Canavanine utilization and detoxification in bacteria

Doctoral thesis for obtaining the

academic degree

Doctor of Natural Sciences

(Dr. rer. nat.)

submitted by

Franziskus Nikolaus Michael Hauth

at the

Faculty of Sciences

Department of Chemistry

Konstanz, 2022

Date of Defense: 05.05.2023

1. Referee: Prof. Dr. Jörg Hartig

2. Referee: Prof. Dr. David Schleheck

3. Referee: Prof. Dr. Daniel Summerer

Notes to the reader

This doctoral thesis is divided into 5 chapters. Chapter 1 gives an introduction to the topic and provides background and general knowledge relevant to the presented work. Chapter 2 introduces the main part of the work and covers three subchapters. First, chapter 2.1 presents the discovery and description of *Pseudomonas canavaninivorans*, a bacterium that is able to grow on canavanine. In chapter 2.2 the molecular pathway by which the bacterium is able to utilize canavanine is explained. Finally, chapter 2.3 describes how the bacterium circumvents canavanine toxicity in the context of translation. In chapter 3 guanidine riboswitch-associated genes and their role in guanidine metabolism are addressed. All chapters include a short introduction, a result and a discussion or conclusion section. An overall conclusion and an outlook is given in chapter 4. The material and method part finalizes the work in a separate chapter 5.

All experiments described were conducted in the laboratory of Professor Dr. Jörg Hartig (Department of Chemistry, University of Konstanz) or in the laboratories of the stated collaborators from March 2019 to October 2022.

The thesis contains content that has already been published or is submitted for publication. Chapter 1 contains content from all papers mentioned hereafter. Chapters 2.1 and 2.2 are mainly based on the published papers "*Pseudomonas canavaninivorans sp. nov., isolated from bean rhizosphere*" (Hauth et al., Int. J. Syst. Evol. Microbiol. 2022; 72:005203, DOI 10.1099/ijsem.0.005203) and "*Canavanine utilization via homoserine and hydroxyguanidine by a PLP-dependent γ-lyase in Pseudomonadaceae and Rhizobiales"* (Hauth et al., RSC Chemical Biology 2022, DOI: 10.1039/D2CB00128D). Chapter 2.3 and parts of chapter 3 represent work that is published in *"A standalone editing protein deacylates mischarged canavanyl-tRNA^{Arg} to prevent canavanine incorporation into proteins"* (Hauth et al., Nucleic Acids Research 2023, DOI: 10.1093/nar/gkac1197). Other parts of chapter 3 have been published in "*Widespread bacterial utilization of guanidine as nitrogen source*" (Sinn & Hauth et al., Molecular microbiology, 2021; DOI: 10.1111/mmi.14702) and "*A variant of guanidine-IV riboswitches exhibits evidence of a distinct ligand specificity"* (Lenkeit et al. RNA biology 2023, DOI: 10.1080/15476286.2022.2160562) or have been submitted for publication in "*Crystal structure of a GCN5 related N-Acetyltransferase from Lactobacillus curiae" (Hauth & Fleming et al.,* Acta Crystallographica Section F, submitted.*)*. Consent has been given by the authors and the journals to reprint text and figures.

The experimental part of "*Widespread bacterial utilization of guanidine as nitrogen source*" was shared work done together with Malte Sinn and Jörg Hartig (University of Konstanz), while the major part of the bioinformatic analysis was conducted by Zasha Weinberg (University of Leipzig).

The acetyltransferase crystallization experiments were performed in collaboration with Jennifer Fleming and Olga Mayans (University of Konstanz). I overexpressed and purified protein. Crystal trials were set up in joined work with Jennifer Fleming. The crystallographic model was build by Jennifer Fleming and Olga Mayans.

Additional files like excel tables of proteome data can be found online in the supplementary material of the stated publications. The raw data of the crystallographic model can be obtained upon justified request.

Table of contents

List of Tables X
List of Schemes X
List of Figures XI

Zusammenfassung 1

Summary 2

Chapter 1: Introduction 3
The central dogma of biology 3
Gene expression in bacteria 4
Ensuring faithful translation 5
Gene expression regulation mechanisms 8
General aspects of bacterial metabolism 10
Canavanine 11

Chapter 2: Canavanine utilization and detoxification in bacteria 13
Chapter 2.1: Discovery and taxonomic description of *Pseudomonas canavaninivorans* 13
Introduction 13
Results 14
Enrichment of canavanine-degrading bacteria 14
Discovery of Pseudomonas canavaninivorans 15
Description of Pseudomonas canavaninivorans sp. nov. 19
Conclusion 20
Chapter 2.2: Canavanine utilization in rhizosphere associated bacteria 21
Introduction 21
Results 21
Analysis of canavanine-dependent gene expression 21
Characterization and reaction mechanism of canavanine- 23
Characterization of homoserine-dehydrogenase, aldehyde-dehydrogenases and ammonium-aspartate-lyase 27
Distribution of canvanine -γ-lyase in prokaryotic genomes 28
Discussion 30

Chapter 2.3: Canavanine detoxification in the context of translation 33
Introduction 33
Results 33
ArgRS does not discriminate efficiently against canavanine 33
Comparative proteomics reveal the upregulation of a standalone B3/4 editing domain-like protein upon canavanine stimulation 35

B3/4 protein prevents incorporation of canavanine into the proteome 37
Discussion 38

Chapter 3: Guanidine riboswitch-associated gene functions 40
General Introduction 40
Chapter 3.1: A guanidine riboswitch-associated carboxylase pathway enables the utilization of guanidine as sole nitrogen source in bacteria 41
Introduction 41
Results 42
Gene functions under control of guanidine riboswitches 42
Guanidine riboswitch-associated ABC transporters could function as importers 43
Enrichment of guanidine utilizers 44
Analysis of guanidine-dependent gene expression 45
Characterization of carboxylase enzymes 47
Distribution of urea and guanidine carboxylase enzymes 49
Distribution of guanidine-utilizing carboxylases in metagenomes/habitats 49
Discussion 51
Chapter 3.2: Beyond nutrition – an additional role of guanidine in nature 53
Introduction 53
Results 54
Discussion 58

Chapter 4: Conclusions 60

Chapter 5: Material and Methods 63
Material 63
Enzymes 66
Kits and size ruler 67
Laboratory Equipment 68
Software 69
Consumables 69
Methods 70
General methods 70
Methods Chapter 2.1: Discovery and taxonomic description of Pseudomonas canavaninivorans 73
Methods Chapter 2.2: Canavanine utilization in rhizosphere associated bacteria 75
Methods Chapter 2.3: Canavanine detoxification in the context of translation 79
Methods Chapter 3.1: A guanidine riboswitch-associated carboxylase pathway enables the utilization of guanidine as sole nitrogen source in bacteria 81
Methods Chapter 3.2: Beyond nutrition – an additional role of guanidine in nature 84

References 86

Data availability statement 95
Abbreviations 124

Danksagung 126

List of Tables

Table 1: *Classification of the aaRS into two classes. Adapted from Rubio et al. (2020) 8*
Table 2*: Digital DNA-DNA hybridization calculations for HB002^T and closely related type strains..... 17*
Table 3: *TrueBac ID results showing the average nucleotide identity and the sequence similarity of the respective genes of HB002^T with closely related species. 17*
Table 4*: Differential characteristic phenotype of P. canavaninivorans and closely related species..... 18*
Table 5: *Guanidine Riboswitches in Isolate Genomes 46*
Table 6: *Michaelis Menten kinetic parameters of Uca and Gca.. ..... 48*
Table 7: *Buffers and solutions. 63*
Table 8: *Composition of media and agars 65*
Table 9: *Stock solutions of antibiotics and compounds. 66*
Table 10: *Enzymes and corresponding buffers. 66*
Table 11: *Kits and size ruler. 67*
Table 12: *Phusion PCR reaction setup. 71*
Table 13: *Thermo cycler program for Phusion DNA polymerase. 71*

List of Schemes

Scheme 1: *Crick's version of the „central dogma" of biology as a flow of information.. ..... 4*
Scheme 2: *Simplified overview of bacterial metabolism. 11*
Scheme 3: *A potential reaction mechanism of canavanine-γ-lyase 24*
Scheme 4: *Scheme of the proposed activity of CtdA.. ..... 39*
Scheme 5*: Bacterial canavanine and guanidine metabolism elucidated within this thesis.. ..... 61*

List of Figures

Figure 1: *The aminoacylation reaction.* 6
Figure 2: *Growth of isolate HB002 on canavanine.* 14
Figure 3*: Phylogenetic tree based on 16S rDNA gene.* 16
Figure 4: *Phylogenetic tree based on genome sequences.* 16
Figure 5: *Genomic loci of induced proteins in P. canavaninivorans* 22
Figure 6: *Aminotransferase assays.* 22
Figure 7: *NADH production by either wt or K213A CanγL in the HD coupled assay* 25
Figure 8: *UV/Vis spectra of 50 µM wt or K213A CanγL after the addition of 10 mM canavanine.* 26
Figure 9: *^{1}H-NMR of the reaction of homoserine to fumarate via homoserine dehydrogenase (HD), aldehyde dehydrogenases (AH) and ammonium-aspartate-lyase (AAL).* 27
Figure 10: *Novel canavanine degradation pathway in P. canavaninivorans.* 28
Figure 11: *Phylogenetic distribution of CanγL based on a homology search.* 29
Figure 12: *Enzymatic assays with rhizobial canavanine-γ-lyase.* 29
Figure 13: *Substrate specificity of the arginine-tRNA-synthetase of P. canavaninivorans* 34
Figure 14: *Representative radio screen of TLC separated 3' ^{32}P labelled (aminoacylated) AMP.* 34
Figure 15: *Position of the B3/4 protein in the genome of P. canavaninivorans and R. leguminosarum relative to the canavanine degradation operon.* 35
Figure 16: *Deacylase activity of the B3/4 protein.* 36
Figure 17*: Relevance of canavanyl-tRNA^{Arg}-editing in vivo.* 38
Figure 18*: Association of guanidine-I, -II, -III and -IV riboswitch classes with specific gene functions.* 43
Figure 19: *Bacterial utilization of guanidine as nitrogen source.* 44
Figure 20: *Scheme of the investigated guanidine assimilation pathway* 46
Figure 21: *Substrate specificity of the two urea/guanidine carboxylase enzymes* 48
Figure 22: *Occurrence of Gdx^{P} and Gca^{P} activities regulated by guanidine riboswitches* 50
Figure 23: *Genes frequently associated with GuaRS-IV/ GGAM-1 RNAs.* 53
Figure 24*: Riboswitch associated CtdA and Galactosidase assay* 55
Figure 25*: DTNB assay with purified GNAT from L. curiae.* 55
Figure 26: *Structure of the GNAT of L. curiae derived by crystallography.* 57

Zusammenfassung

Die stickstoffreiche Verbindung Guanidin ist in der Natur weit verbreitet. Eine Vielzahl von Molekülen enthält eine Guanidinogruppe, von Aminosäuren wie Arginin und Canavanin bis hin zur Nukleobase Guanin und Sekundärmetaboliten wie Streptomycin. Guanidin wird von mindestens vier verschiedenen Klassen von Riboswitchen erkannt, die in Bakterien weit verbreitet sind. Es gibt jedoch nur wenige Erkenntnisse über die Quelle und die physiologische Rolle von Guanidin in der Natur. Im Rahmen dieser Arbeit wurde Canavanin, das δ-oxa-Analogon von Arginin, als Quelle für Guanidin untersucht. Dabei wurde ein neuartiger Canavanin-abbauender *Pseudomonade* isoliert, der aufgrund seiner Fähigkeit, auf Canavanin als einziger Kohlenstoff- und Stickstoffquelle zu wachsen, den Namen *Pseudomonas canavaninivorans* erhielt. In dieser Arbeit wird das Bakterium charakterisiert und auf seine spezifischen phänotypischen Merkmale untersucht, die es von eng verwandten Stämmen unterscheiden. Desweiteren wird dargestellt, wie das Bakterium die Toxizität von Canavanin umgeht, die auf dem Einbau in Proteine und die Arginase-vermittelte Hydrolyse zu Canalin beruht. Der erste spezifische Canavanin-Abbauweg in Bakterien wird im Detail beschrieben. Das Schlüsselenzym, eine PLP-abhängige Canavanin-Lyase, wird charakterisiert und weitere Enzymaktivitäten, die die Canavanin-Verwertung ermöglichen, werden aufgeklärt. Darüber hinaus werden die Verbreitung des neuen Stoffwechselwegs und seine Bedeutung in der Natur diskutiert. Außerdem wird dargestellt, wie *P. canavaninivorans* eine fehlerfreie Translation in der Gegenwart von Canavanin erreicht. Die Arginin-tRNA-Synthetase des Bakteriums ist nicht in der Lage, zwischen Arginin und Canavanin als Substraten zu unterscheiden und verwendet daher ein eigenständiges spezifisches canavanyl-tRNA^{Arg}-hydrolisierendes Protein, um sich vor Fehleinbau zu schützen. Wir schlagen daher Canavanyl-tRNA^{Arg}-Deacylase (CtdA) als systematischen Namen für das Protein vor. Diese Ergebnisse sind die Ersten, die eine Editing-Aktivität gegenüber fehlbeladener tRNA^{Arg} zeigen und geben Beispiel dafür, wie in der Natur korrekte Translation gewährleistet wird. Abschließend werden Guanidin-Riboswitch assoziierte Genfunktionen aufgeklärt, um Einblicke in die Physiologie und den Zweck des Moleküls zu erhalten. Es wird nicht nur gezeigt, dass Guanidin als Stickstoffquelle genutzt werden kann, sondern auch, dass Guanidin-Riboswitches häufig mit dem neu beschriebenen CtdA Protein assoziiert sind. Über die Verwendung als Stickstoffquelle hinaus deuten die Ergebnisse auf eine intrinsische Verbindung zwischen Canavanin und Guanidin-Stoffwechsel hin. Insgesamt stellen die Daten einen Paradigmenwechsel von der Riboswitch-gesteuerten Entgiftung von Guanidin zur Aufnahme und Assimilation dieser Verbindung und zu ihrer Verwendung als Indikatormolekül für Guanidin-haltige Moleküle wie Canavanin dar.

Summary

The nitrogen rich compound guanidine is common in nature. A variety of molecules contain a guanidino group, ranging from amino acids like arginine and canavanine to the nucleobase guanine and secondary metabolites like streptomycin. Guanidine is sensed by at least four different classes of riboswitches that are widespread in bacteria. However, limited insights into the source and physiological role of guanidine in nature exist. For this thesis, canavanine, the δ-oxa-analogue of arginine, was explored as a source of guanidine. Thereby, a novel canavanine-degrading *Pseudomonad* was isolated and due to its ability to grow on canavanine as sole carbon and nitrogen source it was named *Pseudomonas canavaninivorans*. Within this work the bacterium is characterized employing polyphasic taxonomy and screened for its specific phenotypic traits that allow distinguishing it from closely related type strains. Furthermore, it is presented how the bacterium circumvents the toxicity of canavanine, which originates from incorporation into proteins as well as arginase-mediated hydrolysis to canaline that forms stable oximes with carbonyls. First, a specific canavanine degradation pathway is described in great detail. The key enzyme, a PLP-dependent canavanine-γ-lyase, was subjected to extensive characterization and further enzyme activities that facilitate canavanine utilization are elucidated. In addition, the distribution of the novel pathway and its implications in nature are discussed. Second, it is presented how *P. canavaninivorans* achieves error-free translation. The arginine-tRNA-synthetase of the bacterium is not able to discriminate between arginine and canavanine as substrates, so it uses a standalone specific canavanyl-tRNA^{Arg} editing protein to protect itself from misincorporation. We therefore propose canavanyl-tRNA^{Arg} deacylase (CtdA) as the systematic name for the editing protein. The results are the first to show editing activity towards mischarged tRNA^{Arg} and add to the puzzle of how faithful translation is ensured in nature. Finally, guanidine riboswitch-associated gene functions are explained in order to get insights into the compound's physiology and purpose. Not only is it shown that guanidine can be used as a nitrogen source but also guanidine riboswitches are often associated with the newly described CtdA. Beyond being used as a source of nitrogen, the findings hint towards an intrinsic connection between canavanine and guanidine metabolism. In sum, the data suggests a paradigm shift from riboswitch-controlled detoxification of guanidine to the uptake and assimilation of this enigmatic nitrogen-rich compound and to its usage as an indicator molecule for guanidine-group containing molecules like canavanine.

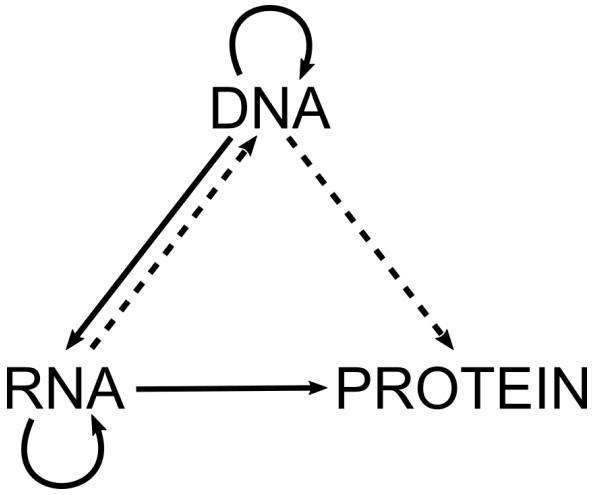
Chapter 1: Introduction

“From elephant to butyric acid bacterium – it is all the same”. The handwritten lecture note by Albert Jan Kluyver (1888-1956), a Dutch microbiologist and chemist, shortly summarizes Kluyvers idea of the “Unity theory of biochemistry” (1926). It states that all general enzymatic reactions needed to support and maintain life within cells and organisms are similar at a molecular level. Similarly, the Nobel laureate Arthur Kornberg stated in 2000: “[…] mechanisms and molecules have been preserved in bacteria, fungi, plants, and animals, essentially intact through billions of years of Darwinian evolution. I regard this insight as one of the great revelations of the 20th century” (1). However, these common theme is played in very different ways between organisms, as most living cells have adapted to a certain habitat or ecological niche. Within this thesis I will present the discovery of an hithero undescribed bacterium that thrives in the rhizosphere of legume plants and has developed strategies to detoxify and even feed on the toxic plant metabolite canavanine. Moreover, a pathway is presented that allows bacteria to use the remarkably stable compound guanidine as nitrogen source and evidence is presented that suggests a more prominent role of guanidine in nature. These findings extend our knowledge of bacterial metabolism in so different areas as central carbon and nitrogen metabolism and gene expression. Therefore, I will introduce the reader to the universal concept of how life on earth maintains and processes information and the basics of bacterial metabolism as background knowledge for a broad understanding of the presented work.

The central dogma of biology

The three main macromolecules in all living cells are deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins. In a lecture held in 1957, Francis Crick (one of the co-discoverers of the DNA structure), introduced the central dogma of biology which proposes the flow of information transfers within cells between DNA, RNA and protein (Scheme 1). First genetic information can be replicated, meaning that the information within the DNA is transferred again into DNA, which was proven experimentally by the above mentioned Nobel prize winner Kornberg with his isolation of the enzyme DNA polymerase I (2). Then, the information of DNA can be transferred into RNA, which is the first step in protein synthesis, called transcription. Following, RNA can be translated into protein, which is the second step of protein biosynthesis and called translation. Last but not least Crick envisioned that RNA can be replicated to RNA (e.g. as in RNA virus replication). Generalized, within DNA the genetic information of an organism is stored. Upon transcription the information is put into RNA, which is translated to produce proteins. The latter then obtain a variety of functions, reaching from the catalysis of metabolic reactions, to DNA replication, to transport or to providing structure to cells, thereby enabling

life as we know it. In 1970 also the information transfer from RNA to DNA was proven by the discovery of the enzyme reverse transcriptase (3,4), as already imagined by Crick before. The process of information flow can also be termed as gene expression and is explained in the following in more detail.



Scheme 1: Crick's version of the „central dogma" of biology as a flow of information within a system. Solid arrows represent the [in his opinion] probable transfers, dotted arrows possible transfers (5).

Gene expression in bacteria

Gene expression in bacteria consist of a two-step mechanism, transcription and translation, and they are linked in time and space (6), which is important as the bacterial half-life time of mRNA is on average only a few minutes (7). Transcription describes the process when the genetic information stored in DNA is transferred into RNA. The reaction is performed by an enzyme called RNA polymerase (RNAP) which use as single stranded DNA molecule as a template. As a standalone enzyme RNAP consists of a catalytic domain (E), but together with a sigma factor (σ) it forms the so called holoenzyme (E σ). σ factors ensure specific promoter recognition and there exist several competing factors which respond to different environmental conditions. For example, in *E. coli* the housekeeping factor σ^{70} is responsible for general transcription initiation of unconditional genes and genes required for all growth conditions. In contrast, for instance σ^{32} is responsible specifically to induce genes related to heat shock response, while σ^{28} regulates motility genes (8). Promoter recognition and binding is the first stage of the three stages of transcription: initiation, elongation and termination. For elongation the RNAP evolves into the so called elongation complex that is characterized by promoter escape and productive RNA synthesis (8). Transcript elongation by RNAP is discontinuous and interrupted by pauses that play key regulatory roles (9). Finally, transcription termination takes place at so called termination sites resulting in the release of the newly synthetized RNA strand (so called messenger RNA (mRNA)) and the fall-off of RNAP of the DNA template strand. Termination can be either Rho-independent (called intrinsic) or Rho-dependent. Intrinsic termination involves the formation of a RNA hairpin which is followed by a polyU track. In combination the hairpin causes the RNA polymerase to halt transcription while the polyU track results in very weak binding causing the above mentioned disassemblence of the elongation complex. Rho-dependent termination relies on the binding of the protein Rho to RNA. Rho is

a RNA/DNA helicase that dissociates RNAP from DNA template to release RNA, deriving energy by hydrolyzing adenosine triphosphate (ATP) through its RNA-dependent ATPase activity to bring about termination (10).

The second step of gene expression, translation, is fulfilled by ribosomes. They consist of a small (30S) and a large (50S) subunit which together assemble on the mRNA that is to be translated. Both subunits contain ribosomal RNA and ribosomal proteins. As for transcription, translation also occurs in the same three stages: initiation, elongation and termination. Upon initiation the ribosome binds to the ribosome binding site (RBS) of the mRNA, a process that is facilitated by several initiation factors. Within the mRNA three nucleotides comprise a so called codon and within the genetic code each amino acid is represented by one or several codons. Hence, the sequence of codons represents the amino acid sequence of the protein which is synthesized. Transfer RNAs (tRNA) are the bridging molecules used as linkers between codon and amino acid. They have a so called anticodon region were the nucleobases used in the region are complementary to the nucleobases of the mRNA codon thereby allowing Watson-Crick-base pairing in the ribosome. tRNAs transport activated amino acids which can then be used by the ribosome to elongate the forming polypeptide chain. Generally, translation elongation starts with the help of an initiator tRNA at the so called start codon (normally AUG) after the RBS. Following, aminoacylated tRNAs are recruited to the ribosome in association with GTP-bound elongation factors (EF) to synthesis the polypeptide chain which then is folding into protein. Finally, translation is terminated when the ribosome encounters a stop codon (normally UAG, UGA or UAA) which leads to the release of the polypeptide chain and the dissemblance of the ribosome. The regulation of translation is of key importance to the cell, as the aminoacylation of a single tRNA consumes an ATP and as upon translocation of the ribosome along the mRNA the GTP of GTP-EF gets hydrolysed. Therefore, the synthesis of protein has a high demand in energy making it very costly for the cell. Thus, the process of tRNA loading and its challenges will be described in more detail in the next part of the introduction, as faithful translation is a major topic in this work.

Ensuring faithful translation

The translation of genetic information into functional proteins is one of the most important information transfers in all living cells. The first step in the process of protein synthesis is the activation and loading of amino acids onto their cognate tRNAs, a reaction catalysed by the family of aminoacyl-tRNA synthetases (aaRSs). A total of 24 aaRSs have been described up to date, one for each encoded proteinogenic amino acid (except for lysine, for which there are two) plus pyrollyl-tRNA syntethase and phosphoseryl-tRNA synthetase (11,12). aaRSs catalyse the esterification of an amino acid to the 3’ end of the corresponding tRNA. Thereby they

produce the building blocks needed by the ribosome for protein synthesis. To load a tRNA, the cognate amino acid must first be activated (Figure 1A). To do so the amino acid and an ATP molecule bind to the catalytic site of the aaRS. Then a nucleophilic attack of the α-carboxylate oxygen of the amino acid to the α-phosphate group of ATP takes place, resulting in an amino-acyl-adenylate (aa-AMP) and pyrophosphate. Second, a hydroxyl group of the tRNA's 3' end adenine 76 attacks the carbonyl carbon of the aa-AMP, forming aminoacyl-tRNA and releasing AMP (Figure 1B). Although usually tRNA is not required for the amino acid activation step certain synthetases like arginine-tRNA synthetase require the binding of tRNA as a prerequisite for amino acid activation (13). Throughout evolution two classes of aaRSs have evolved independently and they can be divided based on their structure and their catalytic mechanism of amino acid activation. An overview regarding the classification of the synthetases is given in Table 1.

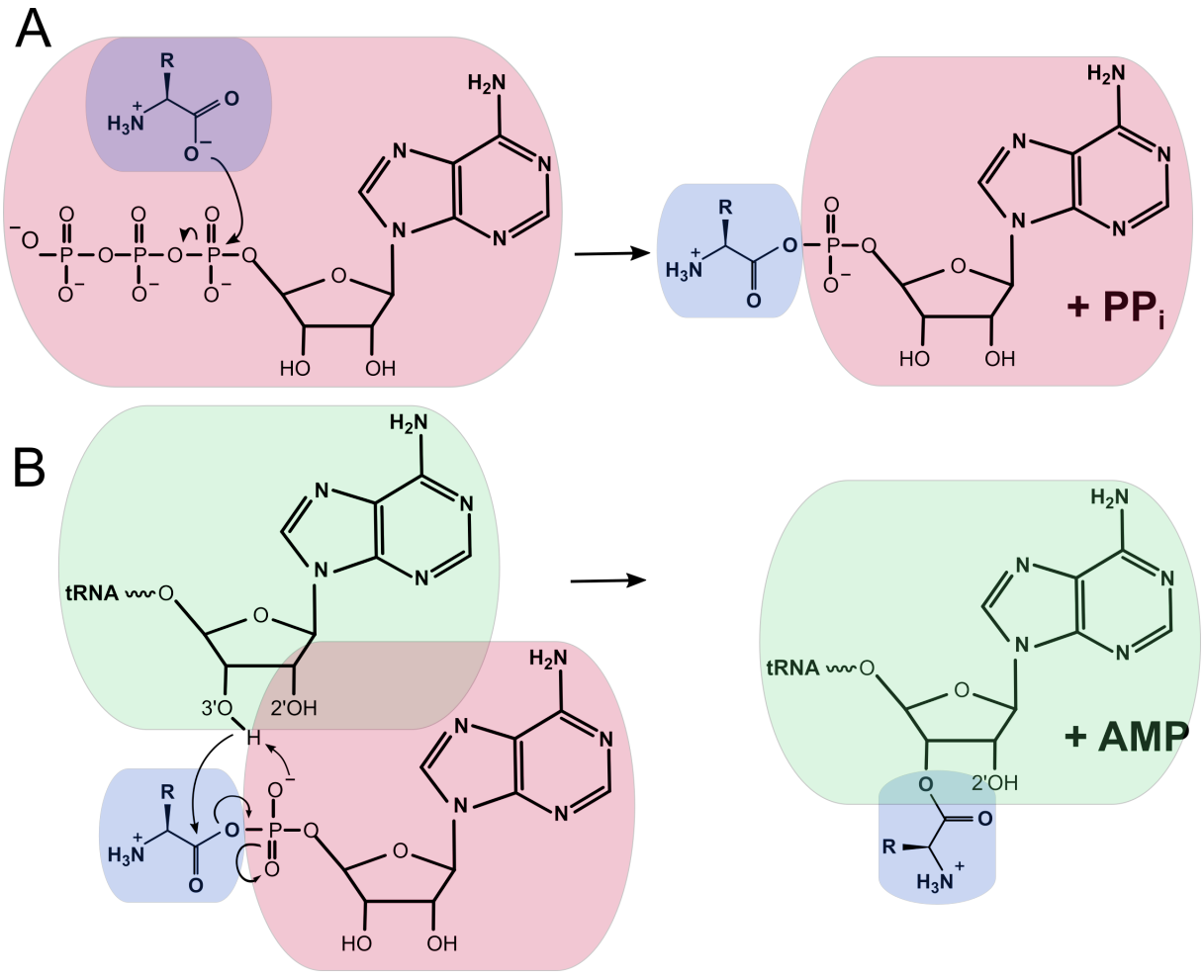


Figure 1: The aminoacylation reaction. In the first step (A), the amino acid (blue) is activated with ATP (red) in the synthetase active site (not depicted), forming aminoacyl-AMP and releasing PP_i. In the second step (B) the amino acid is transferred to the tRNA (green) and AMP is released (depicted in the image is the transfer to the 3'-OH characteristic for class I aaRS, while in class II transfer happens at the 2'OH with a 3'OH attack in the second step). Figure with slight modifications taken from Rubio et al. (2020) (11).

Ensuring the loading of the correct amino acid onto a specific tRNA is a substantial challenge, considering the number of different tRNAs and the choice between 22 proteinogenic and the countless non-proteinogenic amino acids that exist in nature (14). While the selection against non-cognate tRNAs is more readily accomplished due to the size of the protein-RNA contact interface (12) and specific identity elements (15), the discrimination against non-cognate amino acids is more challenging as amino acids are small molecules and only differ in their side chains. However, aaRS have evolved a double-sieve mechanism to avoid the production of mischarged tRNAs (16,17). First, aaRSs have a high specificity in the synthetic site, where the amino acids are activated by linkage to AMP and subsequently transferred to the matching tRNA. Second, several aaRS contain a proof-reading (termed 'editing') site, that hydrolyses incorrectly loaded aminoacyl-tRNAs before they are released from the enzyme. While the former is highly selective to prevent errors, the latter exhibits low selectivity to cleave all mischarged tRNAs whereas the desired product is not cleaved (18). Discrimination and editing activities have been discovered and described in great detail in both classes of aaRS (11,12,18–21). Still, discrimination and editing can meet their limits, especially when the non-cognate amino acid is slightly smaller than the cognate one and tRNA loading cannot be prevented on steric grounds (12). Furthermore, only half of the aaRSs have been found to display editing activity up to date (11). Impaired editing is severe and can affect cell growth or induce pathologies and apoptosis as shown for bacteria (22), yeast (23), mammalian cell culture (24) and mice (25). For several tRNAs where the aaRS does not have editing activity, autonomous standalone editing proteins have been identified in prokaryotes as well as in eukaryotes.

Standalone editing proteins catalyse the *in trans* editing of mischarged tRNAs and have been identified to be homologs of editing domains contained in some aaRSs. Still, the number of experimentally characterized standalone editing proteins is very limited. ThrRS-ed is a free-standing protein (homolog to the editing domain of ThrRS) found in archaea acting on Ser-tRNA^{Thr} (26,27). AlaXps (homologs of the editing domain of AlaRS) are editing factors found in all domains of life and cleave Ser- and Gly-tRNA^{Ala} (28,29). D-aminoacyl-tRNA deacylases (DTDs) also have been shown to edit Gly-tRNA^{Ala} in multiple species (30,31), in addition to their deacylation activity of D-Tyr-tRNA^{Tyr}. Finally, members of the INS superfamily (which can be divided into at least six further subgroups: INS, YbaK, ProXp-ala, ProXp-x, ProXp-ST1 and ProXp-ST2 (32)), clear mischarged tRNA^{Pro} (33–35). Taken together, standalone editing domain like proteins have been described to enhance proof reading in a multitude of organisms and therefore constitute a third "sieve" or mechanism of defence against accumulation of mischarged tRNAs (33). In chapter 2.3 of this work the discovery of an undescribed standalone editing domain-like protein that hydrolyzes canavanylated tRNA^{Arg} is presented.

Table 1: Classification of the aaRS into two classes. Adapted from Rubio et al. (2020) (11).

	Class I	Class II
aaRSs with described editing activity	MetRS, ValRS, LeuRS, IleRS	SerRS, ThrRS, AlaRS, ProRS, LysRS-II, PheRS,
aaRSs with no described editing activity	CysRS, ArgRS, GluRS, GlnRS, LysRS-I, TyrRS, TrpRS	GlyRS, HisRS, AspRS, AsnRS, PylRS, SepRS
Features	Rossman fold catalytic domain	Seven β-strands catalytic domain
	Minor grove approach to tRNA	Major groove approach to tRNA
	Transfer amino acid to the 3'OH	Transfer amino acid to the 2'OH
	Bind ATP in an extended configuration	Bind ATP in a bend configuration
	aa-tRNA release is the limiting step	Amino acid activation is the limiting step

Gene expression regulation mechanisms

To prevent unnecessary consumption of energy and the misuse of valuable resources bacterial cells obtain a tight regulation of gene expression. Thereby they are able to adapt to a fluctuating, competitive and frequently stressful environment (8). Regulation takes place on transcriptional as well as on translational/post-transcriptional level. On a transcriptional level the investigation of bacterial gene regulation was sparked with the discovery of the *Escherichia coli* lac operon by Jacob and Monod in 1959 (36). An operon represents a single transcription unit, meaning that a whole set of genes is expressed from one regulatory control region. Thereby cells can facilitate gene expression of a metabolic cassette of enzymes needed to utilize a present compound. The lac operon encodes enzymes that allow the metabolic usage of lactose. Within the operon a so called repressor (LacI) is expressed that hampers the successful binding of RNAP to the promoter of the catabolic enzymes, which prevents their expression, but LacI can be inactivated if it is binding lactose. Additionally, a catabolite activator protein (CAP) is binding to the promoter in form of cAMP-CAP which enhances the binding affinity of

RNAP, but cAMP is only produced in high concentrations in the absence of glucose. Thereby the cell can regulate that only in the absence of the preferred sugar glucose and the presence of lactose as an alternative source of nutrition the enzymes needed for lactose utilization are expressed. The phenomena that the availability of glucose prevents the usage of alternative carbon sources is termed carbon catabolic repression (CCR) and can also be found in other bacteria for different compounds. For example, in *Pseudomonas* species CCR is exerted by the CbrA/B-CrcZ-Crc golabal regulator system by the preferred C_4-dicarboxylates succinate, fumarate and malate (37). Numerous repressors and activators have been described up to date which can work, for example, as allosteric proteins that bind inducers (e.g. LacI-IPTG, which is used to induce overexpression of proteins) or that are regulated by mechanism like phosphorylation (e.g. OmpR (38)) or oxidation (e.g. OxyR (39)). Other ways of regulation include DNA modification (e.g. DNA methylation (40)) or the involvement of small ligands that interact directly with RNAP (e.g. ppGpp (41)). Noteworthy, over 6% of the open reading frames in *Pseudomonas aeruginosa* are annotated as putative transcription factors (42,43), which emphazises the importance of gene expression regulation. Also, bacteria that obtain different lifestyles have been shown to have more sigma and alternative transcription factors than intracellular pathogens that live in more stable circumstances (44), mirroring the need of different regulatory elements in diverse and changing habitats.

Post-transcriptionally, mRNA stability has a key role in regulating gene expression. mRNAs that are not actively translated are prone to Rho-dependent transcription termination (8) and degradation by the degradosome (45). In contrast, the tri- or diphosphate structure of the 5' end of bacterial mRNA is under discussion to have a protective function towards degradation (46), comparable to the eukaryotic 5' cap site. Furthermore, RNA-binding proteins bind to specific mRNAs and thereby modulate gene expression via various mechanisms: (i) directly affecting the susceptibility of the target RNA to degradation, (ii) modify the accessibility of the RBS for ribosome binding (either positively or negatively), (iii) act as a chaperon for interaction of the target with other effector molecules (including small RNAs) and (iv) alter the formation of transcription terminator/anti-terminator structures (8,47). Finally, some mRNA molecules use *cis*-acting regulatory mechanisms which are located in their 5' untranslated region (UTR), to regulate gene expression. *Cis*-acting mechanisms include riboswitches, thermoswitches, thermosensors and pH sensors, but for the scope of this thesis riboswitches are of main interest. Riboswitches are complex folded mRNA structures that naturally occur in all three kingdoms of life. They have been found to bind a range of small metabolites and ions and exert regulatory control of transcription, translation, splicing, and RNA stability (48). Riboswitches consist of two functional components: an aptamer domain that specifically senses a certain ligand and an expression platform that modulates gene expression (49). Binding of the ligand leads to a structural rearrangement and can thereby either activate (ON-switch) or repress

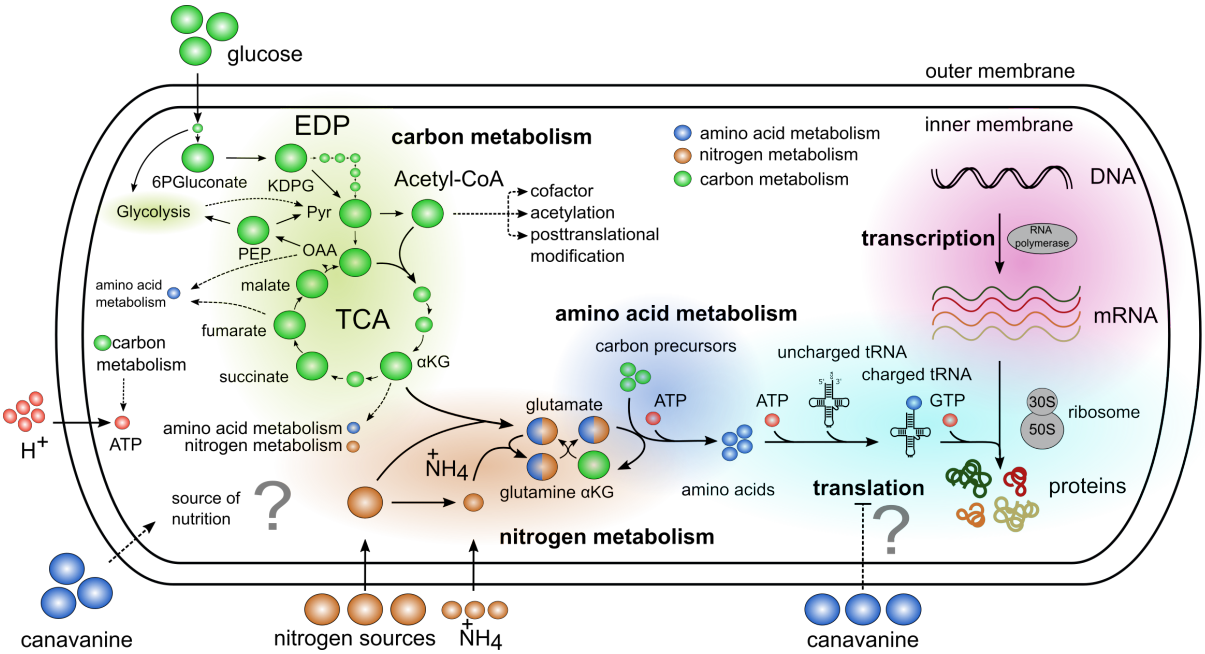
(OFF-switch) gene expression. In Gram-positive bacteria that happens mainly by the formation of a Rho-independent terminator or an anti-termination stem-loop (transcriptional level), whereas in Gram-negative bacteria the regulation is mainly based on trapping or liberating the RBS during translation initiation (47,50,51). A wide collection of ligands has been identified to regulate a riboswitch, ranging from cofactors (e.g. thiamine pyrophosphate), to sugars (e.g. glucoseamine-6-phosphate), to amino acids (e.g. lysine), to elemental ions (e.g. Mn^{2+}), to signalling molecules (e.g. cAMP-GMP) and to RNA precursors (e.g. guanine) (52). Novel riboswitch discovery offers the opportunity to unravel functions of associated genes and give insights into so far hidden pathways. In chapter 3 of this thesis gene functions associated with known members of guanidine riboswitches are described. A more detailed introduction on guanidine will be given there.

General aspects of bacterial metabolism

All organisms replicate and process genetic information but in order to proliferate and grow they need to take up nutrients and generate energy. Thus the major part of metabolism of all cells is based on the uptake of the nutrients carbon and nitrogen and the production of energy in the form of ATP, GTP and NAD(P)H. The nitrogen and carbon sources are then metabolized to produce lipids, polysaccharides, nucleotides and amino acids. The latter two are further used as building blocks for nucleic acids (DNA, RNA) and proteins. Utilizing those compounds, the final aim of metabolism is to sustain cell maintenance and to produce biomass or in other words to grow (53). A simplified overview scheme of bacterial metabolism is shown in Scheme 2.

Different metabolic processes are highly intertwined and many compounds have several roles within a cell. For example, amino acids are not only used in protein synthesis but can also serve as a carbon or nitrogen source. Also, bacteria have adapted to different ecological niches and thereby prefer distinct nutrition and energy sources and evolution has driven bacteria to maximize the nutritional yield of each environment. Several factors are involved in the determination of a “good” or a “bad” source of nutrition. For instance, for the model organism *Escherichia coli* glucose is the preferred sugar, but when the bacterium encounters poor nitrogen sources the normally favoured sugar becomes a bad carbon source and growth is hindered compared to growth on other sugars like arabinose, sorbitol or mannose (54). Moreover, different ways exist in nature to metabolize the same compound. While *E. coli* uses mainly glycolysis to metabolize glucose, in *Pseudomonas* it is mostly metabolized via the conversion to 2-keto-3-deoxy-phosphogluconate and the further oxidation through the Entner-Doudoroff pathway (55) (depicted in Scheme 2). As a result, still many metabolic pathways in bacteria are unknown which makes the discovery of new ones highly interesting, as they help us, for

example, to identify new antibiotics and other bioactive compounds that could be of pharmacological interest or have an environmental impact. Within this thesis the discovery of a hitherto undescribed canavanine degradation pathway is presented (chapter 2.2) which may have important implications regarding the legume rizosphere. As legumes represent a major player in global food security investigating the molecular mechanisms within the rhizosphere and the associated bacteria is of high importance.



Scheme 2: Simplified overview of bacterial metabolism. EDP: Entner-Doudoroff pathway, 6PGluconate: 6-phosphate-gluconate, ATP: adenosine triphosphate, PEP: phosphoenolpyruvate, Pyr: pyruvate, KDPG: 2-keto-3deoxy-phosphogluconate, αKG: α-ketoglutarate, OAA: oxalacetate. Arrows indicate the usual metabolic flow under rich and stress-free growth conditions. The Scheme is partially based on the review of Chubukov et al. (2014) (53).

Canavanine

Non-proteinogenic amino acids represent a large group of metabolites with diverse and often unexplored functions (14). Some non-canonical amino acids act as so-called antimetabolites by mimicking proteinogenic amino acids and thereby interfere with their respective functions. L-canavanine, or δ-oxa-arginine, serves as antimetabolite of L-arginine. It was first discovered when Kitagawa and Tomiyama (1929) assayed urea employing jack bean (*Canavalia ensiformis*) seed urease. Their assay reliably indicated greater urea values and in search for the source of additional urea they identified a basic amino acid as a co extraction product from the jack bean seeds (56). Therefore, it was given the trivial name canavanine and further experiments revealed the structural similarity to arginine (57). The discovery of the specific colori-

metric reaction of pentacyanoamine ammoniumferrate with canavanine enabled extensive surveys of its distribution among plants (58). Although the occurrence of canavanine is limited to legumes (59), a major subfamily of *Fabaceae*, it often serves as the main nitrogen storage compound in seeds with a content of up to ~12% of the seeds´ dry weight and up to 90% of the seed nitrogen allocated to free amino acids (60). During germination, canavanine catabolism produces canaline and urea, which are further degraded to homoserine, CO_2 and ammonia (61). However, canavanine is also found in the root exudate of young plants (62,63). Interestingly, the ability to reduce canaline to homoserine and ammonia is thought to be unique to higher plants (61), opening the legume the possibility to use canavanine as a toxic repellent against herbivores and pathogens. The toxic effects of canavanine originate from the mimicry of arginine, thereby interfering with arginine metabolism and protein functions when incorporated during translation. Arginyl-tRNA synthetases are not able to discriminate between arginine and its antimetabolite resulting in the incorporation of canavanine into nascent polypeptide chains and in the formation of dysfunctional proteins (64,65). Moreover, arginase-catalysed hydrolysis of canavanine yields canaline, which forms stable oximes with α-ketoglutarate and other carbonyl-containing compounds causing deleterious effects in the tricarboxylic acid cycle (66). Canaline is also forming oximes with pyridoxal phosphate (PLP), the activated form of vitamin B_6, thereby inactivating vitamin B_6-containing enzymes (67,68). The fact that almost 1.5% of all genes in most prokaryotic genomes encode PLP-dependent enzymes (69) helps to explain the severity of canaline intoxication in prokaryotes. Some insects have adapted different strategies to circumvent canavanine toxicity. While Drosophila detects it by taste as a repulsive molecule (70), the tobacco budworm *Chloridea virescens* (formerly *Heliothis virescens*) is able to degrade canavanine by a putative, so far uncharacterized gut enzyme (64,71). In bacteria two early reports concerned with canavanine degradation reported its cleavage to homoserine and either guanidine (as mentioned already before) or hydroxyguanidine but the molecular basis for this observation is unknown (72,73). In chapter 2 of this thesis the molecular basis of the latter unknown degradation mechanism will be presented.

Chapter 2: Canavanine utilization and detoxification in bacteria

The following chapter is divided into 3 subchapters that build up on another. First, the way to the discovery of *Pseudomonas canavaninivorans* is presented. The taxonomic position of the novel species is described using a polyphasic approach, including phylogenetic analysis based on 16S rRNA gene and whole genome sequencing, phenotypic characterization and the identification of chemotactic features. Succeeding, the molecular pathway by which the bacterium is able to thrive on canavanine is investigated and the key enzyme of the pathway is studied in great detail. Also the distribution of the pathway in nature is discussed. Finally, the specific detoxification mechanism by which *P. canavaninivorans* ensures faithful translation in the presence of canavanine is described.

Chapter 2.1: Discovery and taxonomic description of *Pseudomonas canavaninivorans*

Introduction

In a quest to identify bacteria that are accumulating guanidine in nature we came across the early publications by Kihara et al. (1955) (72) and Kalyankar et al. (1958) (73). While the former study describes the bacterial cleavage of canavanine to homoserine and guanidine, the latter found canavanine to be cleaved to homoserine and hydroxyguandine. Both authors used bacterial cultures and measured the decline of canavanine concentrations in the growth medium, but detailed descriptions on a molecular and enzymatic level remained elusive, due to the technical limitations of those early years of molecular biochemistry. The bacterium used by Kihara was identified to be *Streptococcus faecalis* (today *Enterococcus faecalis*), a common gut inhabitant and the bacterium used by Kalyankar was categorized to be a *Pseudomonad*. With those findings in mind we decided to shed light onto the mechanisms involved in canavanine cleavage, as they represent possible sources of guanidine in nature. As evolution has driven organisms to adapt to the given conditions in their respective ecological niche we thought to search for canavanine cleaving bacteria in the canavanine rich rhizosphere of legume plants.

Results

Enrichment of canavanine-degrading bacteria

For the purpose of isolating bacteria able to degrade canavanine, rhizosphere soil samples of the legume *Phaseolus coccineus* (runners bean) were collected and serial dilutions streaked out on M9 salts minimal medium plates containing canavanine as sole carbon source. In order to differentiate between bacteria involuntarily using canavanine via the arginase-mediated hydrolysis to canaline and urea and bacteria producing homoserine and guanidine or hydroxyguanidine growing bacterial isolates were further cultivated in liquid medium and culture supernatants analyzed by mass spectrometry (LC-MS) for guanidine and hydroxyguanidine contents. Interestingly, upon growth on 10 mM canavanine as carbon source most isolates produced only small amounts of guanidine (< 500 µM in the supernatant), while accumulating hydroxyguanidine in concentrations up to 8 mM (Figure 2A). The increase of hydroxyguanidine correlated with the decrease of canavanine, suggesting for the isolates a cleavage of canavanine to hydroxyguanidine and homoserine, as it had been observed by Kalyankar for the undescribed *Pseudomonad* (73) or for the tobacco budworm by Berge et al. (71). In a further experiment ammonium salts were omitted as nitrogen source in order to evaluate whether canavanine would serve as sole nitrogen source as well. Isolate HB002 was able to grow on M8 minimal salt medium with canavanine as sole carbon and nitrogen source and no growth differences were observed in optical densities (Figure 2B).

