformation product of metformin in wastewater treatment plants (WWTPs) and we have reported on the biodegradative enzyme (MfmAB) that hydrolyzes metformin to guanylurea (See Chapter 4). However there are minor transformation products (TPs) of metformin that are also detected in the effluent of WWTPs which include 1-methylbiguanide, biguanide and triazines 2,4-diamino-1,3,5-triazine (2,4-DAT), 2-amino-4-methylamino-1,3,5-triazine (2,4-AMT) and 4-amino-2-imino-1-methyl-1,2-dihydro-1,3,5-triazine (4,2,1-AIMT) (Figure 5.1)[23,145].

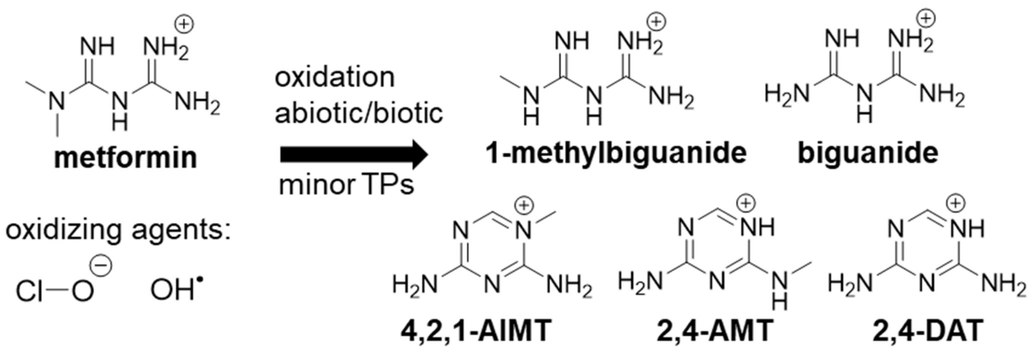


Figure 5.1: Minor transformation products (TPs) of metformin in wastewater treatment

These minor TPs are postulated to be formed by oxidation reactions, whether in the activated sludge by microbes or possibly by reactions with oxidizing agents in the water via an abiotic process. *Ex situ*, chemical reactions with metformin and oxidizing agents like hypochlorite and hydroxy radicals from peroxide and ozone are able to form these minor TPs[145,146]. Chlorination and ozonation are standard practice for WWTPs at the last stage of water treatment to kill microbes and viruses and make it safe for drinking. It is unclear if wastewater microbes are responsible for doing these same transformations or can biodegrade these products once they leave the WWTPs.

The possible pathways are numerous to transform metformin to these minor TPs and necessitates finding the genes responsible and characterizing the enzymes to get the best picture. In Chapter 4, the enzyme responsible for metformin hydrolysis to guanylurea, MfmAB, was found in a *Pseudomonas mendocina* sp. MET-2 that could mineralize metformin but also 1-methylbiguanide and biguanide as the sole nitrogen source[109]. The activity of MfmAB on 1-methylbiguanide and biguanide was four orders of magnitude less than the activity of metformin suggesting another enzyme encoded in the genome may be responsible. Interestingly, the metformin hydrolase genes *mfmA* and *mfmB* are not clustered together with the guanylurea hydrolase gene (*guuH*) in the *Pseudomonas* sp. genome despite being functionally linked. Instead, co-located with *guuH* was a gene, *bguH*, encoding a protein homologous to cytidine deaminase-like protein superfamily which comprises of metal-dependent hydrolases that are known to hydrolyze nucleotide and nucleoside substrates like cytidine and guanosine (Figure 5.2)[52,147]. The hypothesis was then that this *bguH* gene can encode a hydrolase to transform a substrate to guanylurea, most likely being 1-methylbiguanide or biguanide.

Here we report that the *bguH* gene encodes a metal-dependent biguanide hydrolase (BguH) and characterize its kinetics and substrate specificity.

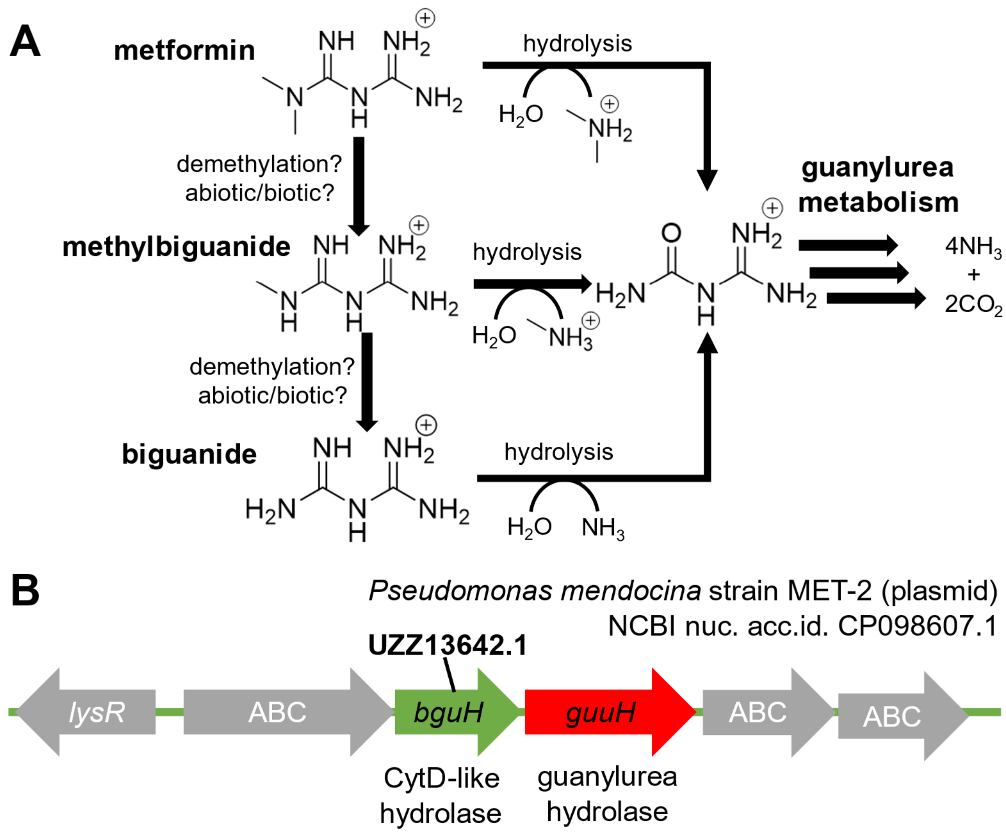


Figure 5.2: Metformin transformation via hydrolytic or demethylation pathways

(A) Potential metabolic pathways for metformin metabolism. Guanylurea is the main transformation product of metformin in WWTP although some reports detect 1-methylbiguanide and biguanide as transformation products. The latter two could be formed from metformin via demethylation by microbes or abiotic factors. Hydrolysis of metformin, 1-methylbiguanide or biguanide can funnel into guanylurea metabolism. (B) Genomic context of the guanylurea hydrolase gene (*guuH*) in a *Pseudomonas mendocina* strain that can grow on metformin, 1-methylbiguanide and biguanide as the sole nitrogen source. Clustered with the guanylurea hydrolase is a gene (*bgu*H) encoding biguanide hydrolase which is homologous to members of the cytidine deaminase-like (CytD-like) protein superfamily.

5.2 Methods

Cloning, expression and purification of BguH

The biguanidase gene, *bguH* from *Pseudomonas mendocina* sp. MET-2 (NCBI accession UZZ13642.1) was codon-optimized and cloned into *Escherichia coli* BL21 DE3 cells (New England Biolabs) using a pET28 vector derivative with kanamycin resistance. The *bguH* gene was cloned with a N-terminal 6x His-tag and inserted, by Gibson assembly, using the NdeI and HindIII restriction sites. The *bguH* genes were expressed by growing cells in terrific broth (TB) medium with 50 µg/mL kanamycin at 37°C and 200 rpm to an OD_{600} of 0.6 in a shake flask. The culture was cooled to 16°C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and, with the same agitation, incubated for 20 hours. Cell pellets were harvested by centrifugation at 1,500 x g for 20 min and then resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 10 mM beta-mercapatoethanol pH 7.4). The cells were lysed using a French Press with three passes at 10,000 psi and the lysate then clarified by centrifugation at 20,000 x g for one hour. Biguanidase was purified from the lysate by using fast protein liquid chromatography (FPLC) and immobilized metal affinity chromatography (IMAC). Using a GE-AKTA FPLC and a GE HisTrap 5 mL column, BguH was purified after running an imidazole gradient from 50 mM to 500 mM and fractions collected. The expression yield for BguH was 10 mg per liter culture. Pooled fractions from the FPLC were buffer exchanged into storage buffer (20 mM HEPES-NaOH, 200 mM NaCl pH 8) using a 15-mL Amicon 10 kDa centrifugal filter. The His-tag of purified BguH was cleaved by adding 1 unit of bovine thrombin protease per mg of BguH in a dilute protein solution between 5-10 mg/mL BguH and the cleavage reaction was placed on a rotator at 4°C overnight. The reaction was concentrated to 2 mL and cleaved BguH was purified by size exclusion chromatography using the AKTA FPLC and a GE Healthcare HiLoad 16/600 Superdex 200 pg column. The column was equilibrated with storage buffer, the sample was injected onto the column, and washed with 1 column volume at 1 mL per min flow rate. BguH eluted as a homodimer with an apparent molecular weight of ~37 kDa and the preparation was then subsequently used in kinetic and crystallization experiments (data not shown).

Enzyme activation and kinetics

Rates of substrate hydrolysis by BguH were determined by a spectrophotometric, coupled-enzyme assay. The coupled-enzyme assay included coupling of enzyme bovine liver L-Glutamic dehydrogenase (GDH) with BguH to measure NADH oxidation that was proportional to ammonia release from BguH. Ammonia release can be used by GDH in reductive amination of 2-oxoglutarate to form L-glutamate, causing NADH oxidation that can be measured by absorbance at 340 nm. For reactions, a 10X coupled enzyme assay master mix was prepared in 50mM HEPES-NaOH pH 8 which had the following components and final concentrations: 0.3 mM NADH disodium salt (Sigma), 5 mM 2-oxoglutarate sodium salt (Aldrich), 0.8 mM adenosine diphosphate sodium salt (Sigma) and 2.5 U/mL GDH from lyophilized powder (Sigma). When testing 1-methylbiguanide or metformin as a substrate, purified guanylurea hydrolase (GuuH, 0.3 mg/mL) was also added to the master mix if ammonia was not directly generated from these substrates with BguH. GuuH hydrolyzes guanylurea to form guanidine and ammonium (See Chapter 3). Methods for expression and purification of GuuH are detailed by Martinez-Vaz et al[109]. The master mix was then diluted with buffer and purified BguH enzyme into wells of 96-well flat-bottom microplates and the reaction was initiated by adding substrate to make a total sample volume of 200 μL. The reactions were monitored, continuously, by absorbance at 340 nm using an Agilent BioTek Synergy HTX microplate reader and initial rates were recorded. Rates of NADH oxidation were calculated assuming a pathlength of 0.56 cm and a molar extinction coefficient for NADH, at 340 nm, of 6220 $M^{-1}cm^{-1}$. BguH enzyme concentrations used in the assay were between 0.01 μg/mL and

10 μg/mL to measure substrate hydrolysis with the rate being linearly dependent to enzyme concentration in this range. Negative controls for the assays included no-enzyme and enzyme with no substrate that resulted in an unchanging amount of NADH in the assay over time. No metal was supplemented in the reactions as incubating BguA with divalent zinc caused precipitation and divalent or trivalent iron inhibited the enzyme. Since numerous homologs of BguH from the cytidine deaminase-like superfamily appear to ubiquitously depend on divalent zinc, BguH was not supplemented with any metal during enzyme kinetic experiments.

HPLC separation and substrates

A reversed-phase HPLC method to separate guanylurea from biguanide was adapted from Lin et al. and detailed by Martinez-Vaz et al. which, in brief, used a C18 column and an isocratic mobile phase of 75:25 (v/v) acetonitrile:10 mM potassium phosphate buffer pH 6.6[109,114]. Biguanide hydrochloride (Synthonix), guanylurea phosphate (TCI Chemicals), metformin hydrochloride (Cayman Chemical), 1-methylbiguanide sulfate (Alfa) and cyanoguanidine (Acros) were obtained with high purity (>97%).

Crystallization of BguH

Initial crystallization conditions were found from the sparse matrix screen JCSG Plus HT (Hampton Research) using sitting drops containing 5 mg/mL BguH and reservoir solutions. Crystals grew from the following conditions, 0.2 M $LiSO_4$, 0.1 M sodium acetate pH 4.5, 30% (w/v) PEG8000; or 2 M ammonium acetate, 0.1 M sodium acetate pH 4.6. Optimization of BguH crystals was done by vapor diffusion in 24-well hanging drop crystallization plates. Crystals grew in a range of conditions at 18°C with 0.2 M lithium sulfate between 11-16% (w/v) PEG 8000 and 0.1 M sodium acetate pH 4.5-6.0 in drops of 1 μL of protein (5-10 mg/mL) with 1 or 2 μL of precipitant and appeared after one day.

5.3 Results

Discovery of a biguanide hydrolase, BguH, that hydrolyzes biguanide to guanylurea and ammonia

The biguanide hydrolase gene, *bguH*, tested in this study came from *Pseudomonas mendocina* sp. MET-2, isolated from activated sludge of the Metropolitan Wastewater Treatment Plant in Saint Paul, Minnesota, United States[109]. The BguH enzyme is related to members of the cytidine deaminase-like protein superfamily that are characterized as metallohydrolases dependent on divalent zinc metal[52,147]. The heterologous expression of BguH in *E. coli* showed no activity on metformin but was active on biguanide in lysates (Figure 5.3). An HPLC method that could separate biguanide from guanlylurea demonstrated that the latter was a reaction product of the BguH reaction.

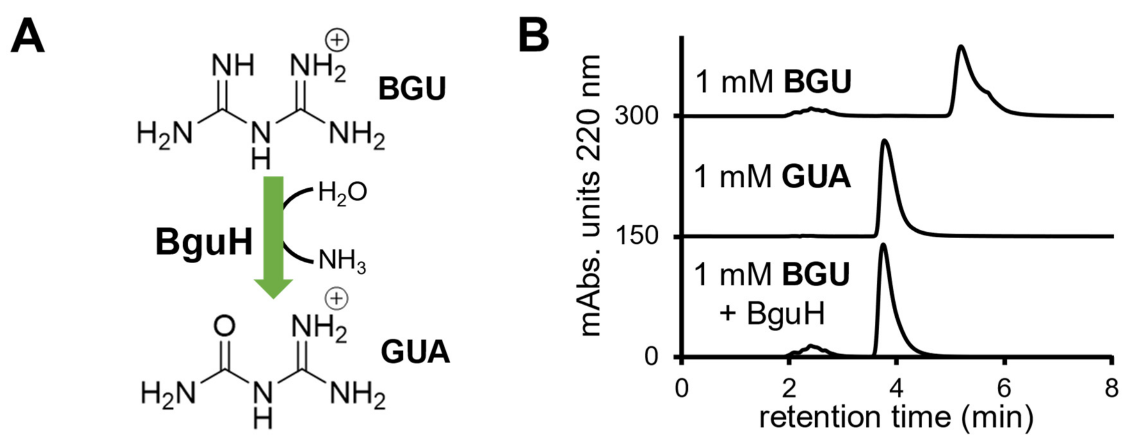


Figure 5.3: BguH is a biguanide hydrolase

(A) Biguanide hydrolase, BguH, hydrolyzes biguanide (BGU) to form ammonia and guanylurea (GUA). (B) HPLC traces of 1 mM biguanide, 1 mM guanylurea and 1 mM biguanide incubated with BguH enzyme. Guanylurea was identified as a reaction product of BguH as determined by a HPLC method that can separate guanylurea from biguanide. *E. coli* lysate with BguH expressed was incubated with 1 mM biguanide for one hour in 20 mM HEPES pH 7.5 before being sampled.

BguH kinetics and exquisite substrate specificity

To characterize the enzyme kinetics, a coupled enzyme assay with glutamate dehydrogenase was developed (Figure 5.4A). The ammonia released by BguH can be consumed in the reductive amination reaction catalyzed by glutamate dehydrogenase with stoichiometric NADH oxidation, that can be measured spectrophotometrically. The pH optimum of BguH, where there is maximal activity, was broad between pH 7-8 (Figure 5.4B). Analysis of the metal content of purified BguH by inductively coupled plasma-optical emission spectroscopy (ICP-OES) showed sub-stoichiometric amounts of zinc and iron. Characterized homologs of the cytidine deaminase-like superfamily appear to ubiquitously depend on divalent zinc for activity[52,147,148].

Michealis-Menten kinetics were determined for BguH with a catalytic efficiency (k_{cat}/K_M) of 4×10^3 $M^{-1}s^{-1}$ at the pH optimum, pH 7.5 (Figure 5.4C). BguH was found to be very specific to biguanide, with no other substrate showing more than 0.2% of the activity observed for biguanide (Table 5.1). 1-methylbiguanide was the second most active while guanylurea, cyanoguanidine being less active and metformin showed no detectable activity. The product of the BguH reaction with 1-methylbiguanide produced guanylurea and methylamine as activity was only detected when coupling guanylurea hydrolase to the coupled-enzyme assay, which liberates an ammonium from guanylurea (See Chapter 3).

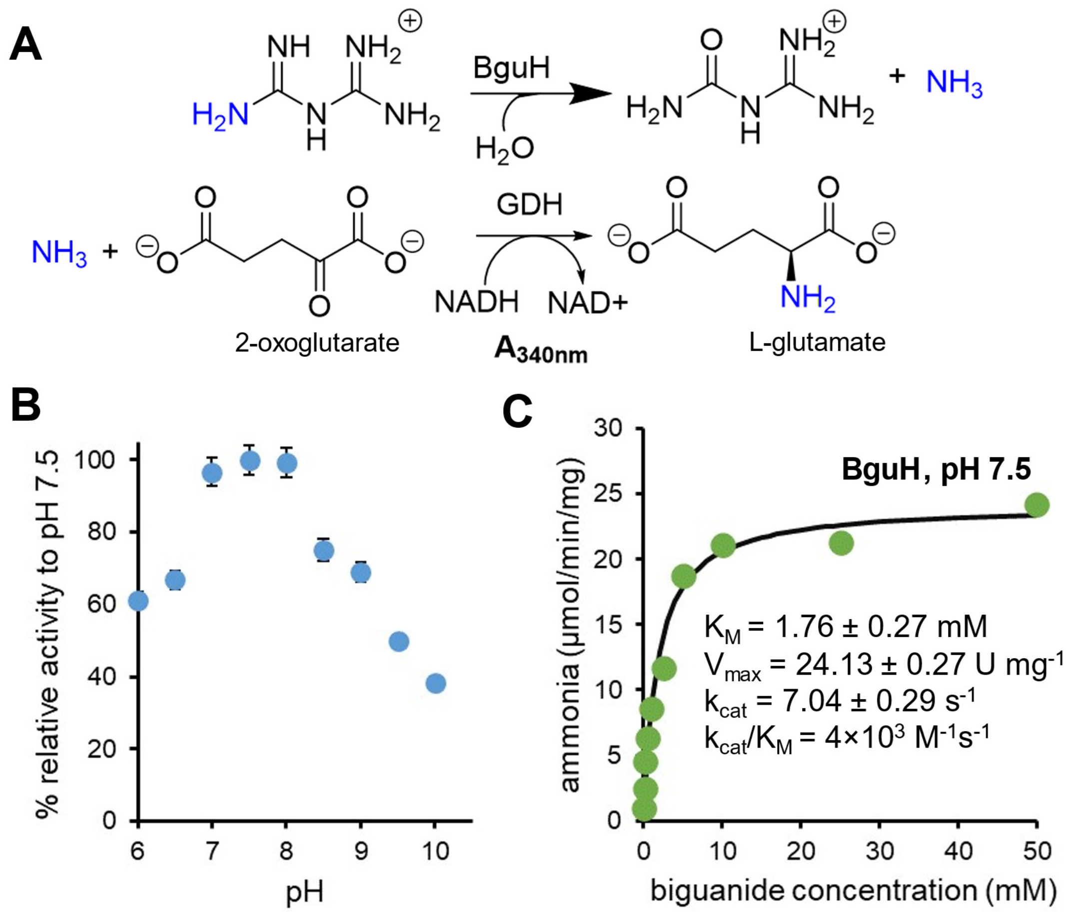


Figure 5.4: Enzyme kinetics of BguH

(A) Coupled-enzyme assay to detect ammonia release. The biguanidase (BguH) reaction releases ammonia when hydrolyzing biguanide. This ammonia can then be used in reductive amination of 2-oxoglutarate oxoglutarate to form glutamate by bovine glutamate dehydrogenase (GDH) which uses NADH as a cofactor. The oxidation of NADH in the GDH reaction can be measured spectrophotometrically by absorbance at 340 nm. (B) pH-Activity dependence of BguH. The pH optimum is broad between pH 7 and pH 8. Enzyme was incubated in different pH buffers for 15 min before adding 20 mM biguanide and sampling after another 15 min. The pH buffer types used, 50 mM, were PIPES (pH 6-7), HEPES (pH 7-8.5) and CHES (8.5-10). (C) Kinetic parameters of BguH. Activity at several biguanide concentrations was measured by observing ammonia release by a coupled-enzyme assay in 100 mM HEPES pH 7.5 buffer. Error bars denote one standard deviation of the mean from averaging two technical replicates. One unit (U) of activity is defined as one µmol ammonia released per min.

Table 5.1: Substrate specificity of BguH

substrate (50 mM)	spec. activity (μmol/min/mg)	specificity ratio
biguanide	24.2 ± 1.01	1
1-methylbiguanide[a]	0.0488 ± 0.0064	2.01×10^{-3}
guanylurea	0.00328 ± 0.00022	1.36×10^{-4}
cyanoguanidine	0.00146 ± 0.00008	6.03×10^{-5}
metformin[a]	n.d	

[a] guanylurea hydrolase added to coupled-enzyme assay; n.d – not detected

Active site of BguH

Modelling the structure of BguH identifies potential enzyme residues that are critical for activity (Figure 5.5). The enzyme conserves the two cysteine residues (C85 and C88) and one histidine residue (H56) that are conserved in the CytD-like superfamily that chelate divalent zinc. The glutamate residue (E58) is also conserved which is implicated in deprotonating the water bound by the metal and transferring the proton to the ammonia leaving group on the substrate (Figure 5.5)[52]. Docking of biguanide into the BguH active site identified residues that may contact the substrate during catalysis. The backbone carbonyls of residues C82 and E83 may stabilize the ammonia leaving group of the substrate and the loop on which these residues lie is held rigid by the proline residue P84 (Figure 5.5). The phenyl ring of residue F84 likely provides VDW contact to the substrate and the aspartate residue may hydrogen bond with biguanide.

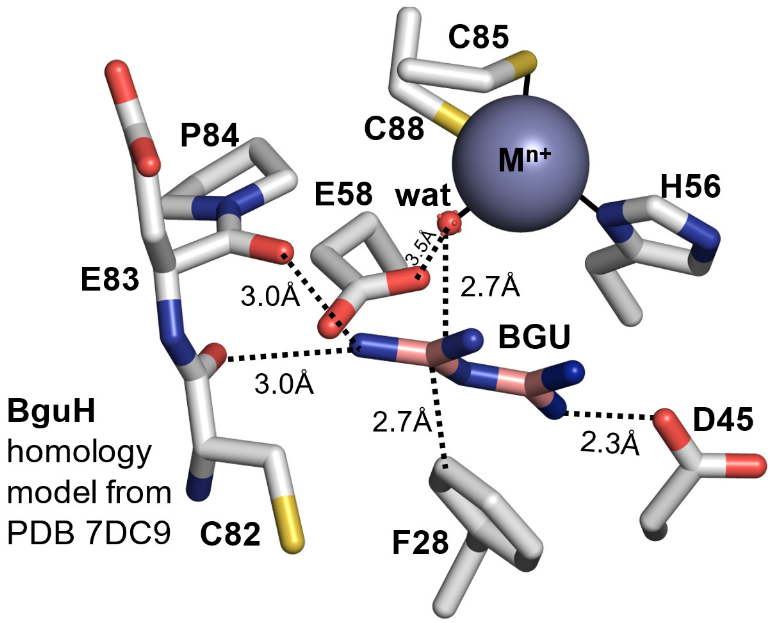


Figure 5.5: Model of BguH active site with biguanide docked

BguH is a metallohydrolase that chelates a metal (M^{n+}) with residues C85, C88 and H56. The metal binds a water molecule (wat) and is activated by residue E58 to attack the substrate. Biguanide is likely to make contact with the sidechains of residues D45 and F28 and the backbone carbonyls of residues E83 and C82. Model of BguH was made by homology model using the structure of guanosine deaminase (PDB 7DC9) as template with 41.2% sequence identity.

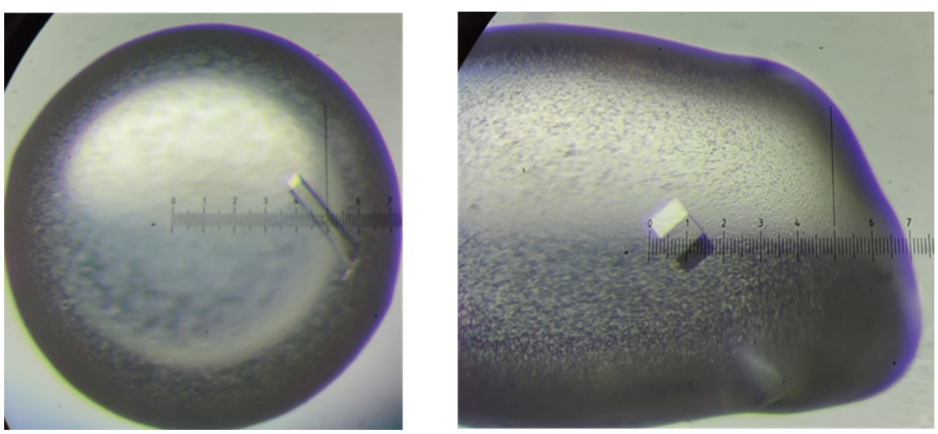


Figure 5.6: Crystals of BguH

Crystallization of BguH

Crystals of BguH were formed from a certain set of conditions to make rod-like crystals with an average diameter of 50 μm (Figure 5.6). The crystals were not subjected to X-ray diffraction experiments so it is not known if these crystals can diffract to determine the structure of BguH.

