UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Biosynthesis and translational control of terminal-alkyne amino acids

Permalink
https://escholarship.org/uc/item/34s446zk

Author
Marchand Benmaman, Jorge A

Publication Date
2019

Peer reviewed|Thesis/dissertation

Biosynthesis and translational control of terminal-alkyne amino acids

By

Jorge A. Marchand Benmaman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michelle C. Y. Chang, Co-chair
Professor Wenjun Zhang, Co-chair
Professor David V. Schaffer
Professor Daniel K. Nomura

Summer 2019

Biosynthesis and Translational Control of Terminal-Alkyne Amino Acids

© 2019

by Jorge A. Marchand Benmaman

Abstract

Biosynthesis and Translational Control of Terminal-Alkyne Amino Acids

by

Jorge A. Marchand Benmaman

Doctor of Philosophy in Chemical and Biomolecular Engineering

University of California, Berkeley

Professor Michelle C. Y. Chang, Co-chair

Professor Wenjun Zhang, Co-chair

Living systems are remarkably efficient in their ability to build complex molecular function, using a relatively small set of chemical functional groups to drive the wide range of life processes. In contrast, synthetic compounds access a much broader scope of structural and reaction diversity to tune desired chemical properties. However, unusual functional groups can be found in Nature. Many pathways for secondary metabolism produce compounds that could be useful for chemical applications if their biosynthesis could be repurposed. Here, we report the discovery and characterization of a unique pathway to produce a terminal alkyne-containing amino acid, β-ethynylserine, by the microbe *Streptomyces cattleya* through an unexpected reaction sequence that includes radical halogenation and oxidative C-C bond cleavage, followed by a desaturation that proceeds through a putative allene intermediate. We also report progress in porting the *bes* pathway into heterologous host, particularly *E. coli*. Furthermore, we present preliminary work on the discovery of a protein who's function is to prevent translation of β-ethynylserine by acting as a tRNA proofreading enzyme. This pathway, and the auxiliary proteins related to control of translation of non-standard amino acids, offers the potential to genetically encode the *de novo* production of halo-, alkene-, and alkyne-labeled proteins and peptides from glucose to provide bioorthogonal functionality for downstream applications.

Table of Contents

Chapter 1: Introduction

Chapter 2: Discovery of the *bes* biosynthetic gene cluster

Chapter 3: Development and optimization of a pathway for production of halo, alkene, and alkyne amino acids

Chapter 4: Translational-machinery control of β-ethynylserine by a trans-editing tRNA deacylase

Appendices

List of Figures, Schemes, and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

αKG	α-ketoglutarate
ATP	adenosine-5´-triphosphate
ACP	*acyl-carrier protein*
ARS	aminoacyl-tRNA synthetase
βes	β-ethynylserine
CuAAC	*copper catalyzed azide alkyne cycloaddition*
bp	base pairs
d	day
Da	daltons
DTT	DL-dithiothreitol
F-Thr	4-fluorothreonine
EIC	extraction ion chromatogram
EDG	electron donating group
EWG	electron withdrawing group
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
HPLC	high pressure liquid chromatography
Glu-Pra	γ-glutamyl-propargylglycine
Glu-βes	γ-glutamyl-β-ethynylserine
Ggt	γ-glutamyl-transpeptidase
kDa	kilodaltons
LB	lysogeny broth
LC	liquid chromatography
LC/MS	liquid chromatography mass spectrometry
LC/MS-MS	liquid chromatography tandem mass spectrometry
Leu	leucine
min	minute
MWCO	molecular weight cutoff
MRM	multiple reaction monitoring
Met	methionine
NADH	nicotinamide adenine dinucleotide
NiNTA	nickel nitrilotriacetic acid
PMSF	phenylmethanesulfonyl fluoride
Pra	propargylglycine
QQQ	triple quadrupole tandem mass spectrometer
QTOF	quadrupole-time of flight mass spectrometer
TCO	trans-cyclooctene

tRNA	transfer RNA
aa-tRNA	aminoacyl-tRNA
tRNA^{Thr}	threonine tRNA
Thr	threonine
TRS	threonyl-tRNA synthetase

Acknowledgments

I would like to acknowledge first the many members of the Michelle Chang Lab that contributed significantly to my training during my graduate career. Special thanks to the endless patience of my earliest mentors in lab: Dr. Joe Gallager, Dr. Brooks Bond-Watts, Dr. Jonathan McMurry, and Dr. Vivian Yu. I was fully dependent on your time and effort in learning synthetic biology and will always be in debt. Other influential mentors include Dr. Chia-I Lin, Dr. Omer Ad and Dr. Mike Blaisse. Thank you all for always having time to engage in deep conversations about chemistry. I also want to thank Monica Neugebauer for joining me on my project during some of the earliest stages when uncertainty was at its peak. Your help was instrumental to the success of my project, and I could not have done it without you. Finally, I want to thank the many members of the Michelle Chang Lab that have joined after me. Sasilada Siriungruang, Dr. Hongjun Dong, Kersh Theva, Ed Koleski, Jason Fang, Doug Millar, Mira Liu, Eli Kissman. Though it may not seem like it, I learn something (either scientific or personal),from everyone and you all have contributed to my own personal development in one way or another.

There are a few mentors who have supported me a long the way. Dr. Pei-Hsun Wu pushed me harder than anyone as an undergraduate and really solidified in me the motivation to pursue an academic career. Prof. Denis Wirtz has also been an extremely influential supporter and mentor, and I thank him for never wavering and always believing. Finally I would like to thank my latest mentor, Prof. Michelle Chang, for her unconditional support. I joined the lab without any experience in synthetic biology or chemistry, but she still entertained my ideas and helped guide my project and training in the best possible direction.

Outside of graduate school, I want to thank my family and friends. My family may never understand what I during graduate school but that is okay, they always made sure to be supportive. I want to single out Dr. Jonathan Bachman for being a lasting friend through the 6 years of my PhD, and supporting me like no one else should need to.

Most importantly, I want to acknowledge/thank/celebrate the largest source of stability and happiness in my life – Erica Fuhrmeister. You followed me here and I do not believe either one of us knew what we were getting into. No chance I could have been successful without your love and support. Nothing makes me happier than seeing you also succeed during your PhD.

Chapter 1: *Introduction*

Introduction

The ability to repurpose biomolecules has lead to technology that forms the basis of a wide range of industries from biotech, to food manufacturing, to bio-renewable fuels and plastics [1–8]. Despite the relatively limited set of building blocks for constructing biomolecules, cells are able to achieve extraordinary functional complexity across a large range of length scales. Indeed, some of the most abundant molecules of life (carbohydrates, proteins, and nucleic acids) are constructed from as few as 4 monomeric units, in the case of DNA or RNA, and 20 monomeric units in the case of proteins. Biomolecules possess unique and complex chemical properties that emerge from their ability to have their sequence encoded precisely by the genetic and biosynthetic machinery. In contrast to how living systems construct molecules, synthetic molecules can utilize an almost unlimited range of functional groups and building blocks, yet remain limited by our technical ability to construct chiral and sequence-specific macromolecules at scale.

In thinking about how to build biomolecules with tailored set of physical, structural, and chemical properties we can imagine designing by modifying the monomeric units of biomolecules. From the perspective of a chemist, however, selectively targeting biomolecules for modification is challenging due to the repetitive nature of the monomers used in biomolecules. Indeed, most organisms utilize the same sets of building blocks (presenting the same kind of functional groups) from small molecule metabolites to macromolecular biomolecules. More so, living systems, and their constitutive natural molecules, face physical constrains that make them difficult to manipulate using chemistry. The solvent of life is water, not organic solvents, which effectively eliminates a large portion of organic chemistry reactions that could be used for reacting with natural systems. Additionally, life has evolved to be stable over a narrow range of temperatures, optimally 5-50 °C, eliminating the possibility of using extreme temperatures for reactivity. Lastly, duration of reaction is an important factor to consider, as many biomolecules are unstable over longer time scales. As such, chemists have taken the approach of devising a framework of reactions that are inherently compatible with modifying living systems and biomolecules. These reactions are known as bioorthogonal reactions.

1.1. Bioorthogonal Reactions

Introduction. The defining feature of bioorthogonal reactions is that they can be used to modify living systems with minimal interference with the biochemistry of life [9–13]. In practice, these reactions need to be selective and react with only the target molecule of interest. The reactions need to be able to be performed in water, as living systems and many biomolecules do not tolerate high levels of organic solvents. They also require rapid reaction rates – and to reach completion or near completion in the time scale of minutes to hours. Generally, these reactions should be limited to a narrow range of temperatures – from room temperature (20 °C) to physiological temperatures (37-52 °C). Furthermore, the pH at which bioorthogonal reactions occur should also be neutral or near neutral. Lastly, it is preferable these reactions should not release byproducts to avoid the accumulation of any molecule that could lead to toxicity or perturbations of the system being studied. While under their strictest definition bioorthogonal reactions can safely be used to modify intact living systems, they can also be used to describe reactions that more generally modify biomolecules *ex-vivo*. To date, many reactions have been developed that fall under the umbrella of being 'bioorthogonal', each having its own set of

advantages and disadvantages. While many reactions have been developed that possess these properties [10,11,13,14], the three reactions relevant to the this work (a terminal alkyne amino acid biosynthetic pathway) are briefly described below.

Cu(I) Catalyzed Azide-Alkyne Cycloaddition. Though Cu(I) can be toxic to living systems, the Cu(I) Catalyzed Azide-Alkyne Cycloaddition (CuAAC, or 'Cu-click') is one of the most widely used reactions for modifying biomolecules (*Figure 1.1*) [15–20]. Aside from toxicity, the CuAAC meets many of the desirable properties outlined previously – fast ($k > 10^1$ M^{-1}s^{-1}), works in water, at room temperature, in neutral pH, and highly selective. In the CuAAC, a terminal alkyne and an azide undergo a Cu(I)-mediated cycloaddition to produce a triazole [20]. The high selectivity of this reaction involving living systems is derived from the rarity of these functional groups in Nature. Terminal alkynes are scarce, notably absent in humans, and only found in a few organisms [21]. Meanwhile, there have been no reports to date of azides being made by a living system. Nonetheless, selectivity is generally conferred by introducing either an alkyne or an azide into the molecule of interest, which can then be modified by a large selection of commercially available azides or alkynes. Since first being described in 2001, the CuAAC has seen a wide range of applications from attachment of fluorophores to creation of antibody-drug conjugates [22–27]. Though faster bioorthogonal reactions have been developed that avoid Cu cytotoxicity, the CuAAC continues to be used for its versatility [28]. The CuAAC reaction has also served as the inspiration for "Cu-free Click" reactions that utilize a strained alkyne substrate to provide the driving force for tetrazole formation in the absence of Cu [29,30].
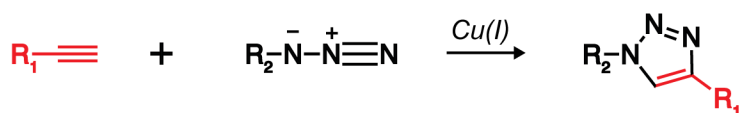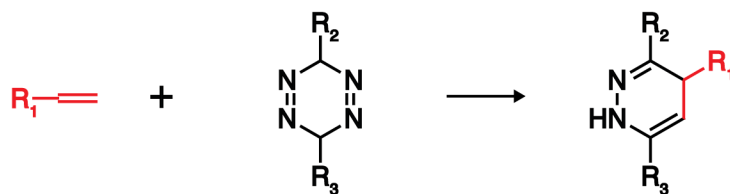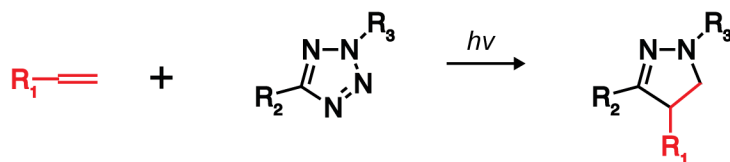
Figure 1.1. Scope of bioorthogonal reactions covered in this work. The CuAAC between a terminal alkyne and an azide, the tetrazine ligation between a terminal alkene and a tetrazole, and the tetrazole photoclick between a terminal alkene and a tetrazole.

Tetrazine Ligation. The tetrazine ligation is a named inverse-electron demand Diels-Alder reaction that occurs between an activated alkene and a di-substituted tetrazine (*Figure 1.1*) [31].

No catalyst is required for performing tetrazine-ligations, which makes this reaction compatible with performing chemistry in living cells [32]. The rate of this reaction can vary widely and depends almost exclusively on the chosen electrophile (tetrazine) and dieneophile (alkene). Tetrazines with strong EWG will react faster than tetrazines with EDG. As for the dienophile, strain, sterics, and stereochemistry all play an important role in determining rate of reaction. One of the most reactive dienophiles, *trans*-cyclooctenes (TCO) are at least 6 orders of magnitude more reactive with tetrazines than terminal alkenes (TCO $k > 800$ $M^{-1}s^{-1}$, terminal alkene $k < 0.001$ $M^{-1}s^{-1}$) [32–34]. Nonetheless, in the context of this work we are interested in using terminal alkenes as a dienophile for tetrazine ligations since terminal alkenes can be biosynthesized. Like azides, tetrazines are also not found in Nature.

Tetrazole Photoclick. While terminal alkenes are a poor substrate (kinetically) in the tetrazine ligation reaction, they can react much more quickly ($k > 0.05$ $M^{-1}s^{-1}$) with tetrazoles under UV light (*Figure 1.1*) [35–37]. In the tetrazole photoclick, UV light leads to a photoinduced cycloreversion (releasing N_2), and generating a reactive 1,3-nitrile imine intermediate. The nitrile imine quickly undergoes a 1,3 dipolar cycloaddition with terminal alkenes, yielding a pyrazoline. Though quick, the requirement of UV to activate the tetrazole limits applications of this reactions to conditions where UV light can penetrate and that are insensitive to UV-induced toxicity. As was the case for azides and tetrazines, tetrazoles are not found to occur naturally in living systems.

1.2. Biosynthesis of acetylenic natural products

Introduction. Acetylenic natural products are a large class of molecules that possess a wide variety of bioactive properties – antitumor, antibiotic, antifungal, and herbicidal [21,38–40]. By definition, acetylenic natural products contain at least one alkyne group, but can also exist as conjugated poly-ynes. Many plants, fungi, archaea, and bacteria posses enzymatic machinery for making acetylenic compounds from simple, primary metabolite precursors but little is known about how they are made. Two early models for formation of acetylenes were 1) desaturation and 2) decarboxylative enol-elimination. Though the latter has been proven to be plausible with synthetic compounds, only the former has been observed in living systems. Indeed, the most well studied examples of biosynthetic pathways for making acetylenic metabolites are through fatty acid desaturation. Here, a fatty acid desaturase use an iron center to catalyze the O_2-dependent oxidation of an existing acyl-alkene-ACP. Even though fatty acid desaturases are well studied in the context of installing internal alkynes on acyl chains, there are few examples of fatty acid desaturases that can form terminal alkynes.

Terminal alkyne formation in Jamaicamide B. One of the few known fatty acid desaturases that has been shown to make a terminal alkyne is JamB [41]. In the biosynthesis of Jamaicamide B, JamB catalyzes the consecutive ω-desaturation of a hexanoyl-ACP to first produce the ω-
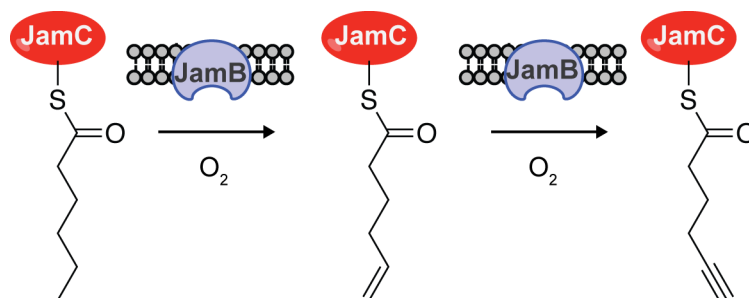
Figure 1.2. JamB is a membrane-bound fatty acid desaturase involved in the installation of terminal alkyne on hexanoyl-ACP during Jamaicamide biosynthesis.

hexenoyl-ACP which is then subsequently desaturatated to the ω-hexynoyl-ACP (*Figure 1.2*) [42]. The ω-hexynoyl-ACP is then passed along additional units of polyketide synthase, and undergoes a series of condensations, reductions, and oxidation to form the final product. Since its discovery, genome mining has revealed other natural product gene clusters that contain JamB homologues capable of forming various chain-length terminal-alkyne acyl-ACPs (from 6-18 carbons in length) [43].

Terminal alkyne amino acids. Though largely of unknown biosynthetic origin, there have been almost a dozen terminal alkyne amino acids that have been observed to be produced by fungi and bacteria (*Figure 1.3, Table 1.1*) [44–50]. The stereochemistry of the amine in these amino acids can be either (*S*), similar to proteinogenic amino acids, or (*R*). Amino acids **1-3** are the only ones that have been observed in *Streptomyces* spp., while **1**, **4-11** have been observed in various species of fungi. **1** (2(S)-3(S)-2-amino-3-hydroxypent-4-ynoic acid, or β-ethnyl-L-serine, βes) was independently observed to be produced by both fungi, *Sclerotium rolfsii*, and bacteria *Sterptomyces cattleya* [47,48]. **2** (2(*S*)-2-aminopent-4-ynoic, or L-propargylgycine, Pra) was discovered in a strain of *Streptomyces* held in a private collection and has never had its genome sequenced [50]. In our own work, we show that *S. cattleya* is capable of making both **1** and **2**. Lastly, **3** (2(*S*)-2-aminobut-3-ynoic acid, or L-ethynylglycine) was reported made by *Streptomyces catenulae*, though in our work we have only observed *S. catenulae* producing **2** [45,46]. It remains a possibility that **3** was structurally misannotated upon its discovery. Both *S. cattleya* and *S. catenulae* have had their genome sequenced, making the discovery for the biosynthetic pathways of **1** and **2** an attractive target for investigation. Though the function of most of these amino acids is currently unknown, it is speculated that they could function as cytotoxic metabolites used in host defense.

1.3. Incorporation of synthetic functional groups into proteins

Introduction. Aminoacyl-tRNA synthetases (ARS) are ATP-dependent enzymes that charge amino acids onto their cognate tRNA [51–54]. Aminoacylates form between the carboxylic acid on the amino acid to either the 2′-OH (Class I) or 3′-OH hydroxyl (Class II) of the 3′-adenosine in tRNA. Roughly each amino acid has its own ARS, which has evolved to selectively acylate its corresponding amino acid onto the appropriate tRNAs. Directed evolution of ARS and ARS/tRNA pairs has been used to modify the amino acid ARS's charge, and onto which tRNA they acylate [55]. Two prominent strategies for modifying the composition of proteins by evolving ARS to recognize non-standard amino acids (nsAA) have thus emerged: residue-

specific incorporation [56–60] and site-specific incorporation [61–70]. While each of these methods is covered in more detail in Chapter 3, I will highlight the important structural features of two pertinent *E. coli* ARS that are discussed in later chapters – the methionyl-tRNA synthetase and the leucyl-tRNA synthetase.

Methyionyl-tRNA Synthetase. The methyionyl-tRNA synthetase (MetRS) is responsible for charging methionine onto a tRNA^{Met}, but has previously been shown to have promiscuous activity. Other amino acids that can be charged by MetRS include azido-homoalanine (Aha), and homo-propargylglycine (Hpg) [56,71]. Both Aha and Hpg have extensively been used in residue incorporation strategies as a way to replace Met residues, found in the translation start of every protein, with amino acids that contain a bioorthogonal tag – either an azide (Aha) or alkyne (Hpg). Though naturally promiscuous, there has been extensive effort at evolving the *E. coli* MetRS to selectively incorporate other amino acids over Met. For example, directed evolution of the MetRS active site lead to the development of NLL-MetRS (MetRS L13N, Y260L, H301L) which canchange amino acid selectivity, previous work has also characterized structural features important for tRNA^{Met} recognition.

Table 1.1. Terminal-alkyne amino acids that have been discovered, the organisms discovered to produce them, and the compound number of corresponding structures. *Streptomyces cattleya* was discovered to be a producer of 2(S)-2-aminopent-4-ynoic acid in this work.

Amino acid	Organism	Compound No.
2(*S*)-3(*S*)-2-amino-3-hydroxypent-4-ynoic acid	*Sclerotium rolfsii, Streptomyces cattleya*	**1**
2(*S*)-2-aminopent-4-ynoic acid	*Streptomyces sp. (HLR-599A), Streptomyces cattleya*, *Amanita pseudoporphyria*	**2**
2(*S*)-2-aminobut-3-ynoic acid	*Streptomyces catenulae*	**3**
2(*S*)-2-aminohex-5-ynoic acid	*Amanita gymnopus, Cortinarius claricolor*	**4**
2(*S*)-2-aminohept-6-ynoic acid	*Amanita gymnopus, Cortinarius claricolor*	**5**
2(*R*)-2-aminohex-5-ynoic acid	*Trogia venenata*	**6**
2(*R*)-2-amino-3-hydroxyhex-5-ynoic acid	*Trogia venenata*	**7**
2(*R*)-4(*R*)-2amino-4-methoxy-5-hexynoic acid	*Trogia venenata*	**8**
2(*S*)-aminohept-4-en-6-ynoic acid	*Amanita pseudoporphyria*	**9**
2-amino-4-methyl-5-hexynoic acid	*Euphoria longan*	**10**
2-amino-4-methoxy-5-hexynoic acid	*Euphoria longan*	**11**
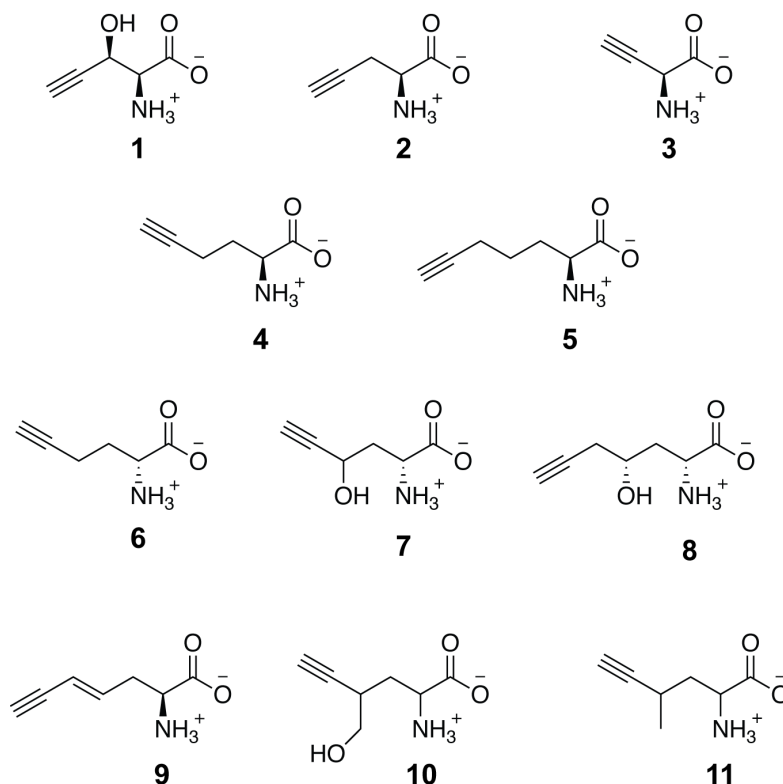
Figure 1.3. Structures of various terminal-alkyne amino acids discovered in fungi and bacteria. Stereochemistry explicitly drawn if known.
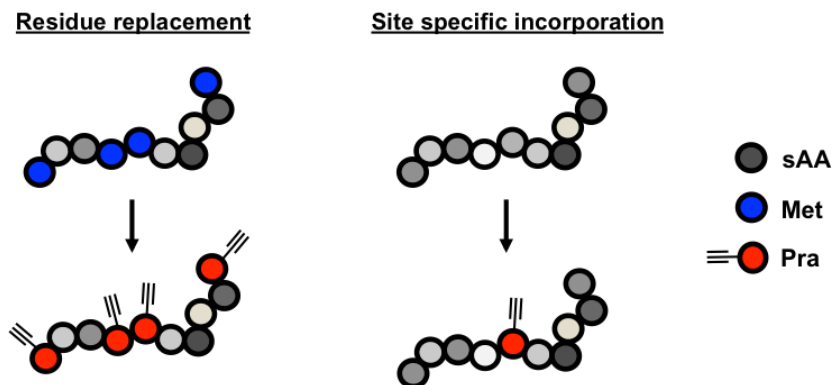
Figure 1.4. Examples distinguishing the two strategies for incorporating non-standard amino acids into polypeptides using engineered ARS. In 'residue replacement', engineered ARS will selectively charge non-standard amino acid onto native tRNA. In the example shown, this leads to replacement of residues for standard amino acid, Met, with a non-standard amino acid, Pra. 'Site specific incorporation' utilizes an engineered ARS/tRNA pair that allows for the genetic encoding non-standard amino acid to the codon of choice (sAA = standard amino acid, Met = methionine, Pra = propargylglycine).

selectively incorporate azido-norleucine (Anl) over Met [72]. Relevant to our own work, other groups have evolved PraRS – a MetRS mutant capable of charging Pra onto tRNA^{Met}. To generate PraRS, researchers used directed evolution at the active site residues (L13, A256, P257, Y260, and H301) followed by random mutagenesis of MetRS and found MetRS L13P A256G P257T Y260Q H301F A331V (PraRS) [73]. In addition to mutating the active site of MetRS to

MetRS selectivity for cognate tRNA^{Met}AUG is partially dictated by the AUG anticodon on the tRNA-Met [74,75]. Two acidic residues along the anti-codon binding site help MetRS discriminate against non-cognate tRNAs. In addition, an aromatic tryptophan residue in this region is necessary for complex formation with cognate tRNA-Met [74–76]. While wild-type *E. coli* MetRS is incapable of amino-acylating suppressor tRNA^{Met}CUA, previous work by others has shown that mutations of the anti-codon binding residues are sufficient to alter MetRS tRNA selectivity *in vitro*, from tRNA^{Met}AUG to tRNA^{Met}CUA [74]. Specifically, MetRS P460R W461A D456K V463A A464M N453I mutant has been shown to both increase MetRS affinity for suppressor tRNA^{Met}CUA and lower its affinity for cognate tRNA^{Met}AUG. Though we have an understanding of how MetRS can be mutated to recognize suppressor tRNAs, an orthogonal MetRS/tRNA^{Met}CUA has yet to be discovered or evolved for site-selective introduction of non-standard amino acids.

Leucyl tRNA Synthetase and a *cis*-editing ARS domains. Analogous to the MetRS, the leucyl-tRNA (LeuRS) synthetase charges leucine onto tRNA^{Leu}. To achieve selectivity, LeuRS utilizes a 'double-sieve' strategy [77]. The LeuRS active site is promiscuous and acylates small amino acids onto tRNA^{Leu}. A second catalytic domain, known as a *cis*-acting editing domain, then hydrolyzes any misacylated aminoacyl-tRNA^{Leu}. The selectivity of the editing step is proposed to be driven by size, whereby the active site sterically excludes Leu-tRNA^{Leu} but allows occupancy of smaller aminoacyl-tRNA^{Leu} substrates. Directed mutagenesis of the LeuRS editing domain has identified residue replacements which abolish the editing activity. Specifically, the LeuRS T252Y mutant has been shown to be incapable of hydrolyzing incorrectly charged tRNA^{Leu}, resulting in an ARS that is more promiscuous for small, aliphatic amino acids [78–80]. Relevant to our own work, LeuRS T252Y has previously been to allow acylation of allylglycine onto tRNA^{Leu}.

Trans-editing ARS domains. In addition to the *cis*-editing domains found on many ARSs (e.g. LeuRS, IleRS, AlaRS) as described above for the LeuRS, standalone *trans*-editing aminoacyl-tRNA hydrolase enzymes catalyze the removal of amino acids from misacylated tRNAs [81–86]. However, there are cases of ARS (e.g. ThrRS, ProRS) from certain organisms utilizing a *trans*-editing domain instead of, or in addition to, the *cis*-editing domain. These organisms (mostly bacteria and archaea) tend to have *trans*-acting editing enzymes which largely fall into two categories: AlaX-like or YbaK-like *trans*-editing enzymes [82,84]. Though much of what we know about *trans*-editing domains is derived from housekeeping enzymes that resolve challenging selectivity problems between the standard proteinogenic amino acids (e.g. Ser vs Cys vs Thr), recent work in our group has discovered a *trans*-editing enzyme, FthB, which has evolved to selectively deacylate a non-standard amino acid produced by the host, 4-fluorothreonine, from tRNA^{Thr} [87]. It remains an open possibility that *trans*-editing domains can be used, in engineered applications, to control the incorporation of non-standard amino acids in proteins in cases where the native editing process is bypassed.

1.4. Conclusion

The concepts reviewed briefly above motivate the work described in this thesis. Bioorthogonal chemistry has had tremendous impact in chemical biology, enabling methods to selectively modify biomolecules and living systems using unique chemical reactivity. Three relevant bioorthogonal reactions to our work are the CuAAC, tetrazine ligation, and tetrazole photoclick reactions, which react selectively with either terminal alkynes or terminal alkenes. There are

examples of natural products, most commonly fatty acids and related polyketides, which contain terminal alkynes and terminal alkenes. Interestingly, terminal alkyne-containing amino acids have also been observed to be produced in some organisms, particularly fungi and soil bacteria, offering an approach to their selective introduction into macromolecules using methods for engineering the residue- or site-selective replacement of the 20 standard proteinogenic amino acids. If we can elucidate how terminal alkyne amino acids are made, we can then envision expanding the genetic code of industrial or model organisms to endow them the capability of making proteins or other natural products with a terminal alkene and alkyne reactive groups.

1.5. References

1. Wen, M., Bond-Watts, B. B. & Chang, M. C. Y. Production of advanced biofuels in engineered E. coli. *Curr. Opin. Chem. Biol.* **17,** 472–9 (2013).

2. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B. & Keasling, J. D. Microbial engineering for the production of advanced biofuels. *Nature* **488,** 320–8 (2012).

3. DiCosimo, R., McAuliffe, J., Poulose, A. J. & Bohlmann, G. Industrial use of immobilized enzymes. *Chem. Soc. Rev.* **42,** 6437–74 (2013).

4. Bhosale, S. H., Rao, M. B. & Deshpande, V. V. Molecular and industrial aspects of glucose isomerase. *Microbiol. Rev.* **60,** 280–300 (1996).

5. Jullesson, D., David, F., Pfleger, B. & Nielsen, J. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.* **33,** 1395–1402 (2015).

6. Zolot, R. S., Basu, S. & Million, R. P. Antibody-drug conjugates. *Nat. Rev. Drug Discov.* **12,** 259–60 (2013).

7. Grosová, Z., Rosenberg, M. & Rebroš, M. Perspectives and applications of immobilised β-galactosidase in food industry - A review. *Czech J. Food Sci.* **26,** 1–14 (2008).

8. Queiroz, A. U. B. & Collares-Queiroz, F. P. Innovation and industrial trends in bioplastics. *Polym. Rev.* **49,** 65–78 (2009).

9. Lim, R. K. V. & Lin, Q. Bioorthogonal chemistry: recent progress and future directions. *Chem. Commun.* **46,** 1589 (2010).

10. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48,** 6974–98 (2009).

11. Best, M. D. Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **48,** 6571–84 (2009).

12. Li, J. & Chen, P. R. Development and application of bond cleavage reactions in bioorthogonal chemistry. *Nat. Chem. Biol.* **12,** 129–137 (2016).

13. Borrmann, A. & van Hest, J. C. M. Bioorthogonal chemistry in living organisms. *Chem. Sci.* **5,** 2123 (2014).

14. Chen, X. & Wu, Y. W. Selective chemical labeling of proteins. *Org. Biomol. Chem.* **14,** 5417–5439 (2016).

15. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chemie - Int. Ed.* **40,** 2004–2021 (2001).

16. Li, S. *et al.* Extent of the oxidative side-reactions to peptides and proteins during CuAAC reaction. *Bioconjug. Chem.* acs.bioconjchem.6b00267 (2016). doi:10.1021/acs.bioconjchem.6b00267

17. Liang, L. & Astruc, D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) 'click' reaction and its applications. An overview. *Coord. Chem. Rev.* **255,** 2933–2945 (2011).

18. Hein, J. E. & Fokin, V. V. Copper-catalyzed azide-alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper (I) acetylides. *Chem. Soc. Rev.* **39,** 1302–1315 (2011).

19. Hong, V., Presolski, S., Ma, C. & Finn, M. G. Analysis and optimization of CuAAC for bioconjugation. *October* **141,** 520–529 (2008).

20. Berg, R. & Straub, B. F. Advancements in the mechanistic understanding of the copper-catalyzed azide-alkyne cycloaddition. *Beilstein J. Org. Chem.* **9,** 2715–50 (2013).

21. Minto, R. & Blacklock, B. Biosynthesis and function of polyacetylenes and allied natural products. *Prog. Lipid Res.* **47,** 233–306 (2008).

22. Wang, X., Huang, B., Liu, X. & Zhan, P. Discovery of bioactive molecules from CuAAC click-chemistry-based combinatorial libraries. *Drug Discov. Today* **21,** 118–132 (2016).

23. Lutz, J. F. & Zarafshani, Z. Efficient construction of therapeutics, bioconjugates, biomaterials and bioactive surfaces using azide-alkyne 'click' chemistry. *Adv. Drug Deliv. Rev.* **60,** 958–970 (2008).

24. Hong, V., Steinmetz, N. F., Manchester, M. & Finn, M. G. Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconjug. Chem.* **21,** 1912–1916 (2010).

25. Rodionov, V. O., Fokin, V. V. & Finn, M. G. Mechanism of the ligand-free Cu(I)-catalyzed azide-alkyne cycloaddition reaction. *Angew. Chemie - Int. Ed.* **44,** 2210–2215 (2005).

26. Liu, J., Xu, Y., Stoleru, D. & Salic, A. Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *Proc. Natl. Acad. Sci.* **109,** 413–418 (2012).

27. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8,** 1128–1137 (2003).

28. Haldón, E., Nicasio, M. C. & Pérez, P. J. Copper-catalysed azide-alkyne cycloadditions (CuAAC): An update. *Org. Biomol. Chem.* 9528–9550 (2015). doi:10.1039/c5ob01457c

29. Codelli, J. A., Baskin, J. M., Agard, N. J. & Bertozzi, C. R. Second-generation difluorinated cyclooctynes for copper-free click chemistry. *J. Am. Chem. Soc.* **130,** 11486–11493 (2008).

30. Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **39,** 1272–1279 (2010).

31. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine ligation: Fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc.* **130,** 13518–13519 (2008).

32. Oliveira, B. L., Guo, Z. & Bernardes, G. J. L. Inverse electron demand Diels-Alder reactions in chemical biology. *Chem. Soc. Rev.* **46,** 4895–4950 (2017).

33. Rieder, U. & Luedtke, N. W. Alkene – tetrazine ligation for imagine cellular DNA. *Angew. Chemie - Int. Ed.* **53,** 9168–9172 (2014).

34. Pidgeon, S. E. & Pires, M. M. Metabolic remodeling of bacterial surfaces via tetrazine ligations. *Chem. Commun.* **51,** 10330–10333 (2015).

35. Wang, Y., Song, W., Hu, W. J. & Lin, Q. Fast alkene functionalization in vivo by

photoclick chemistry: HOMO lifting of nitrile imine dipoles. *Angew. Chemie - Int. Ed.* **48,** 5330–5333 (2009).

36. Song, W., Wang, Y., Qu, J. & Lin, Q. Selective functionalization of a genetically encoded alkene-containing protein via photoclick chemistry in bacterial cells. *J. Am. Chem. Soc.* **130,** 9654–9655 (2008).

37. Song, W. *et al.* A metabolic alkene reporter for spatiotemporally controlled imaging of newly synthesized proteins in mammalian cells. *ACS Chem. Biol.* **5,** 875–885 (2010).

38. Scrimgeour, C. M. Natural acetylenic and olefinic compounds, excluding marine natural products. *Aliphatic Relat. Nat. Prod. Chem.* **2,** 1–19 (1979).

39. Kuklev, D. V, Domb, A. J. & Dembitsky, V. M. Bioactive acetylenic metabolites. *Phytomedicine* **20,** 1145–59 (2013).

40. Dembitsky, V. M. Anticancer activity of natural and synthetic acetylenic lipids. *Lipids* **41,** 883–924 (2006).

41. Edwards, D. J. *et al.* Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chem. Biol.* **11,** 817–833 (2004).

42. Zhu, X., Liu, J. & Zhang, W. De novo biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11,** 115–20 (2015).

43. Zhu, X., Su, M., Manickam, K. & Zhang, W. Bacterial genome mining of enzymatic tools for alkyne biosynthesis. *ACS Chem. Biol.* **10,** 2785–2793 (2015).

44. Shin-Ichi, H. Amino acids from mushrooms. *Prog. Chem. Org. Nat. Prod.* 117–140 (1992).

45. Yoshio, K. *et al.* FR-900130, A novel amino acid antibiotic I. Discovery, taxonomy, isolation, and properties. *J. Antibiot. (Tokyo).* **33,** 125–131 (1980).

46. Yoshio, K. *et al.* FR-900130, A novel amino acid antibiotic II. Isolation and structure eluciation of the acetyl derivative of FR-900130. *J. Antibiot. (Tokyo).* **33,** 132–136 (1980).

47. Sanada, M., Miyano, T. & Iwadare, S. Ethynylserine, an antimetabolite of L-threonine from Streptomyces cattleya. *J. Antiboitics* **39,** 304–305 (1986).

48. Potgieter, H. C., Vermeulen, N. M. J., Potgieter, D. J. J. & Strauss, H. F. A toxic amino acid, 2(S)3(R)-2-amino-3-hydroxypent-4-ynoic acid from the fungus Sclerotium rolfsii. *Phytochemistry* **14,** 1757–1759 (1977).

49. Zhou, Z.-Y. *et al.* Evidence for the natural toxins from the mushroom Trogia venenata as a cause of sudden unexpected death in Yunnan Province, China. *Angew. Chem. Int. Ed. Engl.* **51,** 2368–70 (2012).

50. Scannell, J. P. *et al.* Antimetabolites produced by microorganisms II. L-2-Amino-4-Pentynoic Acid. *J. Antibiot. (Tokyo).* **24,** 239–244 (1971).

51. Woese, C. R., Olsen, G. J., Ibba, M. & Söll, D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64,** 202–236 (2003).

52. Schimmel, P. R. & Söll, D. Aminoacyl-tRNA synthetases: general features and

recognition of transfer RNAs. *Annu. Rev. Biochem.* **48,** 601–648 (2003).

53. Cusack, S. Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* **7,** 881–889 (1997).

54. Nawaz, M. H. & Martinis, S. A. Chemistry of aminoacyl tRNA synthetases. *Wiley Encycl. Chem. Biol.* (2008). doi:10.1002/9780470048672.wecb008

55. Voloshchuk, N. & Montclare, J. K. Incorporation of unnatural amino acids for synthetic biology. *Mol. Biosyst.* **6,** 65–80 (2010).

56. Yuet, K. P. & Tirrell, D. A. Chemical tools for temporally and spatially resolved mass spectrometry-based proteomics. *Ann. Biomed. Eng.* **42,** 299–311 (2014).

57. Hinz, F. I., Dieterich, D. C., Tirrell, D. A. & Schuman, E. M. Noncanonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. *ACS Chem. Neurosci.* **3,** 40–49 (2012).

58. Niehues, S. *et al.* Impaired protein translation in Drosophila models for Charcot-Marie-Tooth neuropathy caused by mutant tRNA synthetases. *Nat. Commun.* **6,** 7520 (2015).

59. Erdmann, I. *et al.* Cell-selective labelling of proteomes in Drosophila melanogaster. *Nat. Commun.* **6,** 7521 (2015).

60. Dieterich, D. C., Link, J. A., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9482–7 (2006).

61. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83,** 379–408 (2014).

62. Zhang, Z., Wang, L., Brock, A. & Schultz, P. G. The selective incorporation of alkenes into proteins in Escherichia coli. *Angew. Chem. Int. Ed. Engl.* **41,** 2840–2 (2002).

63. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science.* **964,** 964–968 (2013).

64. Chatterjee, A., Xiao, H. & Schultz, P. G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 14841–6 (2012).

65. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20,** 1044–1048 (2002).

66. Mehl, R. A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125,** 935–939 (2003).

67. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 225–249 (2006).

68. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126,** 14306–14307 (2004).

69. Anderson, J. C. & Schultz, P. G. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **42,** 9598–9608 (2003).

70.   Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **124,** 9026–9027 (2002).

71.   Beatty, K. E. & Tirrell, D. A. Two-color labeling of temporally defined protein populations in mammalian cells. *Bioorganic Med. Chem. Lett.* **18,** 5995–5999 (2008).

72.   Tanrikulu, I. C. *et al.* Cell-selective metabolic labeling of proteins. *Nat. Chem. Biol.* **5,** 715–717 (2009).

73.   Truong, F., Yoo, T. H., Lampo, T. J. & Tirrell, D. A. Two-strain, cell-selective protein labeling in mixed bacterial cultures. *J. Am. Chem. Soc.* **134,** 8551–8556 (2012).

74.   Schmitt, E., Meinnel, T., Panvert, M., Mechulam, Y. & Blanquet, S. Two acidic residues of E. coli MetARS as negative discriminants towards the binding of non-cognate tRNA anticodons. *J. Mol. Biol.* **233,** 615–628 (1993).

75.   Mechulam, Y. *et al.* Crystal structure of Escherichia coli methionyl-tRNA synthetase highlights species-specific features. *J. Mol. Biol.* **294,** 1287–97 (1999).

76.   Meinnel, T. *et al.* Selection of suppressor methionyl-tRNA synthetases: mapping the tRNA anticodon binding site. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 291–5 (1991).

77.   Yadavalli, S. S. & Ibba, M. Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. in *Advances in protein chemistry and structural biology* **86,** 1–43 (Elsevier Inc., 2012).

78.   Tang, Y. & Tirrell, D. A. Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **41,** 10635–10645 (2002).

79.   Zhai, Y. *et al.* Modulation of substrate specificity within the amino acid editing site of leucyl-tRNA synthetase. *Biochemistry* **46,** 3331–3337 (2007).

80.   Lue, S. W. & Kelley, S. O. An aminoacyl-tRNA synthetase with a defunct editing site. *Biochemistry* **44,** 3010–3016 (2005).

81.   Ahel, I., Korencic, D., Ibba, M. & Söll, D. Trans-editing of mischarged tRNAs. *Proc. Natl. Acad. Sci.* **100,** 15422–15427 (2003).

82.   An, S. & Musier-Forsyth, K. Trans-editing of Cys-tRNAPro by Haemophilus influenzae YbaK protein. *J. Biol. Chem.* **279,** 42359–62 (2004).

83.   Das, M., Vargas-Rodriguez, O., Goto, Y., Suga, H. & Musier-Forsyth, K. Distinct tRNA recognition strategies used by a homologous family of editing domains prevent mistranslation. *Nucleic Acids Res.* **42,** 3943–3953 (2014).

84.   Fukunaga, R. & Yokoyama, S. Structure of the AlaX-M trans-editing enzyme from Pyrococcus horikoshii. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63,** 390–400 (2007).

85.   Novoa, E. M. *et al.* Ancestral AlaX editing enzymes for control of genetic code fidelity are not tRNA-specific. *J. Biol. Chem.* **290,** 10495–503 (2015).

86.   Korencic, D. *et al.* A freestanding proofreading domain is required for protein synthesis quality control in Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10260–5 (2004).

87.   McMurry, J. L. & Chang, M. C. Y.   Fluorothreonyl-tRNA deacylase prevents

mistranslation in the organofluorine producer Streptomyces cattleya . *Proc. Natl. Acad. Sci.* **114,** 11920–11925 (2017).