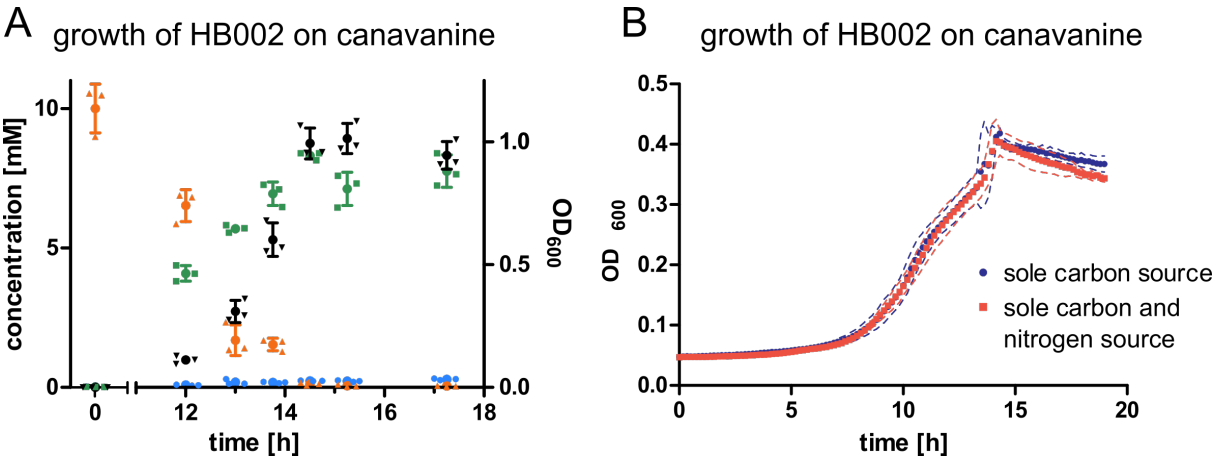


Figure 2: Growth of isolate HB002 on canavanine. **A:** Concentration of canavanine (orange), guanidine (blue) and hydroxyguanidine (green) in the supernatant of isolate HB002 grown on 10 mM canavanine as sole carbon source in M9 minimal medium, values represent the mean of triplicates measured over time, optical density of the corresponding cultures is shown in black, n = 3, error bars = SD. **B:** growth of HB002 in minimal medium with 10 mM canavanine as sole carbon and nitrogen (red) or sole carbon (blue) source, n = 3, dashed line = SD.

Discovery of Pseudomonas canavaninivorans

HB002 was then further investigated by 16S rRNA gene sequencing. Running the sequence against the EzBioCloud 16S database (74) revealed that the bacterium belongs to the genus *Pseudomonas* with its closest relatives being *P. piscium* (P50^T, 100% similarity, 84.4% completeness), *P. bijieensis* (L22-9^T, 99.93% similarity, 100% completeness), (*P. brassicacearum subsp. neoaurantiaca* (ATCC 49054^T, 99.76% similarity, 99.7% completeness), *P. brassicacearum subsp. brassicacearum* (DBK 11^T, 99.63% similarity, 99.9% completeness), *Pseudomonas thivervalensis* (DSM 13194^T, 99.51% similarity, 100% completeness), *Pseudomonas kilonensis (*(DSM 13647^T, 99.39% similarity, 100% completeness) and *Pseudomonas corrugata*, ATCC 29736^T, 99.39% similarity, 98.6% completeness). A summary of the top valid hits and the gene sequences used for the analysis can be found online in the supplementary material of the original publication (75). Next, to further investigate the isolate, the bacterium was targeted by whole genome sequencing, which was done by Novogene (Novogene UK, Cambridge) and the obtained genomic data was subjected to digital DNA-DNA hybridization (dDDH) analysis using the type strain genome server (TYGS) (76–81). A detailed description of the TYGS method can be found online (75). Based on the MASH algorithm (82) in combination with 16S rRNA gene sequence analysis (77,83,84) a phylogenetic tree was build (Figure **3**) which confirmed the preliminary 16S rRNA gene amplification result that strain HB002 clustered with known members of the *P. corrugata* subgroup of *P. fluorescens*. dDDH values for closely related type strains were all below the 70% cut-off value for species delineation (85). Figure 4 shows the phylogenetic tree inferred from the genome BLAST distance phylogeny (GBDP) calculated on the genome sequences. Table 2 shows the calculated dDDH values and the difference in G+C content between the query sequence and closely related type strains of the genus *Pseudomonas*. Average nucleotide identity (ANI) was calculated using the TrueBac^{TM} ID server (86–88) and confirmed the previous dDDH analysis based classification of the isolated strain as belonging to a new species of the genus *Pseudomonas*. The ANI scores of closest relatives were all below the threshold of 95% for species delineation (89). Also, the housekeeper gene *recA* showed high differences in similarity (Table 3). *In silico* PCR on nine marker genes was performed as described by Garrido-Sanz et al. (2017) (90) and, as expected from the previous classification results, strain HB002^T obtained the specific markers of the *P. corrugata* phylogroup. Next, the bacterium was tested towards its physiological traits that allow to differentiate strain HB002^T from closely related species. The key results are listed in Table 4. The results of the whole phenotypic characterization are given in the subsequent species description.

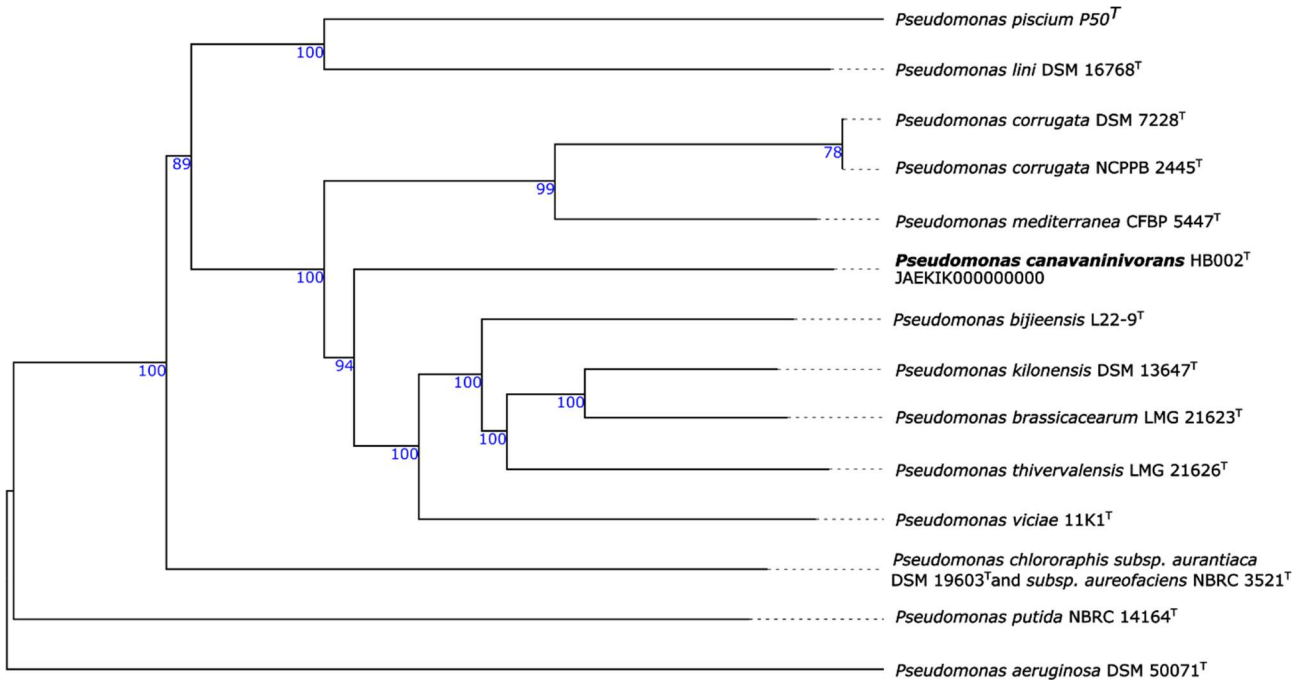


Figure 3: Phylogenetic tree based on 16S rDNA gene. Tree inferred with FastME 2.1.6.1[13] from GBDP distances calculated from 16S rDNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula d5 (for a more detailed description see supplementary material File S2). The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 74.1 %. The tree was rooted at the midpoint [14]. *P. putida* and *P. aeruginosa* were chosen as outgroups. Strain HB002^T = DSM 112525^T = LMG 32336^T is highlighted in bold.

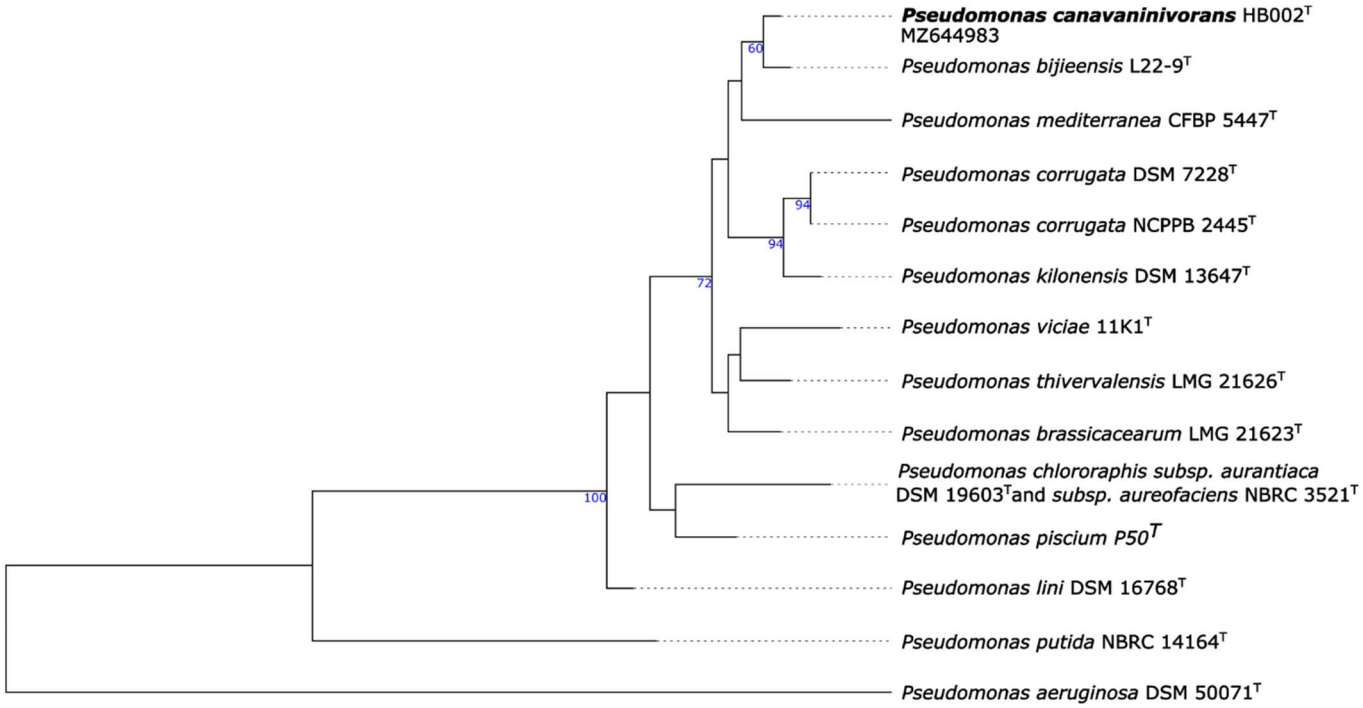


Figure 4: Phylogenetic tree based on genome sequences. Tree inferred with FastME 2.1.6.1[13] from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5 (for a more detailed description see supplementary material File S2). The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 74.1 %. The tree was rooted at the midpoint [14]. *P. putida* and *P. aeruginosa* were chosen as outgroups. Strain HB002^T = DSM 112525^T = LMG 32336^T is highlighted in bold.

Table 2: Digital DNA-DNA hybridization calculations for HB002^T and closely related type strains. The dDDH cut off value for the identification of a novel species is <70%.

Hit taxon	dDDH (d0, in %)	dDDH (d4, in %)	dDDH (d6, in %)	G+C content difference (in %)
Pseudomonas thivervalensis LMG 21626^T	64.6	35.4	57.4	0.14
Pseudomonas bijieensis L22-9^T	60.2	34.2	53.6	0.20
Pseudomonas kilonensis DSM 13647^T	61.0	35.2	54.6	0.19
Pseudomonas brassicacearum LMG 21623^T	60.7	35.1	54.3	0.25
Pseudomonas bijieensis L22-9^T	59.5	33.4	52.7	0.19
Pseudomonas viciae 11K1^T	55.1	33.6	49.4	0.71
Pseudomonas mediterranea CFBP 5447^T	50.6	32.0	45.4	0.18
Pseudomonas corrugata NCPPB 2445^T	50.0	31.1	44.7	0.43
Pseudomonas corrugata DSM 7228^T	49.9	31.1	44.6	0.42
Pseudomonas chloroaphis subsp aureofaciens NBRC 3521^T	29.5	27.2	27.9	1.7
Pseudomonas piscium P50^T	26.7	26.1	25.4	2.55
Pseudomonas lini DSM 16768^T	33.0	26.8	30.5	2.23

Table 3: TrueBac ID results showing the average nucleotide identity and the sequence similarity of the respective genes of HB002^T with closely related species. The ANI cut-off value for the identification of a novel species is <95%.

Hit taxon	ANI (%)	ANI coverage (%)	16S (%)	recA (%)
Pseudomonas brassicacearum subsp. neoaurantiaca ATCC 49054^T	89.08	73.8	99.79	94.87
Pseudomonas thivervalensis DSM 13194^T	89.32	74.8	99.59	N/A
Pseudomonas bijieensis L22-9^T	87.61	48.2	99.93	93.9
Pseudomonas corrugata LMG 2172^T	87.67	64.2	99.58	95.06
Pseudomonas kilonensis DSM 13647^T	89.18	72.9	99.52	94.87
Pseudomonas brassicacearum subsp. brassicacearum LMG 21623^T	89.08	73.0	99.52	94.78
Pseudomonas moorei DSM 12647^T	86.15	41.1	97.85	N/A
Pseudomonas umsongensis LMG 21317^T	86.05	39.4	97.67	N/A
Pseudomonas reinekei DSM 18361^T	85.99	41.5	97.60	N/A

Table 4: Differential characteristic phenotype of *P. canavaninivorans* and closely related species. 1: *P. canavaninivorans* strain HB002^T, 2: *P. bijieensis* L22-9^T (91), 3: *P. corrugata* NCPPB 2445^T [BacDive ID: 13044], 4: *P. viciae* 11K1^T (92), 5: *P. kilonensis* DSM 13647^T (93), 6: *P. brassicacearum* DBK11^T (94), 7: *P. thivervalensis* DSM13194^T (94); data for *P.canavaninivorans* HB002^T was from this study and the data of the related strains was obtained from the respective publication/database; +: positive, -: negative, w: weak positive, v: variable, NA: data not available.

Characteristics	1	2	3	4	5	6	7
API 20NE							
gelatine hydrolysis	-	+	+	-	+	+	+
aesculin hydrolysis	w	-	-	-	-	-	-
assimilation of arabinose	w	+	+	+	+	+	+
assimilation of mannitol	+	+	+	+	+	-	+
assimilation of N-acetyl glucosamine	-	+	+	-	-	+	+
API ZYM							
alkaline phosphatase	+	+	v	+	NA	NA	NA
esterase (C4)	+	+	+	-	NA	NA	NA
esterase lipase (C8)	+	+	+	w	NA	NA	NA
lipase (C14)	-	-	-	+	NA	NA	NA
leucine aryl amidase	+	+	+	-	NA	NA	NA
valine aryl amidase	+	+	v	-	NA	NA	NA
cystine aryl amidase	-	-	-	w	NA	NA	NA
trypsin	-	-	+	+	NA	NA	NA
α-chymotrypsin	-	+	v	+	NA	NA	NA
acid phosphatase	+	-	v	-	NA	NA	NA
napthole AS-BI-phosphohydrolase	+	+	-	-	NA	NA	NA

Description of Pseudomonas canavaninivorans sp. nov.

The type strain HB002^T was isolated from the rhizosphere soil of a runner bean (*Phaseolus coccineus*) field that was collected from the island Reichenau, at Lake Constance, Germany (47.7024177717245 N 9.044449559793435 E) in mid-August 2020. The DNA G+C content of strain HB002^T is 60.02mol%. The Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession number JAEKIK000000000. The 16S rRNA gene sequence for deposition was obtained from the whole genome sequence and deposited at DDBJ/ENA/GenBank under the accession number MZ644983. The type strain was deposited at the Leibnitz institute DSMZ under the accession number DSM 112525 and at the Belgium microorganism collection BCCM/LMG under the accession number LMG 32336.

The bacterium was named *Pseudomonas canavaninivorans* due to its ability to grow solely on canavanine as carbon and nitrogen source. The name is derived from the latin words canavanium and *vorans* (ca.na.va.ni.ni.vo'rans. N.L. neut. n. *canavaninum*, canavanine; L. pres. part. *vorans*, eating; N.L. fem. part. adj. *canavaninivorans*: canavanine-eating).

The rod-shaped, Gram-negative bacterium grew as translucent, round, yellow-beige colored colonies on LB agar. It is motile, 1-1.2 µm wide and 2.5 – 3.3 µm long. It is growing facultative anaerobe in the presence of KNO_3 and growth occurs in 0 - 4.6% (w/v) NaCl, at pH varying from pH 5.5 – 8.0 and within the temperature range of 4 to 37 °C. Fluorescent colonies could be detected on King B but not on King A agar. DNA hydrolysis was weak positive whereas gelatine hydrolysis was negative. Catalase and oxidase tests were both positive. In the API 20NE test the bacterium was positive for the hydrolysis of aesculin and the assimilation of D-glucose, L-arabinose, D-mannose, D-mannitol, gluconate, capric acid, malic acid and citric acid. It was negative for indole production, glucose fermentation, arginine dihydrolase, urease, 4-nitroso-β-d-methyl galactose, N-acetyl glucosamine, maltose, adipic acid and phenylacetic acid. In the API ZYM system it was positive for alkaline phosphomonoesterase, esterase (C4), esterase lipase (C18), leucine arylamidase, acid phosphatase and napthol AS-BI-phosphohydrolase, weak positive for valine arylamidase and negative for lipase (C14), cysteine arylamidase, α-chemotrypsine, α-galactosidase, β-galactosidase, β-glucoronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. The predominant fatty acids of *P. canavaninivorans* were $C_{16:0}$, $C_{17:0}$ cyclo ω7c and $C_{18:1}$ ω7c. The major respiratory quinone was Q9, followed by Q8 and only minor components of Q7 and Q10. A more detailed overview of the fatty acid and respiratory quinone profile can be found online (75).

Conclusion

In this chapter the discovery of a novel *Pseudomonas* species – *P. canavaninivorans* – was presented. Genome analysis and *in silico* gene marker analysis classify it to belong to the subfamily of *P. corrugata* within the group of *P. fluorescens*. The ability to hydrolyze gelatine and trypsin activity can be used to differentiate the bacterium from its known close relatives. *P. canavaninivorans* is able to grow solely on canavanine as carbon and nitrogen source and to do so, it hydrolyzes canavanine and releases hydroxyguanidine into the medium. The molecular mechanism of the cleavage reaction is depicted in the next chapter of this thesis.

Chapter 2.2: Canavanine utilization in rhizosphere associated bacteria

Introduction

Following the discovery of *P. canavaninivorans*, in this chapter a specific bacterial canavanine degradation pathway is described. By performing comparative proteomics, the genes upregulated when growing the bacterium on canavanine could be identified. Within the following the key enzyme of the utilization pathway was studied and further enzymatic activities were investigated that result in the conversion of canavanine via intermediates to fumarate. Finally, the distribution of this novel degradation pathway and its relevance in nature are discussed.

Results

Analysis of canavanine-dependent gene expression

In order to examine the genes involved in canavanine degradation *P. canavaninivorans* was either grown on canavanine or glycerol as sole carbon source and differences in protein expression levels were detected by comparative proteome analysis. Among the 1610 proteins identified (the whole list of identified proteins can be found online in the SI of the original publication (95)) 45 were only expressed upon growth on canavanine. Of those 45 proteins, seven cluster together in the same genetic locus. Among those proteins is a class I/II PLP-dependent aminotransferase (MBJ2347155.1, further referred to as AT) and an aldehyde dehydrogenase (MBJ2347154.1, further referred to as AC) which cluster together with a transcription factor, a guanidine exporter (96) (MBJ2347156.1, further referred to as Gdx) and four subunits of an ATP-binding cassette-type transporter (MBJ2347157.1- MBJ2347160.1, further referred to as ABC transporter) (Figure 5)). In addition, we found a highly upregulated homoserine-dehydrogenase (MBJ2346999.1, further referred to as HD) with yet another aldehyde-dehydrogenase (MBJ2347000.1, further referred to as AH) clustering together and at a separate locus an aspartate-ammonia-lyase (MBJ2345991.1, further referred to as AAL) (Figure 5). We speculated that the potential operon in conjunction with the other enzymes is enabling *P. canavaninivorans* to utilize canavanine as sole carbon source, as they were amongst the most abundant proteins found in the canavanine-grown sample. As a key enzyme in canavanine degradation the AT was the most promising candidate, because individuals of this general class of enzymes are known to catalyze a variety of different reactions with amino acid substrates ranging from decarboxylation, racemization and transamination to replacement and elimination reactions (97).

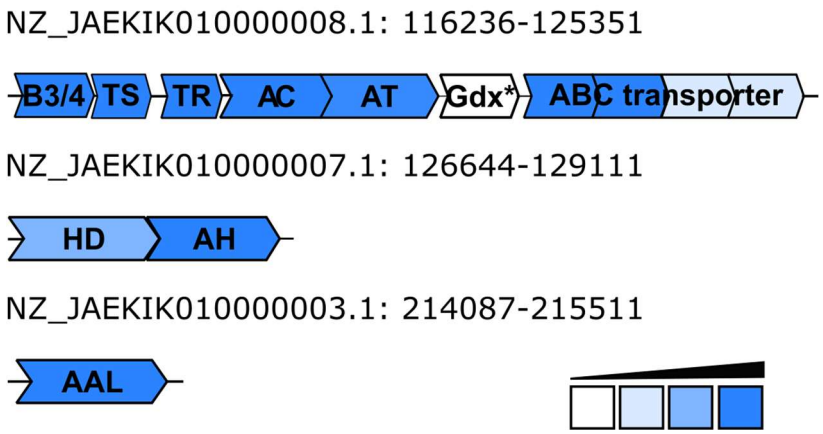


Figure 5: Genomic loci of induced proteins in *P. canavaninivorans* (genome available at NCBI, accession number JAEKIK000000000) upon growth on canavanine compared to growth on glycerol. Except for HD and two of the ABC transporter units all proteins shown were only detected in the canavanine-grown sample. Abbreviations: TR: transcriptional regulator, AC/AH: aldehyde dehydrogenases, AT: class I/II aminotransferase, Gdx: guanidine exporter (*not detected in the proteome analysis), ABC: ATP binding cassette, HD: homoserine dehydrogenase, AAL: ammonium aspartate lyase.

In order to gain more insight into potential reactions performed by the PLP-dependent AT, an advanced sequence search against the RCSD protein data bank was performed. The most similar hit was a methionine-γ-lyase (MγL) of *Clostridium sporonges* (98). We speculated that in analogy to MγL, the identified AT could catalyze the elimination of hydroxyguanidine from canavanine with a subsequent water addition to yield homoserine as additional product (Figure 6A). In order to test whether AT catalyzes the proposed reaction, the enzyme was overexpressed in *E. coli* and purified.

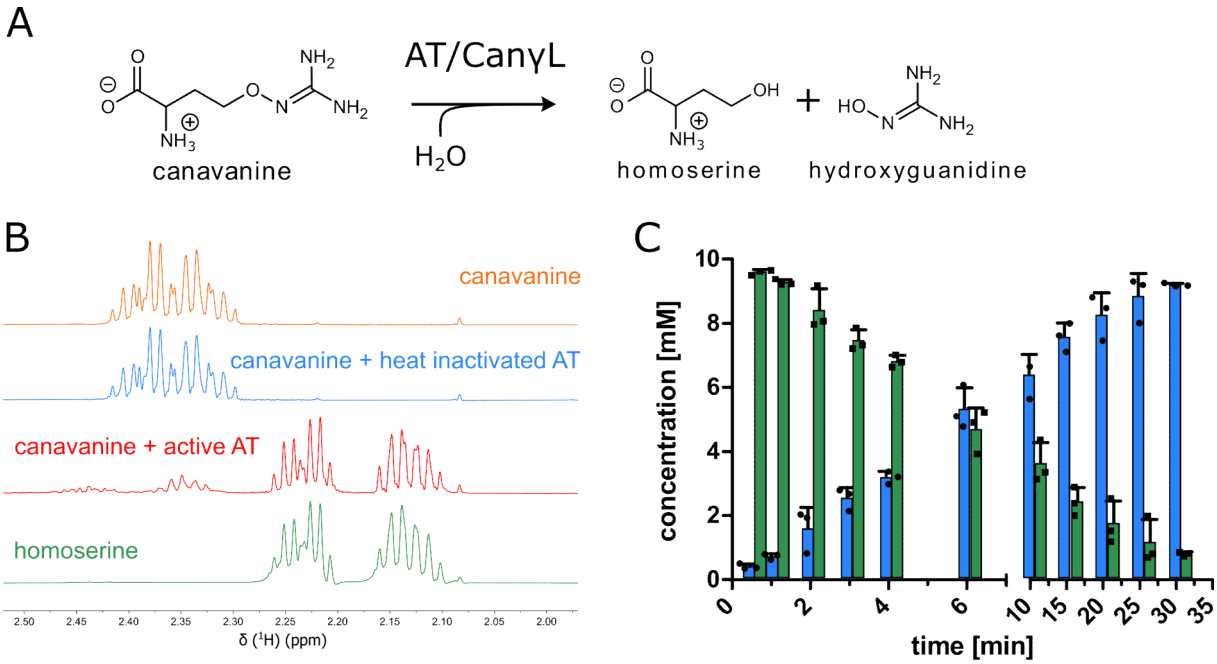


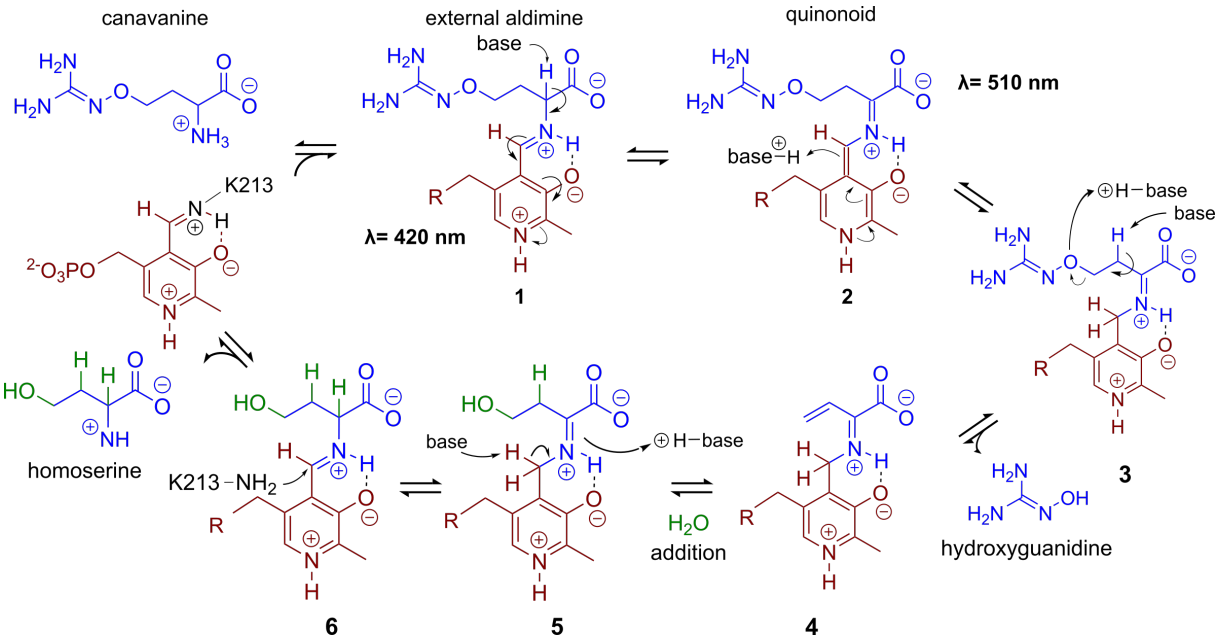
Figure 6: Aminotransferase assays. **A:** Proposed reaction catalyzed by aminotransferase AT (MBJ2347155.1). **B:** ^{1}H-proton NMR spectra of reactions with purified AT showing the conversion of canavanine to homoserine (for complete spectra see (SI Item 1). **C:** MS concentration measurements of canavanine (green) and hydroxyguanidine (blue) in a reaction with purified AT, n = 3, error bars = SD, each data point represents the mean of a technical triplicate.

Characterization and reaction mechanism of canavanine-γ-lyase

In a first experiment the aminotransferase (AT) was incubated with canavanine and the reaction was investigated using proton nuclear magnetic resonance spectrometry (^{1}H-NMR) and MS. Canavanine was converted to homoserine in an enzyme- and time-dependent manner (Figure 6B and SI Figure 1). No changes in canavanine peak intensities were observed when AT was heat-inactivated prior to the reaction (Figure 6B, blue spectrum), while with active enzyme after one hour the majority of canavanine was converted to homoserine (Figure 6B, red spectrum). Furthermore, in consistency with the *in vivo* growth experiment (see chapter 2.1, Figure 2A) MS measurements of reactions with purified AT revealed an increase of hydroxyguanidine with a simultaneous decrease of canavanine over time (Figure 6C). ^{1}H-NMR also showed the formation of small amounts (< 5% of product) of 2-oxobutanoate and ammonium (SI Figure 2), which is in accordance with a potential β,γ-elimination mechanism. For the methionine-γ-lyase-catalyzed reaction of methionine it has been shown before that 2-oxobutanoate and ammonium occur as products (99,100). In the case of the newly identified AT, this reaction only seems to take place as a side reaction and homoserine is found as main product.

In order to shed light on the reaction mechanism of the AT, a hydrogen/deuterium (H/D) exchange experiment was performed. In case the reaction involves the addition of water to the unsaturated bond at the β,γ-position, carrying out the reaction in D_2O should lead to changes in some of the signals of homoserine recorded in the ^{1}H-decoupled ^{13}C-NMR experiment. Specifically, we expected a change of multiplicity for the signal assigned to the carbon atoms in the α and β position as a result of the incorporation of deuterium into the molecule. Indeed, the obtained ^{13}C-NMR data confirmed our hypothesis as the signals assigned to the α and β carbon atoms changed their multiplicity from singlet to triplet with equally intense peaks SI Figure 3. This outcome can be explained by the inclusion of a deuterium atom at the two positions which also confirms the protonation event from solvent D_2O in the last imine isomerization. The experiments let us conclude that the identified AT is a canavanine-γ-lyase (CanγL) and allowed to suggest a likely reaction route Scheme 3. First, canavanine reacts to form aldimine (1) between the alpha amino group of canavanine and the cofactor pyridoxal phosphate (PLP) formerly bound to an internal lysine. Then, protonation/deprotonation results in formation of quinonoid (2) and canavanine ketamine (3) followed by the elimination of hydroxyguanidine from the β - and γ-positions to yield the unsaturated vinylglycine ketamine intermediate (4). Next, water adds to the double bond in the β,γ-position forming the imine equivalent of homoserine (5). Imine isomerization to form the homoserine aldimine (6) and reaction with the internal CanγL-lysine residue results in the release of the product homoserine. Alternatively, if release of the vinylglycine ketimine (4) from PLP takes place before water addition the reaction likely

follows the described mechanism of methionine-γ-lyase (99,100) to yield 2-oxobutanoate and ammonium, thereby explaining the formation of small quantities of these products (SI Figure 2).



Scheme 3: A potential reaction mechanism of canavanine-γ-lyase: Starting with free canavanine, PLP-canavanine aldimine (1) formation, quinonoid (2) formation, canavanine ketimine (3) formation, elimination of hydroxyguanidine resulting in vinylglycine ketimine (4), water addition yielding homoserine ketimine (5), homoserine aldimine (6) formation, and release of homoserine by reformation of PLP-imine with the internal lysine. Wavelengths indicate UV-Vis maxima in UV-Vis data shown in SI figure 8 next to reaction intermediates that could potentially cause the observed absorption.

Next, the substrate specificity of canvanine-γ-lyase (CanγL) was characterized with the canavanine analogue arginine as well as substrates of the mechanistically related enzymes methionine- and cystathionine-γ-lyase. No turnover was detected with arginine nor using the substrates methionine and cystathionine monitored by ^{1}H-NMR (SI Figure 4), concluding that CanγL is highly specific for canavanine degradation. Further experiments at different pHs and temperatures revealed a pH optimum of CanγL between pH 6.9 and 8.0 and an optimal reaction temperature between 34.8 and 37°C (SI Figure 1B&C) which is similar to reported values for the related enzyme methionine-γ-lyase (101,102). A kinetic characterization of CanγL was carried out by coupling the cleavage reaction of canavanine to the upregulated homoserine-dehydrogenase (HD) which catalyzes the conversion of homoserine and NAD^+ to aspartate-semi-aldehyde and NADH. The production of NADH was monitored spectrophotometrically. HD was added in 20-fold access over CanγL to ensure that the CanγL-catalyzed step is rate-limiting (SI Figure 5A). Plotting the initial velocities over the substrate concentration allowed us to calculated K_M and k_{cat}. CanγL converts canavanine with a K_M of 590 ± 20 µM and a k_{cat} of 1.34 ± 0.01 s^{-1}.

To confirm the involvement of PLP as a cofactor the predicted active site lysine (K213), which is responsible for PLP binding, was then mutated to alanine and the mutational effect on the reaction by NADH formation was evaluated. The mutation K213A completely abolished enzymatic activity as NADH formation of K213A with 25 mM canavanine was similar to the background reaction of wild-type CanγL without addition of the substrate canavanine (Figure 7).

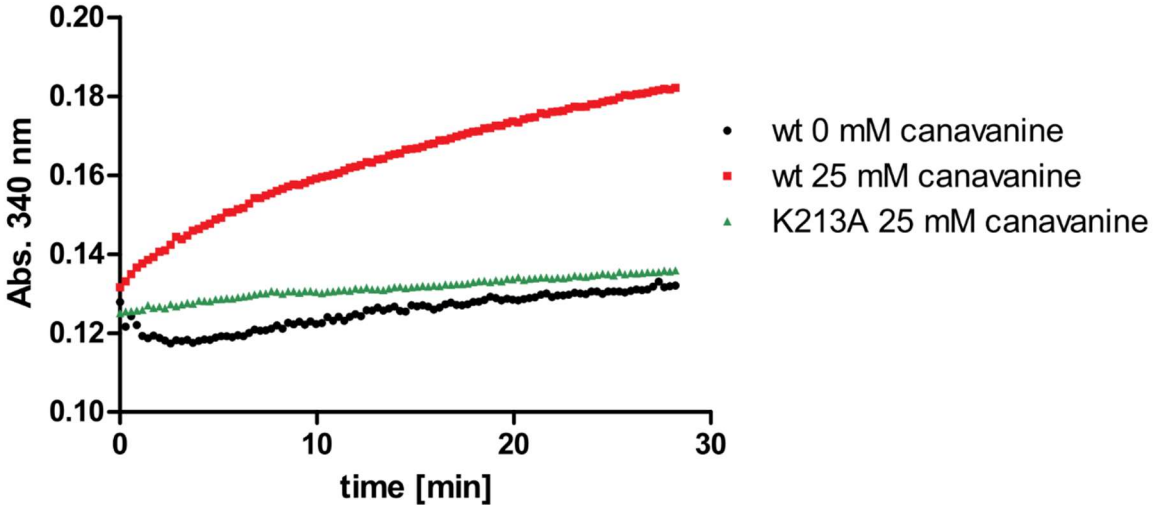


Figure 7: NADH production by either wt or K213A CanγL in the HD coupled assay with 25 or 0 mM canavanine substrate. Each data point represents the mean of a technical triplicate.

Additionally, UV/Vis spectra of the CanγL enzyme reaction were recorded since the PLP aldimine and quinonoid intermediates give rise to characteristic peaks(103,104). Time-dependent increase of presumed aldimine and quinonoid peaks were observed for the wt enzyme whereas a K213A mutant of the enzyme did not show time-dependent changes in peak intensities (Figure 8).

PLP-dependent enzymes are found in several different protein families and fold classes. In order to gain an insight into the structure and fold type of the enzyme, the primary sequence of CanγL was submitted to alpha-fold (105–109) and the resulting model (SI Figure 6A) superimposed with the structure of methionine-γ-lyase (MγL) from *Clostridium sporonges* (PDB code: 5DX5 chain A), as this enzyme represented the closest match in the RSCB protein data bank. The overall fold and structure of CanγL is very similar to MγL (SI Figure 6B), especially in the PLP-binding domain, which assigns the CanγL to the type I-fold architecture of PLP-dependent enzymes. The CanγL PLP domain contains the typical seven stranded mainly parallel β-sheet with +−+++++ directions of the β-strands, which are flanked by seven α-helices on both sides of the β-sheet (98,110). Also the PLP-binding lysine K213 aligned well with K212 of MγL (SI Figure 6C). A PYMOL file of the superimposed model can be found online in the supplementary files of the original publication.

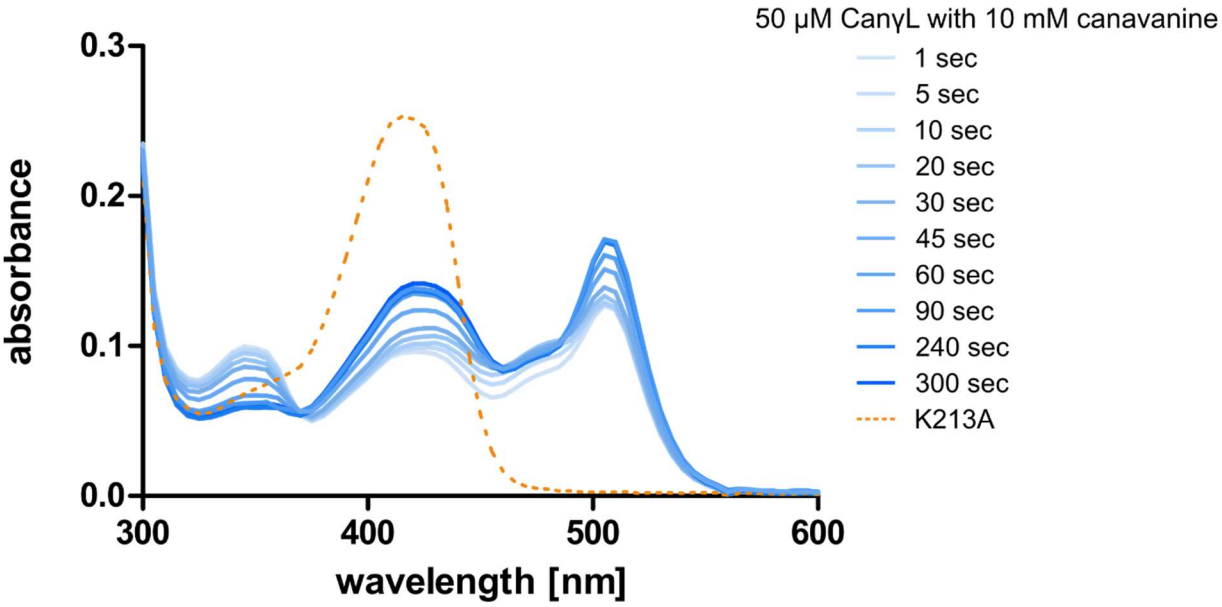


Figure 8: UV/Vis spectra of 50 µM wt or K213A CanγL after the addition of 10 mM canavanine. For mutant K213A a representative spectrum is shown since no changes over reaction time were observable. The spectra display absorption peaks in accordance to characteristic PLP intermediates reported in the literature (103,104): The peak at 420 nm presumably corresponds to the PLP aldimine, while the peak at 510 nm shows the quinonoid intermediate. The assignment of the peak at 350 nm is less straightforward but it could represent the canavanine ketamine. In general, UV/Vis spectra of PLP-dependent enzymes and their many reaction intermediates are often complicated to interpret and further mechanistic studies would be necessary in order to confirm the proposed assignments.

Finally, to validate the proposed role of canavanine-γ-lyase in canavanine assimilation deletion strains were created via homologous recombination. CanγL was deleted together with a second gene annotated as B3/4 domain-like protein (MBJ2347151.1) which we suspected to have an involvement in canavanine metabolism as it was also highly upregulated in the proteome data and its located in the genome in front of the canavanine degradation operon. The deletion of CanγL rendered the strain unable to grow on canavanine as sole carbon source while its growth on glucose was not affected (SI Figure 7). A single deletion strain of the B3/4 domain already resulted in a strong reduction of growth on canavanine, an observation that will be assessed in chapter 2.3 of this thesis. Taken together, the findings describe a novel enzyme acting efficiently and specifically on canavanine by a PLP-dependent β/γ-elimination/addition mechanism yielding the two products homoserine and hydroxyguanidine.

Characterization of homoserine-dehydrogenase, aldehyde-dehydrogenases and ammonium-aspartate-lyase

Next, it was tested whether the enzymatic activity of homoserine-dehydrogenase (HD), aldehyde-dehydrogenase 1 (AH, same operon as HD), aldehyde-dehydrogenase 2 (AC, same operon as CanγL) and ammonium-aspartate-lyase (AAL) would lead to the formation of fumarate since this set of enzymes should enable *P. canavaninivorans* to utilize homoserine via aspartate semialdehyde and aspartate yielding fumarate (111). Indeed, when combining purified HD with either of the two aldehyde-dehydrogenases AH or AC and AAL the formation of fumarate via aspartate was observed (Figure 9 and SI Figure 8). Pairwise sequence alignment of AH and AC by EMBOSS Needle (112) showed 67.4% identity between the two oxidoreductases. Moreover, in the genomic context of *P. brassicacearum* 3Re2-7 and *P. fluorescens* F113, close relatives of *P. canavaninivorans*, the corresponding homolog of HD is found in the same genetic locus as CanγL and AC. Considering these findings, we suggest that the aldehyde-dehydrogenases AH and AC have redundant functions and that in *P. canavaninivorans* a gene duplication and reorganization event has split apart the complete canavanine degradation operon present in other *Pseudomonads* into two separate genomic loci.

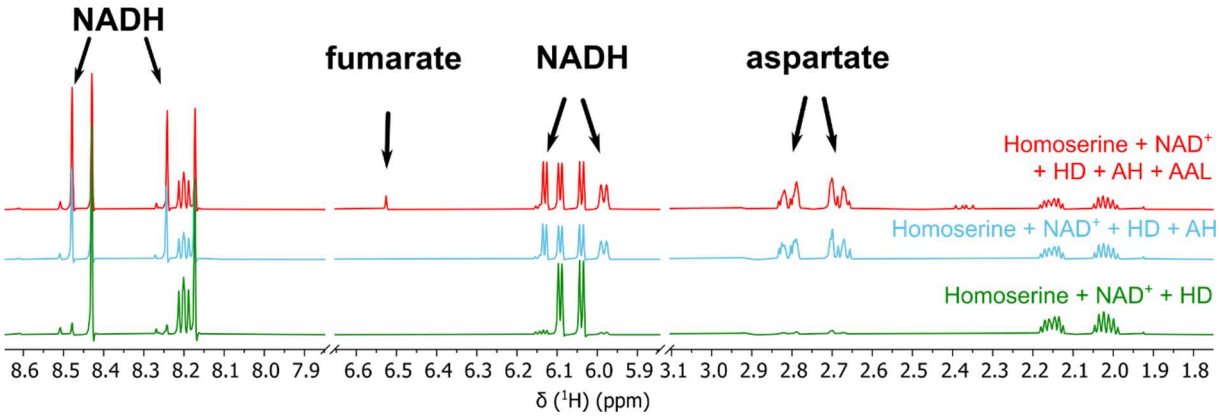


Figure 9: ^{1}H-NMR of the reaction of homoserine to fumarate via homoserine dehydrogenase (HD), aldehyde dehydrogenases (AH) and ammonium-aspartate-lyase (AAL). The reaction with aldehyde dehydrogenase AC and the corresponding reference spectra are shown in SI Figure 8.

It was also observed that the reaction of HD was only dependent on the cofactor NAD^+ and that HD showed no activity utilizing $NADP^+$ (SI Figure 5B). Comparing the sequence of HD to a recently discovered class of homoserine-dehydrogenases showed moderate sequence identity (59.6%) (113). Members of this new class favour the oxidation of homoserine rather than the reverse reaction taking place in the last step of homoserine biosynthesis. The identified HD also displays the reported change in the $NAD(P)^+$ binding motif, explaining why HD rejects $NADP^+$ as co-substrate. In conclusion, the whole pathway from canavanine to fumarate shown in Figure 10 was reconstituted using purified enzymes.

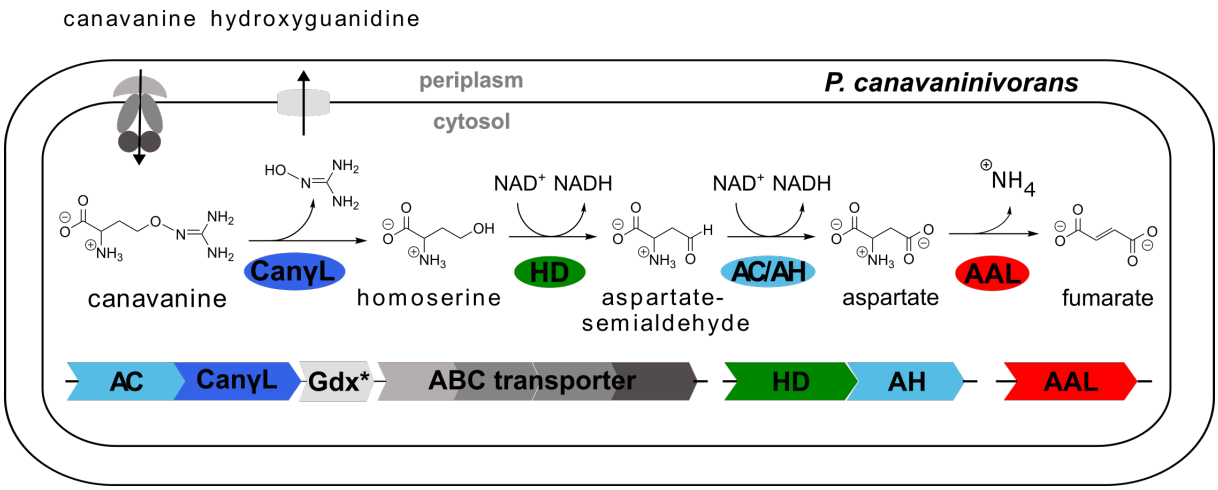


Figure 10: Novel canavanine degradation pathway in *P. canavaninivorans*. AC/AH: aldehyde dehydrogenase, CanγL: canavanine-γ-lyase, Gdx: guanidine exporter, ABC: ATP-binding cassette type transporter, HD: homoserine dehydrogenase, AAL: ammonium-aspartate-lyase. Color indicates that the enzymatic function was experimentally validated in this work.

Distribution of canvanine -γ-lyase in prokaryotic genomes

In order to identify the occurrence of the canavanine degradation activity in other bacteria a neighbor-joining phylogeny search was performed based on canvanine-γ-lyase (CanγL) using the Consurf platform (114) and the BLOSUM62 algorithm (115) and the result were visualized using iTOL v6 (Figure 12A) (116). The novel CanγL enzyme (red) is clearly distant to other validated PLP-dependent enzymes like cystathionine-γ-lyase or methionine-γ-lyase (orange). The most similar homologs to the identified CanγL were found in other *Pseudomonads* (red) and in *Rhizobiales* (blue). Furthermore, when the genomic contexts of the CanγL were compared, only in the genera *Pseudomonas* and *Rhizobium* the homologs clustered together with homologs of the additional enzymes of the pathway (HD, AC/AH, ABC transporters and Gdx) (Figure 12B). Even more striking, in *Rhizobiales* genes coding for AAL and another enzyme (dapA) involved in aspartate semialdehyde metabolism are located in the same genetic locus mirroring that at least in this clade of bacteria canavanine metabolism was optimized by evolution to be represented in a canavanine degradation operon structure

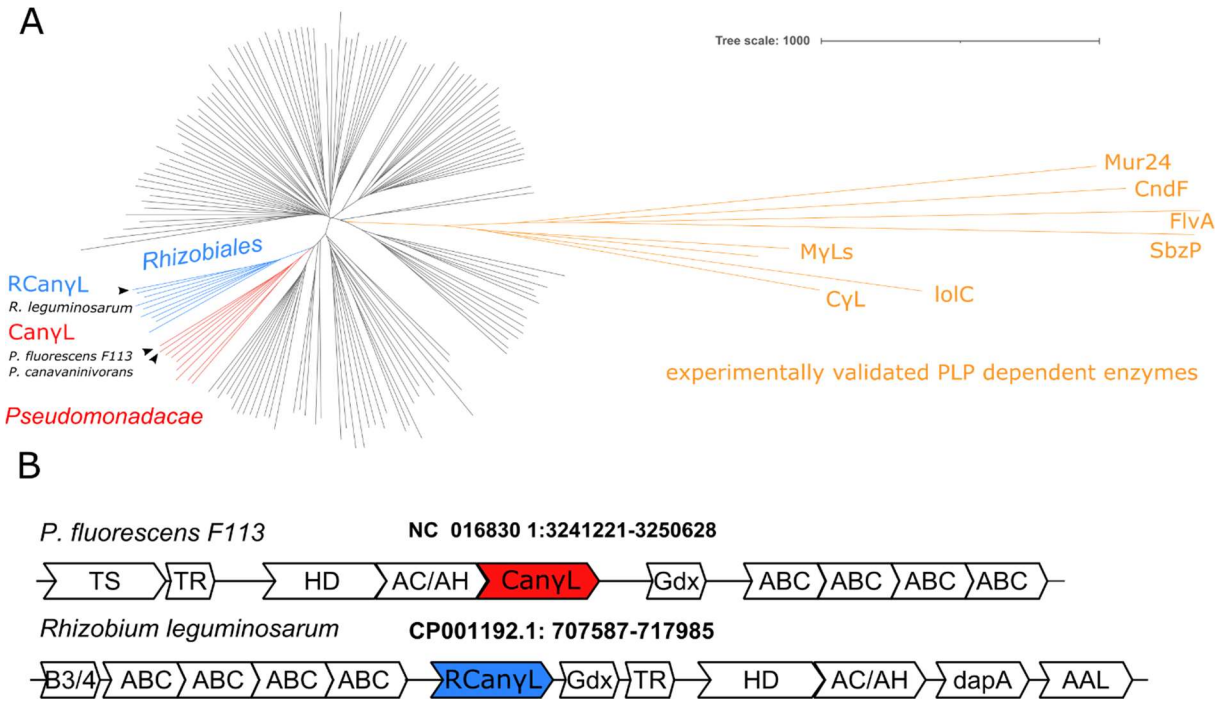


Figure 12: A: Phylogenetic distribution of CanγL based on a homology search. Only for Pseudomonadacae (red) and Rhizobiales (blue) (R)CanγL is found co-localized with genes shown in B. Black clades represent homology-inferred representatives annotated as MγLs, CγLs and other PLP-dependent enzymes. In orange experimentally validated PLP-dependent enzymes are shown, see discussion. **B:** Representatives of the predominant genomic loci of the novel canavanine degradation pathway in Pseudomonadacae and Rhizobiales. Gene annotations: TS: threonine synthase, TR: transcriptional regulator, HD: homoserine dehydrogenase, AC/AH: aldehyde dehydrogenase, AT: aminotransferase (CanγL), Gdx: guanidinium exporter, ABC: importer system, B3/4: potential tRNA editing domain, dapA: 4-hydroxy-tetrahydrodipicolinate synthase, AAL: ammonium-aspartate-lyase.