5.5 Discussion

BguH has a catalytic efficiency of 4×10^3 $M^{-1}s^{-1}$ and displays exquisite specificity for biguanide. Considering the short time of exposure to biguanide in the environment, this is a remarkable feat of recent evolution by microbes to utilize a pervasive pollutant in wastewater. The existence of this function suggests that biguanide and 1-methylbiguanide are seen by microbes in the environment, most likely from demethylation of metformin. BguH activity on 1-methylbiguanide suggests that *Pseudomonas* strain uses this enzyme to metabolize 1-methylbiguanide for growth as a nitrogen source, despite its slow kinetics (Table 5.1)[109]. Gene knockout of the *bguH* gene is still necessary to confirm that this sole gene is responsible for 1-methylbiguanide or biguanide utilization. Comparing BguH and the metformin hydrolase, MfmAB, for specific activity on 1-methylbiguanide, at their respective pH optima, shows that BguH has ten-fold more activity (Table 5.1, Chapter 4). Although since the pH optimum of MfmAB is at pH 9, the specific activity on 1-methylbiguanide at near neutral pH 7.5 is suggested to be 50-fold less (See Chapter 4). This supports the role of BguH to confer the phenotype of growth on 1-methylbiguande and biguanide for microbes.

Genes encoding enzymes that may demethylate metformin or 1-methylbiguanide could not be identified in this study and it is unclear if *Pseudomonas mendocina* sp. MET-2 is able to do this demethylation or by other microbes in the environment. An interesting study would be to quantify the relative expression of MfmAB and BguH in a wastewater treatment plant at the RNA level using quantitative polymerase chain reaction (qPCR) and understand at what stage of water treatment, in the influent, aerobic digestion, anaerobic digestion or effluent where this biodegradation is occurring.

If an enzyme is found to demethylate metformin to 1-methybiguanide in wastewater microbes, it is possible that if the same or homologous enzyme exists in human gut microbes of patients prescribed metformin that there could be drug interaction. The concentration of metformin in the human gut is many orders of magnitude greater than in wastewater and could be transformed by gut microbes to modulate the drug's efficacy[17,18]. Alternatively, the same gut enzymes may have poor activity on metformin but could be competitively inhibited by the drug and disrupt native gut metabolism that is important to the therapeutic effect of the drug or side effects.

Chapter 6

New insights into the action of the pharmaceutical metformin: Targeted inhibition of the gut microbial enzyme agmatinase

Prepared for submission from L.J. Tassoulas, and L.P. Wackett. New insights into the action of the pharmaceutical metformin: Targeted inhibition of the gut microbial enzyme agmatinase. *iScience*, in preparation, 2023.

6.1 Introduction

Metformin is the first-line treatment for type-II diabetes with an unknown mechanism and prescribed to over hundreds of millions of patients each year. Beyond the prescribed usage, metformin has been identified to also have anti-tumor and anti-obesity properties[149,150]. The versatility of the drug to promote healthy outcomes in treatment has prompted intense scrutiny of its mechanism of action in the past two decades. Recent studies identify metformin's interaction with human gut microbes to be responsible for the drug's therapeutic effects[3–5]. In the human gut, the concentration of metformin is estimated to vary between 1 and 10 mM, while the concentration of the drug once absorbed into the portal vein, entering the liver, is reported to be between 10 and 40 μM[5]. The concentration in human tissue is orders of magnitude less than the inhibitory concentrations found for metformin acting on certain human targets. In addition, a study that conducted intravenous administration of metformin to treat type-II diabetes patients, which mainly circumvents the human gut, showed little to no effectiveness[6].

Human gut microbes are known to be key indicators in human health and disease and modulation of the gut microbiome is seen as a new frontier for therapeutic intervention[16]. Targeting the human gut can be problematic due to the variation of the gut microbiome from patient to patient, which may cause variable side effects or potency. Approximately 30% of patients treated with metformin suffer from gastrointestinal symptoms and another 20% of type-II diabetes patients are non-responders to the drug and require alternative medication[19,20]. Elucidating the direct mode of action of metformin in the gut could help reveal key improvements that can improve the therapy and ameliorate side effects. The recommended dose for metformin is 1000-2000 mg daily for patients, which is not metabolized by human enzymes and >80% is excreted into wastewater and is one of the most pervasive pollutants in the world[105].

Recently we discovered the genes and enzymes in microbes responsible for transforming metformin to guanylurea in wastewater treatment plants[151]. This was an effort to determine gene markers that could be used to identify drug metabolism in human gut metagenomes and possibly explain patient variation in drug dosage or side effects. The metformin hydrolase, MfmAB, is homologous to members of the ureohydrolase protein superfamily like agmatinase and arginase which are metal-dependent using binuclear, divalent metals. Searching for these gene markers in human gut metagenomes did not find significant hits that suggested metformin hydrolase was present in human gut microbiota (Figure 6.1A). An alternative hypothesis we conceived was that since metformin hydrolase activity evolved from this family of proteins, there could be sequence-distant gut ureohydrolase homologs of MfmAB that could have slow activity on metformin and possibly be inhibited by metformin via competition with their native substrates.

Here we report that metformin targets gut bacterial agmatinases as a potent, competitive inhibitor that is relevant in the human gut context. We also find that related biguanide analogs buformin, phenformin and galegine also competitively inhibit agmatinase, indicating a shared mechanism. Revealing that metformin targets agmatine metabolism in the gut may have profound implications in type-II diabetes therapy which is discussed in this chapter.

6.2 Methods

Cloning, expression and purification of ureohydrolase homologs

The agmatinase from *Escherichia coli* gene (EcAGM, WP_000105566.1*)* was cloned from the ASKA[152] clone, JW2904, that contained the *speB* gene on an overexpression plasmid which was PCR amplified and cloned into *Escherichia coli* BL21 DE3 cells (New England Biolabs) using a pET28 vector derivative with kanamycin resistance. The gene was cloned with a N-terminal a 6x His-tag and inserted, by Gibson assembly, into the multiple cloning site using the NdeI and HindIII restriction sites. The *Clostridium butyricum* agmatinase (CbAGM, QGH27404.1) was codon-optimized and cloned in the same procedure as EcAGM. The proteins were expressed by growing cells in lysogeny broth (LB) medium supplemented, with 0.5 mM $MnSO_4$, for EcAGM, or 0.5 mM $NiSO_4$l for CbAGM, and 50 µg/mL kanamycin at 37°C and 200 rpm to an OD_{600} of 0.6 in a shake flask. The culture was cooled to 16°C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and, with the same agitation, incubated for 20 hours. Cell pellets were harvested by centrifugation at 1,500 x g for 20 min and stored at −80°C.

For purification, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 10 mM beta-mercapatoethanol pH 7.4).The cells were lysed using a French Press with three passes at 10,000 psi and the lysate then clarified by centrifugation at 20,000 x g for one hour. The proteins were purified from the lysate by using fast protein liquid chromatography (FPLC) and immobilized metal affinity chromatography (IMAC). Using a GE-AKTA FPLC and a GE HisTrap 5 mL column, proteins were purified after running an imidazole gradient from 25 mM to 500 mM and fractions collected. Pooled fractions from the FPLC were buffer exchanged into storage buffer (20 mM HEPES-NaOH, 200 mM NaCl pH 8) using a 15-mL Amicon 10 kDa centrifugal filter. Protein concentrations were determined by Bradford reagent. Aliquots of concentrated protein were flash frozen in liquid nitrogen and stored at −80°C to be then used for kinetic assays. See Appendix A.1, for methods describing the cloning, expression and purification of ARG1, ARG2, AGMAT, GpuA and AgDI proteins.

Substrates and NMR

Metformin hydrochloride (Cayman Chemical), buformin hydrochloride (Enamine), phenformin hydrochloride (Cayman Chemical), L-arginine hydrochloride (Acros), agmatine sulfate (Fluka), 4-guanidinobutyric acid (Fluka) and 3-guanidinopropionic acid (Sigma) were obtained with high purity (>97%).

Galegine hemisulfate was synthesized via the condensation reaction of a thiourea with the alkylamine using a procedure adapted from Williams *et al*[153]. 2-methyl-2-thiopseudourea hemisulfate, 7 mmol, (Sigma) was added to 8 mmol 3-methyl-2-buten-1-amine (A2B Chem) in 5 mL of water and stirred overnight at room temperature. The reaction produced precipitate which was dried on a porcelain plate and washed with methanol. The solids were then recrystallized in methanol to produce galegine hemisulfate in high purity (>95%) as determined by NMR. ^{1}H-NMR (DMSO-d6, 400 MHz) δ (ppm) 1.62 (s, 3H), 1.68 (s, 3H), 3.64 (br, 2H), 5.16 (t, 1H), 7.62 (br, 4H), 8.46 (br, 1H). ^{1}H-NMR experiments were conducted using the Varian Unity Inova 400 MHz NMR system and VnmrQ 2.2 software.

Enzyme kinetics and inhibition

Agmatinase, arginase, guanidinopropionase and guanidinobutyrase activity was measured by tracking urea production using a fixed time-point, colorimetric assay[154]. Enzyme assays were done at room temperature in 100 mM HEPES-NaOH pH 7.5 with 2 mM $MnCl_2$ or 0.1 mM $NiCl_2$ for the Ni^{2+}-dependent CbAGM. Enzyme reactions with 25 μL volume were quenched with 75 μL of the colorimetric acid reagent (120 μM $FeCl_3$, 10 mM phosphoric acid in 20% w/v H_2SO_4) and then to this mixture, 50 μL of the colorimetric color reagent (62 mM 2,3-butanedione monoxime, 3.6 mM thiosemicarbazide in water) was added. The samples were then heated in a PCR thermocycler for 15 minutes at 96°C, in capped tubes, to produce color. Samples were cooled to room temperature and then transferred to 96-well flat-bottom microplates and diluted to 200 μL with deionized water. Using an Agilent BioTek Synergy HTX microplate reader, the absorbance value at 520 nm was measured for all of the samples. Urea standards from 0 μM to 200 μM and no enzyme controls were treated in the same way as the enzyme reactions to quantify urea concentrations and determine background levels.

Inhibition constants (K_i) for were determined for the inhibitors by first obtaining the apparent K_M (K_M^{app}) of agmatine in the presence of inhibitor. The activity of EcAGM and CbAGM was measured at multiple agmatine concentrations (0-50 mM) with or without fixed inhibitor concentrations (0-50 mM) of metformin, phenformin, buformin and galegine. The measured K_M, K_M^{app} for each inhibitor were then used to determine the K_i using equation (1), assuming purely competitive inhibition.

$$K_M^{app} = K_M\left(1 + \frac{[I]}{K_i}\right) \quad (1)$$

Computational modelling and Bioinformatics

Docking metformin, buformin, phenformin and galegine into the active site of EcAGM was done using AutoDock Vina (Version 1.2.5) and ligand restraints were obtained from the ZINC20 database (https://zinc.docking.org/, ZINC000012859773, ZINC000004097425, ZINC000005851063, ZINC000000897460, respectively)[121,122]. The protein receptors were prepared using AutoDockTools4 (Version 4.2.6) with polar hydrogens added, using default charges for standard residues and the partial charges for manganese ions were set to 0.580 according to Neves *et al* [124,155]. The docking was done using the AutoDock4 forcefield with default parameters except the exhaustiveness was set to 100 and the ligand guanidinium torsions were set to rotatable. The AutoDock scores for the best binding mode when docking EcAGM with metformin, buformin, phenformin and galegine were −6.7, −8.6, −4.9, and −6.0, respectively.

Bioinformatics

Ureohydrolase protein superfamily human gut homologs were mined from the Unified Human Gastrointestinal Protein (UHGP) v.2.0.1 catalog (http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/human-gut/v2.0.1/) by first performing a protein BLAST search to retrieve 10,000 sequences that are most similar to the protein sequence of *E. coli* agmatinase (EcAGM)[156]. A sequence similarity network (SSN) was then generated from this set of sequence hits using the EFI-EST tool that performed pairwise BLAST comparisons on the 10,000 related sequences[44]. Cytoscape was used to visualize the clustering in the SSN and identify sequence clusters encoding unique functions[45]. The SSN also included protein sequences of experimentally characterized ureohydrolase enzymes with known function to help annotate the hypothetical function of these clusters. Using this SSN, the cluster of sequences associated with EcAGM and the cluster associated with CbAGM were extracted to identify the taxonomy of the gut microbes that contain an agmatinase related to EcAGM or CbAGM.

To understand the likely relative abundances of EcAGM-like, and CbAGM-like functions in the human gut microbiome, the 16S rRNA profiling data from fecal samples of type-II diabetes patients, taking metformin, as analyzed by Pryor *et al* was used (https://data.mendeley.com/datasets/crmtpmd622/1)[3,8,138,157]. The analysis consisted of mapping 16S rRNA reads from three separate cohort studies (Danish[3], Swedish[157] and Spanish[8]) to 773 human gut microbial genomes used in the metabolic reconstruction tool

AGORA (Assembly of Gut Organisms through Reconstruction and Analysis)[158]. The mapped reads to microbial genomes were normalized to the total amount of reads mapped in each sample to calculate % relative abundances. Microbes with genus and species that matched the organismal taxonomy of the protein sequences pertaining to EcAGM-like, and CbAGM-like functions were put into separate bins.

6.3 Results

Searching for metformin metabolic genes in wastewater and gut metagenomes

The metformin hydrolase genes *mfmA* and *mfmB* encode a nickel-dependent binuclear metalloenzyme that is made from the complex of MfmA of and MfmB proteins with the MfmA subunit containing the active site[151]. Using the MGnify database, which allows for a protein sequence search against hundreds of thousands translated metagenome datasets, found no close matches (>30% seq. id.) to MfmA or MfmB sequences in recorded human gut metagenomes, at the time of writing this dissertation[159]. This inconclusive result led us to develop an alternative hypothesis that, since the metformin hydrolase evolved from the ureohydrolase protein superfamily, there could be very poor activity or possibly inhibition of gut ureohydrolase homologs with metformin.

Agmatinase is an example of a gut microbial ureohydrolase that, in a recent host-microbe study by Pryor *et al*, was linked to metformin action to increase the lifespan of *C. elegans* with *Escherichia coli* (*E. coli*) as an endosymbiont[138]. In the study, deletion of the agmatinase gene in *E. coli* simulated metformin's positive effect and this finding suggested that metformin inhibits agmatinase from *E. coli*, a known gut microbe in humans[138].

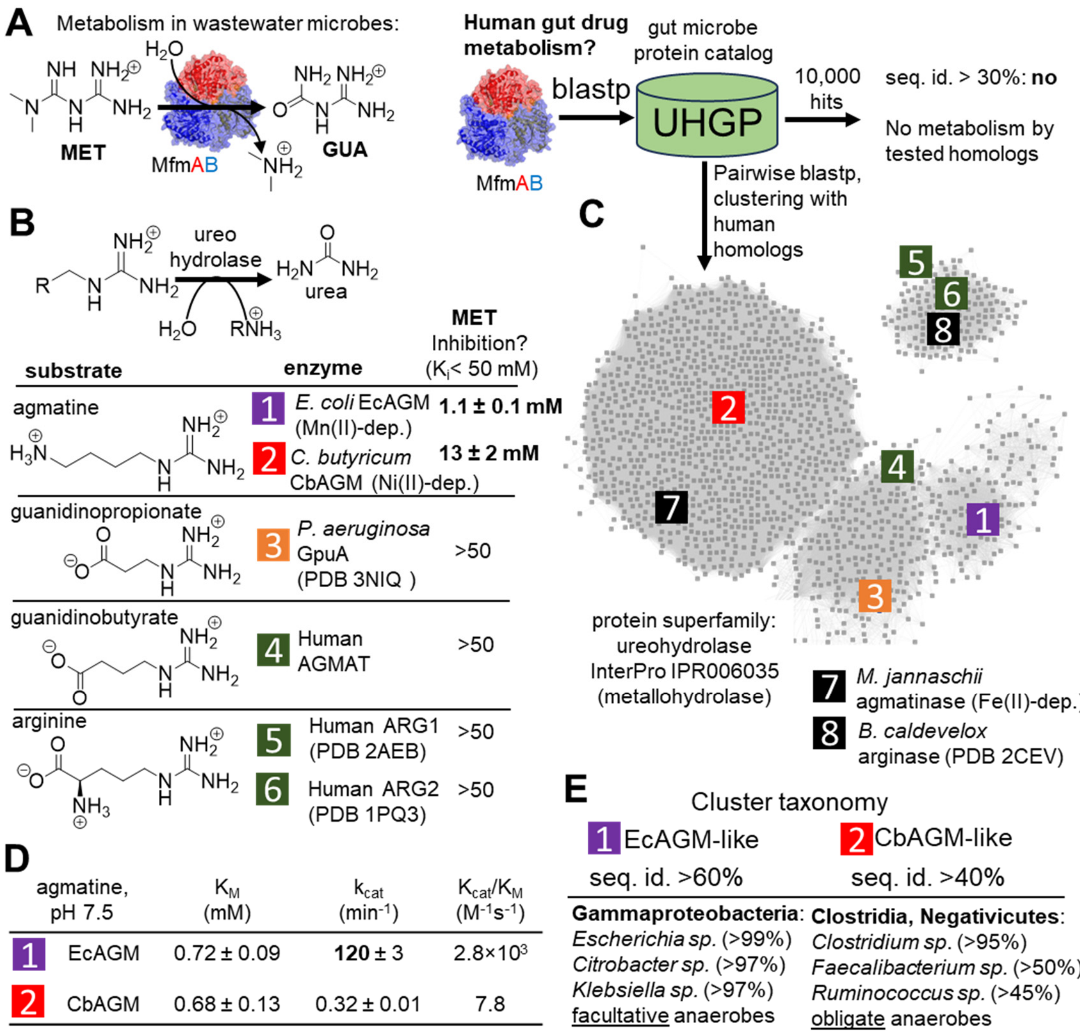


Figure 6.1: Targeted inhibition of gut bacterial agmatinases by metformin (A) Searching for metformin-degrading enzymes in the human gut microbiome. (B) Gut and human homologs of MfmAB tested for inhibition by metformin. Select human and gut microbe homologs were expressed heterologously in *E. coli*, purified and screened for inhibition by metformin in competition with their native substrates. (C) Sequence similarity network (SSN) of gut and human homologs of MfmAB. The E-value cutoff for this SSN was set to 61. (D) Kinetic parameters of EcAGM and CbAGM agmatinases. EcAGM has a superior k_{cat} over CbAGM with an approximately 350 fold difference in catalytic efficiency (k_{cat}/K_M). (E) Taxonomy of EcAGM-like and CbAGM-like sequences. Sequences pertaining to the clusters with CbAGM or EcAGM were extracted from the SSN and taxonomic metadata was evaluated.

Metformin inhibits human gut agmatinases

Metformin is a potent, competitive inhibitor of *E. coli* agmatinase (EcAGM) with an inhibition constant (K_i) of 1 mM at near physiological pH, 7.5 (Figure 1B). This level of inhibition is relevant in the context of the human gut as the concentration of metformin in the gut is estimated to be between 1 and 10 mM[21]. This result prompted us to explore if metformin can inhibit other gut ureohydrolases or possibly even homologs found in human cells.

To discern this, we clustered 10,000 human gut protein sequences from the Unified Human Gastrointestinal Protein catalog (UHGP) that are homologous to EcAGM, along with experimentally characterized ureohydrolases, to create a sequence similarity network (SSN) that can visualize clusters of sequences with unique functions (Figure 6.1C). Annotating the SSN, using the characterized ureohydrolases, the SSN revealed clusters of ureohydrolases, present in the human gut, that likely utilize 3-guanidinopropioniate, 4-guanidinobutyrate, arginine along with agmatine. There are three ureohydrolase homologs in humans which include arginases, ARG1 and ARG2 as well as the enzyme AGMAT that is active on guanidino acids like 4-guanidinobutyrate[51,160,161].

The largest cluster, cluster 2, of gut sequences in the SSN contains a separate subfamily of agmatinase enzymes which have been found to be Ni^{2+}, Co^{2+} or Fe^{2+} dependent and are divergent from the *E. coli* agmatinase which is Mn^{2+} dependent (Figure 6.1C). This was determined when characterizing a sequence from this divergent cluster, from the gut microbe, *Clostridium butyricum* (CbAGM), with 32% sequence identity to EcAGM, which was most activated with the addition of Ni^{2+} but not Mn^{2+} cations (Supp. Fig. A.14). In addition, previously characterized agmatinases from hyperthermophilic archaea, *Pyrococcus horikoshii* and *Methanocaldococcus jannaschii*, were found to be Co^{2+} and Fe^{2+} dependent, respectively, and are also part of the agmatinase subfamily that includes CbAGM (Figure 6.1C, Supp. Table A.7)[162,163].

After assessing the possible functional landscape of ureohydrolases in the human gut, a select set of sequences were screened for inhibition by metformin that could represent most of the sequence space. No metabolism of metformin was found for any gut or human ureohydrolase homolog tested (see Appendix A.1). The screening found no significant inhibition (K_i < 50 mM) by metformin with ureohydrolase function other than agmatinase (Fig. 1B). The competitive inhibition constant of metformin for the Ni^{2+}-dependent CbAGM was 13 mM and was less potent than metformin's effect on EcAGM with a K_i of 1.1 mM. Comparing the kinetic parameters of the two enzymes, EcAGM has more than 350 fold higher k_{cat} but similar K_M compared to CbAGM (Figure 6.1D). The cluster of sequences in the SSN that contains CbAGM, or CbAGM-like sequences, are encoded primarily in Clostridria and Negativicutes microbes (ex. *Clostridium, Faecalibacterium, Ruminococcus*) that are obligate anaerobes whereas the cluster of sequences that contains EcAGM-like sequences are encoded in Gammaproteobacteria (ex. *Escherichia, Citrobacter, Klebsiella*) that are facultative anaerobes (Figures 6.1C, 6.1E).

To estimate the level of inhibition that metformin may have on these agmatinases in the gut, Michaelis-Menten kinetic equations may be used, assuming a steady state. Taking into account of the concentration of agmatine in the gut, which has been reported to be in the micromolar range (ex. 20 μM), and an average metformin concentration of 5 mM, the reduction in the maximum velocity of EcAGM-like and CbAGM-like enzymes is predicted to be 83% and 27%, respectively (Supp. Fig. A.15)[164]. This indicates that inhibition of EcAGM-like agmatinases by metformin is significant while for CbAGM-like it appears to be less so.

Inhibition of E. coli agmatinase by metformin analogs and its parent compound, galegine

To determine whether agmatinase inhibition is a mechanism of action reserved only for metformin, more potent anti-diabetic analogs, phenformin, buformin and metformin's parent compound, galegine, were tested as competitive inhibitors of EcAGM. This revealed that phenformin, buformin and galegine were more potent inhibitors of EcAGM, compared to metformin, with K_i of 0.6, 0.1 and 0.007 mM, respectively (Figure 6.2A). The potency of the analogs to inhibit *E. coli* agmatinase appears to be proportional to their recommended doses to treat type-II diabetes (Figure 6.2B). The recommended dose for metformin is 1000-2000 mg daily while for phenformin and buformin the dose is more than fivefold less than metformin[12]. Galegine has not been tested clinically but in an experiment by Muller *et al*, galegine was administered to several diabetic patients with doses between 25-100 mg as treatment[165].

The competitive inhibition of EcAGM by metformin and its analogs indicates that the inhibitors bind in the enzyme's active site which contains a binuclear manganese cluster. Attempts were made to get an X-ray crystal structure to show the inhibition complex of metformin or galegine in the EcAGM active site but were not successful. Instead, the complex was modelled *in silico* by docking the inhibitors into the EcAGM active site. The docking suggests that the biguanide inhibitors can bind so that some of its guanidinium nitrogen atoms may chelate the binuclear manganese metals whereas other nitrogen atoms can make polar contacts with EcAGM residues Thr242, Thr244 and Glu274 (Figure 6.2C). The exquisite potency of galegine suggests that it too may chelate the binuclear metals, and it is possible that the sp^2 character of the alkenyl bond of galegine may act as a chelator in an η^2 fashion while the guanidinium nitrogen atoms could interact with the same EcAGM residues that are mentioned to bind the biguanide inhibitors (Figure 6.2D).