Chapter 2: *Discovery of the β-ethynylserine biosynthetic gene cluster*

2.1. Introduction

While acetylenic natural products form a relatively large class of compounds (~2,000) [2–4], very few of these contain a terminal alkyne that can be used in the CuAAC reaction, which requires a terminal alkyne as a substrate (*Figure 2.1A*). Interestingly, a set of terminal alkyne-containing amino acids had been reported to be produced by a select group of fungi and bacteria (*Figure 2.1B*) [5–10]. Of the 11 reported terminal-alkyne amino acids, three (Bes, Pra, and ethynylglycine) have been reported to be produced by streptomycetes. Notably, terminal-alkyne amino acids, such as Pra, have previously been used for both residue- [11–13] and site-specific [14–19] labeling of proteins and proteomes. However, this body of previous work has relied on terminal-alkyne amino acids which are first chemically synthesized and then fed to organism. As such, the identification of the biosynthetic pathway for these amino acids could potentially provide a general route to encode the *in situ* production of terminal alkyne-containing proteins directly from glucose.
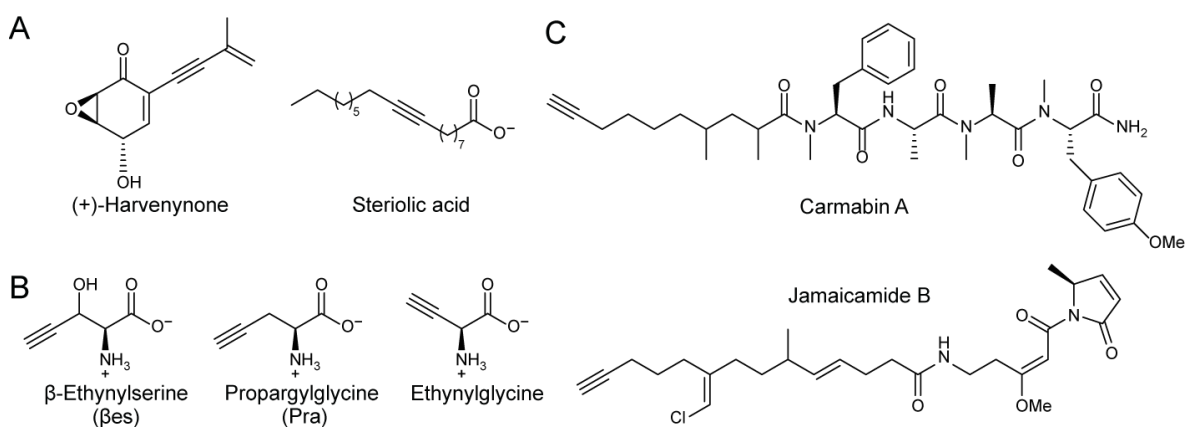
Figure 2.1. Natural products of acetylenic origin. (A) Natural products with internal terminal alkyne. (B) Terminal-alkyne amino acids of streptomyces biosynthetic origin. (C) Terminal alkyne natural products with alkynes from fatty acid origin.

Early isotope-labeling experiments showed that most terminal-alkyne compounds originate from acetate-derived fatty acid or polyketide biosynthetic pathways and were postulated to utilize a fatty acid desaturase-like enzyme to form the triple bond [2]. More recently, fatty acid desaturases were also identified to participate in the formation of the terminal alkyne of the polyketide natural products, Jamaicamide B and Carmabin A [2,20,21] (*Figure 2.1C*). The most well characterized acetylenase from these biosynthetic pathways is JamB, a membrane-bound fatty acid acetylenase which desaturastes a C6-acyl-ACP. Homologues of JamB have also been discovered to accept a wider range of acyl chain lengths [22]. However, the few terminal-alkyne amino acids described to date appear to be structurally quite different from products originating from fatty acids [10]. We thus focused on elucidating the genes for production of β-ethynylserine (βes) in a genetically-tractable soil bacterium, *Streptomyces cattleya*.

17

2.2. Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), yeast extract, malt extract, and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), adenosine monophosphate sodium salt (AMP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), pyruvate kinase, lactate dehydrogenase, isocitrate dehydrogenase, lysozyme, magnesium sulfate heptahydrate, poly(ethyleneimine) solution (PEI), ammonium iron (II) sulfate hexahydrate, β-mercaptoethanol (βME), sodium phosphate dibasic heptahydrate, D-mannitol, apramycin sulfate salt (Am), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), betaine, dimethylsulfoxide (DMSO), pyrodoxal-5′-phosphate (PLP), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), 4-amino-3-penten-2-one (Fluoral-P), L-allylglycine (allylglycine), copper sulfate (CuSO$_4$), Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), maltose, L-propargylglycine (Pra), and L-lysine, L-glutamate, γ-glutamyl-p-nitroaniline, γ-glutamyl-transpeptidase from equine kidney (Ggt), sodium cyanoborohydride, sodium cyanoborodeuteride (96% D), sodium bromide, sodium fluoride, sodium iodide, n-Butyllithium solution 2.5 M in hexanes (n-BuLi), Tetrahydrofuran (anhydrous, THF), Chlorotrimethylsilane (TMS-Cl), BOC-Gly (OH), N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC), 4-(Dimethylamino)pyridine (DMAP), Lithium diisopropylamide (LDA), Tetrabutylammonium fluoride (TBAF), Trimethylsilyl trifluoromethanesulfonate (TMSOTf), and glutamate dehydrogenase were purchased from Sigma-Aldrich (St. Louis, MO). 4-Cl-lysine was purchased from AKos GmbH (Steinen, Germany). (2S)-2-amino-4-chloropent-4-enoic acid (4-Cl-allyglycine) was purchased from Enamine, Ltd (Kiev, Ukraine). (2S)-2-amino-4-bromopent-4-enoic acid (4-Br-allyglycine) and Hygromycin B was purchased from Santa Cruz Biotechnology (Santa Cruz, California). 3-Azido-7-hydroxycoumarin was purchased from AK Scientific, Inc (Union City, CA). Formaldehyde (16% (*w/v*), methanol-free) and PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA synthetase was purchased from Megazyme International (Bray, Ireland). Bacto^{TM} Agar was purchased from BD (Sparks, Maryland) Soy flour was purchased from Berkeley Bowl (Berkeley, CA). Deuterium oxide, L-lysine · 2HCl (^{13}C$_6$, 99%; ^{15}N$_2$, 99%), L-lysine ·2HCl (4,4,5,5-D$_4$, 96-98%), D-dextrose (1-^{13}C 98%), and L-glutamate (^{15}N, 99%) were purchased from Cambridge Isotope Laboratories (Andover, MA). SuperCos 1 Cosmid Vector Kit was purchased from Agilent (Santa Clara, CA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, T4 DNA ligase, Antarctic phosphatase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at either 4 °C for immediate usage or -20 °C for longer term usage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). SoluLyse reagent was purchased from Genlantis (San Diego, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). Strep-

Tactin Superflow resin was purchased from IBA GmbH (Göttingen, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators and 5,000 MWCO regenerated cellulose ultrafiltration membranes were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). TEV protease was purchased from the University of California, Berkeley Macro Lab (Berkeley, CA).

Bacterial strains. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection (Manassas, VA). *Streptomyces lavenduligriseus* (NRRL B-3173), *Streptomyces catenulae* (NRRL B2342), *Streptomyces achromogenes* (NRRL B-2120), *Streptomyces* sp. (NRRL S-1448), and *Streptomyces* sp. (NRRL S-31) were provided by Agricultural Research Service Culture Collection: Bacterial Foodborne Pathogens and Mycology Research Unit (National Center for Agricultural Utilization Research, Peoria, IL). *E. coli* DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production of BesA, BesB, BesC, BesD, BesE. *E. coli* GM272 harboring the non-transmissible, *oriT*-mobilizing plasmid pUZ8002 was used for conjugative plasmid transfer into *S. cattleya.*

Comparative genomics and clustering analysis. Protein genome sequences and associated protein lists for 27 *Streptomyces* spp. were downloaded from NCBI database (NIH). Three species of Streptomycetes have been reported to produce terminal-alkyne amino acids, *S. cattleya, S. catenulae*, and *Streptomyces* sp. HLR-599A. However, *S.* sp. HLR-599A is held in a private collection and its genome has not been sequenced; therefore it was not included in this study. An all-by-all protein pBLAST was performed between *S. cattleya* (chromosome: NC_015876, plasmid: NC_015875) and *S. catenulae* (NZ_JODY00000000) as the in-group and *S. cattleya* + *S. hygroscopicus, S. rimosus, S. griseus, S. albulus , S. filamentosus, S. viridochromogenes, S. lavendulae, S. chartreusis, S. ipomoneae, S. sviceus, S. pristinaespiralis, S. bingchenggensis, S. ghanaensis, S. coelicolor, S. clavuligerus, S. ambofaciens, S. grieseoplanus, S. scabiei, S avermenitilis, S. canus, S. antibioticus, S. roseochromogenus, S. cyaneogriseus, S. prunicolor, S. lydicus* as the out-group. Strict adjacency criteria, within 2 CDSs of best homologue, was used to determine if proteins were clustered in both organisms. Proteins were clustered and visualized using Cytoscape. Proteins that had a lower E-value in the In-group were marked 'unique', while proteins with lower E-value in the Out-group were marked 'common'. Of the initial ~650 clusters of proteins, greater than 3 proteins in size, only 4 were unique between *S. cattleya* and *S. catenulae*. Pipeline and tutorial for comparative clustering analysis is available online as a php script on Github (https://github.com/jamarchand/io).

Cell culture of *Streptomyces* spp. for metabolite analysis. Single colonies of *S. catenulae, S.* sp. NRRL-1448*, S.* sp. NRRL-S31*, S. achromogenes, S. lavendulaegriseus*, *S. cattleya, S. cattleya* Δ*SCAT_0184*::Am^R, *S. cattleya* Δ*SCAT_4823*::Am^R, *S. cattleya* Δ*SCAT_p1525*::Am^R, *S. cattleya* Δ*besA*::Am^R, *S. cattleya* Δ*besB*::Am^R, *S. cattleya* Δ*besC*::Am^R, *S. cattleya* Δ*besD*::Am^R, *S. cattleya* Δ*besE*::Am^R*, S. cattleya* Δ*besB*::Am^R *attB*^{ΦC31}::*ermEp*-besB*, *S. cattleya* Δ*besC*::Am^R *attB*^{ΦC31}::*ermEp*-besC*, and *S. cattleya* Δ*besD*::Am^R *attB*^{ΦC31}::*ermEp*-besD* were picked from soy-mannitol plates and inoculated into GYM pH 7 (25 mL). After 3 d of growth, cultures were sub-cultured into GYM pH 7 (25 mL), and grown for 7-14 d.

Derivatization and characterization of β-ethynylserine and propargylglycine production in cell culture. Culture supernatants were sampled by removing 1 mL of culture broth and

centrifuging at 9,800 × g for 10 min at room temperature. 100 μL of supernatant was added to a solution containing 3-azido-7-hydroxy-coumarin (20 μM), THPTA (100 μM), and sodium ascorbate (10 mM) with and without $CuSO_4$ (50 μM) in 400 μL of PBS (pH 7). After incubating for 10 min, aliquots were analyzed using an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 μm, 2.1 × 50 mm; Agilent) using a linear gradient from 5% to 95% acetonitrile over 5 min at a flow rate of 0.6 mL/min with 0.1% (*v/v*) formic acid, pH 3.5 as the mobile phase. Mass spectra were acquired using an Agilent 6530 QTOF with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

Extraction of polar metabolites from lysates and supernatants of *Streptomyces* spp. cultures. An aliquot of culture broth (1 mL) was removed and centrifuged at 9,800 × g for 10 min at room temperature to pellet cells. The supernatant was then diluted 1:1 in methanol with 1% (v/v) formic acid, centrifuged at 9,800 × g for 10 min to remove precipitate for LC/ MS analysis. Culture pellets were then collected after removal of the supernatant and resuspended by vigorous vortexing in a 50% (*v/v*) solution of methanol (100 μL of methanol/100 mg of wet cell pellet). These solutions were subjected to five freeze-thaw cycles from -80°C (30 min) to -20°C (30 min) with vortexing in between. Solutions were then centrifuged at 9,800 × g for 10 min to remove remaining cell debris before LC/MS analysis. Comparative metabolic experiments of *S. cattleya* wildtype*, ΔbesA, ΔbesB, ΔbesC, ΔbesD,* and *ΔbesE* strains were carried out on cell pellet extracts.

General procedure for high resolution HPLC/MS analysis of polar metabolites. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1 × 100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

Comparative metabolite profiling of *S. cattleya* wildtype, Δ*besA*, Δ*besB*, Δ*besC*, Δ*besD*, and Δ*besE* strains. Raw mass spectra data was converted to mzXML format using MZconvert. Peak detection was processed on XCMS platform [23] from the Nomura Lab. Data-collection parameters were set to default HPLC Orbitrap values (centwave feature detection, 0.5 minimum fraction of samples in one group to be a valid group, *P*-value thresholds for significant feature = 0.05, isotopic ppm error = 5, *m/z* absolute error = 0.015) except the following: ppm = 20, signal/noise threshold = 2, mzwid = 0.025, and with loess non-linear retention time alignment (default settings). List of *m/z* values, retention times, peak intensities, and peak areas were exported to Excel for processing.

***In vitro* assays of BesD and BesD G139D.** Reactions (50 μL) contained L-lysine · HCl (4 mM), sodium αKG (5 mM), sodium ascorbate (5 mM), and $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ (1 mM) in 100 mM MOPS buffer (pH 7). Chloride is introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Various assay components were omitted for control reactions as

20

noted in the data. Reactions were initiated by addition of wild-type or G139D BesD (17.6 µM final concentration) and allowed to proceed for 1 h at room temperature before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. To analyze Cl-lysine instability, samples were quenched at t = 0 and t = 1 h after removal of the enzyme with Pall Nanosep spin filters (10 kDa MWCO).

***In vitro* assay for BesC.** BesC was assayed by either coupling to BesD in order to produce 4-Cl-lysine *in situ* or by using purified 4-Cl-lysine. For coupled reactions, reactions (50 µL) contained BesD (2.9 µM), L-lysine · HCl (4 mM), sodium αKG (5 mM), sodium ascorbate (5 mM), and $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ (1 mM) in 100 mM HEPES (pH 7.5). For reactions using purified 4-Cl-lysine, reactions (50 µL) contained 4-Cl-lysine (0.5 mM), sodium ascorbate (5 mM), and $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ (1 mM) in 100 mM HEPES (pH 7.5). Chloride is introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. For production of allyglycine, the reaction was run without addition of BesD and sodium αKG. Various assay components were omitted for control reactions as notated in the data. Reactions were initiated by addition of BesC (15.7 µM final concentration) and incubated for 1 h at room temperature before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid and centrifuged at 9,800 × *g* for 10 min. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

***In vitro* assay for BesB.** Reactions (100 µL) contained either 4-Cl-allyglycine, 4-Br-allylglycine, or allylglycine (1 mM) and PLP (10 µM) in 100 mM HEPES (pH 7.5). Reactions were initiated by the addition of BesB (2.4 µM final concentration) and incubated at room temperature for 2 h before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

D_2O exchange assay for BesB. BesB (1.2 mg/ml), was exchanged into 50 mM ammonium bicarbonate buffer in 90% D_2O (pH 7.5). 4-Cl-allylglycine or L-propargylglycine was prepared as a 100 mM stock solution in >99% D_2O. Reactions (1 mL) contained 4-Cl-allylglycine (10 mM, 100 mM stock prepared in D_2O) or L-propargylglycine (10 mM, 100 mM Stock prepared in D_2O) and PLP (20 µM) in 50 mM ammonium bicarbonate (pH 7.5 in D_2O). Reactions were initiated by the addition of BesB (4.8 µM final concentration) and incubated for 30 min at room temperature, for 4-Cl-allylglycine, or 24h, for L-propargylglycine, before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. Reaction carried out with L-propargylglycine as a substrate was also analyzed by ^{2}D-NMR (900 MHz, 90% D_2O).

Steady-state kinetic characterization of BesB. Reactions (100 µL) contained varying concentrations of L-Pra (0.1, 0.5, 1 mM) with PLP (20 µM), HEPES (50 mM, pH 7.5). Reactions were initiated by addition of BesB (3.0 µM final concentration). Initial rates were measured by monitoring formation of L-Pra using LC/MS Agilent 6530 QTOF. A standard curve of L-Pra (10 µM – 1 µM) was ran along side samples for quantification.

Imine trapping BesB PLP-intermediates. Reactions (100 µL) contained varying concentrations of 4-Cl-allylglycine (1 mM), 4-Br-allylglycine (1 mM) or L-Pra (1 mM) with PLP (20 µM), HEPES (50 mM, pH 7.5). Reactions were initiated by addition of BesB (3.0 µM final

concentration). Reactions were then quenched, at 5 min and 30 min, with 5 µL of sodium cyanoborohydride or sodium cyanoborodeuteride. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using a Hypercarb (5 µm, 250 Å, 150 × 21.1 mm, Thermo-Fischer Scientific). The mobile phase was set as gradient from 90% H2O, 10% acetonitrile with 10 mM ammonium bicarbonate to 10% H2O, 90% acetonitrile, and 10 mM ammonium bicarbonate over 10 minutes. Mass spectra were acquired using an Agilent 6530 QTOF with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

Halogen exchange reactions with BesB. Purified MBP-BesB was desalted by passage through Zeba spin desalting column and exchanged into 50 mM HEPES (pH 7.5) with various salts: NaF, NaCl, NaBr, or NaI (50 mM). Reactions (100 µL) contained 4-Cl-allylglycine (1 mM) or 4-Br-allylglycine (1 mM) with PLP (20 µM), and buffer HEPES (50 mM, pH 7.5) with 50 mM of either NaF, NaCl, NaBr, or NaI. Reactions were initiated by addition of BesB (3.0 µM final concentration) in the corresponding salt buffer. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Preparation of compound 1′: 1,3-Bis(trimethylsilyl)prop-2-yn-1-ol. 3-(trimethylsilyl)prop-2-yn-1-ol (1.3g, 10 mmol, 5.89 mL), n-BuLi 1.6M in hexane (17.1 mL, 2.75 mmol), trimethylsilylchloride (1.2 mL, 11.2 mmol) were used. Dry THF was added to 3-(trimethylsilyl)prop-2-yn-1-ol under N_2. The solution was cooled to -78 °C while stirring for 10 min. n-BuLi (7.65 mL) was slowly added to solution and allowed to stir for 3 h. After adding trimethylsilylchloride, the solution was allowed to warm to room temperature for 2 h. The solution was then cooled to -45 °C and n-BuLi (9.5 mL) was slowly added. Reaction was stirred at -45 °C for 2h then cooled to -78 °C and quenched with 10% acetic acid. Solution was extracted with ether, neutralized with $NaHCO_3$, and evaporated in vacuo before purification by flash column chromotagraphy on silica gel (n-hexane: AcOEt = 50: 1) to give a light yellow oil. ^{1}H NMR (400 MHz, CDCl3)**:** δ 4.07 (s, 1H, H_1), 0.14 (s, 9H, TMS_1), 0.12 (s, 9H, TMS_2).

Preparation of compound 2′: *tert*-**Butyl ((1,3-bis(trimethylsilyl)prop-2-ynyloxy)carbonyl)methylcarbamate**. 1,3-Bis(trimethylsilyl)prop-2-yn-1-ol (1.05 g, 5.2 mmol), N-Boc-glycine (1.38 g, 7.8 mmol), EDCI (1.2 g, 7.8 mmol) and DMAP (66 mg, 0.5 mmol) were used. Dry DCM was added to a solution containing N-Boc-glycine, DMAP, and 1,3-Bis(trimethylsilyl)prop-2-yn-1-ol while stirring at room temperature under N_2. The solution was cooled to 4 °C for 10 min. EDC was then slowly added and solution was removed from water bath and allowed to warm to room temperature for 1.5 h. The solvent was evaporated in vacuo to give a crude pale yellow mixture. The mixture was extracted with ether and sat. $NaHCO_3$ and evaporated in vacuo. The resulting residue was purified by flash column chromatography on silica gel (n-hexane : AcOEt = 30 : 1) to give the 1,3-bis(trimethylsilyl)prop-2-ynyloxy)carbonyl)methylcarbamate (700 mg, 37%) as white crystals. ^{1}H NMR (400 MHz, CDCl3)**:** δ 5.25 (s, 1H, H_1), 4.99 (brs, 1H, N<u>H</u>Boc), 3.95-3.83 (m, 2H, $H_{1'}$), 1.43 (s, 9H, Boc), 0.13 (s, 9H, TMS_1), 0.12 (s, 9H, TMS_2).

Preparation of compound 3′: 2-(t-Butoxycarbonylamino)-3,5-bis(trimethylsilyl)penta-3,4-dienoic acid by Claisen rearrangement. 1,3-bis(trimethylsilyl)prop-2-ynyloxy)carbonyl)methylcarbamate (700 mg, 1.95 mol), diisopropylamine (0.95ml, 690 mg, 6.85 mol), n-BuLi 1.6M in hexane (3.6 mL, 5.85 mol), zinc chloride 1M in ether (2.4 mL, 2.4 mol) were used. A solution of n-BuLi in hexane (3 equiv) was slowly added to 0.28 mL of

diisopropylamine (3.5 equiv) in THF at -78°C and stirred for 30 min under N_2. Meanwhile, 1 equiv of tert-Butyl (((S)-1,3-bis(trimethylsilyl)prop-2-ynyloxy)carbonyl)methylcarbamate was added to 5 mL of THF under N_2 then slowly added to solution of n-BuLi and diisopropylamine at -78 °C. 1.2 equiv of dry ZnCl in ether was then added and the solution was allowed to stir at -78°C for 30 min. The solution was then allowed to warm to room temperature for 1 h. Reaction was quenched with 1 N HCl and extracted with $MgSO_4$. The solvent was evaporated in vacuo to give a light yellow crude residue. Crude residue was purified by flash chromatography on silica gel (n-hexane : AcOEt = 1 : 1) to give a light yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$)**:** δ 5.03 (brd, J = 8.0 Hz, 1H, N$\underline{H}$Boc), 4.79 (s, 1H, H$_2$), 4.65 (brd, J = 8.0 Hz, 1H, H$_1$) 1.44 (s, 9H, Boc), 0.13 (s, 9H, TMS$_1$), 0.06 (s, 9H, TMS$_1$), 0.12 (s, 9H, TMS$_2$). ESI-MS [M+H]$^+$: calculated for $C_{16}H_{32}NO_4Si_2$, *m/z*, 358.1864 found *m/z* 358.1858.

Preparation of compound 4′: 2-(tert-Butoxycarbonylamino)penta-3,4-dienoate. 2-(t-Butoxycarbonylamino)-3,5-bis(trimethylsilyl)penta-3,4-dienoic (20 mg, 0.056 mmol), acetic acid (2.4 µL), TBAF (105 µL) were used. TBAF and acetic acid were added to a of solution 2-(t-Butoxycarbonylamino)-3,5-bis(trimethylsilyl)penta-3,4-dienoic in THF and allowed to stir for 16 h at room temperature. Reaction was quenched with HCl, extracted with AcOEt, and dried in vacuo to give a crude yellow solid. Crude product was used for later steps without further purification. ESI-MS [M+H]$^+$: calculated for $C_{10}H_{16}NO_4$, *m/z*, 214.1074, found *m/z* 214.1069.

Preparation of compound 5′: allenylglycine. Crude 2-(tert-Butoxycarbonylamino)penta-3,4-dienoate **(**approx. 12 mg, 0.07 mmol), lutidine (100 µL), TMSOTF (100 µL) were used. Lutidine (50 µL) and TMSOTF (50 µL) were added to 2-(tert-Butoxycarbonylamino)penta-3,4-dienoate in DCM and allowed to stir at room temperature. After 1 h, lutidine (50 µL) and TMSOTF (50 µL) were added to the solution and allowed to react for 1 h. Solution was extracted with H_2O and lyophilized to give a white crystal. Final compound was then purified by liquid chromatography using SeQuant ZIC-HILIC Semi-Preparative Columns (5 µm, 200 Å, 150 × 21.1 mm, EMD Millipore) with the following gradient: Buffer A (90 % acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). Linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 60% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 ml / min. ^{1}H NMR (600 MHz, D_2O): δ 5.54 (q, J = 6.6 Hz, 1H, H$_2$), 5.18-5.16 (m, 2H, H$_3$), 4.37-4.35 (m, 1H, H$_1$). ESI-MS [M+H]$^+$: calculated for $C_5H_8NO_2$, *m/z*, 114.0550, found *m/z* 114.0550.

Steady-state kinetic characterization of BesA. Reactions (100 µL) contained varying concentrations of ʟ-Pra (0.4-5 mM) or ʟ-cysteine (1.25-20 mM), with ʟ-glutamate (10 mM), ATP (2.5 mM), $MgCl_2$ (5 mM), phosphoenol-pyruvate (PEP, 10 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/mL), and pyruvate kinase (PK, 10 U/mL) in 100 mM HEPES (pH 7.5). Reactions were initiated by addition of BesA (6.4 µM final concentration). Initial rates were measured by monitoring λ$_{340}$ using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature. Kinetic parameters (k_{cat}, K_M) were determined by fitting the data using Microcal Origin to the equation: $v_o = v_{max}$ [S] / (K_M + [S]), where v is the initial rate and [S] is the substrate concentration. Error bars on graphs represent mean ± s.d. (n = 3). Error in k_{cat}/K_M is calculated by propagation of error from the individual kinetic parameters.

***In vitro* assay for BesA.** Reactions (50 µL) contained ʟ-Pra (2 mM), ʟ-glutamate (5 mM), ATP (2.5 mM), $MgCl_2$ (5 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of BesA (6.4 µM) and incubated at room temperature for 1 h before quenching in 1.5 volumes of

methanol with 1% (v/v) formic acid. Reactions with BesA were compared with the production of γ-Glu-Pra from commercial γ-glutamyl-transpeptidase (Ggt). Ggt reactions (50 µL) contained L-Pra (2 mM), γ-glutamyl-*p*-nitroaniline (5 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of Ggt (1 U) and allowed to proceed for 1 h before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

***In vitro* assay for BesE.** BesE was assayed by coupling to either BesA or Ggt in order to produce Glu-Pra *in situ*. With BesA coupling, reactions (50 µL) contained BesA (6.4 µM), L-Pra (2 mM), L-glutamate (5 mM), ATP (2.5 mM), MgCl$_2$ (5 mM), sodium αKG (5 mM), (NH$_4$)$_2$Fe(SO$_4$)$_2$ · 6H$_2$O (1 mM), and sodium ascorbate (1 mM) in 100 mM MOPS buffer (pH 7). With Ggt coupling, BesA and L-glutamate were replaced with Ggt (1 U) and γ-glutamyl-*p*-nitroaniline (5 mM). Reactions were initiated by addition of BesE (4.4 µM) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

***In vitro* reconstitution of Pra biosynthesis from L-lysine with BesB, BesC, BesD.** Reactions (50 µL) contained BesD (2.9 µM), BesC (15.7 µM), BesB (2.4 µM), sodium αKG (5 mM), sodium ascorbate (5 mM), and (NH$_4$)$_2$Fe(SO$_4$)$_2$ · 6H$_2$O (1 mM), and PLP (10 µM) in 100 mM HEPES (pH 7.5). Chloride is introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Reactions were initiated by addition of L-lysine · HCl (4 mM) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

***In vitro* reconstitution of Glu-βes biosynthesis with BesA, BesB, BesC, BesD and BesE.** Reactions (50 µL) contained BesD (2.9 µM), BesC (15.7 µM), BesB (2.4 µM), BesA (6.4 µM), BesE (4.4 µM), sodium αKG (5 mM), sodium ascorbate (5 mM), and (NH$_4$)$_2$Fe(SO$_4$)$_2$ · 6H$_2$O (1 mM), PLP (10 µM), L-glutamate (5 mM), ATP (2.5 mM), and MgCl$_2$ (5 mM) in 100 mM HEPES (pH 7.5). Chloride is introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Reactions were initiated by addition of L-lysine · HCl (4 mM) and allowed to proceed for 16 h before being quenched in 1.5 volumes of methanol with 1% (*v/v*) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis.

Construction of cosmids for fatty acid desaturase gene disruption in *S. cattleya*. *S. cattleya* NRRL 8057 cosmid library was constructed using the SuperCos 1 Cosmid Vector Kit (Stratagene). Genes *SCAT_p1525, SCAT_0184, SCAT_2136,* and *SCAT_4823* on cosmids were disrupted by lambda-red mediated recombination with an Am[R]/OriT cassette. The Am[R]/OriT cassette was amplified from pIJ773 using primers sets LR-SCAT_0184-F/R, LR-SCAT_2136-F/R, LR-SCAT_4823-F/R, and LR-SCAT_p1525-F/R (*Appendix A1.1*). These cassettes were used to disrupt the fatty acid desaturases in their respective cosmids using the REDIRECT method[34] Apramycin resistant colonies were picked and disruption of *SCAT_0184*, *SCAT_2136*, *SCAT_4823*, *SCAT_p1525* on each cosmid was confirmed by sequencing.

Construction of plasmids for *bes* cluster gene disruption and complementation in *S. cattleya*. Primers for amplifying homology arms for gene disruption were designed to produce amplicons of either 2 kB or 4 kB of homology upstream of start codon and downstream of stop

codon for gene of interest. For gene complementation, primers were designed to clone native *S. cattleya besB*, *besC*, or *besD* genes into pSET152 (Hyg^R) under an ermEp* promoter. Standard molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase using the oligonucleotides listed in Appendix A1.1. For gene disruption, pIJ773-BesA, pIJ773-BesB, pIJ773-BesC, pIJ773-BesD, and pIJ773-BesE were constructed by amplification of 2 kb regions from *S. cattleya* genomic DNA using the respective primers in the K773 primer set where K773-BesX-UP-F/R corresponds to the primers used to generate upstream fragment of gene *besX*, while K773-BesX-DOWN-F/R corresponds to the downstream fragment (*Appendix A1.1*). For gene complementation, pSET152-BesB, pSET152-BesC, and pSET152-BesD were constructed using primer sets pSET152-BesX-F/R. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (Berkeley, CA).

Conjugative transfer of plasmids and cosmids from *E. coli* to *S. cattleya*. Plasmids and cosmids for gene disruption or complementation were transferred into *S. cattleya* by conjugation with *E. coli* GM272 harboring the *oriT*-mobilizing plasmid pUZ8002 using the literature protocol [24,25]. Gene disruption was verified by PCR amplification of the target locus and neighboring loci using primer set 'dck', followed by sequencing to validate correct disruption, generating *S. cattleya* Δ*SCAT_0184*::Am^R, *S. cattleya* Δ*SCAT_4823*::Am^R, *S. cattleya SCAT_p1525*::Am^R, *S. cattleya* Δ*besA*::Am^R, *S. cattleya* Δ*besB*::Am^R, *S. cattleya* Δ*besC*::Am^R, *S. cattleya* Δ*besD*::Am^R, and *S. cattleya* Δ*besE*::Am^R (*Appendix A1.1*). Complementation of *S. cattleya* knockout strains was carried out by introducing an expression cassette for the appropriate *bes* gene driven by the ermEp* promoter into the phage attachment site. *S. cattleya* Δ*besB*::Am^R, *S. cattleya* Δ*besC*::Am^R, and *S. cattleya* Δ*besD*::Am^R were transformed respectively with pSET152-BesB, pSET152-BesC, pSET152-BesD to generate *S. cattleya* Δ*besB*::Am^R *attB*$^{\Phi C31}$::*ermEp**-besB*, *S. cattleya* Δ*besC*::Am^R *attB*$^{\Phi C31}$::*ermEp**-besC*, and *S. cattleya* Δ*besD*::Am^R *attB*$^{\Phi C31}$::*ermEp**-besD*. Complementation was verified by PCR using primers designated to amplify wild-type gene (*Appendix A1.1*).

Construction of plasmids for protein expression. Standard molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase using the oligonucleotides listed in *Appendix A1.1*. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (Berkeley, CA) or the Barker Hall Sequencing Facility at UC Berkeley (Berkeley, CA).

Expression of His$_6$MBP-, His$_{10}$-, and Strep-tagged proteins. *E. coli* BL21 Star (DE3) was transformed with appropriate protein expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics (50 μg/mL kanamycin) in a 2.8 L-baffled shake flask to OD$_{600}$ = 0.05. The cultures were grown at 37°C at 200 rpm to OD$_{600}$ = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.2 mM) and overnight growth at 16°C. For His$_{10}$-BesE, His$_6$MBP-BesC, His$_{10}$-BesD, Strep-BesD, Strep-BesD G139D the medium was supplemented with Fe(SO$_4$) · 7H$_2$O (420 μM) at the time of induction. Cell pellets were harvested by centrifugation at 9,800 × *g* for 7 min at 4ºC and stored at -80ºC.

Expression testing of His$_6$MBP-tagged BesB variants for solubility. Expression testing of His$_6$MBP-BesB was performed by the University of California, Berkeley Macro Lab (Berkeley,

CA). Vector pSV272.1-BesB was transformed into *E. coli* Rosetta2(DE3) pLysS, *E. coli* BL21(DE3) CodonPlus RIL, *E. coli* BL21 Star (DE3), *E. coli* SoluBL21, and *E. coli* Shuffle T7 Express. Transformants were grown overnight in autoinducing Magic Media (Invitrogen), in a 96-well block at 37°C at 200 rpm. Cultures were them moved to 16°C and allowed to continue growing for an additional 40 h. Cells were lysed using BugBuster (EMD Millipore) and proteins purified using Ni-NTA resin (Qiagen). BL21 Star (DE3) was selected as the production host from these studies based on yield of soluble protein and the lack of need for a secondary antibiotic.

Purification of His$_6$MBP-BesB. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5) supplemented with phenylmethane sulfonyl fluoride (PMSF, 0.5 mM). The cell paste was homogenized then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 14,000 psi. The lysate was then centrifuged at 13,500 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (*w/v*) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at 13,500 × *g* for 20 min at 4°C. The remaining soluble lysate was diluted three-fold then passed over amylose column (1 mL resin / 20 g of cell paste) for isolation by gravity flow. Column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5) for 20 column volumes and eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 20 mM βME, 10% (*v/v*) glycerol, 10 mM maltose, pH 7.5). Fractions containing the target protein were pooled by A$_{280\,nm}$ and concentrated using an Amicon Ultra spin concentrator (100 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM NaCl, 20 μM PLP, 20% (*v/v*) glycerol, 1 mM DTT, pH 7.5) using a PD-10 desalting column. Based on concerns over solubility, His$_6$MBP-BesB was used directly for *in vitro* assays. Final protein concentration before storage (1.2 mg/mL) was estimated using the ε$_{280\,nm}$ calculated by ExPASY ProtParam for His$_6$MBP-BesB and measured by nanodrop (ε$_{280\,nm}$ = 112,440 M^{-1} cm^{-1}). Protein was aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

Purification of His$_{10}$-BesA, His$_{10}$-BesD, His$_{10}$-BesE, and His$_6$MBP-BesC. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5) supplemented with PMSF (0.5 mM). The cell paste was homogenized then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 9,000 psi. The lysate was then centrifuged at 13,500 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (*w/v*) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at 13,500 × *g* for 20 min at 4°C. The soluble lysate was incubated with Ni-NTA (0.5 mL resin/g of cell paste) for 45 min at 4°C, then resuspended and loaded onto a column by gravity flow. Column was washed with wash buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5) for 15-20 column volumes. The column was then eluted with elution buffer (50 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5).

For His$_6$MBP-BesC and His$_{10}$MBP-BesD, the protein was dialyzed three times following elution from Ni-NTA column for 1.5 h against dialysis buffer (50 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5) to remove imidazole. After the third round of dialysis, protein was incubated with TEV protease (0.5 mg protease/mg protein) and dialyzed overnight into dialysis buffer. Cleaved

and dialyzed protein was passed through amylose column (4 mL) to remove TEV and MBP. The eluent was then diluted to a final salt concentration of 20 mM NaCl using buffer A (50 mM HEPES, 20% (*v/v*) glycerol, 1 mM DTT, pH 7.5) and loaded onto a 5 mL HiTrap-Q column for ion exchange. The protein was eluted using a gradient from 0-100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, 20% (*v/v*) glycerol, 1 mM DTT, pH 7.5) over 40 min.

Fractions containing the target protein were pooled by $A_{280\,nm}$ and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (*v/v*) glycerol, 1 mM DTT, pH 7.5) using PD-10 desalting columns.

Final protein concentrations before storage were estimated using the $\varepsilon_{280\,nm}$ calculated by ExPASY ProtParam and measured by nanodrop, and are as follows: His$_{10}$-BesA: 5.6 mg/mL ($\varepsilon_{280\,nm} = 40,255$ M^{-1} cm^{-1}), His$_{10}$-BesE: 2.9 mg/mL ($\varepsilon_{280\,nm} = 48,930$ M^{-1} cm^{-1}), BesC: 4.8 mg/mL ($\varepsilon_{280\,nm} = 34,380$M^{-1} cm^{-1}), BesD: 1.6 mg/mL ($\varepsilon_{280\,nm} = 47,565$ M^{-1} cm^{-1}). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

Purification of Strep-BesD and Strep-BesD G139D. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 20% (*v/v*) glycerol, 2.5 mM DTT, pH 8.0) supplemented with PMSF (0.5 mM). The cell paste was homogenized then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 9,000 psi. The lysate was then centrifuged at $13,500 \times g$ for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (*w/v*) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at $13,500 \times g$ for 20 min at 4°C. Avidin was then directly added to the soluble lysate (20 μg/L *E. coli* cell culture OD$_{600}$ = 1). Lysate was then passed over Strep-Tactin Superflow resin (0.5 mL resin /g cell pellet) and washed with 10 column volumes of wash buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 20% (*v/v*) glycerol, 2.5 mM DTT, pH 8.0). Protein was then eluted with 6 column volumes of elution buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 10 mM desthiobiotin, 20% (*v/v*) glycerol, 2.5 mM DTT, pH 8.0). Fractions containing the target protein were pooled by $A_{280\,nm}$ and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (*v/v*) glycerol, 2.5 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\varepsilon_{280\,nm}$ calculated by ExPASY ProtParam and measured by nanodrop and are as follows: Strep-BesD: 8.9 mg/mL ($\varepsilon_{280\,nm} = 47,565$ M^{-1} cm^{-1}), Strep-BesD G139D: 7.3 mg/mL ($\varepsilon_{280\,nm} = 47,565$ M^{-1} cm^{-1}). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

Preparation and purification of 4-Cl-lysine. 4-Cl-lysine was prepared in a 200 μL reaction containing BesD (70 μM), L-lysine · HCl (0.35 mM), αKG (5 mM), (NH$_4$)$_2$Fe(SO$_4$)$_2$·6H$_2$O (1 mM), and sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. Samples were allowed to react for 1 h before quenching in 1.5 volumes of methanol with 1% (*v/v*) formic acid. 4-Cl-lysine was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 μm, 200 Å, 150 × 21.1 mm, EMD Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A gradient was applied as follows: linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 40% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 ml/min. The

fractions containing the 4-Cl-lysine were determined by LC-QTOF MS screening, pooled, and dried by rotary evaporation followed by Speed-vac when the volume reached <1.5 mL.

Preparation and purification of $[^{13}C_6,^{15}N_2]$-4-Cl-lysine methyl ester. $[^{13}C_6,^{15}N_2]$-4-Cl-lysine methyl ester was prepared in a 200 µL reaction containing BesD (70 µM), $[^{13}C_6,^{15}N_2]$- L-lysine · 2HCl (0.35 mM), αKG (5 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ (1 mM), sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. The reaction was quenched after 20 min by transferring into 3 mL of 3 M HCl in methanol followed by a 10 min incubation at 50°C. Samples were then cooled at room temperature for 3 h before neutralizing by dropwise addition of 10 M NaOH. Precipitate was removed by centrifugation at 10,000 × g for 10 min. The $[^{13}C_6,^{15}N_2]$- 4-Cl-lysine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 µm, 200 Å, 150 × 21.1 mm, EMD Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A gradient was applied as follows: linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 40% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 ml/min. The fractions containing the 4-Cl-lysine methyl ester were determined by LC-QTOF MS screening, pooled, and dried by rotary evaporation followed by Speed-Vac when the volume reached <1.5 mL. Samples were resuspended in a mixture of 60% D_3-acetonitrile:40% D_2O (*v/v*) for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS $[M+H]^+$: calculated for $C^{13}C_6H_{16}^{35}Cl^{15}N_2O$ *m/z*, 203.1037, found *m/z* 203.1037. $C^{13}C_6H_{16}^{37}Cl^{15}N_2O$ *m/z*, 205.1007, found *m/z* 205.1008.

NMR characterization of $[^{13}C_6,^{15}N_2]$-4-Cl-lysine methyl ester. A constant-time ^{1}H-^{13}C HSQC experiment (Bruker pulse sequence hsqcctetgpsi) was recorded with carrier frequencies set to 5.48 ppm (^{1}H) and 40 ppm (^{13}C)) and the spectral widths were set to 16 ppm (^{1}H) and 80 ppm (^{13}C). The recycle delay was set to 1.5 s. A 2D ^{1}H-^{13}C version of the HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was recorded with carrier frequencies of 5.48 ppm (^{1}H) and 39 ppm (^{13}C) with spectral widths of 14 ppm (^{1}H) and 75 ppm (^{13}C). The recycle delay was set to 1.8 s. Chemical shifts (ppm) were obtained from 2D-NMR spectra at a resolution too low to observe coupling constants. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. ^{1}H NMR (900 MHz, 60% D_3-acetonitrile:40% D_2O): δ 4.36 (H_α), 4.28 (H_γ), 2.47 ($H_{\beta1}$), 2.39 ($H_{\beta2}$), 2.27 ($H_{\delta1}$), 2.14 ($H_{\delta2}$), 3.26 ($H_{\epsilon1}$), 3.17 ($H_{\epsilon2}$). ^{13}C NMR (900 MHz, 60% D_3-acetonitrile:40% D_2O): δ 56.6 (C_γ), 51.6 (C_α), 37.6 (C_ϵ), 39.1 (C_β), 2.14 (C_δ).

Preparation and purification of ε-lactam cyclization product of 4-Cl-lysine (2b). An enzymatic reaction (250 µL) containing BesD (18 µM), $[^{13}C_6,^{15}N_2]$- L-lysine · 2HCl (5 mM), αKG (10 mM), $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ (1 mM), and sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5) was run for 16 h at room temperature. Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. The reaction was then quenched by transfer into 750 µL of methanol with 1% (*v/v*) formic acid. The precipitate was removed by centrifugation at 10,000 × g for 10 min. $[^{13}C_6,^{15}N_2]$-4-hydroxy-2-amino-ε-lactam was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 µm, 200 Å, 150 × 21.1 mm, EMD Millipore) using the following buffers: Buffer A (90 % acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 90% to 10% Buffer A was applied over 1.5 h at a

flow rate of 5 mL/min. The peak determined to be the ε-lactam by LC/MS was collected manually, lyophilized, and resuspended in 90% d^6-DMSO:10% D$_2$O (v/v) for NMR analysis on a Bruker 900 MHz instrument.