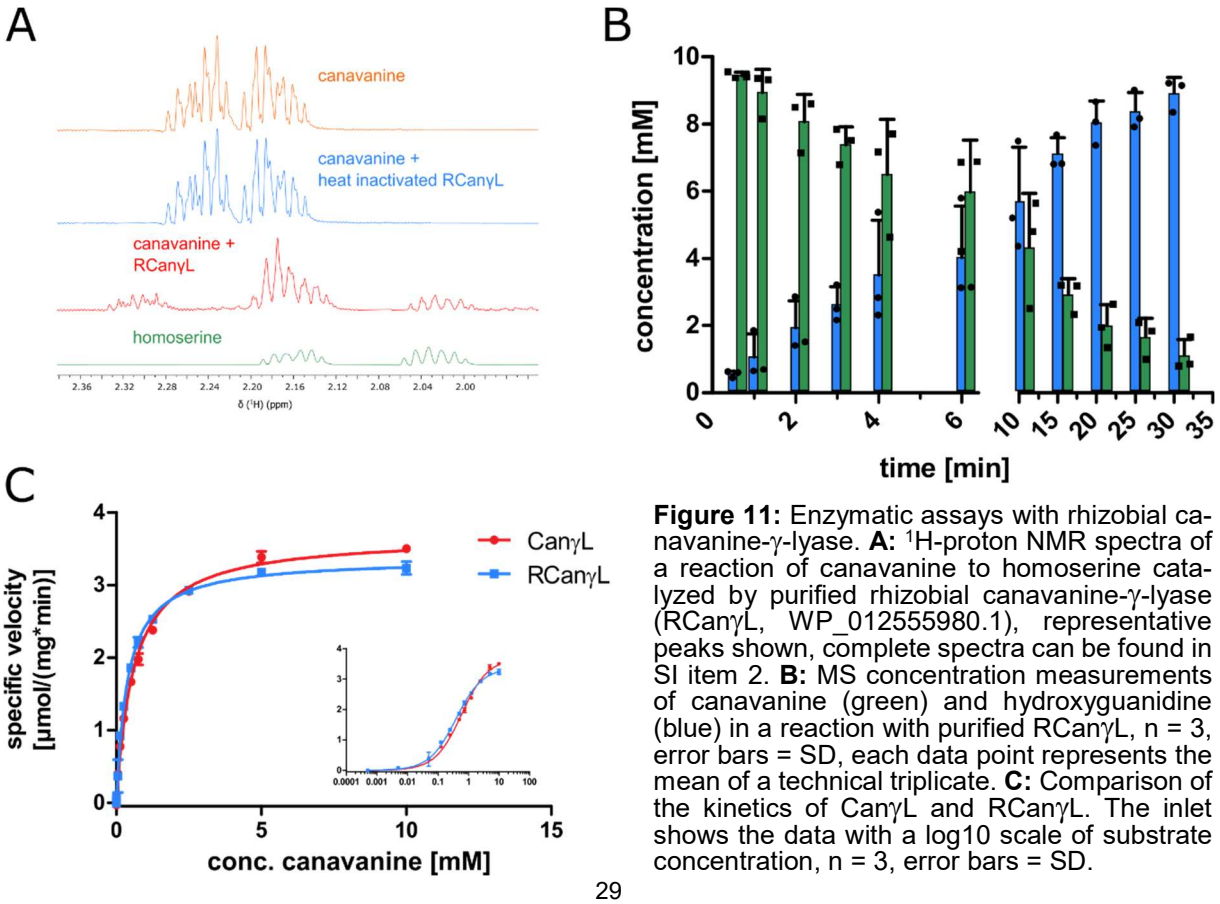


Figure 11: Enzymatic assays with rhizobial canavanine-γ-lyase. **A:** ^{1}H-proton NMR spectra of a reaction of canavanine to homoserine catalyzed by purified rhizobial canavanine-γ-lyase (RCanγL, WP_012555980.1), representative peaks shown, complete spectra can be found in SI item 2. **B:** MS concentration measurements of canavanine (green) and hydroxyguanidine (blue) in a reaction with purified RCanγL, n = 3, error bars = SD, each data point represents the mean of a technical triplicate. **C:** Comparison of the kinetics of CanγL and RCanγL. The inlet shows the data with a log10 scale of substrate concentration, n = 3, error bars = SD.

To test whether the Canγ L homologs identified in *Rhizobiales* are indeed catalyzing the same reaction the respective enzyme of *Rhizobium leguminosarum* (WP_012555980.1) was over-expressed in *E. coli* and canavanine degradation of the purified enzyme was investigated as described before. The rhizobial Canγ L (RCanγ L) also led to the formation of homoserine (Figure 11A) and hydroxyguanidine (Figure 11B) from canavanine. With a K_M of 390 ± 13 µM and a k_{cat} of 1.23 ± 0.011 s^{-1} RCanγ L displayed kinetic parameters similar to Canγ L. Both reactions fitted well to Michaelis-Menten-kinetics (Figure 11C).

Discussion

Taken together, in this chapter it was demonstrated that a set of proteins upregulated upon growth on canavanine composes the first degradation pathway acting specifically on canavanine. The key enzyme, a canavanine-γ-lyase, converts canavanine to homoserine by a PLP-dependent elimination of hydroxyguanidine resulting in a β,γ-unsaturated vinylglycine ketamine intermediate. This intermediate is characteristic for PLP-dependent enzymes of the cystathionine-γ-lyase family although different reaction paths are possible from here (117). For example, in methionine-γ-lyases it leads to the formation of 2-oxobutanoate (99), while in the cystathionine-γ-lyase family-related enzymes CndF (118) and FlvA (119) vinylglycine ketimine reacts as a Michael acceptor with a C-nucleophile. Reactions with N-nucleophiles (IolC (120), Mur24 (121)) and S-nucleophiles (cystathionine-synthase (122), O-acetylhomoserine-sulfhydrylase (123) have been reported as well. Recently, the enzyme SbzP from *Streptomyces* species was described that catalyses Cγ activation with reversed reactivity via an intermediary β, γ-unsaturated quinonoid attacking C4 of the pyridinium ring of β-NAD acting as electrophile (103). In case of the canavanine-γ-lyase described here water is added, resulting in the formation of homoserine. Given the large number of so-far uncharacterized PLP-dependent enzymes with relatively high similarity (Figure 12A) it seems likely that other so-far unrecognized reactions are catalyzed by members of this versatile enzyme class.

The canavanine degradation operon described here facilitates the oxidation of released homoserine to aspartate-semi-aldehyde by a member of a recently discovered new class of homoserine-dehydrogenases (113), followed by oxidation to aspartate by either one of the two aldehyde-dehydrogenases (AC/AH) found induced by canavanine. Then, an ammonium-aspartate-lyase converts aspartate to ammonia and fumarate. The former can be used directly as a source of nitrogen in the cell, while the latter can be used via the tricarboxylic acid cycle

for example as a precursor for amino acids. As already mentioned before fumarate is one of the preferred carbon sources in *Pseudomonas (37)*.

The ABC transporter found in the same operon as CanγL likely functions as a canavanine importer. Since periplasmic substrate-binding proteins are only found associated with importers in bacteria (124), the ABC transporter facilitates the uptake of a substrate. In addition, the substrate-binding domain-containing protein of the upregulated transporter (MBJ2347157.1) is a homolog of an already described arginine-, lysine- and ornithine-binding domain-containing protein of *Salmonella enterica* (WP_079987354.1, 43.5% identity).

The guanidine exporter Gdx also present in the operon likely functions as a hydroxyguanidine exporter since it has been demonstrated recently that Gdx not only efficiently exports guanidine but also amino- and methylguanidine (96). In the growth experiments with 10 mM canavanine as substrate it was observed that hydroxyguanidine is almost stoichiometrically exported with 8 mM measured in the supernatant, even when canavanine serves as the only nitrogen source (Chapter 2.1, Figure 2). The canavanine cleavage product homoserine should be sufficient for supporting growth because the nitrogen from the α-amino group in homoserine satisfies the bacterial biomass need with a carbon:nitrogen ratio of 4:1 (125,126). This explains why the majority of hydroxyguanidine is exported and not used by the bacterium for nitrogen assimilation as we have demonstrated recently for guanidine (127).

The homology search of the canavanine-γ-lyase revealed that the enzyme is also found in *Rhizobiales*. We confirmed that the CanγL homolog of *R. leguminosarum* is indeed degrading canavanine. CanγL was found co-occurring in the same genomic locus with the other enzymes of the described canavanine degradation pathway only in the genera of *Pseudomonas* and *Rhizobiales*. The pathway therefore most likely represents an adaptation to the vicinity of legumes as canavanine is thought to be restricted to these plants (60,61,128). Adding to this picture, it has been shown that the legume *Glycyrrhiza uralensis* exudes canavanine in order to support symbiotic events with rhizobia (62). The authors of the study showed that canavanine induces the expression of the canavanine exporter MsiA in *Mesorhizobium tianshanense*, which confers canavanine resistance. *P. canavaninivorans* also contains a homolog of MsiA (MBJ2350238.1, 40.0% identity, 58.0% similarity), suggesting that certain bacteria might have at least two ways for coping with canavanine: by degrading and/or exporting it. This finding also emphasizes the association of *P. canavaninivorans* with the canavanine-rich legume rhizosphere.

Recently, studying hairy vetch (*Vicia villosa*) it was shown that canavanine is exudated by the roots during early plant development impacting the composition and metabolic activity of the

microbial soil community (63). Furthermore, it was demonstrated for *Sinorhizobium meliloti* that canavanine shuts down the production of the exopolysaccharide EPSII without affecting the growth of the bacterium (129), which has been reported to be a crucial step in the process of nodule invasion (130). Hence, canavanine exudated by legume plants could support nodulation and infection during root development. Canavanine could thereby act both as a signaling molecule for the free-living bacterium to onset its symbiosis with the legume plant as well as serve as a nutrient for both symbiotic and non-symbiotic bacteria in the rhizosphere. *S. meliloti* also contains a homolog of Canγ L (WP_100674444.1, 71.0% identity) in the genetic context of the utilization pathway presented in this work (see operon structure *R. leguminosarum* (Figure 12B).

The finding that the canavanine-γ-lyase deletion strain of *P. canavaninivorans* was unable to grow on canavanine represents additional *in vivo* evidence that the genes described in this work indeed compose a canavanine degradation operon. Also, the poor growth of the B3/4 editing domain-like protein deletion strain raises further questions regarding the function of this protein and its connection to canavanine metabolism, which will be addressed in the next chapter of this thesis.

Even more speculative, canavanine could also play a potential role in establishing the metabolic relationship of the plant and its Rhizobia in analogy to a recently published co-catabolism of arginine and succinate in symbiotic nitrogen fixation (131). The authors found that plants supply not only succinate but also arginine to the nitrogen-fixing bacteria which finally benefits the plant with higher nitrogen yields and the bacteroid with facilitated respiration under the highly acidic conditions present within the symbiosome. Canavanine degradation could be used by the bacterium in a similar fashion to arginine, as the processing of the carbon skeleton would also generate reductive power in form of NADH. Also, recent studies indicate that co-inoculation of Rhizobia with plant growth-promoting rhizobacteria like certain pseudomonads results in larger nitrogen fixation and better growth of the respective legume host under several stress conditions (132–134). Additionally, it was just currently shown for the model plant *Arabidopsis thaliana*, that commensal *Pseudomonas* can protect the plant against very closely related pathogenic *Pseudomonads* and that the host environment indeed affects the microbe-microbe interactions (135). Taken together, modulating the microbiome by offering canavanine and thereby attracting Rhizobia and other plant growth-promoting bacteria capable of utilizing canavanine may foster symbiotic events and help shaping a favorable rhizosphere.

Chapter 2.3: Canavanine detoxification in the context of translation

Introduction

As described in chapter 1, in higher plants canavanine is used as an antimetabolite to arginine and its incorporation into protein can have deleterious effects. Hence, in this chapter it is elucidated how *Pseudomonas canavaninivorans* circumvents the toxicity of canavanine which it encounters in its ecological niche, the legume rhizosphere, and how high translational fidelity is ensured. Interestingly, till now, no editing domain in ArgRS or standalone editing proteins are described that target $tRNA^{Arg}$ (11). Here we report that the native arginine-tRNA-synthetase of *P. canavaninivorans* does not discriminiate sufficiently against canavanine and that a specific standalone editing protein corrects the mischarging. Furthermore, *in vivo* and proteome data confirm that the protein mitigates canavanine toxicity and prevents its incorporation into protein.

Results

ArgRS does not discriminate efficiently against canavanine

The tRNAscan-SE server (136) was used to search for potential $tRNA^{Arg}$ in the genome of *P. canavaninivorans* and three candidate tRNAs were identified. Then, *in vitro*-transcribed and ^{32}P-labelled tRNAs were used to perform aminoacylation reactions employing the ArgRS of *P. canavaninivorans*, which was overexpressed and purified from *E. coli*. Successful arginylation was observed for two of the tRNAs and to a lesser extend also canavanylation (SI Figure **9**). To further test the specificity of the ArgRS several structurally similar compounds were tested as potential substrates. As arginine was tightly binding to ArgRS during protein purification the protocol described by Cvetesic et al. (21) was used to obtain amino acid-free ArgRS. Aminoacylation with arginine and canavanine and to a lesser degree with homoarginine was observed, but no reaction for the negative control agmatine and the substrate analogs canaline, trimethyllysine and D-arginine (Figure 13). Next, the Michaelis-Menten-kinetic parameters of ArgRS with the substrates arginine and canavanine were determined (Figure **14**) to clarify whether the native arginine tRNA synthetase in *P. canavaninivorans* possesses sufficient discriminatory power to prevent potentially harmful canavanyl-charging activity of $tRNA^{Arg}$. Remarkably, the observed K_Ms of the two substrates were very similar (20 and 23 µM), while the observed v_{max} differed by a factor of 7. Hence, the discriminatory factor D ((k_{cat}/K_M (arginine)) /

(K_{cat}/K_M (canavanine))) was calculated to be 7.4. A discriminatory factor of 3000 is assumed to be the lower threshold where mis-incorporation of the non-cognate amino acid does not impact cellular fitness (137–139). Consequently, the finding strongly suggests the need for editing activity towards canavanyl-tRNA^{Arg} in bacteria that are exposed to canavanine. Moreover, the reported half time for canavanyl-tRNA^{Arg} in *in vitro* experiments is ~ 5 minutes compared to ~ 46 min for arginine-tRNA^{Arg} (140), indicating that there might be even less discriminatory potential due to the faster decay of canavanyl-tRNA^{Arg} resulting in a lower observed v_{max}.

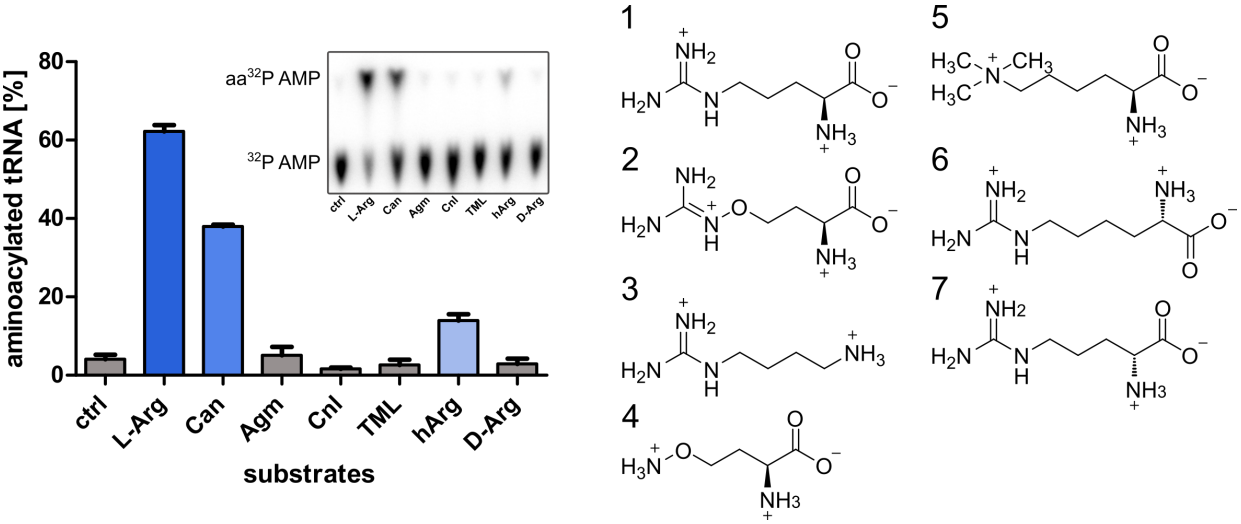


Figure 13: Substrate specificity of the arginine-tRNA-synthetase of *P. canavaninivorans*, error = SD of triplicates. The inlet shows a representative radio screen of the thin layer chromatography separated 32P AMP and aminoacylated 32P AMP. Pictures of all radiographs used to obtain this figure are shown in SI Figure 10. 1: L-arginine, 2: L-canavanine, 3: agmatine, 4: canaline, 5: N-trimethyllysine, 6: homoarginine, 7: D-arginine.

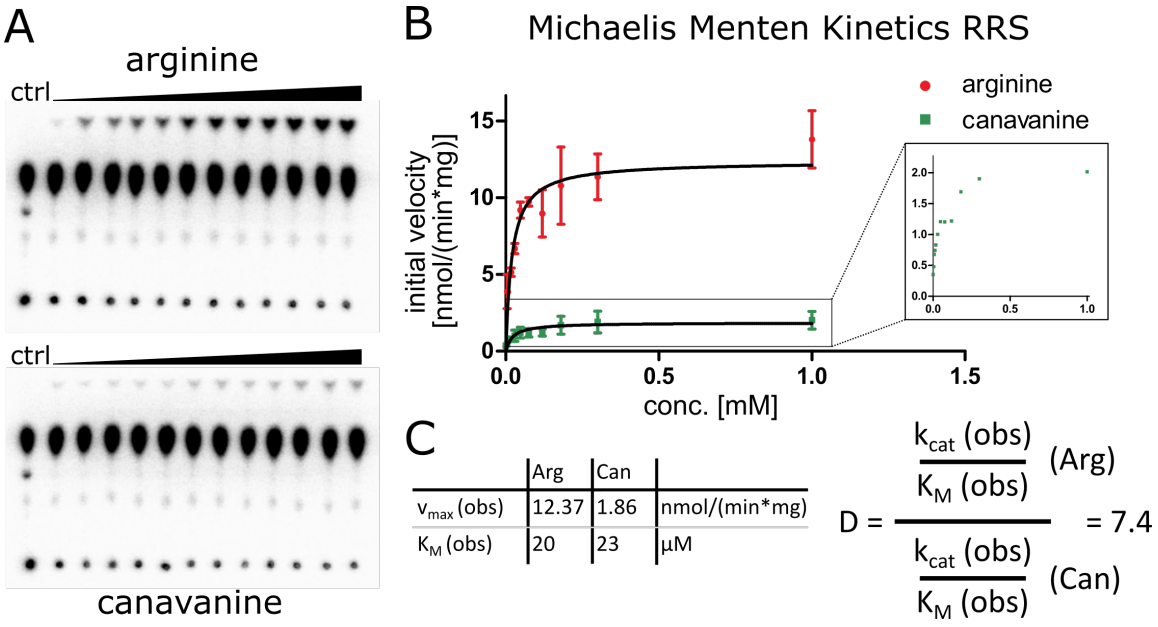


Figure 14: A: Representative radio screen of TLC separated 3' ^{32}P labelled (aminoacylated) AMP. Either arginine or canavanine were added in rising concentrations (0, 4.8, 7.5, 12, 18, 30, 38, 75, 120, 180, 300, 1000 µM) to the acylation reaction. **B:** Reaction velocity plotted over substrate concentration with fit, assuming Michaelis-Menten Kinetics, error = SD of two independent experiments. **C:** Observed Michelis-Menten parameters used for the calculation of the discrimination factor D.

Comparative proteomics reveal the upregulation of a standalone B3/4 editing domain-like protein upon canavanine stimulation

To identify mechanisms that could avoid canavanine incorporation into proteins comparative proteomics of *P. canavaninivorans* cultures grown on canavanine or glycerol as sole carbon source were carried out. Apart from the canavanine utilization pathway described by us recently (95), a protein annotated as B3/4 editing domain-like protein (NCBI MBJ2347151.1, from here on called B3/4 protein) was only detectable in the canavanine-grown culture and not in the culture grown on glycerol. The annotation refers to the similarity to the B3 and B4 domains of the phenylalanine tRNA synthetase subunit β, which were identified to form the editing site of the enzyme (20). Interestingly, the B3/4 protein was located right in front of the canavanine utilization operon, which was also true for a second bacterium, *Rhizobium leguminosarum*. There a homolog of the B3/4 protein (55% identity, 68% similarity) was also co-localized with canavanine degradation genes (Figure 15). Hence, we speculated that this protein is responsible for correcting the mis-charging of $tRNA^{Arg}$ with canavanine. *P. canavaninivorans* as well as *R. leguminosarum* are both natural habitants of the canavanine-rich legume rhizosphere and therefore most likely adapted to circumvent its toxicity.

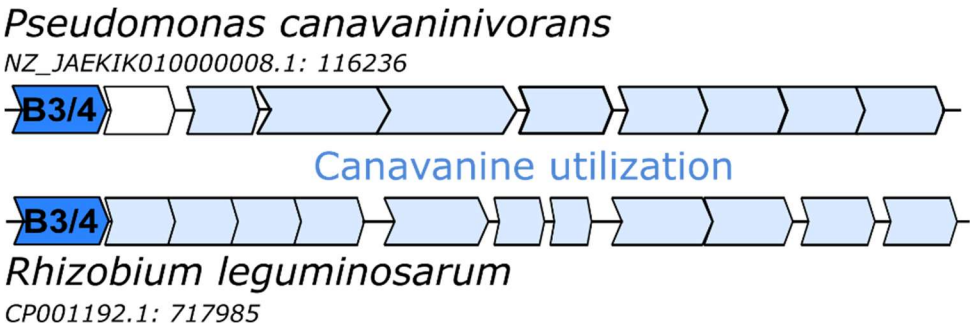


Figure 15: Position of the B3/4 protein in the genome of *P. canavaninivorans* and *R. leguminosarum* relative to the canavanine degradation operon. The number indicates the position in the respective NCBI reference genome.

The B3/4 protein edits canavanylated $tRNA^{Arg}$ in vitro

Upon addition of the B3/4 protein to the aminoacylation reaction with ArgRS, less canavanylated $tRNA^{Arg}$ was observed, while there was no effect on obtained levels of arginyl-$tRNA^{Arg}$ (Figure 3A) indicating that the B3/4 protein can edit canavanyl-$tRNA^{Arg}$ or stimulate editing activity in ArgRS. As the natural modifications of $tRNA^{Arg}$ could play an important role in the recognition process by ArgRS or B3/4 protein and thereby influence aminoacylation and editing the aminoacylation reaction was carried out with $tRNA^{Arg}$ overexpressed in *P. canavaninivorans.* The *in vivo*-produced tRNA comprises a pool of all expressed tRNAs with the desired

tRNA^{Arg} enriched by overexpression. Although changes in the overall charging levels were observed due to the different method of tRNA^{Arg} preparation, editing of canavanyl-tRNA^{Arg} by the B3/4 protein was even more effective compared to the reaction with *in vitro*-transcribed tRNA^{Arg} (Figure 3B). The observed level of canavanylated tRNA upon addition of B3/4 protein was as low as the background without added amino acid, whereas the arginylation was again unaffected. Also, no editing activity was apparent for homoarginylated tRNA^{Arg} (SI Figure 1 and SI Figure 4), indicating that the B3/4 protein activity might be specific for canavanyl-tRNA^{Arg}.

To clarify if the editing protein needs ArgRS or if it has a standalone catalytic activity purified canavanyl-tRNA^{Arg} was incubated with the B3/4 protein. To compensate for the instability of canavanyl-tRNA^{Arg} (140) the samples were only incubated for one minute and B3/4 protein added in high excess (50 µM protein to 3 µM canavanylated tRNA^{Arg}) to achieve rapid deacylation. After incubation with B3/4 protein, almost complete deacylation of canavanylated tRNA^{Arg} was reached for *in vitro* as well as *in vivo*-produced tRNA^{Arg} (Figure 3C), indicating that ArgRS is not needed for editing activity.

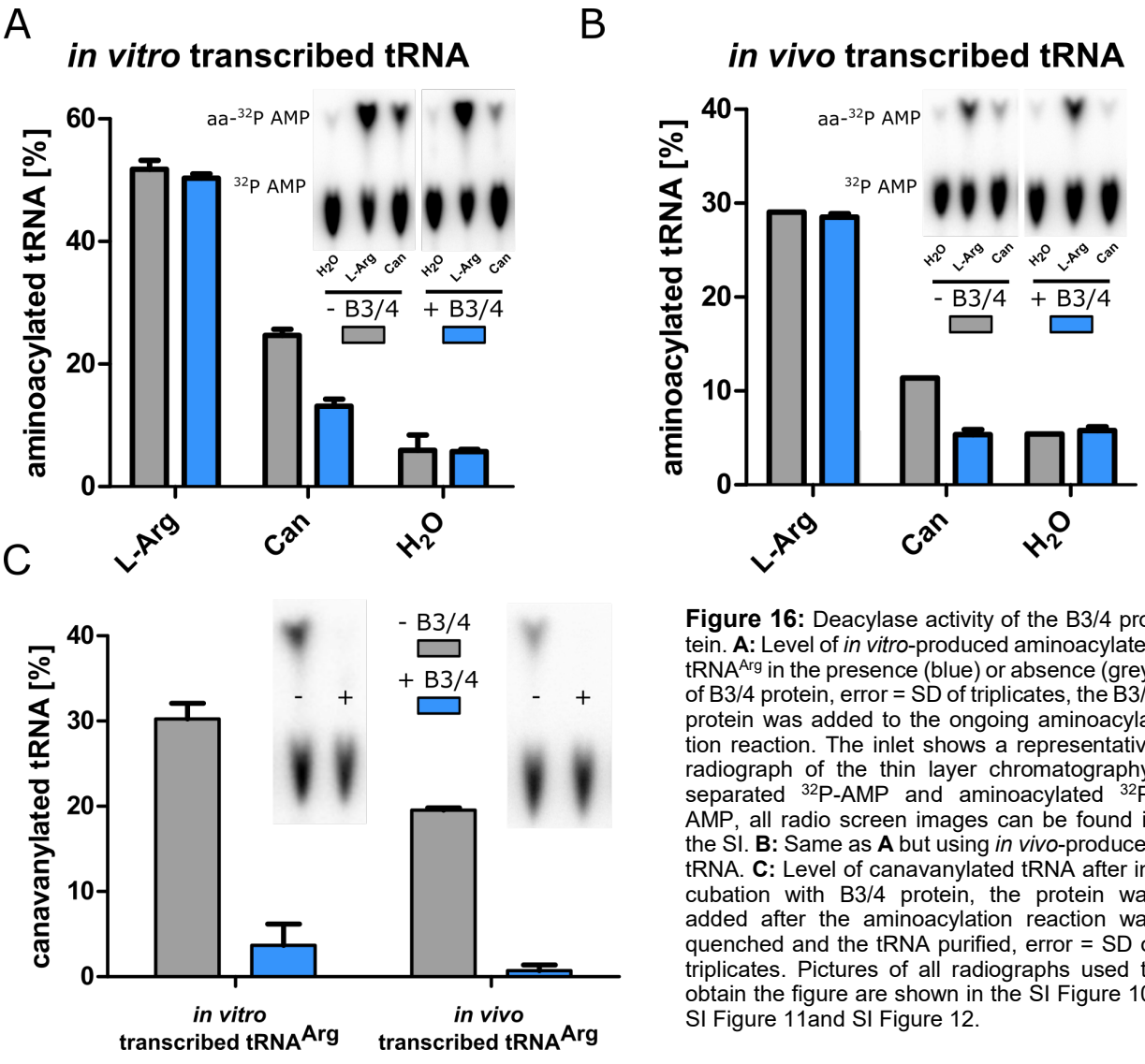


Figure 16: Deacylase activity of the B3/4 protein. **A:** Level of *in vitro*-produced aminoacylated tRNA^{Arg} in the presence (blue) or absence (grey) of B3/4 protein, error = SD of triplicates, the B3/4 protein was added to the ongoing aminoacylation reaction. The inlet shows a representative radiograph of the thin layer chromatography-separated ^{32}P-AMP and aminoacylated ^{32}P-AMP, all radio screen images can be found in the SI. **B:** Same as **A** but using *in vivo*-produced tRNA. **C:** Level of canavanylated tRNA after incubation with B3/4 protein, the protein was added after the aminoacylation reaction was quenched and the tRNA purified, error = SD of triplicates. Pictures of all radiographs used to obtain the figure are shown in the SI Figure 10, SI Figure 11and SI Figure 12.

The B3/4 protein mitigates canavanine toxicity in vivo

To elucidate a potential impact of the B3/4 protein´s activity on bacterial viability and growth, growth experiments were carried out with the *P. canavaninivorans* deletions strains ΔB3/4 and ΔB3/4 ΔCanγL. The double mutation renders *P. canavaninivorans* unable to detoxify canavanine by degradation, thereby promoting its potentially deleterious effect. In M9 minimal medium with canavanine as sole carbon source the wildtype (wt) strain was able to grow, while as expected the double deletion strain ΔB3/4 ΔCanγL was unable to grow on canavanine alone (Figure 17A). In accordance with the proposed editing activity, the ΔB3/4 protein single deletion strain was strongly impaired in growth in comparison to the wildtype strain. To exclude a general growth deficit due to the gene deletions, the three strains were cultivated in minimal medium with glucose as carbon source supplemented with different levels of canavanine. In the absence of canavanine no notable differences in growth were observed, whereas upon addition of 1 mM canavanine, growth rates differed between the strains, and the effect became more pronounced upon addition of 9 mM canavanine. In the presence of canavanine, wildtype *P. canavaninivorans* grew better than the two mutant strains and among these the double deletion strain grew worst.

B3/4 protein prevents incorporation of canavanine into the proteome

To evaluate if the growth impairment was due to the incorporation of canavanine into the proteome and whether the activity of the B3/4 protein is impacting the proteome composition in the presence of canavanine, the wildtype and the two mutant strains were grown in canavanine-containing M9 minimal medium and their proteomes were analysed by mass spectrometry to investigate whether canavanine was incorporated into proteins instead of arginine. The incorporation of canavanine was detected during data evaluation as an arginine modification with a mass shift of +2. Additionally, incorporation of canavanine instead of arginine almost completely blocks the tryptic digest during sample preparation (141). The combination of mass shift and missed cleavage hence provides a reliable way to identify misincorporation of canavanine. Only 0.15% of the detected peptides matched both criteria (mass shift + missed cleavage) in the wt culture without canavanine, whereas in a wt culture with 3.5 mM canavanine the percentage already doubled to 0.3%. In the single deletion strain ΔB3/4 we found approximately 5.6% of all detected peptides canavanylated, which corresponds to a 19-fold change compared to wt, while in the double deletion strain ΔB3/4 ΔCanγL the effect was even more pronounced (10% canavanylated peptides, 33-fold compared to wt) (Figure 17B). The proteome data provides further evidence that the B3/4 protein prevents canavanine incorporation into proteins via deacylation of canavanylated tRNAs. A scheme of the proposed function of B3/4 is presented in Scheme 4.

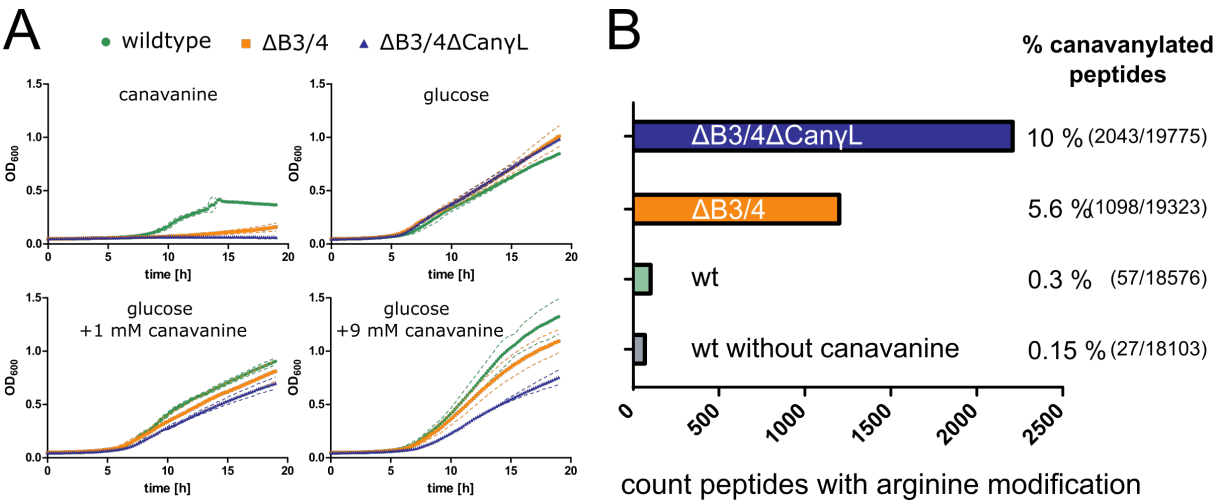
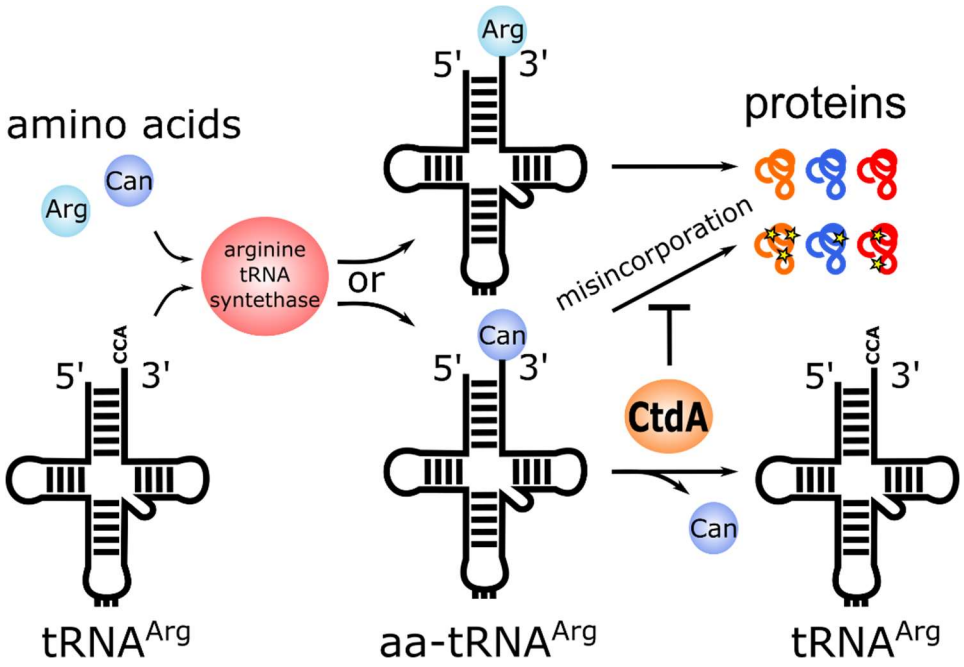


Figure 17: Relevance of canavanyl-tRNA^{Arg}-editing *in vivo.* **a:** Growth curves of *P. canavaninivorans* wt, ΔB3/4 and ΔB3/4 ΔCanγL strains grown on minimal M9 medium with either canavanine, glucose or glucose + canavanine as carbon source, dashed line = SD of triplicates. **b:** Peptides counted with arginine modifications and missed cleavages in a whole proteome analysis. Indicated bacterial strains were grown on minimal M9 medium with glucose and 3.5 mM canavanine, '% canavanylated peptides' shows the number of peptides found to be mass-shifted in combination with a missed tryptic cleavage compared to the overall peptides found in the sample.

Discussion

In this chapter a standalone protein annotated as B3/4 editing domain-like protein that specifically deacylates canavanyl-tRNA^{Arg} is described. The enzyme was found upregulated when the bacterium *Pseudomonas canavaninivorans* grew on canavanine as sole carbon source (95). Aminoacylation experiments showed that ArgRS of *P. canavaninivorans* does not discriminate sufficiently between arginine and canavanine and that the B3/4 protein is editing canavanyl-tRNA^{Arg} as a standalone protein. We therefore propose to rename this specific class of B3/4 proteins to canavanyl-tRNA^{Arg} deacylase (CtdA). Growth experiments with *P. canavaninivorans* gene deletion strains confirm the physiological importance of B3/4 editing activity and proteome data strongly suggest that impaired growth of the deletion strains is due to massive canavanine misincorporation into the proteome. A homolog of CtdA is also found in other rhizosphere inhabitants like *R. leguminosarum* where it is similarly located directly upstream of an operon coding for a canavanine degradation pathway. Furthermore, opportunistic bacteria like *Klebsiella pneumonia* which are attracted to the rhizosphere because it is a nutrient-rich ecological niche (142) obtain a B3/4 protein (Identity 67.5%, Similarity 78.8%, compared to *P. canavaninivorans* CtdA). Hence, the enzymatic activity of CtdA appears to be part of the bacterial response to encountering canavanine and could be crucial for successful rhizosphere colonialization.



Scheme 4: Scheme of the proposed activity of CtdA. Arg = arginine, Can = canavanine. Stars represent canavanine incorporation into proteins, which is presumably toxic to the cell.

The finding of a standalone enzyme able to discriminate arginyl-tRNA^{Arg} and its close structural analog canavanyl-tRNA^{Arg} is in itself remarkable. As mentioned before, the acylated tRNAs are very similar with regard to their size and shape. However, the electron-withdrawing effect of the oxa-group in canavanine changes the pK_a of the guanidinium group drastically from 13.8 to approximately 7 (143,144). A potential basis for discrimination could hence be the different binding of charged and uncharged guanidine residues where only the substrate complex with the uncharged sidechain results in catalytic activity. Moreover, the oxaguanidine group of canavanine could potentially take part in the catalytic mechanism in a co-factor-like manner. Due to the lowered pK_a, the canavanine sidechain could engage in general acid-base catalysis to support cleavage of the ester bond between canavanine and the tRNA. These speculations offer a promising starting point for a more thorough investigation of the catalytic mechanism of this novel subclass of tRNA deacylases with a potentially distinct reaction mechanism.

Chapter 3: Guanidine riboswitch-associated gene functions

General Introduction

The nitrogen rich compound guanidine is used widely in the manufacturing of plastics, as reactant in certain explosives and propellants and as a chaotropic reagent in biochemistry. A plethora of guanidine-containing compounds exist in nature, ranging from arginine, canavanine and creatinine to guanine and large secondary metabolites such as streptomycin. The idea that guanidine might have a prominent but so far overlooked role in natural physiology or metabolism was sparked by a series of publications from Breaker and co-workers who described three classes of riboswitches that respond to guanidine (145–147). The recent discovery of a fourth class of guanidine-riboswitches further supports this idea (148,149). The *ykkC* motif RNA was published as a riboswitch candidate in 2004 (150–152), and is common in various bacterial clades and associated with genes encoding transporters like multidrug efflux pumps, urea carboxylases, purine biosynthesis and amino acid metabolism enzymes. Two additional riboswitch candidates, the mini-*ykkC* (153,154) and the *ykkC*-III RNA motif (155,156) were subsequently identified to be found upstream of several of these genes. Although the three motifs do not share sequence or structural characteristics in the ligand-binding domain, it was hypothesized that they sense the same ligand, based on the extensive overlap of the genetic context (157,158). The wide variety of associated genes hampered the search of the ligand based on the genetic context, but also indicated that its ligand participates in widespread metabolic reactions (150,158). All three *ykkC* motifs, now re-named as guanidine-I, -II and –III riboswitches (145–147), and the recently published guanidine-IV(148,149) riboswitch were verified to respond selectively to guanidine.

Several questions regarding guanidine arose due to the widespread occurrence of guanidine riboswitches in bacteria: 1. What are the sources of guanidine in nature? 2. Is it an end product or can it be utilized? 3. What is the biological role or purpose? Addressing the potential sources, hardly any biotic reactions have been described that produce guanidine. Growth experiments with *E. coli* indicate towards the production of guanidine when grown on minimal medium, but its source is unknown (145). The best studied enzyme that catalyses the production of guanidine is the ethylene-forming enzyme (EFE). There, guanidine is formed *via* the δ-hydroxylation of arginine and subsequent loss of guanidine (159,160). Also, as mentioned already before, two early reports propose the bacterial degradation of canavanine (δ-oxa-arginine) to homoserine and guanidine (161,162), but the molecular mechanisms remain to be elucidated. In the light of the above described canavanine degradation pathway discovery it appears most likely that by comparative proteomics also those mechanisms might be decoded.

Regarding the second question the overwhelming number of exporters associated with guanidine riboswitches suggested mainly the role of an end product, but the recent discoveries of two degradation pathways contradict this notion. First, we and others discovered a widespread bacterial utilization pathway of guanidine as a nitrogen source via carboxylation (127,163), which will be described within the following. Second, Funck et al. discovered the direct hydrolysis of guanidine to urea and ammonium by a Ni^{2+}-dependent guanidine hydrolase (164). Last but not least, considering guanidine's biological purpose, in this chapter data is presented that gives insights into guanidine's role apart from serving as a nitrogen source.

Chapter 3.1: A guanidine riboswitch-associated carboxylase pathway enables the utilization of guanidine as sole nitrogen source in bacteria

Introduction

A gene commonly controlled by guanidine riboswitches encodes certain representatives of "small multidrug resistance" (SMR) transporters such as YkkCD, EmrE and SugE. Subsequent to the discovery of guanidine riboswitches, this specific family of SMR proteins has been demonstrated to function as selective guanidine exporters, termed Gdx (for guanidine exporters) (96). When SugE-type genes occur under guanidine riboswitch control, they will be referred to as Gdx throughout the thesis. Another gene product that is frequently controlled by guanidine riboswitches is annotated as urea carboxylase (Uca). It has been demonstrated already during the initial discovery of guanidine-I riboswitches that these carboxylases can use both urea and guanidine as substrates (145). The studied riboswitch-controlled carboxylase showed a 40-fold lower K_M for guanidine compared to urea but a comparable k_{cat}. Thus, the riboswitch-associated carboxylase annotated as urea carboxylase prefers guanidine as a substrate. Carboxylase enzymes that occur under guanidine riboswitch control will be referred to as guanidine carboxylases (Gca). It was speculated that the carboxylation reaction initiates the degradation of guanidine for the purpose of detoxification (145). When urea is carboxylated, the resulting product allophanate is further hydrolysed by the enzyme allophanate hydrolase (165). Indeed, many guanidine riboswitch-regulated operons contain genes annotated as allophanate hydrolases. However, guanidine carboxylation results in carboxyguanidine, which might require different hydrolysis activities compared to allophanate. Two further genes (annotated as urea carboxylase-associated genes 1 and 2; ucaa1 and 2) often associate with guanidine carboxylase enzymes. A recent report clarified that these two proteins comprise the subunits of a heterodimeric carboxyguanidine deiminase (CgdAB)) enabling its hydrolysis to allophanate and ammonia (163). The work also addresses the question of substrate specificity

of the associated carboxylases and speculates about its involvement of the use of guanidine as nitrogen source.

Taken together, guanidine riboswitches predominantly induce Gdx transporters in order to export this compound from bacterial cells before it reaches problematic concentrations. However, other gene functions controlled by guanidine riboswitches enable the carboxylation and subsequent degradation. In this chapter it is shown, that the carboxylase pathway enables the utilization of guanidine as sole nitrogen source. The isolation and characterization of guanidine-assimilating bacteria is described and it is demonstrated that carboxylase enzymes have evolved which display selectivity for either urea or guanidine. In addition, metagenomics data were analysed in order to characterize habitats that are enriched for carboxylase pathway enzymes for the utilization of guanidine as nutrient.

Results

Gene functions under control of guanidine riboswitches

Since the discovery of widespread riboswitches that induce gene expression in response to the presence of guanidine, the physiology of guanidine has remained a mystery. In order to shed more light on the physiology of guanidine in bacteria, we sought to investigate genes that are commonly associated with the four known riboswitch classes for guanidine. For a comprehensive analysis the information of the genetic context was combined for all four known riboswitch classes in the available genomic data. Since certain gene functions under guanidine riboswitch control seem to occur in conserved operons that contain groups of highly associated genes, the gene categories were grouped into different pathways: As mentioned before the gene classes most frequently controlled by guanidine riboswitches encode certain representatives of “small multidrug resistance” (SMR) transporters, which have been demonstrated to act as Gdx (96). They occur as individual genes or tandem arrangements (reminiscent of their homo- or heterodimeric topology) and appear only sometimes associated with other genes (Figure 18). The second-largest group of genes associated with guanidine riboswitches comprise a pathway that consists in most cases of ATP binding cassette-type (ABC) transporters, accompanied by genes that encode a urea carboxylase, two different urea carboxylase-associated genes, and allophanate hydrolases (Figure 18). We noted that organisms that regulate urea/guanidine carboxylase-containing pathways by guanidine riboswitches often (in 113 of 152 organisms, 74%) also regulate ABC-type transporters in this way (Figure 18). This suggests that the ABC-type transporters, which are not structurally related to Gdx exporters, are somehow functionally connected to the carboxylase pathway gene. If the associated ABC-type transporters are exporters, this would mean that the guanidine-controlled carboxylase operons

contain two different means to detoxify this compound, export as well as modification via carboxylation and subsequent hydrolysis. Moreover, in 52.2% of the organisms with carboxylase pathway-associated ABC-type transporters, a Gdx-type exporter is also found in the same genome (59 from 113 occurrences, see Figure 18). Hence these organisms would encode two different transporters for the export of guanidine. Such a high frequency of redundancy would be surprising. Instead, the ABC transporter could work together with the carboxylase and imports its substrate guanidine. We therefore set out to investigate the guanidine riboswitch-associated carboxylases in more detail.

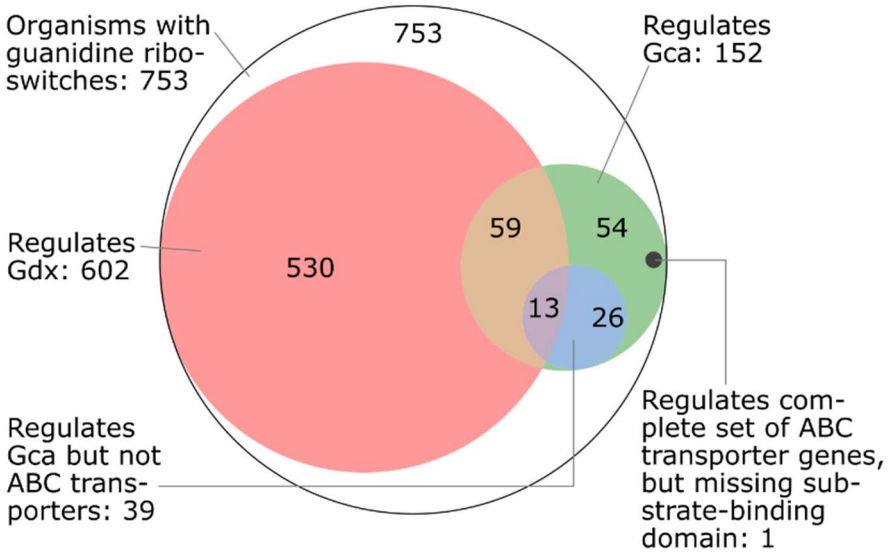


Figure 18: Association of guanidine-I, -II, -III and -IV riboswitch classes with specific gene functions. The largest circle represents all 753 organisms with complete genomes that contain at least one predicted guanidine riboswitch. The sets indicate which organisms have guanidine riboswitches that regulate guanidine exporters Gdx (602 organisms) or guanidine carboxylase Gca genes (152 organisms), and also considers the regulation of ABC transporter genes by guanidine riboswitches. 530 organisms only regulate Gdx by guanidine, 80 organisms only regulate Gca in this way and 72 organisms do both. Of 113 riboswitch-regulated operons encoding ABC transporters, 112 contain a substrate-binding domain.