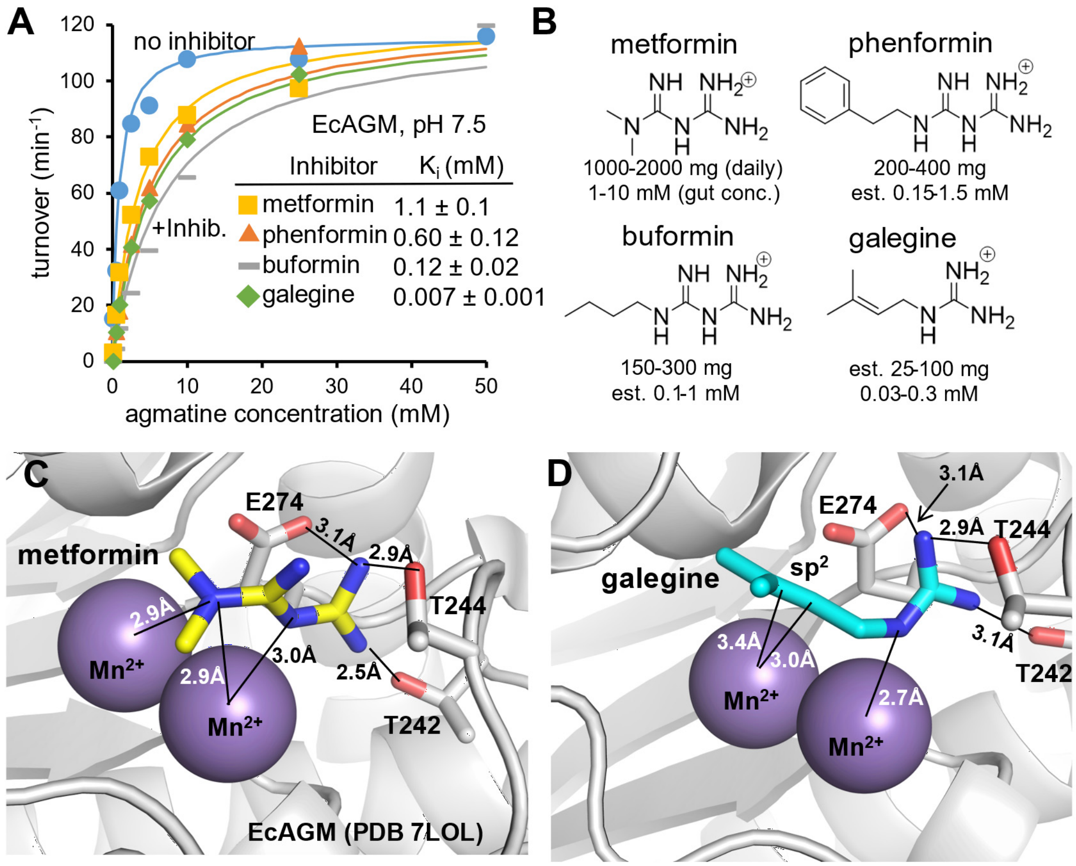
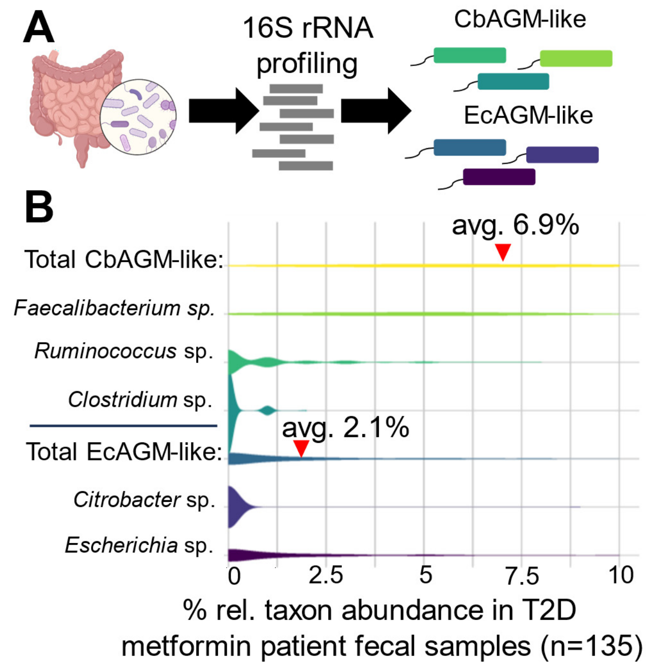


Figure 6.2: Inhibition of *E. coli* agmatinase by metformin analogs. (A) Anti-diabetic analogs phenformin, buformin and galegine are competitive inhibitors of EcAGM. Michealis-Menten kinetics were obtained for EcAGM in the presence or absence of inhibitor and the inhibition constants (K_i) were determined. Activity at several agmatine concentrations was measured by observing urea release using a colorimetry assay in 150 mM HEPES pH 7.5, 2 mM $MnCl_2$. Error bars denote one standard deviation of the mean from two technical replicates. Plot lines show the fit to the Michaelis-Menten equation. The K_i for each inhibitor is expressed as a mean and one standard deviation. (B) Recommended dosage and estimated gut drug concentration of metformin and analogs. Drug potency appears to be correlated with the K_i of the different analogs and estimated gut concentrations are determined relative to metformin's dosage and assuming 75% drug absorption for buformin, phenformin and galegine compared to the known 50% drug absorption for metformin. (C-D) Docking of metformin and galegine into the EcAGM active site. See Supp. Fig. A.16 for docking models of phenformin and buformin.

The mechanism of EcAGM, and other members of the ureohydrolase superfamily, involves metal activated hydroxide to act as the nucleophile in substrate hydrolysis which is demonstrated by the enzyme being the most active in alkaline conditions, where the concentration of hydroxide is higher than neutral pH[129]. To see if the inhibitors are affected by pH and compete with, or are assisted, by the metal bound hydroxide, we determined the inhibition kinetics of metformin and galegine in more acidic, pH 6.5 and more alkaline pH of 8.5. The lower or higher pH resulted in no significant change in the inhibition constants for metformin or galegine that were measured at pH 7.5 (Table A.8).

Metformin inhibition of agmatinase in the context of the human gut

The inhibition by metformin appears most significant for EcAGM-like enzymes and while there are two subfamilies of agmatinases in the human gut, their relative abundance to the whole microbiome and to each other is not clear. To get a good estimation of relative abundance, the taxonomic information of the EcAGM-like and CbAGM-like sequence clusters was cross referenced to the 16S rRNA gene-based profiling of 136 fecal samples obtained from several cohorts of type-II diabetes patients that were prescribed metformin. The profiling data indicated that the most abundant microbes containing EcAGM-like sequence are *Escherichia*, *Citrobacter, Klebsiella* from Gammaproteobacteria (Figure 6.3A). While for CbAGM-like cluster, the profiling data indicated that it is comprised of Clostridiales microbes *Faecalibacterium prasunitzii*, *Ruminococcus* sp. and *Clostridium sp*. The distribution of the relative abundances for CbAGM-like encoding microbes in fecal samples was broad between 0% and 10% while for EcAGM-like microbes the distribution was more narrow between 0% and 2.5% (Figure 6.3B).



C

AGM cluster	est. k_{cat} (min^{-1})	rel. gut AGM abund.	rel. gut AGM activity
EcAGM-like	120	23%	>90%
CbAGM-like	3-0.3	77%	<10%

Figure 6.3: Relative fecal abundances of microbes encoding a CbAGM-like or EcAGM-like agmatinase. (A) Profiling human microbiome using 16S rRNA amplicon sequencing reads. Sequencing reads from microbes with genus-species that matched the taxonomy of EcAGM-like or CbAGM-like protein sequences were counted to estimate relative abundances. (B) Violin plot of relative abundance of CbAGM-like or EcAGM-like encoding microbes in fecal samples of type-II diabetes patients taking metformin. The three most abundant CbAGM-like microbes were *Faecalibacterium*, *Ruminococcus* and *Clostridium* species while for EcAGM-like microbes the two most abundant were *Citrobacter and Escherichia* sp. (C) Estimation of the human gut agmatinase activity contribution by microbes encoding EcAGM-like or CbAGM-like agmatinases. Compounding the relative abundance of CbAGM-like and EcAGM-like microbes with the estimated turnover numbers (k_{cat}) of the enzymes suggests that the majority of gut agmatinase activity (>90%) is contributed by EcAGM-like enzymes in the gut.

A comparison of the means of the two distributions suggests that the CbAGM-like encoding microbes may be more abundant than EcAGM-like encoding microbes in fecal samples. Although EcAGM-like microbes may still contribute the majority of gut agmatinase activity as the turnover number (k_{cat}) of EcAGM at physiological conditions is several orders of magnitude greater than that of characterized CbAGM-like enzymes (Figure 6.3C, Supp. Table A.7). More investigation is necessary to understand levels of gut agmatinase activity *in situ* and what the likely contributions come from CbAGM-like and EcAGM-like agmatinases.

Agmatine produced in the gut is derived from decarboxylation of L-arginine in microbes that encode arginine decarboxylase (ADC). L-arginine comes from dietary or endogenous sources and decarboxylase activity in the gut is highest in the colon, where most gut microbes reside and is an anaerobic environment[166]. Agmatine can be catabolized into putrescine by agmatinase but also by an alternate pathway that is initiated with the agmatine deiminase (AgDI) enzyme (Figure 6.4). Agmatine deiminase is a hydrolytic enzyme like agmatinase but is different in that it is not metal-dependent and yields ammonium and *N*-carbamoylputrescine as products of the reaction[167]. *N*-Carbamoylputrescine is then transformed into putrescine either by putrescine carbamoyltransferase (PCT) or *N*-carbamoylputrescine amidase (NCP) with the latter being hydrolytic and the former produces carbamoylphosphate that can be used by carbamate kinase (CK) to yield adenosine triphosphate (ATP) (Figure 6.4)[168,169]. To see if agmatine catabolism via AgDI is inhibited by metformin or galegine, the previously characterized AgDI homolog from *Enterococcus faecalis* was expressed in *E. coli* and tested *in vitro*. The purified enzyme was active on agmatine but not inhibited by either 50 mM metformin or 5 mM galegine (Supp. Fig. A.17). Putrescine can be utilized by microbes in catabolism to produce metabolites such as γ-aminobutyric acid (GABA) or in biosynthesis of polyamines like spermidine, spermine. The targeted inhibition of gut agmatinase by metformin gives insight that either agmatine, putrescine catabolites or polyamines are effectors of the drug's therapy for type-II diabetes (Figure 6.4).

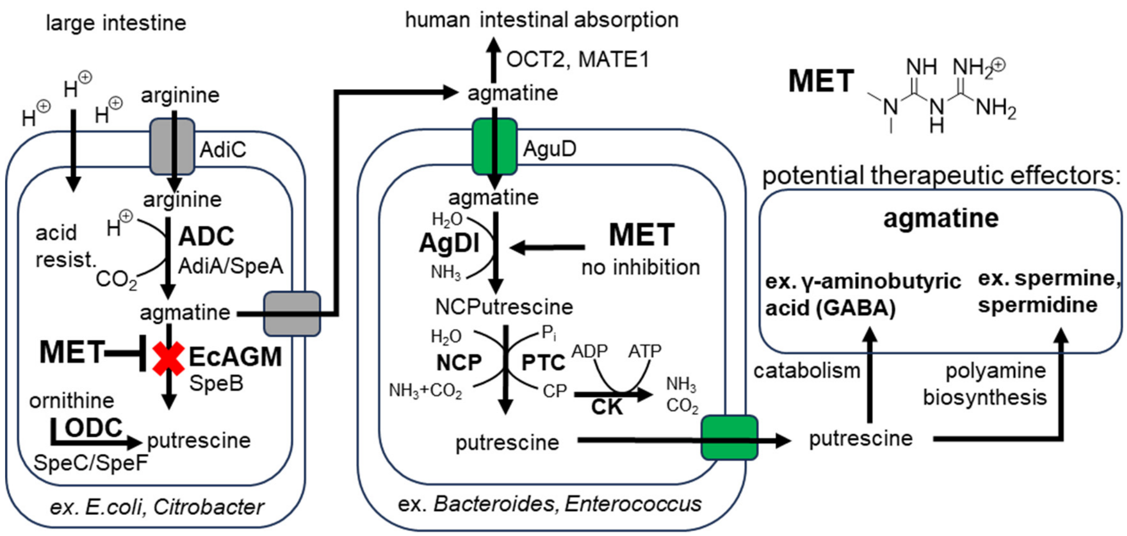


Figure 6.4: Metformin inhibition of agmatinase in the context of the human gut. Agmatine is a product of L-arginine decarboxylation which in *E. coli* or *Citrobacter* serves as an acid resistance mechanism. Arginine and agmatine are known to be transported in and out of the cell by the AdiC antiporter. Metformin inhibition of EcAGM prevents agmatine degradation to putrescine although *E. coli* are able to synthesize putrescine from ornithine via ornithine decarboxylase (ODC). Agmatine is known to be transported into human cells via OCT2 or MATE1 transporters. Agmatine can also be catabolized by gut microbes into putrescine using an alternate pathway that is initiated with agmatine deiminase (AgDI) to produce N-carbamoylputrescine (NCPutrescine). Agmatine deiminase from *Enterococcus faecalis* was not inhibited by metformin or galegine when tested *in vitro* (SI Appendix. Fig. S4). An enzyme, N-carbamoylputrescine amidase (NCP), can hydrolyze this to putrescine in a *Bacteroides* species for example. Alternatively in *E. faecalis*, N-carbamoylputrescine can be converted to putrescine via putrescine transcarbamylase (PTC) which forms carbamoyl phosphate that can be used by carbamate kinase to form an ATP energy equivalent. In these microbes, agmatine and putrescine can be transported in an out of the cell by the AguD antiporter. Putrescine is utilized by microbes in catabolism to produce metabolites such as γ-aminobutyric acid (GABA) or in biosynthesis of polyamines like spermidine, spermine. The targeted inhibition of gut agmatinase by metformin gives insight that either agmatine, putrescine catabolites or polyamines are effectors of the drug's therapy for type-II diabetes.

6.4 Discussion

Metformin targets gut agmatinase activity as a potent, competitive inhibitor

Metformin potently inhibited gut agmatinase activity and not any other homolog of the ureohydrolase protein superfamily or agmatine deiminase that also uses agmatine as a substrate. The agmatinase inhibition constant (K_i) for metformin, at pH 7.5 was around 1 mM which is relevant in the gut where the concentration of metformin is several millimolar and the lumen pH of the small and large intestines varies between pH 6.5-7.5[170].

Inhibition of enzymes *in vitro* by metformin has been described in several prior studies but shows less potent inhibition. Metformin competitively inhibits dihydrofolate reductase (DHFR) from *E. coli* but with a reported K_i of 24 mM[171]. For a human enzyme, serine hydroxymethyltransferase 2 (SHMT2) showed mixed inhibition by metformin with a K_i of 13 mM although the concentration of metformin entering the liver from the portal vein has been measured to be between 10-40 micromolar[5,15]. A study by Ma *et al* found that metformin can bind the human PEN2 protein with a binding constant in the micromolar range and this can cause AMP kinase (AMPK) activation, a hallmark of metformin therapy[14]. However, in the same study biguanide analogs phenformin and buformin were shown to not act in this same manner as metformin. A few studies report metformin inhibiting mitochondrial respiration proteins like complex IV in the hundreds of micromolar but rely on the assumption that metformin can accumulate in the mitochondria of hepatocytes at concentrations several fold higher than in the plasma[13,172].

Agmatine, putrescine and its metabolites as effectors of metformin therapy

The inhibition of gut agmatinase activity suggests either agmatine, putrescine or related metabolites may be effectors in the therapy that metformin delivers (Figure 6.4).

Agmatine is known to act like metformin and lower plasma glucose levels in diabetic mice and also have anti-tumor, anti-aging and neuroprotective effects[138,173]. In a study by Kotagale *et al*, diabetic mice were treated with agmatine and metformin which saw synergistic effects to lower blood glucose and treat impaired cognition of the mice[174]. Agmatine can bind and is an agonist to human imidazoline receptors (I_1, I_2) in the micromolar range and it is purported that this may stimulate the therapeutic effects of the molecule[175,176]. The I_2 receptors specifically are located on the outer membrane of mitochondria but it is still unclear what effects are caused when agmatine binds to these receptors. Another mechanism independent of imidazoline receptors is the effect of agmatine to indirectly inhibit ornithine decarboxylase activity in certain human cell lines and thus polyamine biosynthesis and cell proliferation[177–179]. The effective concentrations for agmatine used in *in vitro* studies to see decreases in cell proliferation were in the range of tens of micromolar. This effect of agmatine is seen as a mechanism to inhibit cancer growth[178].

Metformin does not inhibit agmatine deiminase in the gut which suggests that agmatine can still be catabolized even with metformin treatment. A study by Kitada *et al*, observed that putrescine production in a complex community of gut microbes can be simplified to a collaboration of *E. coli* and *Enterococcus faecalis*, of which the latter encodes agmatine deiminase[137]. The study determined that *E. coli*'s role is to produce agmatine as part of an acid resistance mechanism and then expel agmatine into the media to be consumed by *E. faecalis* to produce putrescine. Agmatine deiminase (AgDI) from *E. faecalis* has superior kinetics parameters (K_M, k_{cat}) compared to EcAGM with a K_M of 35 μM and k_{cat} of 840 s^{-1} which may explain why this synergy is more productive[169]. The study also tested *E. coli* with an agmatinase gene deletion and saw an approximate 50% increase in putrescine production in combination with *E. faecalis*[137]. This experiment was done in culture, and it is important to note that in the context of the gut, the increase in the amount of agmatine exported, due to *E. coli* deficient in agmatinase activity, may be absorbed into the colonic epithelium and only be partially taken up by other gut microbes such as *E. faecalis*. Agmatine is known to be preferentially absorbed into human cells by the human OCT2 and MATE1 transporters[180].

While the *E. coli* agmatinase appears to be a target in the gut for metformin, it is important to mention that a key observation in metformin's effect in altering the gut microbiome of treated patients is the increased abundance of bacterial genera *Escherichia, Enterobacter* and *Citrobacter* , all which encode EcAGM-like agmatinases [3,8,138]. This suggests that inhibition of gut agmatinase activity is not toxic to these microbes with one explanation being that they are still able to produce putrescine and polyamines from another pathway via ornithine decarboxylase (Figure 4). In growth studies of *E. coli* deficient in both pathways to biosynthesize polyamines it was observed that the polyamine spermidine is not needed for aerobic growth but is required for growth in strictly anaerobic conditions, like the large intestine of the human gut[181]. If agmatinase activity is the only target of metformin therapy in the gut then it is possible that positive feedback is coming from the host or other microbes which increases the abundance of EcAGM-like containing microbes. The other key modulations of the gut microbiome by metformin, across many studies, are the increase in short-chain fatty acid (SCFA) producing and mucin-degrading microbiota[3,7,182]. In the aforementioned study by Kitada *et al*, the addition of SCFA producing *Bifidobacterium* microbes in co-culture with *E. coli* and *Enterococcus faecalis* led to an acidification of the medium and increased putrescine production, from *E. coli* and *Enterococcus*, more than two-fold. The acid-resistance mechanism of *E. coli* suggests that SCFA producers in the gut would also promote agmatine production[137]. How metformin-treatment reinforces these modulations of the gut microbiome and whether they are linked remains to be seen but may possibly be probed with this new insight that metformin targets gut agmatinase activity.

According to several reports, putrescine, polyamines or putrescine catabolites derived from the gut may have a myriad of effects that have been relatively understudied compared to agmatine[183,184]. Targeted metabolomics of these compounds in the gut and human tissue is necessary to understand their change in response to metformin treatment. Testing using gnotobiotic and diabetic mice inoculated with *E. coli* deficient in agmatinase activity may be important in revealing metformin's true mechanism in the gut. If agmatine is the true effector of metformin therapy, there are possible opportunities to improve the therapy for the many millions of patients taking metformin. A potential therapy for type-II diabetes may be to administer agmatine in combination with metformin, which may not require as high of dosage as that needed for treatment when using metformin or agmatine alone. A lower dosage would also limit gastrointestinal symptoms of metformin which occurs in 30% of patients[20].

Probiotics like *Bifidobacteria,* that ferment acid metabolites like lactate and short-chain fatty acids, lower intestinal pH and may increase agmatine production of *E. coli*-like bacteria in the gut and could also be co-administered with metformin[137]. One other approach is to develop effective inhibitors of agmatine deiminase to test for type-II diabetes treatment in combination with metformin[185].

Metformin gut drug metabolism

Gut ureohydrolase homologs tested in this study were not active with metformin and were distant in shared sequence identity (<30%) with the metformin hydrolase, MfmAB. This still does not indicate that metformin is not metabolized by gut microbes. The lack of a detected signal may be due to metagenomic sequence coverage in human fecal samples that is not adequate to capture the genes from low-abundance microbes (<1%)[186,187]. In the human gut microbiome, microbes present in low-abundance have been shown to be important in causing extensive changes to the overall community, affecting pathogenicity and secondary metabolism[188]. Quantitative polymerase chain reaction (qPCR) may be a great tool to detect the MfmAB genes in fecal samples and determine if gut drug metabolism may be present.

Chapter 7

Application and Future Research

7.1 Treatment advances for metformin therapy

Metformin treatment is effective and a relatively inexpensive drug for patients and while there are many other small molecule drugs out there to treat type-II diabetes, metformin is exceptional in that there is no risk of causing hypoglycemia unlike other diabetes drugs. Although, there is room for improvement as metformin comes with gastrointestinal side effects. The high daily dosage (1-2 g) of metformin is prone to have off-target effects that include diarrhea, bloating and prolonged usage causes vitamin B12 deficiency[19,20]. These side effects are a key problem for keeping patients compliant on the drug regimen.

The formulation of metformin can control the incidence of side effects while still maintaining drug efficacy[189]. The main formulation used in developed countries is the extended release (XR) form of metformin and has less side effects than the immediate release (IR) version that is cheaper and more available in developing countries[189]. These formulations allow for better toleration of the drug but still require high doses. Elucidating the mechanism of metformin would provide many more avenues to improving the therapy. In Chapter 6, gut microbial agmatinase is identified as a target for inhibition by metformin and other biguanide drugs. Further studies are necessary to understand how significant agmatinase inhibition is for type-II diabetes therapy. This would include testing diabetic mice that are gnotobiotic with agmatine-producing microbes, like *E. coli*, and see if metformin therapy is controlled by the genotype of the microbes. If an agmatinase knockout *E. coli* strain is inoculated in diabetic mice and is found to be therapeutic (lowers plasma glucose) then this may highly suggest agmatinase as a target. In addition, if metformin is administered with the same knockout *E. coli* strain to the mice and no improvement is found than this may indicate that metformin is therapeutic solely through agmatinase inhibition. A similar study has been done looking into metformin and aging in a *C. elegans* model[138]. To better understand metformin and agmatinase in the human context, targeted metabolomics of patients administered metformin would be ideal to quantify agmatine and its catabolites in the gut and bloodstream. This would potentially reveal large cascade effects of agmatinase inhibition in the human gut and identify consistent trends in accumulation of agmatine or reduction in downstream catabolites and polyamines.

If agmatinase is an important drug target, then its protein structure would be a clear basis to design more potent analogs with a similar safety profile. A boronic acid analog of agmatine could be tested and would likely be very potent. This has been done for the enzyme arginase where boronic acid analogs have been reported to be effective at nanomolar concentrations[51]. The likely important criteria for designing a new inhibitor are that the molecule needs to be highly polar to limit intestinal absorption, have tight binding to the target and be relatively inexpensive to make. Achieving all three of these criteria may be difficult but is a clear direction to move towards for metformin drug development.

Another avenue is supplementation of agmatine or probiotics with metformin that may be synergistic and could allow for reduced drug dosage. There is already evidence where synergy is observed for metformin and agmatine in treating diabetic mice to lower blood glucose and treat impaired cognition[174]. Probiotics and metformin co-administration has been tested clinically in many studies and, in summary, shows moderate benefits in glycemic control and great benefits in reducing incidence of gastrointestinal side effects[190]. The probiotics used in most of these clinical studies were predominantly *Lactobacillus* and *Bifidobacteria* which are both known for fermenting acidic products and modulating the luminal pH of the gut[137]s. These probiotics may be beneficial in promoting agmatine production in the gut as it is part of the acid-resistance mechanism of *E. coli*[137]. A possible clinical trial to establish is to have agmatine producing microbes like *E. coli* to be used as a probiotic. This has not been tested clinically at the time of writing this dissertation and if tested may cause strong effects for patients. The *E. coli* strain Nissle 1917 is extensively used as a probiotic and has been found to lower blood glucose in healthy mice[191]. Co-administering metformin with Nissle 1917 may be an interesting clinical trial to establish and potentially observe strong synergy in treating type-II diabetes.

7.2 Discovering gut drug targets at wastewater treatment plants

The journey of this hypothesis driven research throughout this dissertation has been an interesting path. Initially, the goal of this dissertation was to identify metformin metabolizing enzymes in wastewater microbes which would help us find gut metformin metabolism potentially and probe the therapeutic mechanism. However instead we found a connection between metformin metabolizing enzymes and homologous enzymes that metformin can competitively inhibit but not be degraded by. This surprising connection possibly makes one wonder if this could be applied to other oral pharmaceuticals and identify any gut targets. Understanding gut-drug metabolism could explain the variation of side effects that occur from patient to patient[18]. There are many drugs that have therapeutic mechanisms or side effects associated with gut microbiomes[98]. Metagenomic studies have provided some insight into which microbes are affected but actual drug targets remain elusive. One example is acetaminophen, a drug that is generally used for pain relief but still has an unknown mechanism[192]. As acetaminophen and many pharmaceuticals enter wastewater it is likely that wastewater microbes see these compounds and possibly metabolize them[105]. Investigating the wastewater microbiome as opposed to the gut microbiome has benefits in that the wastewater microbes may be easier to cultivate and sampling is not as invasive as human fecal sample collection. Potentially the findings found in this dissertation may encourage other scientists to use a similar approach to translate findings in wastewater to tackle the elusive problems found in pharmacomicrobiomics[192].

7.3 Conclusions

The findings in this dissertation establish a great foundation for future research in wastewater and gut microbiomes as it pertains to metformin. The process of wastewater treatment in the distant future could be augmented with engineered microbes encoded with the metformin-degrading genes. This may be necessary if treated water needs to be fed back into the water supply on a faster time scale. Future work looking at type-II diabetes and gut microbiomes may generate significant findings from the hypothesis developed in this dissertation which finds gut microbial agmatinase a target of metformin. The impact of this work could affect every patient treated with metformin of which there are several hundred million people worldwide. Healthy people are even considering taking this drug as a supplement to slow the aging process which may make this research important in advancing human health in general[193]. Metformin is a wonder drug and the hope is this dissertation makes progress in revealing the therapeutic mechanism that has remained elusive for over a century.

Bibliography

(1) Bailey, C. J. Metformin: Historical Overview. *Diabetologia* **2017**, *60* (9), 1566–1576. https://doi.org/10.1007/s00125-017-4318-z.

(2) Amin, S.; Lux, A.; O'Callaghan, F. The Journey of Metformin from Glycaemic Control to MTOR Inhibition and the Suppression of Tumour Growth. *Br. J. Clin. Pharmacol.* **2019**, *85* (1), 37–46. https://doi.org/10.1111/bcp.13780.

(3) Forslund, K.; Hildebrand, F.; Nielsen, T.; Falony, G.; Le Chatelier, E.; Sunagawa, S.; Prifti, E.; Vieira-Silva, S.; Gudmundsdottir, V.; Krogh Pedersen, H.; Arumugam, M.; Kristiansen, K.; Yvonne Voigt, A.; Vestergaard, H.; Hercog, R.; Igor Costea, P.; Roat Kultima, J.; Li, J.; Jørgensen, T.; Levenez, F.; Dore, J.; Bjørn Nielsen, H.; Brunak, S.; Raes, J.; Hansen, T.; Wang, J.; Dusko Ehrlich, S.; Bork, P.; Pedersen, O. Disentangling Type 2 Diabetes and Metformin Treatment Signatures in the Human Gut Microbiota. *Nat. 2015 5287581* **2015**, *528* (7581), 262–266. https://doi.org/10.1038/nature15766.

(4) Tobar, N.; Rocha, G. Z.; Santos, A.; Guadagnini, D.; Assalin, H. B.; Camargo, J. A.; Gonçalves, A. E. S. S.; Pallis, F. R.; Oliveira, A. G.; Rocco, S. A.; Neto, R. M.; de Sousa, I. L.; Alborghetti, M. R.; Sforça, M. L.; Rodrigues, P. B.; Ludwig, R. G.; Vanzela, E. C.; Brunetto, S. Q.; Boer, P. A.; Gontijo, J. A. R.; Geloneze, B.; Carvalho, C. R. O.; Prada, P. O.; Folli, F.; Curi, R.; Mori, M. A.; Vinolo, M. A. R.; Ramos, C. D.; Franchini, K. G.; Tormena, C. F.; Saad, M. J. A. Metformin Acts in the Gut and Induces Gut-Liver Crosstalk. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *120* (4), e2211933120. https://doi.org/10.1073/PNAS.2211933120/SUPPL_FILE/PNAS.2211933120.SD03.PPTX.