NMR characterization of ε-lactam cyclization product of 4-Cl-lysine (2b). All experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K. The instrument was equipped with a CP TXI cryoprobe. A constant-time ^{1}H-^{13}C HSQC spectrum (Bruker pulse sequence hsqcctetgpsp) was recorded with a ^{13}C evolution period of 13.6 ms, so that signals from ^{13}C attached to two ^{13}C neighbors appeared opposite in phase to those with one ^{13}C neighbor. The carrier frequencies were set to 4.7 ppm (^{1}H) and 40 ppm (^{13}C) and the spectral widths were set to 16 ppm (^{1}H) and 80 ppm (^{13}C). The recycle delay was set to 1.5 s. A 2D ^{1}H-^{13}C version of the HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was recorded with carrier frequencies of 5.5 ppm (^{1}H) and 39 ppm (^{13}C) and the spectral widths were set to 14 ppm (^{1}H) and 75 ppm (^{13}C). The recycle delay was set to 1.8 s. A 2D ^{1}H-^{15}N HSQC experiment (Bruker pulse sequence hsqcetf3gpsi) was recorded with carrier frequencies of 4.65 ppm (^{1}H) and 70 ppm (^{15}N) with spectral widths of 16 ppm (^{1}H) and 140 ppm (^{15}N). The recycle delay was set to 1.0 s. A 2D ^{1}H-^{13}C version of the HNCO experiment (Bruker pulse sequence hncogp3d) was recorded with carrier frequencies of 4.65 ppm (^{1}H), 119.5 ppm (^{15}N) and 176 ppm (^{13}C) and the spectral widths were set to 16 ppm (^{1}H) and 22 ppm (^{13}CO). The recycle delay was 1.0 s. Finally, a 2D ^{1}H-^{13}C version of the HNCA experiment (pulse sequence hncacbgp3d with ^{13}C-^{13}C evolution delay set to 20 μs, effectively making it an HNCA experiment) was recorded with carrier frequencies of 4.65 ppm (^{1}H), 119.5 ppm (^{15}N), and 40 ppm (^{13}C) and spectral widths were set to 14 ppm (^{1}H) and 75 ppm (^{13}C). The recycle delay was set to 1.0 s. Chemical shifts (ppm) were obtained from 2D-NMR spectra at a resolution too low to observe coupling constants. ^{1}H NMR (900 MHz, 90% d^6-DMSO:10% D$_2$O): δ 4.11 (H$_\alpha$), 8.1 (CONH), 3.90 (H$_\gamma$), 3.20 (H$_{\varepsilon 1}$), 3.10 (H$_{\varepsilon 2}$), 2.13 (H$_{\beta 1}$), 2.04 (H$_{\delta 1}$), 1.69 (H$_{\beta 2}$), 1.35 (H$_{\delta 2}$). ^{13}C NMR (900 MHz, 90% d^6-DMSO:10% D$_2$O): δ 172.0 (C=O), 69.4 (C$_\gamma$), 48.2 (C$_\alpha$), 36.6 (C$_\varepsilon$), 36.2 (C$_\beta$), 35.9 (C$_\delta$). ^{15}N NMR (900 MHz, 90% d^6-DMSO:10% D$_2$O): δ 116.3 (CONH).

Detection of ammonia as a coproduct of BesC catalysis. Ammonia was detected using a glutamate dehydrogenase-coupled assay[35]. Reactions (100 μL) contained BesD (10 μM), L-lysine · HCl or ^{13}C$_6$,^{15}N$_2$- L-lysine · 2HCl (10 mM), sodium αKG (5 mM), sodium ascorbate (1 mM), Fe(II)SO$_4$ · 7H$_2$O (0.5 mM), glutamate dehydrogenase (3 U/mL), and NADPH (0.3 mM) in 100 mM HEPES (pH 7.5 Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. Standard curves of 4-Cl-allylglycine, allylglycine, and ^{15}N-glutamate were prepared in the reaction buffer lacking BesD and BesC. Reactions were initiated by addition of BesC (10 μM) and incubated at room temperature for 2 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid containing 5 μM Phe as an internal standard before centrifuging at 9,800 × g for 10 min. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. For spectrophotometric analysis of NADH consumption, reactions were prepared as described above except that glutamate dehydrogenase and NADPH were omitted. After 1 h incubation at room temperature, glutamate dehydrogenase and NADPH were added to 3 U/mL and 0.3 mM, respectively, and the A$_{340}$ was monitored on a SpectraMax M2 Microplate Reader (Molecular Devices).

Detection of formaldehyde as a coproduct of BesC catalysis. Formaldehyde was detected using the Fluoral-P reagent[36]. Reactions (100 μL) contained BesD (5 μM), L-lysine · HCl or

$^{13}C_6, ^{15}N_2$- L-lysine · 2HCl (5 mM), sodium αKG (5 mM), sodium ascorbate (1 mM), Fe(II)SO$_4$ · 7H$_2$O (0.5 mM), and 4-amino-3-penten-2-one (Fluoral-P) (20 mM) in 100 mM HEPES (pH 7.5). Reactions were initiated by addition of BesC (15 μM). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. Standard curves of 4-Cl-allylglycine, allylglycine, and formaldehyde were prepared in a reaction mixture omitting BesD and BesC. After incubating for 16 h at room temperature, the reaction product of formaldehyde and Fluoral-P (3,5-diacetyl-1,4-dihydro-2,6-lutidine) was quantified by fluorescence using a BioTek SynergyMx plate reader (λ_{ex} = 410 nm, λ_{em} = 500 nm) by comparison to a formaldehyde standard curve (0 - 200 μM) . Samples were then quenched in 1.5 volumes of methanol with 1% (*v/v*) formic acid containing 5 μM Phe as an internal standard and analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above to quantify the total alkene product formed.

Succinate formation by BesE. Reactions (100 μL) contained ATP (2.5 mM), MgCl$_2$ (5 mM), phosphoenol-pyruvate (PEP, 1 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL), succinyl CoA synthetase (SCS, 3.2 U/ml), coenzyme A (1 mM), sodium αKG (1 mM), (NH$_4$)$_2$Fe(SO$_4$)$_2$ · 6H$_2$O (0.2 mM), and sodium ascorbate (2 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of BesD (2.9 μM) in the presence and absence of Pra. Initial rates of NADH consumption were measured by monitoring A$_{340}$ using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature.

2.3. Results and discussion

Knockout of fatty acid desaturases. Our initial approach identified fatty acid desaturases as a likely candidate for formation of the terminal alkynes in βes and Pra. Using the annotated genome of *S. cattleya* DSM-46488, we were able to identify four putative fatty acid desaturases: SCAT_0184, SCAT_2136, SCAT_4823, and SCAT_p1525. *SCAT_0184* is adjacent to type I polyketide synthase and may be part of a biosynthetic cluster (*Figure 2.2*). *SCAT_p1525* may also have a metabolic role, although its genomic context is not highly informative. *SCAT_2136* and *SCAT_4823* are clustered with proteins thought to be involved in primary metabolism and rRNA transcription. Targeted knockouts of these four putative fatty acid desaturases was attempted by conjugative transfer of cosmids (~50 kb in size) containing the targeted gene replaced by an apramycin resistance (AccIV) cassette and an OriT. We were unable to identify any deletion strains of *SCAT_2136* suggesting *SCAT_2136* may be an essential fatty acid desaturase. Knockouts of *SCAT_0184*, *SCAT_4823*, and *SCAT_p1525* were found to still produce βes by CuAAC derivatization with 3-azido-7-OH-coumarin followed by LC/QTOF monitoring the βes-coumarin CuAAC product (βes-Cou).

Comparative genomics. We then shifted our attention to comparative genomic approaches to identify uniquely shared gene clusters among *Streptomyces* spp. that produce terminal-alkyne amino acids. Of the two other known producers, only *Streptomyces catenulae* had an available genome sequence and was therefore validated for terminal-alkyne amino acid production (*Figure 2.3*). Upon culturing *S. catenulae* DSM-40258, we discovered this strain produced L-propargylglycine (Pra). Although it was previously reported in literature, we were unable to detect production β-ethynylglycine [9,26]. Interestingly, when we cultured *S. cattleya* we detected Pra in cell-pellet metabolite extracts and βes in supernatant. Given that both *S. cattleya* and *S. catenulae* produced Pra, we hypothesized that both these organisms share a common set of genes for Pra biosynthesis.

Two critical assumptions were being made by using this approach. The first is that both organisms have a set of genes for Pra biosynthesis, and these genes should be unique to these two organisms. Formally, this assumption materializes by finding genes that have more in common with each other, by sequence E-value, than they do with the nearest homologues in other species of Streptomycetes. The second assumption is that the Pra biosynthetic genes should be clustered in both organisms. Unlike eukaryotes, prokaryotes are known to cluster genes related to common metabolic function together in the genome. To test this criteria, a pairwise distance between ORFs was calculated for each gene in *S. cattleya* and mapped to the nearest homologue of adjacent genes in *S. catenulae*. If both sets of genes were within 2 genetic loci, a flexible criteria, they were considered clustered. Therefore, for each gene we were able to group by cluster, and determine if member genes of each cluster are unique to *S. catenulae* and *S. cattleya*, or more commonly found in other species of Streptomycetes.

By using the described approach, we were able to reduce the ~650 total gene clusters conserved between *S. cattleya* and *S. catenulae* to just four unique clusters not found in the other 26 control *Streptomyces* spp. genomes (*Figure 2.4*). Among these four, we focused on one particular cluster that contained gene functions predicted to be associated with amino acid metabolism.

A

Gene	Homologue	%ID/%S	aa
SCAT_0184	Fatty acid desaturase [*Streptomyces uncialis*]	50/64	346
SCAT_2136	Fatty acid desaturase [*Streptomyces megasporus*]	83/91	332
SCAT_4823	Acyl-CoA desaturase [*Streptomyces sp. NRRL S-813*]	98/99	372
SCAT_p1525	ΔFatty acid desaturase [*Streptomyces sp. IMTB 2501*]	86/90	361

B

βes-Cou
[M+H]$^+$
m/z = 333.0831
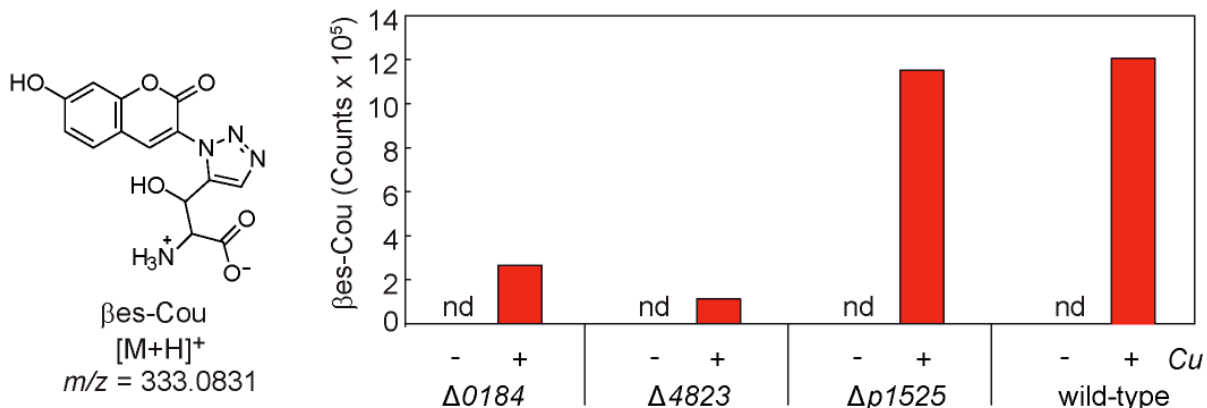
Figure 2.2. (A) A total of four fatty acid desaturases were identified in the genome of *S. cattleya* and targeted for deletion. Percent identity (%ID) and percent similarity (%S) compared to the closest sequenced homologue with the host in brackets. Number of amino acids (aa) in protein also tabulated. *SCAT_0184* is adjacent to type I polyketide synthase and may be part of a biosynthetic cluster. *SCAT_p1525* may also have a metabolic role, although its genomic context is not highly informative. *SCAT_2136* and *SCAT_4823* are clustered with proteins thought to be involved in primary metabolism and rRNA transcription. We were unable to identify any deletion strains of *SCAT_2136* suggesting *SCAT_2136* may be an essential fatty acid desaturase. Knockouts of *SCAT_0184*, *SCAT_4823*, and *SCAT_p1525* were found to still produce βes by CuAAC derivatization with 3-azido-7-hydroxycoumarin followed by LC/QTOF monitoring the βes-coumarin CuAAC product (βes-Cou). The reaction where Cu(I) is omitted provides a negative control since Cu(I) is required to for the cycloaddition to occur. (B) Extracted ion chromatograms of the culture supernatant of *S. catenulae* with and without CuAAC-derivatization monitoring either the Pra-coumarin adduct (Pra-Cou; *m/z* = 317.0880) or Pra (*m/z* = 114.0550) compared to a synthetic Pra standard using LC/QTOF. Chromatograms shown are representative of at least 3 experimental replicates.

32
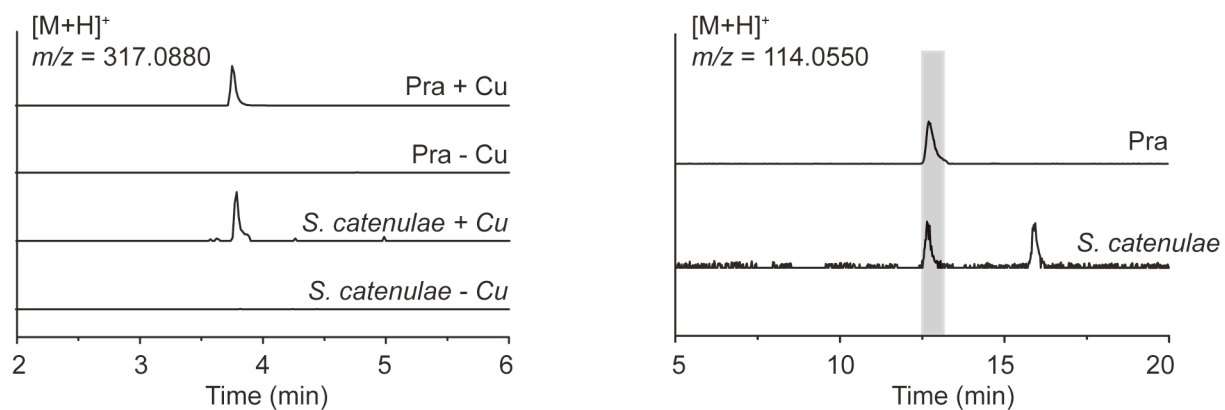
Figure 2.3. Extracted ion chromatograms of the culture supernatant of *S. catenulae* with and without CuAAC-derivatization monitoring either the Pra-coumarin adduct (Pra-Cou; *m/z* = 317.0880) or Pra (*m/z* = 114.0550) compared to a synthetic Pra standard using LC/QTOF. Chromatograms shown are representative of at least 3 experimental replicates.
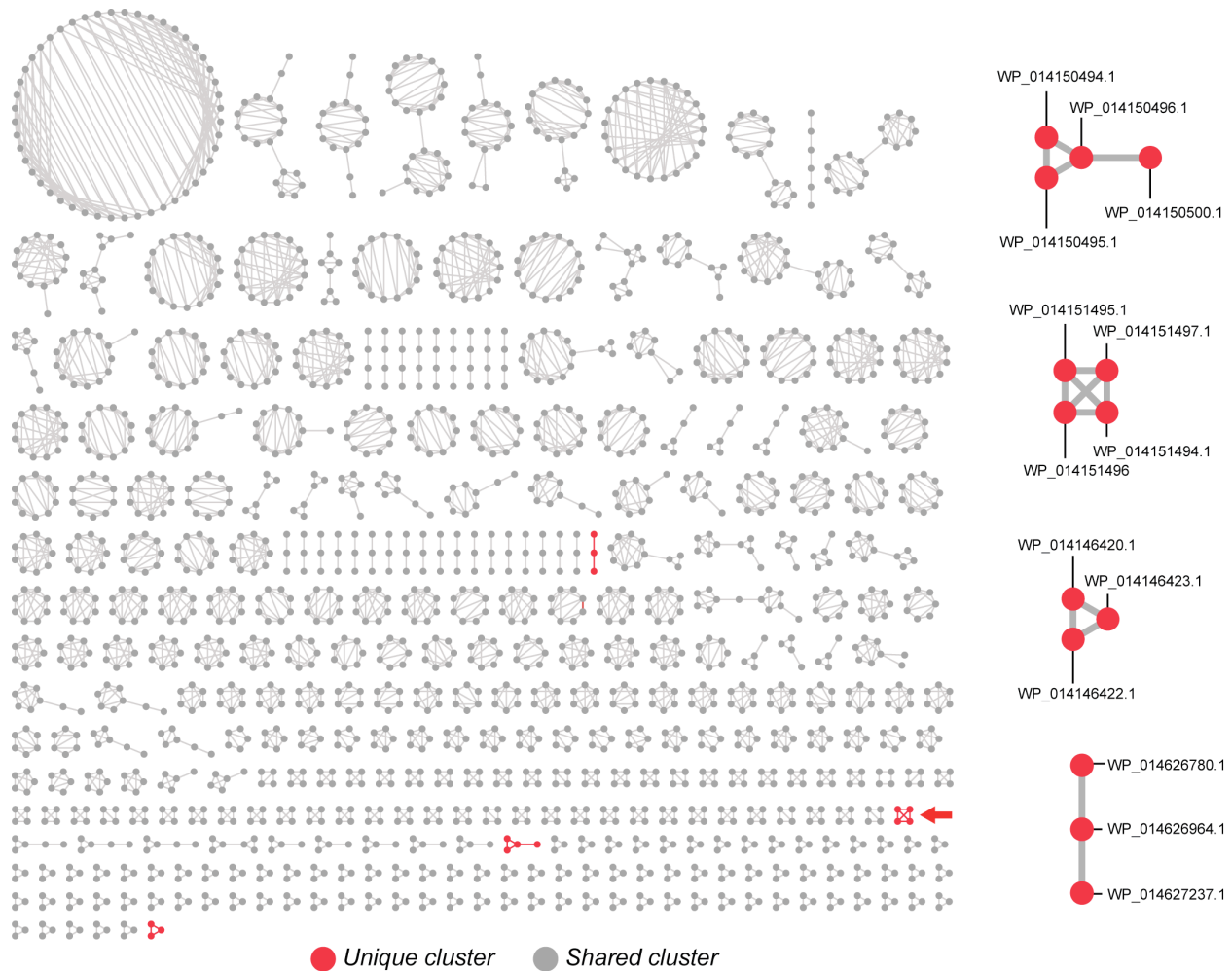
Figure 2.4. We generated a list of approximately 650 gene clusters (> 2 genes) conserved between both species. Clusters were generated using Cytoscape [37]. Nodes are individual gene while edges connect gene that are within 2 protein coding regions from one another. Gene clusters comprise of connected nodes. Connected grey nodes are gene clusters also found in any the other 26 *Streptomyces spp.* used in the analysis. Connected red nodes are gene clusters found to be unique between *S. cattleya* and *S. catenulae*. We then curated this list by removing gene clusters that are more closely related between *S. cattleya* and 26 other *Streptomyces spp.*, narrowing down the number of conserved unique clusters to just four. One of these unique clusters contains member nodes WP_014151494.1, WP_014151495.1, WP_014151496.1, WP_014151497.1, which composes the putative *bes* cluster (indicated with a red arrow).

Validation of putative βes gene cluster. The putative βes biosynthetic cluster features six proteins (*Table 2.1*). BesA possesses a highly conserved ATP-grasp domain and shares homology to carboxylate-amine ligases. BesB is homologous to the pyridoxal-5′-phosphate (PLP)-dependent cystathionine-γ-lyase/*O*-succinylhomoserine-β-synthase family. BesC displays homology to enzymes in the HemeO non-heme iron oxidase superfamily and both BesD and BesE are predicted to be non-heme Fe/α-ketoglutarate (αKG)-dependent oxidases. Finally BesF is a putative EamA-like transporter, with homologues of this protein implicated in amino acid efflux. Searching the non-redundant protein database for additional instances of these genes we identified five additional *Streptomyces* spp. that had not been reported to produce terminal-alkyne amino acids yet were also found to possess the *besABCD* genes (*Figure 2.5, 2.6, Appendix A1.3*). Three of these new strains were found to produce either Pra or βes in cell culture (*Figure 2.7*). Based on these results and the comparison of gene clusters, we propose that BesABCD are responsible for Pra biosynthesis and form the core of the cluster, while BesE is related to the conversion of Pra to βes by hydroxylation at the β-position.

	Gene	aa	Homologue	%ID/%S	Proposed function
←	besF	297	EamA family transporter [*Streptomyces xylophagus*]	69/78	Amino acid transporter
→	besA	511	Carboxylate-amine ligase [*Cylindrospermopsis raciborskii*]	32/50	Amino acid ligase
←	besB	465	Cystathionine beta-lyase [*Opitutaceae* bacterium TSB47]	42/54	Gamma-lyase
←	besC	252	Pyrroloquinoline-quinone synthase [*Marinobacter pelagius*]	44/65	Oxidase
←	besD	268	ArpA protein [*Pseudomonas azotoformans*]	56/71	Halogenase
→	besE	257	Proline hydroxylase [*Streptomyces misionensis*]	54/66	Hydroxylase

Table 2.1. Genes in βes biosynthetic gene cluster with nearest homologue, percent identity and similarity to nearest homologue (%ID/%S), and proposed function.

The biosynthetic genes *besABCDE* were then targeted for deletion from the genome to generate the *S. cattleya* Δ*besA*, Δ*besB*, Δ*besC*, Δ*besD*, and Δ*besE* knockout strains. We derivatized samples of supernatant and metabolites extract with N_3-Coumarin, using CuAAC, to detect formation of terminal alkynes by LC/MS. These experiments showed that the Δ*besB*, Δ*besC*, and Δ*besD* strains did not produce detectable amounts of Pra or βes (*Figure 2.8*). Genomic complementation of the knocked out gene further support the assignment of terminal-alkyne amino acid production to the *bes* gene cluster (*Figure 2.8*). Likely do to polar effects, only the *S. cattleya* Δ*besB*+*besB* showed restored production of Pra and Bes. In comparison, the Δ*besA* strain showed lowered yields of both products, while the Δ*besE* strain still produced Pra but not βes. These results support the hypothesis that *besBCD* are critical to Pra formation, while also suggesting that *besE* is the hydroxylase that converts Pra to βes.

Identification of pathway intermediates through comparative metabolomics. In order to identify additional intermediates, we performed an untargeted comparative metabolomics experiment on the intracellular metabolite extracts of *S. cattleya* wild-type, Δ*besA*, Δ*besB*, Δ*besC*, and Δ*besD* using high resolution LC-MS. The resulting mass spectra were processed using XCMS [23] for feature detection to generate an initial set of ~1,000 lead compounds with high statistical significance ($p < 0.05$) and high fold change (> 2) between wildtype and knockout. By focusing on the metabolites that accounted for the highest degree of variance, we were able to identify a set of putative intermediates and precursors for βes biosynthesis (*Figure 2.9, Figure 2.10*). Based on these findings, we hypothesized that

βes/Pra clusters (*besABCDEF*)
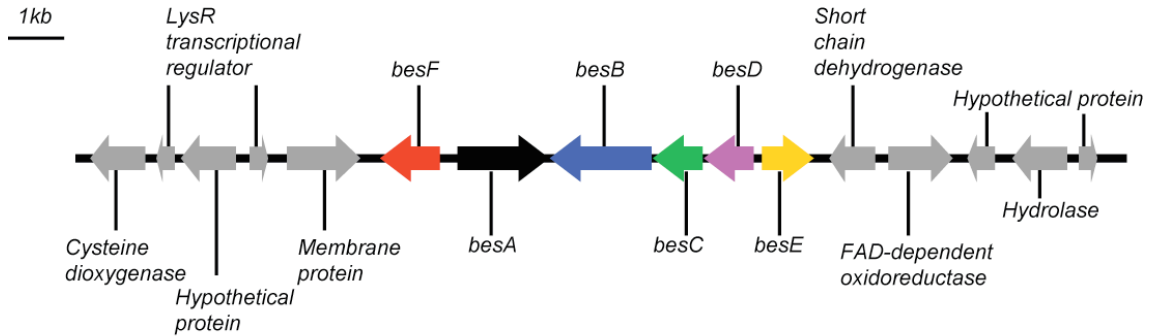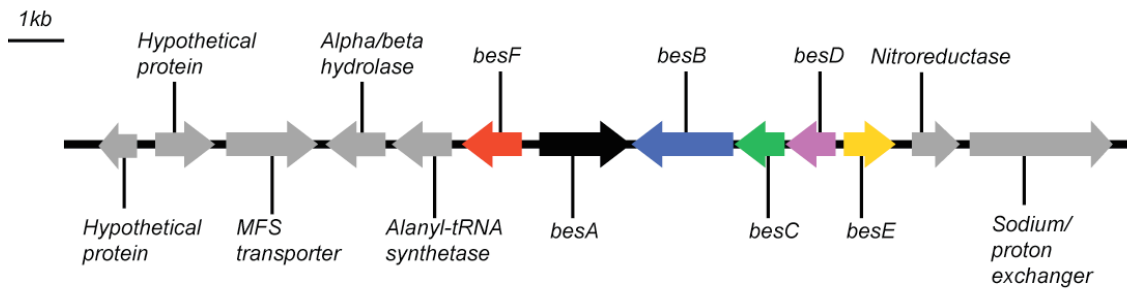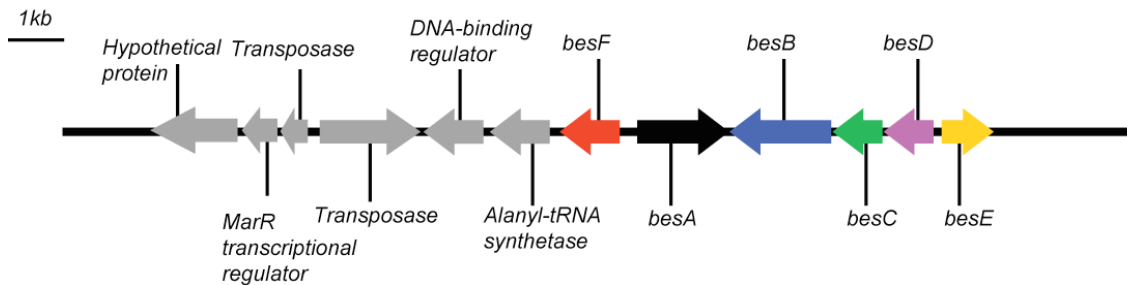
Figure 2.5. Genome clusters and genomic context of βes/Pra biosynthetic clusters (colored) in *S. cattleya, S.* sp. NRRL S-1448*, S. alboverticillatus.* The βes/Pra core cluster contains *besE* and *besF* in addition to *besABCD*, whereas the Pra core cluster contain only *besABCD*. The *S. alboverticillatus besF* is predicted based on partial sequencing of gene from the incomplete genome assembly.

36

Pra clusters (*besABCD*)
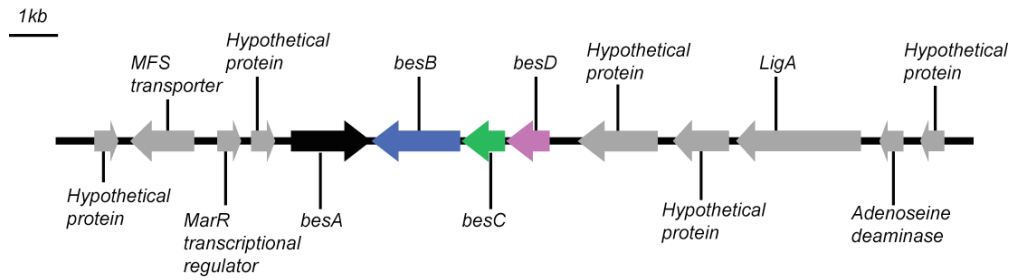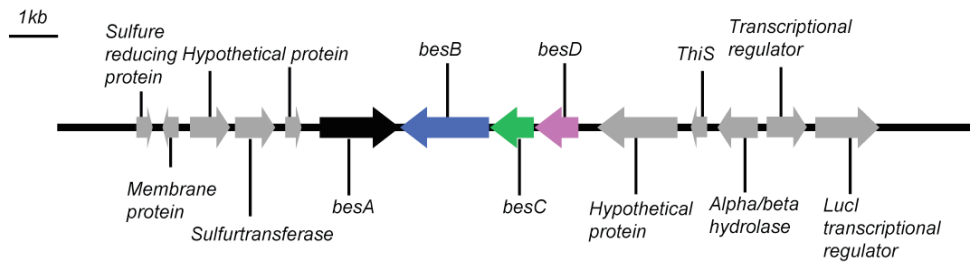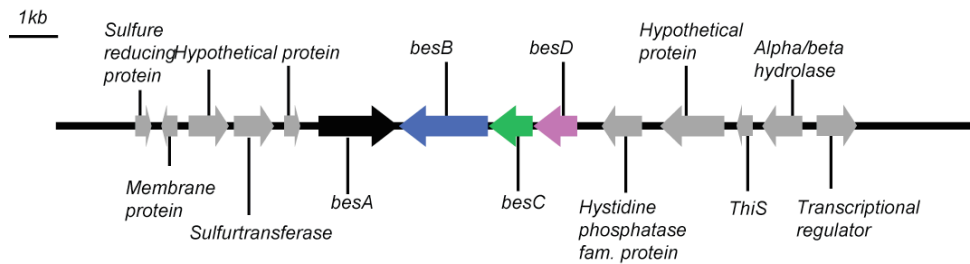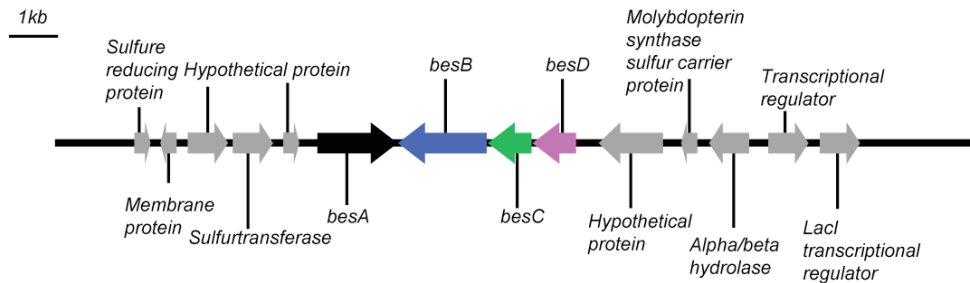
Figure 2.6. Genome clusters and genomic context of Pra biosynthetic clusters (colored) *S. catenulae*, *S. achromogenes, S. lavenduligriseus*, and *S.* sp. NRRL S-31*.*
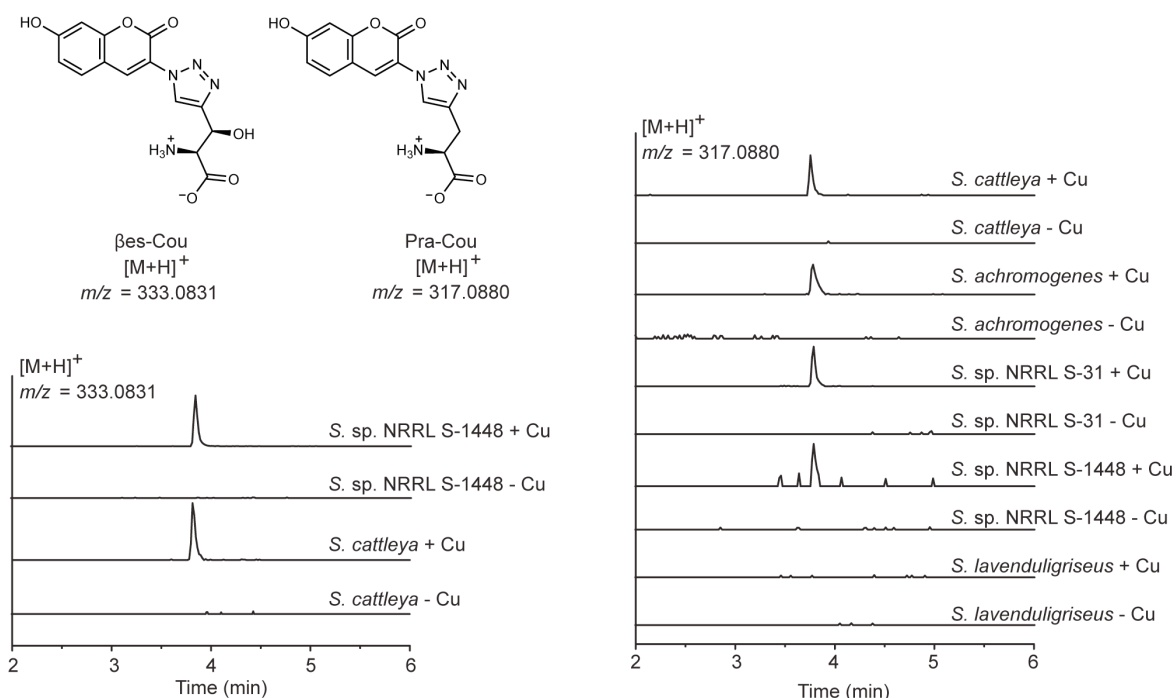
Figure 2.7. Characterization of *Streptomyces* spp. containing the βes/Pra and Pra biosynthetic clusters for their ability to produce terminal alkyne amino acids. The presence of βes and Pra were analyzed by derivatization with 3-azido-7-OH-coumarin by CuAAC and monitoring either the βes-coumarin adduct (βes-Cou; *m/z* = 333.0831) or Pra-coumarin adduct (Pra-Cou; *m/z* = 317.0880) in the presence and absence of Cu(I) using LC/QTOF. The reaction where Cu(I) is omitted provides a negative control since it is required for the cycloaddition to occur. Extracted ion chromatograms of *S. cattleya* and *S.* sp. NRRL S-1448 supernatants showing the production of the βes (βes-Cou; *m/z* = 333.0831) (left). Extracted ion chromatograms of *S. cattleya*, *S. achromogenes*, *S.* sp. NRRL S-31, *S.* sp. NRRL S-1448, and *S. lavenduligriseus* supernatants showing the production of Pra (Pra-Cou; *m/z* = 317.0880; right). *S. lavenduligriseus* was cultured but not found to produce Pra. Attempts to amplify genes in the *bes* cluster failed, suggesting that this particular strain may have lost the *bes* genes. Chromatograms shown are representative of at least 3 experimental replicates.

38
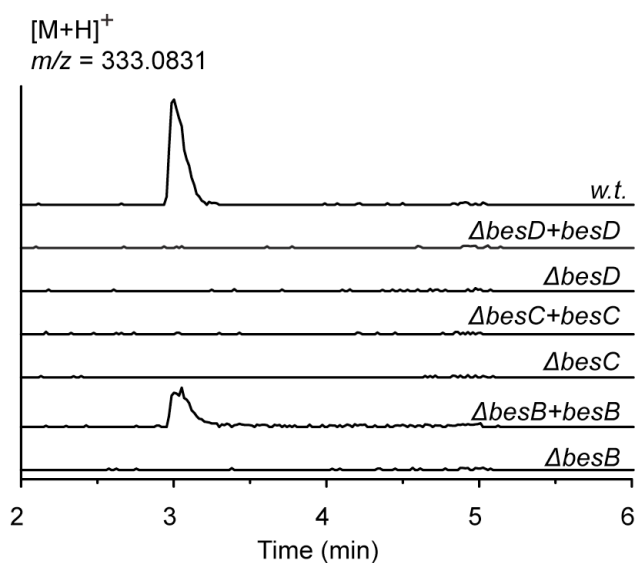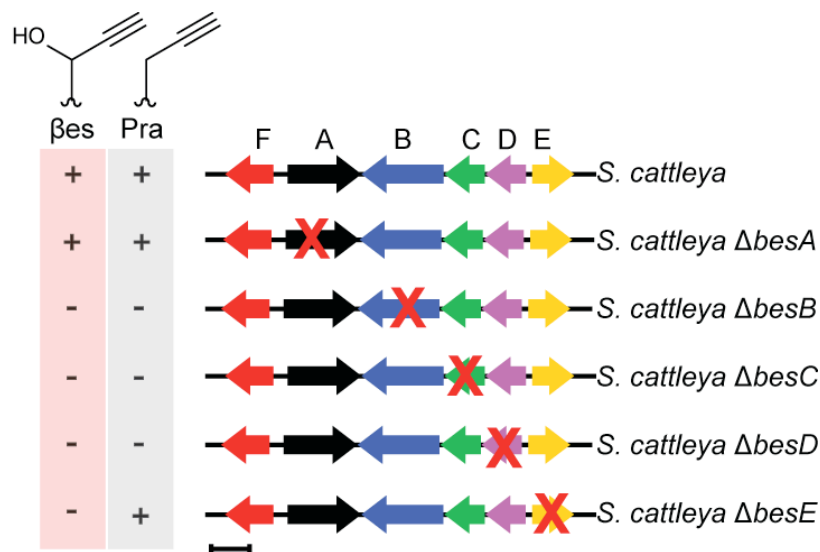
Figure 2.8. (Top) Terminal alkyne production phenotypes of knockouts of each gene in the βes biosynthetic gene cluster from *S. cattleya.* (Bottom) *S. cattleya* knockout strains were complemented with the corresponding wild-type gene expressed from the ErmEp* promoter by insertion into the phage attachment site in the genome. Knockout and complementation strains were cultured for 7 d. The supernatants were derivatized with 3-azido-7-OH-coumarin by CuAAC and analyzed for the βes-coumarin adduct (βes-Cou*; m/z* = 333.0831) by LC/QTOF. Only the *besB* complementation strain appears to recover βes production, possibly due to polar effects related to ordering of the genes in the putative operon (*besD-besC-besB*).
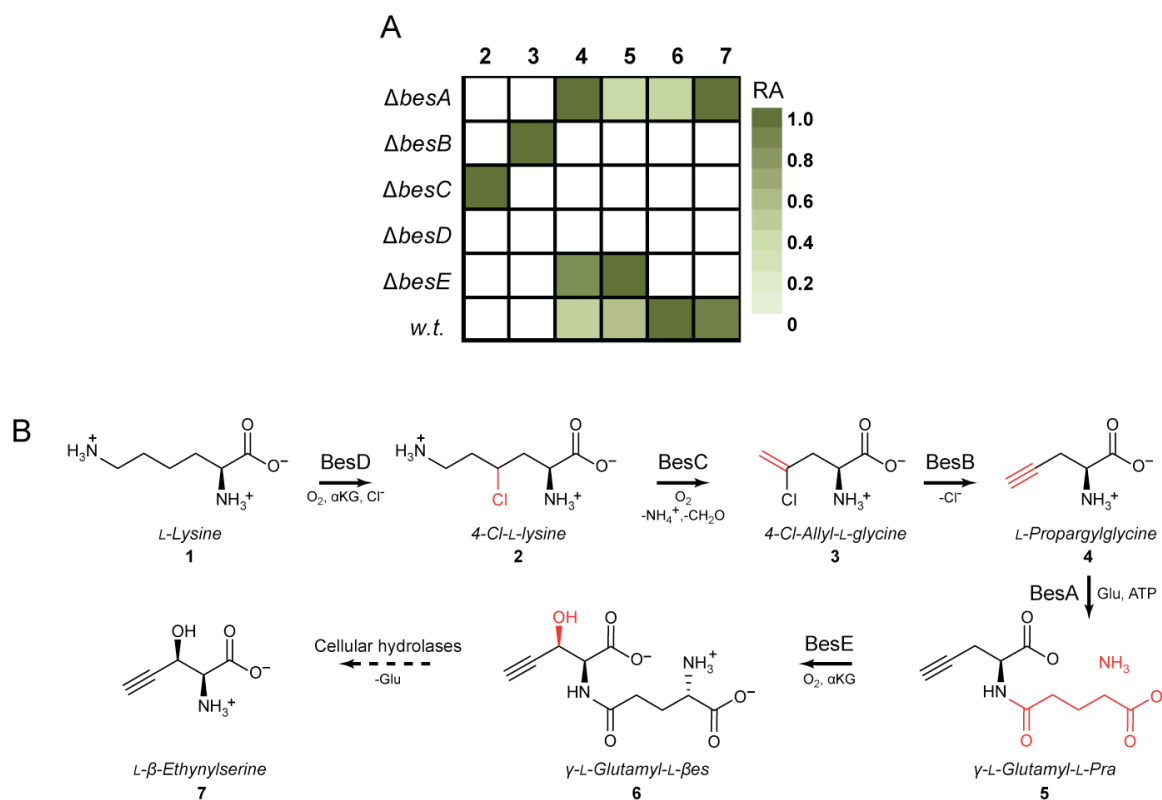
Figure 2.9. Biosynthetic pathway for production of βes from L-lysine. (A) Relative abundance of pathway amino acids found in *S. cattleya* knockouts strains, by average integrated peak area of each compound (n = 5) and normalized by relative abundance (RA) within each species (green, more abundant; white, less abundant). (B) Putative pathway for biosynthesis of β-L-ethynylserine (βes) from L-lysine.

40
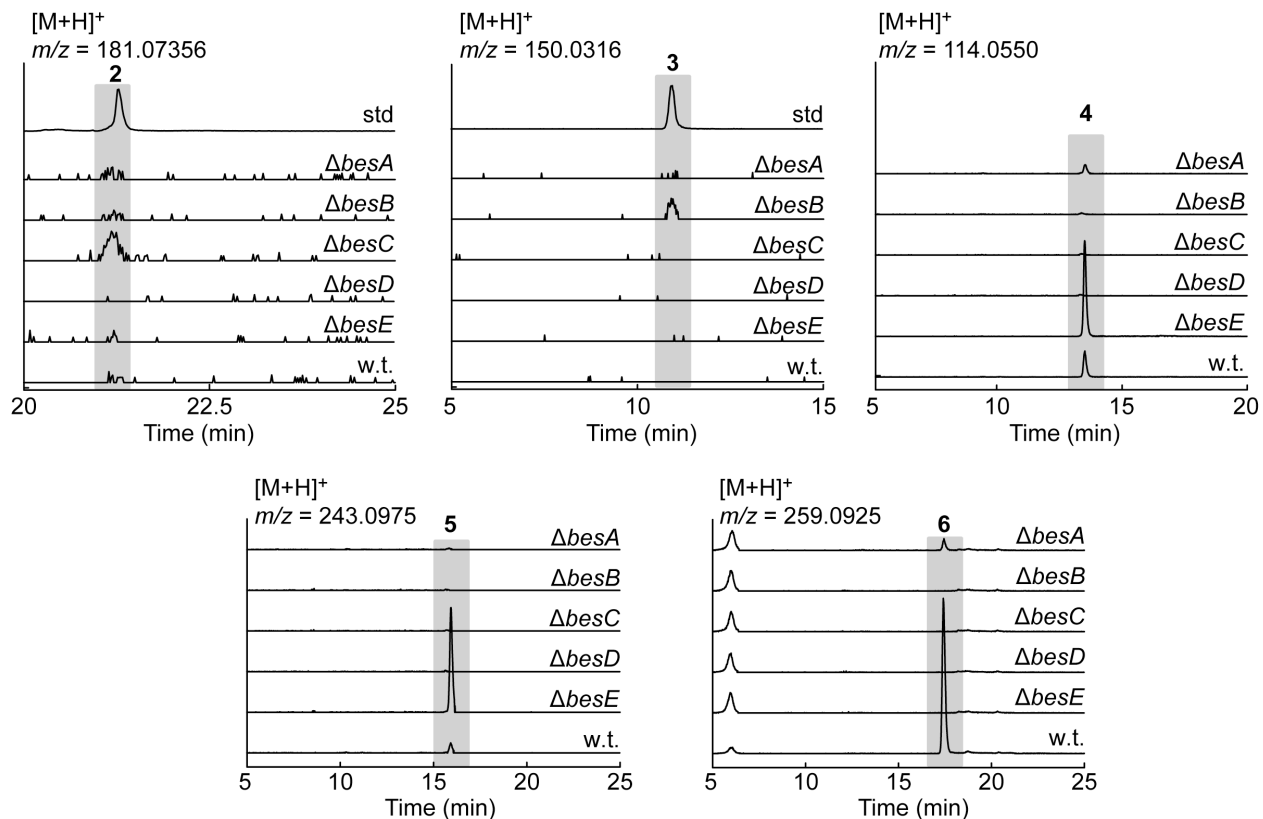
Figure 2.10. Extracted ion chromatograms of pathway intermediates identified using comparative metabolomics from *S. cattleya* pellet extracts for 4-Cl-lysine (2, *m/z* = 181.0738), 4-Cl-allylglycine (3, *m/z* = 150.0316), propargylglycine (4, *m/z* = 114.0550), γ-Glu-Pra (5, *m/z* = 243.0975), and γ-Glu-Bes (6, *m/z* = 259.0925). Chromatograms shown are representative of at least 5 independent biological replicates.