Guanidine riboswitch-associated ABC transporters could function as importers

In order to clarify the role of the ABC transporters associated with the carboxylase pathway, the encoded genes were more carefully analysed. ABC transport systems can either import or export compounds (166,167). In bacteria, these two activities can be distinguished easily by the composition of the subunits: Importers are characterized by the presence of a periplasmic substrate binding protein (in Gram-negative) or a lipid-anchored external protein (Gram-positive) that delivers the transported substrate to the outer side of the ABC-type channel (124,168). In the case of exporters, this domain is not necessary and is lacking. Therefore the guanidine riboswitch-controlled carboxylase operon-associated ABC transporters were further analysed and periplasmic substrate binding domains were found in 112 of the 113 cases (Figure 18). This finding implies that these carboxylase pathway-associated ATP transporters

function rather as importers instead of detoxifying guanidine via export. The prospect of active, ATP hydrolysis-driven import into bacteria opens up the possibility that guanidine is utilized as a nutrient.

Enrichment of guanidine utilizers

Next, in analogy to the isolation of canavanine utilizing bacteria, microorganisms were enriched that are able to utilize guanidine as nitrogen source. Water samples from the lake shore surface sediment of Lake Constance (Bodensee) were collected and filtered and diluted samples were plated on minimal media that contained glycerol as carbon source and guanidine as sole nitrogen source. With undiluted lake water samples, densely overgrown plates were observed at 25°C over 24 h. In diluted samples, many colonies could be identified. From these colonies, three morphologically different ones were chosen for enrichment and subsequent characterization. Individual colonies were passaged several times on selective media in order to obtain homogenous strains. By 16S rRNA gene sequencing the three isolated bacteria were identified as strains of *Raoultella terrigena* (169), *Erwinia rhapontici* (170), and *Klebsiella michiganensis* (171).

In order to characterize the utilization of guanidine as N-source, the three isolated strains were cultivated in liquid culture and growth was monitored. As comparison, the bacteria were grown on minimal medium with glycerol as C-source and supplemented guanidinium chloride (5 mM), urea (10 mM), or ammonium chloride (15 mM) as sole nitrogen source. All strains were able to grow efficiently on guanidine (Figure 19). The bacteria grew slightly slower on guanidine compared to ammonia. Nevertheless, they reached the same maximal optical density indicating that all nitrogen atoms of guanidine can be assimilated.

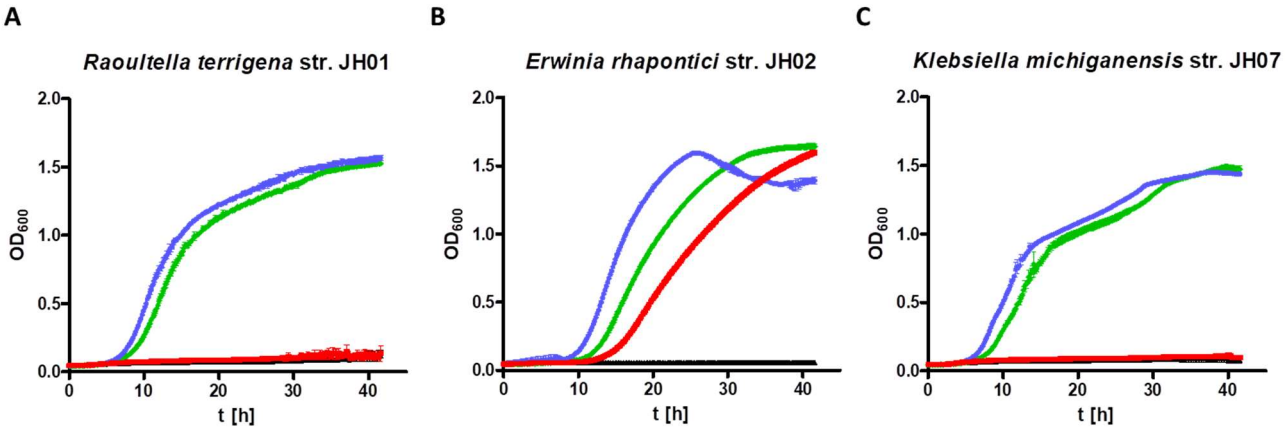


Figure 19: Bacterial utilization of guanidine as nitrogen source. Growth of isolated bacterial strains in minimal medium containing 5 mM guanidine (green), 10 mM urea (red), 15 mM ammonia (blue) and no nitrogen source (black). Values are means of biological triplicates with SD.

Interestingly, only *E. rhapontici* was able to grow on urea as N-source. This result was somewhat unexpected as the isolated *R. terrigena* and *K. michiganensis* strains both encode urease genes, see below. *E. rhapontici*, which is able to utilize urea efficiently, does not encode a urease, but rather two copies of annotated urea carboxylases, see below for a detailed characterization of the two enzymes. We speculated that the difference in growth on urea could be related to the difference of the two means of urea degradation. Urease is a Ni^{2+}-dependent metallohydrolase (172). When the growth experiments were repeated with supplemented Ni^{2+}, both *R. terrigena* and *K. michiganensis* grew to the same optical density (SI Figure 13). The inability of *R. terrigena* and *K. michiganensis* to utilize urea in absence of supplemented Ni^{2+} demonstrates that the encoded riboswitch-controlled guanidine carboxylases (see next paragraph) do not also allow for the utilization of urea but seem to be specific for guanidine.

In order to investigate the molecular basis for the observed utilization of guanidine, the genomes of the three strains were sequenced. The analysis of the resulting genome sequences *Raoultella terrigena* JH1 (NZ_CP050508.1), *Erwinia rhapontici* JH2 (JABANZ000000000.1) and *Klebsiella michiganensis* JH7 (JABANY000000000.1) utilizing TYGS (76) confirmed the taxonomic classification based on the 16S rRNA gene analysis. Next the occurrence of guanidine riboswitches and associated genes in the obtained genomes (Table 5) was analysed. All three strains contain a guanidine carboxylase operon under the control of a guanidine riboswitch. The organization of the genes is very similar among the bacterial strains with the ABC transporter genes and the carboxyguanidine deiminase genes (*cgdAB*) under control of a guanidine-I riboswitch followed by guanidine carboxylase (*gca*) and allophanate hydrolase genes (*atzF*) under the control of a guanidine-II riboswitch (Figure 20A). In *Erwinia rhapontici* str. JH02 *atzF* is missing downstream of the riboswitch-associated *gca.* However, *atzF* is found at a different locus in conjunction with a second urea/guanidine carboxylase gene cluster. Both *R. terrigena* and *K. michiganensis* genomes contain a Gdx-type exporter under guanidine riboswitch control whereas this activity is lacking in *E. rhapontici*.

Analysis of guanidine-dependent gene expression

To explore whether the observed utilization of guanidine as N-source is based on the activity of the riboswitch-controlled gene products, the proteomes of the three isolated strains in response to guanidine were analysed. Each of the isolated bacteria was grown in minimal medium that contained 1% glycerol as carbon source and either 5 mM guanidinium chloride or 15 mM ammonium chloride as nitrogen source. Bacteria were harvested in late exponential phase and the whole proteome was determined by mass spectrometry. In all three bacteria the genes under the control of guanidine riboswitches (Table 5) were highly upregulated (Figure 20B).

Moreover, nitrogen metabolism-related genes were found generally enhanced when bacteria utilized guanidine which could be due to the lack of a preferred nitrogen source.

Table 5: Guanidine Riboswitches in Isolate Genomes

Organism	Genome Accession	Riboswitch	Coordinates[a]	Regulated Genes[b]
Raoultella terrigena str. JH1	CP050508.1	Guanidine-I	2977411-2977313 (-)	Abc-transporter[c], *cgdA, cgdB*
		Guanidine-II	2973102-2973053 (-)	*gca, atzF*
		Guanidine-II	5159112-5159068 (-)	*gdx*
Erwinia rhapontici str. JH2	JABANZ000000000.1	Guanidine-I	402759-402857 (+)	Abc-transporter[c], *cgdA, cgdB*
		Guanidine-II	406997-407043 (+)	*gca*
Klebsiella michiganensis str. JH7	JABANY010000008.1	Guanidine-I	84523-84621 (+)	Abc-transporter[c], *cgdA, cgdB*
		Guanidine-II	88830-88878 (+)	*gca, atzF*
	JABANY010000005.1	Guanidine-II	205465-205421 (-)	*gdx*

a: (+) plus strand; (-) minus strand,
b: *gca*: guanidine carboxylase; *cgdA, cgdB* : Carboxyguanidine deiminases A and B; *atzF*: Allophanate hydrolase; *gdx*: Guanidine exporter
c: Three subunits present: Periplasmic substrate-binding, Transmembrane, and ATP-binding domains

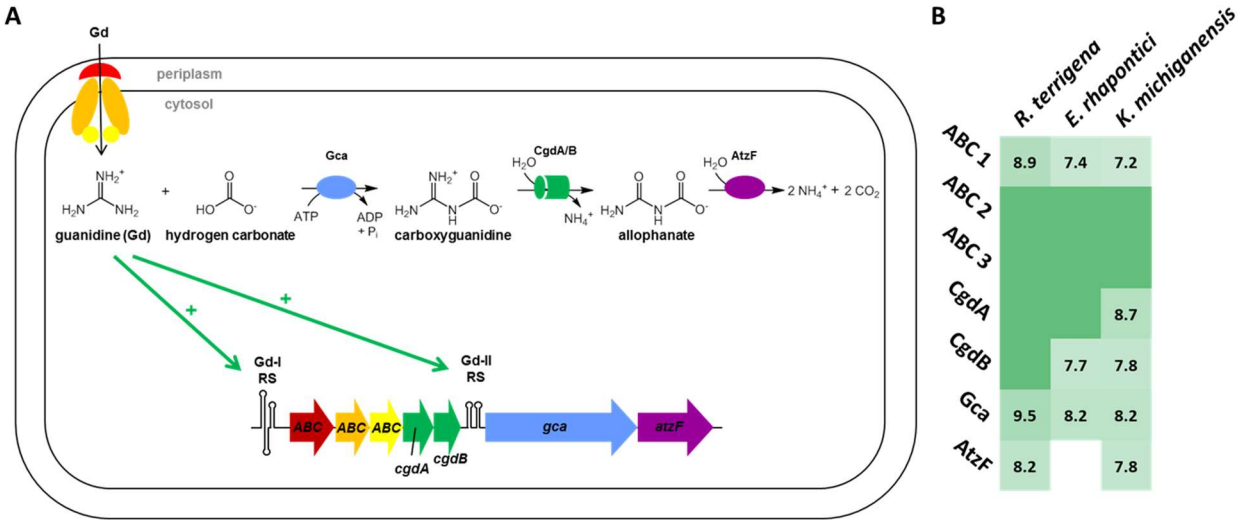


Figure 20: A: Scheme of the investigated guanidine assimilation pathway and its genetic organization in *Raoultella terrigena* str. JH01 that is representative for the three isolated strains, except that *atzF* of *Erwinia rhapontici* str. JH02 is encoded at a different genetic locus. Guanidine induces (green arrows) gene expression by binding to the riboswitches (Gd-I and Gd-II RS). **B:** Proteome data showing the induction of the guanidine assimilation operons during growth on guanidine compared to growth on ammonia. Changes in the protein abundancy are expressed as log2 values. Dark green indicates that proteins were identified only when bacteria were grown on guanidine. ABC 1: periplasmic substrate binding domain; ABC 2: transmembrane domain; ABC 3: ATP-binding cassette.

Characterization of carboxylase enzymes

Within the analysed genome and proteome data of *E. rhapontici*, two different urea/guanidine carboxylase genes were found to be encoded. The first carboxylase (WP_171149239.1) is not riboswitch-controlled, whereas the second (WP_171148480.1) is under control of a guanidine riboswitch, see Table 5. As both were up-regulated upon growth on guanidine their roles in guanidine utilization were further investigated. Both urea/guanidine carboxylases were over-expressed as His-tagged versions in *E. coli* and purified via Ni-NTA in order to determine their substrate preferences. The carboxylation reaction consumes ATP stoichiometrically with regard to substrate turnover (173). The reaction was monitored as described before (165,174,175) by coupling the ATP-consuming carboxylation reaction to the reactions of pyruvate kinase and lactate dehydrogenase, which result in the oxidation of NADH. The decrease of NADH can be monitored spectrophotometrically. By plotting the initial velocity over the substrate concentration the kinetic parameters K_M and k_{cat} could be obtained. The data for both carboxylase enzymes fitted well to Michaelis-Menten-kinetics (Figure 21A&B). The calculated parameters are shown Table 6.

The riboswitch-associated carboxylase (Gca, WP_171148480.1) prefers guanidine over urea with K_M values of 0.058 ± 0.005 mM and 5.3 ± 0.3 mM, respectively. k_{cat} for both substrates is almost the same, 5.9 ± 0.1 s^{-1} for guanidine and 5.3 ± 0.3 s^{-1} for urea. Interestingly, the non-riboswitch-associated carboxylase (Uca, WP_171149239.1) prefers urea (K_M = 0.085 ± 0.009 mM) over guanidine (K_M = 17 ± 4 mM) as substrate and k_{cat} is much lower compared to the guanidine carboxylase (with 0.13 ± 0.01 s^{-1} for guanidine and 0.25 ± 0.01 $s^{-1}$for urea). Thus, the substrate preference of the enzymes is interchanged. The urea carboxylase has a 375-fold higher specificity constant (k_{cat}/K_M) for urea compared to a 90-fold higher specificity constant of guanidine carboxylase for guanidine. As saturation was not reached for the respective poorer substrate, kinetic parameters for those substrates should be taken with care. However, saturation was almost reached as visualized by plotting the data on a linear scale (SI Figure 14). Since *E. rhapontici* encodes no urease enzyme, the non-riboswitch-associated urea carboxylase (WP_171149239.1) is likely responsible for the observed Ni^{2+}-independent utilization of urea as N-source (Figure 19).

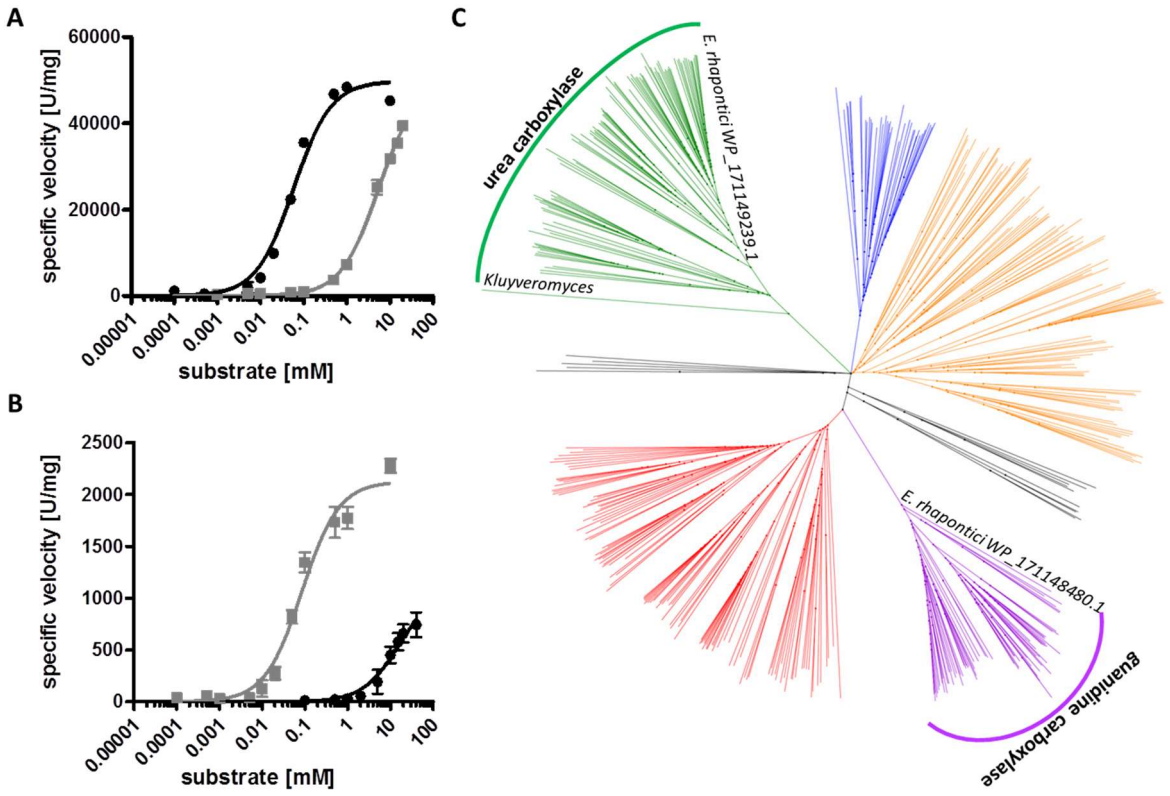


Figure 21: Substrate specificity of the two urea/guanidine carboxylase enzymes encoded by *E. rhapontici* and their phylogenetic relationship. **A:** Characterization of the riboswitch-associated carboxylase WP_171148480.1 and **B:** non-riboswitch-associated carboxylase WP_171149239.1 of *E. rhapontici JH02*. The initial velocity (1 U equals µmoles NAD^+ generated per min) per mg of enzyme is plotted against substrate concentrations for guanidine (black dots) and urea (grey squares). Error bars represent SD from triplicates. **C:** Neighbour joining phylogeny of urea/guanidine carboxylases and their homologues. Homology search and multiple sequence alignment were performed with Consurf (114). The sequences of the urea carboxylase and guanidine carboxylase were aligned subsequently using Jalview(176). Neighbour joining phylogeny was performed with BLOSUM62 (115). The tree was illustrated with iTOL(177).

Table 6: Michaelis Menten kinetic parameters of Uca and Gca. Kinetic parameters were obtained from Michael Menten fit to data in Figure 21.

Enzyme/substrate	k_{cat} [s^{-1}]		K_M [mM]		k_{cat}/K_M [$s^{-1}mM^{-1}$]	
	guanidine	urea	guanidine	urea	guanidine	urea
urea carboxylase WP_171149239.1	0.13 ± 0.01	0.25 ± 0.01	17 ± 4	0.085 ± 0.009	0.008 ± 0.002	3.0 ± 0.2
guanidine carboxylase WP_171148480.1	5.9 ± 0.1	5.3 ± 0.3	0.058 ± 0.005	5.3 ± 0.3	100 ± 4	1.1 ± 0.04

Distribution of urea and guanidine carboxylase enzymes

Previously, it has been noticed that in the substrate binding pocket of the then described urea carboxylase of *O. sagarensis* an aspartic acid contacts the substrate(165,145). Already Nelson et al. hypothesized that the carboxylate in the sidechain of the aspartate is preferably binding the positively charged guanidinium rather than being protonated and binding urea, explaining the lower K_M of guanidinium. Interestingly, another aspartate on the opposite site of the binding pocket contacting the biotin cofactor is replaced by asparagine in the urea carboxylase of *Kluyveromyces lactis* as it is also the case in the non-riboswitch-associated urea carboxylase (WP_171148480.1) of *E. rhapontici* studied above. This aspartate (Asn1330 of the *K. lactis* structure (163) could be indicative for the substrate specificity of urea carboxylases. Hence, a homology search starting from *K. lactis* urea carboxylase was performed followed by a multiple sequence alignment using the Consurf platform(114). Subsequently the urea and guanidine carboxylase sequences from *R. terrigena*, *E. rhapontici* and *K. michiganensis* were aligned and neighbour joined using the BLOSUM62 algorithm (115). The homologue sequences clustered in five major and three minor clades (Figure 21C). Remarkably, *K. lactis* and the non-riboswitch-associated urea carboxylase (WP_171149239.1) clustered in one clade (green), whereas the two studied riboswitch-associated guanidine carboxylases clustered in another clade (purple). All other clades comprise families of yet uncharacterized carboxylases. Hence it was concluded that urea/guanidine carboxylases can be divided in at least two specific enzyme clades: urea carboxylases and guanidine carboxylases. Interestingly, the asparagine residue is only found in the binding pocket of the urea carboxylase (green) clade whereas all other clades in Figure 21C comprise aspartate at that position. If the occurrence of the aspartate residue in the binding pocket is indeed indicative for guanidine specificity, it seems that guanidine carboxylation is the much more widespread activity of this class of enzymes compared to urea carboxylation. However, more representatives from other clades need to be tested in order to support such a conclusion.

Distribution of guanidine-utilizing carboxylases in metagenomes/habitats

So far it was demonstrated that the guanidine-controlled operons encoding ABC-type transporters and carboxylases, carboxyguanidine deiminases, and allophanate hydrolases enable the uptake and assimilation of guanidine. This result is contrasted by widespread occurrence of riboswitch-controlled Gdx-type exporters of guanidine. In order to shed more light on the physiology of guanidine utilization in nature, the occurrence of both pathways in certain bacterial habitats was investigated. When the occurrence of guanidine riboswitches in known organisms is taken into account, one notes that the switches are found widely distributed in many phyla of bacteria. An analysis is complicated by the fact that many bacteria are ubiquitously

distributed and it is not always possible to determine whether a given bacterium has a predominantly water-, soil-, plant- or animal-associated life style. In order to nevertheless extract information about the habitat where guanidine might play a pronounced role, metagenome data were analysed to see whether the occurrence of a riboswitch-controlled activity can be connected to a given environment. Since for these datasets there is always a more or less specific sampling and therefore the isolated sequences are likely to be typical of the given habitat, riboswitches were analysed that occur in metagenome data and the frequency of occurrence of Gdx-type exporters and guanidine carboxylase-type utilization pathways (Gca^P) to the annotated habitat (Figure 22 and online in the supplementary of the original publication (127)) was correlated.

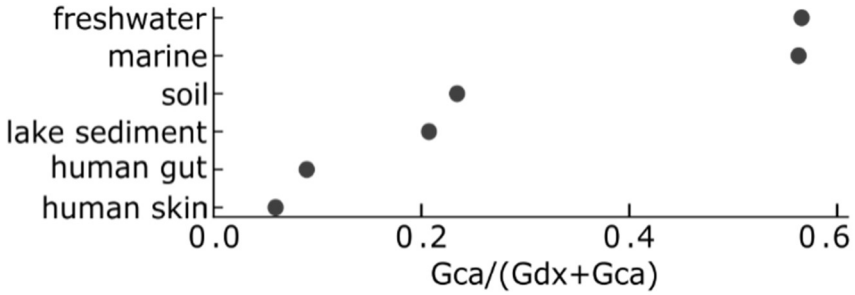


Figure 22: Occurrence of Gdx^P and Gca^P activities regulated by guanidine riboswitches in different environments. Gca^P: the number of occurrences of genes assigned to the guanidine carboxylase pathway that are regulated by guanidine riboswitches in datasets from the given environment. Gdx^P: the number of *gdx* genes controlled by guanidine riboswitches. For the selected environments, Gca^P+Gdx^P is at least 285.

Representative habitats such as human skin and gut metagenome data as well as environmental samples such as soil and aqueous habitats were selected, in which Gdx-type exporters and Gca^P (related) genes under control of guanidine riboswitches were identified. Next, the frequency of the occurrence of Gca^P in the metagenome data in relation to the sum of Gca^P and Gdx^P was calculated, which allowed to gain insight into whether riboswitch-controlled genes in a given habitat prefer guanidine utilization via Gca^P or Gdx^P-mediated export of guanidine, in comparison to other environments. In Figure 22, the X-axis value is zero if all riboswitches in the environment regulate Gdx-type exporter genes, and one if all riboswitches regulate Gca^P related genes. Environments are sorted from those that relatively favour guanidine carboxylase-mediated utilization (freshwater) to those that relatively favour Gdx^P-mediated export (human skin). Interestingly, it seems that the occurrence of the guanidine-utilizing carboxylase pathway negatively correlates with the nutrient- and nitrogen-richness of the respective habitat. Animal-based microbiomes thrive under nutrient-rich conditions, whereas aquatic habitats are often nitrogen-scarce environments(178). Hence, riboswitch-controlled Gca^P activity is more prevalent in N-scarce environments such as fresh water, marine, and soil samples whereas in N-rich habitats such as the human gut guanidine carboxylase pathway genes are found less frequently in comparison to Gdx exporters.

Discussion

In the presented data it was shown that guanidine is utilized by bacteria using riboswitch-controlled carboxylases and hydrolases. ABC-type importers are also often encoded under control of guanidine riboswitches, likely facilitating the efficient uptake of the N-rich compound. Given that the guanidine carboxylases are found widespread in bacteria it seems likely that the utilization of guanidine as N-source is a common activity in many organisms. It seems that a single amino acid in the active site is responsible for determining the substrate specificity of the carboxylase reaction. This finding has also been reported recently when for the first time the role of the associated *ucaa* genes was clarified as carboxyguanidine deiminases (163).

The ATP-dependent carboxylation and subsequent hydrolysis of urea for its utilization has been first described in yeasts and algae in 1968 (179). However, it took until 2004 for the same activity to be described in bacteria (165). It has been noted before that the urease-mediated and the urea carboxylase/allophanate hydrolase-mediated reactions are two apparently redundant means of degrading urea (160). The co-occurrence of bacterial urease and urea carboxylase in one organism led Hausinger to speculate that one of the enzymes might catalyse an alternative reaction. As demonstrated here, it seems that the majority of organisms encode enzymes that should show greater specificity towards guanidine than urea carboxylation, nevertheless carboxylases with higher specificity towards urea do exist, such as the examples in *E. rhapontici* or in *S. cerevisiae* and *C. albicans (163)*. Vice-versa, considering that the two additional bacteria isolated (*K. michiganensis* and *R. terrigena*) are not able to grow efficiently on urea without supplemented Ni^{2+} for urease activation, it seems that the guanidine carboxylases that these two organisms encode are so specific that they are not able to hydrolyse urea in sufficient amounts in order to sustain growth in absence of a urease activity.

With regard to a possible redundancy of urea degradation in organisms that contain both urease and urea carboxylase it might be advantageous for certain bacteria specialized on the utilization of such compounds to invest in the maintenance of genes encoding both systems. Although the urease-dependent direct hydrolysis of urea is more straightforward and seems more energy-economic than the ATP-dependent detour via the carboxylated intermediate, it requires Ni^{2+} as cofactor that might not always be available in sufficient amounts. Such a scenario was observed in our experiments when the two strains *R. terrigena* and *K. michiganensis* did not grow in minimal media with urea as sole N-source unless Ni^{2+} was supplemented. In addition, several additional proteins responsible for the modification of the active site, Ni^{2+} loading, and Ni^{2+} homeostasis are necessary for the activation of urease (180).

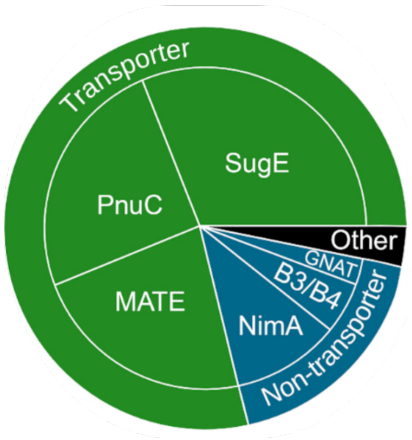
Taken together here it is shown that guanidine utilization is widespread and predominantly found in bacterial organisms living in nutrient-scarce environments. The three isolated bacteria

seem to have specialized on the utilization of alternative nitrogen sources since all of them possess activities for the assimilation of guanidine and urea. Additionally, two of the isolates (*R. terrigena* and *K. michiganensis*) possess genes necessary for N_2 fixation, a rather rare feature among enterobacteria. Guanidine utilization is carried out via carboxylation and subsequent hydrolysis. The carboxylase enzymes appear to be specific for guanidine, although urea-specific homologs also exist, sometimes even in the same organism as we have found for *E. rhapontici*. Interestingly, *R. terrigena* and *K. michiganensis* also contain Gdx-type guanidine exporters. Similar to the Gca pathway enzymes Gdx is also under control of a guanidine-dependent on-riboswitch. It can be speculated that at low nitrogen concentrations guanidine is utilized via the Gca activities and at high nitrogen and guanidine concentrations the Gdx exporter is getting rid of excess guanidine. Such a scenario could be facilitated by placing the expression of the Gca pathway under control of a nitrogen limitation-responsive control mechanism. Some evidence supporting this notion can be seen in the proteome data where nitrogen limitation activities are up-regulated in general when guanidine is offered as sole nitrogen source. Furthermore, a bioinformatics method has been presented in order to assign a given biochemical activity to certain habitats by surveying the occurrence and frequency of certain genes in metagenome data from specific habitats.

Chapter 3.2: Beyond nutrition – an additional role of guanidine in nature

Introduction

In this part of the thesis a potential role of guanidine beyond serving as a nitrogen source is presented. The great majority of guanidine riboswitch-associated genes represent exporters but there are several additional genes which might allow to shed light into the hidden purpose of guanidine, apart from genes like the above described carboxylases that serve nutritional purposes. In their recent publication regarding the discovery of a fourth class of guanidine riboswitch (GuaRS-IV) Lenkeit et al. (2020) (148) analysed the frequency of riboswitch-associated genes and among the afore mentioned transporters the authors found the following genes to be most common: an enzyme belonging to the pyridoxamine 5' phosphate oxidase superfamily (NimA), a GCN5-related N-acetyltransferase protein (GNAT) and to our surprise also a gene annotated to code for a B3/4 editing domain-like protein (Figure 23). Interestingly, genes coding for B3/4 editing domain-like protein were also found to be associated with guanidine riboswitches class I and class IV (Figure 23) (145,148,181). Hence, a representative B3/4 protein (GuaRS-I associated) which obtained 35% sequence identity to CtdA was tested for its ability to hydrolyze canavanyl-tRNA^{Arg} in an aminoacylation assay.



		Frequency among riboswitches (%)			
Name	Description	GGAM-1/ Guan. -IV	Guan. -I	Guan. -II	Guan. -III
SugE	SugE/EmrE/SMR/Gdx transporter family	31	17.6	56.5	82
PnuC	PnuC-like transporter family	25.2	0.09	0.02	0
MATE	Multidrug And Toxic compound Extrusion (MATE) transporter family	22.6	0.1	0.04	0
NimA	Pyridoxamine 5' phosphate oxidase superfamily	10.5	0.08	0	0
B3/4	Phenylalanine- and lysidine- tRNA synthetase domain superfamily	5	6.2	0.02	0
GNAT	GCN5-related N-acetyltransferase family (GNAT)	2.5	0.04	0.02	0.03
Other		3.2			

Figure 23: Genes frequently associated with GuaRS-IV/ *GGAM-1* RNAs. The six conserved protein domains most commonly encoded by genes that are immediately downstream of *GGAM-1* RNAs. Such genes are almost certainly regulated by the riboswitch, in comparison to genes that might be located in extended operons. Three domains function as transporters (green), while the other three do not (blue). Less common domains and domains that did not match the Conserved Domain Database were classified as 'other' (black). Table: Percentage of riboswitches/motifs for all 4 known classes of guanidine riboswitches associated with the described gene. More detailed information can be found in the cited publication, adapted from Lenkeit et al. (2020) (148).

Results

Encouragingly, the B3/4 protein found in *Clostridium perfrigens* (WP_004456252.1) indeed prevented the accumulation of canavanylated $tRNA^{Arg}$ in the aminoacylation assay as effective as CtdA of *P. canavaninivorans* (Figure 24A). Interestingly, the bacteria obtaining B3/4-annotated proteins controlled by guanidine riboswitches are mostly found in microoxic or anaerobic habitats like the gut of higher animals where canavanine exposure could occur via the ingestion of canavanine-producing legumes. The observation that those B3/4 proteins could be involved in canavanine metabolism hints at the possibility that the presence of guanidine acts as an indicator for canavanine exposure. Since canavanine is structurally very similar to arginine it might be more easy to sense a specific degradation product of canavanine instead of the molecule itself. Guanidine detection would then indicate the presence of canavanine, a potentially toxic antimetabolite that warrants additional measures such as an increased proofreading of $tRNA^{Arg}$. In light of this hypothesis, the question arose whether hydroxyguanidine, the so-far confirmed degradation product of canavanine, would also trigger a guanidine riboswitch. Especially guanidine class I riboswitches have been reported to be even more specific with regard to guanidine recognition, even rejecting substrates that are structurally very similar, such as aminoguanidine and methylguanidine (145). In order to characterize the response to hydroxyguanidine, a Gd-I riboswitch from *Pseudomonas pelagia* was inserted into the 5'-UTR of a lacZ gene and its expression in a lacZ-deficient *E. coli* strain in the presence of potential riboswitch ligands was measured using ONPG as substrate. Interestingly, in addition to the expected guanidine also hydroxyguanidine triggered a riboswitch response (Figure 24B) whereas other close analogs did not show induction of gene expression. In addition to using commercial hydroxyguanidine, it was also produced from canavanine by pre-incubation with canavanine-γ-lyase (Can + / Can –). Enzymatically produced hydroxyguanidine also triggered the riboswitch. The finding supports the hypothesis that in some cases guanidine or hydroxyguanidine could act as indicators for the presence of guanidine-producing compounds such as canavanine.

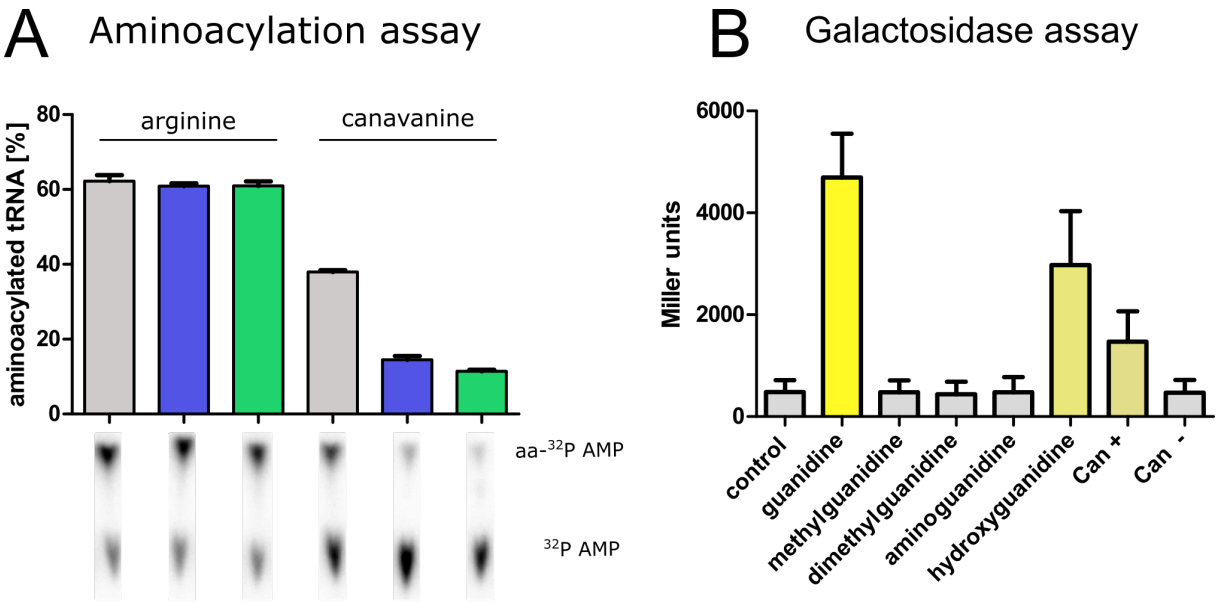


Figure 24: A: Level of *in vitro* produced aminoacylated $tRNA^{Arg}$ in the presence (blue: *P. canavaninivorans*, green: *C. perfrigens*) or absence (grey) of B3/4 protein, error = SD of triplicates. Pictures of all radioscreens used to obtain this figure are shown in (SI Figure 12). B: Miller units represent relative conversion of *ortho*-Nitrophenyl-β-galactoside to the yellow dye *ortho*-Nitrophenol by lacZ. The indicated compounds were added to the growth medium to a final concentration of 2.5 mM. Control: H_2O, Gua: guanidine, Can ±: canavanine incubated with or without CanγL o/n.

Additional evidence supporting the theory of guanidine as a canavanine indicator compound was found when acetylation experiments were performed with a representative GuaRS IV associated GNAT found in *Lactobacillus curiae*. Successful acetylation was observed with canavanine as a substrate and the activity was substrate concentration dependent (Figure 25).

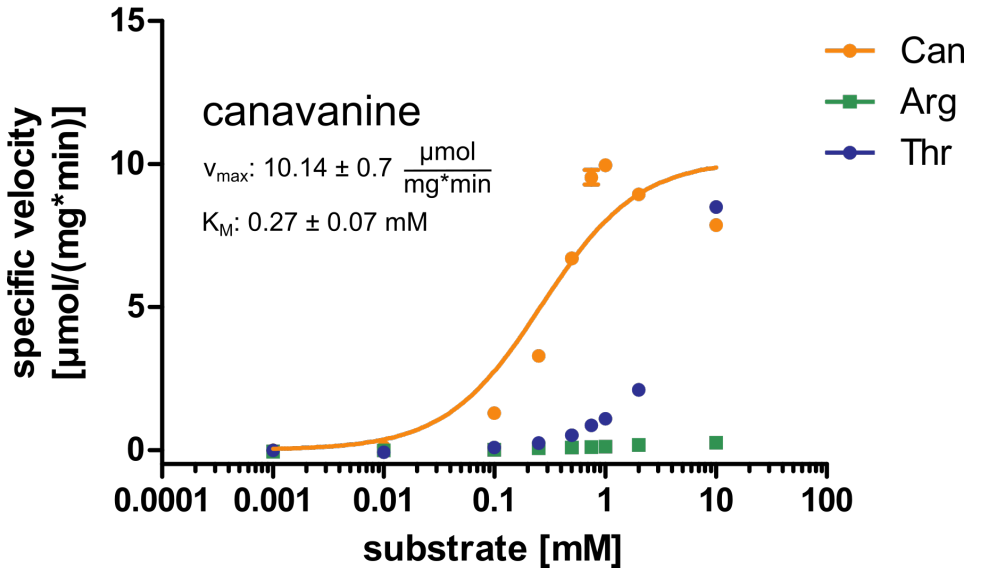


Figure 25: DTNB assay with purified GNAT from *L. curiae*. The fit was performed assuming Michaelis-Menten kinetics although most likely substrate inhibition is taking place at higher concentrations (>1 mM). Can: canavanine, Arg: arginine, Thr: threonine.

Activity at high substrate concentrations was also apparent for threonine and other amino acids (SI Figure 15), which is a common observation for this class of acetyltransferases (182), but hardly any activity was observable for arginine, which is remarkable as the two substrates canavanine and arginine are so similar.

To get insights into the exact mechanism of discrimination crystal trials were set up for the GNAT of *L. curiae.* The protein could be successfully crystallized and a model was built starting from the crystal structure of a GNAT from *Streptococcus mutans* (PDB entry 4E2A). The parameters can be found in SI Table 3. Unfortunately, co-crystallization with the cofactor acetyl-coenzyme A (AcCoA) and the substrate canavanine was not successful up to now, so no experimentally confirmed insights into the exact mechanisms reaction and discrimination mechanism can be presented. Still, the cofactor AcCoA could be modelled into the binding cleft/tunnel. Figure 26 shows the obtained structure with the modelled cofactor, alone (top - left) and superimposed (top - right) to structurally close GNATs, the 3D model showing the surface electrostatics (middle –left: front, middle right: behind, the cross section through the binding cleft/tunnel (bottom left) and the overall model showing all the possible AcCoA binding models within the protein (bottom right). The overall structure of the GNAT corresponds to the general topology found in the enzyme family and it consists of seven β-strands and four α-helices. In accordance to literature, which states that the least conserved secondary structure elements include the termini (183) the secondary elements are not following the typical GNAT composition (β0 β1 α1 α2 β2 β3 β4 α3 β5 α4 β6) (184) but miss a N-terminal and contain an additional C-terminal β-strand (β0 α1 α2 β1 β2 β3 α3 β4 α4 β5 β6). Modelling of the cofactor acetyl-coenzyme A (AcCoa) and a conservation score search using Consurf (114) confirmed the presence of the characteristic pyrophosphate binding loop (P-loop) that is in proximity to the pyrophosphate of the cofactor. Interestingly, in contrast to the reported consensus motif (R/Q-X-X-G-X-A/G) (184, 185) the riboswitch associated GNAT obtains a slight modification and the P-loop is one amino acid shorter (R104-X-G-X-G108). Regarding the second characteristic feature of cofactor binding, the β–bulge, the protein also obtains the V-shaped splaying of the β–sheets β3 and β4 which accommodate the pantothenate moiety of AcCoa. Modelling of canavanine into the cleft/tunnel was not successful because the potential binding site is too spacious.

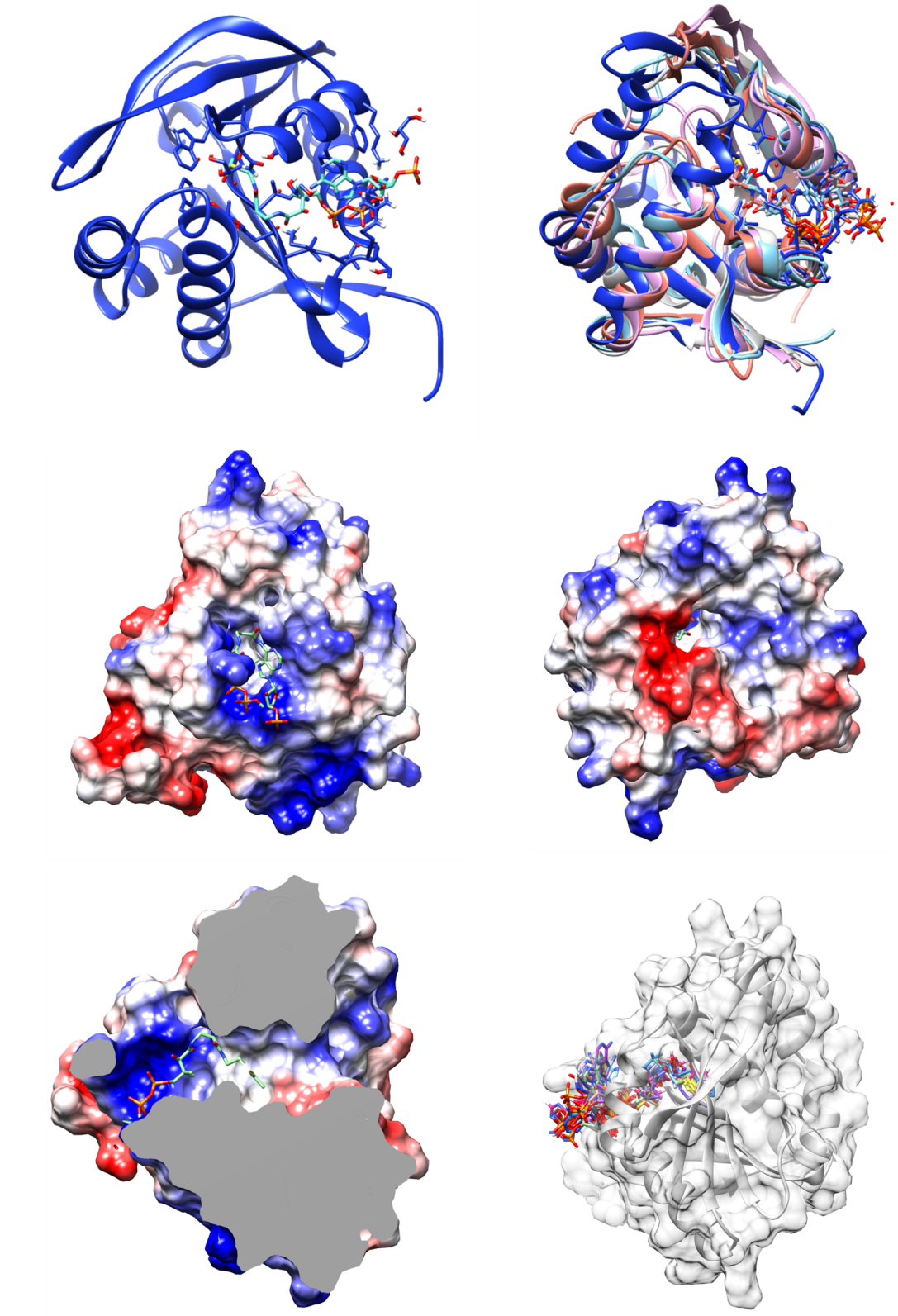


Figure 26: Structure of the GNAT of *L. curiae* derived by crystallography. Top left: Structure with the modelled cofactor Acetyl-Coenzyme A. Top right: Superimposed structure with structurally similar GNATs (PDB entries: 3LD2, 3PP9, 2CNM). Middle left: Front view of the 3D structure showing the electrostatics and the surface of the GNAT. Middle right: Back view. Bottom left: Crosssection of the binding tunnel together with the cofactor. Bottom – right: Overall structure with all potential localizations of Acetyl-Coa.

Discussion

The herein presented data suggest towards a connection between guanidine and canavanine metabolism, as within this thesis it is shown that two of the three most common guanidine IV riboswitch-associated non-transporter gene classes (B3/4 (now CtdA) and GNAT) have enzymatic activity on canavanine. Between the two enzymes the riboswitch-mediated control of CtdA expression gives the strongest hint towards an intricate linkage between canavanine and guanidine metabolism, as for the other enzymes additional work needs to be conducted to fully grasp their metabolic role and purpose. Regarding the GNAT, a recent publication by Bikmetov et al. (186) showed that GNATs exist that have evolved as part of a toxin/antitoxin system and target Gly-tRNA isoacceptors. Thereby they disrupt translation and inhibit cell growth. In light of those findings it is tempting to speculated that canavanyl-$tRNA^{Arg}$ is the true substrate of the GNAT of *L. curiae*, which could also explain the rather spacious cleft found in the predicted binding site of the crystal structure. Hence, in analogy to CtdA the GNAT would protect the bacterium from potential toxic incorporation of canavanine into the proteome, because acetylated canavanyl-$tRNA^{Arg}$ can not be used by the ribosome to synthezise protein. Although further experiments have to be conducted to confirm the hypothesis, either way, by acetylating free canavanine or canavanyl-$tRNA^{Arg}$ *L. curiae* most likely protects itself from canavanine incorporation.

The overall results of this chapter suggest that guanidine is sensed by riboswitches as an indicator compound signalling the presence of guanidine-generating metabolites such as canavanine that can be erroneously loaded onto $tRNA^{Arg}$. The gut inhabitant *E. faecalis* was shown to reduce canavanine to homoserine and guanidine (161), however no molecular characterization of this activity has been reported up to date. The finding that also hydroxyguanidine triggered the guanidine-I riboswitch makes sense in light of the canavanine degradation pathway (chapter 2) that starts with the cleavage of canavanine to homoserine and hydroxyguanidine (95). The hypothesis that guanidine and hydroxyguanidine are detected by riboswitches as indicator compounds for the presence of toxic antimetabolites such as canavanine that are not easily detectable due to structural similarity to arginine adds another physiological function of guanidine sensing.

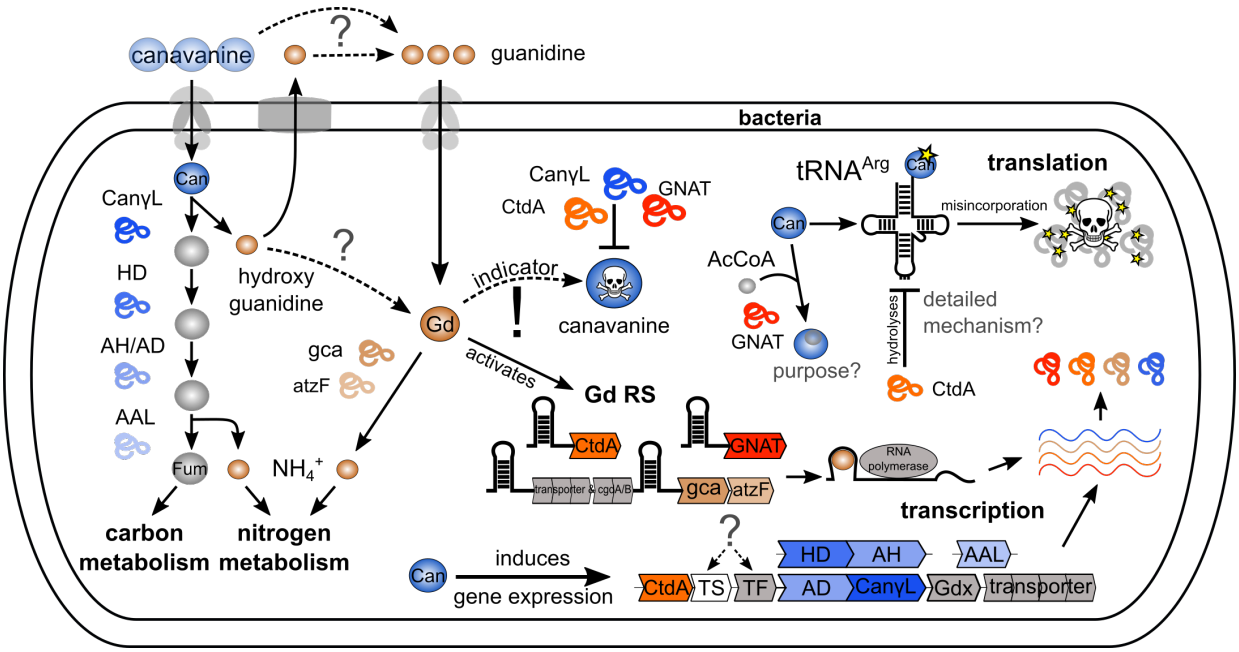
Still the question remains why guanidine instead of canavanine itself is used as the indicator molecule. One explanation could be, that the discrimination between canavanine and arginine may not be achieved by regulatory elements like riboswitches. Under those circumstances the essential amino acid arginine, which is always present in the cell, would constantly trigger a riboswitch response even when there is no canavanine present, which would lead to a waste

of resources and energy. Looking at the already known amino acid sensing riboswitches supports this hypothesis since for example the lysine riboswitch aptamer is among the longest and most complexly folded metabolite sensing aptamers (187). A more complex switch discriminating between arginine and canavanine might just be beyond the limits of ligand sensing by riboswitches.