(5) He, L.; Wondisford, F. E. Metformin Action: Concentrations Matter. *Cell Metabolism*. 2015. https://doi.org/10.1016/j.cmet.2015.01.003.

(6) Sum, C. F.; Webster, J. M.; Johnson, A. B.; Catalano, C.; Cooper, B. G.; Taylor, R. The Effect of Intravenous Metformin on Glucose Metabolism during Hyperglycaemia in Type 2 Diabetes. *Diabet. Med.* **1992**, *9* (1), 61–65. https://doi.org/10.1111/j.1464-5491.1992.tb01716.x.

(7) Foretz, M.; Guigas, B.; Viollet, B. Metformin: Update on Mechanisms of Action and Repurposing Potential. *Nature Reviews Endocrinology*. 2023. https://doi.org/10.1038/s41574-023-00833-4.

(8) Wu, H.; Esteve, E.; Tremaroli, V.; Khan, M. T.; Caesar, R.; Mannerås-Holm, L.; Ståhlman, M.; Olsson, L. M.; Serino, M.; Planas-Fèlix, M.; Xifra, G.; Mercader, J. M.; Torrents, D.; Burcelin, R.; Ricart, W.; Perkins, R.; Fernàndez-Real, J. M.; Bäckhed, F. Metformin Alters the Gut Microbiome of Individuals with Treatment-Naive Type 2 Diabetes, Contributing to the Therapeutic Effects of the Drug. *Nat. Med.* **2017**, *23* (7), 850–858. https://doi.org/10.1038/nm.4345.

(9) Kolodkin-Gal, I.; Behrends, V.; Tian, X.; Wang, F.; Liu, J.; Ye, Y.; Jiang, P.; Huang, C.; Li, J.; Chen, J.; Wang, L.; Lin, Y. Metformin Alters the Chemotaxis and Flagellar Motility of Escherichia Coli. *Front. Microbiol* **2022**, *12*, 792406. https://doi.org/10.3389/fmicb.2021.792406.

(10) Tucker, G. T.; Wesolowski, C. A. Metformin Disposition—A 40-Year-Old Mystery. *British Journal of Clinical Pharmacology*. 2020. https://doi.org/10.1111/bcp.14320.

(11) Sirtori, C. R.; Franceschini, G.; Gailli-Kienle, M.; Cighetti, G.; Galli, G.; Bondioli, A.; Conti, F. Disposition of Metformin (N,N-Dimethylbiguanide) in Man. *Clin. Pharmacol. Ther.* **1978**, *24* (6). https://doi.org/10.1002/cpt1978246683.

(12) Luft, D.; Schmülling, R. M.; Eggstein, M. Lactic Acidosis in Biguanide-Treated Diabetics - A Review of 330 Cases. *Diabetologia* **1978**, *14* (2), 75–87. https://doi.org/10.1007/BF01263444.

(13) LaMoia, T. E.; Butrico, G. M.; Kalpage, H. A.; Goedeke, L.; Hubbard, B. T.; Vatner, D. F.; Gaspar, R. C.; Zhang, X. M.; Cline, G. W.; Nakahara, K.; Woo, S.; Shimada, A.; Hüttemann, M.; Shulman, G. I. Metformin, Phenformin, and Galegine Inhibit Complex IV Activity and Reduce Glycerol-Derived Gluconeogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (10). https://doi.org/10.1073/pnas.2122287119.

(14) Ma, T.; Tian, X.; Zhang, B.; Li, M.; Wang, Y.; Yang, C.; Wu, J.; Wei, X.; Qu, Q.; Yu, Y.; Long, S.; Feng, J. W.; Li, C.; Zhang, C.; Xie, C.; Wu, Y.; Xu, Z.; Chen, J.; Yu, Y.; Huang, X.; He, Y.; Yao, L.; Zhang, L.; Zhu, M.; Wang, W.; Wang, Z. C.; Zhang, M.; Bao, Y.; Jia, W.; Lin, S. Y.; Ye, Z.; Piao, H. L.; Deng, X.; Zhang, C. S.; Lin, S. C. Low-Dose Metformin Targets the Lysosomal AMPK Pathway through PEN2. *Nature* **2022**, *603* (7899). https://doi.org/10.1038/s41586-022-04431-8.

(15) Tramonti, A.; Cuyàs, E.; Encinar, J. A.; Pietzke, M.; Paone, A.; Verdura, S.; Arbusà, A.; Martin-castillo, B.; Giardina, G.; Joven, J.; Vazquez, A.; Contestabile, R.; Cutruzzolà, F.; Menendez, J. A. Metformin Is a Pyridoxal-5′-Phosphate (PLP)-Competitive Inhibitor of SHMT2. *Cancers (Basel).* **2021**, *13* (16). https://doi.org/10.3390/cancers13164009.

(16) Schmidt, T. S. B.; Raes, J.; Bork, P. The Human Gut Microbiome: From Association to Modulation. *Cell* **2018**, *172* (6), 1198–1215. https://doi.org/10.1016/J.CELL.2018.02.044.

(17) Beliaeva, M. A.; Wilmanns, M.; Zimmermann, M. Decipher Enzymes from

Human Microbiota for Drug Discovery and Development. *Curr. Opin. Struct. Biol.* **2023**, *80*, 102567. https://doi.org/10.1016/j.sbi.2023.102567.

(18) Zimmermann, M.; Patil, K. R.; Typas, A.; Maier, L. Towards a Mechanistic Understanding of Reciprocal Drug–Microbiome Interactions. *Mol. Syst. Biol.* **2021**, *17* (3). https://doi.org/10.15252/msb.202010116.

(19) Kahn, S. E.; Haffner, S. M.; Heise, M. A.; Herman, W. H.; Holman, R. R.; Jones, N. P.; Kravitz, B. G.; Lachin, J. M.; O'Neill, M. C.; Zinman, B.; Viberti, G. Glycemic Durability of Rosiglitazone, Metformin, or Glyburide Monotherapy. *N. Engl. J. Med.* **2006**, *355* (23), 2427–2443. https://doi.org/10.1056/NEJMOA066224/SUPPL_FILE/NEJMOA066224SA1.PDF.

(20) Bray, G. A.; Edelstein, S. L.; Crandall, J. P.; Aroda, V. R.; Franks, P. W.; Fujimoto, W.; Horton, E.; Jeffries, S.; Montez, M.; Mudaliar, S.; Pi-Sunyer, F. X.; White, N. H.; Knowler, W. C. Long-Term Safety, Tolerability, and Weight Loss Associated with Metformin in the Diabetes Prevention Program Outcomes Study. *Diabetes Care* **2012**, *35* (4). https://doi.org/10.2337/dc11-1299.

(21) Tucker, G.; Casey, C.; Phillips, P.; Connor, H.; Ward, J.; Woods, H. Metformin Kinetics in Healthy Subjects and in Patients with Diabetes Mellitus. *Br. J. Clin. Pharmacol.* **1981**, *12* (2). https://doi.org/10.1111/j.1365-2125.1981.tb01206.x.

(22) Gabr, R. Q.; El-Sherbeni, A. A.; Ben-Eltriki, M.; El-Kadi, A. O.; Brocks, D. R. Pharmacokinetics of Metformin in the Rat: Assessment of the Effect of Hyperlipidemia and Evidence for Its Metabolism to Guanylurea. *Can. J. Physiol. Pharmacol.* **2017**, *95* (5). https://doi.org/10.1139/cjpp-2016-0329.

(23) Scheurer, M.; Michel, A.; Brauch, H. J.; Ruck, W.; Sacher, F. Occurrence and Fate of the Antidiabetic Drug Metformin and Its Metabolite Guanylurea in the

Environment and during Drinking Water Treatment. *Water Res.* **2012**, *46* (15), 4790–4802. https://doi.org/10.1016/j.watres.2012.06.019.

(24) Golovko, O.; Örn, S.; Sörengård, M.; Frieberg, K.; Nassazzi, W.; Lai, F. Y.; Ahrens, L. Occurrence and Removal of Chemicals of Emerging Concern in Wastewater Treatment Plants and Their Impact on Receiving Water Systems. *Sci. Total Environ.* **2021**, *754*, 142122. https://doi.org/10.1016/j.scitotenv.2020.142122.

(25) Scheurer, M.; Sacher, F.; Brauch, H. J. Occurrence of the Antidiabetic Drug Metformin in Sewage and Surface Waters in Germany. *J. Environ. Monit.* **2009**, *11* (9), 1608–1613. https://doi.org/10.1039/b909311g.

(26) Poursat, B. A. J.; van Spanning, R. J. M.; Braster, M.; Helmus, R.; de Voogt, P.; Parsons, J. R. Biodegradation of Metformin and Its Transformation Product, Guanylurea, by Natural and Exposed Microbial Communities. *Ecotoxicol. Environ. Saf.* **2019**, *182*. https://doi.org/10.1016/j.ecoenv.2019.109414.

(27) Zhang, R.; He, Y.; Yao, L.; Chen, J.; Zhu, S.; Rao, X.; Tang, P.; You, J.; Hua, G.; Zhang, L.; Ju, F.; Wu, L. Metformin Chlorination Byproducts in Drinking Water Exhibit Marked Toxicities of a Potential Health Concern. *Environ. Int.* **2021**, *146*, 106244. https://doi.org/10.1016/j.envint.2020.106244.

(28) He, Y.; Zhang, Y.; Ju, F. Metformin Contamination in Global Waters: Biotic and Abiotic Transformation, Byproduct Generation and Toxicity, and Evaluation as a Pharmaceutical Indicator. *Environmental Science and Technology*. American Chemical Society October 4, 2022, pp 13528–13545. https://doi.org/10.1021/acs.est.2c02495.

(29) Armbruster, D.; Happel, O.; Scheurer, M.; Harms, K.; Schmidt, T. C.; Brauch, H. J. Emerging Nitrogenous Disinfection Byproducts: Transformation of the

Antidiabetic Drug Metformin during Chlorine Disinfection of Water. *Water Res.* **2015**, *79*, 104–118. https://doi.org/10.1016/j.watres.2015.04.020.

(30) Navarathna, D. H. M. L. P.; Harris, S. D.; Roberts, D. D.; Nickerson, K. W. Evolutionary Aspects of Urea Utilization by Fungi. *FEMS Yeast Res.* **2010**, *10* (2), 209–213. https://doi.org/10.1111/j.1567-1364.2009.00602.x.

(31) Strope, P. K.; Nickerson, K. W.; Harris, S. D.; Moriyama, E. N. Molecular Evolution of Urea Amidolyase and Urea Carboxylase in Fungi. *BMC Evol. Biol.* **2011**, *11* (1), 80. https://doi.org/10.1186/1471-2148-11-80.

(32) Whitney, P. A.; Cooper, T. G. Urea Carboxylase and Allophanate Hydrolase: Two Components of a Multienzyme Complex in Saccharomyces Cerevisiae. *Biochem. Biophys. Res. Commun.* **1972**, *49* (1), 45–51. https://doi.org/10.1016/0006-291X(72)90007-1.

(33) Cooper, T. G.; Lam, C.; Turoscy, V. Structural Analysis of the Dur Loci in S. Cerevisiae: Two Domains of a Single Multifunctional Gene. *Genetics* **1980**, *94* (3), 555–580. https://doi.org/10.1093/genetics/94.3.555.

(34) Kanamori, T.; Kanou, N.; Atomi, H.; Imanaka, T. Enzymatic Characterization of a Prokaryotic Urea Carboxylase. *J. Bacteriol.* **2004**, *186* (9), 2532–2539. https://doi.org/10.1128/JB.186.9.2532-2539.2004.

(35) Kanamori, T.; Kanou, N.; Kusakabe, S.; Atomi, H.; Imanaka, T. Allophanate Hydrolase of Oleomonas Sagaranensis Involved in an ATP-Dependent Degradation Pathway Specific to Urea. *FEMS Microbiol. Lett.* **2005**, *245* (1), 61–65. https://doi.org/10.1016/j.femsle.2005.02.023.

(36) Schneider, N. O.; Tassoulas, L. J.; Zeng, D.; Laseke, A. J.; Reiter, N. J.; Wackett, L. P.; Maurice, M. S. Solving the Conundrum: Widespread Proteins Annotated for

Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine. *Biochemistry* **2020**, *59* (35), 3258–3270. https://doi.org/10.1021/acs.biochem.0c00537.

(37) Nelson, J. W.; Atilho, R. M.; Sherlock, M. E.; Stockbridge, R. B.; Breaker, R. R. Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Mol. Cell* **2017**, *65* (2), 220–230. https://doi.org/10.1016/j.molcel.2016.11.019.

(38) Zhao, J.; Zhu, L.; Fan, C.; Wu, Y.; Xiang, S. Structure and Function of Urea Amidolyase. *Biosci. Rep.* **2018**, *38* (1), NA-NA. https://doi.org/10.1042/BSR20171617.

(39) Lin, Y.; Boese, C. J.; St. Maurice, M. The Urea Carboxylase and Allophanate Hydrolase Activities of Urea Amidolyase Are Functionally Independent. *Protein Sci.* **2016**, *25* (10), 1812–1824. https://doi.org/10.1002/pro.2990.

(40) Selengut, J. D.; Rusch, D. B.; Haft, D. H. Sites Inferred by Metabolic Background Assertion Labeling (SIMBAL): Adapting the Partial Phylogenetic Profiling Algorithm to Scan Sequences for Signatures That Predict Protein Function. *BMC Bioinformatics* **2010**, *11* (1), 52. https://doi.org/10.1186/1471-2105-11-52.

(41) Minami, T.; Anda, M.; Mitsui, H.; Sugawara, M.; Kaneko, T.; Sato, S.; Ikeda, S.; Okubo, T.; Tsurumaru, H.; Minamisawa, K. Metagenomic Analysis Revealed Methylamine and Ureide Utilization of Soybean-Associated Methylobacterium. *Microbes Environ.* **2016**, *31* (3), 268–278. https://doi.org/10.1264/jsme2.ME16035.

(42) Oshiki, M.; Araki, M.; Hirakata, Y.; Hatamoto, M.; Yamaguchi, T.; Araki, N. Ureolytic Prokaryotes in Soil: Community Abundance and Diversity. *Microbes Environ.* **2018**, *33* (2), 230–233. https://doi.org/10.1264/jsme2.ME17188.

(43) Zhang, Y.; Rodionov, D. A.; Gelfand, M. S.; Gladyshev, V. N. Comparative Genomic Analyses of Nickel, Cobalt and Vitamin B12 Utilization. *BMC Genomics* **2009**, *10* (1), 78. https://doi.org/10.1186/1471-2164-10-78.

(44) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochim. Biophys. Acta - Proteins Proteomics* **2015**, *1854* (8), 1019–1037. https://doi.org/10.1016/j.bbapap.2015.04.015.

(45) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.

(46) Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7* (10), e1002195-NA. https://doi.org/10.1371/journal.pcbi.1002195.

(47) Haft, D. H.; Loftus, B. J.; Richardson, D. L.; Yang, F.; Eisen, J. A.; Paulsen, I. T.; White, O. TIGRFAMs: A Protein Family Resource for the Functional Identification of Proteins. *Nucleic Acids Res.* **2001**, *29* (1), 41–43. https://doi.org/10.1093/nar/29.1.41.

(48) Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H. C.; Zakai, U. I.; Mitchell, D. A. A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, *13* (5), 470–478. https://doi.org/10.1038/nchembio.2319.

(49) 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.; Liu, H. W. The Amipurimycin and Miharamycin Biosynthetic Gene Clusters: Unraveling

the Origins of 2-Aminopurinyl Peptidyl Nucleoside Antibiotics. *J. Am. Chem. Soc.* **2019**, *141* (36), 14152–14159. https://doi.org/10.1021/jacs.9b03021.

(50) Galkin, A.; Lu, X.; Dunaway-Mariano, D.; Herzberg, O. Crystal Structures Representing the Michaelis Complex and the Thiouronium Reaction Intermediate of Pseudomonas Aeruginosa Arginine Deiminase. *J. Biol. Chem.* **2005**, *280* (40), 34080–34087. https://doi.org/10.1074/jbc.M505471200.

(51) Di Costanzo, L.; Sabio, G.; Mora, A.; Rodriguez, P. C.; Ochoa, A. C.; Centeno, F.; Christianson, D. W. Crystal Structure of Human Arginase I at 1.29-Å Resolution and Exploration of Inhibition in the Immune Response. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (37), 13058–13063. https://doi.org/10.1073/pnas.0504027102.

(52) Betts, L.; Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W. Cytidine Deaminase. the 2.3 Å Crystal Structure of an Enzyme: Transition-State Analog Complex. *J. Mol. Biol.* **1994**, *235* (2), 635–656. https://doi.org/10.1006/jmbi.1994.1018.

(53) Gong, L.; Goswami, S.; Giacomini, K. M.; Altman, R. B.; Klein, T. E. Metformin Pathways: Pharmacokinetics and Pharmacodynamics. *Pharmacogenet. Genomics* **2012**, *22* (11), 820–827. https://doi.org/10.1097/FPC.0b013e3283559b22.

(54) Briones, R. M.; Zhuang, W. Q.; Sarmah, A. K. Biodegradation of Metformin and Guanylurea by Aerobic Cultures Enriched from Sludge. *Environ. Pollut.* **2018**, *243* (Pt A), 255–262. https://doi.org/10.1016/j.envpol.2018.08.075.

(55) Scheen, A. J. Metformin and COVID-19: From Cellular Mechanisms to Reduced Mortality. *Diabetes Metab.* **2020**, *46* (6), 423–426. https://doi.org/10.1016/j.diabet.2020.07.006.

(56) Novelle, M. G.; Ali, A.; Diéguez, C.; Bernier, M.; de Cabo, R. Metformin: A

Hopeful Promise in Aging Research. *Cold Spring Harb. Perspect. Med.* **2016**, *6* (3), a025932-NA. https://doi.org/10.1101/cshperspect.a025932.

(57) Bramante, C. T.; Ingraham, N. E.; Murray, T. A.; Marmor, S.; Hovertsen, S.; Gronski, J.; McNeil, C.; Feng, R.; Guzman, G.; Abdelwahab, N.; King, S.; Meehan, T.; Pendleton, K. M.; Benson, B.; Vojta, D.; Tignanelli, C. J. Observational Study of Metformin and Risk of Mortality in Patients Hospitalized with Covid-19. *medRxiv : the preprint server for health sciences*. 2020, p NA-NA. https://doi.org/10.1101/2020.06.19.20135095.

(58) Trautwein, C.; Kümmerer, K. Incomplete Aerobic Degradation of the Antidiabetic Drug Metformin and Identification of the Bacterial Dead-End Transformation Product Guanylurea. *Chemosphere* **2011**, *85* (5), 765–773. https://doi.org/10.1016/j.chemosphere.2011.06.057.

(59) Bradley, P. M.; Journey, C. A.; Button, D. T.; Carlisle, D. M.; Clark, J. M.; Mahler, B. J.; Nakagaki, N.; Qi, S. L.; Waite, I. R.; VanMetre, P. C. Metformin and Other Pharmaceuticals Widespread in Wadeable Streams of the Southeastern United States. *Environ. Sci. Technol. Lett.* **2016**, *3* (6), 243–249. https://doi.org/10.1021/acs.estlett.6b00170.

(60) Bradley, P. M.; Battaglin, W. A.; Clark, J. M.; Henning, F. P.; Hladik, M. L.; Iwanowicz, L. R.; Journey, C. A.; Riley, J. W.; Romanok, K. M. Widespread Occurrence and Potential for Biodegradation of Bioactive Contaminants in Congaree National Park, USA. *Environ. Toxicol. Chem.* **2017**, *36* (11), 3045–3056. https://doi.org/10.1002/etc.3873.

(61) Briones, R. M.; Sarmah, A. K.; Padhye, L. P. A Global Perspective on the Use, Occurrence, Fate and Effects of Anti-Diabetic Drug Metformin in Natural and Engineered Ecosystems. *Environ. Pollut.* **2016**, *219* (NA), 1007–1020.

https://doi.org/10.1016/j.envpol.2016.07.040.

(62) Tao, Y.; Chen, B.; Zhang, B. (Helen); Zhu, Z. (Joy); Cai, Q. Occurrence, Impact, Analysis and Treatment of Metformin and Guanylurea in Coastal Aquatic Environments of Canada, USA and Europe. *Adv. Mar. Biol.* **2018**, *81* (NA), 23–58. https://doi.org/10.1016/bs.amb.2018.09.005.

(63) Tong, A. Z.; Ghoshdastidar, A. J.; Fox, S. The Presence of the Top Prescribed Pharmaceuticals in Treated Sewage Effluents and Receiving Waters in Southwest Nova Scotia, Canada. *Environ. Sci. Pollut. Res.* **2015**, *22* (1), 689–700. https://doi.org/10.1007/s11356-014-3400-z.

(64) Blair, B. D.; Crago, J. P.; Hedman, C. J.; Treguer, R. J. F.; Magruder, C.; Royer, L. S.; Klaper, R. D. Evaluation of a Model for the Removal of Pharmaceuticals, Personal Care Products, and Hormones from Wastewater. *Sci. Total Environ.* **2013**, *444* (NA), 515–521. https://doi.org/10.1016/j.scitotenv.2012.11.103.

(65) Oosterhuis, M.; Sacher, F.; ter Laak, T. L. Prediction of Concentration Levels of Metformin and Other High Consumption Pharmaceuticals in Wastewater and Regional Surface Water Based on Sales Data. *Sci. Total Environ.* **2013**, *442* (NA), 380–388. https://doi.org/10.1016/j.scitotenv.2012.10.046.

(66) Piai, L.; Blokland, M.; van der Wal, A.; Langenhoff, A. Biodegradation and Adsorption of Micropollutants by Biological Activated Carbon from a Drinking Water Production Plant. *J. Hazard. Mater.* **2020**, *388* (NA), 122028. https://doi.org/10.1016/j.jhazmat.2020.122028.

(67) Niemuth, N. J.; Klaper, R. D. Emerging Wastewater Contaminant Metformin Causes Intersex and Reduced Fecundity in Fish. *Chemosphere* **2015**, *135* (NA), 38–45. https://doi.org/10.1016/j.chemosphere.2015.03.060.

(68) Straub, J. O.; Caldwell, D. J.; Davidson, T.; D'Aco, V.; Kappler, K.; Robinson, P. F.; Simon-Hettich, B.; Tell, J. Environmental Risk Assessment of Metformin and Its Transformation Product Guanylurea. I. Environmental Fate. *Chemosphere* **2019**, *216* (NA), 844–854. https://doi.org/10.1016/j.chemosphere.2018.10.036.

(69) Meador, J. P.; Yeh, A.; Young, G.; Gallagher, E. P. Contaminants of Emerging Concern in a Large Temperate Estuary. *Environ. Pollut.* **2016**, *213* (NA), 254–267. https://doi.org/10.1016/j.envpol.2016.01.088.

(70) Waterhouse, H.; Wade, J.; Horwath, W. R.; Burger, M. Effects of Positively Charged Dicyandiamide and Nitrogen Fertilizer Sources on Nitrous Oxide Emissions in Irrigated Corn. *J. Environ. Qual.* **2017**, *46* (5), 1123–1130. https://doi.org/10.2134/jeq2017.01.0033.

(71) Cassman, N. A.; Soares, J. R.; Pijl, A.; Lourenço, K. S.; van Veen, J. A.; Cantarella, H.; Kuramae, E. E. Nitrification Inhibitors Effectively Target N 2 O-Producing Nitrosospira Spp. in Tropical Soil. *Environ. Microbiol.* **2019**, *21* (4), 1241–1254. https://doi.org/10.1111/1462-2920.14557.

(72) Perreault, N. N.; Halasz, A.; Thiboutot, S.; Ampleman, G.; Hawari, J. Joint Photomicrobial Process for the Degradation of the Insensitive Munition N-Guanylurea-Dinitramide (FOX-12). *Environ. Sci. Technol.* **2013**, *47* (10), 5193–5198. https://doi.org/10.1021/es4006652.

(73) Xiao, Z.; Liu, S.; Zhang, Z.; Mai, C.; Xie, Y.; Wang, Q. Fire Retardancy of an Aqueous, Intumescent, and Translucent Wood Varnish Based on Guanylurea Phosphate and Melamine-Urea-Formaldehyde Resin. *Prog. Org. Coatings* **2018**, *121* (NA), 64–72. https://doi.org/10.1016/j.porgcoat.2018.04.015.

(74) Ito, K.; Hashimoto, Y. Gigartinine : A New Amino-Acid in Red Algae [32]. *Nature* **1966**, *211* (5047), 417. https://doi.org/10.1038/211417a0.

(75) Ito, K.; Miyazawa, K.; Hashimoto, Y. Distribution Of Gongrine and Gigartinine in Marine Algae. *Nippon Suisan Gakkaishi (Japanese Ed.* **1966**, *32* (9), 727–729. https://doi.org/10.2331/suisan.32.727.

(76) Ito, K.; Hashimoto, Y. Occurrence of γ-(Guanylureido)Butyric Acid in a Red Alga, Gymnogongrus Flabelliformis. *Agric. Biol. Chem.* **1965**, *29* (9), 832–835. https://doi.org/10.1271/bbb1961.29.832.

(77) Takagi, M. Low Molecular Nitrogen Compounds of Marine Algae. *Bull. Fac. Fish. Hokkaido Univ.* 1970, p 21. https://doi.org/NA.

(78) Laycock, M. V.; Craigie, J. S. The Occurrence and Seasonal Variation of Gigartinine and L Citrullinyl L Arginine in Chondrus Crispus Stackh. *Can. J. Biochem.* **1977**, *55* (1), 27–30. https://doi.org/10.1139/o77-004.

(79) Mandelbaum, R. T.; Wackett, L. P.; Allan, D. L. Mineralization of the S-Triazine Ring of Atrazine by Stable Bacterial Mixed Cultures. *Appl. Environ. Microbiol.* **1993**, *59* (6), 1695–1701. https://doi.org/10.1128/aem.59.6.1695-1701.1993.

(80) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

(81) Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19* (5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

(82) Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**,

30 (14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

(83) Weatherburn, M. W. Phenol-Hypochlorite Reaction for Determination of Ammonia. *Anal. Chem.* **1967**, *39* (8), 971–974. https://doi.org/10.1021/ac60252a045.