βes biosynthesis is unexpectedly initiated by chlorination of L-lysine (1) at the C_γ position by the BesD halogenase. The resulting 4-chloro-L-lysine (2) is then subject to an unusual oxidative cleavage by the BesC oxidase to form 4-chloro-allylglycine (3). The PLP-dependent enzyme BesB can next catalyze γ-elimination of chloride from this intermediate followed by isomerization to form the first terminal-alkyne product, Pra (4). We propose that Pra is ligated to glutamate by BesA to form a γ-glutamyl-dipeptide (5), which serves as the substrate for the BesE hydroxylase to produce the corresponding βes dipeptide (6) in the last step. Since no other conserved gene was found in these clusters, it is possible that a non-pathway specific peptidyl-transglutaminase or amide hydrolase is responsible for release of free Pra and βes.

BesD is a lysine chlorinase. Inspection of the BesD sequence reveals it to contain the characteristic HXG motif required for halide coordination to the active-site iron in the Fe/αKG-dependent halogenase family, rather than the HXD/E motif of the related family of hydroxylases (*Figure 2.11*) [27]. Indeed when L-lysine was incubated with purified BesD (*Figure 2.12A*), we observed formation of a chlorinated lysine species. Mutating BesD residue G139 to aspartate, involved in the halogenating HXG motif, resulted in loss of formation of Cl-lysine and only formation of OH-lysine (*Figure 2.12B-D)*. When looking at the mass spectra, the major product of the BesD exhibited a characteristic Cl isotopic pattern and could be assigned as Cl-lysine (*Figure 2.13*). At this point, however, the regiochemistry of chlorination was still unknown. In order to determine site of chlorination, we attempted to purify enough Cl-lysine to characterize by NMR. This proved to be a futile effort, as we then discovered that Cl-lysine quickly decomposes to several other compounds.

Halogenase		**H**	**X**	**A/G**			
D	W	H	Q	A	D	T	SyrB2
A	V	H	V	G	N	D	WelO5
G	W	H	W	G	D	F	BesD

Hydroxylase		**H**	**X**	**D/E**			
N	W	H	T	D	V	T	TauD
K	A	H	Y	D	Y	F	3ITQ
F	I	H	Y	D	D	I	BesE

Figure 2.11. BesD is a putative Fe/αKG-dependent halogenase with conserved HXG halogenation motif.

LC/MS analysis of Cl-lysine revealed formation of three degradation products (*Figure 2.14*). We were able to verify that degradation was non-enzymatic by first enzymatically making Cl-lysine and then filtering out His6-BesD using Ni-NTA. Cl-lysine abundance decreased over time, while the three degradation products increased over time. The rate of degradation of Cl-lysine was observed to be slightly higher at higher pH (pH 8) than neutral (pH 7) or slightly acidic (pH 6) (*Figure 2.15A*). Two degradation products had identical mass ($[M+H]^+$ m/z = 145.0975), corresponding to a loss of chloride and a desaturation, while the third degradation product had a mass corresponding to OH-lysine ($[M+H]^+$ m/z = 163.1077). Degradation through an intramolecular carboxylate-mediated cyclization of amino acids had previously been reported [28]. Thus, we propose that degradation of Cl-lysine occurs through an intramolecular nucleophilic attack on the Cγ, leading to formation of a 5-membered lactone (*Figure 2.15B*). The lactone can then either hydrolyze to form 4-OH-lysine, or undergo an ε-amine mediated ring-opening to form a more-stable 7-membered lactam. The lactam also has the possibility of hydrolyzing to form 4-OH-lysine. Based on this proposed degradation mechanism, we subjugated Cl-lysine to O-methyl-esterification to stabilize it for characterization.

To improve signal on the NMR, we used a $^{13}C/^{15}N$-labeled lysine as substrate for chlorination. A modified constant time HSQC ($^{13}C/^1H$) was used to determine the identity of the

Figure 2.12. (A) BesD activity requires L-lysine, Cl⁻, Fe(II) and αKG (4-Cl-lysine, 2; *m/z* 181.0738). Chromatogram shown are representative of at least 3 independent experimental replicates. (B) LC/QTOF traces monitoring production of 4-Cl-lysine (2, *m/z* = 181.0738) in reactions with no enzyme, BesD, or the BesD G139D mutant. Chromatogram shown is representative of 3 experimental replicates. (C) LC/QTOF traces monitoring production of OH-lysine (*m/z* = 163.0738) As discussed previously, a low level of 4-OH-lysine is expected from direct hydroxylation of L-lysine by wild-type BesD, as has been commonly observed in other members of the αKG/Fe-dependent halogenase family. (D) Circular dichroism (CD) spectra of wild-type BesD compared to the G139D mutant suggests that BesD G139D retains a similar fold or secondary structure content (MRE, molar residue ellipticity). Spectra shown is representative of 3 independent replicates.

43
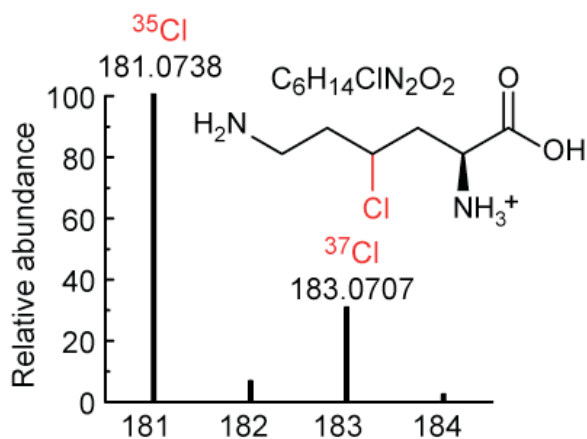
Figure 2.13. Mass spectra of product formed when BesD is incubated with L-lysine and Cl⁻ showing the characteristic chlorine isotope pattern. Spectra shown is representative of at least 3 experimental replicates.
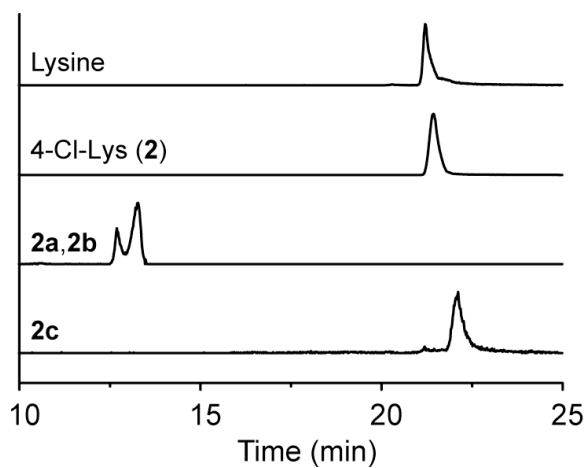
Figure 2.14. Extracted ion chromatograms of the decomposition products observed upon halogenation of L-lysine by BesD by LC/QTOF. (Lysine, *m/z* = 147.1128; 4-Cl-lysine (2), *m/z* = 181.0738; ε-lactam (2a)/γ-lactone (2b), *m/z* = 145.0972; 4-OH-lysine (2c), *m/z* = 163.1077). Chromatogram shown is representative of at least 3 experimental replicates.

44
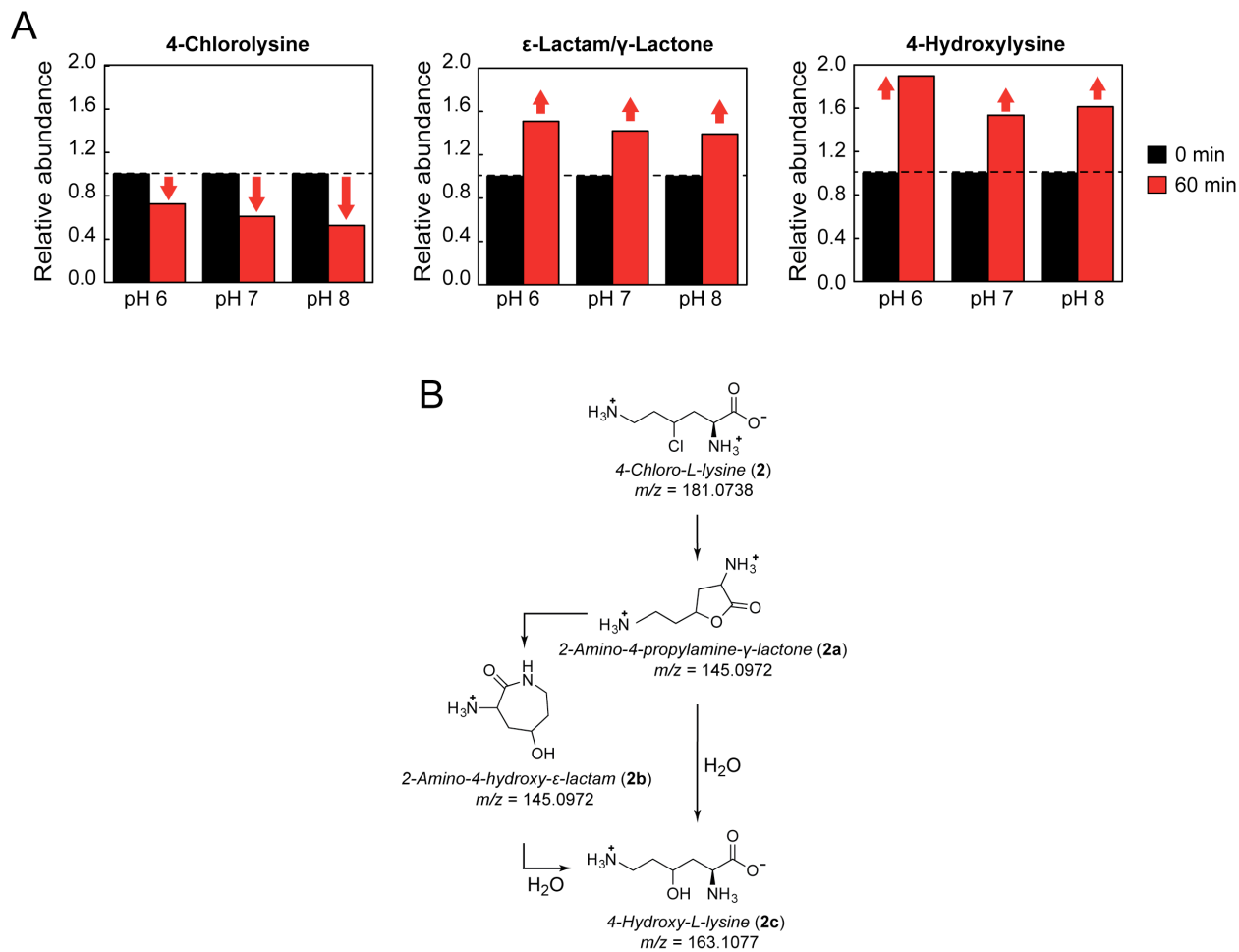
Figure 2.15. (A) Integrated EICs for each of the observed products at pH 6, 7, and 8 following removal of BesD by filtration (Pall Nanosep filter, 10 kDa MWCO). Data are plotted relative to lysine, which is taken to be constant following enzyme removal. After 1 h, Cl-lysine decreases, while the lactam, lactone, and OH-lysine products increase in abundance. (B) 4-Cl-lysine (2) is proposed to undergo an intramolecular nucleophilic attack forming 2-amino-4-propylamine-γ-lactone with an estimated $t_{1/2}$ = 1 h [38]. The γ-lactone is also unstable can form either the 2-amino-4-hydroxy-ε-lactam or 4-OH-lysine. 4-OH-lysine could also arise from direct hydroxylation of L-lysine by BesD, as a low level of hydroxylation is commonly observed in other members of the Fe/αKG-dependent halogenase family.

Cα and Cε, which contain only one adjacent ^{13}C. HCCH-COSY was then used to determine connectivity of the identified carbons. Based on NMR analysis of the methyl-esterified product, the chlorination was determined to occur at the 4-position (*Figure 2.156-D*). Subsequent experiments were performed on the ^{13}C/^{15}N lactam degradation product to confirm we had identified the correct degradation pathway. In addition to a constant time HSQC and HCCH-COSY, we also performed ^{15}N/^{1}H HSQC, ^{13}C/^{1}H HNCO, and ^{13}C/^{1}H HNCA (*Figure 2.17A-E*).

BesC is a Cl-lysine oxidase. In addition to 4-Cl-lysine, we identified a second chlorinated metabolite from comparative metabolomics studies, which had a mass consistent with a loss of –CH$_5$N from 4-Cl-lysine. This compound accumulated in the Δ*besB* strain but was not observed in the Δ*besC* strain (*Figure 2.9*)*,* suggesting that BesC, a putative non-heme dinuclear Fe enzyme (*Figure 2.18*), is the most likely candidate for its production. We thus propose that BesC catalyzes an unusual oxidative C-C bond-cleaving reaction to form a five-carbon terminal-alkene amino acid, 4-Cl-allygllycine (**3**) with eventual release of the methylamine fragment as ammonia and formaldehyde (*Figure 2.19*). To further characterize BesC, we tested the ability of the purified enzyme to produce 4-Cl-allylglycine (**3**) *in vitro* from purified 4-Cl-lysine or coupled reaction with BesD and L-lysine (*Figure 2.20*). The BesC product was confirmed to be **3** by comparison to a commercial standard, and contained a characteristic Cl isotope pattern (*Figure 2.21*).

In investigating BesC, we also explored off-pathway reactions that could be occurring *in vivo*. Specifically, we tested the ability of BesC to oxidize L-lysine to form L-allylglycine (*Figure 2.22*). This reaction worked well *in vitro*, though subsequent competition experiments in BesD-coupled assays showed that Cl-lysine is the preferred substrate for BesC. *In vivo*, we noted that that L-allylglycine was also produced though detection proved to be difficult since L-proline, an abundant amino acid of identical mass, elutes at a similar retention time on the SeQuant ZIC-pHILIC (5 μm, 2.1 × 100 mm; EMD-Millipore) columns used for amino acid analysis.

In order to investigate the fate of the Nε, we developed an assay to capture Nε released as ammonia. Glutamate dehydrogenase (GDH) is an enzyme that catalyzes the transamination of αKG to glutamate, using free ammonia [29]. Since the ammonia being released in the BesD reaction would be isotopically labeled, as ^{15}N, we could then directly observe its incorporation as the Nα in glutamate by mass spectrometry. In a single-pot reaction, we reacted ^{13}C/^{15}N lysine with BesD, BesC, and GDH. Subsequent LC/MS analysis revealed clear mass spectra shift of glutamate, from *m/z* = 148.0603 to *m/z* = 149.0578, confirming ^{15}N was indeed incorporated (*Figure 2.23 A-D*).

A similar LC/MS-based capture assay was developed to trap formaldehyde released in the BesC reaction. Here, fluoral-P was used as a trapping reagent which takes two molecules of fluoral-P and one molecule of formaldehyde to form 3,5-diacetyl-1,4-dihydrolutidine (DDL) and releasing ammonia [30]. Again, using a coupled system we were able to generate ^{13}C/^{15}N 4-Cl-lysine *in situ* using BesD, then added BesC and fluoral-P to capture formaldehyde. LC/MS was then used to detect ^{13}C capture into DDL, confirming that the Cε is released as a byproduct of BesC catalysis (*Figure 2.24*). Both release of ammonia and formaldehyde was found to be stoichiometric with terminal-alkene amino acid formation (*Figure 2.25A & 2.25B*). Taken together, the data show that BesC catalyzes alkene formation with release of formaldehyde and ammonia as the final co-products of the reaction.
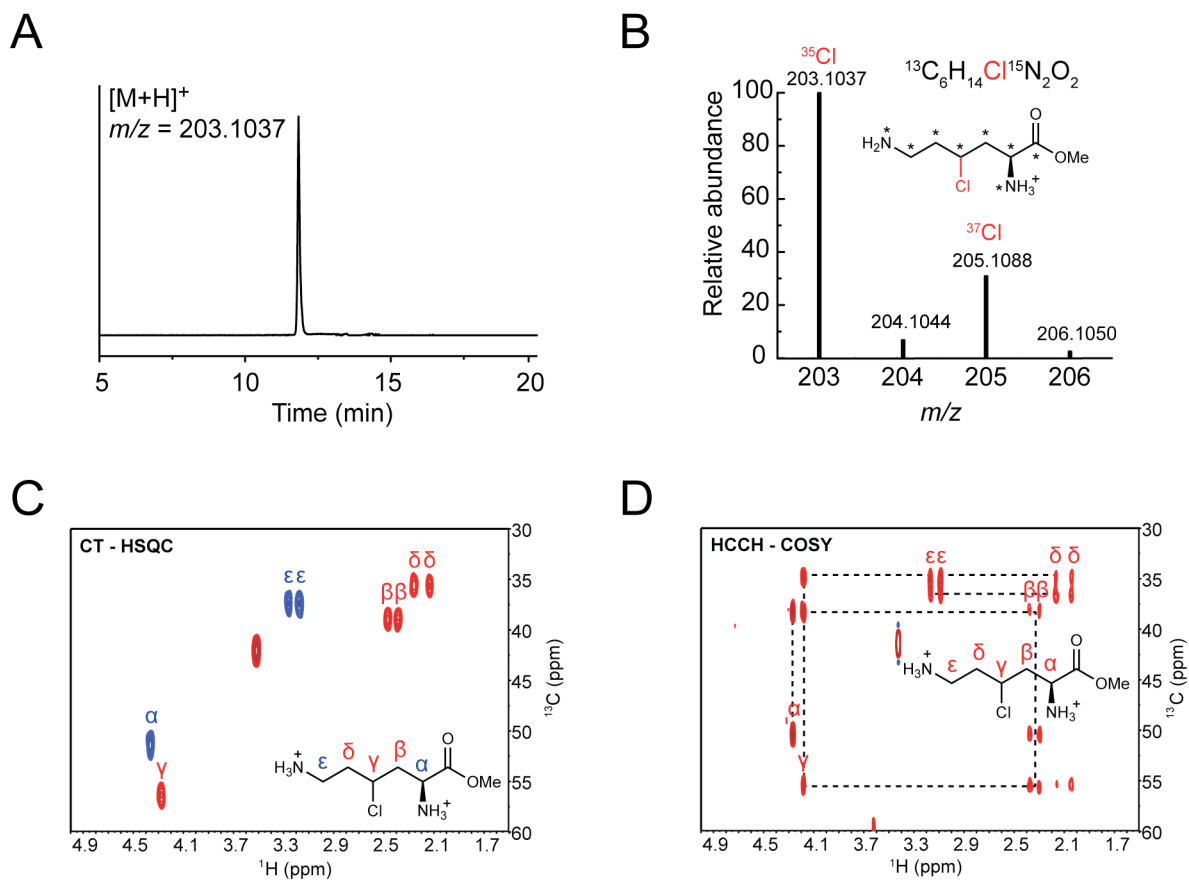
Figure 2.16. (A) BesD was incubated with fully ^{15}N- and ^{13}C-labeled L-lysine, Fe, and αKG for 20 min before quenching with methanolic HCl (3M) to yield the [^{15}N$_2$, ^{13}C$_6$]-4-Cl-lysine methyl ester product. Chloride was provided as NaCl in the enzyme storage buffer. We found that derivatization was necessary in order to avoid intramolecular lactam formation. The derivatized BesD product was then isolated by HPLC and characterized. Extracted ion chromatograms of the [^{15}N$_2$, ^{13}C$_6$]-4-Cl-lysine (*m/z* = 203.1037) produced by BesD. Chromatogram shown is representative of 3 independent experimental replicates. (B) Mass spectra show the characteristic Cl isotope pattern for [^{15}N$_2$, ^{13}C$_6$]-4-Cl-lysine methyl ester produced by BesD. (C) 2D ^{1}H-^{13}C CT-HSQC was used to confirm assignment of C$_\alpha$ and C$_\varepsilon$ of the [^{15}N$_2$, ^{13}C$_6$]-4-Cl-lysine methyl ester produced by BesD. In this experiment carbons with two neighbors (red) have opposite phase of carbons with one or three neighbors (blue). (D) 2D ^{1}H-^{13}C HCCH COSY was used to assign connectivity of carbons of the [^{15}N$_2$, ^{13}C$_6$]-4-Cl-lysine methyl ester produced by BesD.

47
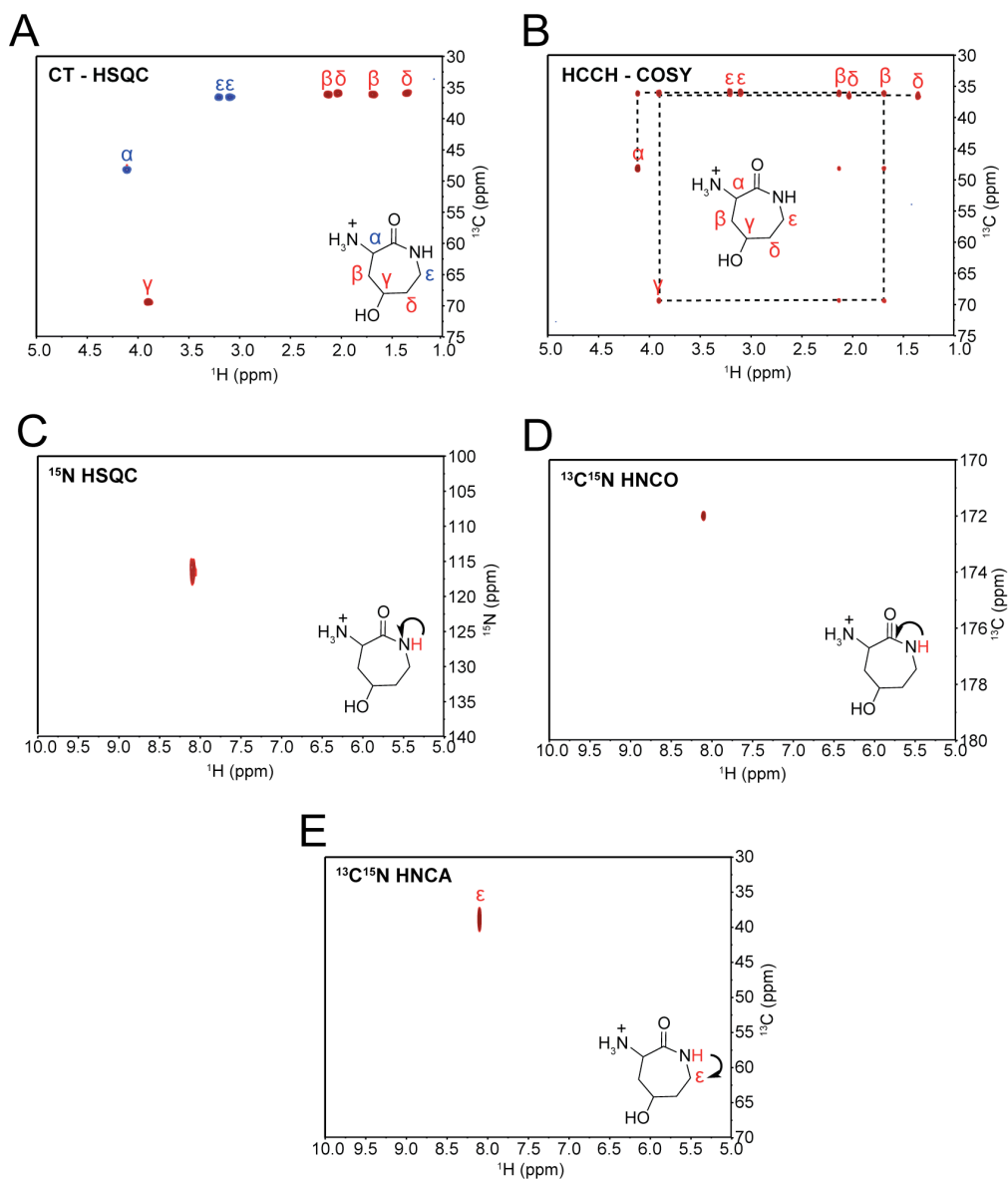
Figure 2.17. (A) For obtaining the lactam structure, BesD was incubated with fully ^{15}N and ^{13}C-labeled L-lysine, Fe, and αKG for 16 h before quenching with 1% (*v/v*) formic acid in methanol to yield the [^{15}N$_2$, ^{13}C$_6$]-lactam as the major product. Chloride was provided as NaCl in the enzyme storage buffer. NMR analysis of the ε-lactam was carried out after HPLC purification. 2D ^{1}H-^{13}C CT-HSQC with a 13.6 ms ^{13}C evolution period was used to confirm assignment of C$_\alpha$ and C$_\epsilon$. In this experiment carbons with two neighbors (red) have opposite phase of carbons with one or three neighbors (blue). (B) ^{1}H-^{13}C HCCH COSY was used to assign backbone connectivity of carbons for the lactam. (C) The ^{1}H-^{15}N HSQC spectrum confirms the presence of the amide in the lactam structure, with reported chemical shift typical of amides (δ 6 - 9 ppm). (D) The 2D ^{1}H-^{13}C HNCO spectrum confirms that the epsilon N is adjacent to a carbonyl species. (E) The 2D ^{1}H-^{13}C HNCA spectrum confirms that the epsilon N is coupled to C$_\epsilon$, forming a cyclic structure.

48
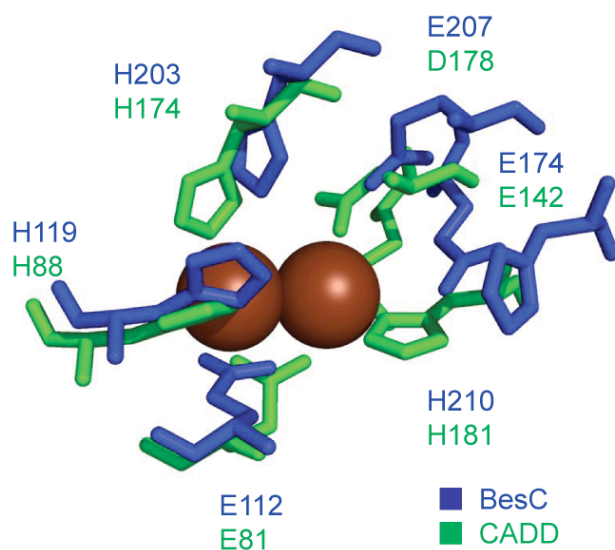
Figure 2.18. BesC shows homology to non-heme diiron proteins such as the Chlamydia protein Associated with Death Domains (CADD, 34% sequence similarity[39]) and stearoyl-ACP desaturase (26% sequence similarity[40]). Overlay of the predicted homology model of BesC (blue) with the CADD crystal structure (green)[41] reveals six conserved putative Fe binding residues in the active site. Putative iron ligands shown in brown. Homology model was generated using Phyre 2.
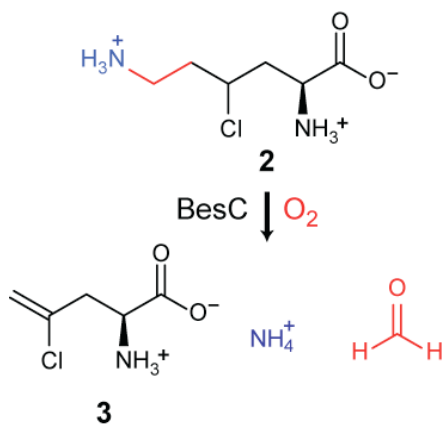
Figure 2.19. Reaction catalyzed by BesC, forming ammonia and formaldehyde as co-products from oxidative cleavage between C_δ and C_ε.

49
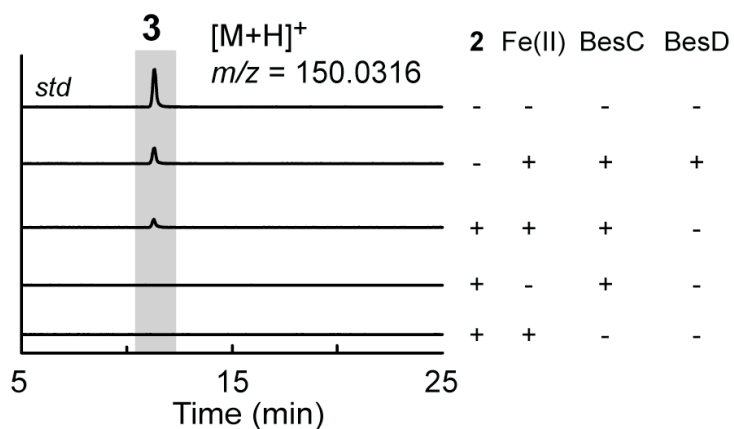
Figure 2.20. BesC catalyzes formation of 4-Cl-allylglycine from 4-Cl-lysine and requires Fe(II) as a cofactor (4-Cl-allylglycine, 3; *m/z* 150.0316). Chromatograms shown are representative of at least 3 experimental replicates.
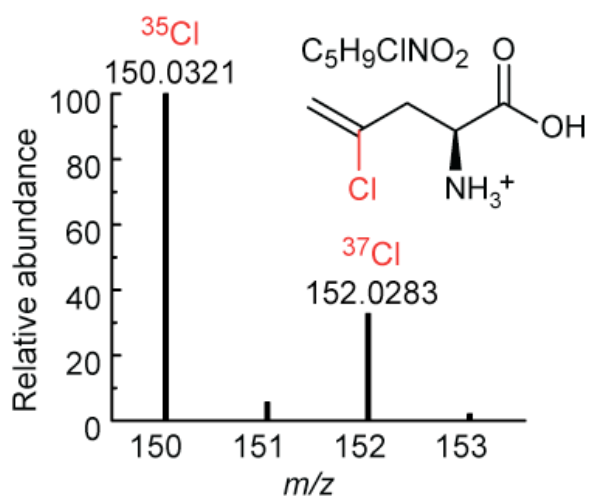
Figure 2.21. Mass spectra of product formed when BesC is incubated with 4-Cl-lysine. Spectra shown is representative of at least 3 experimental replicates.
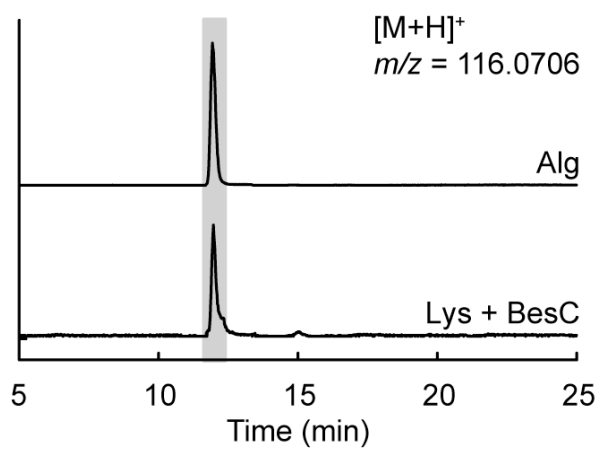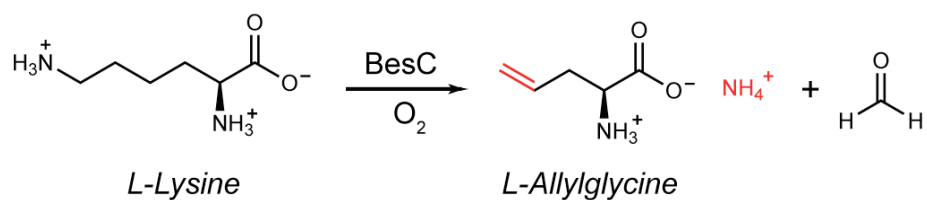
Figure 2.22. BesC is also competent to react with ʟ-lysine directly to produce allylglycine *in vitro*. LC/QTOF traces show allylglycine produced by BesC compared to a synthetic standard (Alg, *m/z* = 116.0706). Chromatogram shown is representative of 3 experimental replicates.
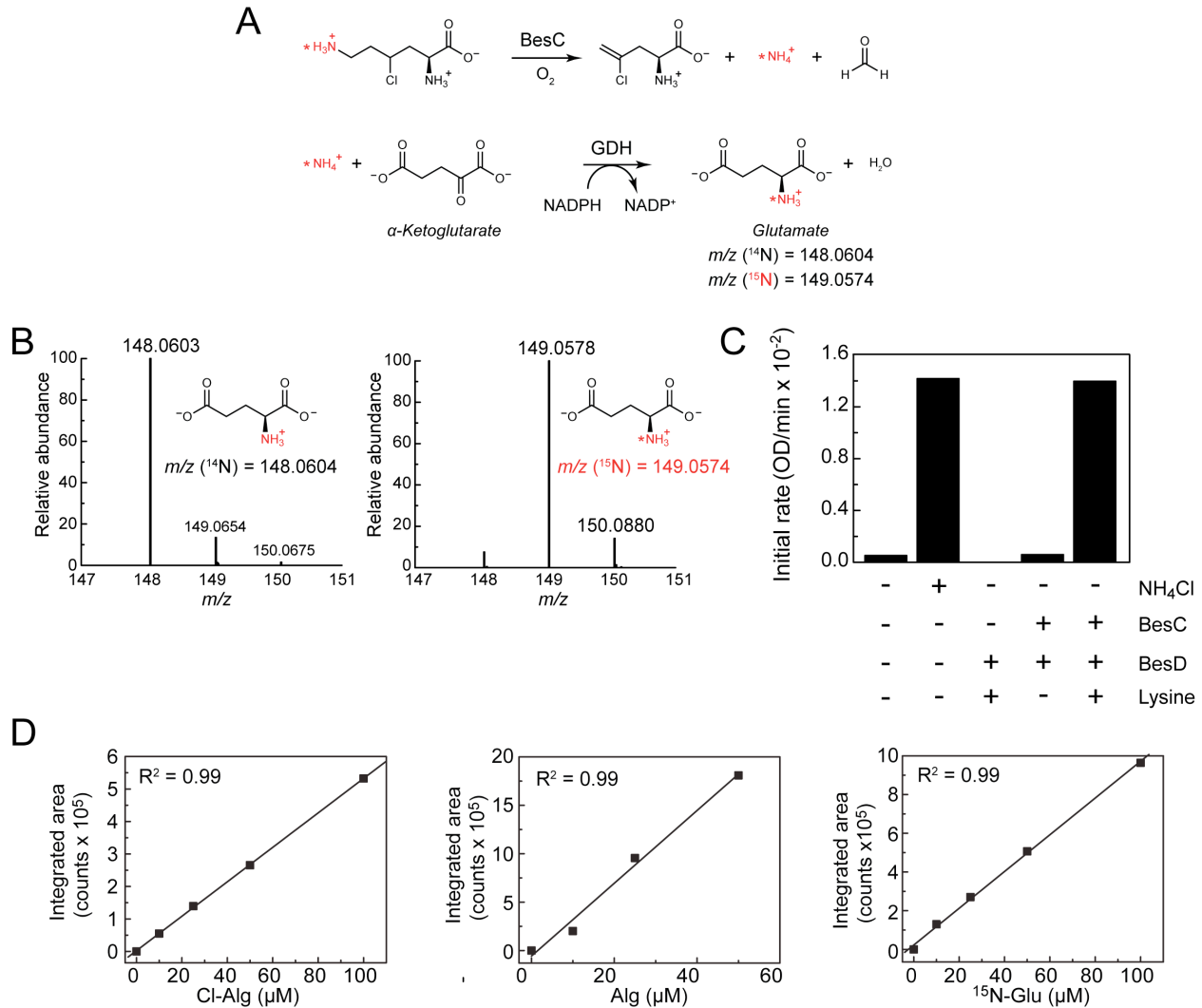
Figure 2.23. (A) The BesC reaction is run as a coupled reaction with BesD starting from L-lysine, αKG, and chloride. To detect the ammonia co-product, we employed glutamate dehydrogenase (GDH), which produces glutamate from ammonia and αKG with concurrent oxidation of NADPH. (B) The release of ammonia from 4-Cl-lysine is monitored through incorporation of the nitrogen from the ^{15}N-labeled lysine substrate into glutamate using LC/QTOF. Positive ionization mass spectrum showing formation of ^{14}N-glutamate or [15]N- L-glutamate when unlabeled L-lysine (left) or $[^{15}N_2, {}^{13}C_6]$-L-lysine (right) are used as the substrate, respectively. Spectra shown are representative of 3 experimental replicates. (C) Consumption of NADPH was observed spectrophotometrically. Reactions contained BesC, BesD, lysine, αKG, sodium ascorbate, Fe(II), and chloride as indicated. After 1 h, NADPH and glutamate dehydrogenase were added and NADPH consumption was monitored by A_{340}. (D) Standard curves constructed by integrating extracted ion counts for the relevant species by LC/QTOF. R^2 value shown is determined by ordinary least squares.
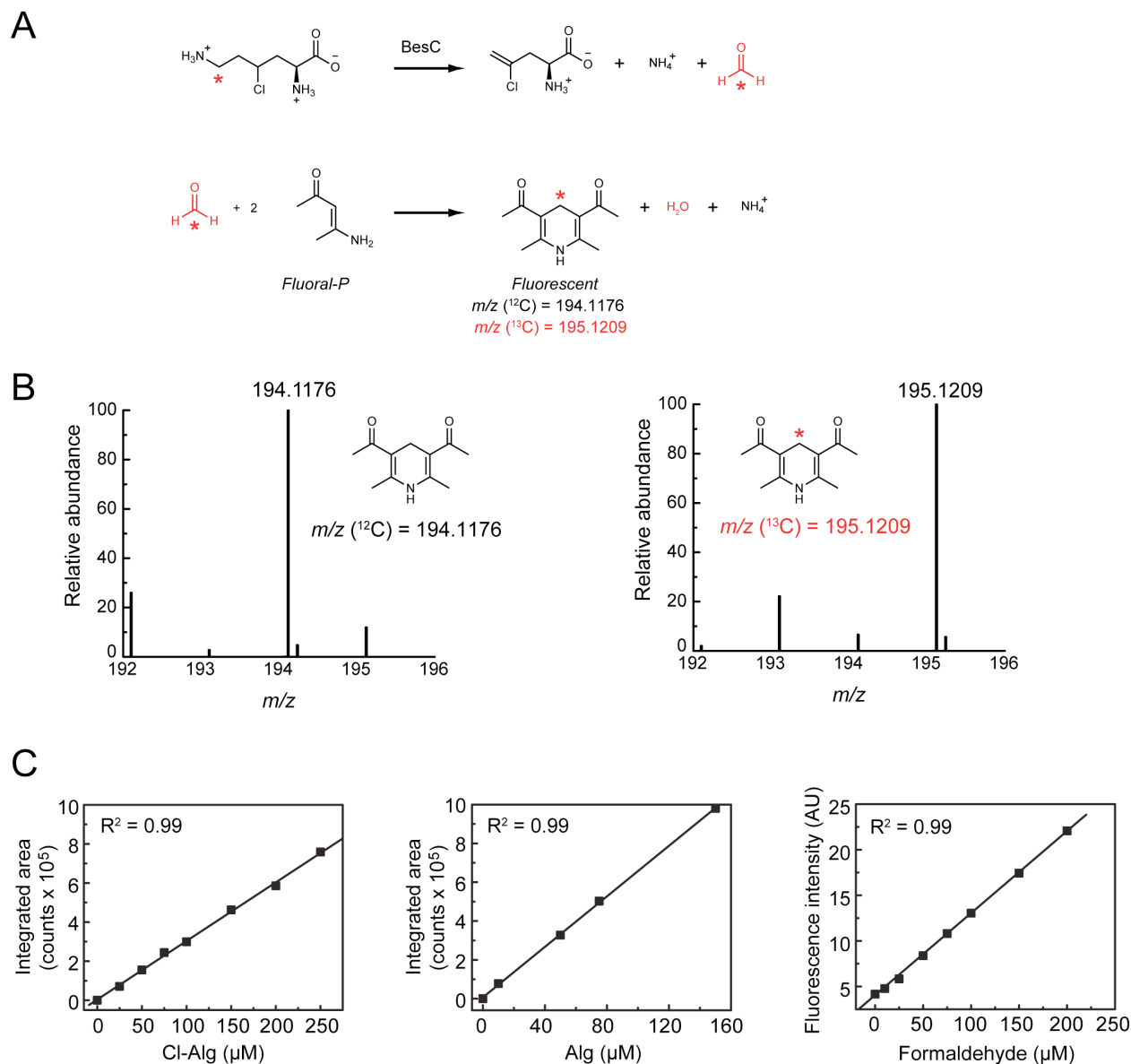
Figure 2.24. Detection of formaldehyde as a co-product of the BesC reaction. The BesC reaction is run as a coupled reaction with BesD starting from L-lysine, αKG, and chloride. (A) To derivatize the putative formaldehyde co-product, we performed the enzymatic reaction in the presence of Fluoral-P. The release of formaldehyde from 4-Cl-lysine was monitored through incorporation of the carbon from ^{13}C-labeled lysine substrate into 3,5-diacetyl-1,4-dihydro-2,6-lutidine (DDL). (B) Positive ionization mass spectrum showing formation of ^{12}C-DDL or ^{13}C-DDL when unlabeled L-lysine (left) or [[15]N$_2$, [13]C$_6$]-L-lysine (right) are used as the substrate, respectively. Spectra shown are representative of 3 experimental replicates. (C) Standard curves for the relevant species were generated by integration of the extracted ion chromatogram from LC/TOF (Cl-Alg, Alg) or by fluorescence quantification (formaldehyde). R^2 value shown was determined by ordinary least squares.
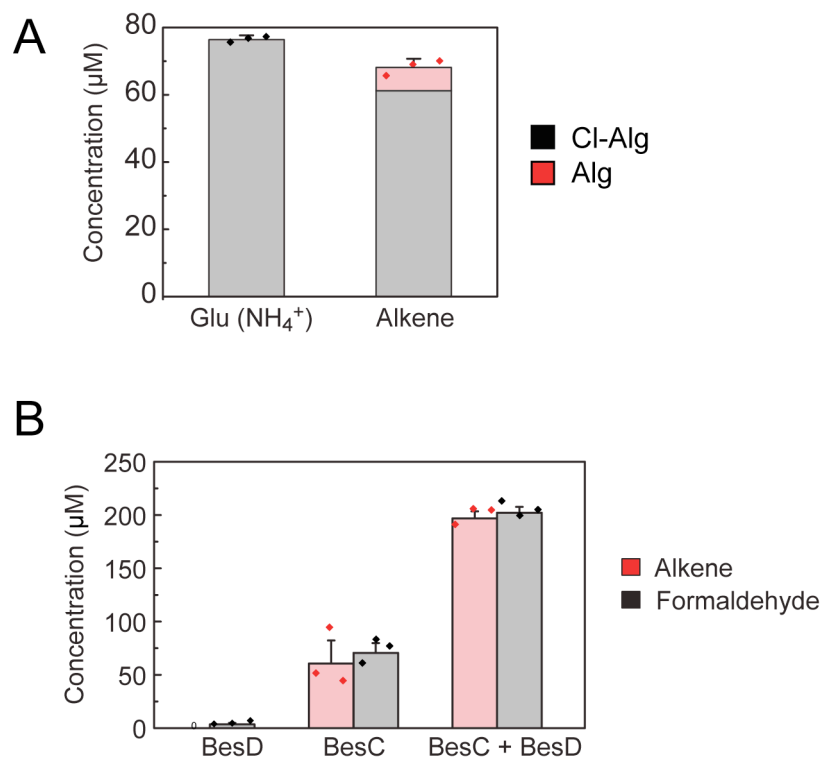
Figure 2.25. (A) Quantification by LC/QTOF showing stoichiometric ammonia release measured as glutamate (Glu) compared total alkene product (4-Cl-allylglycine, Cl-Alg; allylglycine, Alg) in a coupled assay of BesD and BesC with L-lysine using L-glutamate dehydrogenase. Data are mean ± s.d. (n = 3 independent experiments). (B) Quantification of formaldehyde produced from Fluoral-P reaction with 5 mM L-lysine and purified BesD, BesC, or both BesC and BesD. The resulting product was quantified by fluorescence and compared to alkene product formation measured by LC/QTOF and found to be stoichiometric. Data are mean ± s.d. (n = 3 independent experimental replicates).