Chapter 4: Conclusions

Within this thesis, novel insights into bacterial canavanine and guanidine metabolism are presented. A graphic summary of the results is given in Scheme 5. At the starting point of the research, we set out to identify bacteria which would be able to degrade canavanine resulting in the production of (hydroxy) guanidine. Canavanine is an antimetabolite produced by legumes to protect them against herbivores and pathogens and its antimetabolic features originate from its mimickry of arginine. That led to the discovery of a hitherto undescribed bacterium – *Pseudomonas canavaninivorans.* The bacterium belongs to the subgroup of *P. corrugata* withing the intrageneric group of *P. fluorescens*. Its inability to hydrolyse gelatin and to assimilate N-acetyl glucosamine in combination with a positive activity of acid phosphatase and naphthol-AS-BI-phosphohydrolase allow the distinction from closely related type strains. Comparative proteomics of the bacterium resulted in the discovery of the first specific canavanine degradation pathway. *P. canavaninivorans* and other rhizosphere-associated bacteria can degrade canavanine and use it as sole carbon and nitrogen source. The key activity of the pathway is composed of a canavanine-γ-lyase, which yields homoserine and hydroxyguanidine. The former can then be further channelled into the tricarboxylic acid cycle as the C_4-dicarboxylate fumarate by the enzymatic activity of a homoserine dehydrogenase, aldehyde dehydrogenase and aspartate-ammonia-lyase. Thereby, ammonium is also produced and the carbon and nitrogen need of the cell can be met. As the pathway was found to be widespread among rhizosphere-associated bacteria, it could have major implications in the interaction between legumes and bacteria and thereby contribute to the topic of global food security.

In addition to the utilization pathway, a standalone protein annotated as B3/4 editing domain-like protein has been identified. It possesses the activity of hydrolysing canavanyl-$tRNA^{Arg}$. We therefore propose to rename the B3/4 protein canavanyl-$tRNA^{Arg}$ deacylase or CtdA. The protein was not only found in rhizosphere-associated bacteria like *Pseudomonas* and *Rhizobiales*, but also in gut bacteria like *Clostridium perfrigens*. The results suggest that CtdA is the evolutionary response towards the occurrence of canavanine in the respective habitat. The physiological importance of the protein was confirmed by growth experiments with *P. canavaninivorans* and proteome data strongly suggest that impaired growth of the deletion strains is due to canavanine misincorporation into protein. Furthermore, CtdAs are one of the most common non-transporter genes associated with guanidine class I and IV riboswitches. The experimental confirmation presented in this works showing that CtdA edits canavanyl-$tRNA^{Arg}$, hints towards a connection between canavanine and guanidine metabolisms. As an additional guanidine riboswitch-associated gene (GNAT) was also shown to obtain enzymatic activity on canavanine, guanidine might act as an indicator for guanidino-group containg molecules like canavanine, which could be harmful for the cell e.g. because of misincorporation into protein.



Scheme 5: Bacterial canavanine and guanidine metabolism elucidated within this thesis. Can: canavanine, CanγL: canvanine-γ-lyase, HD: homoserine dehydrogenase, AH/AD: aldehyde dehydrogenases, AAL: aspartate-ammonia-lyase, Fum: fumarate, Gd: guanidine, gca: guanidine carboxylase, atzF: allophanate hydrolase, CtdA: canavanyl-tRNA^{Arg} deacylase (former annotated as B3/4 editing domain-like protein). GNAT: GCN5-related N-acetyltransferase, TS: protein annotated as threonine synthase, TF: protein annotated as transcription factor, Gdx: (hydroxyl-) guanidine exporter, Gd RS: guanidine riboswitch, stars represent canavanylation. Solid arrows represent reactions and pathways which are validated by the experimental work of this thesis, dotted arrows represent open questions and hypothesis.

Regarding other purposes of guanidine in nature, gene functions controlled by guanidine riboswitches were elucidated in this thesis. They enable the carboxylation and subsequent degradation of guanidine. The carboxylase pathway allows the utilization of guanidine as sole nitrogen source. Most interestingly, some *Rhizobiales* obtain a guanidine riboswitch-associated gca as well as a CanγL. In contrast to *Pseudomonas canavaninivorans* which mainly exports hydroxyguanidine produced from canavanine, those rhizobia may somehow reduce hydroxyguanidine to guanidine which can then be utilized as a nitrogen source by the described carboxylation pathway. The presented data poses again the intriguing question of the source of guanidine in nature. Given the widespread occurrence of guanidine-sensing riboswitches as well as guanidine-transporting and metabolizing activities, it seems very likely that so-far overlooked, widespread guanidine-producing abiotic or biotic reactions exist in nature.

The presented results show that canavanine is a good candidate for a potential source of guanidine, although more direct and molecular evidence is required. As already mentioned, the cleavage reaction of canavanine to guanidine and homoserine has been observed for the gut inhabitant *E. faecalis* (161), although the mechanism remains elusive. In analogy to the novel canavanine degradation pathway described above, it is most likely that the correspond-

ing enzymes might be discovered by comparative proteomics if gut bacteria are grown anaerobically on canavanine. It could also be possible that the observed reaction uses hydroxyguanidine as an intermediate, which is then further reduced to give rise to guanidine. In *E. coli* the reduction of hydroxyguanidine to guanidine has already been observed in minimal medium growth conditions (188) and also Nelson et al. suggest that guanidine is produced under those conditions (145), although in their experiments the source remained unknown. The discovery of a specific hydroxyguanidine reductase would represent the missing link between the findings presented in this thesis. Still, the ability of guanidine class I riboswitches to also sense hydroxyguanidine already represents a connection between (experimentally confirmed) canavanine degradation to hydroxyguanidine and guanidine itself.

Regarding CtdA, the elucidation of the discrimination and reaction mechanism would be of great interest. The identified editing factor's original annotation derives from its similarity to the B3/4 editing domain of the β subunit of phenylalanine-tRNA-syntethase (PheRS), but preliminary data (not shown) indicates that the crucial catalytic residues described for either the bacterial or the archaeal PheRS editing domain (189) are not present in B3/4 protein. Also, none of the other functionally characterized structural types of editing domains (HxxxH, $N2_A$, YbaK-like) apply, suggesting that the catalytic site of the CtdA could represent a mechanistically novel type of mischarged tRNA-hydrolysing proteins. In the predicted catalytic pocket of CtdA, no histidine or lysine residue is present that could be employed as a general base in a catalytic triad mechanism. As mentioned before, canavanine could take part in the catalysis in a co-factor-like manner and act as the general base because of its lowered pK_a, but until further experimental proof is found his remains highly speculative.

Taken together, this work contributes to a broader understanding of how living cells have adapted to a given habitat like the legume rhizosphere. The findings reveal insights into the molecular mechanism which may stimulate symbiotic events and promote a supportive rhizosphere, both for the bacteria as well as the plant. Moreover, the work deepens our understanding how error free translation is achieved and establishes an intrinsic link between canavanine and guanidine metabolisms. The presented findings regarding guanidine riboswitch-associated gene functions represent a paradigm shift from riboswitch-controlled detoxification of guanidine to the uptake and assimilation of this enigmatic nitrogen-rich compound and open up a variety of research questions to follow up on. And although mechanism and pathways "from elephant to butyric acid bacterium [are] all the same" their uniqueness and perfection (to me) are stunning.

Chapter 5: Material and Methods

Material

Oligonucleotides and chemicals

Chemicals were ordered either from Sigma, Roth or Acros Organics. Synthesized oligonucelotides were purchased from Sigma-Aldrich. Larger fragments were synthesized by Thermo Fisher or Twist Bioscience. A detailed overview of the oligosequences, primers and vectors used can be found in SI Table 1 and the SI Items 3 to 7. Radioactive [α-^{32}P] used for tRNA labelling was obtained from Hartmann Analytic. If not otherwise stated commercial enzymes and kits used were purchased from New England Biolabs and used following the manufacturers' instructions.

Buffers, solutions and media

If not stated otherwise buffers, solutions and media were prepared using MilliQ ultrapure water and sterilized by autocalving or using a 0.22 µm filter.

Table 7: Buffers and solutions.

Name	Composition
2x agarose gel loading buffer	60% (v/v) glycerol, 0.5% (w/v) bromophenol blue, 0.5% (w/v) xylene cyanol
2x RNA denaturing PAGE loading buffer	1 mM EDTA, 80% formamide, 0.5% (w/v) bromophenol blue, 0.5 % (w/v) xylene cyanol
6x SDS loading dye	300 mM Tris-HCl (pH 6.8), 600 mM β-mercaptoethanol, 12% (w/v) SDS, 60% (w/v) glycerol, 0.6% (w/v) bromophenol blue
5x TBE buffer (pH 8.3)	54 g/L Tris base, 27.5 g/L boric acid, 10 mM EDTA
Lysis Buffer (pH 8)	20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 mM imidazole, Pierce protease inhibitor tablet (EDTA-free, one per 10 ml)
Wash Buffer (pH 8)	20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 mM imidazole

Elution Buffer (pH 8)	20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 500 mM imidazole
Buffer exchange buffer	50 mM Tris-HCl (pH 8.0), 150 mM NaCl
5x M9 salts	42.5 g/L Na_2HPO_4 x $2H_2O$, 15 g/L KH_2PO_4, 2.5 g/L NaCl, 5 g/L NH_4Cl
5x M8 salts	42.5 g/L Na_2HPO_4 x $2H_2O$, 15 g/L KH_2PO_4, 2.5 g/L NaCl
100x Trace elements solution	10 mM EDTA, 3 mM $FeCl_3$, 620 µM $ZnCl_2$, 76 µM $CuCl_2$, 42 µM $CoCl_2$, 162 µM H_3BO_3; 8 µM $MnCl_2$
1000x Vitamin Solution	100 mg/l cyanocobalamin, 80 mg/l 4-aminobenzoic acid, 20 mg/l D-(+)-biotin, 20 mg/l niacin, 10 mg/l Ca-D-(+)-pantothenic acid, 30 mg/l pyridoxamine-chloride, 20 mg/l thiamindichlorid
Crush-Soak buffer	200 mM NaCl (1.17 g/ 0.1 L), 1mM EDTA (200 µL/ 0.1 L of 0.5 M EDTA pH 8), 10 mM HEPES (0.238 g/ 0.1 L), adjust to pH 7.5
Ringers solution	125 mM NaCl, 1.5 mM $CaCl_2$ x $2H_2O$, 5 mM KCl, 0.8 mM $Na_{2H}PO_4$, pH 7.4
MS buffer A	10 mM ammonium formate, pH 3.0
MS buffer B	90% acetonitrile, 0.2% formic acid, 10 mM ammonium formate

Table 8: Composition of media and agars

Name	Composition
LB medium	5 g/L NaCl, 5 g/L yeast extract, 10 g/L tryptone, pH 7.0
M9 minimal medium	42 mM Na_2HPO_4 x 2 H_20, 37 mM KH_2PO_4, 86 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (w/v) glucose (200 mL 5x M9 salts, 2 mL 1M $MgSO_4$, 100 µL 1M $CaCl_2$, 20 mL 20% (w/v) glucose to 1 L with H_2O)
M9 w/o C-Source	42 mM Na_2HPO_4 x 2 H_20, 37 mM KH_2PO_4, 86 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ (200 mL 5x M9 salts, 2 mL 1M $MgSO_4$, 100 µL 1M $CaCl_2$) supplemented with C-source
M8, M9 w/o N-Source	42 mM Na_2HPO_4 x 2 H_20, 37 mM KH_2PO_4, 86 mM NaCl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (w/v) glucose (200 mL 5x M8 salts, 2 mL 1M $MgSO_4$, 100 µL 1M $CaCl_2$, 20 mL 20% (w/v) glucose to 1 L with H_2O) supplemented with N-source
LB, M8 and M9 agar	Recipe as above with 15 g/l agar
SOC (super optimal broth with catabolite repression) medium	2% (w/v) tryptone, 0.5% yeast extract, 0.05% (w/v) NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ x 7 H_2O, 20 mM glucose
Simmons citrate agar	1 g/l Sodium citrate, 5 g/l NaCl, 0,2 g/l $MgSO_4$, 1 g/l $(NH_4)H_2PO_4$, 1 g/l K_2HPO_4, 0,08 g/l bromthymolblue, 15 g/l agar, pH 6.8
King A agar	20 g/l gelatine-pepton (pancreatic), 1,4 g/l $MgCl_2$, 10,0 g/l KSO_4, 15,0 g/l Agar, pH 7.2
King B agar	20 g/l pepton, 1.5 g/l $MgSO_4$, 1,5 g/l K_2HPO_4, 10 ml/l glycerin, 15 g/l agar
R2A agar	1 g/l casein (enzymatic digested), 0.5 g/l yeast extract, 0.5 g/l glucose, 0.5 g/l starch, 0.3 g/l K_2HPO_4, 0.024 g/l $MgSO_4$, 0.3 mM sodium pyruvate, 12 g/l agar

Table 9: Stock solutions of antibiotics and compounds. All solutions were sterile filtered through a 0.22 µm filter.

Name	Composition
1000x carbencilline	1 g in 10 mL 50% (v/v) ethanol
1000x kanamycin	3.4 g in 10 mL ethanol
1000x chloramphenicol	3.0 g in 10 mL H_2O
1000x IPTG	0.5 M
10% Rhamnose	10% (w/v) in H_2O

Enzymes

Table 10: Enzymes and corresponding buffers.

Name	Buffer	Manufacturer
Phusion Hot Start II DNA polymerase	5x HF buffer, 5x GC buffer	Thermo Scientific
T4 Quick ligase	2x Quick ligase buffer	NEB
Rnase inhibitor (Ribolock)	n/a	Fermentas
T7 RNA polymerase	5x Transcription buffer	Thermo Scientific
PPase	5x Transcription buffer	Thermo Scientific
DNase I	10x DNase I reaction buffer	Thermo Scientific
rSAP	10x Cut Smart buffer	NEB
T4 polynukleotide kinase	Buffer A PNK	Thermo Scientific
NdeI	10x Cut Smart buffer	NEB
XhoI	10x Cut Smart buffer	NEB
P1 Nuclease	10x Nuclease P1 reaction buffer	NEB

Kits and size ruler

Table 11: Kits and size ruler.

Name	Manufacturer
Zymo DNA Clean & Concentrator Kit	Zymo Research
Zymoclean Gel Recover Kit	Zymo Research
Zyppy Plasmid MiniPrep Kit	Zymo Research
Ni-NTA agarose beads	Qiagen
Gene Ruler 1 kb DNA ladder	Fermentas
Gene Ruler Ultra low range DNA ladder	Fermentas
Page Ruler Prestained Protein ladder	Fermentas
DNeasy Blood and Tissue Kit	Qiagen
API20NE	bioMérieux
API ZYM	bioMérieux
BCA Kit	Thermo Fisher

Laboratory Equipment

Agarose and polyacrylamide gel electrophoresis systems (BioRad)

Amersham Imager 600 (GE Healthcare)

Autoclave (Systec)

Biometra GelDoc (Biometra)

Centrifuge 5810R (Eppendorf)

Electroporator 2510 (Eppendorf)

Erlenmeyer flasks (baffled) for bacterial liquid culture 50 mL, 500 mL, 1 L, 2 L (Duran, Schott)

Heath block, digital (VWR)

Incubators (Infors, Memmert)

Laser scanner Typhoon FLA 7000 (GE Healthcare)

Multichannel pipette (Brand)

NanoQuant plate (Tecan)

Photometer (Eppendorf)

Photostimulable phosphor screens (Fuji)

Pipettes (Eppendorf)

Gel drier (BioRad)

Prominence HPLC (Shimadzu)

SDS-PAGE chamber (BioRad)

Single Quadrupole Mass spectrometer (Shimadzu)

Sonicator (Branson Sonifier 450, Heinemann)

Table top centrifuge mini spin (Eppendorf)

Tecan SPARK plate reader (Tecan)

Thermal shaker (Thermomixer comfort, Eppendorf)

Thermocycler (Biometra)

Ultracentrifuge (Beckman Coulter)

UV-VIS Photodiode Array Detector (Shimadzu)

Vortexer (VWR)

Cary 60 UV-VIS spectrometer (Agilent Technologies)

Polarization Microscope (Zeiss AxioImager)

Software

Clone Manager 9: Design and evaluation of primers, PCR and cloning strategies. Furthermore, it was used for sequence alignment.

Inkscape version 0.92: Arrangement and assignment of figures.

FinchTV was used to assess the quality of sequencing reactions.

GraphPad prism 5 (*HYPNOS*) & OriginLab Origin 8.6 were used for generation of graphs, fitting of curves and statistical analysis

Quantity One: Quantification and visualization of radiographs

TopSpin 4.0 and Mestrenova 14.2.0 were used to assess NMR spectra.

Consumables

1.5 mL and 2 mL plastic tubes (Sarstedt)

PCR tubes (Thermo Scientific)

15 mL and 50 mL platic tubes (BD Bioscience)

24, 48 and 96 deep well plates (Sarstedt)

96 well Intelli plates (Art Robbins Instruments)

10 µL, 200 µL and 1000 µL pipette tips (Sarstedt)

5 mL, 10 mL, 25 mL and 50 mL plastic pipettes (Sarstedt)

Amicon centrifugal filters (Merck)

PEI Cellulose F plates (Supelco, Merck)

Electroporation cuvettes 1 mm (Roth)

Gas permeable adhesive seals (Brand)

Aluminium adhesive seals (Brand)

Petri dishes (Sarstedt)

Sterilizing filters (Sarstedt)

Gloves Nitrile/ Latex (VWR, MaiMed)

Methods

The methods are divided into a general part, followed by chapter specific parts to facilitate reading.

General methods

Cloning

DNA fragments (inserts and backbones) were amplified by Phusion PCR (see Table 12 and Table 13). Detailed information regarding the templates and primers is given in the corresponding sections below. If the amplified plasmid backbone template was further used for Quick ligation (NEB) it was treated with rSAP (NEB) before. To remove methylated template DNA following PCR amplification the product was digested using restriction endonuclease DpnI. 1 µL of DpnI enzyme was added to the PCR reaction mixture in CutSmart buffer and incubated at 37 °C for 20 min. All fragments were purified by agarose gel electrophoresis. Agarose (concentration dependent on the expected fragment size) was dissolved in 0.5x TBE by heating the mixture in a microwave. 1 µL Midori Green (Nippon Genetics Europe GmbH) was added to 40 mL agarose solution for DNA visualization. 0.5% TBE was used as running buffer. Samples were mixed with 6x agarose loading dye. Gel electrophoresis was conducted at 120 V for app. 40 min, depending on the expected DNA size. Gels were analyzed on a gel documentation device under UV light (BioDoc Analyse, Biometra, Analytik Jena). DNA was recovered from agarose gel with the Zyppy Gel Recover Kit (Zymo Research).

Either Gibson assembly or restriction digest followed by Quick ligation was used to obtain functional plasmids, according to the manufacturer's protocols. For Gibson assembly more information can be found under www.neb.com/E2611. For restriction digest the amplified DNA fragments were incubated with the respective restriction enzymes at 37°C in Cutsmart buffer, purified by agarose gel electrophoresis and ligated in a 5:1 insert:backbone ratio using Quick ligase following the manufacturer's protocol (NEB).

Table 12: Phusion PCR reaction setup.

Stock Conc.	Reagent	Volume [µL]	Final Conc.
5x	HF Phusion buffer	10	1x
2 mM	dNTPs (Fermentas)	5	0.2 mM
5 µM	Primer A	5	0.5 µM
5 µM	Primer B	5	0.5 µM
n/a	template DNA	n/a	100-150 ng total
2 U/µL	Phusion high-fidelity DNA polymerase	0.5	0.02 U/µL
n/a	H_2O		to 50 µL total volume

Table 13: Thermo cycler program for Phusion DNA polymerase.

step	temperature [°C]	time [s]	cycles
initial denaturation	98	30	1x
denaturation	98	10	
annealing	app. 58-68°C **	30	30x
extension	72	30/kb	
final-extension	72	5 min	1x

**app. 3°C below T_M

RNA purification via polyacrylamide gel electrophoresis (PAGE)

For RNA purification a 1.5 mm thick 6% denaturing gel was casted. 48 ml of Rotiphorese sequencing gel concentrate (25%), 132 ml 9 M urea and 20 ml 10x TBE in 9 M urea were mixed. The polymerization reaction was started with 1.6 ml 10% (w/v) APS and 80 µl TEMED. Samples were mixed with denaturing RNA loading dye and the gel was run at 65 W for app. 2 h. RNA was visualized by UV shadowing or radiograph. RNA was cut out and incubated with crush-soak buffer for at least 1 h at 37°C under constant agitation. Gel was filtered over glass wool with a syringe and ethanol precipitated. The resulting pellet was washed with ice-cold 70% ethanol.

Sodium dodecyl sulfate (SDS) PAGE

To monitor protein purification samples were analysed using SDS gel electrophoresis. 1 mm thick gels were casted using the BioRad SDS gel electrophoresis equipment. First, 5 ml resolving gel (12-16% (w/v) Rotiphorese Gel 40 (37.5:1), 250 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1%(w/v) TEMED) were set up in water, casted and overlaid with isopropanol. After 30 minutes of polymerization the gel was washed with water and the stacking gel (5% (w/v) Rotiphorese Gel 40 (37.5:1), 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1%(w/v) TEMED) was casted on top of the resolving gel. SDS-gels were run approximately 1 ½ hours at 120 V in 1 x Rotiphorese SDS – PAGE buffer until the running front was near the bottom of the resolving gel.

Microbe strains

E. coli XL10 (endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB laclqZΔM15 Tn10(TetR Amy CmR)]) and *E. coli* BL21(DE3) gold (F– ompT hsdS(r– m–) dcm+ Tetr gal λ(DE3) endA The) cells were maintained in LB medium at 37°C.

P. canavaninivorans (DSM No.: 112525) was maintained in LB medium at 30°C.

Rhizobium leguminosarum (DSM No.: 30141) was grown in Medium 98 (DSMZ) at 26°C.

Methods Chapter 2.1: Discovery and taxonomic description of Pseudomonas canavaninivorans

Enrichment of canavanine-utilizing bacteria

Environmental soil samples were taken in August 2020 from the legume runner bean (*Phaseolus vulgaris*), sampled on the Isle of Reichenau (47 ° 42'08.0 "N 9 ° 02'40.1" E). The soil/root samples were dissolved in Ringers' solution, diluted 1:100 and streaked out on selective M9 salt minimal medium agar plates supplemented with 1 mM canavanine. The next day, single colonies were transferred onto a new plate containing 10 mM canavanine as carbon source and the selection was repeated until homogenous colonies were obtained. Uniformly growing, single colonies were cultivated in liquid media for further analysis and identification. The strains were maintained at -80°C in a suspension of lysogeny broth (LB) supplemented with 50% (w/v) glycerol.

Identification of bacteria degrading canavanine via hydroxyguanidine and homoserine

Isolated strains were grown in M9 salt minimal medium with 10 mM canavanine as sole carbon source at 30°C, while shaking at 200 rpm. After 24 h the cultivation medium as well as the bacterial pellet were analysed by mass spectrometry to identify the levels of canavanine, guanidine and hydroxyguanidine. Bacterial pellet samples were lysed by sonification (1.5 s on/off, 3 minutes, 20% amplitude) on ice with a Branson digital sonifier 450 and centrifuged to remove cell debris. Afterwards samples were analysed by mass spectrometry.

Mass spectrometry

Prominence HPLC system with LCMS-2020 single quadrupole MS (Shimadzu) was applied. Prior to analysis 2 µl of sample was mixed with 18 µl MS buffer A, 2 µl of this solution was injected. Guanidine, hydroxyguanidine and canavanine were separated using a Nucleodur HILIC column (250 mm length x 2 mm i.d., 3 µm particle size, Macherey-Nagel), which was equilibrated with MS buffer A and eluted with a linear gradient of 45% MS buffer B over an 8-min period followed by an isocratic step of 45% MS buffer A for 6 min. The column was operated at 20.0 ± 0.1° C with a flow rate of 0.15 ml/min. MS detection was performed by single ion monitoring (SIM) of corresponding protonated ions in positive ionization mode. LC-MS data were analysed using the LabSolutions software (Shimadzu, Release 5.93). For quantification, stable isotope-labelled guanidine (^{13}C, $^{15}N_3$) was added to the samples. Calibration curves for hydroxyguanidine and canavanine were obtained for the measuring range and used for quantification of these substances.

16S rRNA gene analysis and genome sequencing

Isolates were randomly picked and identified by 16S rRNA gene analysis after colony Phusion PCR using primers 16S_27 and 16S_1492 (190). Amplicons were sequenced by Sanger sequencing (Eurofins, Constance, German) using primer 16S_907 (190) and analysed using the EzBioCloud Database (74). For whole genome sequencing gDNA was isolated using DNeasy Blood and Tissue Kit (Qiagen), following the manufactures protocol for bacterial gDNA isolation. Sequencing was carried out by Novogene (Novogene UK, Cambridge). In short, sequencing libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's recommendations. Sequencing was done using an Illumina NovaSeq PE150 system and de novo assembly was performed using the SOAP denovo software (v. 2.04). Genome coverage was 159.0x with a G+C content of 61.04 mol% which is close to the genus average of 61.2%(191). Average nucleotide identity was calculated using the TrueBac ID server (86) and digital DNA-DNA hybridization was conducted using the Type (Strain) Genome Server (76).

Phenotypic and chemotaxonomic characterization

Strain HB002^T was routinely incubated at 30°C with 200 rpm shaking. Growth was positive on LB agar, Simmons citrate agar, R2A agar, M8 minimal salt agar supplemented with canavanine as sole carbon and nitrogen source. The test for DNA and gelatine hydrolysis was conducted by using DNA(192) and gelatine hydrolysis test agar. The ability to produce pigments was tested on King A and King B agar(193). Catalase activity was determined by bubble production using 3% (v/v) hydrogen peroxide and oxidase activity using Kovács' reagent(194). Anaerobic growth behavior was tested in liquid M9 (without addition of ammonium) minimal salt medium with 10 mM canavanine as sole carbon source and 10 mM KNO_3 as electron acceptor. All media and cultivation bottles were flushed with N_2 prior to inoculation. To assess the temperature range supporting growth, HB002^T was cultivated at 4, 18, 25, 30 and 37°C and OD_{600} was monitored. To determine the pH optimum of growth the strain was grown in phosphate buffered M9 minimal salt medium with pH values ranging from 4 to 9.6. Salt tolerance was tested by growth in the presence of different NaCl concentrations, ranging from 0 – 5 M. To assess the phenotypic traits of strain HB002^T the commercial kits API20NE and API ZYM (bioMérieux) were used, following the manufacturer's protocol and compared to closely related strains. Motility, size and shape were checked by polarization microscopy (Zeiss Axiolmagar) after growing cells in LB medium at 30°C for 24 h. The cellular fatty acid and respiratory quinone analysis were carried out by the identification service Leibniz-Institute DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkultur GmbH (Braunschweig, Germany).

Methods Chapter 2.2: Canavanine utilization in rhizosphere associated bacteria

Comparative proteomics of P. canavaninivorans grown on glycerol or canavanine as carbon source

P. canavaninivorans was grown in M9 salt minimal media with 10 mM canavanine or 13 mM glycerol as sole carbon source at 30°C. Same amounts of cells were harvested after app. 18 h and were lysed as described before. The total protein amount was determined with the BCA Kit (Thermo Fisher) according to the manufacturer's protocol. 50 µg of protein was used for each proteome analysis sample which was carried out by the Proteomics facility of the University of Konstanz. In short, protein samples were delivered in-gel, reduced by dithiothreitol and alkylated using chloroacetamide, followed by tryptic digest. Digested proteins were analyzed on a QExactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to an Easy-nLC 1200 Nanoflow-liquid chromatography system (Thermo Fisher Scientific, Bremen, Germany). Raw data was evaluated using the software proteome discoverer 1.4 (Thermo Fisher scientific) and compared to the whole proteome of *P. canavaninivorans*.

Construction of overexpression plasmids

Full length genes of *P. canavaninivorans* (Canγ**L** (MBJ2347155.1), HD (MBJ2346999.1), AC (MBJ2347154.1), AH (MBJ2347000.1) and AAL (MBJ2345991.1) were amplified from gDNA by Phusion (NewEngland Biolabs) PCR and either Gibson overhangs or NdeI and XhoI restriction sites were introduced. For the rhizobial Canγ**L** (WP_012555980.1) the corresponding codon optimized gene sequence with NdeI and XhoI restriction sites was synthesized by Twist Biosciences. Canγ**L** The K213A mutant of Canγ**L** was constructed by whole plasmid PCR on the Canγ**L** wt vector. All inserts were cloned into the overexpression vector pET28a (KmR) (SI Item 3), which bears a N-terminal 6x HisTag. The resulting plasmids were amplified in *E. coli XL10* and verified by sequencing (GATC, Eurofins).

Protein expression

For protein overexpression the respective strain containing the pET28a overexpression plasmid with the gene of interest was grown with suitable antibiotic overnight in lysogeny broth at 37°C, 200 rpm, diluted 1:100 and cultivated at 37°C, 200 rpm, to an OD_{600} of approx. 0.5. After induction with 0.5 mM isopropyl-β-D-thiogalactopyranosid the cultures were incubated at 18°C for approx. 16 hours. Cells were harvested by centrifugation and stored at -20°C. Active enzyme was purified by standard nickel metal affinity chromatography. In short, frozen pellet was resuspended in lysis buffer and left on ice for

30 minutes. Cell lysis was completed by 2 cycles of sonication (1.5 s on/off, 3 minutes, 20% amplitude) on ice, after which the lysate was centrifuged to remove cell debris and insoluble material. The supernatant fraction was filtered through a 0.2 µm filter and loaded onto high performance Ni-NTA-resin at a flow rate of 1 ml per minute. The resin was washed with wash buffer and the protein eluted with elution buffer. The protein elution was buffer exchanged into buffer without imidazole using PD10-desalting columns (GE Healthcare), concentrated by Amicon centrifugal filters (Merck), snap frozen in liquid nitrogen and stored at -80°C upon further usage. SDS-page samples were taken from all steps to monitor enzyme purification and to check purity. When needed proteins were additionally processed by TEV cleavage to remove the His-Tag, followed by reversed Ni-NTA and gel filtration chromatography on a Superdex 75 16/60 or Superdex 200 16/600 column.

Enzymatic assays

To determine the kinetic parameters of CanγL (or RCanγL) 20 µl of different canavanine concentrations were placed in a 96-well transparent flat plate. The reaction was started with 180 µl of reaction buffer (0.1 µM CanγL, 2.4 µM HD, 0.5 mM NAD+, 100 mM NaCl, 50 mM TrisHCl buffer, pH 8.0) and the absorbance at 340 nm was monitored. A calibration curve with different levels of NADH was measured simultaneously to enable the calculation of substrate turnover. To calculate KM and kcat the initial velocities were plotted over the respective canavanine concentration and fitted by Michaelis-Menten Kinetics using GraphPad Prism 6.0 and equation 1:

$$v = \frac{v_{max}[S]}{K_M + [S]} \qquad \text{equation 1}$$

where v corresponds to the reaction velocity, vmax to the maximum reaction velocity, S to the substrate concentration and K_M to the substrate concentration at half v_{max}.

The *in vitro* conversion of canavanine to hydroxyguanidine measured by MS was conducted by mixing 10 mM canavanine with 0.1 µM CanγL (or RCanγL) in 50 mM TrisHCl buffer, 100 mM NaCl, pH 8.0. The reaction took place at room temperature. At different time points samples were taken and quenched with MS buffer A. Afterwards, samples were directly injected into the MS quadrupole system and measured as described before.

Activity of K213A was tested by the NADH coupled assay described above with 25 mM canavanine in the reaction assay.

NMR measurements

Standard NMR samples consisted of 10 mM substrate in 50 mM TrisHCl, 100 mM NaCl, 200 µM PLP, pH 8.0. Enzymes (normally 0.1 µM) and cofactors (NAD^+ or $NADP^+$) were added as needed specifically for the experiment conducted. 7 µM trimethylsilylpropionate (TMSP) was added as an internal reference. ^{1}H-NMR measurements were conducted by the NMR facility of the University of Konstanz on a Bruker Avance III 600 MHz spectrometer with a TCI-H/C/N-triple resonance cryoprobe head. Spectra were measured at 300 K with 10% D_2O as field lock. Water suppression was achieved using the WATERGATE sequence. Spectra were analysed using TopSpin 4.0 and Mestrenova 14.2.0. For the hydrogen/deuterium exchange experiment a ^{1}H-decoupled natural abundance ^{13}C experiment was performed and the signal was locked on C_6D_6.

*UV/Vis spectra of the Canγ*L reaction*

To monitor the characteristic bands of the PLP cofactor during the CanγL reaction we obtained UV-Vis spectra. The reaction was measured in a quartz cuvette (50 µM enzyme in 50 mM TrisHCl, 100 mM NaCl, pH 8.0) and started by the addition of 10 mM canavanine. The experiment was performed in a Cary 60 UV-Vis spectrometer (Agilent Technologies) in continuous scan mode at a rate of 24000 nm/min, over the range of 600-300 nm in 5nm steps at room temperature.

Alpha fold modelling

The primary sequence of CanγL was submitted to the alpha fold server (105–109) and the resulting first rank model was visualized and superimposed using PyMol version 2.0 (195).

Phylogenetic analysis of canavanine-γ-lyase

Homology search and multiple sequence alignment were performed with the Consurf platform (114). Multiple sequence alignment was built using MAFFT on homologs collected from UNIREF90 database using BLAST/PSI-BLAST (3 iterations, E-value 0.0001). Subsequently, a phylogenetic tree was generated based on neighbour joining with BLOSUM62 (115). The phylogenetic tree was illustrated with iTOL (177).

Homologous recombination of P. canavaninivorans

To generate a gene knockout mutant of *P. canavaninivorans* by homologous recombination we used the method described by Huang and Wilks(196). In short, first the upstream and downstream sequences (500 nts on each site) of the respective gene to be deleted were cloned into the pEX18Gm sacB suicide plasmid (donation from Prof. Herbert Schweizer, University of Florida, USA, SI Item 5) by Gibson assembly. The resulting plasmid was transformed into *P. canavaninivorans* by electroporation. Electro competent cells were obtained by repeatedly washing overnight grown *P. canavaninivorans* cells with 1 mM $MgSO_4$. After recovery in SOC medium cells were plated out on LB with gentamycin. Single colonies that integrated the chromosome by homologous recombination were identified by colony PCR using a primer that binds on the pEX18 Gm plasmid and a primer binding on the genome. Deletion mutants were obtained via sacB mediated sucrose counter selection and the recombination event was confirmed by sequencing.

Methods Chapter 2.3: Canavanine detoxification in the context of translation

Production of labelled tRNA^{Arg}

tRNA was produced following the protocol by Avcilar-Kucukogze et al. (197). In brief, for *in vitro* transcribed tRNA two overlapping single-stranded oligonucleotides encoding the 5' end sequence of the sense strand with an upstream T7 RNA polymerase promoter sequence and the 3' end sequence of the antisense strand were annealed and filled using the Large Klenow Fragment of DNA polymerase. The resulting double-stranded DNA template was *in vitro* transcribed using T7 RNA polymerase followed by incubation with DNase I to digest the DNA template. The tRNA was purified by excision from 6% polyacrylamide gel electrophoresis (PAGE). Excised bands were incubated in crush and soak buffer (200 mM NaCl, 1 mM EDTA, 10 mM Hepes, pH 7.5), filtered through glass wool and tRNA was extracted by ethanol precipitation. For the over-production of tRNA^{Arg} in *Pseudomonas canavaninivorans* the tRNA^{Arg} sequence (GCGCCCGTAGCTCAGCTGGATAGAGCATCCGCCTTCTAAGCGGATG GTC GCA GGTTCGAGTCCTGCCGGGTGCGCCA) together with a terminator (TATTCTAGAAAGGC GCCAGATTTAACGGTCTGGC GCCTTTGCTTTAAAC) was inserted into the rhamnose-inducible expression vector pJeM1 (198) (Addgene #135088, SI Item 4). *Pseudomonas canavaninivorans* was transformed with the resulting plasmid by electroporation and transformants were selected by kanamycin resistance. A single colony was pre-cultured overnight, diluted 1:100 and grown at 30°C to an OD_{600} of 0.4. Then, rhamnose was added to induce tRNA production and the culture was grown for approximate 16 hours. Cells were collected by centrifugation and tRNA was extracted as described by Avcilar-Kucukogze et al. (197). After extraction the total tRNAs enriched in tRNA^{Arg} were further purified by 6% PAGE. tRNA was 3' α-^{32}P-ATP labelled using purified *E.coli* cca tRNA nucleotidyltransferase (overexpression clone from the ASKA collection, (199)), following the protocol by Evans et al (200). In summary, tRNA was first refolded by heating and slow cooling to ensure the right tRNA conformation and then treated with cca adding enzyme in a reaction where the equilibrium is shifted to exchange the terminal adenosine with α-^{32}P-ATP by the addition of inorganic phosphatase and CTP. For the production of 5' γ-^{32}P labelled tRNA, unlabelled purified tRNA was first dephosphorylated using shrimp alkaline phosphatase and then labelled with γ-^{32}P-ATP using T4 polynucleotide kinase following the manufacturer's instructions. In both cases labelled tRNA was purified by PAGE.

Protein overexpression

The full length open reading frames of ArgRS (MBJ2348292.1) and B3/4 protein (MBJ2347151.1) were amplified from *P. canavaninivorans* and cloned into pET28a as described above. The gene coding the B3/4 protein of *Clostridium perfrigens* together with restriction sites was synthesized by Twist Biosciences and cloned into pET28a using Quick ligation. Protein overexpression was conducted as described before.

Aminoacylation assay

Immediately prior to the aminoacylation reaction unlabelled tRNA was spiked with freshly prepared ^{32}P labelled tRNA and then refolded by heating and stepwise cooling. For reactions with different substrates in the presence or absence of B3/4 protein, a master mix was prepared containing ~3 µM tRNA, 3 mM ATP, 0.0026 U/µl PPase, 0.13 µg/µl bovine serum albumin, 0.1 mM DTT, reaction buffer (50 mM HEPES, 25 mM KCl, 15 mM $MgCl_2$, pH 7.5) and 6 µM ArgRS enzyme. 1 µM B3/4 protein was added to the master mix as indicated. Concentrations represent the final assay concentrations. The aminoacylation reactions were started by mixing 1 µl of 2 mM substrate with 5 µl of master mix followed by incubation at 37 °C or room temperature for 15 minutes. Afterwards the reaction was quenched by the addition of 12 µl quenching buffer (1.2 M NaOAc, 0.1% [w/v] SDS, pH 4.0) and (aminoacylated) tRNA was digested using P1 nuclease for 1 hour at room temperature. Then, 1 µl of digested sample was seperated by thin layer chromatography, using polyethylenimine cellulose F plates (Supelco, Merck) and 100 mM ammonium acetate in 5% (v/v) acetic acid as running buffer. After separation the plates were dried and radiographs were recorded with a phosphorimager (GE Healthcare Life Science). Signal intensities were evaluated and quantified using ImageJ (201).

Bacterial growth assays

Cells were grown in LB medium overnight and diluted to an OD_{600} of 0.005 into minimal M9 salt medium as carbon source, either 10 mM canavanine or 0.4% (w/v) glucose with 0, 1 or 9 mM canavanine were added from filter-sterilized stocks.

Proteome analysis

Cells were grown overnight at 30 °C in M9 salt minimal media with 0.4% (w/v) glucose as carbon source. Additionally, 3.5 mM canavanine were added to the respective samples. Cultures were harvested by centrifugation and further processed as described before. Proteomics analysis was again performed by the Proteomics facility of the University of Konstanz.

Methods Chapter 3.1: A guanidine riboswitch-associated carboxylase pathway enables the utilization of guanidine as sole nitrogen source in bacteria

Enrichment of guanidine-utilizing bacteria

Environmental sample was taken in late September 2019 from the lake shore sediment of the Lake of Constance (47°41'44.2"N 9°11'35.1"E). Sediment was rinsed with lake water and filtered. The sample was sequentially diluted to obtain single colonies for selection. Dilutions were streaked out on selective minimal media agar plates supplemented with 5 mM guanidine as the sole nitrogen source and 1% v/v glycerol as carbon source. Plates were incubated at RT for 96 h. Single colonies were picked and transferred onto a new plate. Selection was repeated until homogenous colonies were obtained.

Growth analysis of guanidine-utilizing bacteria

Isolated strains were grown in minimal medium with 5 mM guanidine, 10 mM urea or 15 mM NH_4Cl as the sole carbon source, respectively. Growth was monitored in a 96 well-plate in biological triplicates. The medium was inoculated to OD_{600} = 0.0005. The growth medium was covered with M20 silicon oil to allow gas exchange but avoid evaporation. Plates were shaken at 200 rpm at 30°C in a TECAN reader and OD_{600} was measured every 10 min until stationary phase was reached.

Cloning and protein overexpression

The full length gene of urea (*uca1*, WP_171149239.1) and guanidine (*uca2*, WP_171148480.1) carboxylase were amplified from gDNA of *E. rhapontici.* Cloning, protein overexpression and protein purification was performed as described before.

Enzymatic assay of carboxylases

Carboxylation activity (ATP cleavage) was measured by monitoring the coupled activity of pyruvate kinase and lactate dehydrogenase, as described before (Kanamori et al., 2004). The enzyme assay was performed in 20 mM TrisHCl (pH8.0), 200 mM NaCl, 2mM DTT, containing Uca or Gca (final conc. 1 µM), 8 mM MgCl2, 8 mM NaHCO3, different concentrations of guanidine or urea, 1 mM phosphoenolpyruvate, 0.8 mM NADH and 5 U/ml pyruvate kinase/lactate dehydrogenase enzyme mix (Sigma). The reaction was started by the addition of 1 mM ATP at 25°C and the decrease in NADH absorbance (340 nm) monitored in 96-well plates using a TECAN spark plate reader at a final assay volume of 50 µl. Technical triplicates were run for each substrate concentration and the slope of absorbance decrease over time was used to evaluate the rate of substrate carboxylation. Velocity was defined such as that 1U of activity equalizes 1 µmol NAD^+ formation per minute. Kinetic parameters were determined by fitting a

plot of velocity vs. substrate concentration using GraphPad Prism 6 assuming Michaelis-Menten kinetics.

Phylogenetic Analysis of urea and guanidine carboxylases

Homology search and multiple sequence alignment were performed with Consurf(114) platform based on *K. lactis* crystal structure (PDB: 3VA7). Multiple sequence alignment was built using MAFFT on homologues collected from UNIREF90 database using BLAST/PSI-BLAST (3 iterations, E-value 0.0001). Jalview(176) was used for the alignment of the sequences of the guanidine carboxylases and urea carboxylase from our strains. Subsequently, a phylogenetic tree was generated based on neighbour joining with BLOSUM62(115). The phylogenetic tree was illustrated with iTOL(177).

Analysis of guanidine operons in genomes and metagenomes

All complete bacterial genomes in version 87 of the RefSeq nucleotide database(202) were analysed. Complete genomes were defined as those whose accession begins with “NC_”. The guanidine-I, -II, -III riboswitches were searched using the standard procedure in Rfam [cite https://pubmed.ncbi.nlm.nih.gov/33211869]. For guanidine-I, -II and -III, we used Rfam entries RF00442, RF1068 and RF01763, respectively. Guanidine-IV riboswitch locations were taken from our recent publication (148). Guanidine-I riboswitches are highly similar to riboswitches with other ligand specificities (157). To extract only guanidine-I riboswitches, we looked at the two nucleotides immediately following the conserved CAC sequence (145). We accepted only sequences in which these nucleotides were GG. To reduce false positive riboswitch predictions, we also eliminated guanidine-II sequences unless they had the tetramer ACGR in both hairpins. We also enumerated guanidine-III nucleotides that were at least 97% conserved and not predicted to form a Watson-Crick base pair, and eliminated sequences that deviated from these conserved nucleotides. If the distance between a riboswitch and the first downstream gene was at most 700 nucleotides away, and the gene is encoded in the same direction, we assumed that the gene was regulated by the riboswitch. Subsequent genes were presumed to be co-transcribed if they were also encoding in this strand, and were located no more than 500 nucleotides from the previous gene. These maximum distances are conservatively high to ensure all relevant genes would be found. Genes were functionally classified based on conserved protein domains in version 32.0 of the Pfam database (203). In addition to assigning Gdx-type exporters and Gca-type carboxylase pathways, a third activity (termed Agmat for Agmatinase-like proteins) was included in the analysis since it is also often controlled by guanidine riboswitches(see SI of the original publication and and SI Table 2). However, since it´s function is unclear and it is not connected to the Gdx and Gca activities, the occurrence and function of the Agmat pathway was not further pursued. We manually defined a mapping between

guanidine-associated gene functions and Pfam entries. We also defined a mapping between gene functions and pathways. Riboswitches were deemed to control a pathway when they appeared to regulate at least one gene function that is unambiguously associated with that pathway, i.e., are not components of ABC transporters. Metagenomes were downloaded from various sources, predominantly IMG/M (204) and GenBank (202), and we classified them into environmental categories based on available metadata. Due to inconsistencies in metadata, environmental categories were largely created manually. The locations of genes were predicted by MetaProdigal (205).

Methods Chapter 3.2: Beyond nutrition – an additional role of guanidine in nature

Cloning and protein overexpression

The gene coding the B3/4 protein of *Clostridium perfrigens* together with restriction sites was synthesized by Twist Biosciences and cloned into pET28a using Quick ligation. The sequence of the DNA fragment is shown in SI Item 7.

Protein overexpression and purification was conducted as described before.

Aminoacylation assay

The aminoacylation assay was performed as described before.

In vivo riboswitch activity assay

The guanidine class I riboswitch sequence of *Pseudomonas pelagia* CL-AP6 (DSM No. 25163, AROI01000023.1, nucleotides 57628 to 57864) was cloned under the control of the constitutive ON promoter lysC into the pQE vector in front of a lacZ gene. Primers and the vector map can be found in the SI Table 1 and SI Item 6. Correct assembly of the construct was verified by sequencing and the resulting plasmid was used to transform lacZ-deficient *E. coli* strain ER2566 (NEB). Cells containing the pQE_GuaIRS_lacZ plasmid were grown overnight and then diluted to an OD_{600} of 0.05. Cells were further cultivated at 37 °C until they reached an OD_{600} of approximately 0.3 upon which 200 µl of culture were transferred to a 96-well plate. Candidate riboswitch ligands were added to 2.5 mM and the cultures were overlaid with 30 µl silicon oil to avoid evaporation while allowing gas exchange. Galactosidase activity assays with o-nitrophenyl-β-galactoside (ONPG) were conducted following the protocol by Zhang and Bremer (1995) (206) with slight modifications. The plate was further incubated at 37 °C to allow expression of lacZ for 1 h after which OD_{600} was measured again and 20 µl of each sample were transferred into 80 µl permeabilization solution (100 mM Na_2HPO_4, 20 mM KCl, 2 mM $MgSO_4$, 0.8 mg/ml CTAB, 0.4 mg/ml sodium deoxycholate, 5.4 µl/ml β-mercaptoethanol). Then, 15 µl of permeabilized sample was mixed with 90 µl of ONPG substrate solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/ml ONPG, 2.7 µl/ml β-mercaptoethanol) to detect lacZ activity. ONPG solution was warmed to 30 °C prior to addition. The reaction was stopped after 30 min by the addition of 105 µl 1 M Na_2CO_3. To evaluate colour development, the absorbance at 420 nm was measured and Miller units were calculated according to equation 2:

$$Miller\ units = 1000 * \frac{Ab_{420}}{((OD_{600}\ of\ cultures\ sample)*(volume\ [0.02\ ml])*(reaction\ time))} \quad \text{equation 2}$$

Acetylation assay

For the acetylation assay Ellman's reagent (5,5′-dithiobis-(2-nitrobenzoic acid) or DTNB) was used. 5 µl of the respective substrate were mixed with 20 µl acetylation reaction mix in a 96 well plate. After incubation for 15 minutes at room temperature 75 µl acetylation stop buffer were added, followed by the addition of 100 µl of acetylation DTNB reagent. After 5 min of incubation at room temperature the absorbance at 420 nm was measured. For quantification different concentrations of CoA were used to obtain a calibration curve. To calculate K_M and k_{cat} the initial velocities were plotted over the respective substrate concentration and fitted by Michaelis-Menten kinetics as illustrated before.

Crystallization

Protein purification of *L. curiae* GNAT was performed as described above. The final polishing step was achieved using a size exclusion chromatography on a Superdex S75 16/60 column (GE Healthcare) equilibrated in buffer exchange buffer. Purity was assessed to be >95% using SDS-PAGE stained with Coomassie blue.