(84) Awaya, J. D.; DuBois, J. L. Identification, Isolation, and Analysis of a Gene Cluster Involved in Iron Acquisition by Pseudomonas Mendocina Ymp. *BioMetals* **2008**, *21* (3), 353–366. https://doi.org/10.1007/s10534-007-9124-5.

(85) Dbasy, C. L. Specificity of the Action of Urease. *J. Am. Chem. Soc.* **1946**, *68* (8), 1664–1665. https://doi.org/10.1021/ja01212a506.

(86) Cameron, S. M.; Durchschein, K.; Richman, J. E.; Sadowsky, M. J.; Wackett, L. P. New Family of Biuret Hydrolases Involved in s -Triazine Ring Metabolism. *ACS Catal.* **2011**, *1* (9), 1075–1082. https://doi.org/10.1021/cs200295n.

(87) Davidi, D.; Longo, L. M.; Jabłońska, J.; Milo, R.; Tawfik, D. S. A Bird's-Eye View of Enzyme Evolution: Chemical, Physicochemical, and Physiological Considerations. *Chem. Rev.* **2018**, *118* (18), 8786–8797. https://doi.org/10.1021/acs.chemrev.8b00039.

(88) Wang, T. T.; Bishop, S. H.; Himoe, A. Detection of Carbamate as a Product of the Carbamate Kinase-Catalyzed Reaction by Stopped Flow Spectrophotometry. *J. Biol. Chem.* **1972**, *247* (14), 4437–4440. https://doi.org/10.1016/s0021-9258(19)45003-5.

(89) Liu, Y.; Budelier, M. M.; Stine, K.; St Maurice, M. Allosteric Regulation Alters Carrier Domain Translocation in Pyruvate Carboxylase. *Nat. Commun.* **2018**, *9* (1), 1384. https://doi.org/10.1038/s41467-018-03814-8.

(90) Esquirol, L.; Peat, T. S.; Wilding, M.; Lucent, D.; French, N. G.; Hartley, C. J.;

Newman, J.; Scott, C. Structural and Biochemical Characterization of the Biuret Hydrolase (BiuH) from the Cyanuric Acid Catabolism Pathway of Rhizobium Leguminasorum Bv. Viciae 3841. *PLoS One* **2018**, *13* (2), e0192736-NA. https://doi.org/10.1371/journal.pone.0192736.

(91) Tassoulas, L. J.; Elias, M. H.; Wackett, L. P. Discovery of an Ultraspecific Triuret Hydrolase (TrtA) Establishes the Triuret Biodegradation Pathway. *J. Biol. Chem.* **2021**, *296* (NA), 100055-NA. https://doi.org/10.1074/jbc.RA120.015631.

(92) Bradley, P. M.; Journey, C. A.; Button, D. T.; Carlisle, D. M.; Huffman, B. J.; Qi, S. L.; Romanok, K. M.; van Metre, P. C. Multi-Region Assessment of Pharmaceutical Exposures and Predicted Effects in USA Wadeable Urban-Gradient Streams. *PLoS One* **2020**, *15* (1), 1–25. https://doi.org/10.1371/journal.pone.0228214.

(93) Posselt, M.; Mechelke, J.; Rutere, C.; Coll, C.; Jaeger, A.; Raza, M.; Meinikmann, K.; Krause, S.; Sobek, A.; Lewandowski, J.; Horn, M. A.; Hollender, J.; Benskin, J. P. Bacterial Diversity Controls Transformation of Wastewater-Derived Organic Contaminants in River-Simulating Flumes. *Environ. Sci. Technol.* **2020**, *54* (9), 5467–5479. https://doi.org/10.1021/acs.est.9b06928.

(94) Boxall, A. B. A.; Sinclair, C. J.; Fenner, K.; Kolpin, D.; Maund, S. J. When Synthetic Chemicals Degrade in the Environment. *Environ. Sci. Technol.* **2004**, *38* (19), 368A-375A. https://doi.org/10.1021/es040624v.

(95) Ning, J.; Ai, S.; Cui, L. Dicyandiamide Has More Inhibitory Activities on Nitrification than Thiosulfate. *PLoS One* **2018**, *13* (8), e0200598-NA. https://doi.org/10.1371/journal.pone.0200598.

(96) Wu, L.; Ning, D.; Zhang, B.; Li, Y.; Zhang, P.; Shan, X.; Zhang, Q.; Brown, M. R.; Li, Z.; Van Nostrand, J. D.; Ling, F.; Xiao, N.; Zhang, Y.; Vierheilig, J.; Wells,

G. F.; Yang, Y.; Deng, Y.; Tu, Q.; Wang, A.; Zhang, T.; He, Z.; Keller, J.; Nielsen, P. H.; Alvarez, P. J. J.; Criddle, C. S.; Wagner, M.; Tiedje, J. M.; He, Q.; Curtis, T. P.; Stahl, D. A.; Alvarez-Cohen, L.; Rittmann, B. E.; Wen, X.; Zhou, J.; Acevedo, D.; Agullo-Barcelo, M.; Alvarez, P. J. J.; Alvarez-Cohen, L.; Andersen, G. L.; de Araujo, J. C.; Boehnke, K. F.; Bond, P.; Bott, C. B.; Bovio, P.; Brewster, R. K.; Bux, F.; Cabezas, A.; Cabrol, L.; Chen, S.; Criddle, C. S.; Deng, Y.; Etchebehere, C.; Ford, A.; Frigon, D.; Sanabria, J.; Griffin, J. S.; Gu, A. Z.; Habagil, M.; Hale, L.; Hardeman, S. D.; Harmon, M.; Horn, H.; Hu, Z.; Jauffur, S.; Johnson, D. R.; Keller, J.; Keucken, A.; Kumari, S.; Leal, C. D.; Lebrun, L. A.; Lee, J.; Lee, M.; Lee, Z. M. P.; Li, Y.; Li, Z.; Li, M.; Li, X.; Ling, F.; Liu, Y.; Luthy, R. G.; Mendonça-Hagler, L. C.; de Menezes, F. G. R.; Meyers, A. J.; Mohebbi, A.; Nielsen, P. H.; Ning, D.; Oehmen, A.; Palmer, A.; Parameswaran, P.; Park, J.; Patsch, D.; Reginatto, V.; de los Reyes, F. L.; Rittmann, B. E.; Noyola, A.; Rossetti, S.; Shan, X.; Sidhu, J.; Sloan, W. T.; Smith, K.; de Sousa, O. V.; Stahl, D. A.; Stephens, K.; Tian, R.; Tiedje, J. M.; Tooker, N. B.; Tu, Q.; Van Nostrand, J. D.; Vasconcelos, D. D. los C.; Vierheilig, J.; Wakelin, S.; Wang, A.; Wang, B.; Weaver, J. E.; Wells, G. F.; West, S.; Wilmes, P.; Woo, S. G.; Wu, L.; Wu, J. H.; Wu, L.; Xi, C.; Xiao, N.; Xu, M.; Yan, T.; Yang, Y.; Yang, M.; Young, M.; Yue, H.; Zhang, B.; Zhang, P.; Zhang, Q.; Zhang, Y.; Zhang, T.; Zhang, Q.; Zhang, W.; Zhang, Y.; Zhou, H.; Zhou, J.; Wen, X.; Curtis, T. P.; He, Q.; He, Z.; Brown, M. R. Global Diversity and Biogeography of Bacterial Communities in Wastewater Treatment Plants. *Nat. Microbiol.* **2019**, *4* (7), 1183–1195. https://doi.org/10.1038/s41564-019-0426-5.

(97) Wu, B.; Chen, M. Y.; Gao, Y. C.; Hu, J. L.; Liu, M. Z.; Zhang, W.; Huang, W. H. In Vivo Pharmacodynamic and Pharmacokinetic Effects of Metformin Mediated by the Gut Microbiota in Rats. *Life Sci.* **2019**, *226* (NA), 185–192. https://doi.org/10.1016/j.lfs.2019.04.009.

(98) Vich Vila, A.; Collij, V.; Sanna, S.; Sinha, T.; Imhann, F.; Bourgonje, A. R.; Mujagic, Z.; Jonkers, D. M. A. E.; Masclee, A. A. M.; Fu, J.; Kurilshikov, A.; Wijmenga, C.; Zhernakova, A.; Weersma, R. K. Impact of Commonly Used Drugs on the Composition and Metabolic Function of the Gut Microbiota. *Nat. Commun.* **2020**, *11* (1), 362. https://doi.org/10.1038/s41467-019-14177-z.

(99) Robinson, S. L.; Badalamenti, J. P.; Dodge, A. G.; Tassoulas, L. J.; Wackett, L. P. Microbial Biodegradation of Biuret: Defining Biuret Hydrolases within the Isochorismatase Superfamily. *Environ. Microbiol.* **2018**, *20* (6), 2099–2111. https://doi.org/10.1111/1462-2920.14094.

(100) Poelarends, G. J.; Almrud, J. J.; Serrano, H.; Darty, J. E.; Johnson, W. H.; Hackert, M. L.; Whitman, C. P. Evolution of Enzymatic Activity in the Tautomerase Superfamily: Mechanistic and Structural Consequences of the L8R Mutation in 4-Oxalocrotonate Tautomerase. *Biochemistry* **2006**, *45* (25), 7700–7708. https://doi.org/10.1021/bi0600603.

(101) Poelarends, G. J.; Whitman, C. P. Evolution of Enzymatic Activity in the Tautomerase Superfamily: Mechanistic and Structural Studies of the 1,3-Dichloropropene Catabolic Enzymes. *Bioorg. Chem.* **2004**, *32* (5), 376–392. https://doi.org/10.1016/j.bioorg.2004.05.006.

(102) Seffernick, J. L.; Wackett, L. P. Rapid Evolution of Bacterial Catabolic Enzymes: A Case Study with Atrazine Chlorohydrolase. *Biochemistry* **2001**, *40* (43), 12747–12753. https://doi.org/10.1021/bi011293r.

(103) Shapir, N.; Mongodin, E. F.; Sadowsky, M. J.; Daugherty, S. C.; Nelson, K. E.; Wackett, L. P. Evolution of Catabolic Pathways: Genomic Insights into Microbial s-Triazine Metabolism. *J. Bacteriol.* **2007**, *189* (3), 674–682. https://doi.org/10.1128/JB.01257-06.

(104) Wackett, L. P.; Robinson, S. L. The Ever-Expanding Limits of Enzyme Catalysis and Biodegradation: Polyaromatic, Polychlorinated, Polyfluorinated, and Polymeric Compounds. *Biochem. J.* **2020**, *477* (15), 2875–2891. https://doi.org/10.1042/BCJ20190720.

(105) Wilkinson, J. L.; Boxall, A. B. A.; Kolpin, D. W.; Leung, K. M. Y.; Lai, R. W. S.; Galban-Malag, C.; Adell, A. D.; Mondon, J.; Metian, M.; Marchant, R. A.; Bouzas-Monroy, A.; Cuni-Sanchez, A.; Coors, A.; Carriquiriborde, P.; Rojo, M.; Gordon, C.; Cara, M.; Moermond, M.; Luarte, T.; Petrosyan, V.; Perikhanyan, Y.; Mahon, C. S.; McGurk, C. J.; Hofmann, T.; Kormoker, T.; Iniguez, V.; Guzman-Otazo, J.; Tavares, J. L.; de Figueiredo, F. G.; Razzolini, M. T. P.; Dougnon, V.; Gbaguidi, G.; Traore, O.; Blais, J. M.; Kimpe, L. E.; Wong, M.; Wong, D.; Ntchantcho, R.; Pizarro, J.; Ying, G. G.; Chen, C. E.; Paez, M.; Martınez-Lara, J.; Otamonga, J. P.; Pote, J.; Ifo, S. A.; Wilson, P.; Echeverrıa-Saenz, S.; Udikovic-Kolic, N.; Milakovic, M.; Fatta-Kassinos, D.; Ioannou-Ttofa, L.; Belusova, V.; Vymazal, J.; Cardenas-Bustamante, M.; Kassa, B. A.; Garric, J.; Chaumot, A.; Gibba, P.; Kunchulia, I.; Seidensticker, S.; Lyberatos, G.; Halldorsson, H. P.; Melling, M.; Shashidhar, T.; Lamba, M.; Nastiti, A.; Supriatin, A.; Pourang, N.; Abedini, A.; Abdullah, O.; Gharbia, S. S.; Pilla, F.; Chefetz, B.; Topaz, T.; Yao, K. M.; Aubakirova, B.; Beisenova, R.; Olaka, L.; Mulu, J. K.; Chatanga, P.; Ntuli, V.; Blama, N. T.; Sherif, S.; Aris, A. Z.; Looi, L. J.; Niang, M.; Traore, S. T.; Oldenkamp, R.; Ogunbanwo, O.; Ashfaq, M.; Iqbal, M.; Abdeen, Z.; O'Dea, A.; Morales-Saldaña, J. M.; Custodio, M.; de la Cruz, H.; Navarrete, I.; Carvalho, F.; Gogra, A. B.; Koroma, B. M.; Cerkvenik-Flajs, V.; Gombac, M.; Thwala, M.; Choi, K.; Kang, H.; Celestino Ladu, J. L.; Rico, A.; Amerasinghe, P.; Sobek, A.; Horlitz, G.; Zenker, A. K.; King, A. C.; Jiang, J. J.; Kariuki, R.; Tumbo, M.; Tezel, U.; Onay, T. T.; Lejju, J. B.; Vystavna, Y.; Vergeles, Y.; Heinzen, H.; Perez-Parada, A.; Sims, D. B.; Figy, M.; Good, D.; Teta, C. Pharmaceutical Pollution of

the World's Rivers. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (8). https://doi.org/10.1073/pnas.2113947119.

(106) Trautwein, C.; Berset, J. D.; Wolschke, H.; Kümmerer, K. Occurrence of the Antidiabetic Drug Metformin and Its Ultimate Transformation Product Guanylurea in Several Compartments of the Aquatic Cycle. *Environ. Int.* **2014**, *70* (NA), 203–212. https://doi.org/10.1016/j.envint.2014.05.008.

(107) Barros, S.; Alves, N.; Pinheiro, M.; Ribeiro, M.; Morais, H.; Montes, R.; Rodil, R.; Quintana, J. B.; Coimbra, A. M.; Santos, M. M.; Neuparth, T. Are Fish Populations at Risk? Metformin Disrupts Zebrafish Development and Reproductive Processes at Chronic Environmentally Relevant Concentrations. *Environ. Sci. Technol.* **2023**, *57* (2). https://doi.org/10.1021/acs.est.2c05719.

(108) Elizalde-Velázquez, G. A.; Herrera-Vázquez, S. E.; Gómez-Oliván, L. M.; García-Medina, S. Health Impact Assessment after Danio Rerio Long-Term Exposure to Environmentally Relevant Concentrations of Metformin and Guanylurea. *Chemosphere* **2023**, *341*. https://doi.org/10.1016/j.chemosphere.2023.140070.

(109) Martinez-Vaz, B. M.; Dodge, A. G.; Lucero, R. M.; Stockbridge, R. B.; Robinson, A. A.; Tassoulas, L. J.; Wackett, L. P. Wastewater Bacteria Remediating the Pharmaceutical Metformin: Genomes, Plasmids and Products. *Front. Bioeng. Biotechnol.* **2022**, *10*, 2291. https://doi.org/10.3389/fbioe.2022.1086261.

(110) Chaignaud, P.; Gruffaz, C.; Borreca, A.; Fouteau, S.; Kuhn, L.; Masbou, J.; Rouy, Z.; Hammann, P.; Imfeld, G.; Roche, D.; Vuilleumier, S. A Methylotrophic Bacterium Growing with the Antidiabetic Drug Metformin as Its Sole Carbon, Nitrogen and Energy Source. *Microorganisms* **2022**, *10* (11), 2302. https://doi.org/10.3390/microorganisms10112302.

(111) Li, T.; Xu, Z. J.; Zhou, N. Y. Aerobic Degradation of the Antidiabetic Drug

Metformin by Aminobacter Sp. Strain NyZ550. *Environ. Sci. Technol.* **2022**, *57*, 1510–1519. https://doi.org/10.1021/acs.est.2c07669.

(112) Hillmann, K. B.; Niehaus, T. D. Genome Sequences of Two Pseudomonas Isolates That Can Use Metformin as the Sole Nitrogen Source. *Microbiol. Resour. Announc.* **2022**, *11* (9). https://doi.org/10.1128/mra.00639-22.

(113) Tassoulas, L. J.; Robinson, A.; Martinez-Vaz, B.; Aukema, K. G.; Wackett, L. P. Filling in the Gaps in Metformin Biodegradation: A New Enzyme and a Metabolic Pathway for Guanylurea. *Appl. Environ. Microbiol.* **2021**, *87* (11), 1–13. https://doi.org/10.1128/AEM.03003-20.

(114) Lin, M. C.; Lin, J. H.; Wen, K. C. Detection and Determination of Phenformin in Chinese Medicinal Capsules by GC-MS and HPLC. *J. Food Drug Anal.* **2001**, *9* (3), 139–144. https://doi.org/10.38212/2224-6614.2793.

(115) Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry Land Cell Constants. *J. Appl. Crystallogr.* **1993**, *26* (pt 6), 795–800. https://doi.org/10.1107/S0021889893005588.

(116) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2126–2132. https://doi.org/10.1107/S0907444904019158.

(117) Vagin, A.; Teplyakov, A. MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **1997**, *30* (6), 1022–1025. https://doi.org/10.1107/S0021889897006766.

(118) Winn, M. D.; Isupov, M. N.; Murshudov, G. N. Use of TLS Parameters to Model Anisotropic Displacements in Macromolecular Refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2001**, *57* (1), 122–133.

https://doi.org/10.1107/S0907444900014736.

(119) 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. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 355–367. https://doi.org/10.1107/S0907444911001314.

(120) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(121) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20 - A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60* (12), 6065–6073. https://doi.org/10.1021/acs.jcim.0c00675.

(122) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61* (8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.

(123) Sindhikara, D. J.; Roitberg, A. E.; Merz, K. M. Apo and Nickel-Bound Norms of the Pyrococcus Horikoshii Species of the Metalloregulatory Protein: NikR Characterized by Molecular Dynamics Simulations. *Biochemistry* **2009**, *48* (50), 12024–12033. https://doi.org/10.1021/bi9013352.

(124) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D.

S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.

(125) O'Fágáin, C.; Cummins, P. M.; O'Connor, B. F. Gel-Filtration Chromatography. *Methods in molecular biology (Clifton, N.J.)*. 2011. https://doi.org/10.1007/978-1-60761-913-0_2.

(126) Funck, D.; Sinn, M.; Fleming, J. R.; Stanoppi, M.; Dietrich, J.; López-Igual, R.; Mayans, O.; Hartig, J. S. Discovery of a Ni2+-Dependent Guanidine Hydrolase in Bacteria. *Nature* **2022**, *603* (7901), 515–521. https://doi.org/10.1038/s41586-022-04490-x.

(127) Sekowska, A.; Danchin, A.; Risler, J. L. Phylogeny of Related Functions: The Case of Polyamine Biosynthetic Enzymes. *Microbiology* **2000**, *146* (8), 1815–1828. https://doi.org/10.1099/00221287-146-8-1815.

(128) Khangulov, S. V.; Sossong, T. M.; Ash, D. E.; Dismukes, G. C. L-Arginine Binding to Liver Arginase Requires Proton Transfer to Gateway Residue His141 and Coordination of the Guanidinium Group to the Dimanganese(II,II) Center. *Biochemistry* **1998**, *37* (23), 8539–8550. https://doi.org/10.1021/bi972874c.

(129) Salas, M.; Rodríguez, R.; López, N.; Uribe, E.; López, V.; Carvajal, N. Insights into the Reaction Mechanism of Escherichia Coli Agmatinase by Site-Directed Mutagenesis and Molecular Modelling: A Critical Role for Aspartate 153. *Eur. J. Biochem.* **2002**, *269* (22), 5522–5526. https://doi.org/10.1046/j.1432-1033.2002.03255.x.

(130) Lewis, C. A.; Wolfenden, R. The Nonenzymatic Decomposition of Guanidines and Amidines. *J. Am. Chem. Soc.* **2014**, *136* (1), 130–136. https://doi.org/10.1021/ja411927k.

(131) Hyung, J. A.; Kyoung, H. K.; Lee, J.; Ha, J. Y.; Hyung, H. L.; Kim, D.; Yoon, H. J.; Kwon, A. R.; Se, W. S. Crystal Structure of Agmatinase Reveals Structural Conservation and Inhibition Mechanism of the Ureohydrolase Superfamily. *J. Biol. Chem.* **2004**, *279* (48), 50505–50513. https://doi.org/10.1074/jbc.M409246200.

(132) Dowling, D. P.; Di Costanzo, L.; Gennadios, H. A.; Christianson, D. W. Evolution of the Arginase Fold and Functional Diversity. *Cell. Mol. Life Sci.* **2008**, *65* (13), 2039–2055. https://doi.org/10.1007/s00018-008-7554-z.

(133) Cavalli, R. C.; Soprano, D. R.; Ash, D. E.; Burke, C. J.; Kawamoto, S. Mutagenesis of Rat Liver Arginase Expressed in Escherichia Coli: Role of Conserved Histidines. *Biochemistry* **1994**, *33* (35), 10652–10657. https://doi.org/10.1021/bi00201a012.

(134) Carvajal, N.; Orellana, M. S.; Salas, M.; Enríquez, P.; Alarcón, R.; Uribe, E.; López, V. Kinetic Studies and Site-Directed Mutagenesis of Escherichia Coli Agmatinase. a Role for Glu274 in Binding and Correct Positioning of the Substrate Guanidinium Group. *Arch. Biochem. Biophys.* **2004**, *430* (2), 185–190. https://doi.org/10.1016/j.abb.2004.07.005.

(135) Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D. S.; Milo, R. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **2011**, *50* (21), 4402–4410. https://doi.org/10.1021/bi2002289.

(136) Nattrass, M.; Alberti, K. G. M. M. Biguanides. *Diabetologia*. Springer-Verlag February 1978, pp 71–74. https://doi.org/10.1007/BF01263443.

(137) Kitada, Y.; Muramatsu, K.; Toju, H.; Kibe, R.; Benno, Y.; Kurihara, S.; Matsumoto, M. Bioactive Polyamine Production by a Novel Hybrid System

Comprising Multiple Indigenous Gut Bacterial Strategies. *Sci. Adv.* **2018**, *4* (6), 62–89. https://doi.org/10.1126/sciadv.aat0062.

(138) Pryor, R.; Norvaisas, P.; Marinos, G.; Best, L.; Thingholm, L. B.; Quintaneiro, L. M.; De Haes, W.; Esser, D.; Waschina, S.; Lujan, C.; Smith, R. L.; Scott, T. A.; Martinez-Martinez, D.; Woodward, O.; Bryson, K.; Laudes, M.; Lieb, W.; Houtkooper, R. H.; Franke, A.; Temmerman, L.; Bjedov, I.; Cochemé, H. M.; Kaleta, C.; Cabreiro, F. Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy. *Cell* **2019**, *178* (6), 1299-1312.e29. https://doi.org/10.1016/j.cell.2019.08.003.

(139) De Velasco, D. A. O. G.; Su, A.; Zhai, L.; Kinoshita, S.; Otani, Y.; Ohwada, T. Unexpected Resistance to Base-Catalyzed Hydrolysis of Nitrogen Pyramidal Amides Based on the 7-Azabicyclic[2.2.1]Heptane Scaffold. *Molecules* **2018**, *23* (9). https://doi.org/10.3390/molecules23092363.

(140) Yamada, S. Structure and Reactivity of a Highly Twisted Amide. *Angew. Chemie Int. Ed. English* **1993**, *32* (7), 1083–1085. https://doi.org/10.1002/anie.199310831.

(141) Irving, H.; Williams, R. J. P. Order of Stability of Metal Complexes. *Nature* **1948**, *162* (4123), 746–747. https://doi.org/10.1038/162746a0.

(142) Sydor, A. M.; Lebrette, H.; Ariyakumaran, R.; Cavazza, C.; Zamble, D. B. Relationship between Ni(II) and Zn(II) Coordination and Nucleotide Binding by the Helicobacter Pylori [NiFe]-Hydrogenase and Urease Maturation Factor HypB. *J. Biol. Chem.* **2014**, *289* (7), 3828–3841. https://doi.org/10.1074/jbc.M113.502781.

(143) Kaluarachchi, H.; Chan Chung, K. C.; Zamble, D. B. Microbial Nickel Proteins. *Natural Product Reports*. 2010, pp 681–694. https://doi.org/10.1039/b906688h.

(144) Higgins, K. A.; Carr, C. E.; Maroney, M. J. Specific Metal Recognition in Nickel Trafficking. *Biochemistry* **2012**, *51* (40), 7816–7832. https://doi.org/10.1021/bi300981m.

(145) Tisler, S.; Zwiener, C. Aerobic and Anaerobic Formation and Biodegradation of Guanyl Urea and Other Transformation Products of Metformin. *Water Res.* **2019**, *149*, 130–135. https://doi.org/10.1016/j.watres.2018.11.001.

(146) Collin, F.; Khoury, H.; Bonnefont-Rousselot, D.; Thérond, P.; Legrand, A.; Jore, D.; Gardès-Albert, M. Liquid Chromatographic/Electrospray Ionization Mass Spectrometric Identification of the Oxidation End-Products of Metformin in Aqueous Solutions. *J. Mass Spectrom.* **2004**, *39* (8). https://doi.org/10.1002/jms.656.

(147) Jia, Q.; Zeng, H.; Li, H.; Xiao, N.; Tang, J.; Gao, S.; Zhang, J.; Xie, W. The C-Terminal Loop of Arabidopsis Thaliana Guanosine Deaminase Is Essential to Catalysis. *Chem. Commun.* **2021**, *57* (76), 9748–9751. https://doi.org/10.1039/D1CC03042F.

(148) Burke, A. J.; Birmingham, W. R.; Zhuo, Y.; Thorpe, T. W.; Zucoloto da Costa, B.; Crawshaw, R.; Rowles, I.; Finnigan, J. D.; Young, C.; Holgate, G. M.; Muldowney, M. P.; Charnock, S. J.; Lovelock, S. L.; Turner, N. J.; Green, A. P. An Engineered Cytidine Deaminase for Biocatalytic Production of a Key Intermediate of the Covid-19 Antiviral Molnupiravir. *J. Am. Chem. Soc.* **2022**, *144* (9), 3761–3765. https://doi.org/10.1021/jacs.1c11048.