BesB is a PLP-dependent acetylenase. We next turned our attention to formation of the terminal-alkyne functional group of Pra. Based on the accumulation of **3** upon deletion of the *besB* gene (*Figure 2.9A*), the PLP-dependent BesB enzyme was assigned as the acetylenase that catalyzes the γ-elimination of chloride from **3** to achieve substrate desaturation. *In vitro* assays with purified BesB show both that Pra can be produced either directly from a 4-Cl-allyglycine standard or by a coupled assay with BesCD and that the presence of a γ-halogen as a leaving group is required (*Figure 2.26A & 2.26B*).
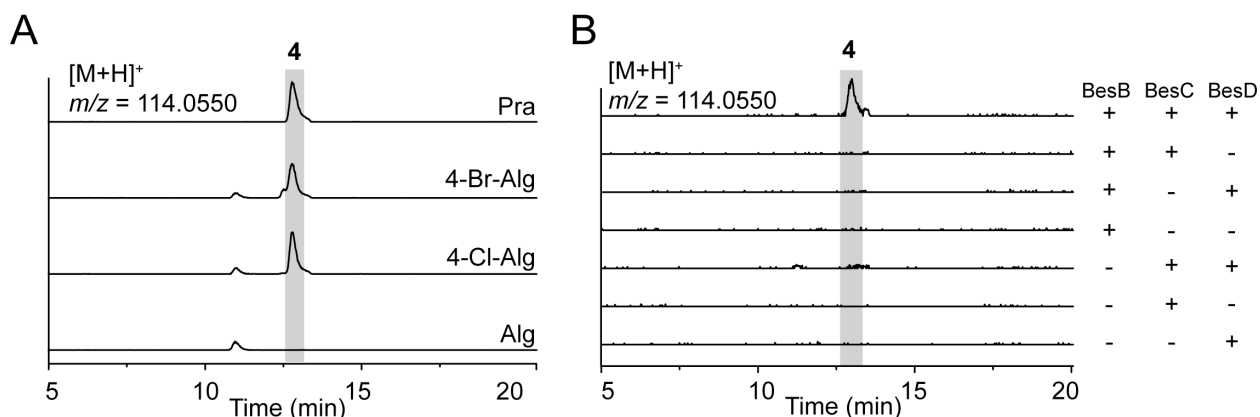
Figure 2.26. *In vitro* characterization of BesB and proposed mechanism of terminal alkyne formation. (A) LC/QTOF analysis of *in vitro* reactions containing BesB, PLP, and substrate. When the substrate is 4-Cl-Alg or 4-Br-Alg, propargylglycine (4, *m/z* = 114.0550) is observed. However, no propargylglycine is observed if allylglycine is used as a substrate. Chromatograms shown are representative of at 3 independent experimental replicates. (B) *In vitro* reconstitution of propargylglycine (4, *m/z* = 114.0550) formation from L-lysine with purified BesB, BesC, and BesD and Fe(II), αKG, chloride, PLP, and ascorbate monitored by LC/QTOF. Chromatograms shown are representative of at 3 independent experimental replicates.

As an initial probe of BesB mechanism, we used high-resolution mass spectroscopy to measure deuterium exchange along the backbone of the substrate. Using 4-Cl-allylglycine (**3**) as a substrate in D$_2$O resulted in formation of [M+2D]$^+$ Pra, suggesting that at least two deprotonation-protonation events occur during the course of BesB catalysis (*Figure 2.26A*). Our model would be consistent with a canonical cystathionine β-lyase/cystathionine γ-synthase mechanism where C$_\alpha$-deprotonation leads to eventual formation of a 4-Cl-allylglycine-ketimine intermediate (*Figure 2.27B*, *Figure 2.28*) [31]. A second deprotonation on the amino acid at C$_\beta$ would initiate elimination of chloride to form an allene intermediate, which could subsequently isomerize to the terminal alkyne before eventual conversion back to aldimine and release from PLP. Importantly, observing the proton exchange at two carbons eliminates a 1-3-hydride shift as the mechanism for terminal alkyne formation from a terminal allene. Instead, we think it is likely that an enzyme-meditated isomerization occurs. This proposed mechanism is based on previous studies of Pra as a mechanism-based inhibitor for PLP enzymes based on its ability to isomerize to the corresponding allene [32,33]. In order to gather more evidence supporting the putative mechanism for BesB, we attempted various strategies for detecting any of the proposed intermediates.
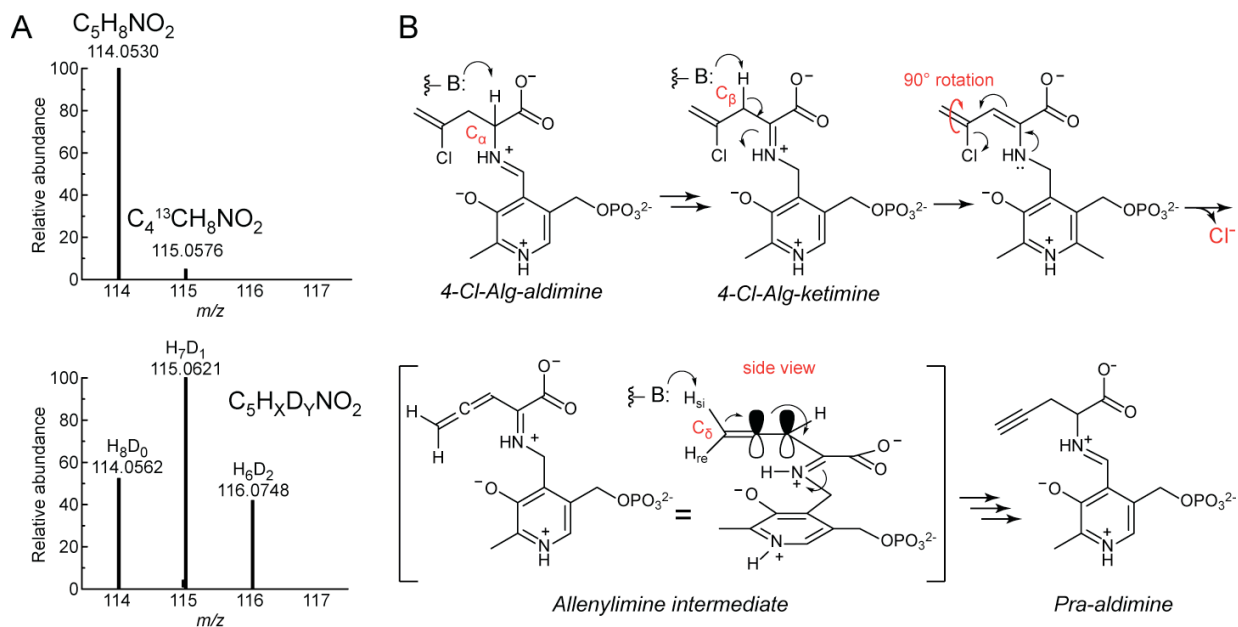
Figure 2.27. (A) Mass spectra showing deuterium incorporation at two positions of Pra (**4**) when BesB reaction is carried out in D_2O compared to H_2O. (top) Extracted ions for the BesB reaction run in H_2O. (Expected *m/z*: $C_5H_8NO_2$ 114.0550 [100%], $C_4{}^{13}CH_8NO_2$ 115,0580 [5.9%]; Observed *m/z*: 114.0530 [100%], 115.0576 [4.8%]). (bottom) Extracted ions for the BesB reaction run in D_2O (Observed *m/z*: 114.0562 [52.3%], 115.0558 [3.2%, not labeled for clarity], 115.0621 [100%], 116.0748 [41.8%]). Spectra shown is representative of 3 experimental replicates. (B) Putative mechanism for BesB-catalyzed formation of Pra (**4**) from 4-Cl-allyglycine (**3**). Formation of a covalent external aldimine adduct between PLP and 4-Cl-allylglycine followed by C_α deprotonation and protonation of the quinonoid intermediate (not shown) is proposed to lead to formation of the initial ketimine intermediate. C_β deprotonation results in elimination of chloride upon 90° bond rotation to form an allene intermediate. Subsequent C_δ deprotonation could then initiate isomerization of the allene to form the terminal alkyne. Following deprotonation of the PLP-bound alkyne to regenerate the quinonoid form (not shown), reprotonation of C_α reforms the external aldimine, from which Pra can be released as the free amino acid (not shown).

56
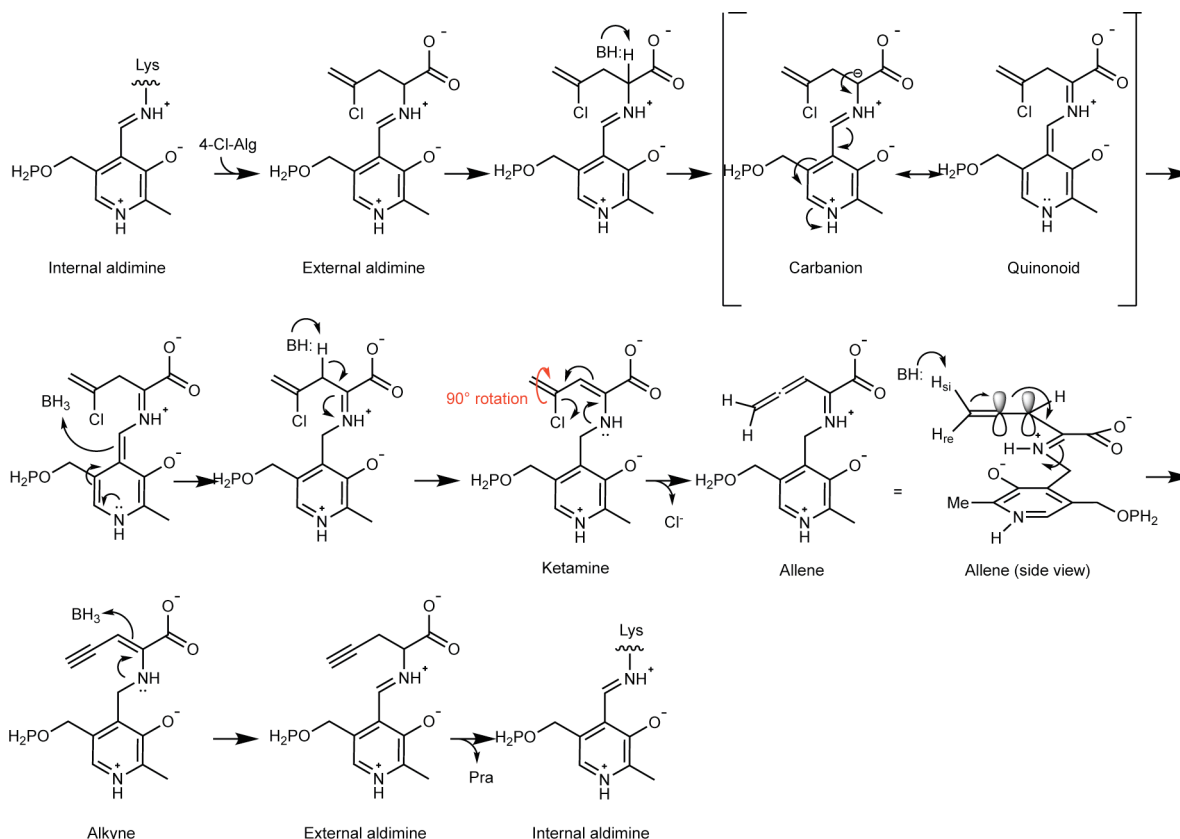
Figure 2.28. A mechanistic model for BesB is proposed in analogy to cystathionine γ-synthase and related PLP mechanisms that form allene-intermediates[42]. In the first step, the external 4-Cl-allylglycine-aldimine is formed. Deprotonation at the C_α-position yields the 4-Cl-allylglycine-carbanion or quinonoid or intermediate. Deprotonation at the C_β-position produces the 4-Cl-allylglycine-ketamine intermediate. Elimination of chloride can then occur upon a 90° rotation to break planarity/aromatization to for a proposed terminal-allene intermediate. Based on work using Pra as a mechanistic inhibitor of PLP enzymes, isomerization between the terminal alkyne and allene can occur. At this time, we suggest that the reverse can occur to convert the PLP-bound allene to the terminal alkyne, yielding a PLP-bound Pra intermediate. The C_δ deprotonation to initiate this step is facilitated by overlap of the C_δ-H bond with the conjugated C_β-C_γ and PMP-imine π-bonds. Subsequent steps then allow for regeneration of the external aldimine with Pra, which is then released from the cofactor.

57

Our proposed mechanism for BesB involves multiple imine-containing intermediates, which could potentially be trapped and directly detected by LC/MS. Previous mechanistic studies into PLP-enzyme intermediates have used sodium cyanoborohydride (NaBH₃CN) as an imine-selective reducing agent (*Scheme 2.1*) [34]. By quenching BesB reactions with NaBH₃CN, we were able to directly detect three distinct PLP species (*Figure 2.29*): free PLP (as pyridoxal-5-phosphate), Cl-Alg-ketamine (reduced form of aldimines, ketamine, or quinonoid species bound to substrate), and Pra-ketamine (reduced form of aldimine, ketamine, or quinonoid species bound to product). Interestingly, each of these compounds had an additional MS signature that corresponded to fragmentation of PLP-intermediate at the α-amine, with the charge migrating to the PLP fragment. We next tried to use this additional MS signature to distinguish between various forms of the PLP-imines. Our proposed mechanism included two forms of imines, one where the imine double bond is formed between the α-nitrogen and α-carbon and
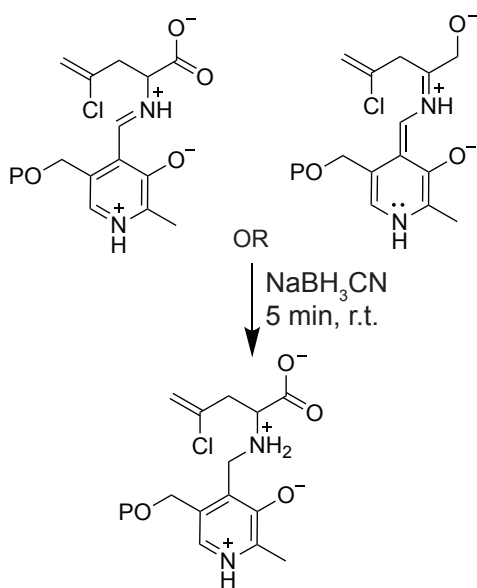
Scheme 2.1. Reduction of PLP intermediates by imine reduction using NaBH₃CN. Two possible species, an aldimine and a quinonoid, would yield a product of identical mass.

another where the imine is formed between the α-nitrogen and the PLP 4′-carbon. We can distinguish these two species by reducing the imine with a deuterated reducing agent, sodium cyanoborodeuteride (NaBD₃CN), and checking the spectra of the PLP-fragment for deuteration (*Scheme 2.2*). Performing these experiments, we found ~ 90% of the PLP-fragments were deuterated, suggesting that the majority of imine species are found as aldimines (data not shown). Unfortunately, background protonation, from impurities in NaBD₃CN, made it difficult to definitively determine if quinonoid or aldimine species were being trapped.

In the proposed mechanism, BesB catalyzes two deprotonations: first at C_α and then at C_β, which subsequently leads to elimination of chloride to form a terminal-allene intermediate. When we ran reactions of BesB using 4-Br-allylglycine as a substrate in 50 mM NaCl, we noticed formation of a significant amount of 4-Cl-allylglycine. Since 4-Cl-allylglycine is not observed when 4-Br-allylglycine is incubated overnight in 50 mM NaCl in the absence of enzyme, it seems as if halogen exchange occurs enzymatically. Working towards understanding halogen substitution reactions by BesB, we carried out reactions varying the γ-halogen, either 4-Cl- or 4-Br-allylglycine, in different salt solutions, either NaF, NaCl, NaBr, or NaI. We were able to observe formation of 4-I-allylglycine when either 4-Cl- or 4-Br-allylglycine were used as substrates (*Figure 2.30*). We also observed that more 4-I-allylglycine being formed when Br was the leaving group than was formed when Cl was the leaving group. When 4-Cl-allylglycine was a substrate and NaBr was the salt, 4-Br-allylglycine was observed as a side product. Likewise, when 4-Br-allylglycine was a substrate and NaCl was the salt, 4-Cl-allylglycine was observed. 4-F-allylglycine was not observed for any of the substrates when the reaction was carried out using NaF as a salt. These results suggest that nucleophiles can add back to an electrophilic intermediate generated upon initial elimination of halide.

58
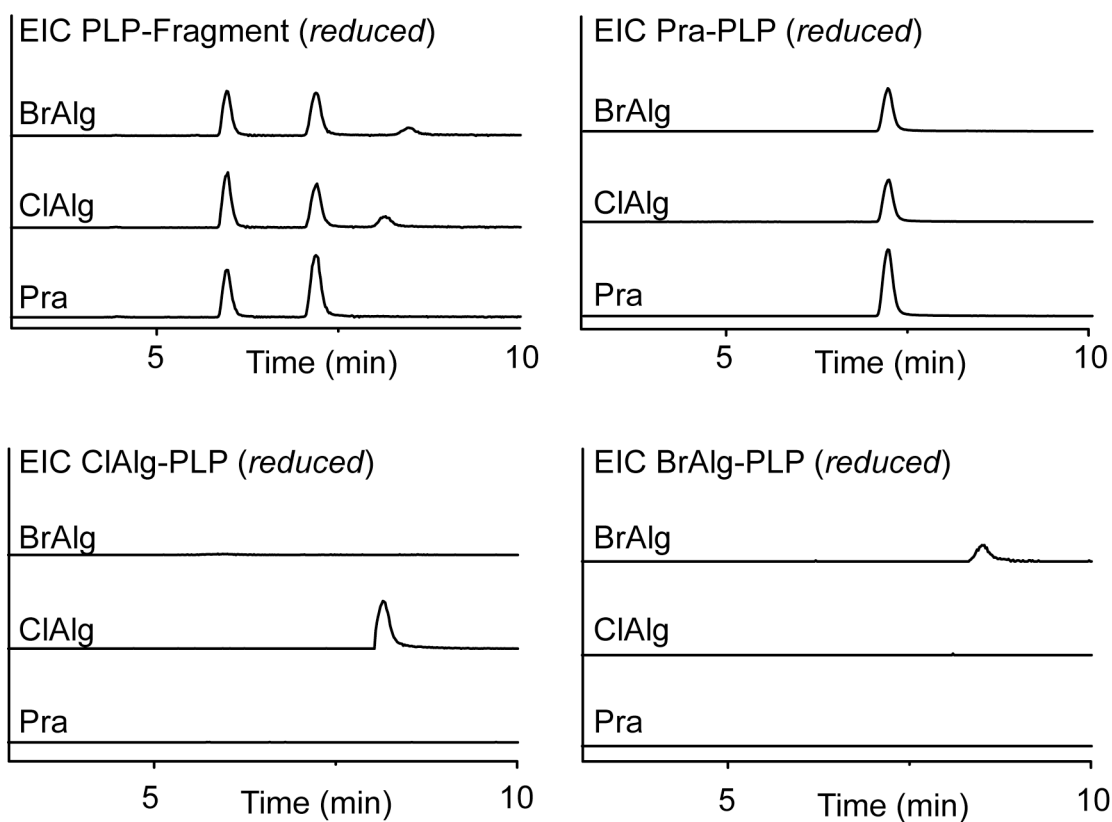
Figure 2.29. LC/QTOF analysis of *in vitro* reactions containing BesB, PLP, and substrates (either 4-Br-allylglycine, 4-Cl-allylglycine, or L-propargylglycine) quenched with NaBH₃CN. EIC for reduced PLP-Fragment (top left), propargylglycine-PLP (top right), Cl-allylglycine-PLP (bottom left), and Br-allylglycine-PLP (bottom right) shown.
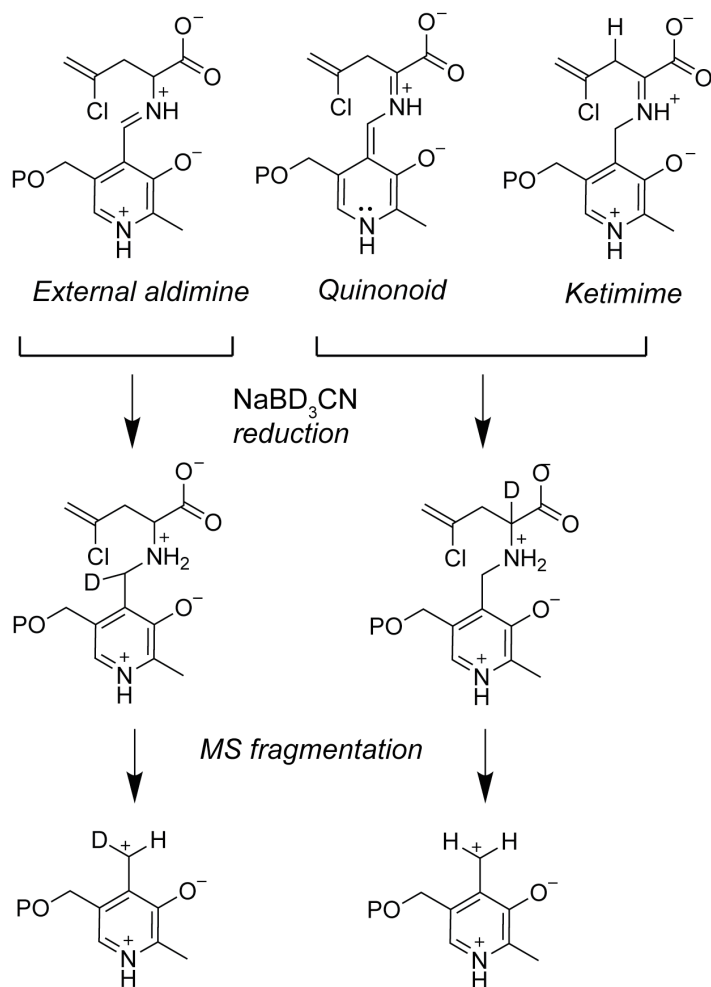
Scheme 2.2. Reduction of imines with NaBD$_3$CN can be used to distinguish between reduced aldimine species and reduced quinonoid/ketimine species. Direct detection performed by looking for PLP fragmentation on LC/MS.
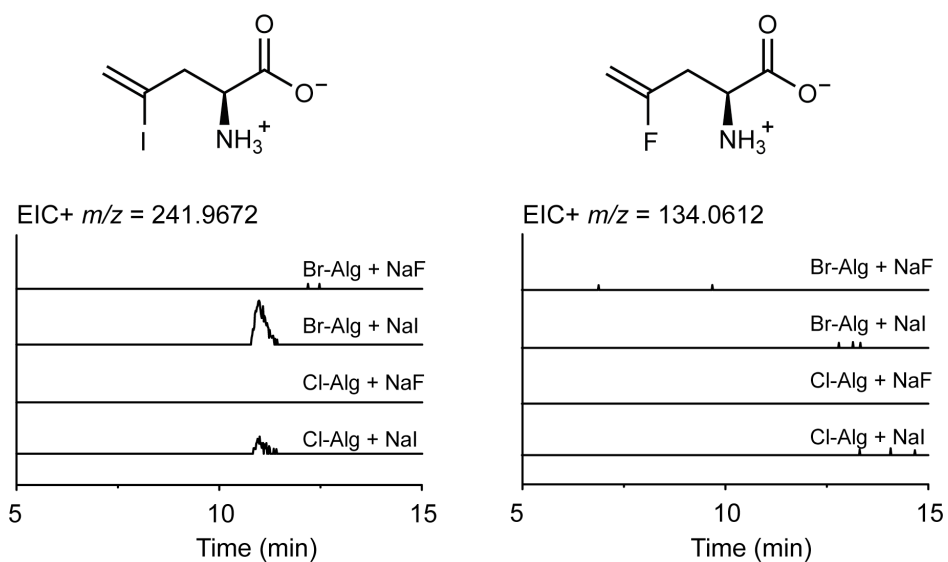
Figure 2.30. BesB is capable of catalyzing the substitution of the Cγ-halogen in 4-X-allylglycine (X = Cl, Br). LC/QTOF analysis of *in vitro* reactions containing BesB, a Cγ-halogenated substrates (either 4-Br-allylglycine, 4-Cl-allylglycine) in buffer containing either NaF or NaI. EIC for reduced 4-I-allylglycine (left) and 4-F-allylglycine (right) shown.

We had also proposed that formation of the terminal alkyne from the terminal allene occurs enzymatically. An alternative possibility is that BesB makes allenylglycine as a final product, which then isomerizes in solution to Pra. To rule out spontaneous isomerization of allenylglycine, we synthesized DL-allenylglycine (*Scheme 2.3*) [35]. From allenylglycine, we were also able to make propyldiene-glycine through a base-catalyzed isomerization (*Scheme 2.4*). Pra, allenlylglycine were resolvable by LC/MS, giving distinct elution times when separated by ZIC-pHILIC (*Figure 2.31*). We saw no formation of Pra when allenylglycine was incubating at room temperature or 37 °C overnight, ruling out spontaneous isomerization as a possibility. Another possibility is that BesB makes allenylglycine, which is subsequently isomerized enzymatically as the free amino acid within the active site of BesB. To test this hypothesis, we incubated allenylglycine with MBP-BesB overnight at room temperature and 37 °C and saw no conversion to Pra (data not shown). Therefore, we believe that if allenylglycine is formed as an intermediate species, it forms and is isomerized while still bound to PLP-cofactor.
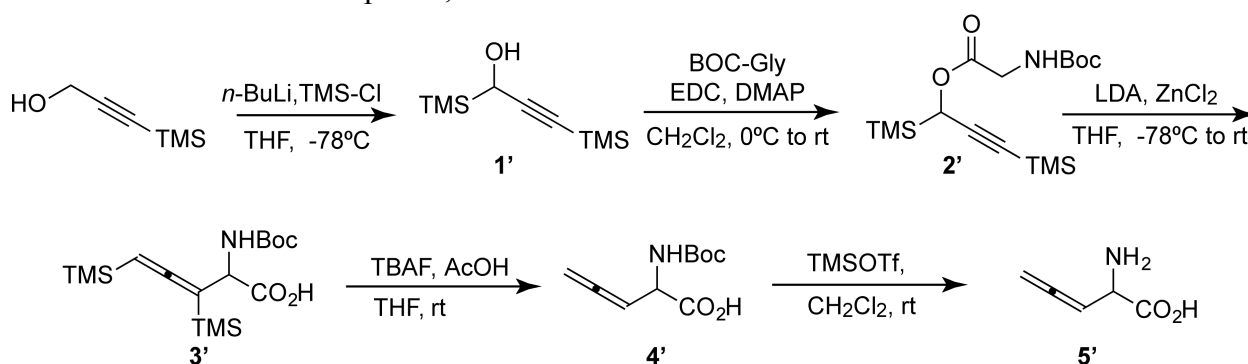
Scheme 2.3. Synthesis of allenylglycine. Adapted from T. Okada e*t al.*
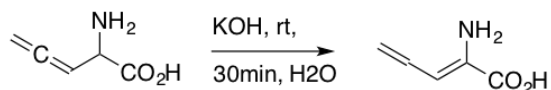
Scheme 2.4. Synthesis of propyldiene-glycine.

Thermodynamically, terminal alkynes and terminal allenes lie at similar G° [36]. The mechanism we proposed involves an allene-to-alkyne isomerization as one of the final steps. We next decided to investigate if BesB is capable of isomerizing Pra when Pra is used as a substrate. Experimentally, we tested this hypothesis by running a BesB reaction using Pra as the substrate, at high concentration, in D_2O. When analyzed by LC/MS, we observed formation of D_3-Pra (*Figure 2.32*). Since Pra has four non-acidic protons, we then needed to investigate which positions were exchanged with deuterium. Using ^{2}H-NMR, we were able to determine that Cα, Cγ, and one of the Cβ protons had exchanged with deuterium (*Figure 2.33-2.35*). While we did not directly determine the stereochemistry of the exchanged Cβ deuterium, it is likely this deuterium lies anti to the α-amine. A homology model of BesB, generated using Phyre2 [37], reveals the catalytic lysine (K322) is on the anti-face of amine-PLP intermediate (*Figure 2.36A*). We believe this residue is responsible for protonation and deprotonation of Cα and Cβ during BesB catalysis.

62
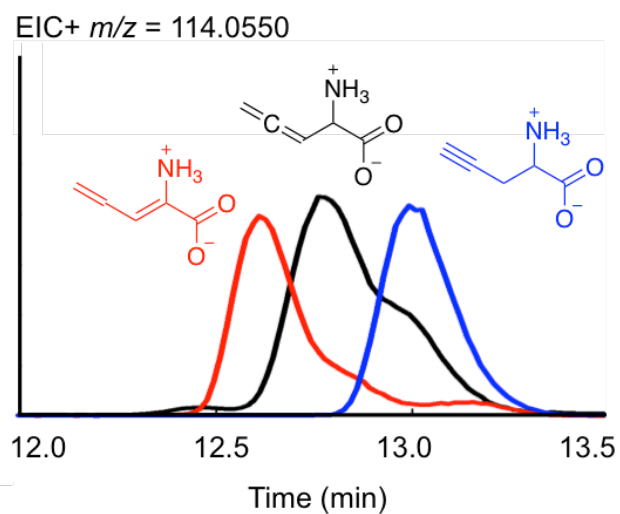
Figure 2.31 Resolving isomers of propargylglycine by mass spec. EIC for m/z = 114.0550, corresponding to propargylglycine, allenylglycine, or propyldiene-glycine, shown. Compounds ran independently and have distinct elution times.
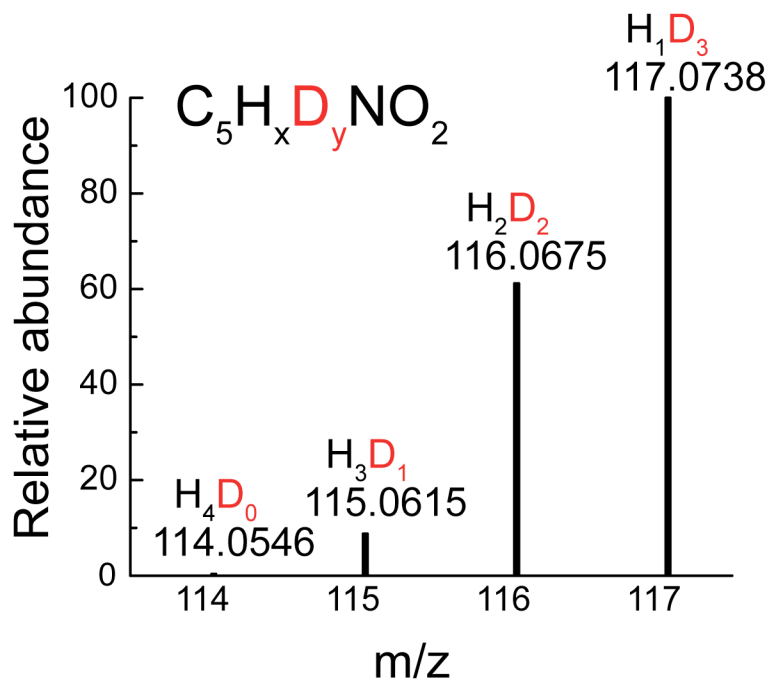
Figure 2.32. Mass spectra of L-propargylglycine incubated with BesB in D_2O overnight. Out of four possible non-acidic protons in Pra, three are seen to exchange by mass spectrometry.
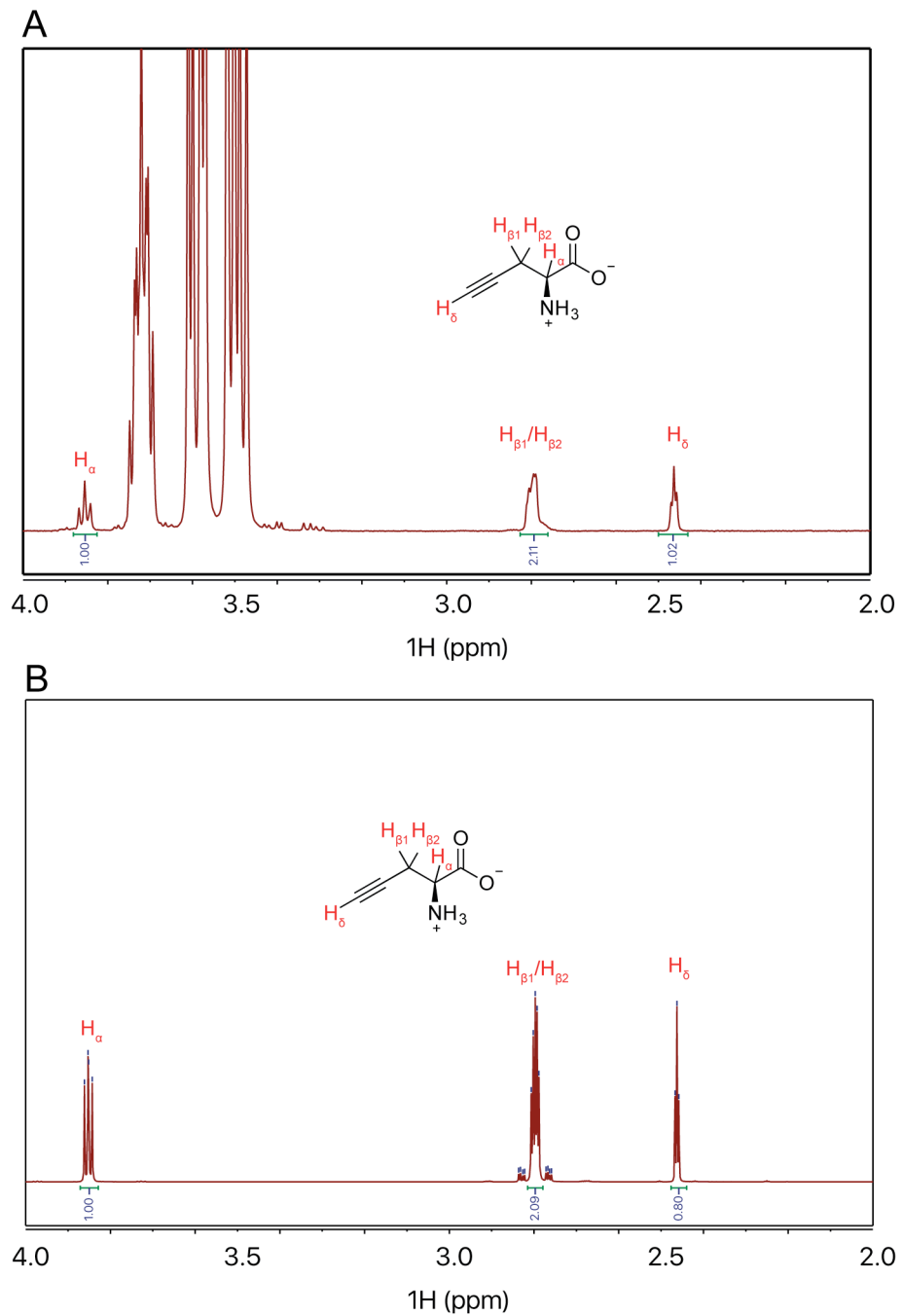
Figure 2.33. (A) 1D ^{1}H NMR of L-propargylglycine (1 mM) incubated with BesB in H_2O. Hα, Hβ, and Hδ assigned. (B) 1D ^{1}H NMR of a L-propargylglycine (1 mM) standard.
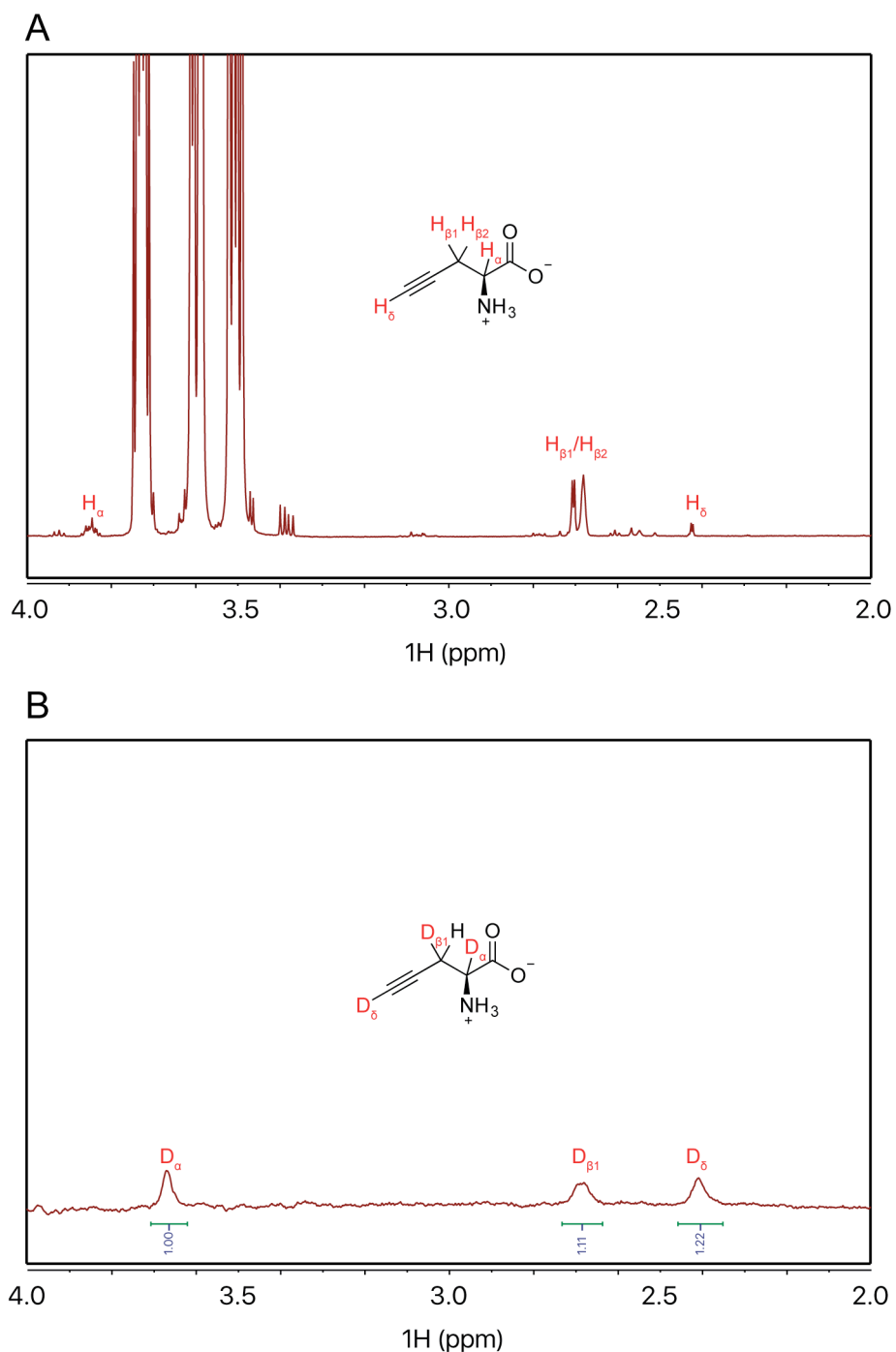
Figure 2.34. (A) 1D ^{1}H NMR of L-propargylglycine (1 mM) incubated with BesB in D_2O overnight. Hα, Hβ, and Hδ assigned. (B) 1D ^{2}H NMR of L-propargylglycine (1 mM) incubated with BesB in D_2O overnight. Three deuterated protons clearly visible, and correspond to Hα, Hδ, and one of two Hβ.
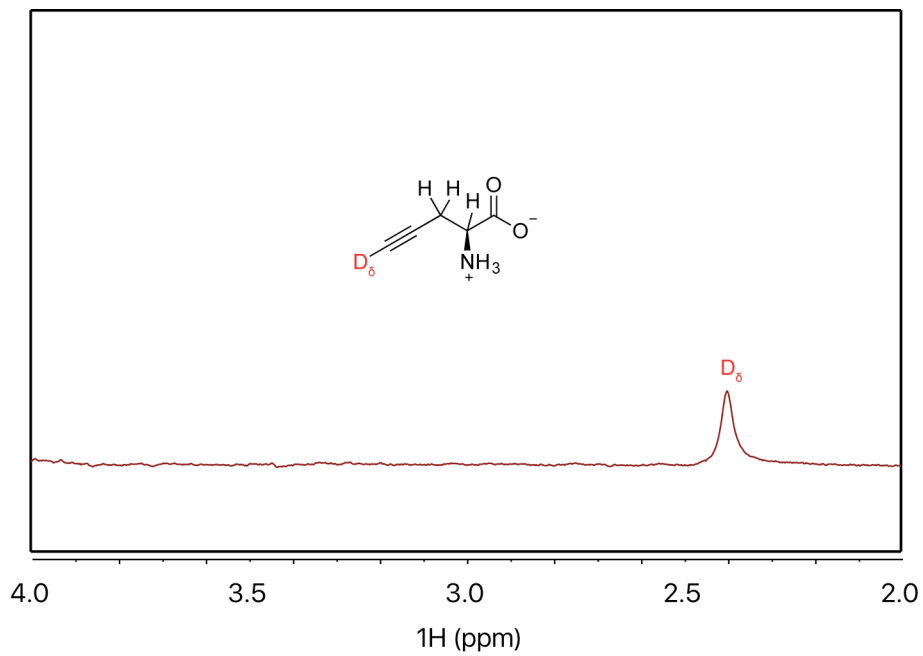
Figure 2.35. 1D ^{2}H NMR of L-propargylglycine (1 mM) incubated in D_2O for 7 d in the absence of BesB. A single proton is thought to exchange with solvent, corresponding to Hδ.

In attempt to directly detect the putative terminal-allene intermediate, we proposed enzymatically trapping the allene via a nucleophilic attack using a neighboring active site residue. Previous groups have observed that L-Pra serves as a mechanistic, suicidie inhibitor of various PLP-dependent enzymes by forming a terminal-allene intermediate, which is subsequently attacked by a neighboring nucleophile [32,33,38]. A crystal structure of cystathionine γ-lyase, covalently inhibited by L-Pra, shows a conserved tyrosine covalently bound to the PLP-inhibitor. In many PLP enzymes, the corresponding conserved tyrosine serves to π-stack with PLP. In BesB, the corresponding residue, F224, is a phenylalanine (*Figure 2.36B*). Making the corresponding mutant BesB-F224Y resulted in loss of activity for the enzyme. Unfortunately, we did not find any evidence, by whole protein MS nor shotgun proteomics, that BesB had been covalently inhibited. Interestingly, when we analyzed the resulting reactions by LC/MS, we did observe a small amount of product eluting in the region of allenylglycine/propargylglycine region (*data not shown*). That BesB-F224Y lacks in its ability to isomerize allenylglycine-PLP to propargylglycine-PLP remains an open possibility.

As a point of general interest for future work with BesB, we attempted to measure rate of Pra formation to estimate Micheális-Menton parameters k_{cat} and K_m. Without a reliable coupled-fluorescent assay, we opted to use discontinuous measurements of Pra by analyzing samples using LC/MS. We estimate $K_M \approx 0.22 \pm 0.12$ mM while $k_{cat} \approx 4.7 \pm 0.6$ min^{-1} (*Figure 3.37*). Future efforts at studying BesB kinetics might be able to take advantage of spectroscopic techniques, such as IR or Raman scattering, to directly measure rate of alkyne formation in real time.