Crystals were formed at 18°C in 96-well Intelli-plates (Art Robbins Instruments) set with a sitting drops of 200:200 ul and 100:200 ul protein to reservoir (1 M Potassium sodium tartrate, 0.1 M MES/ Sodium hydroxide pH 6.0). For X-ray irradiation, crystals were cryoprotected with paratone prior to flash vitrification in liquid nitrogen. X-ray diffraction data were collected on beamline PXI at the Swiss Light Source synchrotron (Villigen, CH) and indexed using XDS (207). Phasing of diffraction data was achieved by molecular replacement conducted in Phaser (208) using a putative acetyltransferase from *Streptococcus mutans* (Identity 38.6%, Similarity 59.4%) pruned to common atoms with SCULPTOR (209) as a search model (PDB entry 4E2A). Manual model building was in COOT (210) and model refinement was carried out in Phenix.refine (211) using isotropic B-factors and TLS parameters. The final model was assed using Molprobity (212).

References

1. Kornberg, A. (2000) Ten commandments: lessons from the enzymology of DNA replication, *Journal of bacteriology,* **182,** 3613–3618.
2. Lehman, I.R., Bessman, M.J., Simms, E.S. and Kornberg, A. (1958) Enzymatic Synthesis of Deoxyribonucleic Acid, *The Journal of biological chemistry,* **233,** 163–170.
3. Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature,* **226,** 1209–1211, https://www.nature.com/articles/2261209a0.
4. Temin, H.M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus, *Nature,* **226,** 1211–1213.
5. Crick, F. (1970) Central dogma of molecular biology, *Nature,* **227,** 561–563.
6. McGary, K. and Nudler, E. (2013) RNA polymerase and the ribosome: the close relationship, *Current opinion in microbiology,* **16,** 112–117. First published on Feb 22, 2013.
7. Rauhut, R. and Klug, G. (1999) mRNA degradation in bacteria, *FEMS microbiology reviews,* **23,** 353–370.
8. Bervoets, I. and Charlier, D. (2019) Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology, *FEMS microbiology reviews,* **43,** 304–339.
9. Artsimovitch, I. and Landick, R. (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals, *Proceedings of the National Academy of Sciences of the United States of America,* **97,** 7090–7095.
10. Banerjee, S., Chalissery, J., Bandey, I. and Sen, R. (2006) Rho-dependent transcription termination: more questions than answers, *Journal of microbiology (Seoul, Korea),* **44,** 11–22.
11. Rubio Gomez, M.A. and Ibba, M. (2020) Aminoacyl-tRNA synthetases, *RNA (New York, N.Y.),* **26,** 910–936. First published on Apr 17, 2020.
12. Perona, J.J. and Gruic-Sovulj, I. (2014) Synthetic and editing mechanisms of aminoacyl-tRNA synthetases, *Topics in current chemistry,* **344,** 1–41.
13. Ganesh, R.B. and Maerkl, S.J. (2022) Biochemistry of Aminoacyl tRNA Synthetase and tRNAs and Their Engineering for Cell-Free and Synthetic Cell Applications, *Frontiers in bioengineering and biotechnology,* **10,** 918659. First published on Jul 1, 2022.
14. Fichtner, M., Voigt, K. and Schuster, S. (2017) The tip and hidden part of the iceberg: Proteinogenic and non-proteinogenic aliphatic amino acids, *Biochimica et biophysica acta. General subjects,* **1861,** 3258–3269. First published on Aug 20, 2016.
15. Giegé, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic acids research,* **26,** 5017–5035.
16. Fersht, A.R. (1975) Demonstration of two active sites on a monomeric aminoacyl-tRNA synthetase. Possible roles of negative cooperativity and half-of-the-sites reactivity in oligomeric enzymes, *Biochemistry,* **14,** 5–12.
17. Fersht, A.R. and Dingwall, C. (1979) Evidence for the double-sieve editing mechanism in protein synthesis. Steric exclusion of isoleucine by valyl-tRNA synthetases, *Biochemistry,* **18,** 2627–2631.
18. Zivkovic, I., Ivkovic, K., Cvetesic, N. and Gruic-Sovulj, I. (2021) Negative catalysis by the editing domain of class I aminoacyl-tRNA synthetases.
19. Martinis, S.A. and Boniecki, M.T. (2010) The balance between pre- and post-transfer editing in tRNA synthetases, *FEBS letters,* **584,** 455–459.
20. Roy, H., Ling, J., Irnov, M. and Ibba, M. (2004) Post-transfer editing in vitro and in vivo by the beta subunit of phenylalanyl-tRNA synthetase, *The EMBO journal,* **23,** 4639–4648. First published on Nov 4, 2004.
21. Cvetesic, N. and Gruic-Sovulj, I. (2017) Synthetic and editing reactions of aminoacyl-tRNA synthetases using cognate and non-cognate amino acid substrates, *Methods (San Diego, Calif.),* **113,** 13–26. First published on Oct 3, 2016.
22. Bacher, J.M. and Schimmel, P. (2007) An editing-defective aminoacyl-tRNA synthetase is mutagenic in aging bacteria via the SOS response, *Proceedings of the National Academy of Sciences of the United States of America,* **104,** 1907–1912. First published on Jan 30, 2007.
23. Zhang, H., Wu, J., Lyu, Z. and Ling, J. (2021) Impact of alanyl-tRNA synthetase editing deficiency in yeast, *Nucleic acids research,* **49,** 9953–9964.
24. Nangle, L.A., Motta, C.M. and Schimmel, P. (2006) Global effects of mistranslation from an editing defect in mammalian cells, *Chemistry & biology,* **13,** 1091–1100.
25. Lee, J.W., Beebe, K., Nangle, L.A., Jang, J., Longo-Guess, C.M., Cook, S.A., Davisson, M.T., Sundberg, J.P., Schimmel, P. and Ackerman, S.L. (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration, *Nature,* **443,** 50–55. First published on Aug 13, 2006.
26. Beebe, K., Merriman, E., Ribas De Pouplana, L. and Schimmel, P. (2004) A domain for editing by an archaebacterial tRNA synthetase, *Proceedings of the National Academy of Sciences of the United States of America,* **101,** 5958–5963. First published on Apr 12, 2004.

27. Korencic, D., Ahel, I., Schelert, J., Sacher, M., Ruan, B., Stathopoulos, C., Blum, P., Ibba, M. and Söll, D. (2004) A freestanding proofreading domain is required for protein synthesis quality control in Archaea, *Proceedings of the National Academy of Sciences of the United States of America,* **101,** 10260–10265. First published on Jul 6, 2004.

28. Chong, Y.E., Yang, X.-L. and Schimmel, P. (2008) Natural homolog of tRNA synthetase editing domain rescues conditional lethality caused by mistranslation, *The Journal of biological chemistry,* **283,** 30073–30078. First published on Aug 22, 2008.

29. Fukunaga, R. and Yokoyama, S. (2007) Structure of the AlaX-M trans-editing enzyme from Pyrococcus horikoshii, *Acta crystallographica. Section D, Biological crystallography,* **63,** 390–400. First published on Feb 21, 2007.

30. Pawar, K.I., Suma, K., Seenivasan, A., Kuncha, S.K., Routh, S.B., Kruparani, S.P. and Sankaranarayanan, R. (2017) Role of D-aminoacyl-tRNA deacylase beyond chiral proofreading as a cellular defense against glycine mischarging by AlaRS, *eLife,* **6.** First published on Mar 31, 2017.

31. Routh, S.B., Pawar, K.I., Ahmad, S., Singh, S., Suma, K., Kumar, M., Kuncha, S.K., Yadav, K., Kruparani, S.P. and Sankaranarayanan, R. (2016) Elongation Factor Tu Prevents Misediting of Gly-tRNA(Gly) Caused by the Design Behind the Chiral Proofreading Site of D-Aminoacyl-tRNA Deacylase, *PLoS biology,* **14,** e1002465. First published on May 25, 2016.

32. Kuzmishin Nagy, A.B., Bakhtina, M. and Musier-Forsyth, K. (2020) Trans-editing by aminoacyl-tRNA synthetase-like editing domains, *The Enzymes,* **48,** 69–115. First published on Sep 8, 2020.

33. An, S. and Musier-Forsyth, K. (2004) Trans-editing of Cys-tRNAPro by Haemophilus influenzae YbaK protein, *The Journal of biological chemistry,* **279,** 42359–42362. First published on Aug 20, 2004.

34. Chen, L., Tanimoto, A., So, B.R., Bakhtina, M., Magliery, T.J., Wysocki, V.H. and Musier-Forsyth, K. (2019) Stoichiometry of triple-sieve tRNA editing complex ensures fidelity of aminoacyl-tRNA formation, *Nucleic acids research,* **47,** 929–940.

35. Liu, Z., Vargas-Rodriguez, O., Goto, Y., Novoa, E.M., Ribas de Pouplana, L., Suga, H. and Musier-Forsyth, K. (2015) Homologous trans-editing factors with broad tRNA specificity prevent mistranslation caused by serine/threonine misactivation, *Proceedings of the National Academy of Sciences of the United States of America,* **112,** 6027–6032. First published on Apr 27, 2015.

36. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins, *Journal of Molecular Biology,* **3,** 318–356.

37. Valentini, M. and Lapouge, K. (2013) Catabolite repression in Pseudomonas aeruginosa PAO1 regulates the uptake of C4 - dicarboxylates depending on succinate concentration, *Environmental microbiology,* **15,** 1707–1716. First published on Dec 18, 2012.

38. Yoshida, T., Qin, L., Egger, L.A. and Inouye, M. (2006) Transcription regulation of ompF and ompC by a single transcription factor, OmpR, *The Journal of biological chemistry,* **281,** 17114–17123. First published on Apr 17, 2006.

39. Pedre, B., Young, D., Charlier, D., Mourenza, Á., Rosado, L.A., Marcos-Pascual, L., Wahni, K., Martens, E., La G de Rubia, A. and Belousov, V.V. *et al.* (2018) Structural snapshots of OxyR reveal the peroxidatic mechanism of H2O2 sensing, *Proceedings of the National Academy of Sciences of the United States of America,* **115,** E11623-E11632. First published on Nov 21, 2018.

40. Low, D.A., Weyand, N.J. and Mahan, M.J. (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence, *Infection and immunity,* **69,** 7197–7204.

41. Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M.N., Hosaka, T., Ochi, K., Yokoyama, S. and Vassylyev, D.G. (2004) Structural Basis for Transcription Regulation by Alarmone ppGpp, *Cell,* **117,** 299–310.

42. Wang, T., Sun, W., Fan, L., Hua, C., Wu, N., Fan, S., Zhang, J., Deng, X. and Yan, J. (2021) An atlas of the binding specificities of transcription factors in Pseudomonas aeruginosa directs prediction of novel regulators in virulence, *eLife,* **10.** First published on Mar 29, 2021.

43. Lecoutere, E., Verleyen, P., Haenen, S., Vandersteegen, K., Noben, J.-P., Robben, J., Schoofs, L., Ceyssens, P.-J., Volckaert, G. and Lavigne, R. (2012) A theoretical and experimental proteome map of Pseudomonas aeruginosa PAO1, *MicrobiologyOpen,* **1,** 169–181.

44. Perez-Rueda, E., Hernandez-Guerrero, R., Martinez-Nuñez, M.A., Armenta-Medina, D., Sanchez, I. and Ibarra, J.A. (2018) Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors, *PloS one,* **13,** e0195332. First published on Apr 3, 2018.

45. Carpousis, A.J. (2007) The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E, *Annual review of microbiology,* **61,** 71–87.

46. Luciano, D.J., Vasilyev, N., Richards, J., Serganov, A. and Belasco, J.G. (2017) A Novel RNA Phosphorylation State Enables 5' End-Dependent Degradation in Escherichia coli, *Molecular cell,* **67,** 44-54.e6. First published on Jun 29, 2017.

47. Duval, M., Simonetti, A., Caldelari, I. and Marzi, S. (2015) Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics, *Biochimie,* **114,** 18–29. First published on Mar 17, 2015.

48. Serganov, A. and Nudler, E. (2013) A decade of riboswitches, *Cell,* **152,** 17–24.

49. Nudler, E. and Mironov, A.S. (2004) The riboswitch control of bacterial metabolism, *Trends in biochemical sciences,* **29,** 11–17.

50. Dambach, M.D. and Winkler, W.C. (2009) Expanding roles for metabolite-sensing regulatory RNAs, *Current opinion in microbiology,* **12,** 161–169. First published on Feb 26, 2009.

51. Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T. and Vaitkevicius, K. *et al.* (2009) The Listeria transcriptional landscape from saprophytism to virulence, *Nature,* **459,** 950–956. First published on May 17, 2009.

52. Breaker, R.R. (2022) The Biochemical Landscape of Riboswitch Ligands, *Biochemistry,* **61,** 137–149. First published on Jan 24, 2022.

53. Chubukov, V., Gerosa, L., Kochanowski, K. and Sauer, U. (2014) Coordination of microbial metabolism, *Nature reviews. Microbiology,* **12,** 327–340. First published on Mar 24, 2014.

54. Bren, A., Park, J.O., Towbin, B.D., Dekel, E., Rabinowitz, J.D. and Alon, U. (2016) Glucose becomes one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal levels of cAMP, *Scientific reports,* **6,** 24834. First published on Apr 25, 2016.

55. Rojo, F. (2010) Carbon catabolite repression in Pseudomonas : optimizing metabolic versatility and interactions with the environment, *FEMS microbiology reviews,* **34,** 658–684. First published on Mar 10, 2010.

56. Kitagawa, M. and Tomiyama, T. (1929) A new amino-compound in the jack bean and a corresponding new ferment.(I)*, *The Journal of Biochemistry,* **11,** 265–271.

57. Kitagawa, M. and Yamada, H. (1932) Studies on a di-amino acid, canavanin (II), *The Journal of Biochemistry,* **16,** 339–350.

58. Fearon, W.R. and Bell, E.A. (1955) Canavanine: detection and occurrence in Colutea arborescens, *The Biochemical journal,* **59,** 221–224.

59. Rosenthal, G.A. (1977) The biological effects and mode of action of L-canavanine, a structural analogue of L-arginine, *The Quarterly review of biology,* **52,** 155–178.

60. Rosenthal, G.A. (1977) Nitrogen allocation for L-Canavanine synthesis and its relationship to chemical defense of the seed, *Biochemical Systematics and Ecology,* **5,** 219–220.

61. Rosenthal, G.A. (1990) Metabolism of l-Canavanine and l-Canaline in Leguminous Plants, *Plant physiology,* **94,** 1–3.

62. Cai, T., Cai, W., Zhang, J., Zheng, H., Tsou, A.M., Xiao, L., Zhong, Z. and Zhu, J. (2009) Host legume-exuded antimetabolites optimize the symbiotic rhizosphere, *Molecular microbiology,* **73,** 507–517. First published on Jul 7, 2009.

63. Mardani-Korrani, H., Nakayasu, M., Yamazaki, S., Aoki, Y., Kaida, R., Motobayashi, T., Kobayashi, M., Ohkama-Ohtsu, N., Oikawa, Y. and Sugiyama, A. *et al.* (2021) L-Canavanine, a Root Exudate From Hairy Vetch (Vicia villosa) Drastically Affecting the Soil Microbial Community and Metabolite Pathways, *Frontiers in microbiology,* **12,** 701796. First published on Sep 27, 2021.

64. Melangeli, C., Rosenthal, G.A. and Dalman, D.L. (1997) The biochemical basis for L-canavanine tolerance by the tobacco budworm Heliothis virescens (Noctuidae), *Proceedings of the National Academy of Sciences of the United States of America,* **94,** 2255–2260.

65. Rosenthal, G.A. and Dahlman, D.L. (1986) L-Canavanine and protein synthesis in the tobacco hornworm Manduca sexta, *Proceedings of the National Academy of Sciences of the United States of America,* **83,** 14–18.

66. Cooper, A.J. (1984) Oxime formation between α-keto acids and l-canaline, *Archives of Biochemistry and Biophysics,* **233,** 603–610.

67. Rosenthal, G.A. and Dahlman, D.L. (1990) Interaction of L-canaline with ornithine aminotransferase of the tobacco hornworm, Manduca sexta (Sphingidae), *The Journal of biological chemistry,* **265,** 868–873.

68. Rahiala, E.-L., Kekomäki, M., Jänne, J., Raina, A. and Räihä, N. (1971) Inhibition of pyridoxal enzymes by l-canaline, *Biochimica et Biophysica Acta (BBA) - Enzymology,* **227,** 337–343.

69. Percudani, R. and Peracchi, A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes, *EMBO reports,* **4,** 850–854.

70. Mitri, C., Soustelle, L., Framery, B., Bockaert, J., Parmentier, M.-L. and Grau, Y. (2009) Plant insecticide L-canavanine repels Drosophila via the insect orphan GPCR DmX, *PLoS biology,* **7,** e1000147. First published on Jun 30, 2009.

71. Berge, M.A., Rosenthal, G.A. and Dahlman, D.L. (1986) Tobacco budworm, Heliothis virescens [Noctuidae] resistance to l-canavanine, a protective allelochemical, *Pesticide Biochemistry and Physiology,* **25,** 319–326.

72. Kihara, H., Prescott, J.M. and Snell, E.E. (1955) The bacterial cleavage of canavanine to homoserine and guanidine, *The Journal of biological chemistry,* **217,** 497–503.

73. Kalyankar, G.D., Ikawa, M. and Snell, E.E. (1958) The enzymatic cleavage of canavanine to homoserine and hydroxyguanidine, *The Journal of biological chemistry,* **233,** 1175–1178.

74. Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H. and Chun, J. (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies, *International journal of systematic and evolutionary microbiology,* **67,** 1613–1617. First published on May 30, 2017.

75. Hauth, F., Buck, H. and Hartig, J.S. (2022) Pseudomonas canavaninivorans sp. nov., isolated from bean rhizosphere, *International journal of systematic and evolutionary microbiology,* **72**.

76. Meier-Kolthoff, J.P. and Göker, M. (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, *Nature communications,* **10,** 2182. First published on May 16, 2019.

77. Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P. and Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions, *BMC bioinformatics,* **14,** 60. First published on Feb 21, 2013.

78. Meier-Kolthoff, J.P., Hahnke, R.L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B. and Goodwin, L.A. *et al.* (2014) Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy, *Standards in genomic sciences,* **9,** 2. First published on Dec 8, 2014.

79. Lefort, V., Desper, R. and Gascuel, O. (2015) FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program, *Molecular biology and evolution,* **32,** 2798–2800. First published on Jun 30, 2015.

80. Kreft, L., Botzki, A., Coppens, F., Vandepoele, K. and van Bel, M. (2017) PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization, *Bioinformatics (Oxford, England),* **33,** 2946–2947.

81. Farris, J.S. (1972) Estimating Phylogenetic Trees from Distance Matrices, *The American Naturalist,* **106,** 645–668.

82. Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S. and Phillippy, A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash, *Genome biology,* **17,** 132. First published on Jun 20, 2016.

83. Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T. and Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes, *Nucleic acids research,* **35,** 3100–3108. First published on Apr 22, 2007.

84. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications, *BMC bioinformatics,* **10,** 421. First published on Dec 15, 2009.

85. Stackebrandt, E. and Goebel, B.M. (1994) Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology, *International journal of systematic and evolutionary microbiology,* **44,** 846–849.

86. Ha, S.M., Kim, C.K., Roh, J., Byun, J.H., Yang, S.J., Choi, S.B., Chun, J. and Yong, D. (2019) Application of the Whole Genome-Based Bacterial Identification System, TrueBac ID, Using Clinical Isolates That Were Not Identified With Three Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) Systems, *Annals of laboratory medicine,* **39,** 530–536.

87. Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.-H. and McDermott, P.F. *et al.* (2019) Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates, *Antimicrobial agents and chemotherapy,* **63.** First published on Oct 22, 2019.

88. Liu, B., Zheng, D., Jin, Q., Chen, L. and Yang, J. (2019) VFDB 2019: a comparative pathogenomic platform with an interactive web interface, *Nucleic acids research,* **47,** D687-D692.

89. Richter, M. and Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition, *Proceedings of the National Academy of Sciences of the United States of America,* **106,** 19126–19131. First published on Oct 23, 2009.

90. Garrido-Sanz, D., Arrebola, E., Martínez-Granero, F., García-Méndez, S., Muriel, C., Blanco-Romero, E., Martín, M., Rivilla, R. and Redondo-Nieto, M. (2017) Classification of Isolates from the *Pseudomonas fluorescens* Complex into Phylogenomic Groups Based in Group-Specific Markers, *Frontiers in microbiology,* **8,** 413. First published on Mar 15, 2017.

91. Liang, J., Wang, S., Yiming, A., Fu, L., Ahmad, I., Chen, G. and Zhu, B. (2021) *Pseudomonas bijieensis sp. nov*., isolated from cornfield soil, *International journal of systematic and evolutionary microbiology,* **71.** First published on Feb 2, 2021.

92. Zhao, H., Ma, Y., Wu, X. and Zhang, L. (2020) *Pseudomonas viciae sp. nov*., isolated from rhizosphere of broad bean, *International journal of systematic and evolutionary microbiology,* **70,** 5012–5018.

93. Sikorski, J., Stackebrandt, E. and Wackernagel, W. (2001) *Pseudomonas kilonensis sp. nov*., a bacterium isolated from agricultural soil, *International journal of systematic and evolutionary microbiology,* **51,** 1549–1555.

94. Achouak, W., Sutra, L., Heulin, T., Meyer, J.M., Fromin, N., Degraeve, S., Christen, R. and Gardan, L. (2000) *Pseudomonas brassicacearum sp. nov.* and *Pseudomonas thivervalensis sp. nov*., two root-associated bacteria isolated from *Brassica napus* and *Arabidopsis thaliana*, *International journal of systematic and evolutionary microbiology,* **50 Pt 1,** 9–18.

95. Hauth, F., Buck, H., Stanoppi, M. and Hartig, J.S. (2022) Canavanine utilization via homoserine and hydroxyguanidine by a PLP-dependent γ-lyase in Pseudomonadaceae and Rhizobiales, *RSC Chem. Biol.*

96. Kermani, A.A., Macdonald, C.B., Gundepudi, R. and Stockbridge, R.B. (2018) Guanidinium export is the primal function of SMR family transporters, *Proceedings of the National Academy of Sciences of the United States of America,* **115,** 3060–3065. First published on Mar 5, 2018.

97. Liang, J., Han, Q., Tan, Y., Ding, H. and Li, J. (2019) Current Advances on Structure-Function Relationships of Pyridoxal 5'-Phosphate-Dependent Enzymes, *Frontiers in molecular biosciences,* **6,** 4. First published on Mar 5, 2019.

98. Revtovich, S., Anufrieva, N., Morozova, E., Kulikova, V., Nikulin, A. and Demidkina, T. (2016) Structure of methionine γ-lyase from Clostridium sporogenes, *Acta crystallographica. Section F, Structural biology communications,* **72,** 65–71. First published on Jan 1, 2016.

99. Sato, D., Shiba, T., Karaki, T., Yamagata, W., Nozaki, T., Nakazawa, T. and Harada, S. (2017) X-Ray snapshots of a pyridoxal enzyme: a catalytic mechanism involving concerted 1,5-hydrogen sigmatropy in methionine γ-lyase, *Scientific reports,* **7,** 4874. First published on Jul 7, 2017.

100. Sato, D. and Nozaki, T. (2009) Methionine gamma-lyase: the unique reaction mechanism, physiological roles, and therapeutic applications against infectious diseases and cancers, *IUBMB life,* **61,** 1019–1028.

101. Faleev, N.G., Alferov, K.V., Tsvetikova, M.A., Morozova, E.A., Revtovich, S.V., Khurs, E.N., Vorob'ev, M.M., Phillips, R.S., Demidkina, T.V. and Khomutov, R.M. (2009) Methionine gamma-lyase: mechanistic deductions from the kinetic pH-effects. The role of the ionic state of a substrate in the enzymatic activity, *Biochimica et biophysica acta,* **1794,** 1414–1420. First published on Jun 6, 2009.

102. Hafkenscheid, J.C. and Kohler, B.E. (1986) Effects of temperature on measurement of aspartate aminotransferase and alanine aminotransferase in commercial control sera, *Clinical chemistry,* **32,** 184–185.

103. Barra, L., Awakawa, T., Shirai, K., Hu, Z., Bashiri, G. and Abe, I. (2021) β-NAD as a building block in natural product biosynthesis, *Nature.* First published on Dec 8, 2021.

104. Caulkins, B.G., Bastin, B., Yang, C., Neubauer, T.J., Young, R.P., Hilario, E., Huang, Y.M., Chang, C.A., Fan, L. and Dunn, M.F. *et al.* (2014) Protonation states of the tryptophan synthase internal aldimine active site from solid-state NMR spectroscopy: direct observation of the protonated Schiff base linkage to pyridoxal-5'-phosphate, *Journal of the American Chemical Society,* **136,** 12824–12827. First published on Sep 3, 2014.

105. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022) ColabFold: Making Protein folding accessible to all, *Nature Methods*.

106. Mirdita, M., Steinegger, M. and S"oding, J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches, *Bioinformatics,* **35,** 2856–2858.

107. Mirdita, M., den Driesch, L. von, Galiez, C., Martin, M.J., S"oding, J. and Steinegger, M. (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments, *Nucleic acids research,* **45,** D170-D176.

108. Mitchell, A.L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M.R., Kale, V., Potter, S.C. and Richardson, L.J. *et al.* (2019) MGnify: the microbiome analysis resource in 2020, *Nucleic acids research*.

109. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Z'idek, A. and Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold, *Nature*.

110. Jansonius, J.N. (1998) Structure, evolution and action of vitamin B6-dependent enzymes, *Current Opinion in Structural Biology,* **8,** 759–769.

111. Gottschalk, G. (1986) Bacterial Metabolism. Springer New York, New York, NY.

112. Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite, *Trends in Genetics,* **16,** 276–277.

113. Liang, X., Deng, H., Bai, Y., Fan, T.-P., Zheng, X. and Cai, Y. (2021) Characterization of a Novel Type Homoserine Dehydrogenase Only with High Oxidation Activity from Arthrobacter nicotinovorans.

114. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures, *Nucleic acids research,* **33,** W299-302.

115. Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks, *Proceedings of the National Academy of Sciences of the United States of America,* **89,** 10915–10919.

116. Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic acids research,* **49,** W293-W296.

117. Brzović, P., Holbrook, E.L., Greene, R.C. and Dunn, M.F. (1990) Reaction mechanism of Escherichia coli cystathionine gamma-synthase: direct evidence for a pyridoxamine derivative of vinylglyoxylate as a key intermediate in pyridoxal phosphate dependent gamma-elimination and gamma-replacement reactions, *Biochemistry,* **29,** 442–451.

118. Chen, M., Liu, C.-T. and Tang, Y. (2020) Discovery and Biocatalytic Application of a PLP-Dependent Amino Acid γ-Substitution Enzyme That Catalyzes C-C Bond Formation, *Journal of the American Chemical Society,* **142,** 10506–10515. First published on Jun 1, 2020.

119. Yee, D.A., Kakule, T.B., Cheng, W., Chen, M., Chong, C.T.Y., Hai, Y., Hang, L.F., Hung, Y.-S., Liu, N. and Ohashi, M. *et al.* (2020) Genome Mining of Alkaloidal Terpenoids from a Hybrid Terpene and Nonribosomal Peptide Biosynthetic Pathway, *Journal of the American Chemical Society,* **142,** 710–714. First published on Jan 3, 2020.

120. Faulkner, J.R., Hussaini, S.R., Blankenship, J.D., Pal, S., Branan, B.M., Grossman, R.B. and Schardl, C.L. (2006) On the sequence of bond formation in loline alkaloid biosynthesis, *Chembiochem : a European journal of chemical biology,* **7,** 1078–1088.

121. Cui, Z., Overbay, J., Wang, X., Liu, X., Zhang, Y., Bhardwaj, M., Lemke, A., Wiegmann, D., Niro, G. and Thorson, J.S. *et al.* (2020) Pyridoxal-5′-phosphate-dependent alkyl transfer in nucleoside antibiotic biosynthesis, *Nat Chem Biol,* **16,** 904–911, https://www.nature.com/articles/s41589-020-0548-3.

122. Borcsok, E. and Abeles, R.H. (1982) Mechanism of action of cystathionine synthase, *Archives of Biochemistry and Biophysics,* **213,** 695–707.

123. Murooka, Y., Kakihara, K., Miwa, T., Seto, K. and Harada, T. (1977) O-alkylhomoserine synthesis catalyzed by O-acetylhomoserine sulfhydrylase in microorganisms, *Journal of bacteriology,* **130,** 62–73.

124. Berntsson, R.P.-A., Smits, S.H.J., Schmitt, L., Slotboom, D.-J. and Poolman, B. (2010) A structural classification of substrate-binding proteins, *FEBS letters,* **584,** 2606–2617. First published on Apr 20, 2010.

125. Milo, R., Jorgensen, P., Moran, U., Weber, G. and Springer, M. (2010) BioNumbers--the database of key numbers in molecular and cell biology, *Nucleic acids research,* **38,** D750-3. First published on Oct 23, 2009.

126. Heldal, M., Norland, S., Fagerbakke, K.M., Thingstad, F. and Bratbak, G. (1996) The elemental composition of bacteria: A signature of growth conditions?, *Marine Pollution Bulletin,* **33,** 3–9.

127. Sinn, M., Hauth, F., Lenkeit, F., Weinberg, Z. and Hartig, J.S. (2021) Widespread bacterial utilization of guanidine as nitrogen source, *Molecular microbiology.* First published on Feb 15, 2021.

128. Rosenthal, G.A. (2001) L-Canavanine: a higher plant insecticidal allelochemical, *Amino acids,* **21,** 319–330.

129. Keshavan, N.D., Chowdhary, P.K., Haines, D.C. and González, J.E. (2005) L-Canavanine made by Medicago sativa interferes with quorum sensing in Sinorhizobium meliloti, *Journal of bacteriology,* **187,** 8427–8436.

130. Mendrygal, K.E. and González, J.E. (2000) Environmental regulation of exopolysaccharide production in Sinorhizobium meliloti, *Journal of bacteriology,* **182,** 599–606.

131. Flores-Tinoco, C.E., Tschan, F., Fuhrer, T., Margot, C., Sauer, U., Christen, M. and Christen, B. (2020) Co-catabolism of arginine and succinate drives symbiotic nitrogen fixation, *Molecular systems biology,* **16,** e9419.

132. Farssi, O., Saih, R., El Moukhtari, A., Oubenali, A., Mouradi, M., Lazali, M., Ghoulam, C., Bouizgaren, A., Berrougui, H. and Farissi, M. (2021) Synergistic effect of Pseudomonas alkylphenolica PF9 and Sinorhizobium meliloti Rm41 on Moroccan alfalfa population grown under limited phosphorus availability, *Saudi journal of biological sciences,* **28,** 3870–3879. First published on Apr 2, 2021.

133. Korir, H., Mungai, N.W., Thuita, M., Hamba, Y. and Masso, C. (2017) Co-inoculation Effect of Rhizobia and Plant Growth Promoting Rhizobacteria on Common Bean Growth in a Low Phosphorus Soil, *Frontiers in plant science,* **8,** 141. First published on Feb 7, 2017.

134. Matse, D.T., Huang, C.-H., Huang, Y.-M. and Yen, M.-Y. (2020) Effects of coinoculation of Rhizobium with plant growth promoting rhizobacteria on the nitrogen fixation and nutrient uptake of Trifolium repens in low phosphorus soil, *Journal of Plant Nutrition,* **43,** 739–752.

135. Shalev, O., Karasov, T.L., Lundberg, D.S., Ashkenazy, H., Pramoj Na Ayutthaya, P. and Weigel, D. (2022) Commensal Pseudomonas strains facilitate protective response against pathogens in the host plant, *Nature ecology & evolution.* First published on Feb 24, 2022.

136. Chan, P.P. and Lowe, T.M. (2019) tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences, *Methods in molecular biology (Clifton, N.J.),* **1962,** 1–14.

137. Fersht, A.R. (1977) Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase, *Biochemistry,* **16,** 1025–1030.

138. Mascarenhas, A.P., An, S., Rosen, A.E., Martinis, S.A. and Musier-Forsyth, K. (2009) Fidelity Mechanisms of the Aminoacyl-tRNA Synthetases. InKöhrer, C. and RajBhandary, U.L. (eds.), *Protein Engineering.* Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 155–203.

139. Reynolds, N.M., Lazazzera, B.A. and Ibba, M. (2010) Cellular mechanisms that control mistranslation, *Nature reviews. Microbiology,* **8,** 849–856.

140. Igloi, G.L. and Schiefermayr, E. (2009) Amino acid discrimination by arginyl-tRNA synthetases as revealed by an examination of natural specificity variants, *The FEBS journal,* **276,** 1307–1318.

141. Crine, P. and Lemieux, E. (1982) Incorporation of canavanine into rat pars intermedia proteins inhibits the maturation of pro-opiomelanocortin, the common precursor to adrenocorticotropin and beta-lipotropin, *The Journal of biological chemistry,* **257,** 832–838.

142. Holden, N. (2019) You Are What You Can Find to Eat: Bacterial Metabolism in the Rhizosphere, *Current issues in molecular biology,* **30,** 1–16. First published on Aug 2, 2018.

143. Boyar, A. and Marsh, R.E. (1982) l-Canavanine, a paradigm for the structures of substituted guanidines, *J. Am. Chem. Soc.,* **104,** 1995–1998.

144. Fitch, C.A., Platzer, G., Okon, M., Garcia-Moreno, B.E. and McIntosh, L.P. (2015) Arginine: Its pKa value revisited, *Protein science : a publication of the Protein Society,* **24,** 752–761. First published on Mar 22, 2015.

145. Nelson, J.W., Atilho, R.M., Sherlock, M.E., Stockbridge, R.B. and Breaker, R.R. (2017) Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class, *Molecular cell,* **65,** 220–230. First published on Dec 15, 2016.

146. Sherlock, M.E. and Breaker, R.R. (2017) Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria, *Biochemistry,* **56,** 359–363. First published on Jan 6, 2017.

147. Sherlock, M.E., Malkowski, S.N. and Breaker, R.R. (2017) Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria, *Biochemistry,* **56,** 352–358. First published on Jan 6, 2017.

148. Lenkeit, F., Eckert, I., Hartig, J.S. and Weinberg, Z. (2020) Discovery and characterization of a fourth class of guanidine riboswitches, *Nucleic acids research,* **48,** 12889–12899.

149. Salvail, H., Balaji, A., Yu, D., Roth, A. and Breaker, R.R. (2020) Biochemical Validation of a Fourth Guanidine Riboswitch Class in Bacteria.

150. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N. and Jona, I. *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control, *Proceedings of the National Academy of Sciences of the United States of America,* **101,** 6421–6426. First published on Apr 19, 2004.

151. Battaglia, R.A., Price, I.R. and Ke, A. (2017) Structural basis for guanidine sensing by the ykkC family of riboswitches, *RNA (New York, N.Y.),* **23,** 578–585. First published on Jan 17, 2017.

152. Reiss, C.W., Xiong, Y. and Strobel, S.A. (2017) Structural Basis for Ligand Binding to the Guanidine-I Riboswitch, *Structure,* **25,** 195–202, https://www.sciencedirect.com/science/article/pii/S0969212616303628.

153. Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F. and Sudarsan, N. *et al.* (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline, *Nucleic acids research,* **35,** 4809–4819. First published on Jul 9, 2007.

154. Huang, L., Wang, J. and Lilley, D.M.J. (2017) The Structure of the Guanidine-II Riboswitch, *Cell chemical biology,* **24,** 695-702.e2. First published on May 18, 2017.

155. Weinberg, Z., Nelson, J.W., Lünse, C.E., Sherlock, M.E. and Breaker, R.R. (2017) Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity, *PNAS,* **114,** E2077-E2085.

156. Huang, L., Wang, J., Wilson, T.J. and Lilley, D.M.J. (2017) Structure of the Guanidine III Riboswitch, *Cell chemical biology,* **24,** 1407-1415.e2. First published on Oct 5, 2017.

157. Sherlock, M.E. and Breaker, R.R. (2020) Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling, and gene control processes, *RNA (New York, N.Y.),* **26,** 675–693.

158. Meyer, M.M., Hammond, M.C., Salinas, Y., Roth, A., Sudarsan, N. and Breaker, R.R. (2011) Challenges of ligand identification for riboswitch candidates, *RNA biology,* **8,** 5–10. First published on Jan 1, 2011.

159. Fukuda, H., Ogawa, T., Tazaki, M., Nagahama, K., Fujiil, T., Tanase, S. and Morino, Y. (1992) Two reactions are simultaneously catalyzed by a single enzyme: The arginine-dependent simultaneous formation of two products, ethylene and succinate, from 2-oxoglutarate by an enzyme from Pseudomonas syringae, *Biochemical and Biophysical Research Communications,* **188,** 483–489.

160. Hausinger, R.P. (2004) Metabolic versatility of prokaryotes for urea decomposition, *Journal of bacteriology,* **186,** https://pubmed.ncbi.nlm.nih.gov/15090490/.

161. Kihara, H., Prescott, J.M. and Snell, E.E. (1955) The bacterial cleavage of canavanine to homoserine and guanidine, *The Journal of biological chemistry,* **217,** 497–503.

162. Natelson, S. and Sherwin, J.E. (1979) Proposed mechanism for urea nitrogen re-utilization: relationship between urea and proposed guanidine cycles, *Clinical chemistry,* **25,** 1343–1344.

163. Schneider, N.O., Tassoulas, L.J., Zeng, D., Laseke, A.J., Reiter, N.J., Wackett, L.P. and Maurice, M.S. (2020) Solving the Conundrum: Widespread Proteins Annotated for Urea Metabolism in Bacteria Are Carboxyguanidine Deiminases Mediating Nitrogen Assimilation from Guanidine, *Biochemistry,* **59,** 3258–3270. First published on Aug 25, 2020.

164. Funck, D., Sinn, M., Fleming, J.R., Stanoppi, M., Dietrich, J., López-Igual, R., Mayans, O. and Hartig, J.S. (2022) Discovery of a Ni2+-dependent guanidine hydrolase in bacteria, *Nature,* **603,** 515–521. First published on Mar 9, 2022.

165. Kanamori, T., Kanou, N., Atomi, H. and Imanaka, T. (2004) Enzymatic Characterization of a Prokaryotic Urea Carboxylase, *Journal of bacteriology,* **186,** 2532–2539.

166. Locher, K.P. (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters, *Nature structural & molecular biology,* **23,** 487–493.

167. Wilkens, S. (2015) Structure and mechanism of ABC transporters, *F1000Prime Reports,* **7**.

168. Maqbool, A., Horler, R.S.P., Muller, A., Wilkinson, A.J., Wilson, K.S. and Thomas, G.H. (2015) The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity, *Biochemical Society transactions,* **43,** 1011–1017.

169. Schicklberger, M., Shapiro, N., Loqué, D., Woyke, T. and Chakraborty, R. (2015) Draft Genome Sequence of Raoultella terrigena R1Gly, a Diazotrophic Endophyte, *Genome announcements,* **3,** https://pubmed.ncbi.nlm.nih.gov/26067957/.

170. Huang, H.C., Erickson, R.S., Yanke, L.J., Hsieh, T.F. and Morrall, R.A.A. (2003) First Report of Pink Seed of Lentil and Chickpea Caused by Erwinia rhapontici in Canada, *Plant disease,* **87,** 1398.

171. Saha, R., Farrance, C.E., Verghese, B., Hong, S. and Donofrio, R.S. (2013) Klebsiella michiganensis sp. nov., a new bacterium isolated from a tooth brush holder, *Current microbiology,* **66,** 72–78.

172. Boer, J.L., Mulrooney, S.B. and Hausinger, R.P. (2014) Nickel-dependent metalloenzymes, *Archives of Biochemistry and Biophysics,* **544,** 142–152.

173. Fan, C., Chou, C.-Y., Tong, L. and Xiang, S. (2012) Crystal Structure of Urea Carboxylase Provides Insights into the Carboxyltransfer Reaction*, *The Journal of biological chemistry,* **287,** 9389–9398.

174. Roon, R.J. and Levenberg, B. (1972) Urea amidolyase. I. Properties of the enzyme from Candida utilis, *The Journal of biological chemistry,* **247,** 4107–4113.

175. Roon, R.J., Hampshire, J. and Levenberg, B. (1972) Urea amidolyase. The involvement of biotin in urea cleavage, *The Journal of biological chemistry,* **247,** 7539–7545.

176. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics,* **25,** 1189–1191. First published on Jan 16, 2009.

177. Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic acids research,* **47,** W256-W259.

178. Erica B. Young and John A. Berges Nitrogen stress in the marine environment: from scarcity to surfeit, *Stressors in the Marine Environment.* Oxford University Press.

179. Roon, R.J. and Levenberg, B. (1968) An adenosine triphosphate-dependent, avidin-sensitive enzymatic cleavage of urea in yeast and green algae, *The Journal of biological chemistry,* **243,** 5213–5215.

180. Farrugia, M.A., Macomber, L. and Hausinger, R.P. (2013) Biosynthesis of the urease metallocenter, *The Journal of biological chemistry,* **288,** 13178–13185.

181. Salvail, H., Balaji, A., Yu, D., Roth, A. and Breaker, R.R. (2020) Biochemical Validation of a Fourth Guanidine Riboswitch Class in Bacteria, *Biochemistry,* **59,** 4654–4662. First published on Nov 25, 2020.

182. Kuhn, M.L., Majorek, K.A., Minor, W. and Anderson, W.F. (2013) Broad-substrate screen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity, *Protein science : a publication of the Protein Society,* **22,** 222–230. First published on Dec 17, 2012.

183. Salah Ud-Din, A.I.M., Tikhomirova, A. and Roujeinikova, A. (2016) Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT), *International Journal of Molecular Sciences,* **17**.

184. Burckhardt, R.M. and Escalante-Semerena, J.C. (2020) Small-Molecule Acetylation by GCN5-Related N-Acetyltransferases in Bacteria, *Microbiology and molecular biology reviews : MMBR,* **84.** First published on Apr 15, 2020.

185. Favrot, L., Blanchard, J.S. and Vergnolle, O. (2016) Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation, *Biochemistry,* **55,** 989–1002. First published on Feb 9, 2016.

186. Bikmetov, D., Hall, A.M.J., Livenskyi, A., Gollan, B., Ovchinnikov, S., Gilep, K., Kim, J.Y., Larrouy-Maumus, G., Zgoda, V. and Borukhov, S. *et al.* (2022) GNAT toxins evolve toward narrow tRNA target specificities, *Nucleic acids research,* **50,** 5807–5817.

187. Serganov, A. and Patel, D.J. (2009) Amino acid recognition and gene regulation by riboswitches, *Biochimica et biophysica acta,* **1789,** 592–611. First published on Jul 18, 2009.

188. Walker, J.B. and Walker, M.S. (1959) The Enzymatic Reduction of Hydroxyguanidine, *The Journal of biological chemistry,* **234,** 1481–1484.

189. Sasaki, H.M., Sekine, S., Sengoku, T., Fukunaga, R., Hattori, M., Utsunomiya, Y., Kuroishi, C., Kuramitsu, S., Shirouzu, M. and Yokoyama, S. (2006) Structural and mutational studies of the amino acid-editing domain from archaeal/eukaryal phenylalanyl-tRNA synthetase, *Proceedings of the National Academy of Sciences of the United States of America,* **103,** 14744–14749. First published on Sep 26, 2006.

190. Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study, *Journal of bacteriology,* **173,** 697–703.

191. Hesse, C., Schulz, F., Bull, C.T., Shaffer, B.T., Yan, Q., Shapiro, N., Hassan, K.A., Varghese, N., Elbourne, L.D.H. and Paulsen, I.T. *et al.* (2018) Genome-based evolutionary history of *Pseudomonas spp*, *Environmental microbiology,* **20,** 2142–2159. First published on Jul 22, 2018.

192. Jeffries, C.D., Holtman, D.F. and Guse, D.G. (1957) Rapid method for determining the activity of microorganisms on nucleic acids, *Journal of bacteriology,* **73,** 590–591.

193. King, E.O., Ward, M.K. and Raney, D.E. (1954) Two simple media for the demonstration of pyocyanin and fluorescin, *The Journal of laboratory and clinical medicine,* **44,** 301–307.

194. Kovacs, N. (1956) Identification of *Pseudomonas pyocyanea* by the oxidase reaction, *Nature,* **178,** 703.

195. Schrödinger, L.L. (2015) The PyMOL Molecular Graphics System, Version 1.8.

196. Huang, W. and Wilks, A. (2017) A rapid seamless method for gene knockout in Pseudomonas aeruginosa, *BMC microbiology,* **17,** 199. First published on Sep 19, 2017.

197. Avcilar-Kucukgoze, I., Gamper, H., Hou, Y.-M. and Kashina, A. (2020) Purification and Use of tRNA for Enzymatic Post-translational Addition of Amino Acids to Proteins, *STAR protocols,* **1,** 100207. First published on Dec 9, 2020.

198. Jeske, M. and Altenbuchner, J. (2010) The Escherichia coli rhamnose promoter rhaP(BAD) is in Pseudomonas putida KT2440 independent of Crp-cAMP activation, *Applied microbiology and biotechnology,* **85,** 1923–1933. First published on Sep 30, 2009.

199. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H. and Mori, H. (2005) 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 research : an international journal for rapid publication of reports on genes and genomes,* **12,** 291–299. First published on Jan 9, 2006.

200. Evans, M.E., Clark, W.C., Zheng, G. and Pan, T. (2017) Determination of tRNA aminoacylation levels by high-throughput sequencing, *Nucleic acids research,* **45,** e133.

201. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis, *Nature Methods,* **9,** 671–675.

202. (2015) Database resources of the National Center for Biotechnology Information, *Nucleic acids research,* **43,** https://pubmed.ncbi.nlm.nih.gov/25398906/.

203. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S. and Richardson, L.J. *et al.* (2020) Pfam: The protein families database in 2021, *Nucleic acids research*.

204. Chen, I.-M.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J.R. and Seshadri, R. *et al.* (2019) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes, *Nucleic acids research,* **47,** D666-D677.

205. Hyatt, D., LoCascio, P.F., Hauser, L.J. and Uberbacher, E.C. (2012) Gene and translation initiation site prediction in metagenomic sequences, *Bioinformatics (Oxford, England),* **28,** 2223–2230.

206. Zhang, X. and Bremer, H. (1995) Control of the Escherichia coli rrnB P1 promoter strength by ppGpp, *The Journal of biological chemistry,* **270,** 11181–11189.

207. Kabsch, W. (2010) XDS, *Acta crystallographica. Section D, Biological crystallography,* **66,** 125–132. First published on Jan 22, 2010.

208. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software, *Journal of applied crystallography,* **40,** 658–674. First published on Jul 13, 2007.

209. Bunkóczi, G. and Read, R.J. (2011) Improvement of molecular-replacement models with Sculptor, *Acta crystallographica. Section D, Biological crystallography,* **67,** 303–312. First published on Mar 18, 2011.

210. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot, *Acta crystallographica. Section D, Biological crystallography,* **66,** 486–501. First published on Mar 24, 2010.

211. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S. and McCoy, A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix, *Acta crystallographica. Section D, Structural biology,* **75,** 861–877. First published on Oct 2, 2019.

212. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J. and Chen, V.B. *et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation, *Protein science : a publication of the Protein Society,* **27,** 293–315. First published on Nov 27, 2017.

213. Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M. and Ostell, J. (2016) NCBI prokaryotic genome annotation pipeline, *Nucleic acids research,* **44,** 6614–6624. First published on Jun 24, 2016.

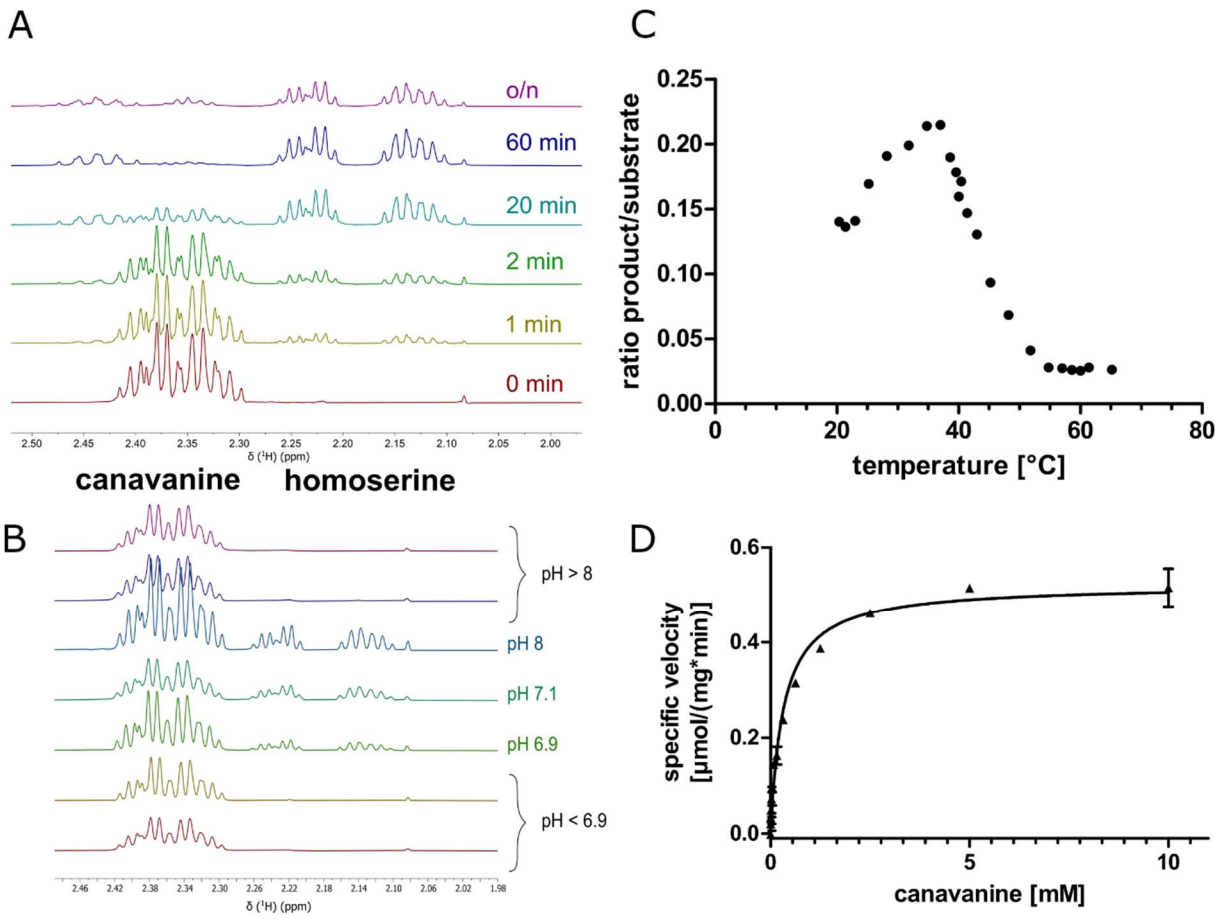
214. Li, W., O'Neill, K.R., Haft, D.H., DiCuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M.K. and Durkin, A.S. *et al.* (2021) RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation, *Nucleic acids research,* **49,** D1020-D1028.