(149) Tzoulaki, I.; Molokhia, M.; Curcin, V.; Little, M. P.; Millett, C. J.; Ng, A.; Hughes, R. I.; Khunti, K.; Wilkins, M. R.; Majeed, A.; Elliott, P. Risk of Cardiovascular Disease and All Cause Mortality among Patients with Type 2 Diabetes Prescribed Oral Antidiabetes Drugs: Retrospective Cohort Study Using

UK General Practice Research Database. *BMJ* **2009**, *339* (7736). https://doi.org/10.1136/bmj.b4731.

(150) Bo, S.; Ciccone, G.; Rosato, R.; Villois, P.; Appendino, G.; Ghigo, E.; Grassi, G. Cancer Mortality Reduction and Metformin: A Retrospective Cohort Study in Type 2 Diabetic Patients. *Diabetes, Obes. Metab.* **2012**, *14* (1). https://doi.org/10.1111/j.1463-1326.2011.01480.x.

(151) Tassoulas, L. J.; Rankin, J. A.; Elias, M. H.; Wackett, L. P. An Ultra-Specific Di-Nickel Enzyme Evolved to Metabolize the Pharmaceutical Metformin: Impacts for Wastewater and Human Microbiomes. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, in review.

(152) Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. Complete Set of ORF Clones of Escherichia Coli ASKA Library (A Complete Set of E. Coli K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res.* **2005**, *12* (5). https://doi.org/10.1093/dnares/dsi012.

(153) Williams, S.; Shah, S.; Hakki, Z.; Stewart, A. Compounds for the Treatment of Respiratory Diseases. WO/2018/201192, November 8, 2018. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2018201192&redirectedID=true (accessed 2023-06-26).

(154) Geyer, J. W.; Dabich, D. Rapid Method for Determination of Arginase Activity in Tissue Homogenates. *Anal. Biochem.* **1971**, *39* (2), 412–417. https://doi.org/10.1016/0003-2697(71)90431-3.

(155) Neves, R. P. P.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Parameters for Molecular Dynamics Simulations of Manganese-Containing Metalloproteins. *J. Chem. Theory Comput.* **2013**, *9* (6), 2718–2732.

https://doi.org/10.1021/ct400055v.

(156) Almeida, A.; Nayfach, S.; Boland, M.; Strozzi, F.; Beracochea, M.; Shi, Z. J.; Pollard, K. S.; Sakharova, E.; Parks, D. H.; Hugenholtz, P.; Segata, N.; Kyrpides, N. C.; Finn, R. D. A Unified Catalog of 204,938 Reference Genomes from the Human Gut Microbiome. *Nat. Biotechnol.* **2021**, *39* (1). https://doi.org/10.1038/s41587-020-0603-3.

(157) Karlsson, F. H.; Tremaroli, V.; Nookaew, I.; Bergström, G.; Behre, C. J.; Fagerberg, B.; Nielsen, J.; Bäckhed, F. Gut Metagenome in European Women with Normal, Impaired and Diabetic Glucose Control. *Nat. 2013 4987452* **2013**, *498* (7452), 99–103. https://doi.org/10.1038/nature12198.

(158) Magnúsdóttir, S.; Heinken, A.; Kutt, L.; Ravcheev, D. A.; Bauer, E.; Noronha, A.; Greenhalgh, K.; Jäger, C.; Baginska, J.; Wilmes, P.; Fleming, R. M. T.; Thiele, I. Generation of Genome-Scale Metabolic Reconstructions for 773 Members of the Human Gut Microbiota. *Nat. Biotechnol.* **2017**, *35* (1). https://doi.org/10.1038/nbt.3703.

(159) Richardson, L.; Allen, B.; Baldi, G.; Beracochea, M.; Bileschi, M. L.; Burdett, T.; Burgin, J.; Caballero-Perez, J.; Cochrane, G.; Colwell, L. J.; Curtis, T.; Escobar-Zepeda, A.; Gurbich, T. A.; Kale, V.; Korobeynikov, A.; Raj, S.; Rogers, A. B.; Sakharova, E.; Sanchez, S.; Wilkinson, D. J.; Finn, R. D. MGnify: The Microbiome Sequence Data Analysis Resource in 2023. *Nucleic Acids Res.* **2023**, *51* (1 D). https://doi.org/10.1093/nar/gkac1080.

(160) Cama, E.; Colleluori, D. M.; Emig, F. A.; Shin, H.; Kim, S. W.; Kim, N. N.; Traish, A. M.; Ash, D. E.; Christianson, D. W. Human Arginase II: Crystal Structure and Physiological Role in Male and Female Sexual Arousal. *Biochemistry* **2003**, *42* (28), 8445–8451. https://doi.org/10.1021/bi034340j.

(161) Sinn, M.; Stanoppi, M.; Hauth, F.; Fleming, J. R.; Funck, D.; Mayans, O.; Hartig, J. S. Guanidino Acid Hydrolysis by the Human Enzyme Annotated as Agmatinase. *Sci. Rep.* **2022**, *12* (1), 22088. https://doi.org/10.1038/s41598-022-26655-4.

(162) Miller, D.; Xu, H.; White, R. H. A New Subfamily of Agmatinases Present in Methanogenic Archaea Is Fe(II) Dependent. *Biochemistry* **2012**, *51* (14), 3067–3078. https://doi.org/10.1021/bi300039f.

(163) Goda, S.; Sakuraba, H.; Kawarabayasi, Y.; Ohshima, T. The First Archaeal Agmatinase from Anaerobic Hyperthermophilic Archaeon Pyrococcus Horikoshii: Cloning, Expression, and Characterization. *Biochim. Biophys. Acta - Proteins Proteomics* **2005**, *1748* (1). https://doi.org/10.1016/j.bbapap.2004.12.010.

(164) Molderings, G. J.; Burian, M.; Homann, Jü.; Nilius, M.; Göthert, M. Potential Relevance of Agmatine as a Virulence Factor of Helicobacter Pylori. *Dig. Dis. Sci.* **1999**, *44* (12), 2397–2404. https://doi.org/10.1023/A:1026662316750/METRICS.

(165) Müller, H.; Reinwein, H. Zur Pharmakologie Des Galegins. *Naunyn. Schmiedebergs. Arch. Exp. Pathol. Pharmakol.* **1927**, *125* (3–4), 212–228. https://doi.org/10.1007/BF01862957.

(166) Pereira, F. C.; Berry, D. Microbial Nutrient Niches in the Gut. *Environmental Microbiology*. 2017. https://doi.org/10.1111/1462-2920.13659.

(167) Nakamura, A.; Ooga, T.; Matsumoto, M. Intestinal Luminal Putrescine Is Produced by Collective Biosynthetic Pathways of the Commensal Microbiome. *Gut Microbes* **2019**, *10* (2), 159–171. https://doi.org/10.1080/19490976.2018.1494466.

(168) Shimokawa, H.; Sakanaka, M.; Fujisawa, Y.; Ohta, H.; Sugiyama, Y.; Kurihara, S. N-Carbamoylputrescine Amidohydrolase of Bacteroides Thetaiotaomicron, a

Dominant Species of the Human Gut Microbiota. *Biomedicines* **2023**, *11* (4). https://doi.org/10.3390/biomedicines11041123.

(169) Llácer, J. L.; Polo, L. M.; Tavárez, S.; Alarcón, B.; Hilario, R.; Rubio, V. The Gene Cluster for Agmatine Catabolism of Enterococcus Faecalis: Study of Recombinant Putrescine Transcarbamylase and Agmatine Deiminase and a Snapshot of Agmatine Deiminase Catalyzing Its Reaction. In *Journal of Bacteriology*; 2007; Vol. 189. https://doi.org/10.1128/JB.01216-06.

(170) Evans, D. F.; Pye, G.; Bramley, R.; Clark, A. G.; Dyson, T. J.; Hardcastle, J. D. Measurement of Gastrointestinal PH Profiles in Normal Ambulant Human Subjects. *Gut* **1988**, *29* (8). https://doi.org/10.1136/gut.29.8.1035.

(171) Gabel, S. A.; Duff, M. R.; Pedersen, L. C.; DeRose, E. F.; Krahn, J. M.; Howell, E. E.; London, R. E. A Structural Basis for Biguanide Activity. *Biochemistry* **2017**, *56* (36). https://doi.org/10.1021/acs.biochem.7b00619.

(172) Madiraju, A. K.; Erion, D. M.; Rahimi, Y.; Zhang, X. M.; Braddock, D. T.; Albright, R. A.; Prigaro, B. J.; Wood, J. L.; Bhanot, S.; MacDonald, M. J.; Jurczak, M. J.; Camporez, J. P.; Lee, H. Y.; Cline, G. W.; Samuel, V. T.; Kibbey, R. G.; Shulman, G. I. Metformin Suppresses Gluconeogenesis by Inhibiting Mitochondrial Glycerophosphate Dehydrogenase. *Nature* **2014**, *510* (7506), 542–546. https://doi.org/10.1038/nature13270.

(173) Piletz, J. E.; Aricioglu, F.; Cheng, J. T.; Fairbanks, C. A.; Gilad, V. H.; Haenisch, B.; Halaris, A.; Hong, S.; Lee, J. E.; Li, J.; Liu, P.; Molderings, G. J.; Rodrigues, A. L. S.; Satriano, J.; Seong, G. J.; Wilcox, G.; Wu, N.; Gilad, G. M. Agmatine: Clinical Applications after 100 Years in Translation. *Drug Discovery Today*. 2013. https://doi.org/10.1016/j.drudis.2013.05.017.

(174) Kotagale, N.; Rahangdale, S.; Borkar, A.; Singh, K.; Ikhar, A.; Takale, N.;

Umekar, M.; Taksande, B. Possible Involvement of Agmatine in Neuropharmacological Actions of Metformin in Diabetic Mice. *Eur. J. Pharmacol.* **2021**, *907*, 174255. https://doi.org/10.1016/J.EJPHAR.2021.174255.

(175) Hwang, S. L.; Liu, I. M.; Tzeng, T. F.; Cheng, J. T. Activation of Imidazoline Receptors in Adrenal Gland to Lower Plasma Glucose in Streptozotocin-Induced Diabetic Rats. *Diabetologia* **2005**, *48* (4). https://doi.org/10.1007/s00125-005-1698-2.

(176) Li, G.; Regunathan, S.; Barrow, C. J.; Eshraghi, J.; Cooper, R.; Reis, D. J. Agmatine: An Endogenous Clonidine-Displacing Substance in the Brain. *Science (80-.).* **1994**, *263* (5149). https://doi.org/10.1126/science.7906055.

(177) Satriano, J.; Matsufuji, S.; Murakami, Y.; Lortie, M. J.; Schwartz, D.; Kelly, C. J.; Hayashi, S. I.; Blantz, R. C. Agmatine Suppresses Proliferation by Frameshift Induction of Antizyme and Attenuation of Cellular Polyamine Levels. *J. Biol. Chem.* **1998**, *273* (25). https://doi.org/10.1074/jbc.273.25.15313.

(178) Wolf, C.; Brüss, M.; Hänisch, B.; Göthert, M.; Von Kügelgen, I.; Molderings, G. J. Molecular Basis for the Antiproliferative Effect of Agmatine in Tumor Cells of Colonic, Hepatic, and Neuronal Origin. *Mol. Pharmacol.* **2007**, *71* (1). https://doi.org/10.1124/mol.106.028449.

(179) Isome, M.; Lortie, M. J.; Murakami, Y.; Parisi, E.; Matsufuji, S.; Satriano, J. The Antiproliferative Effects of Agmatine Correlate with the Rate of Cellular Proliferation. *Am. J. Physiol. - Cell Physiol.* **2007**, *293* (2). https://doi.org/10.1152/ajpcell.00084.2007.

(180) Winter, T. N.; Elmquist, W. F.; Fairbanks, C. A. OCT2 and MATE1 Provide Bidirectional Agmatine Transport. *Mol. Pharm.* **2011**, *8* (1). https://doi.org/10.1021/mp100180a.

(181) Chattopadhyay, M. K.; Tabor, C. W.; Tabor, H. Polyamines Are Not Required for Aerobic Growth of Escherichia Coli: Preparation of a Strain with Deletions in All of the Genes for Polyamine Biosynthesis. *J. Bacteriol.* **2009**, *191* (17). https://doi.org/10.1128/JB.00381-09.

(182) De La Cuesta-Zuluaga, J.; Mueller, N. T.; Corrales-Agudelo, V.; Velásquez-Mejía, E. P.; Carmona, J. A.; Abad, J. M.; Escobar, J. S. Metformin Is Associated with Higher Relative Abundance of Mucin-Degrading Akkermansia Muciniphila and Several Short-Chain Fatty Acid-Producing Microbiota in the Gut. *Diabetes Care* **2017**, *40* (1). https://doi.org/10.2337/dc16-1324.

(183) Bui, T. I.; Britt, E. A.; Muthukrishnan, G.; Gill, S. R. Probiotic Induced Synthesis of Microbiota Polyamine as a Nutraceutical for Metabolic Syndrome and Obesity-Related Type 2 Diabetes. *Frontiers in Endocrinology*. 2023. https://doi.org/10.3389/fendo.2022.1094258.

(184) Kulkarni, A.; Anderson, C. M.; Mirmira, R. G.; Tersey, S. A. Role of Polyamines and Hypusine in β Cells and Diabetes Pathogenesis. *Metabolites*. 2022. https://doi.org/10.3390/metabo12040344.

(185) Jones, J. E.; Causey, C. P.; Lovelace, L.; Knuckley, B.; Flick, H.; Lebioda, L.; Thompson, P. R. Characterization and Inactivation of an Agmatine Deiminase from Helicobacter Pylori. *Bioorg. Chem.* **2010**, *38* (2), 62–73. https://doi.org/10.1016/j.bioorg.2009.11.004.

(186) Rodriguez-R, L. M.; Konstantinidis, K. T. Estimating Coverage in Metagenomic Data Sets and Why It Matters. *ISME J.* **2014**, *8* (11), 2349–2351. https://doi.org/10.1038/ismej.2014.76.

(187) Quince, C.; Curtis, T. P.; Sloan, W. T. The Rational Exploration of Microbial Diversity. *ISME J.* **2008**, *2* (10), 997–1006. https://doi.org/10.1038/ismej.2008.69.

(188) Cena, J. A. de; Zhang, J.; Deng, D.; Damé-Teixeira, N.; Do, T. Low-Abundant Microorganisms: The Human Microbiome's Dark Matter, a Scoping Review. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 1. https://doi.org/10.3389/fcimb.2021.689197.

(189) Buse, J. B.; DeFronzo, R. A.; Rosenstock, J.; Kim, T.; Burns, C.; Skare, S.; Baron, A.; Fineman, M. The Primary Glucose-Lowering Effect of Metformin Resides in the Gut, Not the Circulation: Results from Short-Term Pharmacokinetic and 12-Week Dose-Ranging Studies. *Diabetes Care* **2016**, *39* (2), 198–205. https://doi.org/10.2337/dc15-0488.

(190) Paquette, S.; Thomas, S. C.; Venkataraman, K.; Appanna, V. D.; Tharmalingam, S. The Effects of Oral Probiotics on Type 2 Diabetes Mellitus (T2DM): A Clinical Trial Systematic Literature Review. *Nutr. 2023, Vol. 15, Page 4690* **2023**, *15* (21), 4690. https://doi.org/10.3390/NU15214690.

(191) Chavkin, T. A.; Pham, L. D.; Kostic, A. E. Coli Nissle 1917 Modulates Host Glucose Metabolism without Directly Acting on Glucose. *Sci. Rep.* **2021**, *11* (1). https://doi.org/10.1038/s41598-021-02431-8.

(192) Clayton, T. A.; Baker, D.; Lindon, J. C.; Everett, J. R.; Nicholson, J. K. Pharmacometabonomic Identification of a Significant Host-Microbiome Metabolic Interaction Affecting Human Drug Metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (34). https://doi.org/10.1073/pnas.0904489106.

(193) Barzilai, N.; Crandall, J. P.; Kritchevsky, S. B.; Espeland, M. A. Metformin as a Tool to Target Aging. *Cell Metab.* **2016**, *23* (6), 1060. https://doi.org/10.1016/J.CMET.2016.05.011.

(194) Nishihara, K.; Kanemori, M.; Kitagawa, M.; Yanagi, H.; Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE

and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in Escherichia Coli. *Appl. Environ. Microbiol.* **1998**, *64* (5), 1694–1699. https://doi.org/10.1128/aem.64.5.1694-1699.1998.

(195) Proctor, L. M.; Creasy, H. H.; Fettweis, J. M.; Lloyd-Price, J.; Mahurkar, A.; Zhou, W.; Buck, G. A.; Snyder, M. P.; Strauss, J. F.; Weinstock, G. M.; White, O.; Huttenhower, C. The Integrative Human Microbiome Project. *Nature* **2019**, *569* (7758), 641–648. https://doi.org/10.1038/s41586-019-1238-8.

Appendix A

Supplementary Information

A.1 Supplementary methods:

Cloning, expression and purification of ureohydrolase homologs

The agmatinase from *Pseudomonas aeruginosa* guanidinopropionase (GpuA, WP_003112934.1), *Enterococcus faecalis* agmatine deiminase (AgDI, WP_002363185.1), and human arginase 1 (ARG1, NP_000036.2) were codon-optimized and cloned into *Escherichia coli* BL21 DE3 cells (New England Biolabs) using a pET28 vector derivative with kanamycin resistance. The genes were cloned with a N-terminal a 6x His-tag and inserted, by Gibson assembly, into the multiple cloning site using the NdeI and HindIII restriction sites. The homologs, human arginase 2 (ARG2, NP_001163.1) and human guanidino acid hydrolase (AGMAT, NP_079034.3) were cloned as truncation variants for optimal expression in *E. coli* as the native sequences contain signal peptide sequences to direct their proteins to the mitochondria. Truncation variants ARG2 $\Delta_{1\text{-}24}/\Delta_{331\text{-}354}$ and AGMAT $\Delta_{1\text{-}35}$ were made as done in prior studies and codon-optimized, cloned in the same procedure as above[160,161].

The proteins were expressed by growing cells in terrific broth (TB) medium supplemented, with 0.5 mM $MnSO_4$ (for ARG1, GpuA, ARG2) or without metal (AgDI), and 50 µg/mL kanamycin at 37°C and 200 rpm to an OD_{600} of 0.6 in a shake flask. The culture was cooled to 16°C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and, with the same agitation, incubated for 20 hours. Cell pellets were harvested by centrifugation at 1,500 x g for 20 min and stored at −80°C.

The soluble expression of AGMAT Δ_{1-35} required the co-expression of GroES-GroEL chaperones using a co-transformed plasmid, pGro7 (Takara Bio) with chloramphenicol resistance and expression induced by arabinose[194]. AGMAT was expressed by growing cells in TB medium, supplemented with 0.5 mM $MnSO_4$, and 50 μg/mL kanamycin and 30 μg/mL chloramphenicol at 37°C and 200 rpm to an OD_{600} of 0.4 in a shake flask. The culture was cooled to 16°C and GroES-GroEL chaperones were induced with 0.005% (w/v) arabinose and one hour later, AGMAT was induced with 1 mM IPTG and, with the same agitation, incubated for 20 hours. Cell pellet of AGMAT was harvested as stored as done for the other proteins.

For purification, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 10 mM beta-mercapatoethanol pH 7.4).The cells were lysed using a French Press with three passes at 10,000 psi and the lysate then clarified by centrifugation at 20,000 x g for one hour. The proteins were purified from the lysate by using fast protein liquid chromatography (FPLC) and immobilized metal affinity chromatography (IMAC). Using a GE-AKTA FPLC and a GE HisTrap 5 mL column, proteins were purified after running an imidazole gradient from 25 mM to 500 mM and fractions collected. Pooled fractions from the FPLC were buffer exchanged into storage buffer (20 mM HEPES-NaOH, 200 mM NaCl pH 8) using a 15-mL Amicon 10 kDa centrifugal filter. Aliquots of concentrated protein were flash frozen in liquid nitrogen and stored at −80°C to be then used for kinetic assays. For metal reconstitution experiments, CbAGM enzyme was stripped of metal by incubating by buffer exchanging the enzyme with 1 mM 1,10-phenanthroline, 2 mM EDTA in storage buffer using a 15-mL Amicon 10 kDa centrifugal filter.

HPLC analysis for metformin transformation

A reversed-phase HPLC method to separate guanylurea from metformin was adapted from Lin et al. and detailed by Martinez-Vaz et al. which, in brief, used a C18 column and an isocratic mobile phase of 75:25 (v/v) acetonitrile:10 mM potassium phosphate buffer pH 6.6[109,114]. To determine if metformin could be transformed by enzymes EcAGM, CbAGM, GpuA, AGMAT, ARG1 and ARG2, 1 mM metformin was incubated with 10 μg of purified enzyme in 20 mM CHES pH 9 overnight and then assayed by HPLC. No transformation of metformin was seen for any of the enzymes tested.

Enzyme kinetics of agmatine deiminase

Agmatine deiminase activity by AgDI was measured by following ammonia production using a spectrophotometric, coupled-enzyme assay. Agmatine deiminase produces N-carbamoylputrescine and ammonia which the latter can be used with bovine liver L-Glutamic dehydrogenase (GDH) to reductively aminate 2-oxoglutarate to form L-glutamate, which causes NADH oxidation that can be measured by absorbance at 340 nm. For reactions, a 10X coupled enzyme assay master mix was prepared in 50 mM HEPES-NaOH pH 7.5 which had the following components and final concentrations: 0.3 mM NADH disodium salt (Sigma), 5 mM 2-oxoglutarate sodium salt (Aldrich), 0.8 mM adenosine diphosphate sodium salt (Sigma), 2.5 U/mL GDH from lyophilized powder (Sigma). The master mix was then diluted with buffer (150 mM HEPES-NaOH pH 7.5) and purified AgDI enzyme into wells of 96-well flat-bottom microplates and the reaction was initiated by adding substrate to make a total sample volume of 200 μL. The reactions were monitored, continuously, by absorbance at 340 nm using a microplate reader and initial rates were recorded. Rates of NADH oxidation were calculated assuming a pathlength of 0.56 cm and a molar extinction coefficient for NADH, at 340 nm, of 6220 $M^{-1}cm^{-1}$. Negative controls for the assays included no-enzyme and enzyme with no substrate that resulted in an unchanging amount of NADH in the assay over time.

A.2 Supplementary figures and tables:

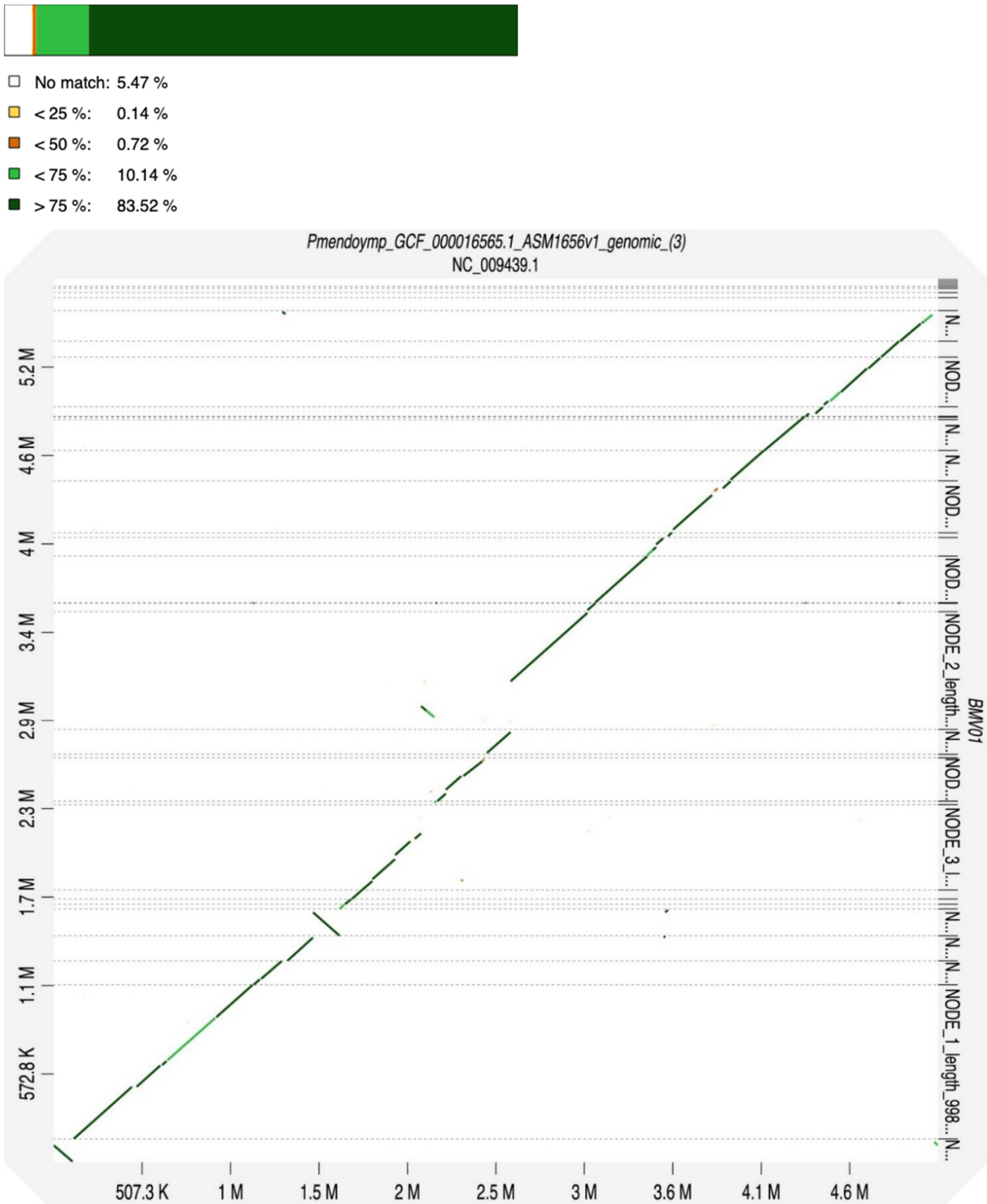


Figure A.1. Dot-plot comparing genes of *P. mendocina* strains GU and ymp.