BesA is a Pra-γ-glutamyl ligase. From the terminal-alkyne amino acid profile of the different streptomycete producers, along with the Δ*besE* knockout strain (*Figure 2.5, 2.6, 2.7A, 2.38*), it seemed reasonable to believe that hydroxylation of Pra by BesE yields βes (**7**). However, no βes was observed when purified BesE was incubated with Pra in the presence of Fe, ascorbate, and αKG. We thus looked for an alternative substrate for BesE, which we were able to identify as a glutamate-Pra dipeptide (**5**) through comparative metabolomics of *S. cattleya* Δ*besE* and wild-type strains (*Figure 3.39*). This compound was also present at low concentration in the Δ*besA* knockout. Upon fragmentation, characteristic fragments consistent with an amide bond cleavage were detected (*Figure 3.40*). Comparison of the products of purified BesA and a commercial γ-glutamyl transpeptidase further supports to product assignment as γ-Glu-Pra (*Figure 3.41*). Additional biochemical characterization indicates that BesA is selective for Pra over other common amino acids (*Figure 3.42, 3.43*). Interestingly BesA is conserved among these Pra biosynthetic gene clusters, suggesting that Glu-Pra formation has a physiological function, perhaps to reduce side reactions of Pra [33,38] or to facilitate transport.

BesE is a Glu-Pra hydroxylase. Having discovered that Glu-Pra is an on-pathway metabolite, we then incubated Glu-Pra with BesE, Fe, ascorbate, and αKG. Reactions were analyzed by LC/MS, and production of Glu-Bes was observed. These results show that Glu-Pra, and not Pra, is the substrate for BesE (*Figure 4.44*) near a terminal-alkyne gives rise to an intriguing question: How is BesE able to create a stable radical so close to a terminal-alkyne without rearrangement? Further work in characterizing BesE should probe this question, as well as attempt to understand the residues involved in terminal alkyne recognition.
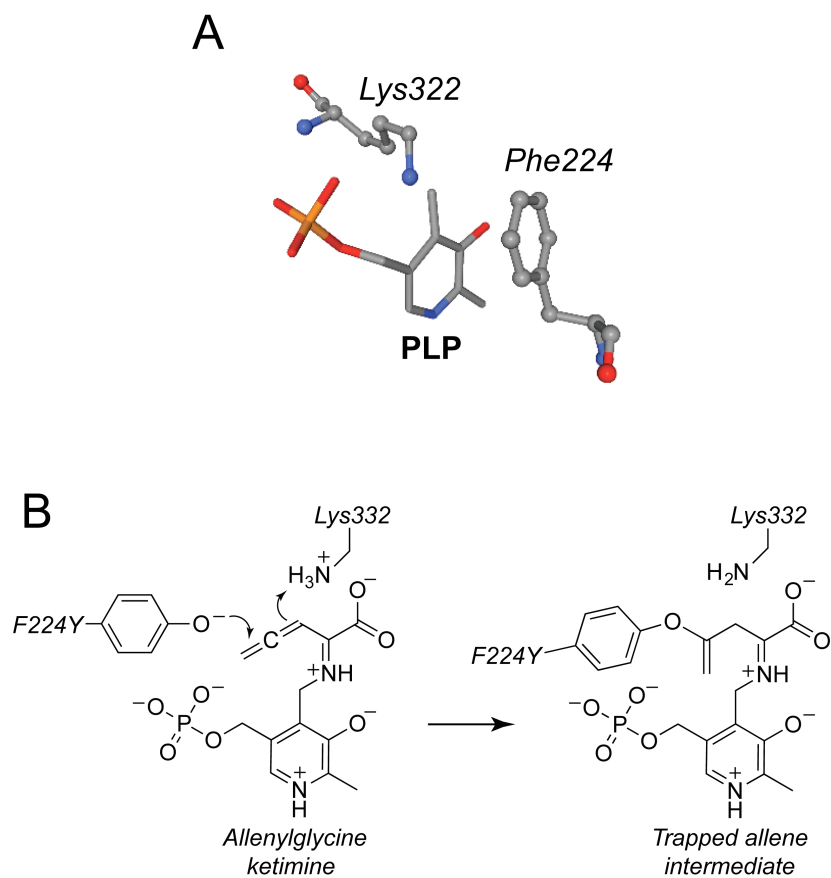
Figure 2.36. (A) Modeled active site of BesB contains a Phe224 which likely interacts with PLP. In other PLP enzymes, this residue is a tyrosine. (B) Mechanistic trapping of allene intermediate of BesB by using BesB F224Y mutant.
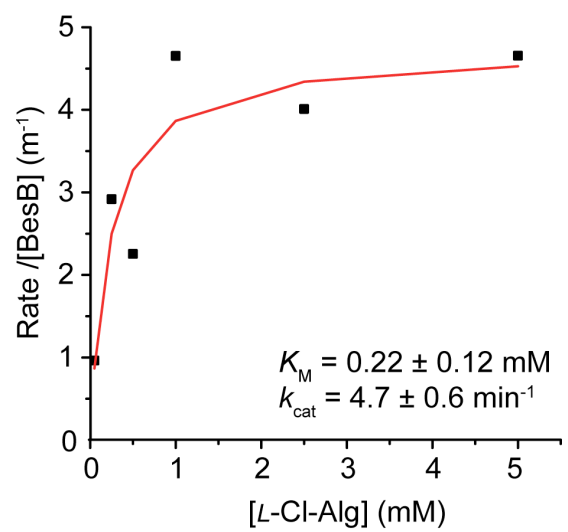
Figure 2.37. K_M and k_{cat} estimates for BesB reaction with L-Cl-allylglycine.
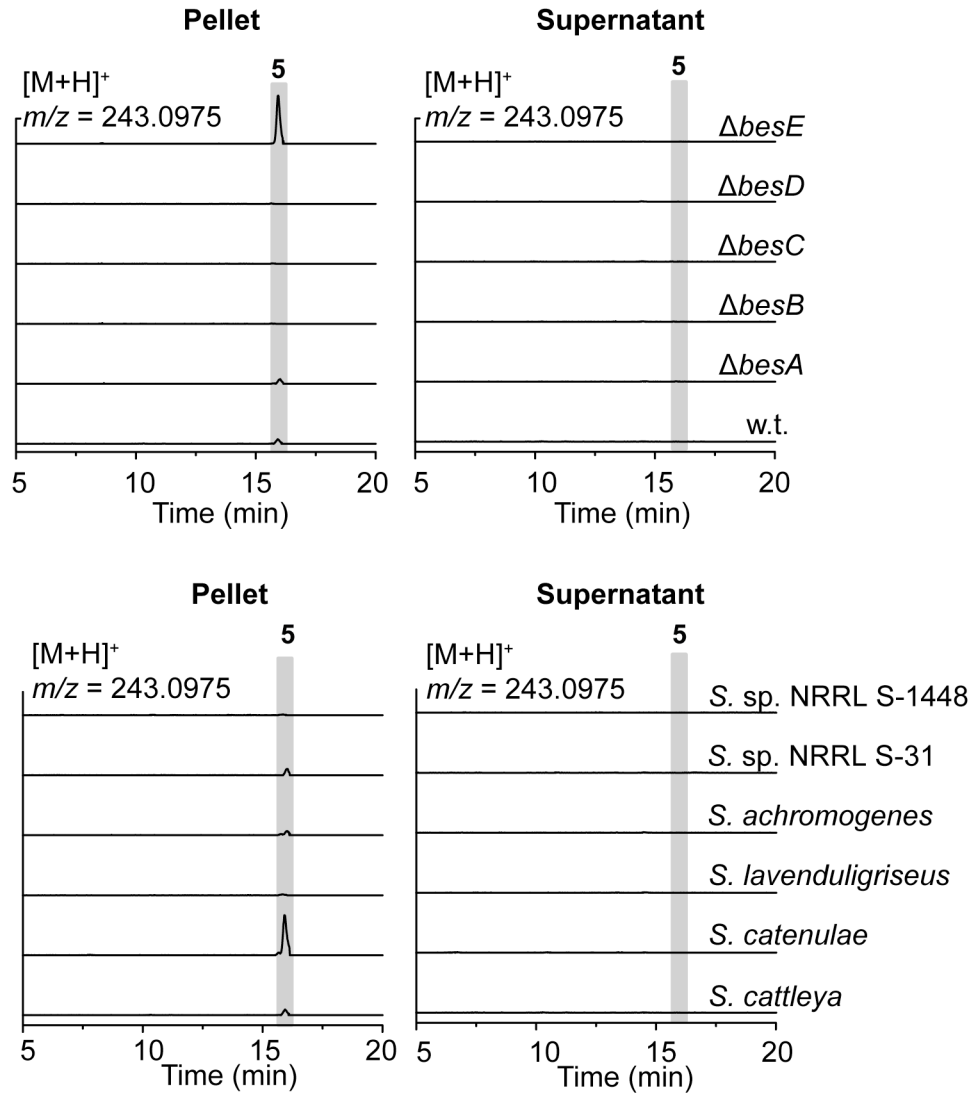
Figure 2.38. LC/QTOF traces for γ-Glu-Pra (*m/z* = 243.0975) from metabolites extracted from cell pellets and supernatants of *S. cattleya* w.t. and *bes* knockout strains as well as *S.* sp. NRRL S-1448, *S.* sp. NRRL S-31, *S. achromogenes, S. lavenduligriseus,* and *S. catenulae*. γ-Glu-Pra was detected in the cell pellets of *S. cattleya* w.t., the Δ*besA* and Δ*besE* knockout strains, *S. catenulae, S. achromogenes*, *S.* sp. NRRL S-31, and *S.* sp. NRRL S-1448. Lack of detection in the culture supernatants suggests that it is not a secreted metabolite. Chromatograms shown are representative of at least 3 independent experimental replicates
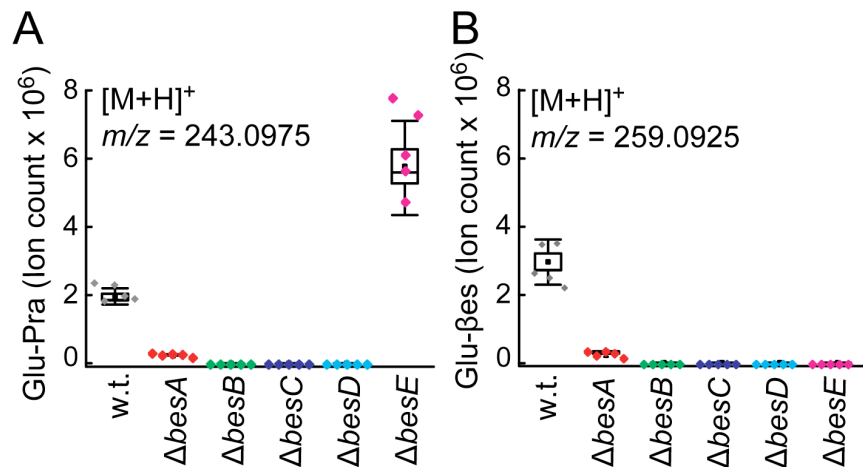
Figure 2.39. (A) Relative abundance of the γ-Glu-Pra dipeptide in *S. cattleya* wildtype and pathway knockout strains (n = 5 biologically independent experiments; dot = mean; line = median; box edges = 25th/75th percentile; error bars = 95th percentile). (B) Relative abundance of the γ-Glu-βes dipeptide in *S. cattleya* wildtype and pathway knockout strains (n = 5 biologically independent experiments; dot = mean; line = median; box edges = 25th/75th percentile; error bars = 95th percentile).
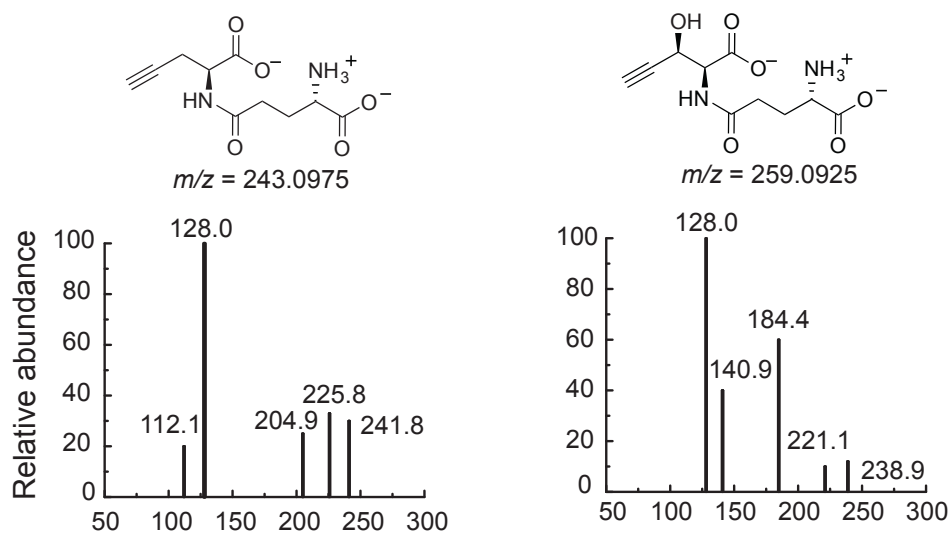
Figure 2.40. Fragmentation of γ-Glu-Pra (left) and γ-Glu-βes (right) in negative ionization mode by LC/QQQ shows fragments characteristic of γ-glutamyl-dipeptides (*m/z* = 128.0). Spectra shown are representative of at least 3 independent experimental replicates.
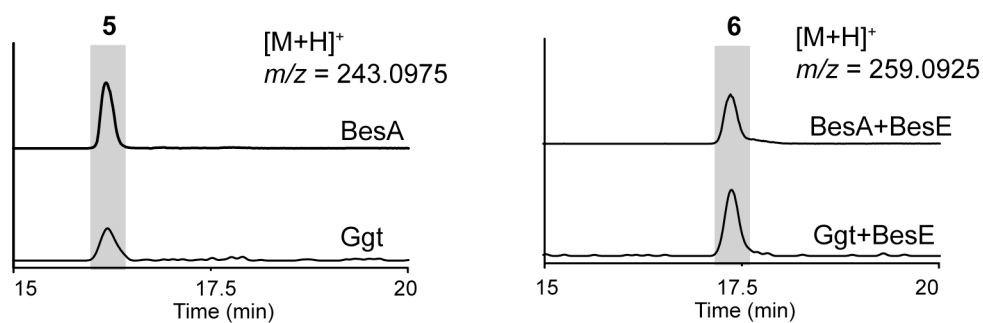
Figure 2.41. LC/MS traces for enzymatically synthesized γ-Glu-Pra (5, *m/z* =243.0975) by either Ggt or BesA and LC/QTOF traces for enzymatically synthesized γ-Glu-βes (6, *m/z* = 259.0925) by either Ggt + BesE or BesA + BesE. Chromatograms shown are representative of at least 3 independent experimental replicates.
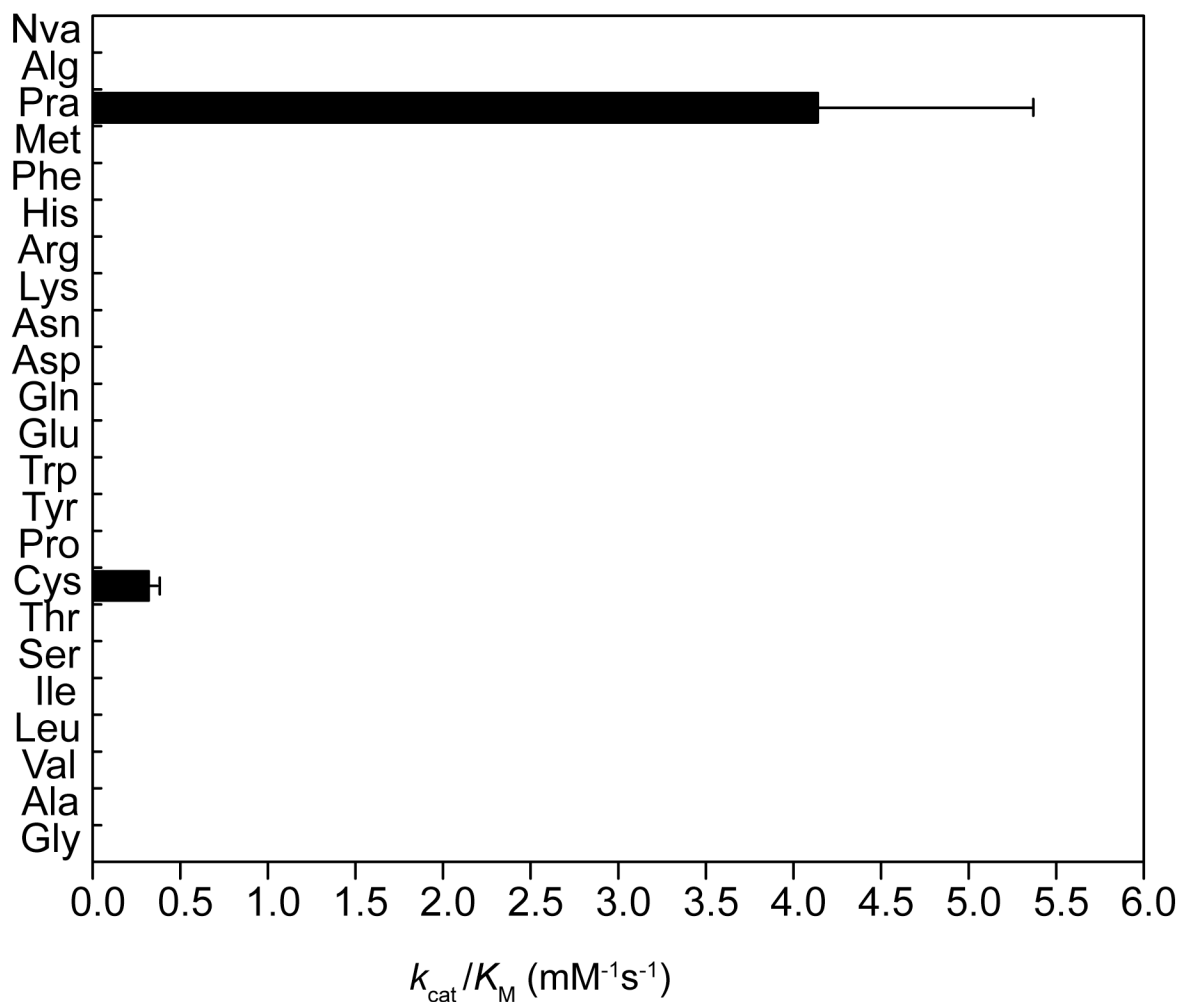
Figure 2.42. Steady-state kinetic analysis for BesA reactions between L-glutamate and various L-amino acid partners (Nva, norvaline; Alg, allylglycine). Each bar represents the k_{cat}/K_M value for the corresponding amino acid calculated from the individual kinetic terms (k_{cat} and K_M) obtained from non-linear curve fitting of data to the Michaelis-Menten equation. Data shown are mean ± standard error of the mean (n = 3 experimental replicates) obtained by propagation. Only reactions between L-glutamate and L-propargylglycine or L-cysteine showed any detectable activity

Pra

Cys
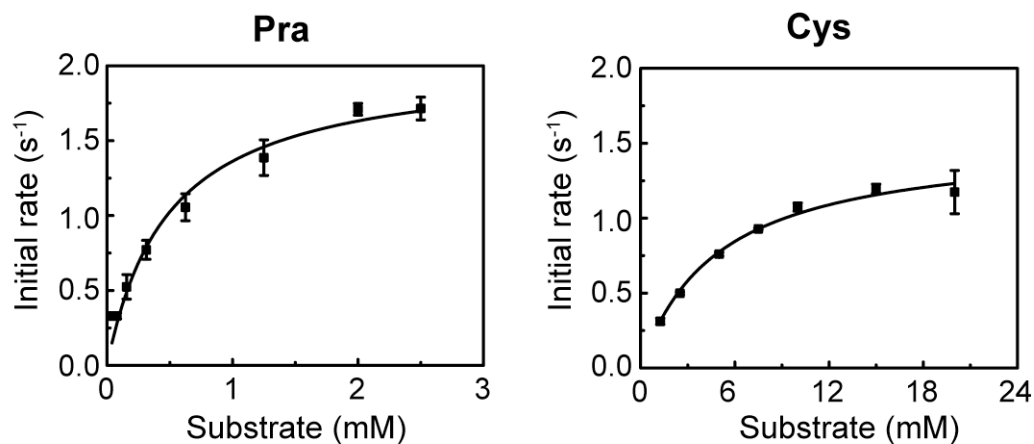
	k_{cat} (s⁻¹)	K_M (mM)	k_{cat}/K_M (mM⁻¹s⁻¹)
Pra	2.0 ± 0.1	0.49 ± 0.09	4 ± 1
Cys	1.53 ± 0.06	4.9 ± 0.6	0.3 ± 0.1

Figure 2.43. Steady state kinetic analysis of BesA with 10 mM glutamate and L-propargylglycine or L-cysteine as substrates. Data are mean ± s.d. (n = 3 experimental replicates). Table contains k_{cat}, K_M, and k_{cat}/K_M calculated by non-linear curve fitting to the Michaelis-Menten equation. Data are mean ± s.e. Error in k_{cat}/K_M is obtained by propagation from the individual kinetic terms.
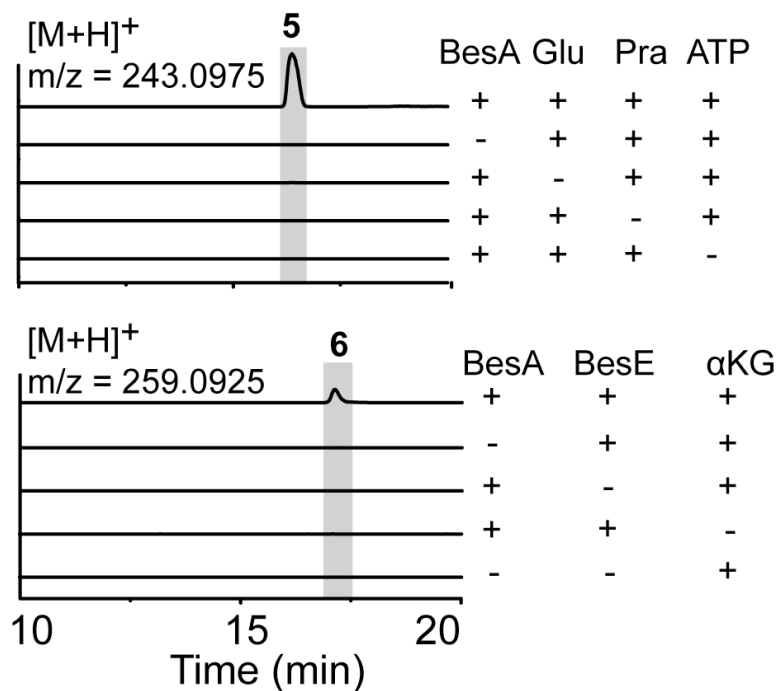
Figure 2.44. (Top) Extracted ion chromatograms for γ-Glu-Pra in reactions containing a combination of BesA, BesE, and cofactors (γ-Glu-Pra, **5**; *m/z* 243.0975). (Bottom) Extracted ion chromatograms for γ-Glu-βes in reactions containing a combination of BesA, BesE, and cofactors (γ-Glu-βes, **6**; *m/z* 259.0925). Chromatograms shown are representative of at least 3 independent experimental replicates.

76

Reconstitution of βes pathway *in vitro*. Elucidation of the final committed step, hydroxylation by BesE, finally allowed us to fully reconstitute the biosynthesis of Glu-βes (**6**) *in vitro* (*Figure 2.45*). In a single pot reaction with L-lysine, purified protein (BesA, BesB, BesC, BesD, and BesE), and necessary co-factors (αKG, Fe, NaAsc, PLP, ATP, L-Glu) we were able to successfully make Glu-βes *in vitro*. While the intended goal had been to make free Glu-βes, we believe other enzymes, found ubiquitously in life, can carry out Glu-βes hydrolysis. One such candidate enzyme is Ggt. How *S. cattleya* and other Glu-βes/Glu-Pra producing species of Streptomyces hydrolyze these amino acid dipetides remains open to speculation.
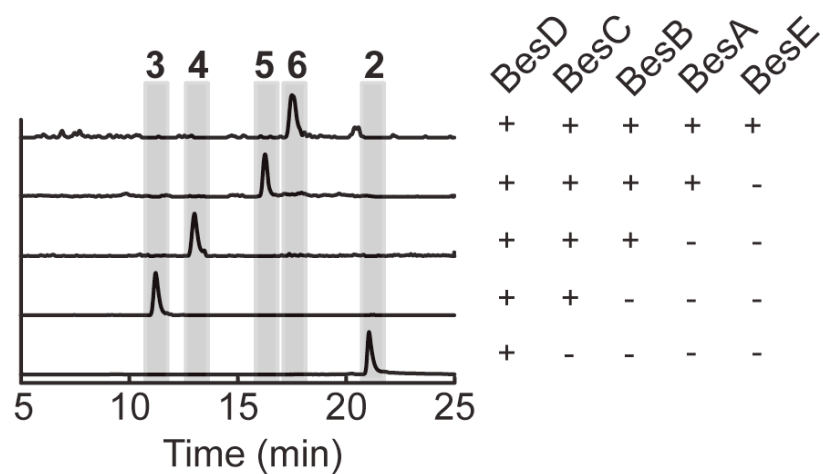
Figure 2.45 *In vitro* reconstitution of βes biosynthesis. Extracted ion chromatogram for each pathway intermediate in the presence of corresponding enzymes required to catalyze each step (**2**, *m/z* 181.0738; **3**, *m/z* 150.0316; **4**, *m/z* 114.550; **5**, *m/z* 243.0975; **6**, *m/z* 259.0925). Peaks are normalized for each trace to the most abundant ion.

2.4. Conclusion

Though first discovered in the 1970s and 1980s, the biosynthetic origin of terminal-alkyne amino acids had eluded various attempts at their elucidation. This work is first in uncovering how terminal-alkyne amino acids, βes and Pra, can be enzymatically synthesized from primary metabolites, particularly L-lysine. In doing so, we have identified and characterized a unique biosynthetic pathway for the production of halo, terminal alkene, and terminal-alkyne amino acids which includes several unusual chemical transformations and intermediates. are several enzymes with interesting reactivity including an amino acid halogenase (BesD), di-iron oxidase that oxidatively cleaves C-C bond (BesC), and a new type of acetylenase (BesB). The biochemistry uncovered in this work has

Looking forward, the subsequent introduction of these amino acids, which contain functional groups not commonly found in natural systems, can be attempted. Various opportunities exist for their incorporation into peptides and proteins (either site- or residue-specifically) for downstream diversification by bioorthogonal chemistries. These functional groups give access to chemical reaction space ranging from alkyne-mediated CuAAC 'Click' reactions, olefin metathesis or polymerization, and halogen-dependent reactions. The ability to manipulate protein and peptide structure, function, and physical properties through chemical reactions leaves open the possibility for developing hybrid natural systems capable of interacting with the synthetic chemical space.

2.5. References

1. Marchand, J. A. *et al*. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature*. doi:10.1038/s41586-019-1020-y

2. Minto, R. & Blacklock, B. Biosynthesis and function of polyacetylenes and allied natural products. *Prog. Lipid Res*. **47,** 233–306 (2008).

3. Scrimgeour, C. M. Natural acetylenic and olefinic compounds, excluding marine natural products. *Aliphatic Relat. Nat. Prod. Chem*. **2,** (1979).

4. Kuklev, D. V, Domb, A. J. & Dembitsky, V. M. Bioactive acetylenic metabolites. *Phytomedicine* **20,** 1145–59 (2013).

5. Potgieter, H. C., Vermeulen, N. M. J., Potgieter, D. J. J. & Strauss, H. F. A toxic amino acid, 2(S)3(R)-2-amino-3-hydroxypent-4-ynoic acid from the fungus Sclerotium rolfsii. *Phytochemistry* **14,** 1757–1759 (1977).

6. Sanada, M., Miyano, T. & Iwadare, S. Ethynylserine, an antimetabolite of L-threonine from Streptomyces cattleya. *J. Antiboitics* **39,** 304–305 (1986).

7. Scannell, J. P. *et al*. Antimetabolites produced by microorganisms II. L-2-amino-4-pentynoic acid. *J. Antibiot. (Tokyo)*. **24,** 239–244 (1971).

8. Zhou, Z.-Y. *et al*. Evidence for the natural toxins from the mushroom Trogia venenata as a cause of sudden unexpected death in Yunnan Province, China. *Angew. Chem. Int. Ed. Engl*. **51,** 2368–70 (2012).

9. Yoshio, K. *et al*. FR-900130, A novel amino acid antibiotic I. Discovery, taxonomy, isolation, and properties. *J. Antibiot. (Tokyo)*. **33,** 125–131 (1980).

10. Shin-Ichi, H. Amino acids from mushrooms. *Prog. Chem. Org. Nat. Prod*. 117–140 (1992).

11. De Simone, A., Acevedo-Rocha, C. G., Hoesl, M. G. & Budisa, N. Towards reassignment of the methionine codon AUG to two different noncanonical amino acids in bacterial translation. *Croat. Chem. Acta* **89,** 1–11 (2016).

12. Dieterich, D. C. *et al*. Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc*. **2,** 532–40 (2007).

13. Truong, F., Yoo, T. H., Lampo, T. J. & Tirrell, D. A. Two-strain, cell-selective protein labeling in mixed bacterial cultures. *J. Am. Chem. Soc*. **134,** 8551–8556 (2012).

14. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chemical Reviews* **114,** 4764–4806 (2014).

15. Uttamapinant, C. *et al*. Genetic Code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. *J. Am. Chem. Soc*. **137,** 4602–4605 (2015).

16. De Graaf, A. J., Kooijman, M., Hennink, W. E. & Mastrobattista, E. Nonnatural amino acids for site-specific protein conjugation. *Bioconjug. Chem*. **20,** 1281–1295 (2009).

17. Sakamoto, K. Site-specific incorporation of an unnatural amino acid into proteins in

mammalian cells. *Nucleic Acids Res*. **30,** 4692–4699 (2002).

18. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol*. **20,** 1044–1048 (2002).

19. Bindman, N. A. & van der Donk, W. A. A general method for fluorescent labeling of the N-termini of lanthipeptides and its application to visualize their cellular localization. *J. Am. Chem. Soc*. **135,** 10362–71 (2013).

20. Zhu, X., Liu, J. & Zhang, W. De novo biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol*. **11,** 115–20 (2015).

21. Edwards, D. J. *et al*. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chem. Biol*. **11,** 817–833 (2004).

22. Zhu, X., Su, M., Manickam, K. & Zhang, W. Bacterial genome mining of enzymatic tools for alkyne biosynthesis. *ACS Chem. Biol*. **10,** 2785–2793 (2015).

23. Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. XCMS online: A web-based platform to process untargeted metabolomic data. *Anal. Chem*. **84,** 5035–5039 (2012).

24. Gust, B. *et al*. λ Red-mediated genetic manipulation of antibiotic-producing Streptomyces. *Advances in Applied Microbiology* **54,** 107–128 (2004).

25. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci*. **100,** 1541–1546 (2003).

26. Yoshio, K. *et al*. FR-900130, A novel amino acid antibiotic II. Isolation and structure eluciation of the acetyl derivative of FR-900130. *J. Antibiot. (Tokyo)*. **33,** 132–136 (1980).

27. Wong, S. D. *et al*. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499,** 320 (2013).

28. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436,** 1191–1194 (2005).

29. Fisher, H. F. L-glutamate dehydrogenase from bovine liver. *Methods Enzymol*. **113,** 16–27 (1985).

30. Campton, B. J. & Purdy, W. C. Fluoral-P, a member of a selective family of reagents for aldehydes. *Anal. Chim. Acta* **119,** 349–357 (1980).

31. Spring, J. G. C. *et al*. 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 (1990).

32. Sun, Q. *et al*. Structural basis for the inhibition mechanism of human cystathionine gamma-lyase, an enzyme responsible for the production of H2S. *J. Biol. Chem*. **284,** 3076–3085 (2009).

33. Marcotte, P. & Walsh, C. Vinylglycine and propargylglycine: Complementary suicide substrates for L-amino acid oxidase and D-amino acid oxidase. *Biochemistry* **15,** 3070–3076 (1976).

34. Marini, M. a *et al*. A critical examination of the reaction of pyridoxal 5-phosphate with human hemoglobin Ao. *Biopolymers* **28,** 2071–2083 (1989).

35. Okada, T., Oda, N., Suzuki, H., Sakaguchi, K. & Ohfune, Y. Synthesis of optically active α-(allenyl)- and α-substituted-α-(allenyl)glycines. *Tetrahedron Lett*. **51,** 3765–3768 (2010).

36. Yu, S. & Ma, S. How easy are the syntheses of allenes? *Chem. Commun. (Camb)*. **47,** 5384–5418 (2011).

37. Kelly, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc*. **10,** 845–858 (2015).

38. Abeles, R. H. & Walsh, C. T. Acetylenic enzyme inactivators. Inactivation of gamma-cystathionase, in vitro and in vivo, by propargylglycine. *J. Am. Chem. Soc*. **434,** 3–4 (1973).

Chapter 3: *Development and optimization of a pathway for production of halo, alkene, and alkyne amino acids*

3.1. Introduction

Organisms use a standard set of 20 proteinogenic building blocks to build the proteins and enzymes that give life its functional complexity. Still, exceptions exist where additional non-standard amino acids are employed to expand the biochemical reactivity or structural-functional space of proteins. The most widespread of these exceptions is selenocysteine, an amino acid found in all three domains of life that has a lower pK_a and lower reduction potential than cysteine [2,3]. In archaea, certain species have been discovered that can make and translate pyrrolysine, which is believed to be involved in expanding catalytic function of certain methyltransferases [4,5]. The discovery of a naturally-expanded genetic code (selenocysteine and pyrrolysine) has forced us to reimagine what type of catalysis proteins could perform, and what life might look like, if living organisms could translate proteins composed of amino acids beyond the canonical 20.

Efforts at expanding the genetic code with non-standard amino acids (nsAA) have largely fallen into two categories: residue-replacement and site-specific incorporation [6]. The former strategy aims to replace all instances of one of the 20 proteinogenic amino acids in all proteins with a nsAA [7]. Residue-replacement has served as a powerful tool to study translation, as newly synthesized proteins containing a nsAA can be differentially separated from the remainder of the proteome through the use of an enrichment tag [8–11]. The latter strategy looks to target the incorporation of a nsAA to one position on one protein of interest [12]. This targeted approach has been used to incorporate over 100 different nsAA into proteins, with applications ranging from enhancing catalytic activity to providing a reactive handle for click chemistry [13–22]. While both these strategies have been widely successful, they rely on using nsAA that are first chemically synthesized and then fed to organisms.

The previously described βes pathway provides the biosynthetic potential to expand the available amino acid building blocks organisms are able to make using their preferred carbon source (*Figure 3.1*). Coupling the biosynthesis of βes pathway amino acids to residue-replacement or site-specific incorporations offers an avenue to engineer organisms that can autonomously expand their proteinogenic chemical space to include halo-, alkene-, and alkyne functional groups. One challenge moving forward is translating what we have learned about the βes pathway *in vitro* to optimize production and building pathways *in vivo*.

In this section, we detail our efforts at lifting the βes pathway from soil bacterium *S. cattleya* to industrial organism *Escherchia coli*. We first focus on optimizing production titers for diverse set of amino acids (4-Cl-lysine, allylglycine, 4-Cl-allylglycine, and Pra) by exploring changes to media conditions, genes, and regulation (*Figure 3.1*). As a proof-of-concept, we then show that we can obtain residue-replacement of Met with Pra in *E. coli* using an engineered aminoacyl-tRNA synthetase. Finally, we discuss ongoing efforts for site-specific incorporation of Pra using an engineered MetRS/tRNA^{Met}CUA pair, and preliminary work towards residue-replacement of Leu with allylglycine.
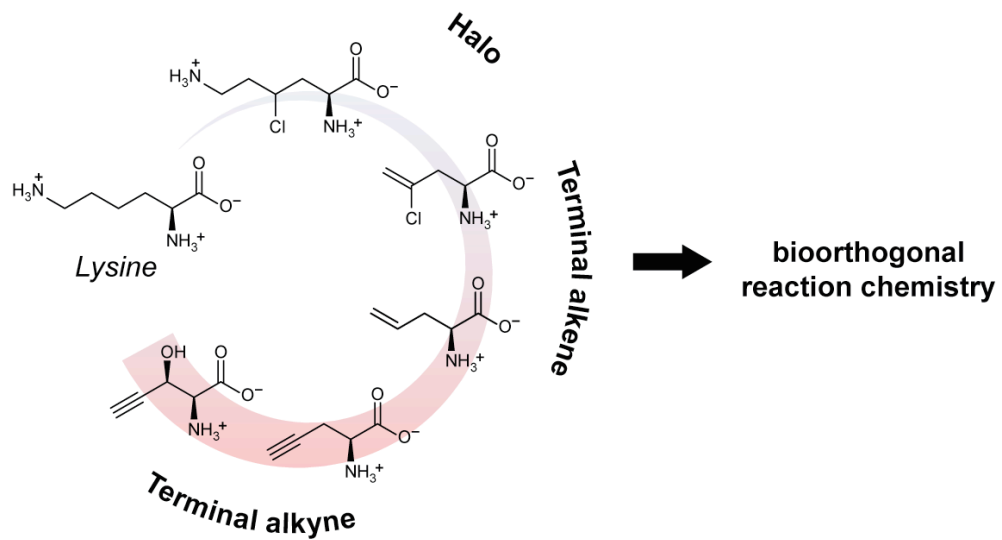
Figure 3.1. Amino acids from the βes pathway can be made *de novo* from lysine, and be used for bioorthogonal reactions.

3.2. Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), yeast extract, malt extract, and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Ammonium iron (II) sulfate hexahydrate, β-mercaptoethanol (βME), sodium phosphate dibasic heptahydrate, dimethylsulfoxide (DMSO), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), L-allylglycine (allylglycine), copper sulfate (CuSO$_4$), Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), L-propargylglycine (Pra), and L-lysine, 5-(3-Bromo-phenyl)-2H-tetrazole, and glutamate dehydrogenase were purchased from Sigma-Aldrich (St. Louis, MO). 4-Chloro-lysine was purchased from AKos GmbH (Steinen, Germany). (2S)-2-amino-4-chloropent-4-enoic acid (4-Cl-allylglycine) was purchased from Enamine, Ltd (Kiev, Ukraine). 3-Azido-7-hydroxycoumarin was purchased from AK Scientific, Inc (Union City, CA). Formaldehyde (16% (*w/v*), methanol-free) and PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA synthetase was purchased from Megazyme International (Bray, Ireland). Bacto^{TM} Agar was purchased from BD (Sparks, Maryland). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 µM in 10 mM Tris-HCl, pH 8.5, and stored at either 4 °C for immediate usage or -20 °C for longer term usage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). SoluLyse reagent was purchased from Genlantis (San Diego, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators and 5,000 MWCO regenerated cellulose ultrafiltration membranes were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). TEV protease was purchased from the University of California, Berkeley Macro Lab (Berkeley, CA). Cy5-Tetrazine was purchased from Click Chemistry Tools (Scottsdale, AZ).

Bacterial strains. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection (Manassas, VA). *E. coli* DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production of BesA, BesB, BesC, BesD, BesE. KEIO collection strain *E. coli* BW1125 Δ*metB* (JW3910) was used to test the effects of Met competition on Pra incorporation. *E. coli* B834(DE3) was used for *in vivo* studies involving incorporation of Pra into proteins using engineered pathway. *E. coli* Rosetta2 (DE3) pLysS, BL-21 (DE3) CodonPlusRIL, BL21 (DE3) Star, SoluBL-21, and Shuffle T7 Express were used for expression testing of BesB. *E. coli* TUNERS (DE3) was one strain used for promoter strength expression testing of BesB.

Construction of plasmids for *in vivo* production of amino acids, and incorporation of Pra in *E. coli*. Standard molecular biology techniques were used to carry out plasmid construction

using *E. coli* DH10B-T1R as the cloning host. All PCR amplifications were carried out with Phusion polymerase using the oligonucleotides listed in Table S1. For production with BesD or BesC from *P. fluorescens*, primer sets p16b-BesD-Fluor-F/R or p16-BesC-Fluor-F/R were used to amplify respective genes from *P. fluorescens* and clone resulting PCR product into pET16b using NdeI/BamHI cut sites. For expression of BesB under arabinose inducible promoter, BesB from *S.* sp. NRRL S-1448 was cloned into pBAD24 using pBAD24-BesB-F/R primers with Gibson ligation using NdeI/HindIII cut sites. For expression of BesB homologues from *Opitutaceae bacterium TSB47*, *Leptospira kirschneri*, or *Verrucomicrobia bacterium*, the corresponding synthetic gene (sBesB-OB, sBesB-LK, or sBesB-VB) was purchased and cloned into pSV272.1 using Gibson ligation and SfoI cut site. For expression of BesB under trc promoter, BesB from *S.* sp. NRRL S-1448 was cloned into pTrc33 using pTrc33-BesB-F/R primers with Gibson ligation using KpnI/XhoI cut sites. For expression of BesB under glnS promoter, F primer p16bglns-BesB-F (containing glnS promoter) and p16bglns-BesB-R was used to amplify BesB from S. sp. NRRL S-1448 and cloned into pET16b with Gibson ligation using BglII/XbaI cut sites. For BesC/BesD co-expression, the BesD homologue from *P. fluorescens* was cloned into pETDUET-1 using BamHI/HindIII cut sites while BesC homologue from *P. fluorescens* was cloned into NdeI/KpnI sites to form pETDUET-BesC.BesD. The codon-optimized BesB homologue from *S.* sp. NRRL S-1448 (sBesB), including T7 promoter region, was cloned from pSV272.1-MBP-BesB-opt and inserted into the pETDUET-BesC.BesD KpnI site to generate pPra. praGro was generated by first cloning PraRS into pBAD24 NdeI/HindIII cut site, then using PCR to amplify fragment that extends from the araP region to the stop codon. This resulting fragment was cloned into pGro7 at the XmaI/HindIII cut site to generate praGro. To construct pREP-GFP-CUA and pREP-GFP, primer pREP-CUA-F/pREP-R or pREP-F/R were used to amplify eGFP and cloned into pet28a using Gibson ligation with NcoI/HindIII cut sites. To construct pET16b-MetRS-CUA, primer sets MetRS-CUA-A/B-F/R were used to amplify two fragments of MetRS from *E. coli* DH10B genomic DNA. PCR fragments were subsequently cloned into pET16b with Gibson ligation using XbaI/NcoI cut sites. To construct pEVOL-MetRS-CUA, MetRS-CUA was amplified from pET16b-MetRS-CUA using pEVOL-MetRS-CUA-A-F/R and pEVOL-MetRS-CUA-B-F/R primer sets, and cloned into pEVOL using Gibson ligation using either BglII/SalI or Nde/PstI cut sites respectively. Met-tRNA-CUA was cloned into pEVOL-MetRS by cloning synthetic oligo pEVOL-Met-tRNA-CUA (*Appendix A1.3*) with Gibson ligation using BclI/XhoI cut site. To construct PraRS-CUA, the same primer sets as MetRS-CUA were used but PraRS (instead of MetRS) was used as a template for amplification. To construct pEVOL-PraRS-CUA, the same primer sets as pEVOL-MetRS-CUA were used, but PraRS-CUA (instead of MetRS) was used a template for PCR amplification. To generate N_T-fused BesB variants, BesB from *S.* sp. NRRL S-1448 was amplified and cloned with Gibson ligation into either pet41b (for GST fusion) or pet32a (for Trx fusion) using NcoI/XhoI or KpnI/HindIII cut sites. To generate truncation library of BesB (Δ5aa, Δ10aa, Δ15aa, Δ20aa, Δ25aa, Δ30aa, Δ35aa, Δ40aa, Δ45aa, and Δ50aa), BesB from *S.* sp. NRRL S-1448 was amplified using forward primer sets p16b-BesBTrunc-X with p16b-BesBTrunc-R and cloned into pET16b using Gibson ligation with NcoI/BamHI cut sites. LeuRS-T252Y was cloned into pBAD24a by amplifying two LeuRS fragments from *E. coli* DH10B genomic DNA with primer sets pBAD24-LeuRS-T252Y-A/B-F/R and cloning them with Gibson ligation using NcoI/XbaI cut sites. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (Berkeley, CA) or the Barker Hall Sequencing Facility at UC Berkeley (Berkeley, CA).