215. Haft, D.H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M.K. and Gonzales, N.R. *et al.* (2018) RefSeq: an update on prokaryotic genome annotation and curation, *Nucleic acids research,* **46,** D851-D860.

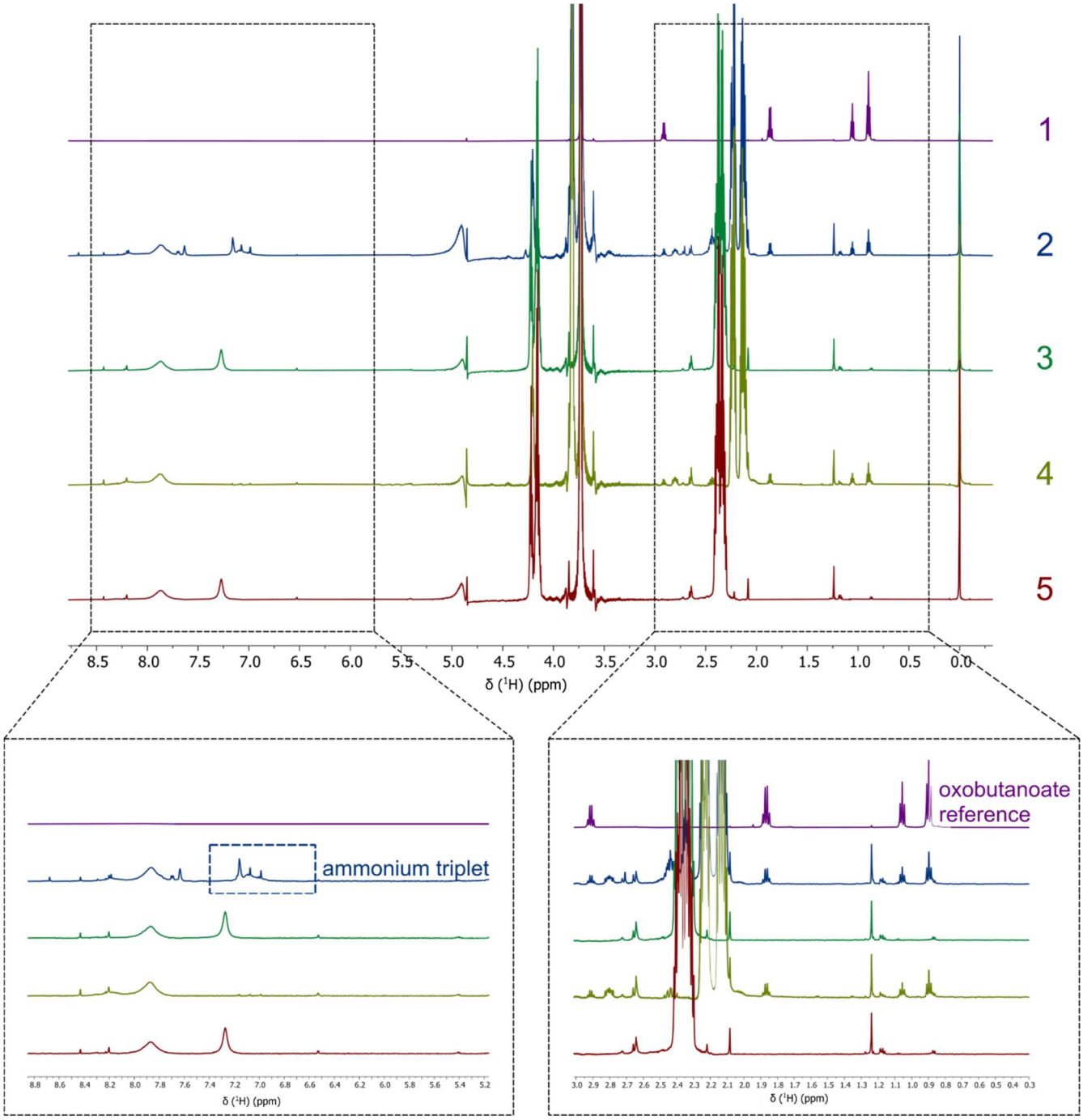
Data availability statement

The Whole Genome Shotgun (WGS) project of the *Pseudomonas canavaninivorans* strain HB002^T (= DSM 112525T = LMG 32336T) was deposited at DDBJ/ENA/GenBank under the accession number JAEKIK000000000 and annotated with the NCBI prokaryotic genome annotation pipeline[213–215]. The 16S rRNA gene sequence of strain HB002^T was deposited at DDBJ/ENA/GenBank under the accession number MZ644983.

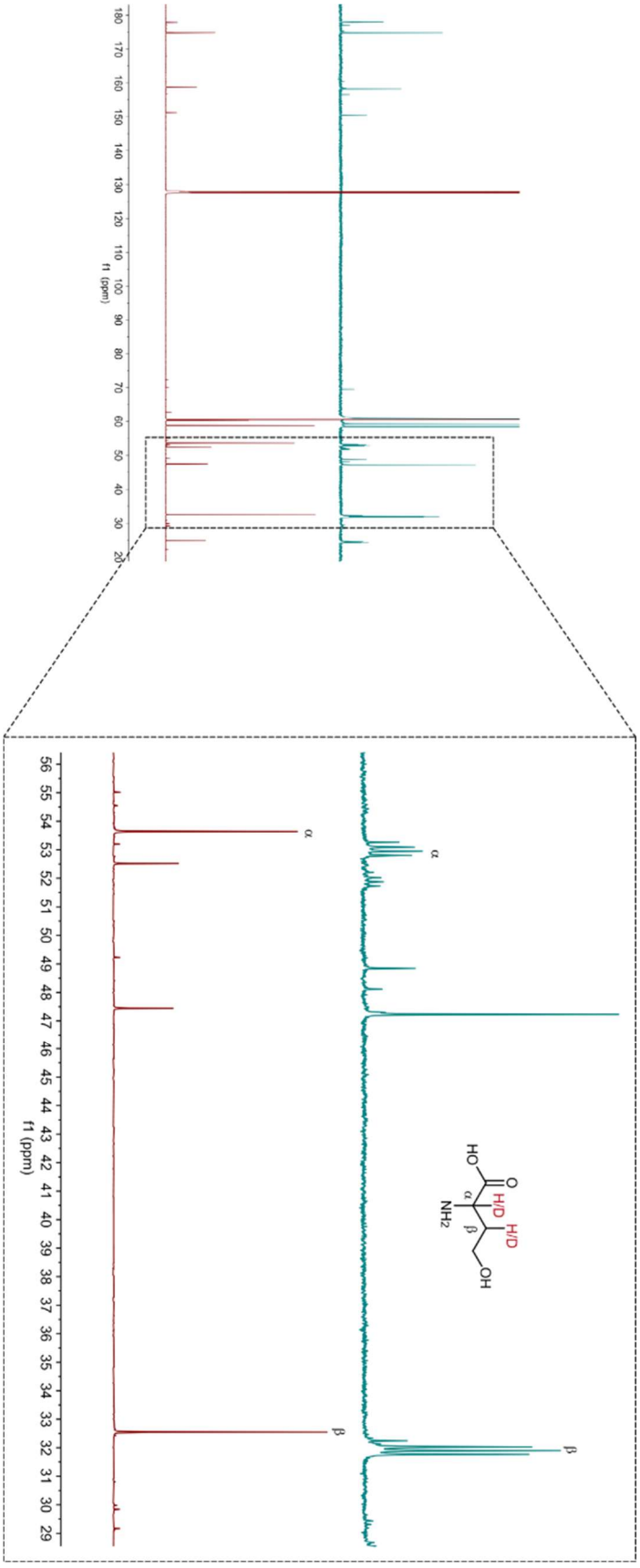
Supplementary material



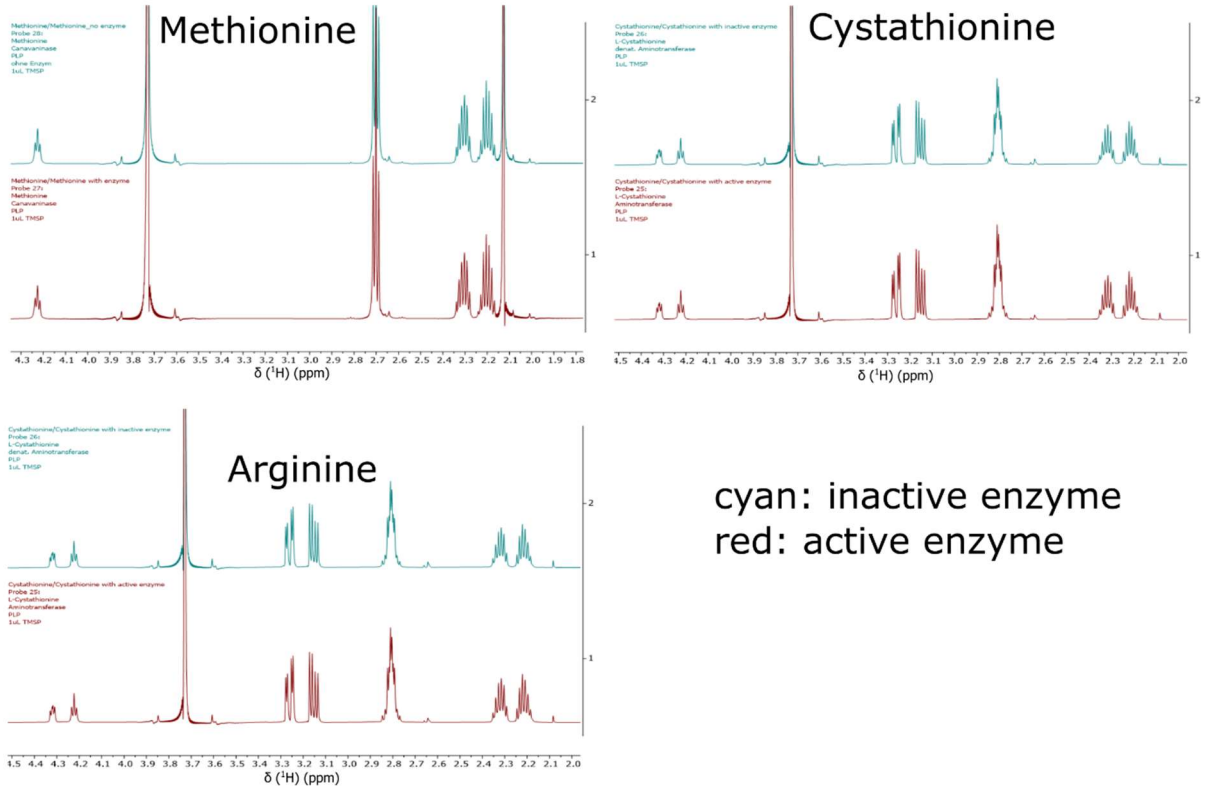
SI Figure 1: CanγL characteristics: **A:** ^{1}H NMR spectra overlay of CanγL reactions stopped at different time points. **B:** ^{1}H NMR spectra overlay of CanγL reactions at different pH values, canavanine (2.42-2.30 ppm), homoserine (2.26 – 2.10 ppm). **C:** Ratio of product/substrate conversion at different temperatures, CanγL reactions were stopped after a certain time by the addition of perchloric acid and measured by MS. **D:** Specific velocities of CanγL at different substrate concentrations and Michaelis Menten kinetics fit.



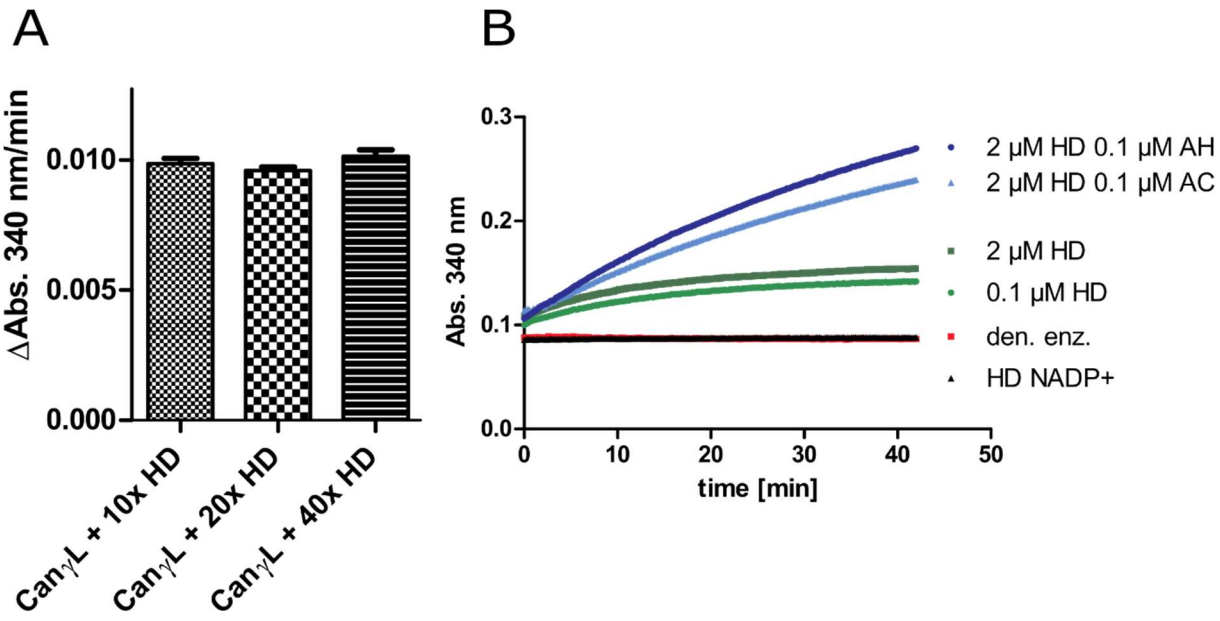
SI Figure 2: Full ^{1}H-NMR spectra of CanγL reaction. **1:** Oxobutanoate reference, **2:** Canavanine with active CanγL **3:** Canavanine with inactivated CanγL, **4:** Homoserine with active enzyme **5:** Canavanine reference, left bottom panel: formation of ammonium triplet, right bottom panel: formation of oxobutanoate



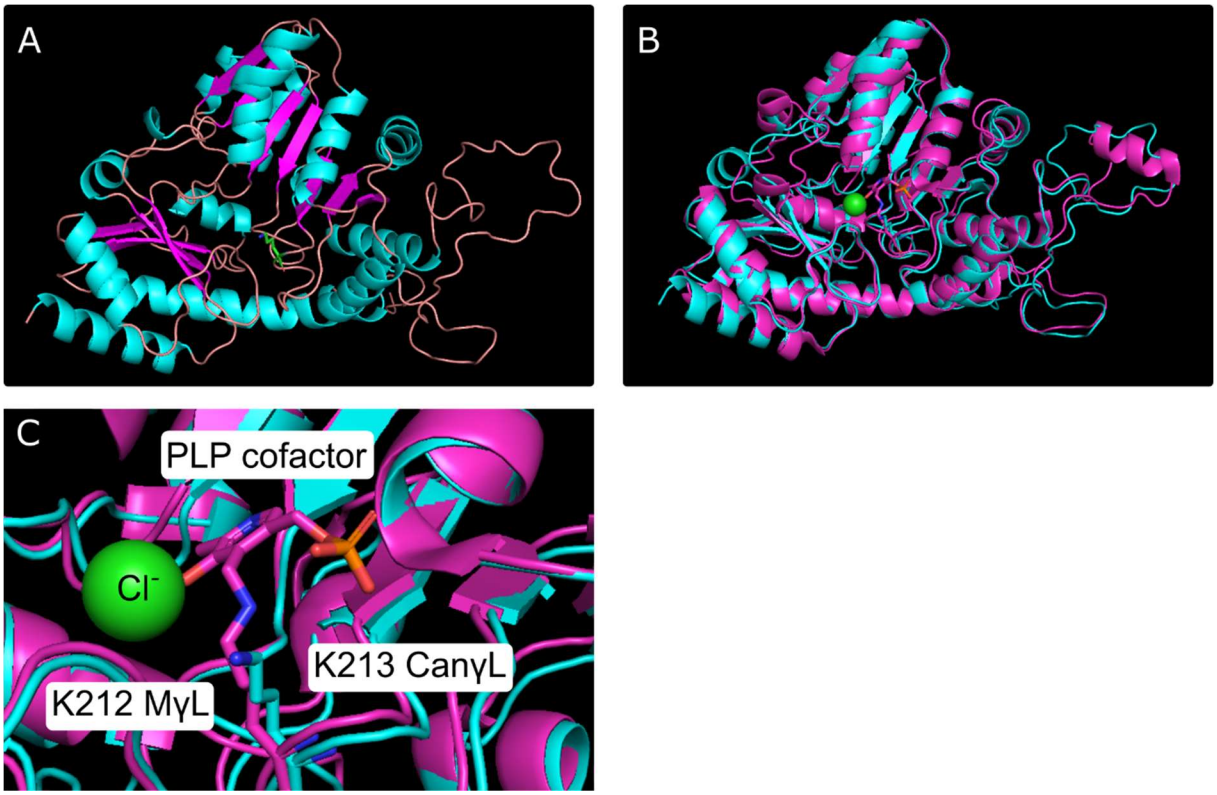
SI Figure 3: Natural abundance ^{13}C-NMR hydrogen/deuterium (H/D) exchange experiment, upper spectrum: canavanine cleavage reaction in D_2O, lower spectrum: canavanine cleavage reaction in H_2O, α/β correspond to the respective canavanine carbon



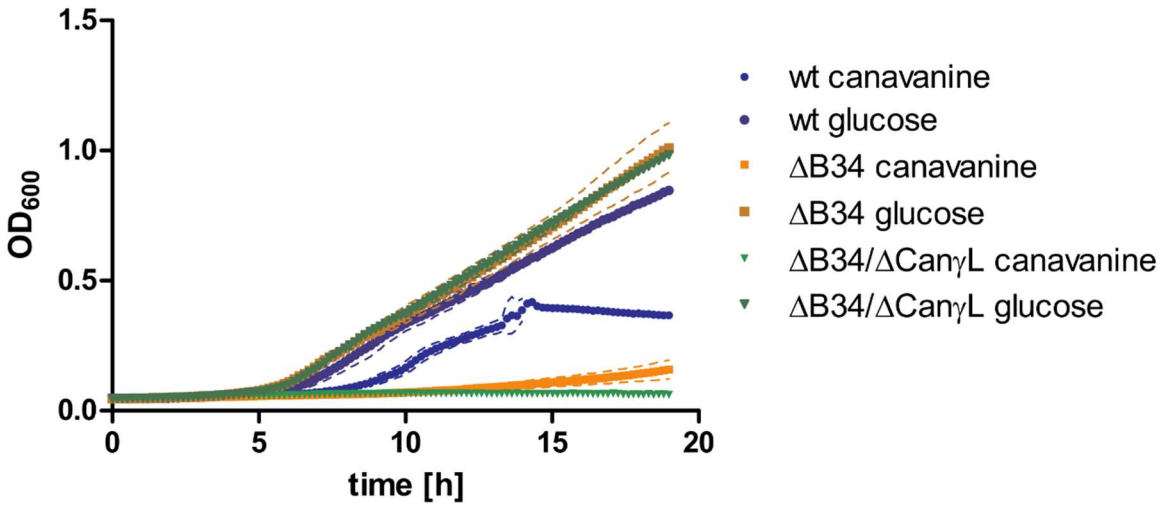
SI Figure 4: ^{1}H-NMR spectra indicating substrate specificity of CanγL, spectra of methionine, cystathionine and arginine with active (red) or heat-inactivated (cyan), no differences in peak intensities or any additional or shifting peaks could be observed comparing active or heat-inactivated CanγL.



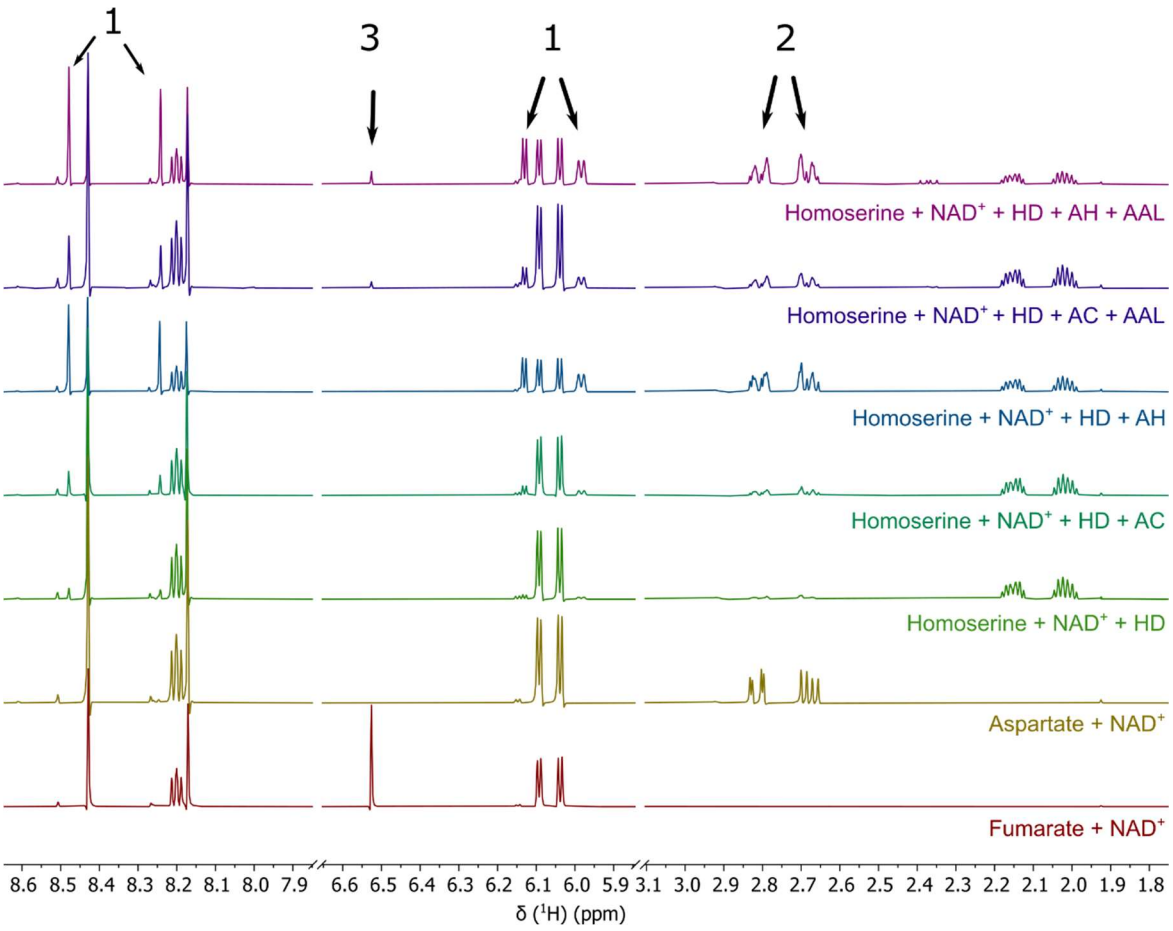
SI Figure 5: A: Determination of HD excess over CanγL for *in vitro* assay. **B:** NADH dependency of HD and further coupled *in vitro* reaction of HD with either of the aldehyde dehydrogenases AC or AH. Each data point represents the mean of a technical triplicate. Den. enz. corresponds to a heat inactivated HD sample.



SI Figure 6: Fold type identification of CanγL. **A:** Alphafold (106,105,107–109) model visualized using PyMol (195). Colours represent secondary structures, bronze: loop, turquois: α-helices, pink: β-sheets. **B:** CanγL model superimposed with methionine-γ-lyase from *Clostridium sporonges* (PBD code 5DX5 chain A), green ball: Cl^- found in the crystal structure of MγL. **C:** Zoom on the PLP cofactor with the PLP binding lysines of CanγL (K213) and MγL (K212).



SI Figure 7: Growth curve of wt, ΔB3/4 and ΔB3/4Δ CanγL strains of *P. canavaninivorans*. The knockout strains were generated by homologous recombination as described in the methods part of this manuscript. Strains were grown on minimal medium with either canavanine or glucose as sole carbon source. Dashed line = SD, n = 3.

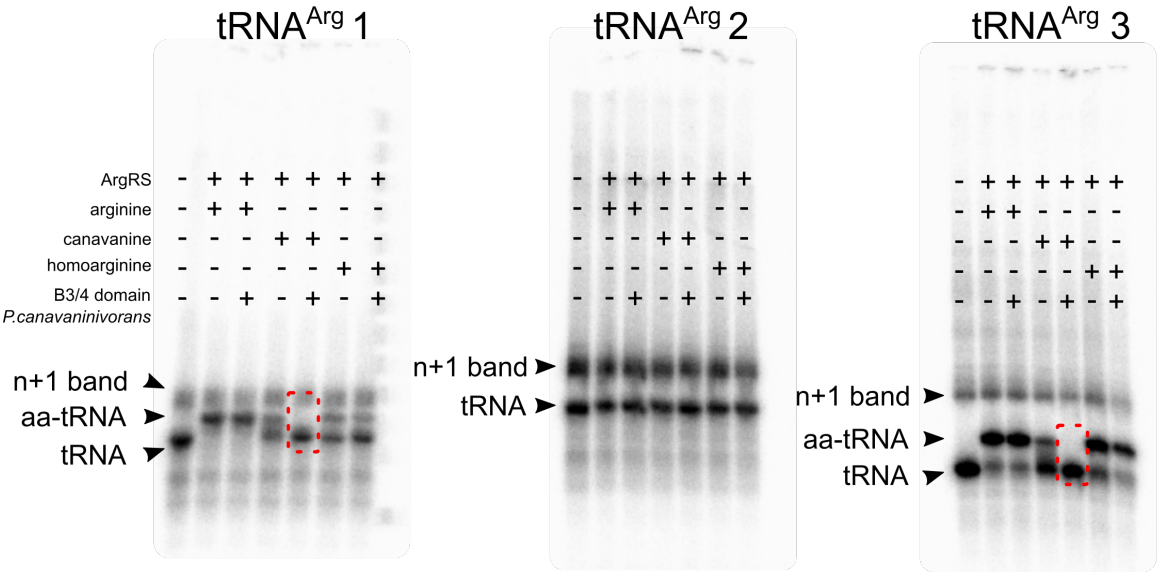


SI Figure 8: Full ^{1}H-NMR of the reaction of homoserine to fumarate via homoserine dehydrogenase (HD), either of the two aldehyde dehydrogenases (AC or AH) and ammonium-aspartate-lyase (AAL). 1: Peaks corresponding to NADH, 2: Peaks corresponding to aspartate, 3: Peak corresponding to fumarate.

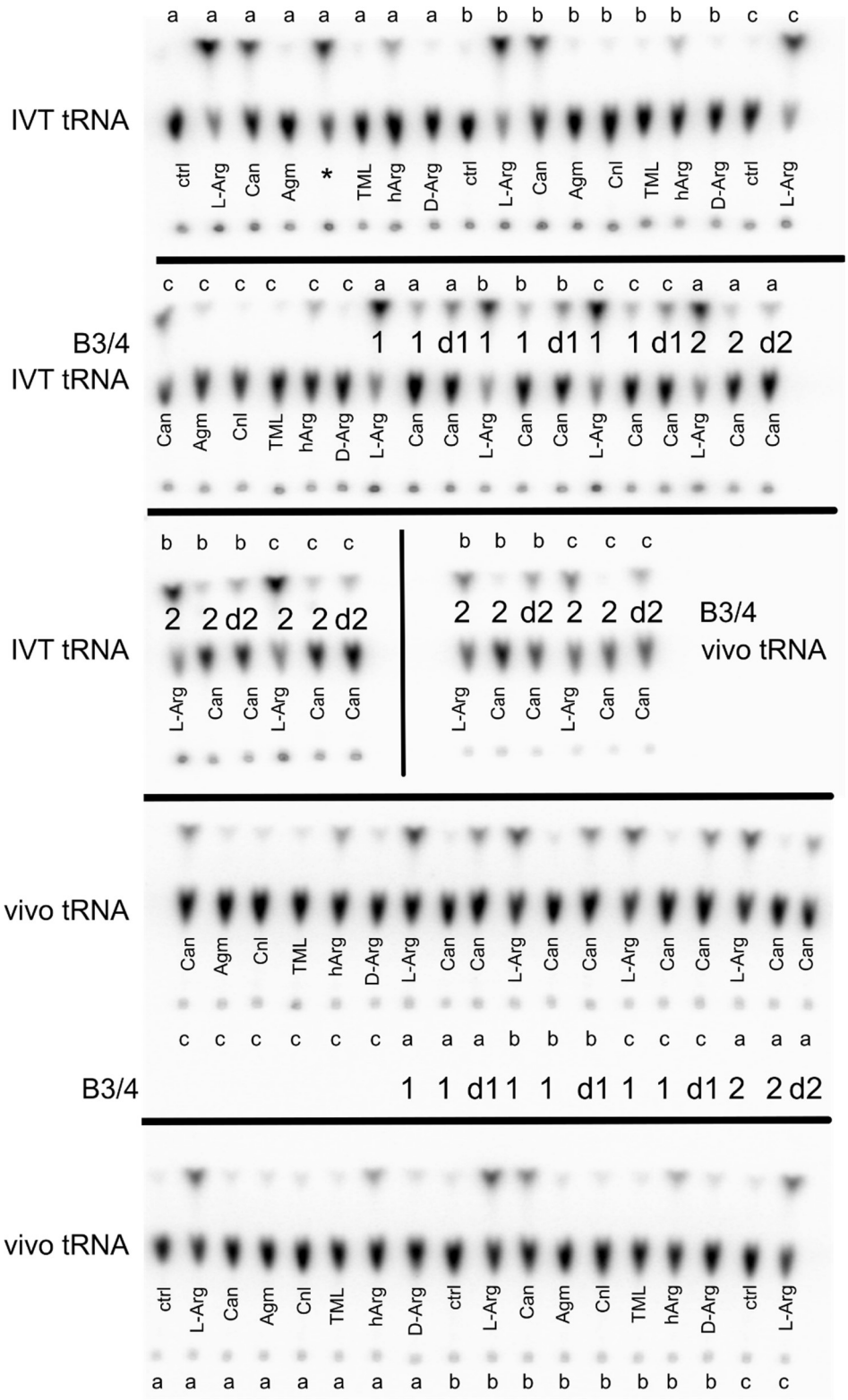
tRNA^{Arg} 1: 5’ – GTCTCAGTAGCTCAATTGGATAGAGCATCCCCCT**CCT**AAGGGGAAGGTTGGCAGTTCGAACCTGCCCTGGGACACCA – 3'

tRNA^{Arg} 2: 5’ – GCACCAGTAGCTCAGCTGGATAGAGTACTGCCCT**CCG**AAGGCAGGGGTCGTGGGTTCGAATCCCGCCTGGTGCACCA – 3'

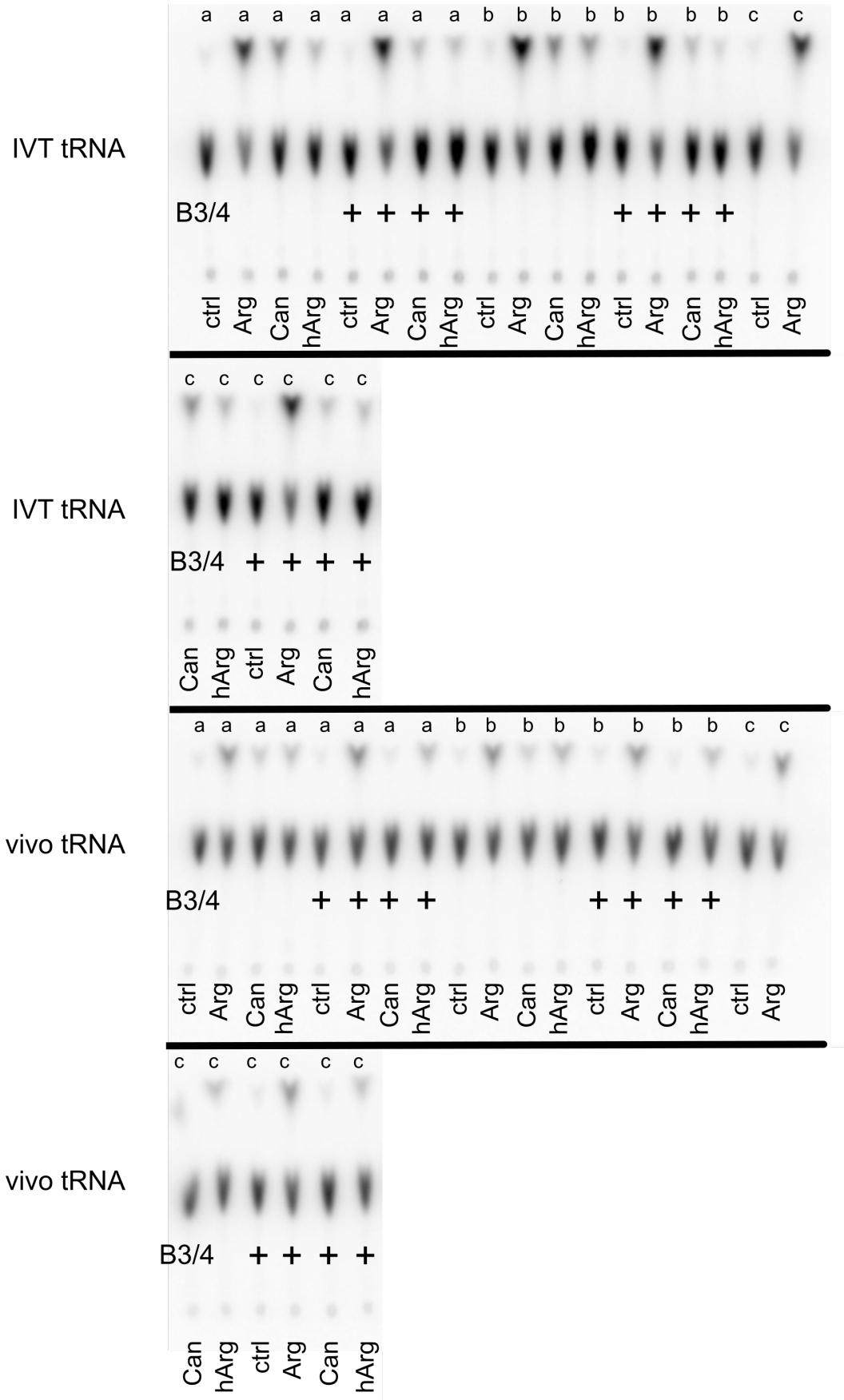
tRNA^{Arg} 3: 5’ – GCGCCCGTAGCTCAGCTGGATAGAGCATCCGCCT**TCT**AAGCGGATGGTCGCAGGTTCGAGTCCTGCCGGGTGCGCCA – 3'



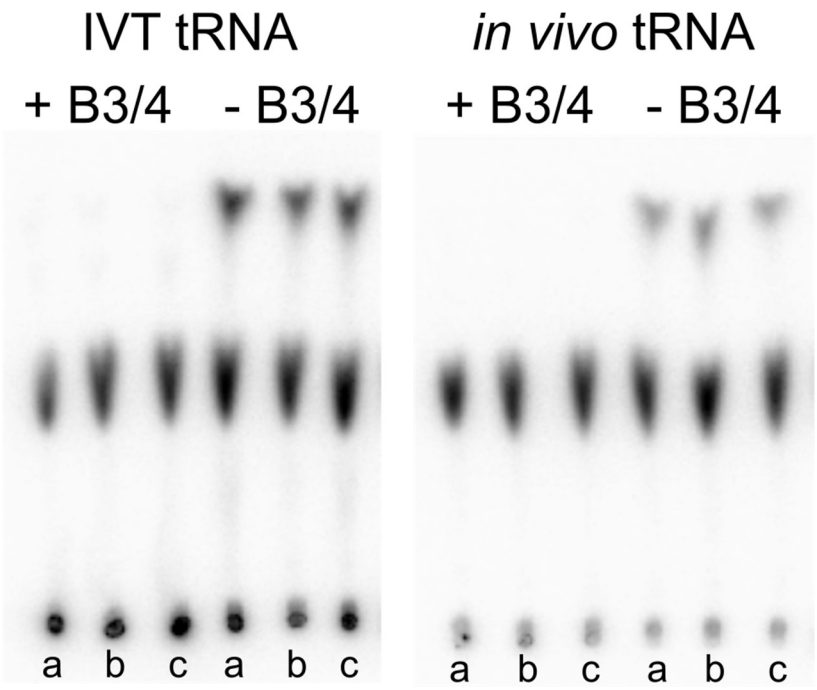
SI Figure 9: Radio screen of an acid PAGE separating the 5’ end γ-^{32}P labelled (aminoacylated) tRNAs. The sequences of the predicted tRNAs are shown above. The n+1 band is a common artefact of *in vitro* produced RNAs, but can be reduced by using modified primers. Aminoacylation reactions were performed in the presence or absence of the respective substrate and B3/4 domain like protein for 2h at 37°C. Only canavanylated but neither arginylated or homoarginylated tRNA was hydrolysed by the B3/4 domain like protein, see red box.



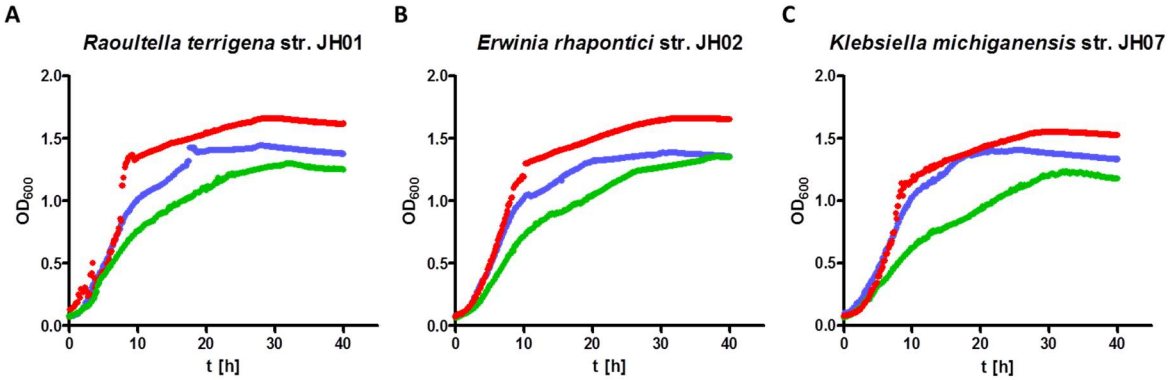
SI Figure 10: Radio screens of TLC separated 3'α^{32}P (aminoacylated) AMP of triplicate experiments (a, b, c). Either *in vitro* or *in vivo* produced tRNA was used in the aminoacylation experiment, which was conducted in the presence of B3/4 domain like protein as indicated. B3/4 1 represents the domain from *P. canavaninivorans*, 2 the domain from *C. perfrigens*, d1/d2 indicates that the domain was heat incubated prior to the aminoacylation reaction, the substrate used in the reaction is shown above the TLC loading spot: ctrl: H_2O, Arg: L-Arginine, Can: canavanine, Agm: negative control agmatine, Cnl: canaline, TML: N-trimethyllysine, hArg: homoarginine, D-Arg: D-arginine, * pipetting error during spotting on the TLC plate.



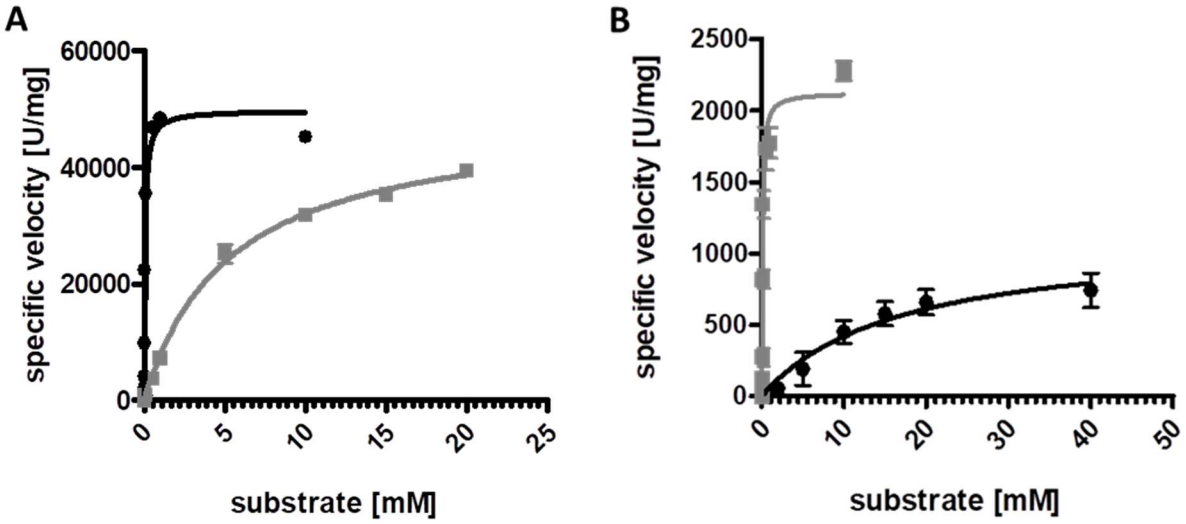
SI Figure 11: Radioscreens of TLC separated 3'α^{32}P (aminoacylated) AMP of triplicate experiments (a, b, c). Either *in vitro* or *in vivo* produced tRNA was used in the aminoacylation experiment, which was conducted in the presence (+) or absence ( ) of B3/4 domain like protein from *P. canavaninivorans*. , the substrate used in the reaction is shown beneath the TLC loading spot: ctrl: H_2O, Arg: L-Arginine, Can: canavanine, hArg: homoarginine.



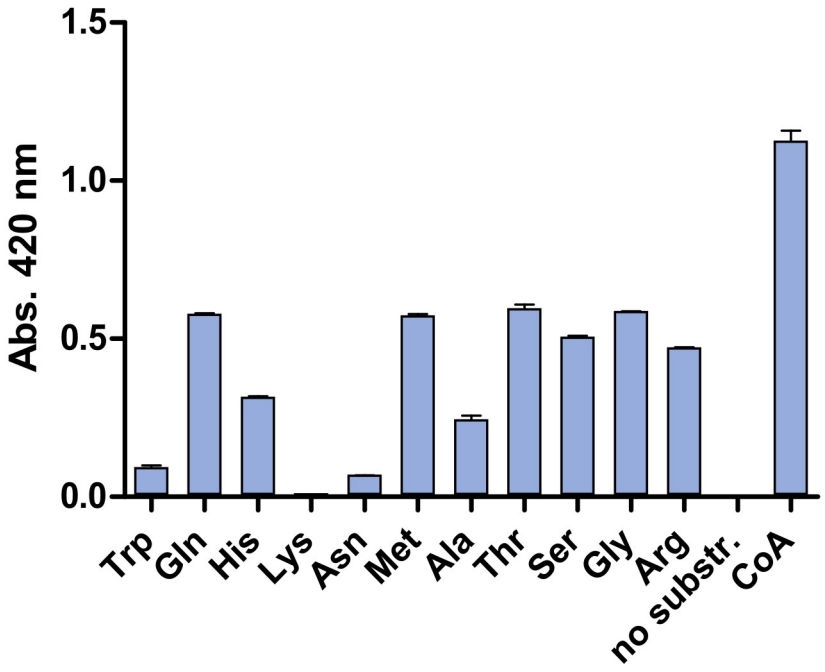
SI Figure 12: Radio screen pictures of TLC separated 3' α^{32}P labelled (canavanylated) AMP of triplicate experiments (a, b, c). Purified canavanylated $tRNA^{Arg}$ was incubated with or without 50 µM of B3/4 domain like protein from *P. canavaninivorans* for one minute, quenched and then spotted onto a TLC plate.



SI Figure 13: Growth of isolated bacterial strains in minimal medium supplemented with Ni^{2+}. Nitrogen source was 5 mM guanidine (green), 10 mM urea (red), 15 mM ammonia (blue), supplemented with 10 µM Ni^{2+}. Values are means of biological duplicates.

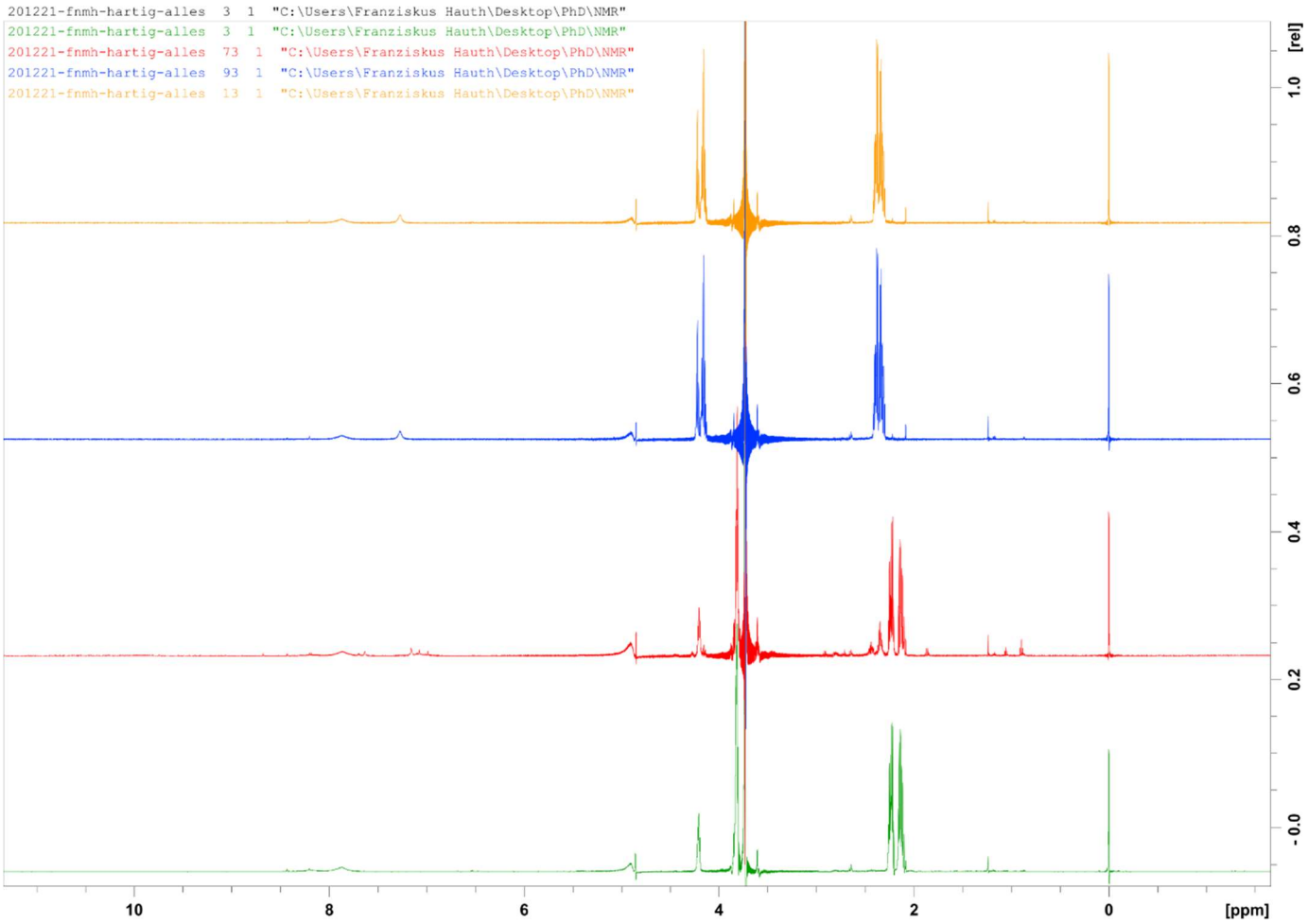


SI Figure 14: Plots of the UC and GC kinetics characterization. Data as shown in Figure 21 A and B plotted against a linear instead of log10 scale.