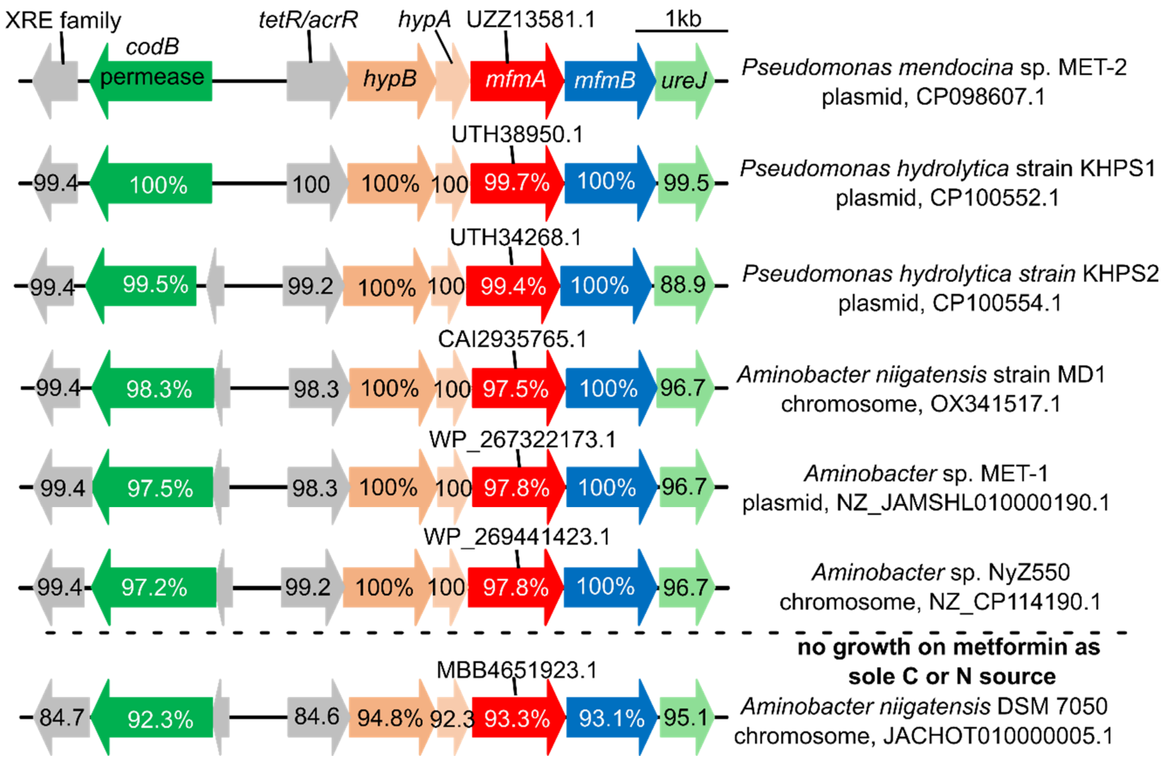


Figure A.2: Conservation of the metformin degrading operon in microbes characterized for growth on metformin

Metformin degradation has been characterized in six *Aminobacter* and *Pseudomonas* strains that have their genomes sequenced to date and show very similar genomic contexts around the *mfmA, mfmB, hypA, hypB* genes characterized in this study [109–112]. Along with these genes, there are several other genes that are co-occurring in all the strains: A putative metformin transporter (*codB*), nickel transporter (*ureJ*) and transcription factors (XRE and *tetR*) are highly conserved. The protein sequences of each gene were aligned to the corresponding, homologous sequence in *Pseudomonas mendocina* sp. MET-2 and the percent sequence identities are reported. Interestingly, several *Aminobacter* strains, including *Aminobacter niigatensis* DSM 7050 do not grow on metformin but contains the same set of genes as the other strains (see Figure S10) [110]. The DSM 7050 strain has ~93% seq. id. with the MET-2 strain while the other metformin degrading strains share >97% seq. id with MET-2. NCBI accession identifiers for MfmA protein sequences are labeled above the *mfmA* genes and the nucleotide accession IDs are labeled next to the strain names.

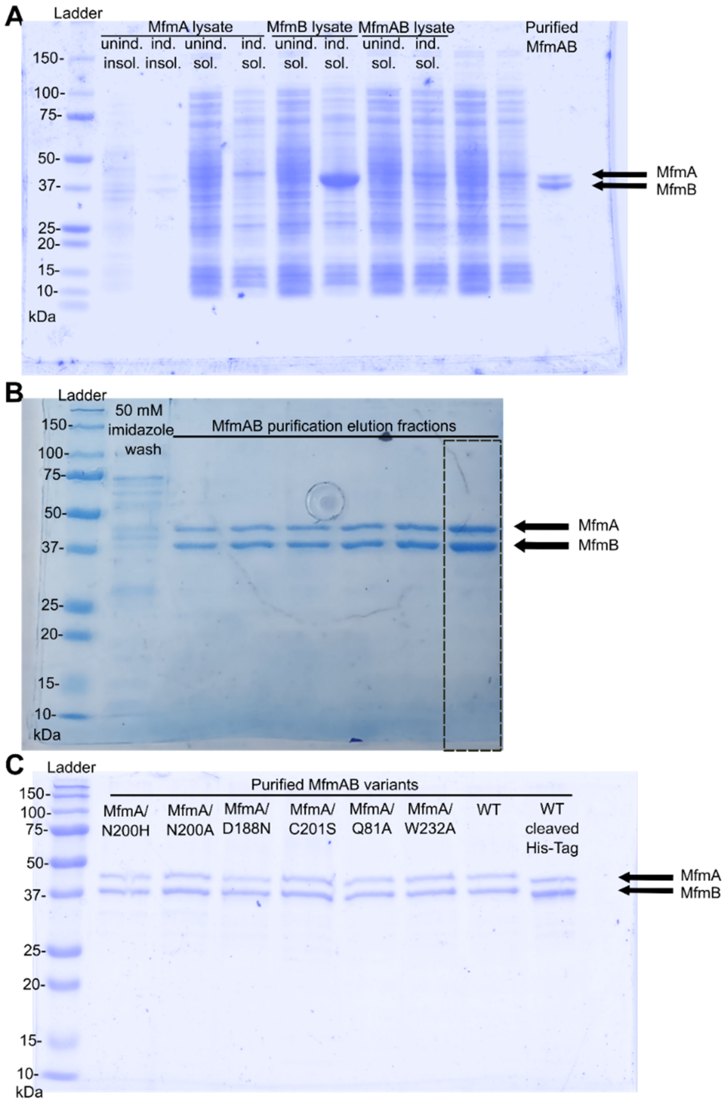


Figure A.3: Stained, denaturing polyacrylamide gels of MfmAB proteins

(A) Expression of MfmA and MfmB, individually and together in *E. coli* lysates. There was no significant expression of MfmA in the soluble (sol.) or insoluble (insol.) fractions of the lysate when comparing cultures induced (ind.) or uninduced (unind.) with IPTG. Significant, soluble expression of MfmB was observed in the lysate and slight over-expression was observed for the co-expressed MfmAB construct. (B) Purification of MfmAB using Ni-NTA chromatography. The gel lane labeled by a dashed box was the lane used in Fig. 1C of the main text. (C) Purification of MfmAB variants.

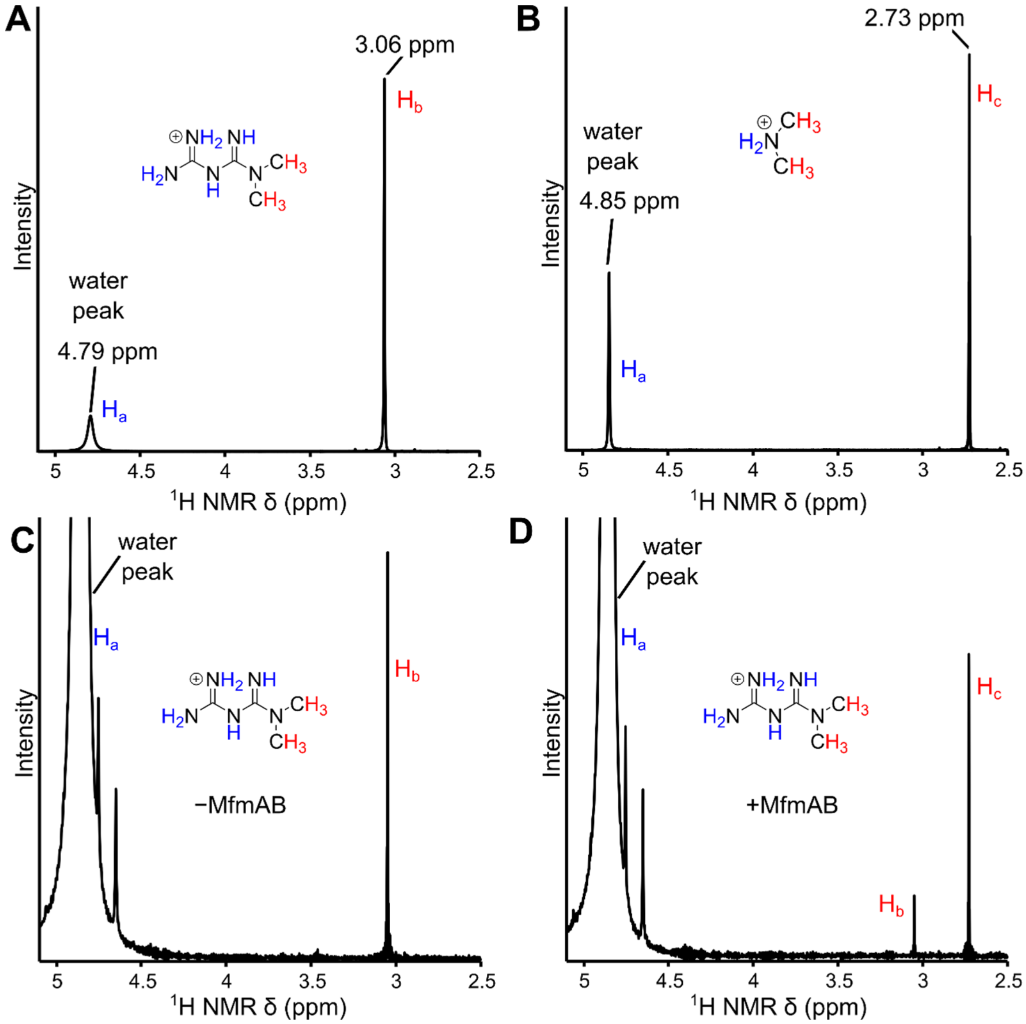


Figure A.4: ^{1}H NMR spectra of metformin, dimethylamine and transformation of metformin by MfmAB in D_2O. NMR spectra of metformin hydrochloride and dimethylamine hydrochloride in D_2O

Two shifts were observed for each compound, one corresponding to the water peak that also contain exchangeable protons (H_a) labeled in blue, and a second shift corresponding to methyl protons upfield (H_b and H_c) labeled in red. (A) Spectrum of metformin hydrochloride (50 mM) in 50 mM ammonium formate pH 8.5 buffer with 20% (v/v) D_2O. (B) Spectrum of dimethylamine hydrochloride (50 mM) in 50 mM ammonium formate pH 8.5 buffer with 20% (v/v) D_2O. (C-D) Spectrum of metformin hydrochloride (50 mM) in 50 mM ammonium formate, 200 mM NaCl pH 8.5, 1 mM $NiCl_2$ with 20% (v/v) D_2O before (−MfmAB) and after the addition of 200 µg purified metformin hydrolase (+MfmAB) after one hour.

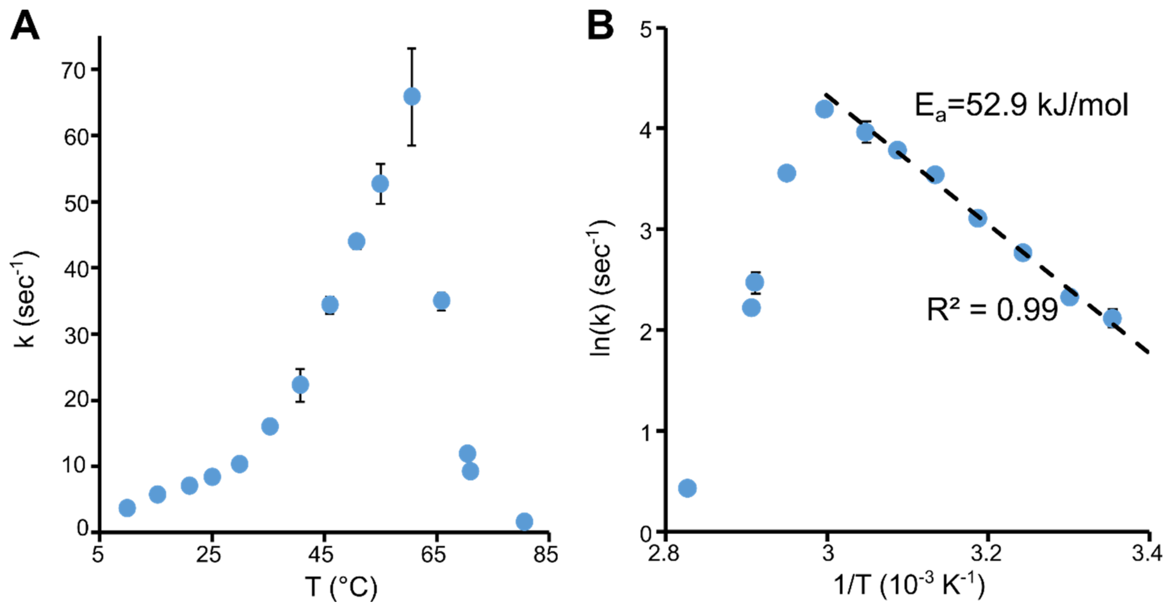


Figure A.5: Temperature dependence of activity of MfmAB working on metformin

(A) Temperature-Activity dependence of MfmAB. 15 mM metformin hydrochloride was incubated at various temperatures in 50 mM HEPES, 1 mM $NiCl_2$ pH 8 before purified MfmAB was added and the reaction sampled after one minute. (B) Arrhenius plot from temperature-activity data showing the linear fit and the estimated activation energy of metformin hydrolysis catalyzed by MfmAB. Error bars denote one standard deviation of the mean from averaging two technical replicates.

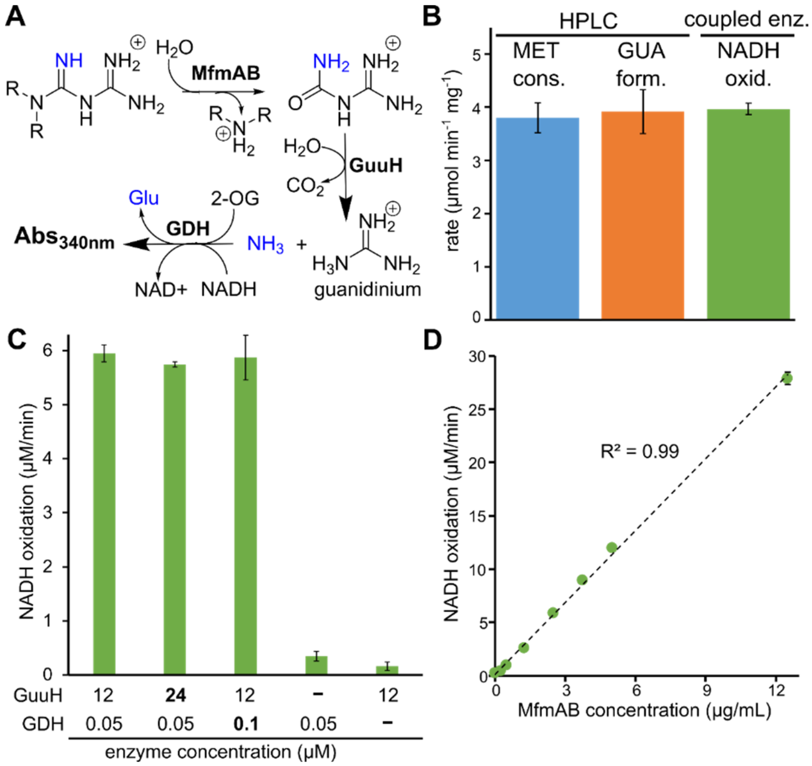


Figure A.6: Coupled enzyme assay development for measuring MfmAB kinetics

(A) Coupled enzymes used in continuous, spectrophotometric assay. Guanylurea produced by MfmAB during metformin hydrolysis, or from other biguanides, can be hydrolyzed by guanylurea hydrolase (GuuH) to form ammonia and guanidinium. The ammonia generated can then be used in reductive amination of 2-oxoglutarate to form glutamate by bovine glutamate dehydrogenase (GDH) which uses NADH as a cofactor. The oxidation of NADH in the GDH reaction can be measured, spectrophotometrically by absorbance at 340 nm. (B) Comparison of activity measurements by HPLC and the coupled enzyme assay. Using HPLC, rates of metformin consumption (cons.) and guanylurea formation (form.) can be observed in the presence of MfmAB and are comparable to the rate measured by the coupled enzyme assay. (C) Coupled enzyme assay development. The coupled enzymes, GuuH and GDH, were added in great excess, compared to MfmAB, to observe metformin hydrolysis, catalyzed by MfmAB. Adding an additional molar equivalent of GuuH or GDH, to the assay mixture used in this study, did not increase the reaction rate observed. Only adding GuuH or GDH individually saw a very slow rate that can be attributed to background NADH oxidation. (D) MfmAB enzyme loading is directly proportional to the activity observed in the coupled enzyme assay. Error bars denote one standard deviation of the mean from averaging two technical replicates.

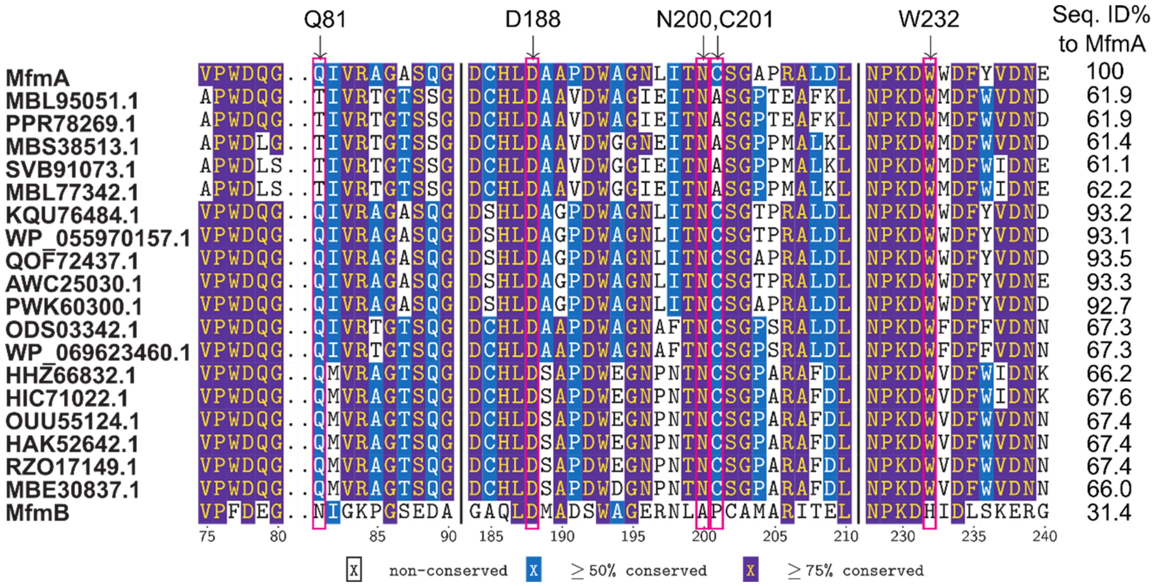


Figure A.7: Multiple sequence alignment of MfmA and homologous sequences with sequence identity greater than 60%

Obtaining a set of sequences using this sequence threshold (>60%) was considered with the assumption that these sequences could all share the same function while being divergent enough to highlight highly conserved active site residues. Positions in the alignment that are boxed in red were subject to mutagenesis in MfmA and these were chosen due to their high conservation in this set of sequences and also structurally they are located in the active site. Residue numbering is based on the sequence of MfmA.

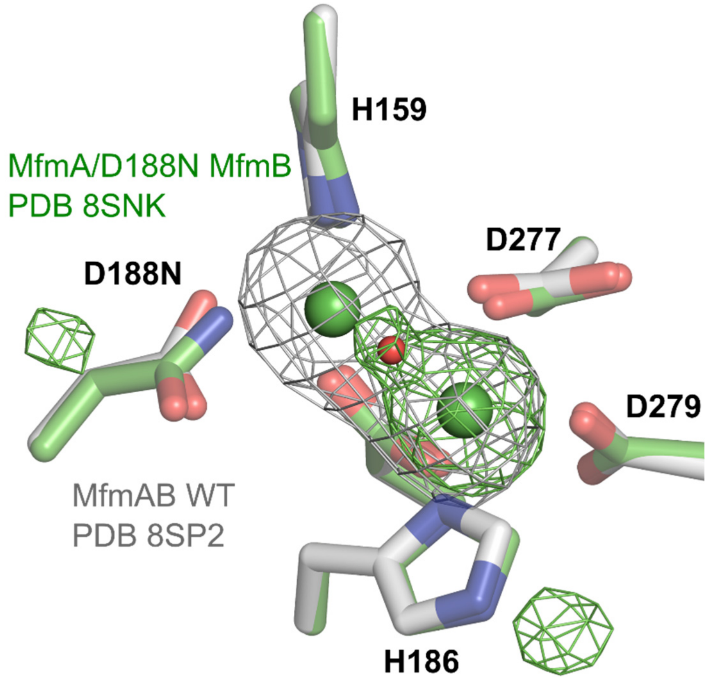


Figure A.8: Overlay of omit maps of the active site binuclear metals for MfmAB wildtype and the MfmA/D188N MfmB variant

F_o-F_c difference density maps contoured at 5σ shown as green and gray mesh for MfmA/D188N and MfmAB WT, respectively. The omit maps for the nickel atoms (shown as green spheres) show that the MfmA/D188N variant only has occupancy for one of the two metal binding sites. For the MfmAB wildtype, both sites are occupied. Additional density is observed between the metal binding sites and is modeled as a water molecule shown as a red sphere.

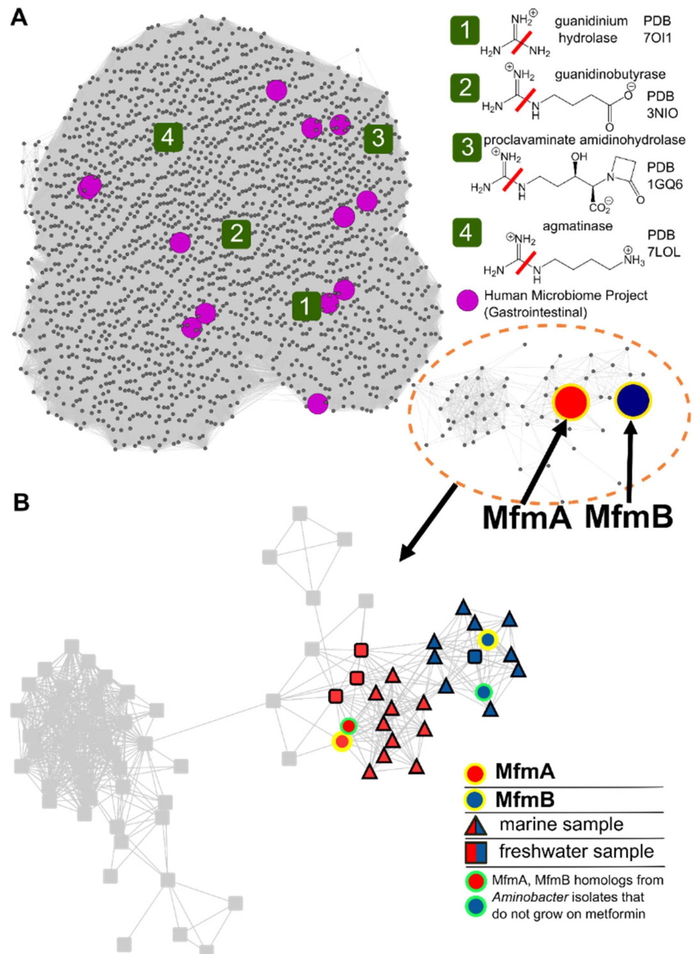


Figure A.9: Sequence similarity network (SSN) of protein sequences related to MfmA and MfmB from NCBI and UniProt databases

(A) SSN of 10,000 related sequences to MfmA from the ureohydrolase superfamily. MfmA and MfmB sequences cluster together and both are distantly related (<30% seq. id.) to characterized enzymes guanidinium hydrolase, guanidinobutyrase, proclavaminate amidinohydrolase and agmatinase in the ureohydrolase superfamily. Protein sequences found in the Human Microbiome Project's catalog of gastrointestinal microbes are also annotated [195]. SSN built using EFI-EST with a cutoff of E−65. (B) SSN of the MfmA and MfmB protein sequence clusters. Sequences in these clusters predominantly come from marine or freshwater samples. It is important to note that some of these sequences in the clusters of MfmA and MfmB come from *Aminobacter* isolates that do not grow on metformin as a carbon or nitrogen source. SSN built using EFI-EST with a cutoff of E−67.