Expression testing of His₆MBP-tagged BesB variants for solubility. Expression testing of His$_6$MBP-BesB was performed by the University of California, Berkeley Macro Lab (Berkeley, CA). Vector pSV272.1-BesB was transformed into *E. coli* Rosetta2(DE3) pLysS, *E. coli* BL21(DE3) CodonPlus RIL, *E. coli* BL21 Star (DE3), E. coli SoluBL21, and E. coli Shuffle T7 Express. Transformants were grown overnight in autoinducing Magic Media (Invitrogen), in a 96-well block at 37°C at 200 rpm. Cultures were them moved to 16°C and allowed to continue growing for an additional 40 h. Cells were lysed using BugBuster (EMD Millipore) and proteins purified using Ni-NTA resin (Qiagen). BL21 Star (DE3) was selected as the production host from these studies based on yield of soluble protein and the lack of need for a secondary antibiotic.

General procedure for high resolution HPLC/MS analysis of polar metabolites. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1 × 100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

General procedure for high resolution HPLC/MS analysis of non-polar metabolites. Samples containing non-polar metabolites were analyzed using an Agilent 1290 UPLC on a Poroshell 120 EC-C18 (5 μm, 2.1 × 30 mm; Agilent) using the following buffers: Buffer A (100% water, 0.1% v/v formic acid) and Buffer B (100% Acetonitrile). A linear gradient from 100% to 0% Buffer A over 5 min was applied at a flow rate of 0.6 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

***In vivo* production of Pra in *E. coli* from 4-Cl-allylglycine.** In experiments where *in vivo* BesB flux was being optimized, a general procedure was followed. Single colonies of E. coli BL21 Star) transformed with plasmids for truncation screening (p16B-BesB-Trunc-X), solubility tag screening (pSV272.1-BesB, pET32a-BesB, or pet41b-BesB), promoter variants (pET16-BesB, pBAD24-BesB, pTrc33-BesB, or p16b-glnS-BesB), codon optimized variant (pSV272.1-sBesB), chaperone screening (pET16b-sBesB co-expressed with pGro7, pG-KJE8, pKJE7, pGTf2, or pTf16) were inoculated into TB (10 mL) containing the appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into TB (25 mL, pH 6.5) in a 250 mL-baffled shake flask to OD$_{600}$ = 0.05 and grown at 37°C and 200 rpm until the OD$_{600}$ reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.2 mM) and adding 4-Cl-allylglycine (1 mM). Cultures were grown 24-72 h at 16°C and 200 rpm. At time points where Pra production was being measured, supernatant samples were analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis.

Adaptive evolution of an acid-tolerant *E. coli* BL21 strain. Single colonies of *E. coli* BL21 Star (DE3) transformed with pET16b-BesD were inoculated into TB (10 mL, pH 5.0) containing the appropriate antibiotics and grown at 37°C with shaking at 200 rpm for 48 h. 100 μL of cell culture was then passaged into fresh TB (10 mL, pH 5.0), and grown again at 37°C with shaking at 200 rpm for 48 h. E. coli were passaged 3 and 6 more times to generate acid tolerate strains after the 4th and 7th passages, which were collected and frozen in 25% glycerol, -80C, for further testing.

Expression and Purification of BesD and BesC. General procedure for expression of BesD and BesC can be found in Chapter for under section "Expression of His$_6$MBP-, His$_{10}$-, and Strep-tagged proteins", while procedure for purification can be found under "Purification of His$_{10}$-BesA, His$_{10}$-BesD, His$_{10}$-BesE, and His$_6$MBP-BesC.

General procedure for *in vivo* production of 4-Cl-lysine, 4-Cl-allylglycine, allylglycine, Pra biosynthesis in E. coli. Single colonies of *E. coli* BL21 Star (DE3) transformed with pET16b-BesD (from *Pseudomonas fluorescens*, WP_016975823), pET16b-BesC (from *Pseudomonas fluorescens*, WP_080628534), pETDUET-BesC.BesD, or pPra were inoculated into TB (10 mL, pH 6.5) containing the appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into TB or LB (25 mL, pH 6.5 if 4-Cl-lysine production were required) in a 250 mL-baffled shake flask to OD$_{600}$ = 0.05 and grown at 37°C and 200 rpm until the OD$_{600}$ reached 0.6 - 0.8. Cultures were cooled on ice for 20 min before inducing with IPTG (0.2 mM) and then cultured for 72 h at 16°C and 200 rpm. Supernatant samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis.

General procedure for Pra incorporation into the *E. coli* proteome using PraRS. Single colonies of *E. coli* B834(DE3) transformed with pBAD24-PraRS are grown overnight in LB with appropriate antibiotics at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into LB with appropriate antibiotics (25 mL) with 0.2% (w/v) L-arabinose in a 250 mL-baffled shake flask to OD$_{600}$ = 0.05 and grown at 37°C and 200 rpm until the OD$_{600}$ reached 0.6 to 0.8. Cells were washed twice with 0.9% (w/v) NaCl by pelleting and resuspension. The cell pellet was then resuspended in M9 media containing 33.7 mM Na2HPO4, 22 mM KH2PO4, 8.55 mM NaCl, 9.35 mM NH4Cl, 0.4% glucose, 100 mg/L of the 19 proteinogenic amino acids (-Met), Pra (100 μM – 1 mM), 1 mM MgSO4, 300 μM CaCl2, and 1 mg/ml thiamine to OD$_{600}$ = 0.4 - 0.6 and cultured for 24 – 48 h at 16°C and 200 rpm. Evidence of Pra incorporation can be seen as early as 30 min after adding Pra, with increased labeling observed over time.

***In vivo* incorporation of Pra into the *E. coli* proteome.** Single colonies of *E. coli* B834(DE3) transformed with appropriate plasmids (Empty plasmid control: pETDUET-1/pGro7, no endogenous Pra control: pETDUET-1/praGro, no PraRS control: pPra/pGro7, or pPra/praGro) were inoculated into LB Cb Cm (10 mL) and grown overnight at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into LB with appropriate antibiotics (25 mL) with 0.2% (w/v) L-arabinose in a 250 mL-baffled shake flask to OD$_{600}$ = 0.05 and grown at 37°C and 200 rpm until the OD$_{600}$ reached 0.6 to 0.8. Cultures were cooled on for 20 min before inducing with IPTG (0.2 mM) and then cultured for 48 h at 16°C and 200 rpm. Cells were washed twice with 0.9% (w/v) NaCl by pelleting and resuspension. The cell pellet was then resuspended in M9 media containing 33.7 mM Na$_2$HPO$_4$, 22 mM KH$_2$PO$_4$, 8.55 mM NaCl, 9.35 mM NH$_4$Cl, 0.4% glucose, 100 mg/L of the 19 proteinogenic amino acids (-Met), 50 mg/L Met, 1 mM MgSO$_4$, 300

μM CaCl$_2$, and 1 mg/ml thiamine to OD$_{600}$ = 0.4 - 0.6 and cultured for 24 – 48 h at 16°C and 200 rpm. For positive controls, Pra (0.2 mM) was also added.

Derivatization and visualization of alkyne-labeled proteins with Tamra-azide. Cell cultures (1 mL) were collected and centrifuged at 9,800 × g for 7 min at 4ºC. The resulting cell pellet was resuspended in SoluLyse (150 μL) and incubated for 1 h at room temperature. Samples were centrifuged at 9,800 × g for 5 min and the supernatant (50 μL) was mixed with an assay solution (50 μL) containing Tamra-azide (40 μM), THPTA (100 μM), and sodium ascorbate (10 mM) in PBS (pH 7) with or without CuSO4 (50 μM). The labeling reaction was incubated for 1 h at room temperature. Laemmli buffer was added and samples were heated at 95ºC for 5 min before analyzing by SDS-PAGE (MiniProtean 8-16% TGX gels; BioRad) using a constant current (30 A) and maximum voltage (200 V) for 1.5 h. Gels were imaged for fluorescence using a ChemiDoc MP Imager (Biorad; λex = 520-545 nm, λem = 577-613 nm, exposure = 150 s).

Proteomic analysis of *E. coli* with endogenous Pra production. Cell pellets for *E. coli* B834(DE3) pPra praGro and B834(DE3) pETDUET-1 pGro7 were submitted to UC Davis Proteomics Core at the UC Davis Genome Center for bottom-up proteomic analysis as positive and negative controls, respectively. Total protein was extracted and subjected to tryptic digest and iodoacetamide treatment before bottom-up proteomic analysis using a Thermo LC-Q Exactive. Following tandem mass spectra extraction, charge state deconvolution, and deisotoping, MS/MS spectra were analyzed using X! Tandem (The GPM; Tandem Alanine 2017.2.1.4). X! Tandem was searched with a fragment ion mass tolerance of 20 ppm and a parent ion tolerance of 20 ppm. The following modifications were specified as fixed: carbamidomethyl of Cys and Se-Cys (iodoacetamide cap). The following modifications were specified as variable: Met-36 (Met → Pra) of Met and the N-terminus, Glu → Pyro-Glu of the N-terminus, ammonia loss from the N-terminus, Gln → Pyro-Glu of the N-terminus, deamidation of Asn and Gln, oxidation of Mn and Trp, and dioxidation of Met and Trp. Scaffold 4.8.7 (Proteome Software Inc.; Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 99.0% probability by the Scaffold Local FDR algorithm. Peptide identifications were also required to exceed specific database search engine thresholds. Protein identifications were accepted if they could be established at greater than 99.0% probability to achieve an FDR less than 5.0% and contained at least 1 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Amber suppression experiments with pEVOL-MetRS-CUA. Single colonies of *E. coli* BL21 (DE3) co-transformed with pEVOL-Met-CUA and pREP-GFP-CUA or pREP-GFP are grown overnight in LB with appropriate antibiotics at 37°C with shaking at 200 rpm. Overnight cultures were inoculated into LB with appropriate antibiotics (25 mL) with 0.2% (w/v) L-arabinose, induce expression of MetRS-CUA, in a 250 mL-baffled shake flask to OD$_{600}$ = 0.05 and grown at 37°C and 200 rpm until the OD$_{600}$ reached 0.6 to 0.8. Cultures were cooled on for 20 min before inducing with IPTG (0.2 mM) and then cultured for 24h. GFP fluorescence monitored by measuring fluorescence (λ$_{ex}$ = 488 nm, λ$_{em}$ = 509nm) using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature with 485nm/20nm excitation and 530nm/25nm emission filter sets.

Tetrazine-ligation reactions with allylglycine. L-Allylglycine (100 µM) was added to a reaction containing Cy5-Tetrazine (1 mM) with 50 mM HEPES (pH 7.5) in water and allowed to react for 1-24h at room temperature or 90 °C. As negative controls, either L-allylglycine or Cy5-Tetrazine was left out of the reaction. Samples were then analyzed using general procedure for non-polar metabolite analysis described previously.

Tetrazole-photoclick reactions with allylglycine. L-Allylglycine (100 µM) was added to a reaction containing 5-(3-Bromo-phenyl)-2H-tetrazole (1 mM) with 50 mM HEPES (pH 7.5) in water. The Br in 5-(3-Bromo-phenyl)-2H-tetrazole facilitates detection by MS since Br has a unique isotopic signature. A UV lamp was used to shine light ($\lambda = 302$nm) on the sample for 10 minutes. As negative controls, either L-allylglycine or 5-(3-Bromo-phenyl)-2H-tetrazole was left out of the reaction. Samples were then analyzed using general procedure for non-polar metabolite analysis described previously.

3.3. Results and discussion

Optimization of 4-Cl-lysine production in *E. coli*. While we had been successful in reconstituting Pra biosynthesis *in vitro*, obtaining detectable production of βes pathway metabolites proved to be challenging *in vivo*. Attempting to heterologously express BesD from *S. cattleya* in *E. coli* yielded only trace amounts (<5 μM) of 4-Cl-lysine in the culture broth (*Figure 3.2*) after 6 h, which became depleted after 24 h (*Figure 3.2 & 3.3*). Using BLAST, we identified homologues of BesD that had a high degree of similarity (>50%) in sequence, but were found in *Pseudomonas* spp. While expressing BesD homologue from *Pseudomonas fluorescens* did not lead to meaningful increase in 4-Cl-lysine titers, we opted to use this gene for *in vivo* optimization since it lead to higher soluble expression in *E. coli* than the *S. cattleya* variant.
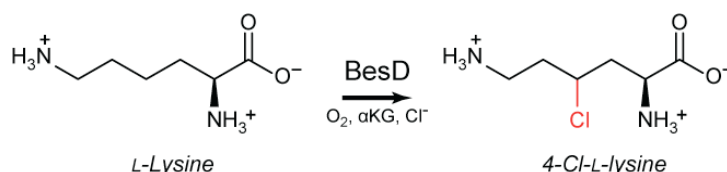
Figure 3.2. Conversion of L-lysine to 4-Cl-L-lysine by BesD.
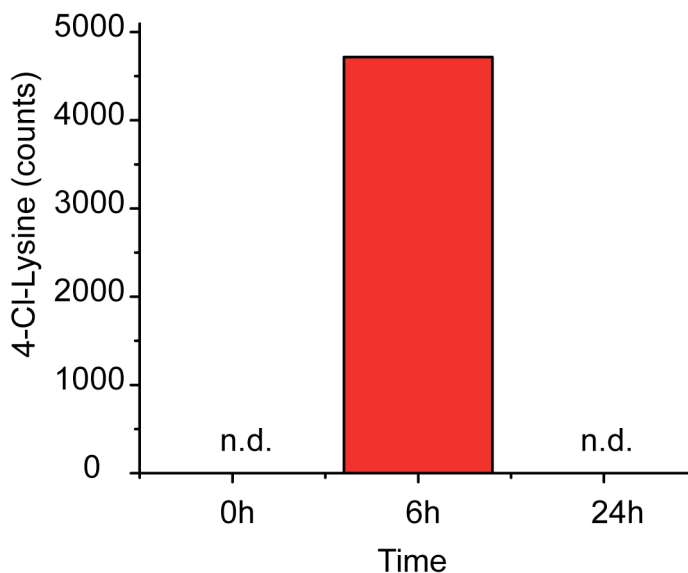
Figure 3.3. Integrated area for EIC of 4-Cl-lysine ([M+H]+ m/z = 181.0738) at varying points after expression of BesD (from *S. cattleya*) in *E. coli* BL21. Data shown is from metabolites extracted from cell pellets. The amount of 4-Cl-lysine detected is estimated to be below 5 μM.

Previously, we had identified a decomposition pathway for 4-Cl-lysine, whereby in slightly basic conditions 4-Cl-lysine undergoes a series of intramolecular rearrangements including cyclization to form a 5-membered lactone, ring expansion to form a 7-membered lactam, and hydrolysis to form 4-OH-lysine (*Figure 2.15B*). We hypothesized that this decomposition pathway was physiologically relevant *in vivo*, and perhaps the reason we were unable to produce detectable

92
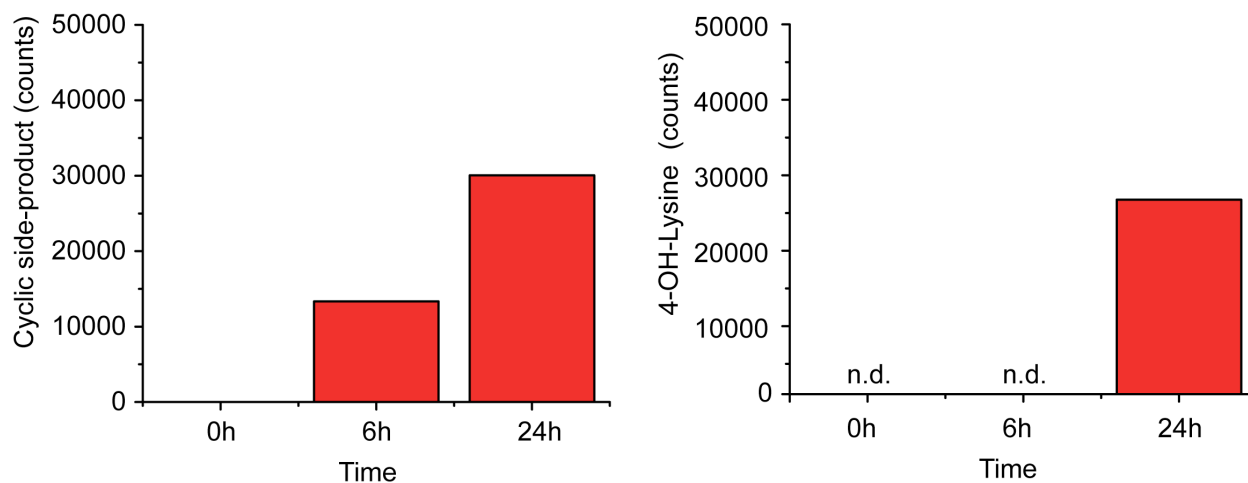
Figure 3.4. Integrated area for EIC of lactone/lactam cyclic side product ([M+H]+ m/z = 145.0927) and 4-OH-lysine ([M+H]+ m/z = 163.1077) at varying points after expression of BesD (from *S. cattleya*) in *E. coli* BL21. Data shown is from metabolites extracted from cell pellets.
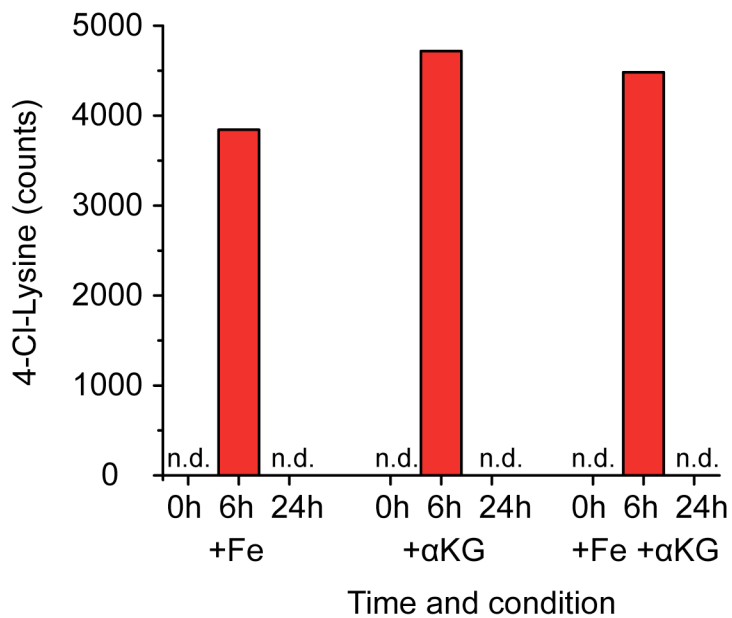
Figure 3.5. Integrated area for EIC of 4-Cl-lysine ([M+H]+ m/z = 181.0738) at varying points after expression of BesD (from *S. cattleya*) in *E. coli* BL21 in conditions where Fe (50 μM), αKG (5 mM), or both are added to production cultures. Data shown is from metabolites extracted from cell pellets. The amount of 4-Cl-lysine detected is estimated to be below 5 μM.

amounts of 4-Cl-lysine. Indeed, when search for decomposition metabolites in *E. coli* culture broths expressing BesD, we observed increasing amount of decomposition products compared to 4-Cl-lysine (*Figure 3.4*).

In an effort to increase *in vivo* productivity of BesD, we fed *E. coli* cultures excess amounts of the necessary co-factors and substrates required in BesD catalysis; namely Fe, αKG, and lysine. However we saw little change in 4-Cl-lysine production when Fe, αKG, or both were fed in excess, suggesting there may be other reasons why BesD is unable to accumulate (*Figure 3.5*).

As mentioned, we had previously observed that decomposition of 4-Cl-lysine accelerates in basic conditions (*Figure 2.15A*). We then decided to measure the pH of *E. coli* grown in terrific broth (TB) to investigate the range of pH *E. coli* is exposed to during BesD production. While TB media is initially pH 7.2 (as prepared), over 72 h the pH increased to 8.5 (*Figure 3.6*). We then decreased the initial pH of TB to 6.5, 6.0, 5.5, and 5.0 and measured production of 4-Cl-lysine. In all four of these conditions, titers of 4-Cl-lysine increased at least 25-fold (as measured by 4-Cl-lysine EIC counts) after 24h (*Figure 3.7.*). The highest total production of 4-Cl-lysine was observed at pH 6.5. Cultures grown at pH 5.5 had the highest specific production (counts of 4-Cl-lysine / OD_{600}), but grew to an OD_{600} of 2 while cultures at pH 6.5 grew to an OD_{600} greater than 8 (*Figure 3.7D*).
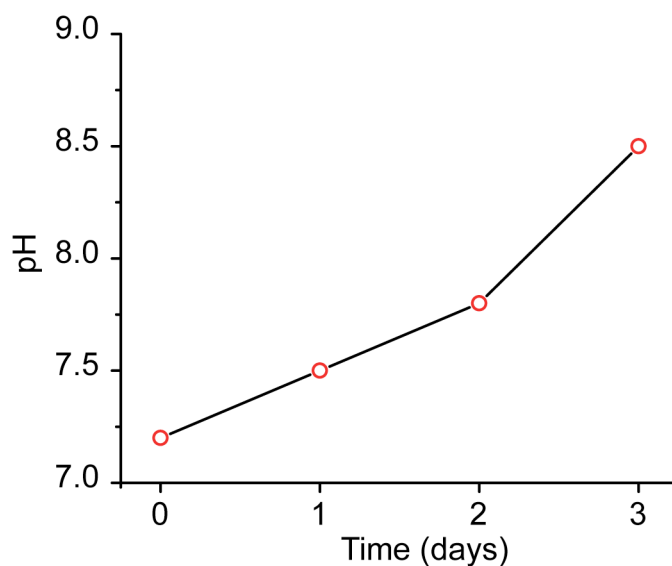
Figure 3.6. Example of a pH profile for a production culture of *E. coli* BL21 grown in TB. Cultures become more alkali over time.

To improve *in vivo* 4-Cl-lysine production, we hypothesized that if we could adapt *E. coli* to grow to higher OD_{600} in acidic conditions (< pH 5.5), then we should be able to obtain even higher titers. We seeded cultures at pH 5.0 and passaged them daily. While the initial OD_{600} after 24 h for the parent strain was < 1, after four passages cells were growing to an OD_{600} > 8 (*Figure 3.8A*). Unfortunately, we did not observe an increase in 4-Cl-lysine production as we had expected **(*data not shown*)**. Upon further investigation, we discovered that our acid-adapted *E. coli* cultures were more efficient at neutralizing the pH of the medium. After 24 h, acid-adapted *E. coli* cultures seeded at pH 5.0 had reached a pH > 6, and pH > 7 after 48h, whereas the parent strain of *E. coli* maintained a pH of 5.0 throughout the course of the experiment (*Figure 3.8B*).
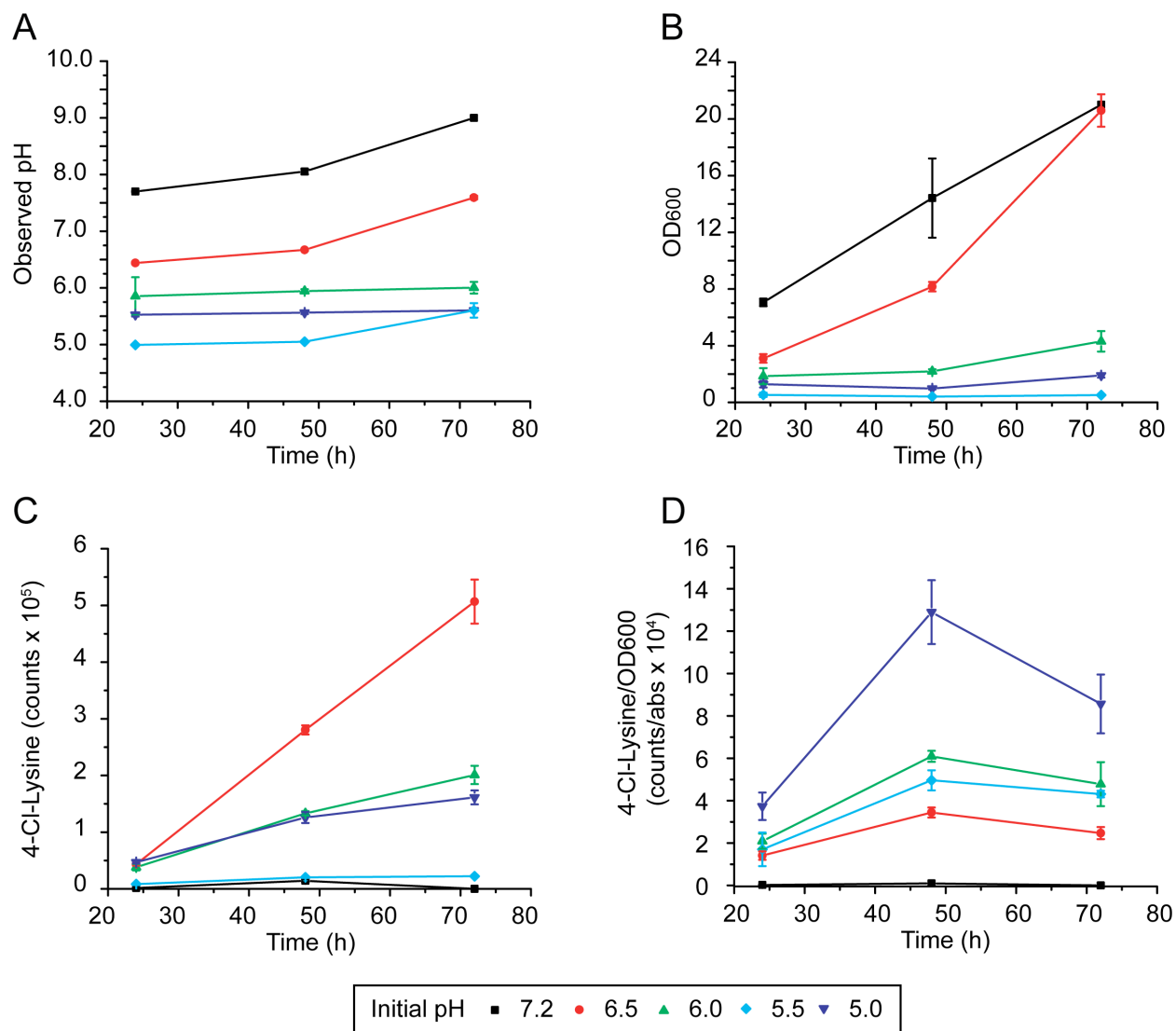
Figure 3.7. Production of 4-Cl-lysine in acidic media. *E. coli* BL21 pET16b-BesD were cultured in TB at varying initial pH: 7.2 (as prepared), 6.5, 6.0, 5.5, and 5.0. (A) Observed pH over time for production culture over the course of 72 h. (B) OD_{600} of production cultures over the course of 72h. (C) Integrated area for EIC of 4-Cl-lysine ($[M+H]^+$ m/z = 181.0738). Samples are from culture supernatants. (D) Integrated area of 4-Cl-Lysine over OD_{600}, used as an estimate of specific productivity, for production cultures over 72 h. Error bars shown are standard deviation of 3 experimental replicates.

95
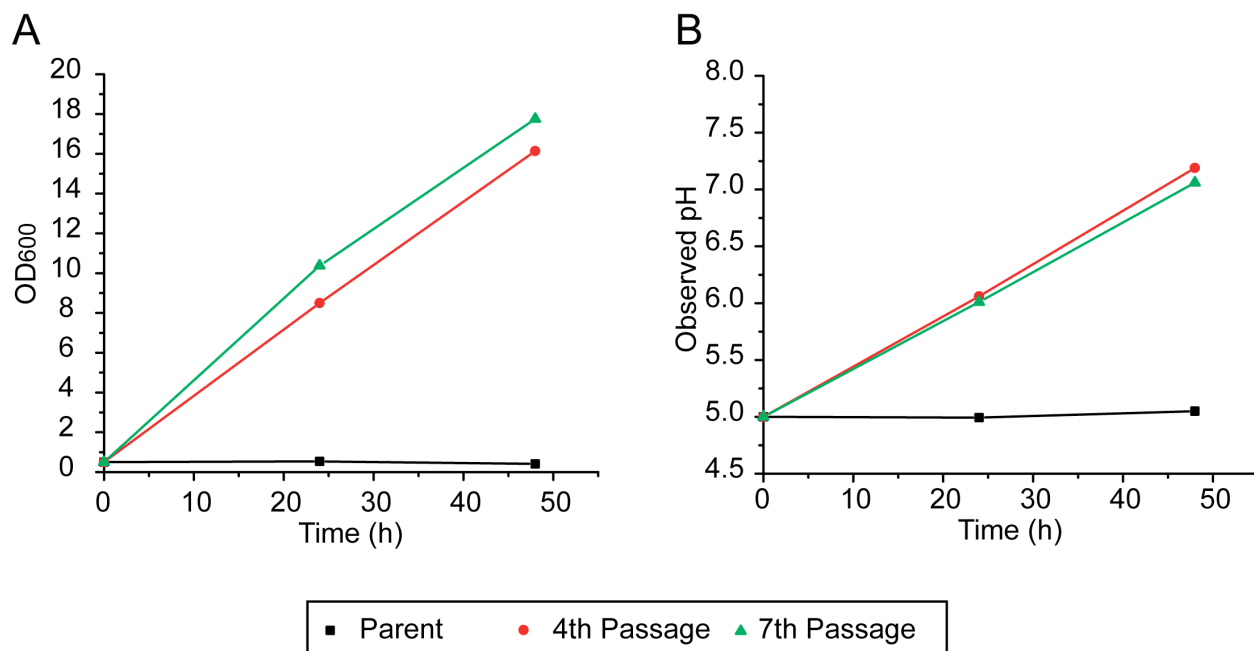
Figure 3.8. Adaptive evolution of an acid-tolerant *E. coli* strain. *E. coli* BL21 (DE3) pET16b-BesD were cultured at in TB (pH 5.0) and passaged every 48 h to fresh TB (pH 5.0). (A) Observed OD_{600} of cell cultures showing that acid-adapted *E. coli* can grow to higher OD600 than parent strain. (B) Observed pH of cell cultures showing that acid-adapted *E. coli* is more effective at neutralizing the pH of media as compared to parent.
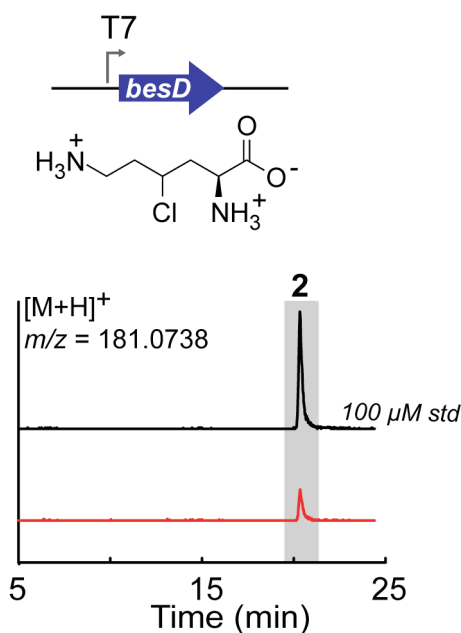
Figure 3.9. Extracted ion chromatograms showing *in vivo* production of 4-Cl-lysine (**2**, *m/z* 181.0738), in *E. coli* BL-21 Star (DE3) pET16b-BesD, grown in LB. 4-Cl-lysine titers of 50 µM are reached within 48 h by expression of *besD* from *P. fluorescens*. Chromatograms shown are representative of at least 3 experimental replicates.

Through optimization of media conditions, we were able to obtain 4-Cl-lysine titers of approximately 50 µM in *E. coli* after 48 h (*Figure 3.9*). Follow-up *in vivo* production experiments that require BesD were therefore performed in TB at pH 6.5 or in LB (pH 6.5 as prepared), and used the BesD homologue from *P. fluorescens*.

Optimization of allylglycine production in *E. coli*. In the βes pathway, BesC catalyzes the oxidative cleavage of 4-Cl-lysine to 4-Cl-allylglycine. In the absence of 4-Cl-lysine, BesC can catalyze the conversion of lysine to allylglycine (*Figure 3.10*). From expressing BesC for *in vitro* experiments, we had noticed a considerable increase in protein yield (4 mg/L → 40 mg/L) if 'rare' tRNAs, using pRARE2, were also expressed. When *E. coli* pET16b-BesC (from *S. cattleya*) was co-transformed pRARE2, we saw allylglycine titers of approximately 50 µM (*Figure 3.11 & 3.12*). Switching to the *P. fluorescencs besC* gene gave us similar results to using *S. cattleya besC* + pRARE2 (*Figure 3.11*). We noticed that when detecting using our traditional LC-MS analysis methods on ZIC-pHILIC, proline (Pro, same nominal mass as allylglycine, $[M+H]^+$ m/z = 116.0706) eluted 10-30s earlier than allyglycine (*Figure 3.12*). Depleting Pro in the media, by growing cells in M9 with no Pro, or running the ZIC-pHILIC without guard column were quick workarounds implemented to detect allylglycine without overlapping peaks. Subsequent *in vivo* pathways that utilize BesC use the *P. fluorescens* homologue.

Optimization of 4-Cl-allylglycine production in *E. coli*. In the native βes pathway, BesC oxidatively cleaves 4-Cl-lysine between C_δ and C_ε to form 4-Cl-allylglycine (*Figure 3.13*). When co-expressing BesD and BesC from *S. cattleya* in TB, we were unable to detect 4-Cl-allylglycine in culture broths. Surprisingly, we were able to produce 4-Cl-allylglycine by combining lysates of BesD- and BesC-overexpressing *E. coli* cells (*Figure 3.14*). One possibility for the observed *in vitro* activity is that the pH of the lysis buffe (pH 7.0) was low enough to support accumulation of 4-Cl-lysine. These lysate experiments were useful in helping us understand which cofactors or substrates might be limiting in the BesD-BesD pathway. Ultimately, co-expression of BesD and BesC homologues from *P. fluorescens* in a slightly acidic media (TB pH 6.5 or LB) were sufficient to obtain considerable yields of 4-Cl-allylglycine *in vivo* (*Figure 3.15*). We were able to see titers upwards of 200 µM if cultures were grown for > 3d, suggesting this pathway could be used for sustain high levels of amino acid production.

Optimization of Pra production in *E. coli*. Our first attempt at producing Pra in *E. coli* consisted of overexpressing *besD*, *besC*, and *besB* from *S. cattleya* in TB media (pH 7.2). Under these conditions, we saw no detectable amount of Pra in culture broth after 72h (*Figure 3.16*). Using BesD and BesC from *P. fluorescens*, and BesB from *S. cattleya,* and TB media at pH 6.5 we were able to achieve production of ~25 µM Pra after 48h, and ~50 µM after 72 h (*Figure 3.16-17*). While an improvement over our initial set of conditions, the majority of 4-Cl-allylglycine (> 100 µM) remained unconverted. The remainder of this section will focus on improving *in vivo E. coli* Pra titers by considering BesB as the likely bottleneck of this engineered metabolic pathway. In optimizing this step, 4-Cl-allylglycine is fed cultures.

Previously, we had found homologues of *bes* genes (*besD* and *besC* from *P. fluorescens*) that expressed better in *E. coli*. A pBLAST search of BesB homologues yielded 6 other close matches – all from other species of streptomycetes that contain the βes/Pra pathways (*Table 3.1*). To find the best expressing BesB variant, we performed an initial expression screen of the orthologs from *S. catenulae, S.* sp. NRRL S-31, *S.* sp. NRRL S-1448, and *S. achromogenes*. We
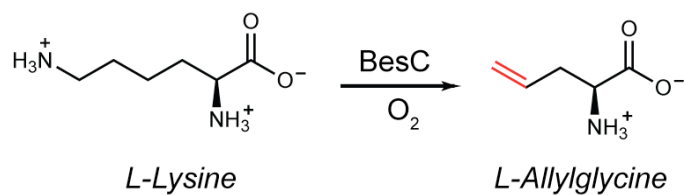
Figure 3.10. Conversion of L-lysine to L-allylglycine by BesC.
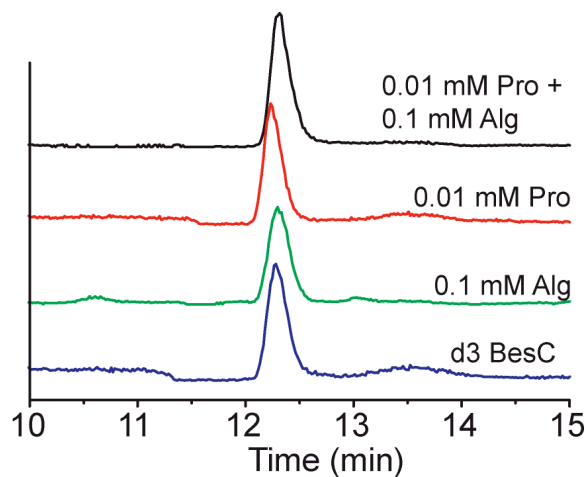
Figure 3.11. Extracted ion chromatograms showing *in vivo* production of allylglycine (m/z = 106.0706), in *E. coli* BL-21 Star (DE3), expressing either *besC* (from *S. cattleya*) with pRARE2 or *besC* from (*P. fluorescens*). Proline has the same monoisotopic mass as allylglycine and elutes at a similar retention time.
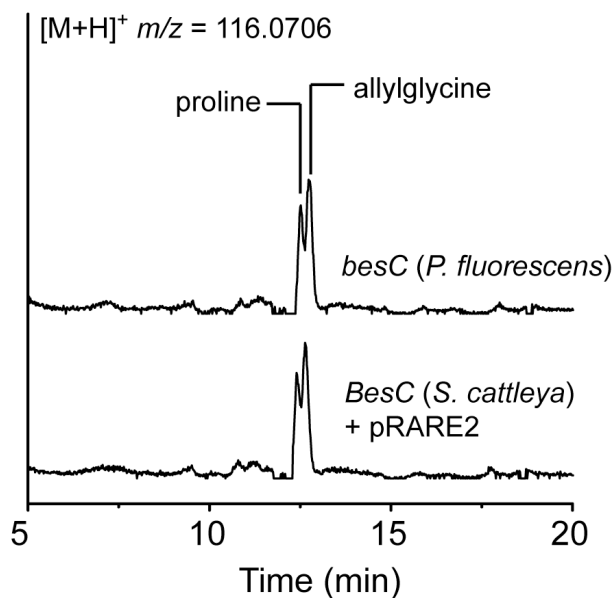
Figure 3.12. Extracted ion chromatograms for allylglycine/proline (m/z = 116.0706) highlighting the similar retention time between both compounds. Standards of proline and allylglycine were used to confirm identity of peak, with proline eluting slightly before allylglycine. Representative day 3 culture of BL21 expressing *besC* from S. cattleya with pRARE2, in M9 (-Pro) media is also shown.
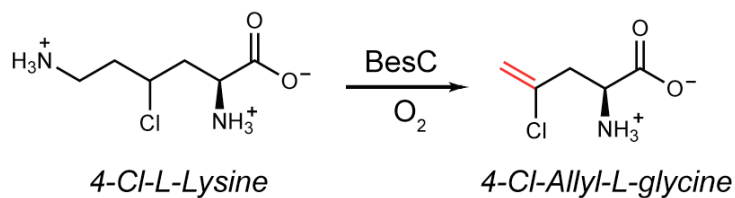
Figure 3.13. Conversion of 4-Cl-L-lysine to 4-Cl-L-allylglycine by BesC.

99
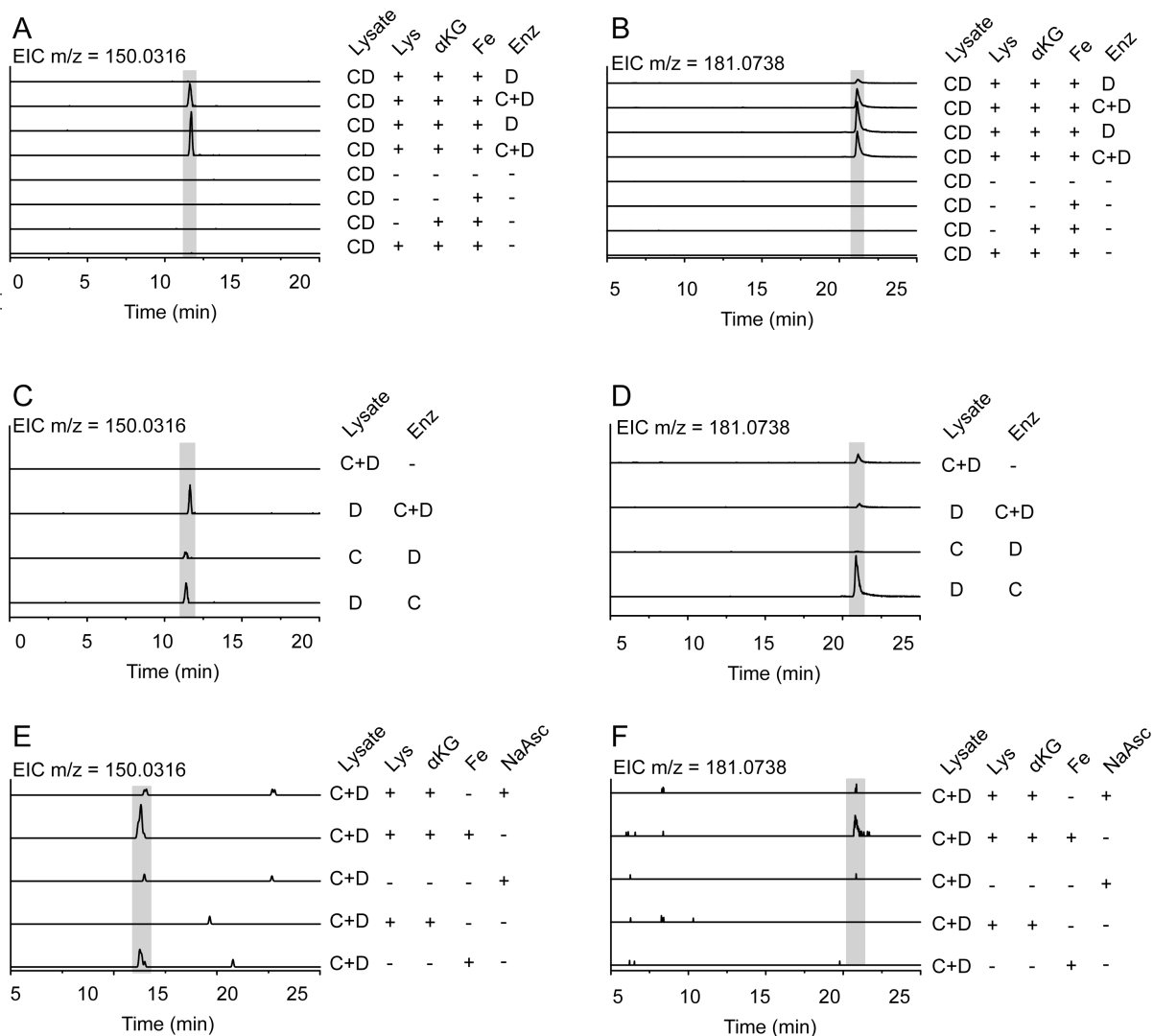
Figure 3.14. Lysate assay to determine which enzyme or co-factor my be rate limiting in vivo for 4-Cl-allylglycine production. *E. coli* BL21 cultures expressing BesD, BesC or both BesC and BesD from (*S. cattleya*) were lysed by bead beating and centrifuged to separate soluble fraction. Lysine (1 mM), αKG (5 mM), sodium ascorbate (5 mM), ammonium iron sulfate (50 μM), purified BesD (5 μM), purified BesC (5 μM), or other soluble lysate fractions (from cells expressing BesC, BesD, or both) were added to the soluble fraction and allowed to react at room temperature for 2h. LC-MS was then used to analyze samples for either 4-Cl-allylglycine (*m/z* = 150.0316) or 4-Cl-lysine (*m/z* = 181.0738). (A) 4-Cl-allylglycine or (B) 4-Cl-lysine produced when lysates co-expressing BesC and BesD are mixed with additional lysine, αKG, Fe, or purified BesC/BesD. (C) 4-Cl-allylglycine or (D) 4-Cl-lysine produced when lysates expressing BesC, BesD, or both BesC and BesD are mixed with purified BesD, BesC, or both BesD and BesC. (E) 4-Cl-allylglycine or (F) 4-Cl-lysine produced when lysates BesC is mixed with lysate expressing BesD are additional lysine, αKG, Fe, or NaAsc is added.