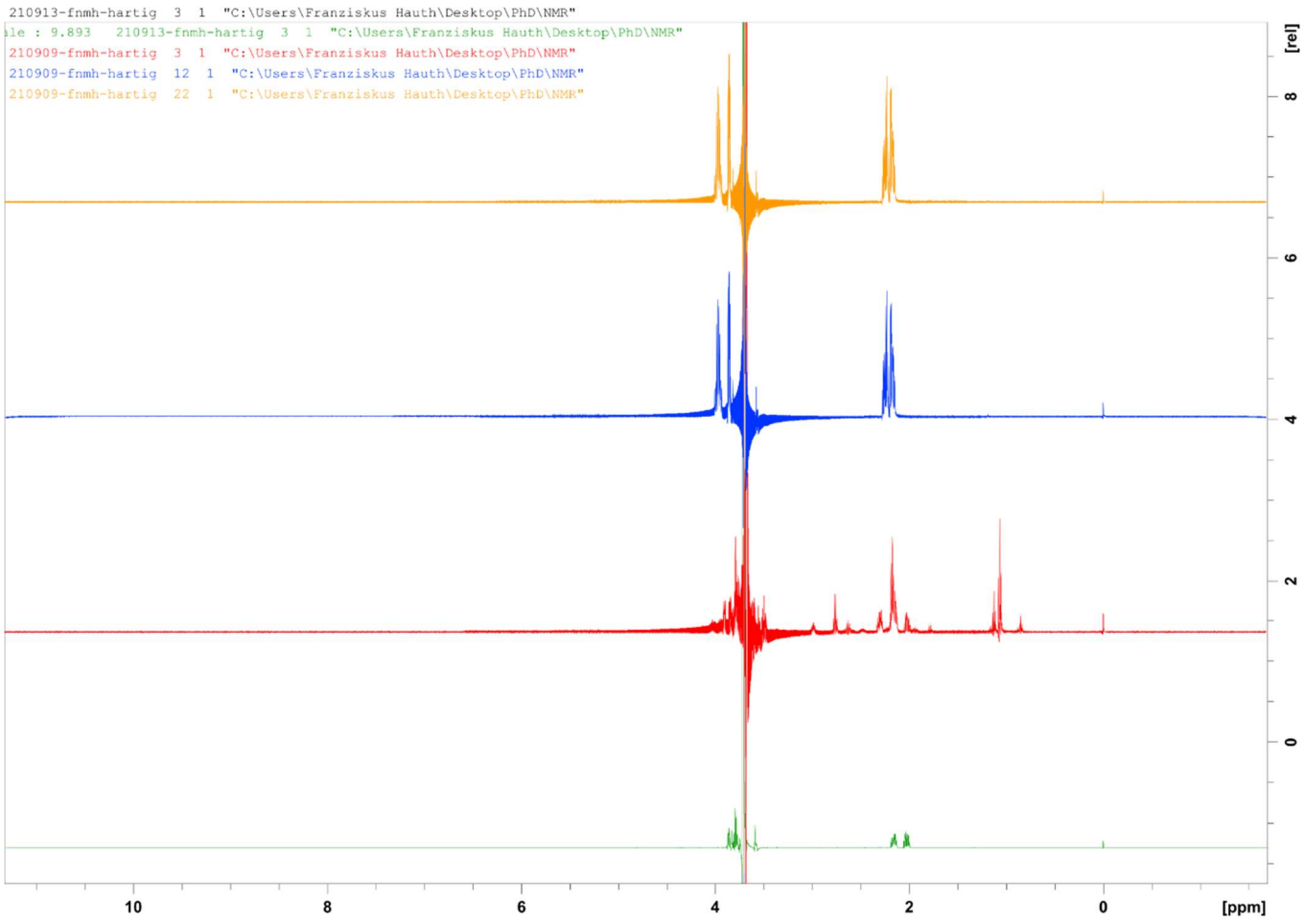


SI Figure 15: Amino acids accepted by *L. curiae* GNAT, mean of technical triplicates, error = SD.

SI Item 1: Full ^{1}H NMR spectra of canavanine degradation by CanγL: **green:** 10 mM homoserine reference, **red:** Reaction with 10 mM canavanine and active CanγL enzyme, **blue:** Reaction with 10 mM canavanine and heat inactivated CanγL enzyme, **orange:** 10 mM canavanine reference. Peak at 0 ppm correspondes to the internal standard TSMP, signals around 3.75 ppm correspond to Tris buffer.



SI Item 2: Full ^{1}H NMR spectra of canavanine degradation by RCanγL: **green:** 10 mM homoserine reference, **red:** Reaction with 50 mM canavanine and active RCanγL enzyme, **blue:** Reaction with 50 mM canavanine and heat inactivated RCanγL enzyme, **orange:** 50 mM canavanine reference. Peak at 0 ppm correspondes to the internal standard TSMP, signals around 3.75 ppm correspond to Tris buffer.



SI Item 3: pet28a_TEV_backbone (5326 bp) sequence and overview plasmid map

GCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGC
CATCATCATCATCATCACAGCAGCGGCGGTGAGAATCTTTATTTTCAGGGCCATATGTTTTTTTTTTTTTTTTTTTTTCTCGAGCACCACCACCACCACCACTGA
GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGG
GTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACC
GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCC
TTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGA
TTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTG
CGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT
CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCT
GCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGA
ATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGC
GCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTC
ACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGT
CGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATC
GGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGC
CTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGT
TTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCC
GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGA
ACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCG
CACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGC
GGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGC
GGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGC
TATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGC
TGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGAT
TCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCC
TGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGA
ACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATA
CAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACA
CGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTA
AGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAA
CGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGG
TCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCC
ACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCC
GCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTC
ACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTT
TGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCG
CGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACT
CCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCG
CTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCA
GAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTG
CGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCG
CCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAA
TGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCG
GCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATT
CGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATG
GTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGAT
CTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGA
TCCC

SI Item 4: pJeM1_backbone (7119 bp) sequence

ACCTTCGGGAGCGCCTGAAGCCCGTTCTGGACGCCCTGGGGCCGTTGAATCGGGATATGCAGGCCAAGGCCGCCGCGATCATCAAGGCCGTGGGCGAAAAGCTG
CTGACGGAACAGCGGGAAGTCCAGCGCCAGAAACAGGCCCAGCGCCAGCAGGAACGCGGGCGCGCACATTTCCCCGAAAAGTGCCACCTGGGATGAATGTCAGC
TACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCA
AGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGA
TCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATG
ACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGA
ATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTG
CTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATC
CGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATC
AGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATG
GTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTG
GCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACT
CTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGG
ACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTG
GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTTTTTATGCATTAATCTTTCTGCGAATTGAGATGACGCCA
CTGGCTGGGCGTCATCCCGGTTTCCCGGGTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCT
GGAGAAGATATTGCGCATGACACACTCTGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCACGAT
GCCTCATCACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTATCCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAAT
CACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCATGCTCAAGCTGA
CCGATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGCCCCCTGCCAGTCAAGATTC
AGCTTCAGACGCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCAGCATGAATGTAAAAGAGATCGCCACGGGT
AATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAAAATCATGTGTATGTTCAGCAAAGACATCTTGCGGATAAC
GGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAAGTTTTAACTGATGCGCCACCGTGGCTACCTCGGCCAGAGAACGAAGTTGATT
ATTCGCAATATGGCGTACAAATACGTTGAGAAGATTCGCGTTATTGCAGAAAGCCATCCCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAA
AGCGTCGAAAAGTGGTTACTGTCGCTGAATCCACAGCGATAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCGCAGG
CGGTTCAGGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCT
CATCGGCAAAATGGTCCTCCAGCCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTG
CATAAACAAGATCTCGCGACTGGCGGTCGAGGGTAAATCATTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAA
CGCGCCAGTGAGACGGATACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATTGGTCA
GACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTAAACACATGAATACCC
GTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCGCCACGGACGCGTTACCAGACGGAA
AAAAATCCACACTATGTAATACGGTCATACTGGCCTCCTGATGTCGTCAACACGGCGAAATAGTAATCACGAGGTCAGGTTCTTACCTTAAATTTTCGACGGAAAACC
ACGTAAAAAACGTCGATTTTTCAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCACATCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTC
CCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTAATGAACAATTCTTAAGAAGGAGATATACATATG
ACCATGATTACGCATCATCATCATCATCATGGATCCAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC
CACAAGTTCAGCGTGTACGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCAC
CCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAA
CTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC
ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGG
ACGAGCTGTACAAGTAAGCTTACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGG
TAATAATTGACGATATGATCATTTATTCTGCCTCCCAGAGCCTGATAAAAACGGTGAATCCGTTAGCGAGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGC
TCCATGCACCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGAGGCGGCTACAGCCGATAGTCTGGAACAGCGCACTTACGGGTTGCTGCGCAACC
CAAGTGCTACCGGCGCGGCAGCGTGACCCGTGTCGGCGGCTCCAACGGCTCGCCATCGTCCAGAAAACACGGCTCATCGGGCATCGGCAGGCGCTGCTGCCCG
CGCCGTTCCCATTCCTCCGTTTCGGTCAAGGCTGGCAGGTCTGGTTCCATGCCCGGAATGCCGGGCTGGCTGGGCGGCTCCTCGCCGGGGCCGGTCGGTAGTTG
CTGCTCGCCCGGATACAGGGTCGGGATGCGGCGCAGGTCGCCATGCCCCAACAGCGATTCGTCCTGGTCGTCGTGATCAACCACCACGGCGGCACTGAACACCG
ACAGGCGCAACTGGTCGCGGGGCTGGCCCCACGCCACGCGGTCATTGACCACGTAGGCCGACACGGTGCCGGGGCCGTTGAGCTTCACGACGGAGATCCAGCG
CTCGGCCACCAAGTCCTTGACTGCGTATTGGACCGTCCGCAAAGAACGTCCGATGAGCTTGGAAAGTGTCTTCTGGCTGACCACCACGGCGTTCTGGTGGCCCATC
TGCGCCACGAGGTGATGCAGCAGCATTGCCGCCGTGGGTTTCCTCGCAATAAGCCCGGCCCACGCCTCATGCGCTTTGCGTTCCGTTTGCACCCAGTGACCGGGC
TTGTTCTTGGCTTGAATGCCGATTTCTCTGGACTGCGTGGCCATGCTTATCTCCATGCGGTAGGGGTGCCGCACGGTTGCGGCACCATGCGCAATCAGCTGCAACT
TTTCGGCAGCGCGACAACAATTATGCGTTGCGTAAAAGTGGCAGTCAATTACAGATTTTCTTTAACCTACGCAATGAGCTATTGCGGGGGGTGCCGCAATGAGCTGT
TGCGTACCCCCCTTTTTTAAGTTGTTGATTTTTAAGTCTTTCGCATTTCGCCCTATATCTAGTTCTTTGGTGCCCAAAGAAGGGCACCCCTGCGGGGTTCCCCCACGC
CTTCGGCGCGGCTCCCCCTCCGGCAAAAAGTGGCCCCTCCGGGGCTTGTTGATCGACTGCGCGGCCTTCGGCCTTGCCCAAGGTGGCGCTGCCCCCTTGGAACC
CCCGCACTCGCCGCCGTGAGGCTCGGGGGGCAGGCGGGCGGGCTTCGCCCTTCGACTGCCCCCACTCGCATAGGCTTGGGTCGTTCCAGGCGCGTCAAGGCCA
AGCCGCTGCGCGGTCGCTGCGCGAGCCTTGACCCGCCTTCCACTTGGTGTCCAACCGGCAAGCGAAGCGCGCAGGCCGCAGGCCGGAGGCTTTTCCCCAGAGAA
AATTAAAAAAATTGATGGGGCAAGGCCGCAGGCCGCGCAGTTGGAGCCGGTGGGTATGTGGTCGAAGGCTGGGTAGCCGGTGGGCAATCCCTGTGGTCAAGCTC
GTGGGCAGGCGCAGCCTGTCCATCAGCTTGTCCAGCAGGGTTGTCCACGGGCCGAGCGAAGCGAGCCAGCCGGTGGCCGCTCGCGGCCATCGTCCACATATCCA
CGGGCTGGCAAGGGAGCGCAGCGACCGCGCAGGGCGAAGCCCGGAGAGCAAGCCCGTAGGGCGCCGCAGCCGCCGTAGGCGGTCACGACTTTGCGAAGCAAA
GTCTAGTGAGTATACTCAAGCATTGAGTGGCCCGCCGGAGGCACCGCCTTGCGCTGCCCCCGTCGAGCCGGTTGGACACCAAAAGGGAGGGGCAGGCATGGCGG
CATACGCGATCATGCGATGCAAGAAGCTGGCGAAAATGGGCAACGTGGCGGCCAGTCTCAAGCACGCCTACCGCGAGCGCGAGACGCCCAACGCTGACGCCAGC
AGGACGCCAGAGAACGAGCACTGGGCGGCCAGCAGCACCGATGAAGCGATGGGCCGACTGCGCGAGTTGCTGCCAGAGAAGCGGCGCAAGGACGCTGTGTTGG
CGGTCGAGTACGTCATGACGGCCAGCCCGGAATGGTGGAAGTCGGCCAGCCAAGAACAGCAGGCGGCGTTCTTCGAGAAGGCGCACAAGTGGCTGGCGGACAAG
TACGGGGCGGATCGCATCGTGACGGCCAGCATCCACCGTGACGAAACCAGCCCGCACATGACCGCGTTCGTGGTGCCGCTGACGCAGGACGGCAGGCTGTCGG
CCAAGGAGTTCATCGGCAACAAAGCGCAGATGACCCGCGACCAGACCACGTTTGCGGCCGCTGTGGCCGATCTAGGGCTGCAACGGGGCATCGAGGGCAGCAAG
GCACGTCACACGCGCATTCAGGCGTTCTACGAGGCCCTGGAGCGGCCACCAGTGGGCCACGTCACCATCAGCCCGCAAGCGGTCGAGCCACGCGCCTATGCACC
GCAGGGATTGGCCGAAAAGCTGGGAATCTCAAAGCGCGTTGAGACGCCGGAAGCCGTGGCCGACCGGCTGACAAAAGCGGTTCGGCAGGGGTATGAGCCTGCCC
TACAGGCCGCCGCAGGAGCGCGTGAGATGCGCAAGAAGGCCGATCAAGCCCAAGAGACGGCCCGAG

SI Item 5: pEX18Gm_backbone (5831 bp) sequence

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGG
GCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAATCGGCATTTTCTTTTGCGTTTTTA
TTTGTTAACTGTTAATTGTCCTTGTTCAAGGATGCTGTCTTTGACAACAGATGTTTTCTTGCCTTTGATGTTCAGCAGGAAGCTAGGCGCAAACGTTGATTGTTTGTCT
GCGTAGAATCCTCTGTTTGTCATATAGCTTGTAATCACGACATTGTTTCCTTTCGCTTGAGGTACAGCGAAGTGTGAGTAAGTAAAGGTTACATCGTTAGGATCAAGA
TCCATTTTTAACACAAGGCCAGTTTTGTTCAGCGGCTTGTATGGGCCAGTTAAAGAATTAGAAACATAACCAAGCATGTAAATATCGTTAGACGTAATGCCGTCAATC
GTCATTTTTGATCCGCGGGAGTCAGTGAACAGATACCATTTGCCGTTCATTTTAAAGACGTTCGCGCGTTCAATTTCATCTGTTACTGTGTTAGATGCAATCAGCGGT
TTCATCACTTTTTTCAGTGTGTAATCATCGTTTAGCTCAATCATACCGAGAGCGCCGTTTGCTAACTCAGCCGTGCGTTTTTTATCGCTTTGCAGAAGTTTTTGACTTT
CTTGACGGAAGAATGATGTGCTTTTGCCATAGTATGCTTTGTTAAATAAAGATTCTTCGCCTTGGTAGCCATCTTCAGTTCCAGTGTTTGCTTCAAATACTAAGTATTT
GTGGCCTTTATCTTCTACGTAGTGAGGATCTCTCAGCGTATGGTTGTCGCCTGAGCTGTAGTTGCCTTCATCGATGAACTGCTGTACATTTTGATACGTTTTTCCGTC
ACCGTCAAAGATTGATTTATAATCCTCTACACCGTTGATGTTCAAAGAGCTGTCTGATGCTGATACGTTAACTTGTGCAGTTGTCAGTGTTTGTTTGCCGTAATGTTTA
CCGGAGAAATCAGTGTAGAATAAACGGATTTTTCCGTCAGATGTAAATGTGGCTGAACCTGACCATTCTTGTGTTTGGTCTTTTAGGATAGAATCATTTGCATCGAATT
TGTCGCTGTCTTTAAAGACGCGGCCAGCGTTTTTCCAGCTGTCAATAGAAGTTTCGCCGACTTTTTGATAGAACATGTAAATCGATGTGTCATCCGCATTTTTAGGAT
CTCCGGCTAATGCAAAGACGATGTGGTAGCCGTGATAGTTTGCGACAGTGCCGTCAGCGTTTTGTAATGGCCAGCTGTCCCAAACGTCCAGGCCTTTTGCAGAAGA
GATATTTTTAATTGTGGACGAATCGAACTCAGGAACTTGATATTTTTCATTTTTTTGCTGTTCAGGGATTTGCAGCATATCATGGCGTGTAATATGGGAAATGCCGTAT
GTTTCCTTATATGGCTTTTGGTTCGTTTCTTTCGCAAACGCTTGAGTTGCGCCTCCTGCCAGCAGTGCGGTAGTAAAGGTTAATACTGTTGCTTGTTTTGCAAACTTTT
TGATGTTCATCGTTCATGTCTCCTTTTTTATGTACTGTGTTAGCGGTCTGCTTCTTCCAGCCCTCCTGTTTGAAGATGGCAAGTTAGTTACGCACAATAAAAAAAGACC
TAAAATATGTAAGGGGTGACGCCAAAGTATACACTTTGCCCCTTTACACATTTTAGGTCTTGCCTGCTTTATCAGTAACAAACCCGCGCGATTTACTTTTCGACCTCATT
CTATTAGACTCTCGTTTGGATTGCAACTGGTCTATTTTCCTCTTTTGTTTGATAGAAAATCATAAAAGGATTTGCAGACTACGGGCCTAAAGAACTAAAAAATCTATCT
GTTTCTTTTCATTCTCTGTATTTTTTATAGTTTCTGTTGCATGGGCATAAAGTTGCCTTTTTAATCACAATTCAGAAAATATCATAATATCTCATTTCACTAAATAATAGTG
AACGGCAGGTATATGTGATGGGTTAAAAAGGATCGATCCTCTAGCTAGAGTCGATCTTCGCCAGCAGGGCGAGGATCGTGGCATCACCGAACCGCGCCGTGCGCG
GGTCGTCGGTGAGCCAGAGTTTCAGCAGGCCGCCCAGGCGGCCCAGGTCGCCATTGATGCGGGCCAGCTCGCGGACGTGCTCATAGTCCACGACGCCCGTGATT
TTGTAGCCCTGGCCGACGGCCAGCAGGTAGGCCGACAGGCTCATGCCGGCCGCCGCCGCCTTTTCCTCAATCGCTCTTCGTTCGTCTGGAAGGCAGTACACCTTG
ATAGGTGGGCTGCCCTTCCTGGTTGGCTTGGTTTCATCAGCCATCCGCTTGCCCTCATCTGTTACGCCGGCGGTAGCCGGCCAGCCTCGCAGAGCAGGATTCCCG
TTGAGCACCGCCAGGTGCGAATAAGGGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCCCGGCTGACGCCGTTGGATACACCAA
GGAAAGTCTACACGAACCCTTTGGCAAAATCCTGTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAATGACCCCGAAGCAGGGTTATGCAGCGGA
AAAGCGCTGCTTCCCTGCTGTTTTGTGGAATATCTACCGACTGGAAACAGGCAAATGCAGGAAATTACTGAACTGAGGGGACAGGCGAGAGACGATGCCAAAGAGC
TACACCGACGAGCTGGCCGAGTGGGTTGAATCCCGCGCGGCCAAGAAGCGCCGGCGTGATGAGGCTGCGGTTGCGTTCCTGGCGGTGAGGGCGGATGTCGATA
TGCGTAAGGAGAAAATACCGCATCAGGCGCATATTTGAATGTATTTAGAAAAATAAACAAAAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCT
GCTTAATTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCGCAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAG
GAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGAC
CCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACCACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGACCG
CTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGC
GATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTA
AAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA
ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCA
CTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAG
AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGA
CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTG
TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGC
AGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG
ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAATCGATGCGAATTGGCCGCGGCGTTGTGAC
AATTTACCGAACAACTCCGCGGCCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGGTCGATATCAAAGTGCATCACTTCTTCCCGTATGC
CCAACTTTGTATAGAGAGCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGCACCAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAG
CGCGGTGGCAATGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATATAGATCTCACTACGCGGCTGCTCAAACCTGGGCAGAACGTAAGCCGCGAG
AGCGCCAACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATGAATGTCTTACTACGGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGT
GGCTACGTCTCCGAACTCACGACCGAAAAGATCAAGAGCAGCCCGCATGGATTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTGAGCCA
CCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGCTGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGCGTAACGCGCTT
GCTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAACAGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGTTCGGTCAAGGTTCTGGA
CCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTACGAACCGAACAGGCTTATGTCAATTCGGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCA
TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCT
AAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

SI Item 6: pQE_lysCpromoter_lacZ_backbone (5968 bp) sequence

TCCGCATGACCCCATCGTTGACAACCGCCCCGCTCACCCTTTATTTATAAATGTACACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAA
ACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAG
CCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTC
AAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACT
CGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTC
GGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGC
AGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCG
CTTTAATGATGATTTCAGCCGCGCTGTACTGGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACG
CAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAACTG
TGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTG
CGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGA
CGATGGTGCAGGATATCCTGCTGATGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGG
CCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTAAC
GCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAA
ATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCC
CTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTT
GGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACG
GCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGC
TGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTG
CACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAG
CCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCT
GGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCA
ATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACA
TTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCA
GATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTC
AAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGC
TCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAG
CGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAGTCAACAGCAACTGAT
GGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGT
ATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTCTGGTGTCAAAAATAAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
TGCTGCCACCGCTGAGCAATACCGCGGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAAC
AGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGA
GAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAA
TCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCC
TGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATTCTAGAGCTGCCTCGCGCGTTTCGGTGATGAC
GGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGG
CGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATG
CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGT
AAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCT
TTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG
AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG
GTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCC
CCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTG
AGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA
AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA
TGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGC
GTATCACGAGGCCCTTTCGTCTTCAC

SI Item 7: DNA fragments ordered by Twist Bioscience or Thermo Fisher, green: NdeI restriction site, cyan: XhoI restriction site, AGATA (yellow) overhangs facilitate the enzymatic digest

Rhizobial canavanine-γ-lyase (Twist)

AGATACATATGCCAAATCCCGACAAGCACAATTTCTCATCCCTCTCATTCGGCACTTTGGCGGTCCATGGTGGCAA
TGAGATTGACAAGACTTCAGGGGCAATCCGTACCCCGATCGTCATGGCAAATTCCTATTCCCTCCCCTACGACCC
CTCAACGATGGACTGGTCGGACACCGAGGAGCCATCCTATACCCGCAACTCGGGACACAACCAGATTTGCCTGC
AGCGCAAATTGGCCGCGATGGAGCGCGGGGAAGACGCCGCTGTTTTCGCAACAGGTGTCGCGGCGTTGCATGC
CGTGTTCTTCACATTTCTGAAATCCGGCGATCATGTGATCGTGGGCGATGTGACATATGAGGCAGTGTGGAGATT
GTTCGCAGAACTCCTTCCCGAGCGCTACAACATCGAAGCTACTTTCGTGGATATGGGGAATATGGACGCCGTCCG
GGCAGCAGTCCGTCCGAAAACCAAACTCATTCACACCGAGACCGTCGCTAATCCCACAACGAAGGTCGCCGATAT
CGCCGCTCTCGTCTCCATTGCTAAGGACGCGGGCGCCTTGCTGTCGGTGGACTCAACCTTTACTCCGCCTCCGTT
CTTTCGCCCGCTGGAGCTTGGCGTTGACCTCGTCATCCATTCCCTGACGAAGTACATCAATGGTCATGGCGATGC
GATGGGTGGCGTCGTGATCGGTTCCAAGAAGCTCGTACATCACATCAAGGCCGATGCACTTGTCGATCTCGGCG
GCACGATTTCCCCGTTCAACGCTTGGCTTATAACCCGCGGCTCGGTCACGCTTCCGCTCAGACTGAAGCAACAAT
TCTCAACAGCCGAGATCGTCGCCAGATATCTTGAAAGCGAGAACCGTCTTGCATTTGTTACGTATCCGGGCCTTG
AGCGTCATGATCAACACCATCTGGCGAAGGCTCAGTTCGGCGGAAAAGGATATGGGGCAATGATGGCGTTCGCT
GTCGATGGCGACCCTGACACCCAGAACCGATTTGTGTCAAACCTCAAGGTTATCACGTCCGCTGTTTCACTTGGT
CACGATGAGACCCTCATTGTCCATGTCGGCGGTGGCGGACGCGGCGGTGCAGAGAGGTATCCATTGAACTTCCA
GAAATATGGGCATCTCCGTCTGTCTATTGGCCTGGAAGACCCCGAGGACCTCATCGCCGACATCAAGAACGCACT
CGACCTTACATTTGCGTGACTCAGATA

CtdA/B34 editing domain-like protein *Clostridium perfrigens* (Twist)

AGATACATATGAAAAAGTTTATTATTGAGGATGATTTTTGGGAGCTATTCCCTAATGCAAAGATAGGTATTATCACTT
GCAATGGCATAAATAATACTATAAAAGATGAGAACCAATACCAGGATATGCTTTCTCAAGGAGAAAAGGAAGCTTT
AACTCACTTGCCTAATGAAGAATTTAGTAGCAATGAAGTTATTAAGGTATGGAGAGATGCATTTAAAAAGTTTAAGA
CTAAAAAAGGAGCTAGATCATCAATTGAAGCTTTATTAAAGAGGGTTTCTACTGGTAAAGGATTAGGAACAATTAAC
CCACTAGTTGATATTTACAACTTTATATCATTAAAGTATGCCATGCCTTGTGGTGGAGAGGATATGGATAAGTTTAT
TGGAGATATAAGATTAACTAAGGCAACAGGAGATGAAAGTTTTATTACCTTAGGATCAGATAAAAGTGAGCCACCC
TATGAGGGAGAAATAGTTTATAAAGATGATGAAGGAGCTATTTGCAGATGCTGGAATTGGCGTGAGTCAGTAAGAA
CTATGTTAACAGAAGACACAAAAAATGCTTTCTTATGCATTGAGTTAGTTGATGAAAAAAGAGAAAAAGATTTTGAA
AATGCCTTAAAAGAGTTATCTCAATTAGTAGAGGAAAACTTAGGTGGAAAATCTGAGATTTCAATACTTCATATTAAT
AATAAGGAAGCTATAATCTGACTCGAGAGATA

Acetyltransferase *Lactobacillus curiae* (Thermo)

AGATACATATGATGGAACTGAGCATCAAACGCTGCGATATTAATGATCTGGATCAGCTGGTTGATATTGCCATTGA
AACCTTTGTTGATACCTTTCTGCCGAACAACAAACAGAAAGATATCGATCAGTATGTGATCAACGCCTTCAAAAGC
AGCAAACTGATGGATGAACTGCATAATCCGGAAAGCCAGTTCTTCTTTATCTATTCCGATGAGGAACTGGCAGGTT
ATCTGAAAGTTAATGTTGGCACCGCACAGACCGAAGATATGGGTCCTGATAGCTTTGAAGTTCAGCGTATTTATGT
GCGCCAGAAATTCAAACGTACCGGTGTTGGTACAGAACTGATGACCCGTGCAATTCAGCTGGCAAAACGTGCAAA
AAAGAAACAGGTTTGGTTAGGCGTGTGGGAAAAAAATGTTAACGCGCAGAAATTCTACGAGAAGTTCGGTTTCATT
AAAACCGGCAGCCATAAATTTCTGATGGGTAGCACCCCGAACCATGATTGGATTATGACCAAACAGCTGTAACTC
GAGAGATA

SI Table 1: Primers and oligos used in this thesis.

internal primer name	*used for*	*5' mod*	*sequence*	*description*
FH138	cloning of pQE_Gua I RS_lacZ		TTATAAATGTACAC-Cctggcaccaagcgtgc	fwd primer to generate the ykkC *P. pelagia* motif for Gibson cloning into the pQE vector between the lysC promoter and lacZ
FH140	cloning of pQE_Gua I RS_lacZ		TGAATCCGTAATCAT-ggtccaactcctactcggtg	revprimer to generate the ykkC *P.pelagia* motif for Gibson cloning into the pQE vector between the lysC promoter and lacZ
FH186	cloning of protein overexpression construct		GCTAGCATGACTGGTGG ACAGCAAATGGG	fwd primer to amplify the linear backbone fragment for Gibson cloning
FH187	cloning of protein overexpression construct		ATGGCTGCCGCGCGGCA C	rev primer to amplify the linear backbone fragment for Gibson cloning
FH188	cloning of protein overexpression construct		tggtgccgcgcgg-cagccatATGTTTAATAC-CGTACTCATCGCTAAC	fwd primer to amplify the linear insert fragment of uca1 for Gibson cloning from *Erwinia rhapontici*
FH189	cloning of protein overexpression construct		tgtccaccag-tcatgctagcTTATTCCAGCC ATAGCAGGG	rev primer to amplify the linear insert fragment of uca1 for Gibson cloning from *Erwinia rhapontici*
FH190	cloning of protein overexpression construct		tggtgccgcgcgg-cagccatATGTTTACAAAAC TGCTGATTGCTAATC	fwd primer to amplify the linear insert fragment of uca2 for Gibson cloning from *Erwinia rhapontici*
FH191	cloning of protein overexpression construct		tgtccaccag-tcatgctagcTTAAAC-GCCCATCACCACC	rev primer to amplify the linear insert fragment of uca2 for Gibson cloning from *Erwinia rhapontici*
FH201	cloning of protein overexpression construct		tggtgccgcgcggcagccatAT-GGCTTTTACCACAGGG	fwd primer to amplify the linear insert fragment of atzF for Gibson cloning from *Erwinia rhapontici*
FH202	cloning of protein overexpression construct		tgtccaccagtcat-gctagcTTAAACATGTGAA-GATACCTCG	rev primer to amplify the linear insert fragment of atzF for Gibson cloning from *Erwinia rhapontici*
FH223	cloning of protein overexpression construct		AGATACATATGCGTCCTG ATAAAAAGAACCTGTCCT CGC	fwd primer to amplify CangL from P. *canavaninivorans* with NdeI restriction site
FH224	cloning of protein overexpression construct		TATCTCTCGAGTTAGTTG AAGGTTTCATCCAACG	rev primer to amplify CangL from *P. canavaninivorans* with XhoI restriction site
FH225	cloning of protein overexpression construct		AGATACATATGAATAATT TGAGTCTCTCGCTCG	fwd primer to amplify AC from *P. canavaninivorans* with NdeI restriction site
FH226	cloning of protein overexpression construct		TATCTCTCGAGTCAGAAG GATGTTGTTGTCTTGC	rev primer to amplify AC from *P. canavaninivorans* with XhoI restriction site
FH251	16S rRNA gene sequencing		AGAGTTTGATCCTGGCT CA	16S_27
FH252	16S rRNA gene sequencing		CGGCTACCTTGTTACGA C	16S_1497
FH253	16S rRNA gene sequencing		CCGTCAATTCMTTTGAGT TT	16S_907

FH256	cloning of protein overexpression construct	agatacatATGCTGTCAG-TACTGCCGTT	fwd primer to amplify B3/4 of *P. canavaninivorans* and add restriction site NdeI to clone into pET28a
FH257	cloning of protein overexpression construct	tatctctcgagTCAGT-GAGCGCCAGGAC	rev primer to amplify B3/4 of *P. canavaninivorans* and add restriction site XhoI to clone into pET28a
FH270	cloning of protein overexpression construct	CTCGAGCACCACCACCA C	fwd primer to amplify backbone for cloning via gibson into pET28a
FH271 _1	cloning of protein overexpression construct	ATGGCCCTGAAAATAAAA GATTCTCAC	rev primer to amplify backbone for cloning via gibson into pET28a
FH279	cloning of protein overexpression construct	tggtcccactgaaggctt-gaCTCGAGCACCACCAC-CAC	fwd primer to amplify pet28a backbone for Gibson cloning of ammonium aspartate lyase *P.canavaninivorans*
FH280	cloning of protein overexpression construct	ATGGCCCTGAAAATAAA GATTCTCAC	rev primer to amplify pet28a backbone for Gibson cloning of ammonium aspartate lyase *P.canavaninivorans*
FH281	cloning of protein overexpression construct	atctttattttcagggccatAT-GTCCTCCGCTGCATCATT C	fwd primer to amplify the AAL insert for Gibson cloning of ammonium aspartate lyase *P.canavaninivorans*
FH282	cloning of protein overexpression construct	TCAAGCCTTCAGTGGGA C	rev primer to amplify the AAL insert for Gibson cloning of ammonium aspartate lyase *P.canavaninivorans*
FH297	cloning of protein overexpression construct	atctttattttcagggccatAT-GAAAGACACCATTCGCC AG	fwd primer to amplify ArgS from *P. canavaninivorans* and clone it by Gibson into pET28a
FH298	cloning of protein overexpression construct	tggtggtggtggtgctcgagTTA-CATGCGCTCCAGCGTTT C	rev primer to amplify ArgS from *P. canavaninivorans* and clone it by Gibson into pET28a
FH299	IVT tRNA	TAATACGACTCACTATAG GTCTCAGTAGCTCAATTG GATAGAGCATCCCCCTC CTAAGGG	fwd primer for IVT of tRNA 1
FH30	sequencing primer	GCCTCAGGCATTTGAGA AGC	sequencing primer to verify successful cloning into pJeM1 vector
FH300	IVT tRNA	TGGTGTCCCAGGGCAGG TTCGAACTGCCAACCTTC CCCTTAGGAGGGGGATG CTCTATCC	rev primer for IVT of tRNA 1
FH301	IVT tRNA	TAATACGACTCACTATAG GCACCAGTAGCTCAGCT GGATAGAGTACTGCCCT CCGAAGGC	fwd primer for IVT of tRNA 22
FH302	IVT tRNA	TGGTGCACCAGGCGGGA TTCGAACCCACGACCCC TGCCTTCGGAGGGCAGT ACTCTATCC	rev primer for IVT of tRNA 2
FH303	IVT tRNA	TAATACGACTCACTATAG GCGCCCGTAGCTCAGCT GGATAGAGCATCCGCCT TCTAAGCG	fwd primer for IVT of tRNA 3
FH304	IVT tRNA	TGGCGCACCCGGCAGGA CTCGAACCTGCGACCAT CCGCTTAGAAGGCGGAT GCTCTATCC	rev primer for IVT of tRNA 3

FH308	homologues recomb.		CTTTCTCGACATGGAACAGC	fwd primer binding on the genome upstream of the B34 homologous region (see primer 3, Paper: A rapid seamless method for gene knockout in *P. aeruginosa*)
FH309	homologues recomb.		CGGTGCTCAACAAACGC	rev primer binding on the genome upstream of the B34 homologous region (see primer 4, Paper: A rapid seamless method for gene knockout in *P.aeruginosa*)
FH338	homologues recomb.		GCCAGTGCCAAGCTTGCATG	fwd primer to amplify the backbone pEX18Gm for insertion of HR1 and HR2 via gibson
FH339	homologues recomb.		CGTCGTTTTACAACGTCGTGAC	rev primer to amplify the backbone pEX18Gm for insertion of HR1 and HR2 via gibson
FH340	homologues recomb.		cacgacgttg-taaaacgacgGGCGAGCAGGCGTGCAGT	fwd primer to generate the HR1 insert of the B3/4 homologous region, cloned with fragment HR2 into the pEX18Gm vector
FH341	homologues recomb.		aacctgttgtCAAGGTT-CTCCAGCGATCCAAATTGG	rev primer to generate the HR1 insert of the B3/4 homologous region, cloned with fragment HR2 into the pEX18Gm vector
FH342	homologues recomb.		gagaaccttgACAACAGGTT-GTCCCGCAG	fwd primer to generate the HR2 insert of the B3/4 homologous region, cloned with fragment HR1 into the pEX18Gm vector
FH343	homologues recomb.		catgcaagctt-ggcactggcGTTAAAAGCCTTGATCGCCG	rev primer to generate the HR2 insert of the B3/4 homologous region, cloned with fragment HR1 into the pEX18Gm vector
FH344	IVT tRNA	5' [mU][mG]	UGGCGCACCCGGCAGGACTCGAACCTGCGACCATCCGCTTAGAAGGCGGATGCTCTATCC	5' modified primer FH304 to reduce the n+1 product of IVT of tRNA
FH350	cloning of protein overexpression construct	5' phos	GCGTACATCAACGGCCAC	fwd primer to mutate K213A in the CanγL by whole plasmid PCR and quick ligation
FH351	cloning of protein overexpression construct		GGTCAACGAGTGCACGAC	rev primer to mutate K213A in the CanγL by whole plasmid PCR and quick ligation
FH362	homologues recomb.		cacgacgttg-taaaacgacgGGCGGCCAAGCGCGCTGA	fwd primer to generate the HR1 insert of the CangL homologous region, cloned with fragment HR2 into the pEX18Gm vector
FH363	homologues recomb.		atcagccaagGGGTATTT-CCTCAGAAGGATGTTGTTGTCTTGCCG	rev primer to generate the HR1 insert of the CangL homologous region, cloned with fragment HR2 into the pEX18Gm vector
FH364	homologues recomb.		ggaaatacccCTTGGCTGAT-CAGGTCGC	fwd primer to generate the HR2 insert of the CangL homologous region, cloned with fragment HR1 into the pEX18Gm vector
FH365	homologues recomb.		catgcaagctt-ggcactggcAACGATATTA-CGAGCAAACCG	rev primer to generate the HR2 insert of the CangL homologous region, cloned with fragment HR1 into the pEX18Gm vector

FH368	homologues recomb.		CGGCCCCATGATCAGC	fwd primer binding on the genome upstream of the CangL homologous region (see primer 3, Paper: A rapid seamless method for gene knockout in P.aeruginosa.)
FH369	homologues recomb.		CGCGGTTTTACGAATATGGC	rev primer binding on the genome upstream of the CangL homologous region (see primer 4, Paper: A rapid seamless method for gene knockout in P.aeruginosa.)
FH370	tRNA enrichment in *P. canavaninivorans*	5' phos	AGCGGATGGTCGCAGGTTCGAGTCCTGCCGGGTGCGCCAATGCTTACCGGTTTATTGAC	generation of an tRNA21 overexpression plasmid by introducing the sequence via two whole plasmid PCRs into vector pJeM1: fwd primer of PCR1
FH371	tRNA enrichment in *P. canavaninivorans*		TAGAAGGCGGATGCTCTATCCAGCTGAGCTACGGGCGCCTACGACCAGTCTAAAAAGCG	generation of an tRNA21 overexpression plasmid by introducing the sequence via two whole plasmid PCRs into vector pJeM1: rev primer of PCR1
FH372	tRNA enrichment in *P. canavaninivorans*	5' phos	ACGGTCTGGCGCCTTTGCTTTAAACATGCTTACCGGTTTATTGAC	generation of an tRNA21 overexpression plasmid by introducing the sequence via two whole plasmid PCRs into vector pJeM1: fwd primer of PCR2
FH373	tRNA enrichment in *P. canavaninivorans*		TAAATCTGGCGCCTTTCTAGAATATGGCGCACCCGGCAGGACTCG	generation of an tRNA21 overexpression plasmid by introducing the sequence via two whole plasmid PCRs into vector pJeM1: rev primer of PCR2
FH082	sequencing primer		GGCGCTTTTTAGACTGGTCG	sequencing primer to verify successful cloning into pJeM1 vector
HB010	cloning of protein overexpression construct		CTCGAGCACCACCACCAC	fwd primer to amplify the linear backbone fragment for Gibson cloning
HB011	cloning of protein overexpression construct		CATATGGCCCTGAAAATAAAGATTCTCAC	rev primer to amplify the linear backbone fragment for Gibson cloning
HB014	cloning of protein overexpression construct		tttattttcagggccatatgATGACTGAATACAAACTGGCTCTAG	fwd primer to amplify the linear insert fragment of HD for Gibson cloning from *P. canavaninivorans*
HB015	cloning of protein overexpression construct		tggtggtggtggtgctcgagTCACAGGGCTTGCTCCTG	rev primer to amplify the linear insert fragment of HD for Gibson cloning from *P. canavaninivorans*
HB016	cloning of protein overexpression construct		tttattttcagggccatatgGTGAACGTATCGCGCCTG	fwd primer to amplify the linear insert fragment of AH for Gibson cloning from *P. canavaninivorans*
HB017	cloning of protein overexpression construct		tggtggtggtggtgctcgagTCACCCCAACGTATTCAAAC	rev primer to amplify the linear insert fragment of AH for Gibson cloning from *P. canavaninivorans*
SP04	sequencing primer		TCCCCATCGGTGATGTC	sequencing primer to verify successful cloning into pET28a vector
SP10	sequencing primer		CTAGTTATTGCTCAGCGG	sequencing primer to verify successful cloning into pET28a vector
SP13	sequencing primer		CGGATAACAATTTCACACAG	sequencing primer to verify successful cloning into pEX18Gm vector
SP26	sequencing primer		GGCGATTAAGTTGGGTAACG	sequencing primer to verify successful cloning into pEX18Gm vector
SP42	sequencing primer		TCCGGATATAGTTCCTCCT	sequencing primer to verify successful cloning into pET28a vector

SI Table 2: Functions-and-PFAM. “Pathway(s)” which of the three guanidine-relevant pathways the domain belongs to. Domains unique to one pathway are listed before ambiguous domains. “Domain name”: our name for the domain. Note: current databases do not distinguish between Gdx proteins and homologous proteins that function as multidrug exporters. We assume that all proteins of this family that are regulated by guanidine-riboswitches function as Gdx proteins. Similarly, current databases to not distinguish between guanidine carboxylase proteins and possible urea carboxylase proteins. We assumed that all proteins in this family are specific to guanidine if their genes are regulated by a guanidine riboswitch. “Unique to pathway”: indicates whether the domain is unique to one of the three proposed pathways, or whether it spans multiple pathways. “Pfam accession(s)”: the entries in Pfam we used to define the domain. Accessions beginning with “PF” refer to families. Accessions beginning with “CL” refer to clans, which correspond to multiple families. Proteins that match any listed family or clans, according to Pfam’s parameters, were assumed to correspond to the domain.

Pathway(s)	Domain name	Unique to pathway	Pfam accession
Guanidine Export (Gdx^P)	Guanidine exporter (Gdx) *	Yes	CL0184
Guanidine carboxylation (Gca^P)	Guanidine carboxylase, N-terminus	Yes	PF00289
Guanidine carboxylation (Gca^P)	Guanidine carboxylase, C-terminus	Yes	PF02785
Guanidine carboxylation (Gca^P)	Carboxyguanidine deiminase	Yes	PF09347
Guanidine carboxylation (Gca^P)	Allophanate hydrolase 1	Yes	PF02626
Guanidine carboxylation (Gca^P)	Allophanate hydrolase 2	Yes	PF02682
Guanidine carboxylation (Gca^P)	PotE	Yes	PF13520
Guanidine carboxylation (Gca^P)	Btu periplasmic domain	Yes	CL0043
Guanidine carboxylation (Gca^P)	Ecf periplasmic domain	Yes	CL0315
Agmatinase related ($Agmat^P$)	Arginase/Agmatinase-like	Yes	PF00491
Agmatinase related ($Agmat^P$)	HypA	Yes	PF01155
Agmatinase related ($Agmat^P$)	HypB	Yes	PF02492
Gca^P, $Agmat^P$	ABC transporter, periplasmic domain	No	CL0177
Gca^P, $Agmat^P$	ABC transporter, permease domain	No	PF00528
Gca^P, $Agmat^P$	ABC transporter, ATPase domain	No	PF00005

SI Table 3: Crystal trials parameters for GNAT *L. curiae.*

	GNAT
PDB code	n/a
Space Group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	60.68 37.08 83.650
α, β, γ (°)	90.00 97.67 90.00
Molecules in asymmetric unit	2
X-ray data	
X-ray source	SLS PX1
Detector	EIGER 16M X
Wavelength (Å)	1.00002
Resolution (Å)	36.48 - 2.0 (2.072 - 2.0)
Unique Reflections	24776 (2410)
Multiplicity	6.9 (7.2)
Completeness, %	97.46 (95.18)
<I/σ (I)>	10.60 (2.70)
R_{meas}, %	0.1593 (0.8667)
$CC_{1/2,}$ %	0.995 (0.861)
Refinement	

	GNAT
Reflections work / R-free sets	24764 (2409)
Protein residues / waters	353 / 158
Ligands	3 x sodium tartrate; 5 x glycerol
$R_{factor}/R_{free,}$ %	18/24
RMSD bond length (Å) / angle (°)	0.013/1.5
Ramachandran **Favored / allowed/ outliers %**	99.42/0.58/0

Abbreviations

AAL	aspartate ammonia lyase
aaRS	aminoacyl-tRNA synthetase
ABC	ATP binding cassette
AC/AH	aldehyde dehydrogenase
ANI	average nucleotide identity
ArgRS	arginine tRNA syntethase
AT	aminotransferase
ATP	adenosine triphosphate
B3/4 pro-tein	B3/4 editing domain-like protein
BLAST	basic local alignment search tool
cAMP	cyclic adenosine monophosphate
CanγL	canavanine-γ-lyase
CAP	catabolite activator protein
CCR	carbon catabolite repression
CgdAB	carboxyguanidine deiminase
CγL	cystathionine-γ-lyase
C-source	carbon source
CtdA	canavanyl-tRNA^{Arg} deacylase
dDDH	digital DNA-DNA hybridization
DNA	deoxyribonucleic acid
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
E. coli	*Escherichia coli*
EF	elongation factor
EFE	ethylene-forming enzyme
GBDP	genome BLAST distance phylogeny
Gca	guanidine carboxylases
Gdx	guanidine exporter
Gdx	guanidine exporters
GNAT	GCN5-related N-acetyltransferase protein
GTP	guanosine triphosphate
GuaRS	guanidine riboswitch
HD	homoserine dehydrogenase
MγL	methionine-γ-lyase
mRNA	messenger RNA
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NMR	nuclear magnetic resonance
N-source	nitrogen source
ONPG	*o*-Nitrophenyl-β-D-galactopyranosid
P. canavaninivorans	*Pseudomonas canavaninivorans*
PCR	polymerase chain reaction
PLP	pyridoxal phosphate
RBS	ribosome binding site
RNA	ribonucleic acid
RNAP	RNA polymerase
SMR	small multidrug resistance

tRNA	transfer RNA
TYGS	type strain genome server
Uca	urea carboxylase
Ucaa1	urea carboxylase-associated gene 1
Ucca2	urea carboxylase-associated gene 2
UTR	untranslated region

Danksagung

An erster Stelle möchte ich mich bei meinem Doktorvater Jörg Hartig für die Möglichkeit bedanken, in seiner Arbeitsgruppe zu promovieren. Die freie Hand die er mir während meiner Promotion zur Bearbeitung einer Vielzahl von Themen und Projekten ließ, hat es mir ermöglicht, gerne zur Arbeit zu kommen. Sein Vertrauen in meine Arbeit gepaart mit einem offenen Ohr, Tipps und konstruktiver Kritik, waren die perfekte Mischung für die Zeit meiner Promotion.

Danke auch an die weiteren Mitglieder meines Prüfungskomitees, Valentin Wittmann für den Vorsitz und David Schleheck für das Begutachten meiner Arbeit.

Desweiteren möchte ich David Schleheck und Florian Stengel danken, die mich im Rahmen der Graduiertenschule KoRS-CB wissenschaftlich betreut haben.

Der Graduiertenschule KoRS-CB gilt mein Dank nicht nur für die finanzielle Unterstützung für den Besuch einer Konferenz, sondern auch für die Organisation diverser Kurse, Events und Retreats. Der Austausch mit anderen Wissenschaftler*innen und die Vernetzung mit anderen Promovierenden an der Universität Konstanz waren (nicht nur wissenschaftlich) eine Bereicherung.

Weiterhin möchte ich unseren Kollaborationspartner*innen der verschiedenen Projekte für die wertvolle und zuverlässige Zusammenarbeit danken: Zasha Weinberg (Universität Leipzig), Jennifer Fleming und Olga Mayans (Universität Konstanz).

Ein ganz großer Dank geht an die gesamte Arbeitsgruppe Hartig. Danke an Malte und Dietmar die immer für Diskussionen und Austausch bereitstanden. Danke an Astrid, die durch ihr Organisationstalent für Ordnung und Struktur sorgte, und stehts bereit war, mir beim Klonieren unter die Arme zu greifen. Danke an Dmitry für Massemessungen und Pufferzubereitung. Und natürlich danke an alle anderen Mitglieder der AG, die nicht nur fachlich unterstützten, sondern darüber hinaus mit Kaffeepausen, Biergartenbesuche, AG Ausflügen und Abenden am See auch für Freizeiterholung sorgten. Ich hatte das große Glück im Labor nicht nur mit Kollegen*innen, sondern vorallem mit Freunden arbeiten zu können.

Zuletzt ein großer Dank an meine Familie und meine Freunde, die mich auf meinem gesamten (Ausbildungs-)Weg immer unterstützt haben. Och tack så mycket Frederike.