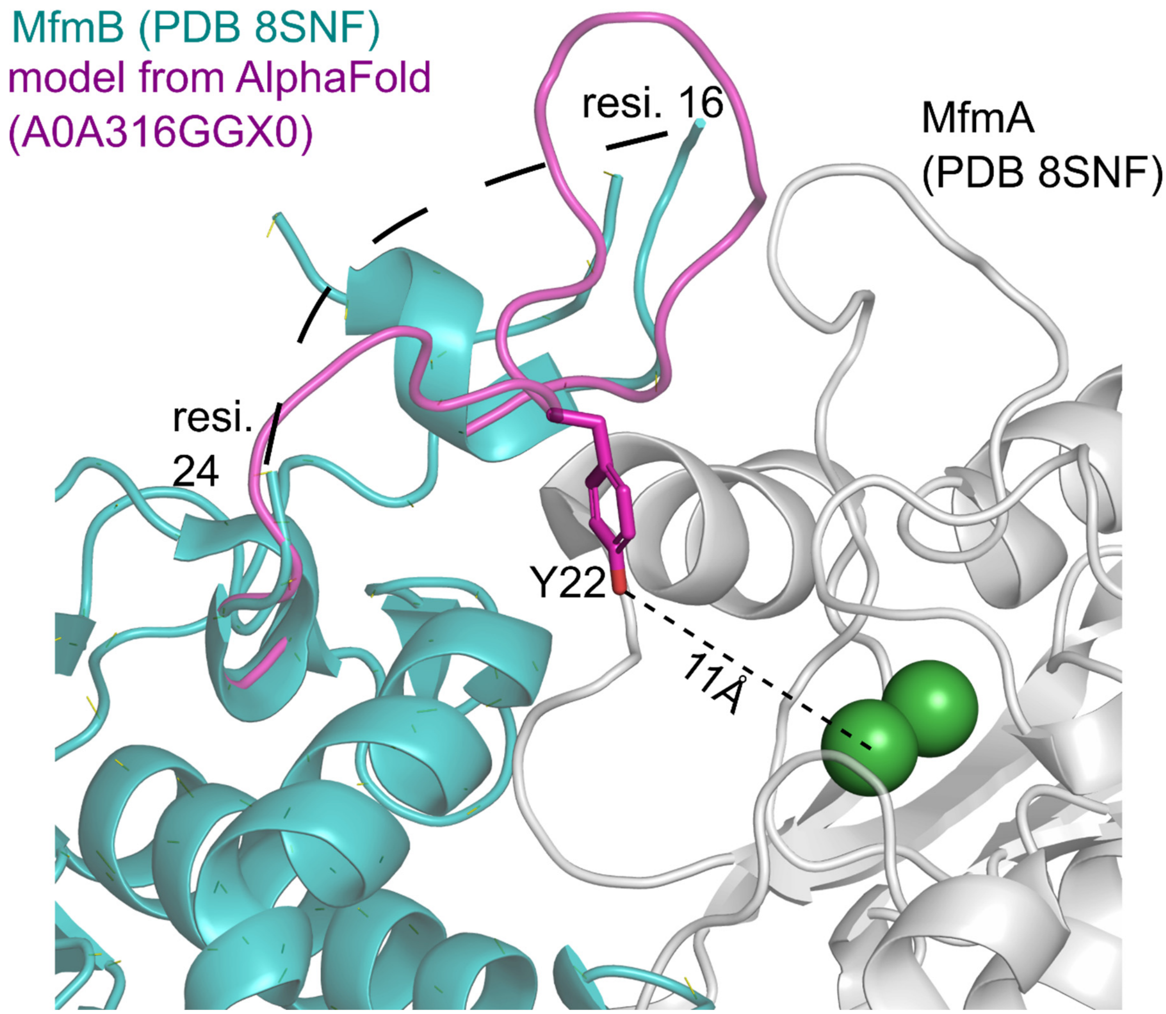


Figure A.10: Disordered N-terminal loop of MfmB may play a role in gating the active site of MfmA

Residues 16-24 of MfmB chains could not be resolved in x-ray crystal structures (black dashes). An AlphaFold model for this loop region considers the prediction to be low (50-60 pLDDT) and suggests that residues in this loop do not likely interact with the substrate in the MfmA active site [120]. Instead, this loop may be involved in the gating of substrates and products.

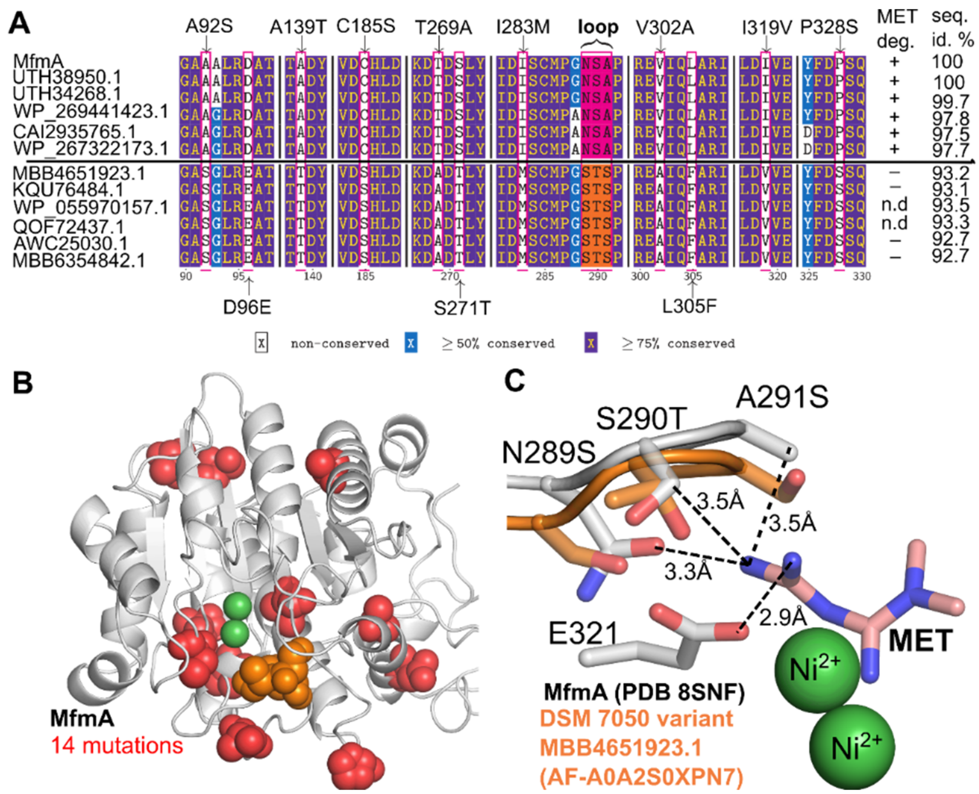


Figure A.11: Evolution of MfmAB from close homologs that are not active on metformin

(A) Multiple sequence alignment of MfmA with closely homologous sequences (>92% seq. id.) that are encoded in genomes from organisms that can grow (+) or not grow (−) on metformin as the sole carbon or nitrogen. The alignment reveals 14 residues that are different in MfmA compared to homologs that are likely not active on metformin. Substitutions on one active site loop (residues 289-291) stands out. Residue numbering is based on the sequence of MfmA. The names of the strains that did not grow on metformin as the sole carbon source were *Aminobacter niigatensis* DSM 7050 (MBB4651923.1), *Aminobacter* sp. Root100 (KQU76484.1), *Aminobacter niigatensis* MSH1 (AWC25030.1) and *Aminobacter aganoensis* DSM 7051 (MBB6354842.1) [110]. n.d – not determined. (B) Mutation positions between MfmA and putative homologs that are not active on metformin are globally distributed. The mutation positions are shown as red and orange spheres displayed onto MfmA. (C) Overlay of active site loops from MfmA and the putative homolog from *Aminobacter niigatensis* DSM7050. Three residue positions N289, S290 and A291 of MfmA may provide contacts within 3.5 Å of metformin, according to the docking model. These positions are mutated in the DSM 7050 variant that is predicted to not be active on metformin.

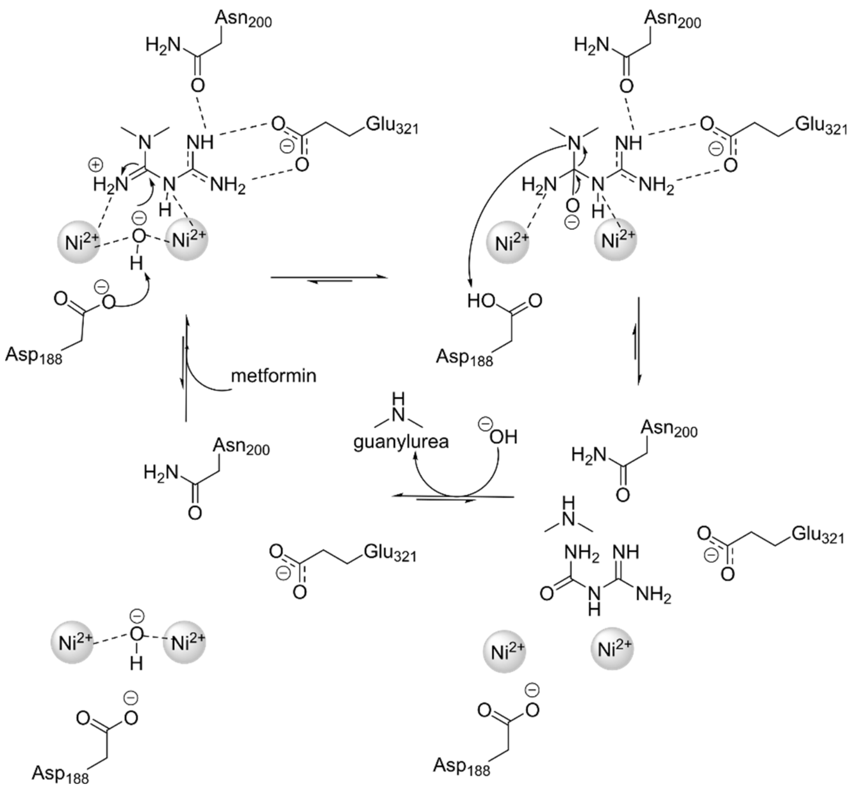


Figure A.12: Putative mechanism of metformin hydrolysis catalyzed by MfmAB

The catalytic residue, D188 of MfmAB, is proposed to abstract the proton from the hydroxide bound by the binuclear metals thereby activating hydroxide to attack metformin. The abstracted proton could then be involved in being transferred to the dimethylamine leaving group to promote its elimination, completing the enzyme turnover. The Michaelis complex is proposed to have metformin in a pre-attack conformation where the molecule is contorted to make it more prone to nucleophilic attack by water. This contortion could be stabilized by substrate binding residues E321 and N200 and the binuclear metals.

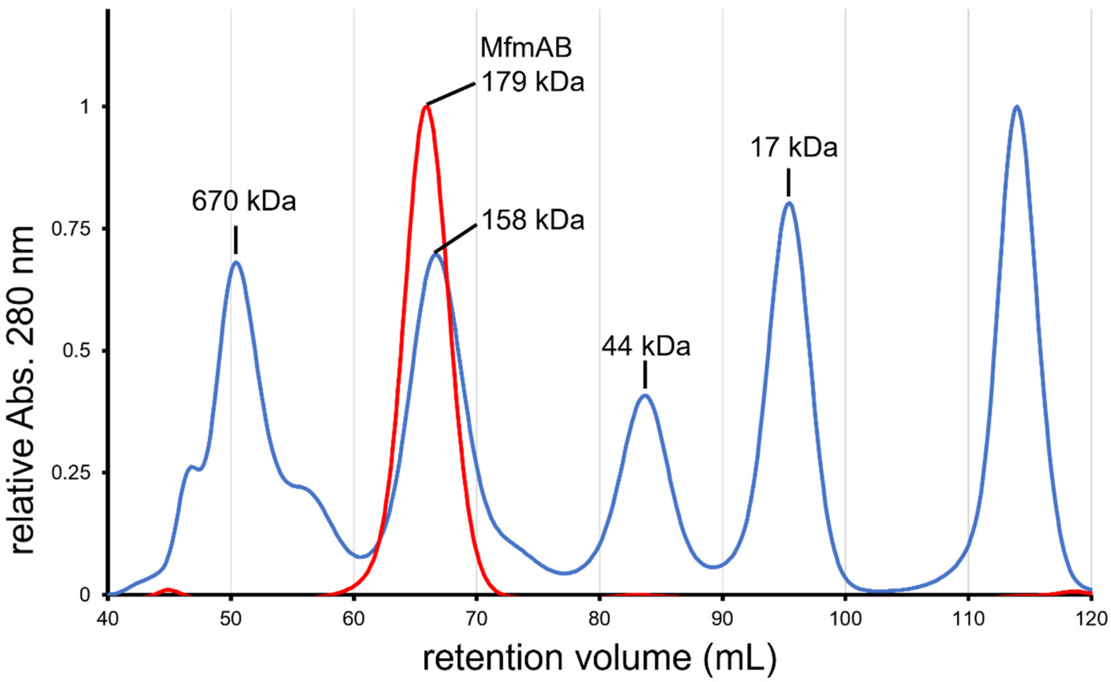


Figure A.13: Size exclusion chromatography of metformin hydrolase

Overlay of chromatograms of protein standards with known molecular weight (blue) and purified metformin hydrolase (red). Metformin hydrolase (MfmAB) eluted with a apparent molecular weight of 179 ± 3 kDa. The protein standards used for the calibration curve were bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa) and equine myoglobulin (17 kDa).

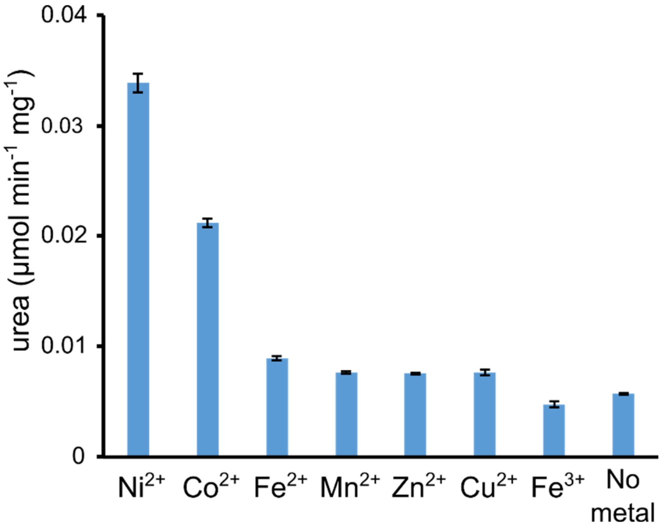


Figure A.14: Metal-activity dependence of CbAGM

Enzyme stripped of metal was reconstituted with 200 µM of various metals and indicates that the enzyme is most activated with Ni^{2+} and Co^{2+} cations. Activity was measured in the presence of 1 mM 2-mercaptoethanol in 100 mM HEPES-NaOH pH 7.5, 10 mM agmatine using the urea colorimetric assay. Error bars indicate one standard deviation of the mean from averaging of two technical replicates.

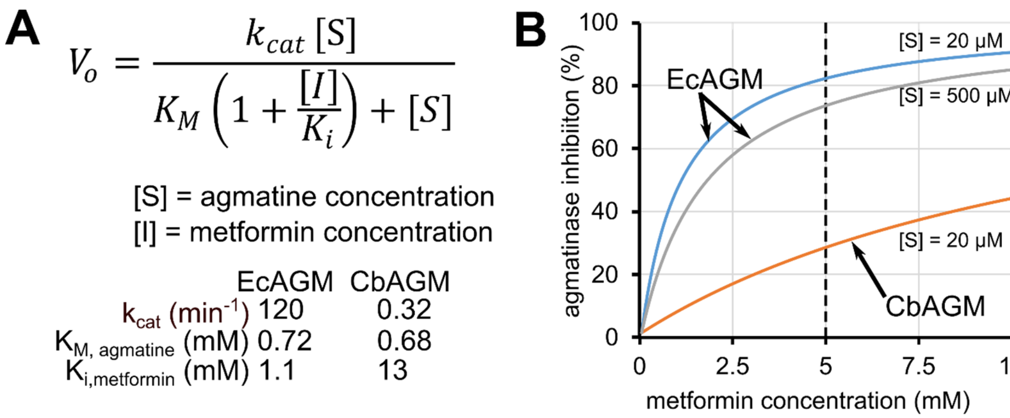


Figure A.15: Modelling of agmatinase inhibition assuming steady state kinetics

(A) Inhibition kinetics assuming steady state. (B) Agmatinase inhibition versus metformin concentration. Assuming a steady state concentration of 20 μM for agmatine, at 5 mM metformin EcAGM and CbAGM are inhibited by 83% and 27%, respectively.

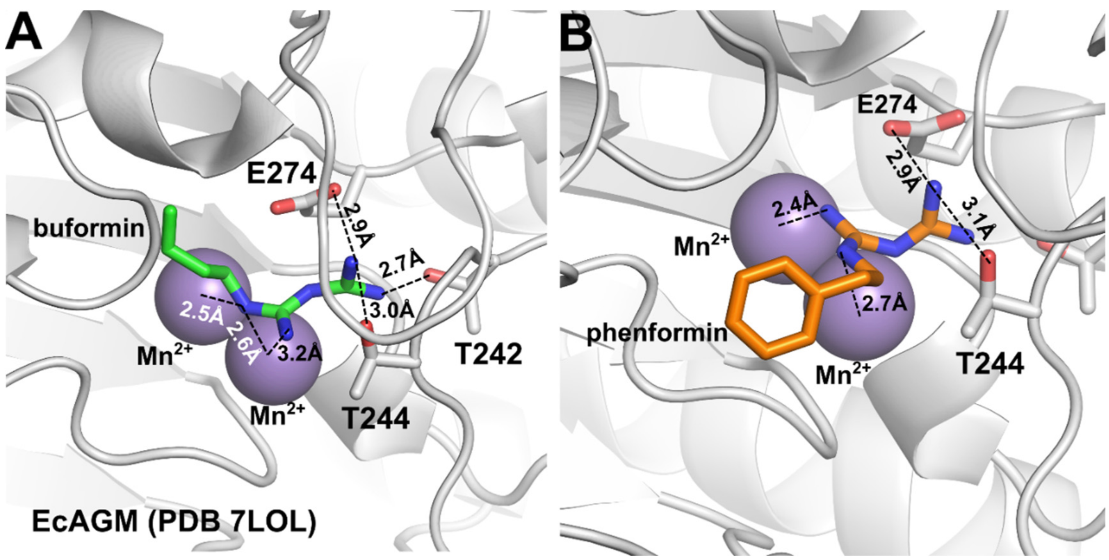


Figure A.16: Docking models of buformin and phenformin in the active site of EcAGM

(A,B) The models suggest that some of the inhibitor guanidinium atoms have polar contacts with residues T242, T244 and E274 of EcAGM and other guanidinium atoms chelate the binuclear manganese in the active site.

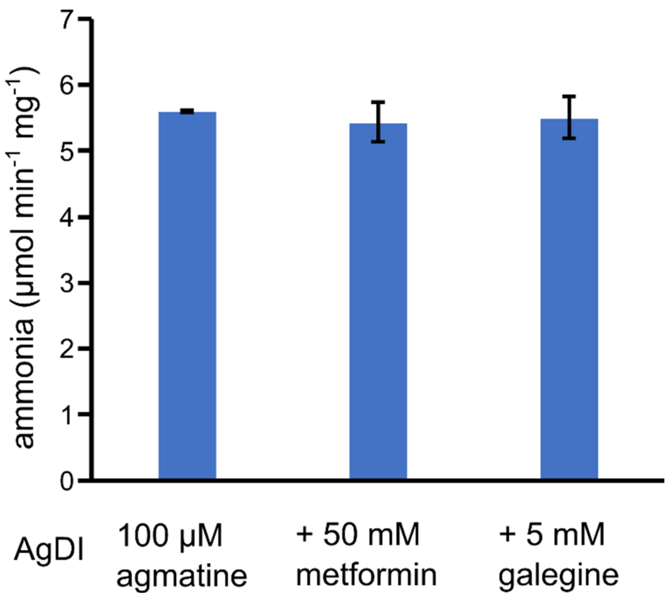


Figure A.17: Metformin and galegine do not inhibit agmatine deiminase from *Enterococcus faecalis*

Activity measured with purified AgDI in 50 mM HEPES-NaOH pH 7.5 using the GDH coupled enzyme assay. Error bars indicate one standard deviation of the mean from averaging of two technical replicates.

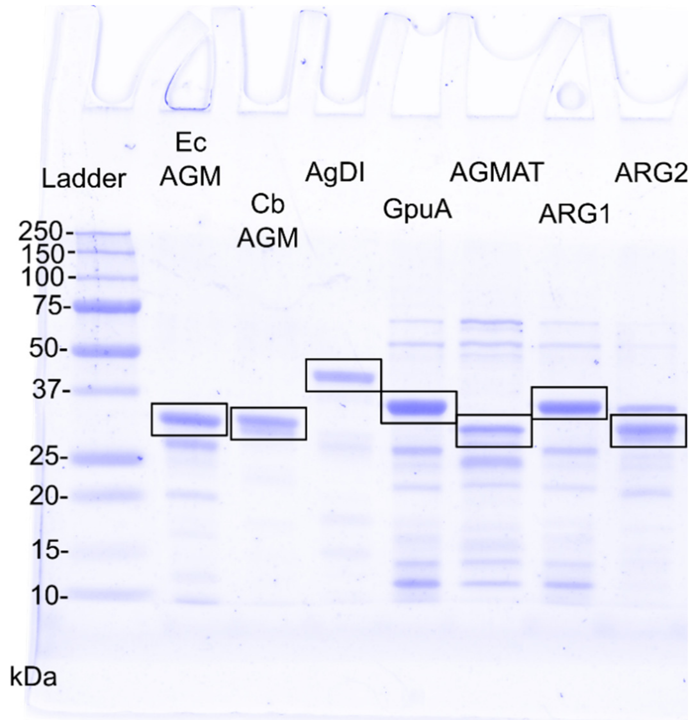


Figure A.18: Stained, denaturing polyacrylamide gel of proteins used in Chapter 6

Table A.1: Growth of *Pseudomonas mendocina* strain GU in different nitrogen rich compounds similar to guanylurea or containing the guanylurea moiety

compound	**generation time (min)**
guanidine	81
urea	92
N-free	no growth
metformin	no growth
agmatine	66
cyanoguanidine	no growth
guanylurea	44

Table A.2: Number of genes in common and sequence commonality between *P. mendocina* strains GU and ymp genome-encoded proteins

Number of genes in *P. mendocina* strain GU (out of 5378 predicted coding genes) similar to strain ymp	% Identity to *P. mendocina* corresponding gene in strain ymp	% of similar genes (out of 5378 predicted coding genes)
4175	≥ 95%	77.6%
106	90-94.9%	2.0%
41	80-89.9%	0.8%
24	70-79.9%	0.4%
24	60-69.9%	0.4%
40	50-59.9%	0.7%
77	40-49.9%	1.4%
165	30-39.9%	3.1%
61	20-29.9%	1.2%
665	Larger genome has no match in strain ymp	12.4%

Table A.3: Activity of biuret hydrolase, BiuH, measured by ammonia release (Berthelot reaction) of wild-type (WT) and Q212E mutant with biuret and guanylurea

Variant	biuret specific activity (U/mg)	guanylurea specific activity (U/mg)
Herb. sp. BH-1 BiuH WT	12±1	n.d
Herb. sp. BH-1 BiuH Q212E	0.25±0.01	0.35±0.02

n.d – not detected; 1U = 1 μmol substrate/min at pH 8.0 and 25˚C with 1 mM substrate

Table A.4: Substrate specificity of MfmAB

50 mM substrate	specific activity (μmol/min/mg)	specificity ratio
metformin	4.55±0.09	1
buformin	0.018±0.0007	3.9×10^{-3}
1,1-dimethylguanidine	0.014±0.0003	3.1×10^{-3}
methylguanidine	0.0040±0.00006	8.9×10^{-4}
phenformin	0.0039±0.0004	8.6×10^{-4}
1-methylbiguanide	0.0037±0.00026	8.1×10^{-4}
biguanide	0.0023±0.00047	5.0×10^{-4}
guanylurea	0.0011±0.00015	2.3×10^{-4}
arginine	0.00053±0.000048	1.2×10^{-4}
guanidine	0.00031±0.00017	6.7×10^{-5}
agmatine	0.00014±0.000014	3.1×10^{-5}
4-guanidinobutyric acid	n.d	
creatine	n.d	

n.d - not detected

Activity values are shown as the mean and one standard deviation from averaging two technical replicates.

Table A.5: Kinetic parameters of MfmAB and with His-tag cleaved

Enzyme	Substrate	K_M (μM)	k_{cat} (1/s)	k_{cat}/K_M ($M^{-1}s^{-1}$)
MfmAB	metformin	530±60	5.2±0.1	$9.8\pm1.8\times10^{3}$
MfmAB CleavedHis	metformin	486±43	5.9±0.1	$1.2\pm0.4\times10^{4}$

MfmAB - *Pseudomonas mendocina* sp. MET-2 metformin hydrolase

Table A.6: Summary of X-ray data collection and refinement for MfmAB

PDB	8SNF	8SP2	8SNK
Crystal	MfmAB + 0.25mM $NiCl_2$	MfmAB	MfmA/D188N MfmB
Data collection			
Beamline	23-ID-D	24-ID-E	23-ID-D
Wavelength ($\AA$)	1.033	0.979	1.033
Space Group	P1	P1	C2
Dimension a, b, c ($\AA$)	83.4, 96.4, 96.7	103.5, 107.9, 114.5	102.7, 162.3, 152.7
Dimension α, β, γ (^{o})	115.5, 106.2, 101.1	93.4, 97.6, 98.1	90.0, 101.1,90.0
Resolution ($\AA$)	$81.53-2.30(2.40-2.30)^a$	$74.61-2.20(2.30-2.20)^a$	$80.00-1.85(1.95-1.85)^a$
No. observed reflections	238631 (28597)	461309 (55575)	931929 (137582)
No. unique reflections	99100 (11503)	224389 (27288)	206743 (30240)
R_{meas} (%)	5.0 (34.4)	5.5 (83.6)	8.3 (52.7)
CC(1/2)	0.998 (0.960)	0.999 (0.745)	0.998 (0.937)
I/σ	13.30 (3.07)	12.75 (1.59)	10.49 (2.84)
Completeness (%)	91.3 (88.6)	91.2 (89.0)	99.2 (99.8)
Redundancy	2.41 (2.49)	2.06 (2.04)	4.51 (4.55)
Refinement			
Resolution ($\AA$)	2.30	2.20	1.85
R_{work}/R_{free}	0.1719/0.2194	0.2125/0.2570	0.1885/0.2197
No. atoms			
Protein	15458	31104	15438
Ligand/Ion	4	8	2
Water	330	910	1456
Mean B-factors ($\AA^2$)	78.0	50.9	29.0
R.m.s deviations			
Bond lengths ($\AA$)	0.007	0.005	0.009
Bond angles (^{o})	1.45	1.31	1.61
Ramachandran plot (%)			
Core	95.4	96.1	96.5
Allowed	4.3	3.4	3.3
Disallowed	0.4	0.4	0.2

[a] Numbers in parentheses refer to the highest-resolution shell

Table A.7: Kinetic parameters of CbAGM-like agmatinase enzymes

agmatinase	metal	pH	T(°C)	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	ref
C. butyricum	Ni^{2+}	7.5	25	0.68±0.13	0.32±0.01	7.8	this study
M. jannaschii	Fe^{2+}	9.3	70	2.4±1.2	34±5.2	2.3×10^3	Miller et al
		7.5	25		<2 (est.)		estimate
P. horikoshii	Co^{2+}	11	50	0.53	46.5	1.5×10^4	Goda et al
		7.5	25		<2 (est.)		estimate

references: Miller et al[162]; Goda et al[163]

Table A.8: Metformin and galegine inhibition of *E. coli* agmatinase at various pH

pH	K_M (mM)	k_{cat} (min^{-1})	K_i (mM) metformin	K_i (mM) galegine
6.5	4.96±0.59	75.1±0.9	1.18±0.15	0.00803±0.00593
7.5	0.718±0.089	120±3	1.14±0.12	0.00748±0.00212
8.5	0.148±0.019	180±12	1.25±0.10	0.00859±0.00346