100
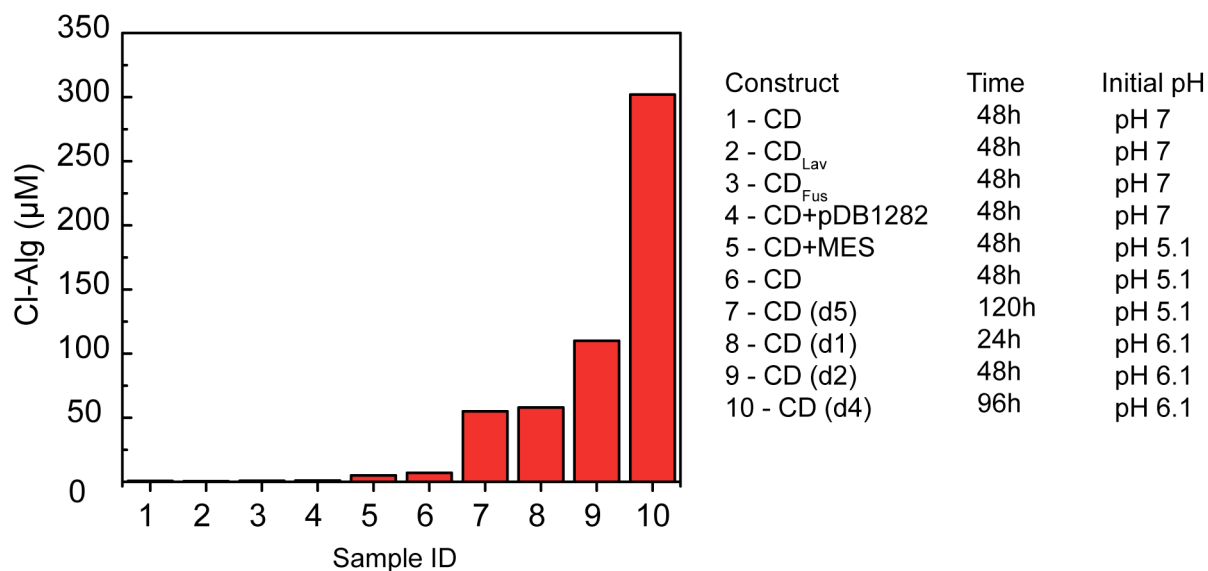
Figure 3.15. 4-Cl-allylglycine titers in optimized media conditions. E. coli BL21 expressing a variety of petDUET-BesD-BesC constructs at varying initial media pH. Constructs are as followed: CD = petDUET-BesD-BesC from *P. fluorescens*; CD_{Lav} = petDUET-$BesD_{Lav}$-BesC where BesD is S. lavenduligriseus homologue and BesC is from *P. fluorescens;* CDfus = pET16bCDfus where BesC fused at the N_T of BesD and both genes are from *P. fluorescens;* CD+pDB1282 = petDUET-BesD-BesC from *P. fluorescens* with co-expression of pDB1282; CD+MES = = petDUET-BesD-BesC from *P. fluorescens* with added MES (100 mM) as buffer.
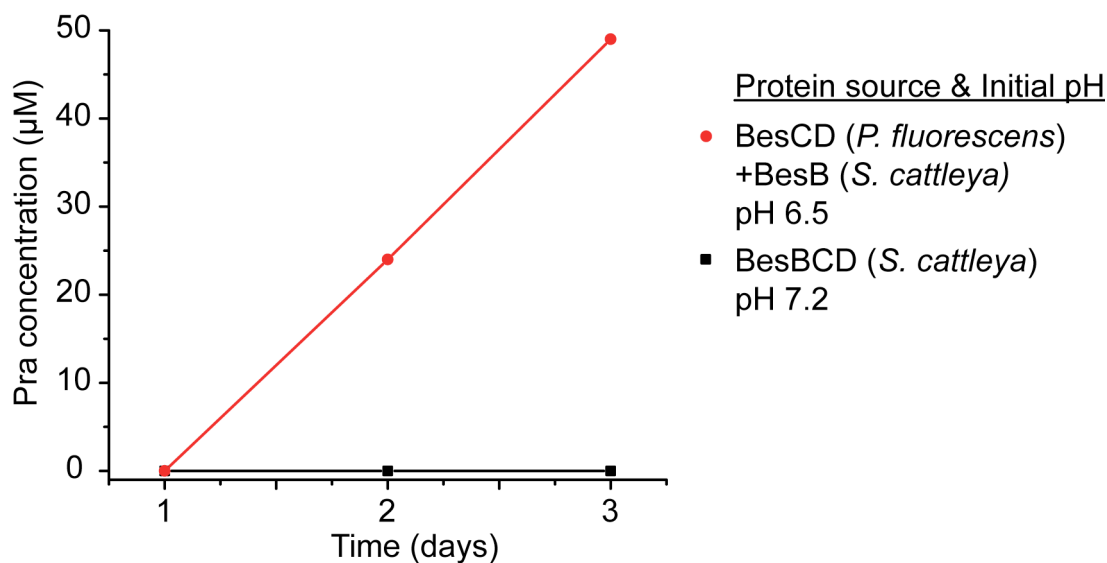
Figure 3.16. Pra titers in optimized media conditions for production of 4-Cl-allylglycine. *E. coli* BL21 expressing BesC and BesD from *P. fluorescens* and BesB from *S. cattleya* grown in TB (initial pH 6.5) is able to make ≈ 50 µM Pra after 3 days, while no Pra is detected when using BesBCD from *S. cattleya* and culturing cells in TB (initial pH 7.2).

found that BesB from *S.* sp. NRRL S-1448 had the highest level of soluble expression (*Appendix Figure A2.5*). Outside the genus Streptomyces, additional BesB homologues in *Opitutaceae bacteria*, *Leptospira spp.*, *Verruucomicrobia bacterium* were also identified (*Table 3.1*). However, we were unable to observe any *in vivo* conversion of 4-Cl-allylglycine to Pra when these genes were heterologously expressed in *E. coli*. We next decided to screen various expression strains of *E. coli* to see if any would lead to higher soluble expression of BesB. We tested Rosetta2 (DE3) pLysS, BL21 (DE3) CodonPlusRIL, BL21 (DE3) Star, SoluBL-21, and Shuffle T7 Express *E. coli* strains and found that BL-21 (DE3) Star, an *rne* truncation mutant, gave the highest amount of soluble expression (*Figure 3.18*). We therefore moved forward with using *besB* from *S.* sp. NRRL S-1448 for further optimization, and used BL21 (DE3) Star as our expression strain of *E. coli*.

While BesB from *S.* sp. NRRL S-1448 expressed better than the *S. cattleya* BesB homologue, most of the protein expressed was showing up in the insoluble fraction. One way to improve soluble protein expression is by attaching an N_T-solubility tag. We tested three N_T-solubility tags: maltose binding protein (MBP, pSV272.1), glutathione-S transferase (Gst, pET41b), and thioredoxin (Trx, pET32a). We observed the greatest improvement *in vivo* Pra production when MBP was used as the N_T tag, corresponding to approximately 2 fold increase over untagged BesB (data not shown). Therefore, in experiments where the full pathway is reconstituted in *E. coli*, an MBP fusion of BesB was used.
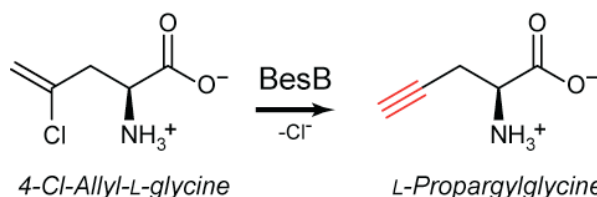
Figure 3.17. Conversion of 4-Cl-allylglycine to L-propargylglycine by BesB.

Table 3.1. BesB homologues found by protein BLAST against the NCBI non-redundant protein database. Source organism, protein accession number, and percent identity/similarity (%ID/%S) shown.

Organism	Accession	%ID/%S
Streptomyces catenulae	WP_030285993.1	73/79
Streptomyces alboverticillatus	WP_086569406.1	69/78
Streptomyces sp. NRRL S-31	WP_030744376.1	69/77
Streptomyces sp. NRRL S-1448	WP_030410683.1	73/81
Streptomyces lavenduligriseus	WP_078637771.1	69/76
Streptomyces achromogenes	WP_078844341.1	68/75
Opitutaceae bacterium TSB47	WP_068769754.1	42/59
Leptospira kirschneri	WP_020763855.1	35/56
Verrucomicrobia bacterium	RFC49611.1	46/62

Independent of attaching an N_T solubility tag, we also attempting truncating the N_T of BesB to see if we could generate a more soluble BesB variant. BesB is a 511 amino acid protein, with a conserved PLP domain beginning around residue 225, with the catalytic lysine is located at residue 322. Homology models of BesB (Phyre2) place the N_T domain far away enough from the catalytic center that it might not be required for activity. We therefore generated N_T truncations of BesB Δ5aa, Δ10aa, Δ15aa, Δ20aa, Δ25aa, Δ30aa, Δ35aa, Δ40aa, Δ45aa, and Δ50aa. We then tested each of these truncations for its ability to convert 4-Cl-allylglycine to Pra *in vivo* (*Figure 3.19*). All variants tested displayed reduced (Δ5aa, Δ10aa) or no activity (all other truncations) compared to wildtype.

Promoter strength is thought to have a large impact on heterologously expressed proteins. Using a strong promoter, such as T7, can lead to high protein yields but may also lead to an increase insoluble protein [23]. In some cases, weaker promoters can improve soluble protein yield by decreasing the translational stress inside the cell [24]. We therefore tested the effects of weakening promoter strength through two strategies. Our first approach relied on testing out a variety of promoters of varying strength (Strongest to Weakest: T7, Trc, Ara, GlnS). Of these promoters tested, T7 gave the highest conversion of 4-Cl-allylglycine to Pra (*Figure 3.20*). For the second approach, we used an *E. coli* strain with a lac permease mutation that switches T7 induction to be dependent on IPTG concentration. We found conversion of 4-Cl-allylglycine to Pra to be highest at lower concentrations of IPTG (*Figure 3.21.*), suggesting that perhaps slightly tuning down T7 promoter strength could have beneficial effects for our pathway. While the results of promoter strength testing were informative, we made no changes to the expression system we were using based on these experiments and continued using T7 promoter.
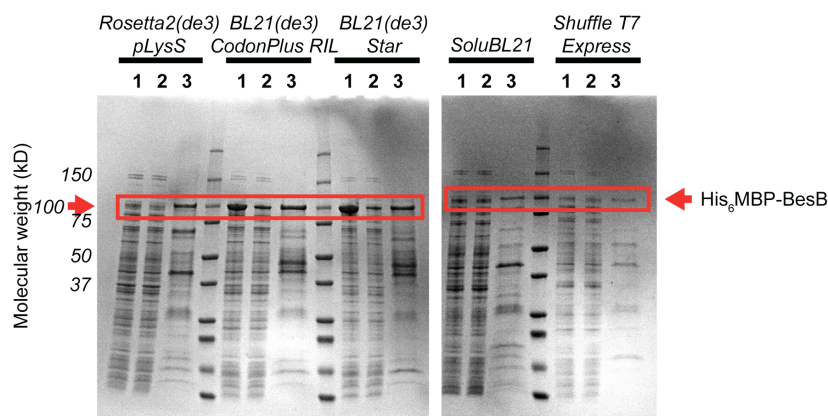
Figure 3.18. SDS-PAGE of His₆MBP-BesB solubility screen from *S.* sp. NRRL S-1448. BesB from *S. cattleya* was found to be highly insoluble, therefore the orthologue *S.* sp. NRRL S-1448 was cloned and purified. Lanes correspond for each cell line tested (**1**, pre-induction; **2**, post-induction; **3**, elution from Ni-NTA column. Highest, soluble expression observed in *E. coli* BL21(DE3)-Star.

At this point in BesB expression optimization, we decided to codon optimize the *S.* sp. NRRL S-1448 *besB* gene for *E. coli* expression (*sbesB*, *Appendix A1.1*). Codon optimization is thought to alleviate cellular requirements for rarely used tRNAs, and can help improve protein expression [25]. Consistent with our initial screen of BesB expression in *E. coli* Rosetta2 and CodonPlus strains, we did not see a significant increase in 4-Cl-allylglycine conversions when the codon-optimized variant was used (data not shown). Nonetheless, we continued to work with this synthetic variant since the reduced GC content of the *sbesB* gene facilitated cloning.
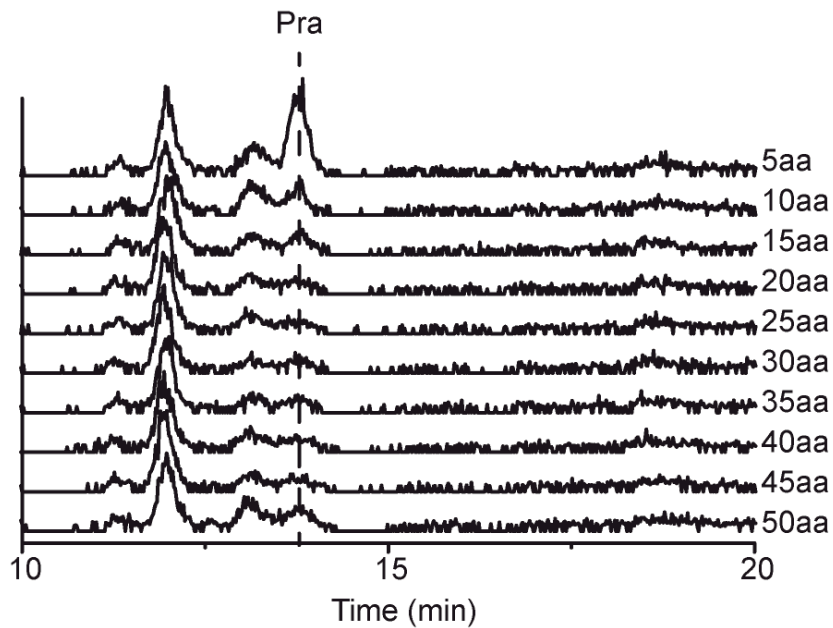
Figure 3.19. Extracted ion chromatogram for Pra (*m/z* = 114.0550) from *E. coli* BL21 cultures supernatants, expressing truncations of BesB Δ5aa, Δ10aa, Δ15aa, Δ20aa, Δ25aa, Δ30aa, Δ35aa, Δ40aa, Δ45aa, and Δ50aa. Peak where Pra elutes is indicated by dashed line. Samples shown were collected 24 h after induction.
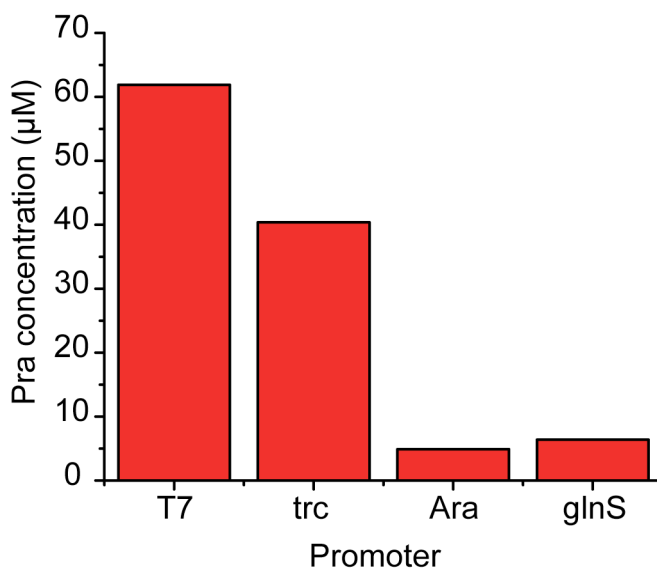
Figure 3.20. Pra concentration from *E. coli* BL21 pSV272.1-BesB cultures supernatants, expressing BesB under control of various promoters. Samples were collected 48 h after induction.

104
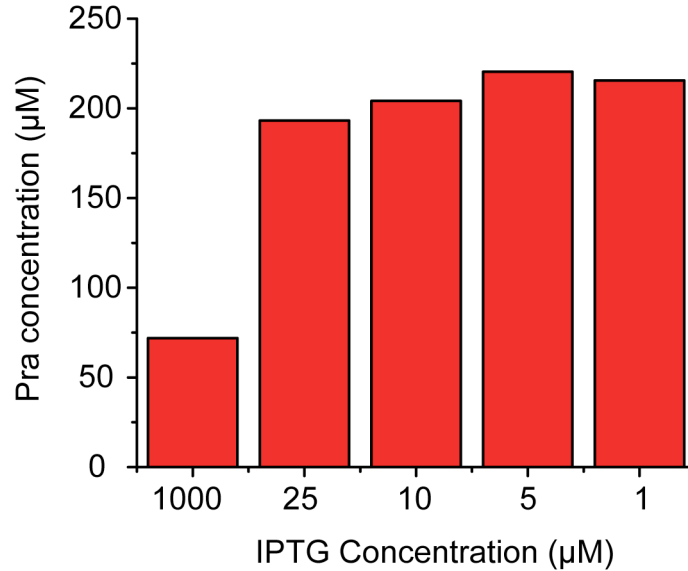
Figure 3.21. Pra concentration from *E. coli* TUNERS pSV272.1-BesB cultures supernatants, expressing BesB under control varying concentrations of IPTG. Samples shown were collected 96h after induction.

Table 3.2. List of chaperones plasmids tested for improving BesB soluble expression and the genes they encode.

Plasmid	Chaperone
pG-KJE8	*groEL, groES, dnaK, dnaJ, grpE*
pKJE7	*dnaK, dnaJ, grpE*
pGro7	*groEL, groES*
pG-Tf2	*groEL, groES, tig*
pTf16	*tig*
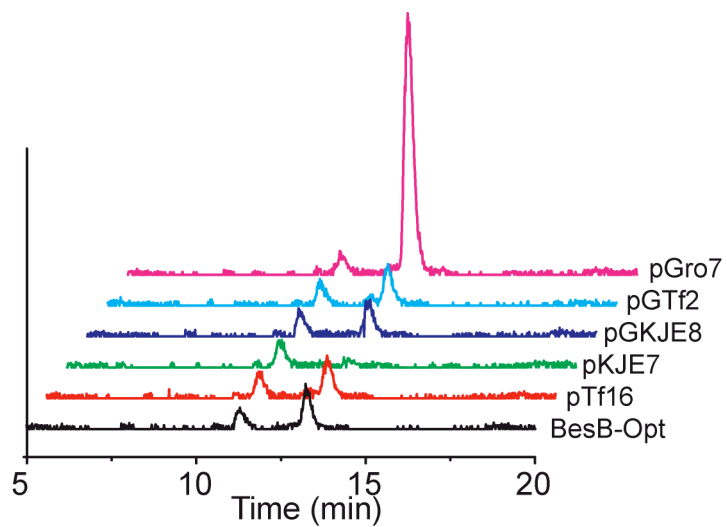
Figure 3.22. Optimization of BesB functional expression with co-expression of various *E. coli* protein chaperones. Production of propargylglycine (*m/z* = 114.550) was monitored by LC/QTOF in *E. coli* BL21 Star (DE3) pSV272.1-MBP-BesB cultures fed with substrate (1 mM 4-Cl-allylglycine). Cells were harvested after 24 h of growth in TB. Chromatograms shown are representative of 3 independent biological replicates.

As a final attempt at improving soluble expression of BesB, we experimented with co-expression of various chaperones. During protein translation, there are various pathways that can lead insoluble expression (protein misfolding) or protein degradation. Chaperones (GroEL, GroES, DnaK, DnaJ, GrpE, Tig) can assist with proper folding [26]. We co-expressed various combinations of chaperones (*Table 3.2*) with our BesB expression vector and measured conversion of 4-Cl-allylglycine to Pra. When pGro7 was used as a chaperone plasmid, containing *groEL* and *groES*, we observed approximately 7 fold increase in 4-Cl-allylglycine conversion (*Figure 3.22*). Using other chaperone plasmids showed only marginal improvements over no-chaperone condition. From these expriments, we decided to carry forward co-expressing pGro7 in our future *in vivo* work when high BesB expression is also required.

Optimization of BesD, BesC, and BesB pathway in *E. coli* and generation of pPra. Having optimized *in vivo* production of 4-Cl-lysine, 4-Cl-allylglycine, and Pra independently, we then worked on assembling and optimizing a *besDCB* pathway. Our goal was to show that we could successfully produce Pra, a terminal-alkyne amino acid, in *E. coli* using only glucose as a carbon source. In practice, we generated a single plasmid (pPra) that contained *besD* (from *P. fluorescens*), *besC* (from *P. fluorescents*), and *MBP-besB* (codon optimized, from *S.* sp. NRRL S-1448) all under T7 promoters (*Figure 3.23*). Each promoter generates a transcript starting at the individual gene and terminating after *besB*. This design maximizes transcripts containing *besB*, which we found had helped improve conversion of lysine to Pra (*Figure 3.24)*. While putting *besB* at the end of transcripts helped with pathway flux, we saw no difference in having *besD* or *besC* first (*Figure 3.25*). Under our optimized set of conditions, *E. coli* BL21 (DE3) Star expressing pPra (pPra = pPra-2) and pGro7 grown in LB at pH 6.5 was able to produce between 50-100 μM Pra after 48h of growth (*Figure 3.26*). At these titers, we believed that Pra, produced from lysine, could be coupled with translational machinery for incorporating Pra into proteins.
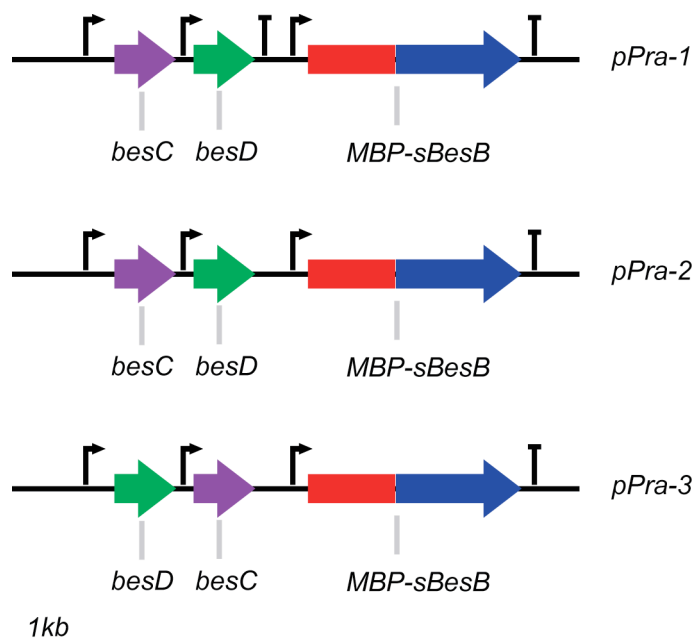
Figure 3.23. Pathway design for production of Pra from lysine in *E. coli*. pPra series of vectors are pET-DUET-based and carry T7 promoters and terminators, as indicated. Genes for *besC* and *besD are P. fluorescens* homologues, while BesB is the MBP-fusion of codon optimized, *S.* sp. NRRL S-1448 homologue (MBP-sBesB).
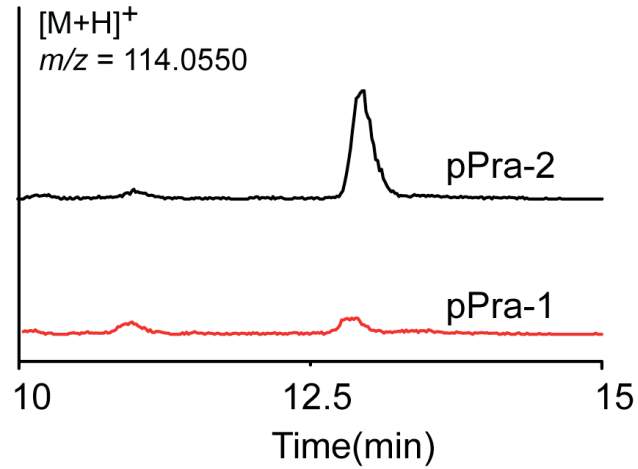
Figure 3.24. Representative extracted ion chromatogram for Pra (*m/z* = 114.0550) from *E. coli* BL21 pPra-1 and pPra-2. Smples shown were collected 48h after induction.
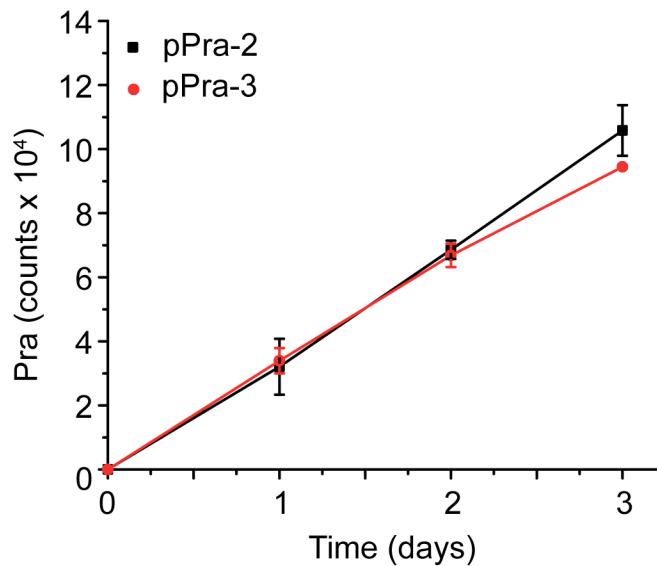
Figure 3.25. Time course production for Pra (integrated counts of Pra, *m/z* = 114.0550) from *E. coli* BL21 pPra-2 and pPra-3 over the course of three days. Error bars shown are standard deviation (n=3).
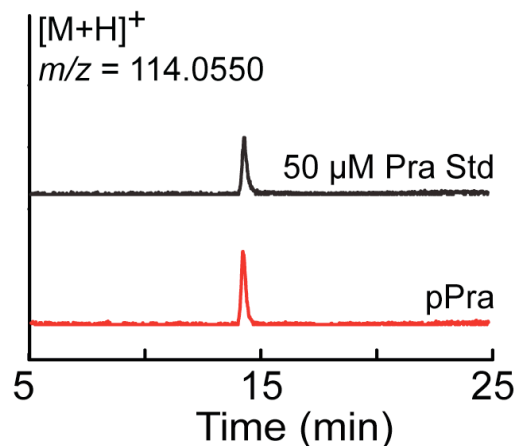
Figure 3.26. Representative extracted ion chromatogram for Pra (*m/z* = 114.0550) from *E. coli* BL21 pPra showing > 50 μM Pra produced after 48h.

Met residue replacement by Pra in *E. coli*. Having optimized a pathway for the production of Pra in *E. coli*, we next moved to develop a strategy for integrating Pra into proteins. We first looked to investigate if we could replace select amino acid residues with Pra using an engineered aminoacyl-tRNA synthetase (ARS) that has altered amino acid specificity. In particular, we aimed to use a previously-engineered ARS, PraRS, that had been shown to selectively acylate Pra over Met onto tRNA^{Met} [27]. *In vivo*, formation of Pra-tRNA^{Met} was sufficient to translate proteins where some of the Met residues are replaced by Pra. Since all proteins contain at least one Met, from translation initiation, it is possible to incorporate Pra into any newly translated protein. To visualize Pra incorporation into proteins, we used CuAAC chemistry to attach a fluorescent dye (TAMRA-azide) to alkyne-containing proteins. SDS-PAGE gels can then be imaged using a fluorescent filter to visualize derivitized proteins. Additionally, we used shotgun proteomics to directly identify peptides where Pra replaced Met.

To make our pathway design compatible with Pra incorporation, we had to make a few changes. First we integrated PraRS under an arabinose-inducible promoter into pGro7, creating praGro, which allowed us to simultaneously induce production of GroEL, GroES, and PraRS (*Figure 3.27*). However, when we fed Pra to cultures of BL21 praGro grown in LB we saw low levels of incorporation even at high amounts of Pra (>1 mM) feeding. One possible explanation for low levels of observed incorporation is completion from Met and wildtype MetRS. We saw an improvement in labeling when BL21 praGro was grown in chemical defined media (M9 + Pra) without any methionine. Unfortunately, the amount of Pra required to see significant labeling was still too high compared to the amount that we could produce *in vivo* using pPra (~50 μM). BL21 has the biosynthetic machinery to make its own Met.

To test if endogenous Met biosynthesis was competing with Pra incorporation, we switched to using *E. coli* BW1125 Δ*metB* (JW3910 from KEIO collection) which is unable to biosynthesize Met. Using this cell line, we could see incorporation of Pra, with Pra feeding as low as 50 μM. Next we lysogenized *E. coli* BW1125 Δ*metB* with λDE3 to make *E. coli* BW1125 (DE3) Δ*metB*. We found that BW1125 (DE3) Δ*metB* pPra makes approximately 10 μM Pra after 48 h. We decided BW1125 was not a suitable strain for Pra production from pPra, and instead switched to using *E. coli* B834, a BL21 parent strain that is naturally deficient in Met biosynthesis. Using *E. coli* B834 pPra, we saw comparable production to BL21 pPRa (~50 μM after 48 h, *Figure 3.28*).

109

Finalizing media condition optimization, we first grew cells and expressed proteins in TB, then switched cells to M9 (M9aa, containing 100 mg/L of each amino acid except Met, and 50 mg/L of Met). These conditions strike a balance between the higher protein expression levels for pPra and praGro grown in TB, with lower levels of Met formulated in our M9aa.

Putting together the pathway for Pra biosynthesis (pPra), Pra incorporation (praGro), a suitable host (*E. coli* B834), and optimized media conditions (M9aa), we were able to label proteins with *de novo* synthesized Pra (*Figure 3.29-30*). Our results show that only when both pPra and praGro are expressed, we see incorporation of Pra into proteins. Our pathway shows less incorporation, by fluorescent-PAGE gel, compared to conditions where Pra (200 μM) is fed exogenously to the cells. While we had previously seen pSV272.1-MBP-sBesB titers of Pra as high as 50-100 μM in LB, in M9aa we only observed ~25μM Pra produced after 48 h (*Figure 3.31*). Nonetheless, direct evidence of Pra incorporation was visible by proteomics. Bottom-up proteomics performed on our B834 pPra praGro samples found convincing examples of peptides where Met residues were replaced by Pra. Three examples, corresponding to peptides from overexpressed proteins GroEL, BesC, and BesD, are shown (*Figure 3.32*).

Efforts towards site-specific incorporation of Pra into proteins. A second strategy for Pra incorporation involves the site-specific, as opposed to residue-specific, incorporation of non-standard amino acid (nsAA) using an engineered ARS and tRNA pair [21]. Various groups have successfully incorporated nsAA using pyrrolysyl- or tyrosyl-derived ARS/tRNA pairs. Since Pra is much smaller than pyrrolysine or tyrosine, we opted to evolve a MetRS/tRNA pair for Pra site-selective incorporation. Two key pieces of previous literature form the basis for this work: 1) As previously described, Tuong *et al.* [27] had shown that MetRS could be evolved to acylate Pra onto tRNA$^{\text{Met}}$ and 2) Schmitt *et. al.* [28] and Meinnel *et. al.* [29] had shown that changes to MetRS tRNA-binding residues could change the specificity from tRNA$^{\text{Met}}$ CAU to amber suppressor tRNA$^{\text{Met}}$CUA. If these sets of mutations could be combined, we would have the basis for a Pra-specific, amber suppressor RS/tRNA pair.

Schmitt *et. al.* had reported various mutations which would alter MetRS anti-codon specificity for tRNA$^{\text{Met}}$. In particular, they found that MetRS mutations P460R, W461A, D456K, V463A, A464M, N453I (MetRS-CUA) altered relative $k_{\text{cat}}/K_{\text{M}}$ from >6×10^{-6} to 1.5 (tRNA$^{\text{Met}}$CUA to tRNA$^{\text{Met}}$CAU). Since these values were reported *in vitro*, we decided to test out the mutations *in vivo* to measure how effective the MetRS-CUA was at amber suppression. We constructed the *MetRS-CUA* gene and cloned two copies of it (one under constitutive promoter and another under arabinose inducible promoter), alongside tRNA$^{\text{Met}}$CUA, into pEVOL to generate pEVOL-MetCUA. To report on amber suppression, we constructed an eGFP-based reporter, pREP-GFP-CUA, with a CUA codon near the N_T (eGFP-N_T-CUA). Successful translation of eGFP would therefore lead to an increase in fluorescent signal, and allows us to measure eGFP fluorescence (*ex: A_{480}, em: A_{508}*) as a proxy for MetRS-CUA suppressor activity *(Figure 3.33)*. Compared to the fluorescence detected when expressing wildtype eGFP, we found that cells expressing pEVOL-MetCUA showed 30% of the fluorescence (*Figure 3.34*). Evolved ARS from other groups have shown as high as 80% relative fluorescence of eGFP-N_T-CUA compared to wildtype eGFP when performing similar experiments [30]. However, when we introduced the PraRS mutations, which confer Pra selectivity, into our suppressor MetRS-CUA to generate PraRS-CUA, we observed no activity towards Pra, suggesting these two sets of mutations are incompatible. Future work in this area should focus on evolving an orthogonal

Figure 3.27. Design of praGro. PraRS was cloned into praGro7 between groEL and Ara terminator. Design consists of two arabinose inducible promoters driving expression of GroES, GroEL, and PraRS with a single terminator.
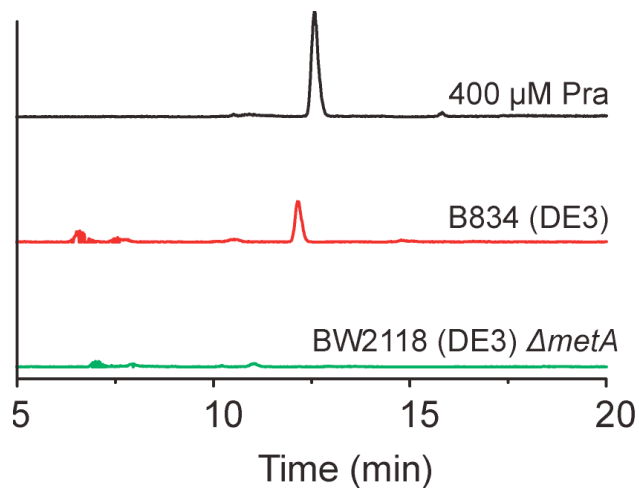
Figure 3.28. Extracted ion chromatogram for Pra (m/z = 114.0550) from *E. coli* B834 pPra or E. coli BW2118 (DE3) Δ*metA* pPra. After 48 h, B834 is able to produce ≈ 50 µM Pra, while there was no detectable Pra in BW2118 (DE3) Δ*metA* cultures. Peaks slightly shifted due to poor column equilibration.

111
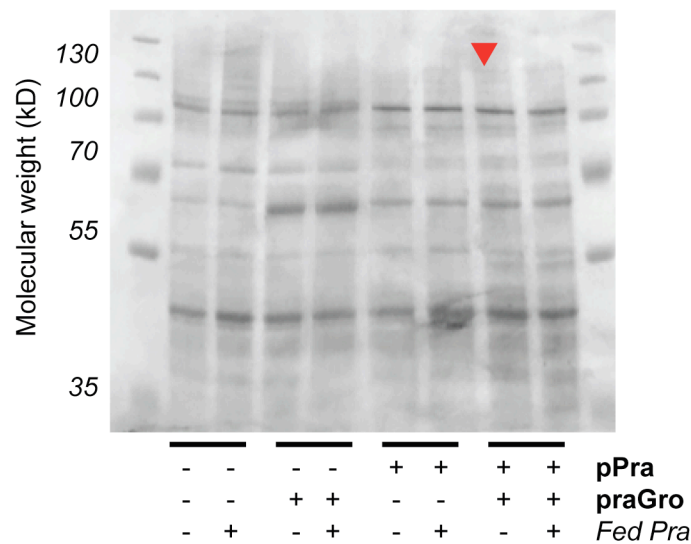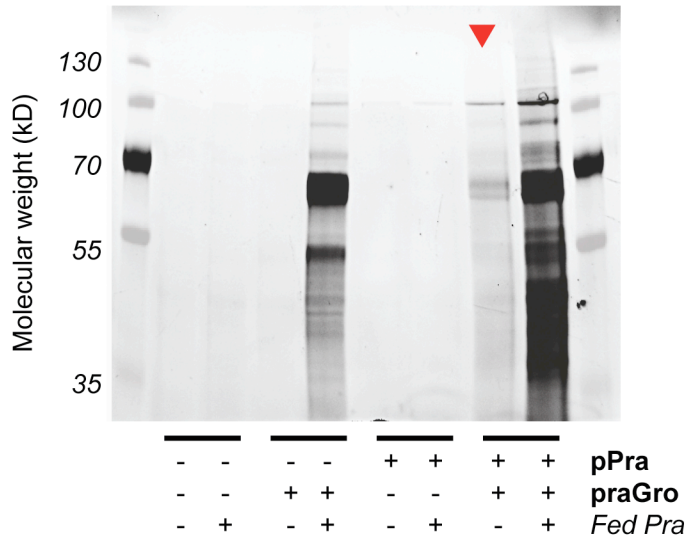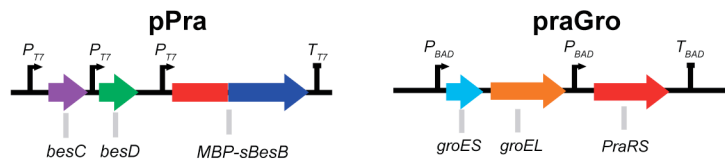
Figure 3.29. (Top) Co-expression of Pra biosynthetic genes (pPra) along with the PraRS (praGro) in a heterologous host leads to proteome incorporation of Pra. *E. coli* B834(DE3) pPra praGro was cultured for 2 d along with the respective empty vector controls and Pra feeding controls. Cell lysate was then reacted with Tamra-azide dye by CuAAC before analysis by SDS-PAGE in fluorescence mode (λ_{Ex} = 546 nm, λ_{Em} = 565 nm). Red arrow indicates lane *E. coli* B834(DE3) pPra praGro, showing incorporation of endogenously-synthesized Pra. Ladder consists of 70 kD marker that can be visualized by fluorescence overlaid with the Coomassie-stained ladder from the same gel for reference. (Bottom) SDS-PAGE gel from showing Pra incorporation into the *E. coli* proteome by CuAAC derivatization and fluorescence imaging stained with Coomasie dye to visualize total protein loading in each lane. Gels shown are representative of 6 independent biological replicates.
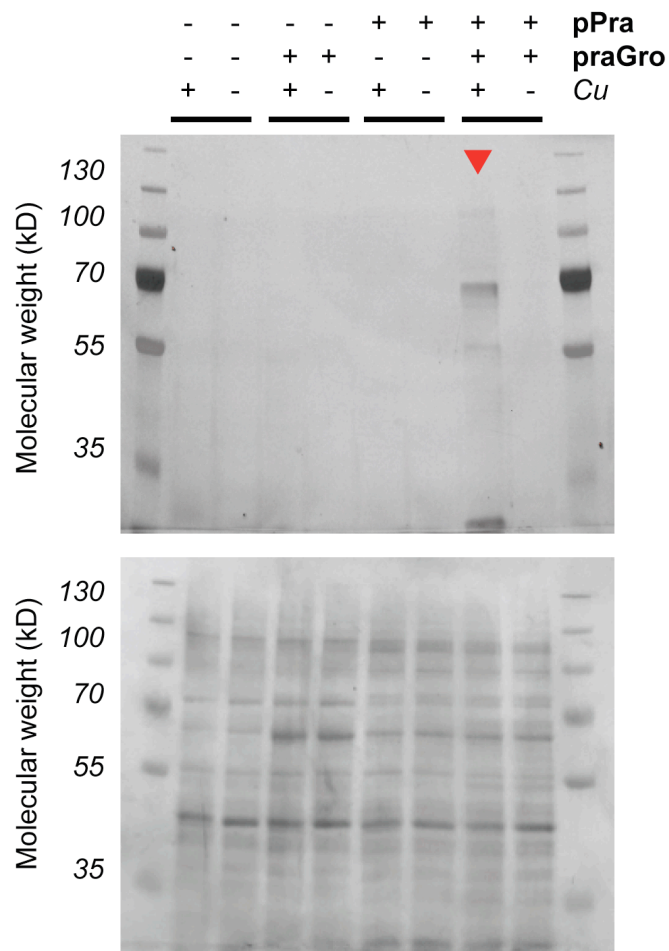
Figure 3.30. Additional controls for the SDS-PAGE gel from Figure 6D showing that successful derivatization with Tamra-azide requires endogenous Pra production, PraRS, and Cu. In this experiment, *E. coli* B834(DE3) pPra praGro was cultured for 2 d along with the respective empty vector controls and no Cu controls. The cell lysates were then derivatized with Tamra-azide dye by CuAAC before analysis by SDS-PAGE under fluorescence mode. Reactions without Cu act a negative control since Cu is required for CuAAC. Ladder consists of 70 kD marker that can be visualized by fluorescence overlaid with the Coomassie-stained ladder from the same gel for reference. Gel shown is representative of 6 independent biological replicates.
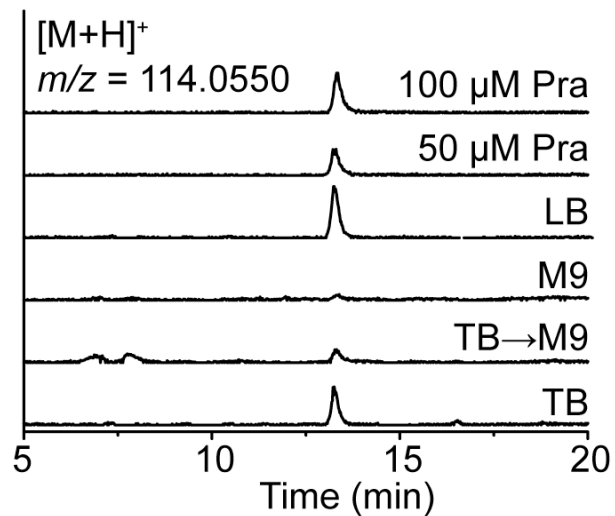
Figure 3.31. BesB functional expression was tested in different media conditions as it is the limiting factor in the biosynthesis of propargylglycine. Expression was tested by analyzing production of propargylglycine (**4**, *m/z* = 114.550) by LC/QTOF in *E. coli* BL21 Star (DE3) pSV272.1-MBP-BesB cultures fed with substrate (1 mM Cl-Alg, **3**). Cells were harvested after 24 h of growth. Chromatograms shown are representative of 3 independent biological replicates.

114
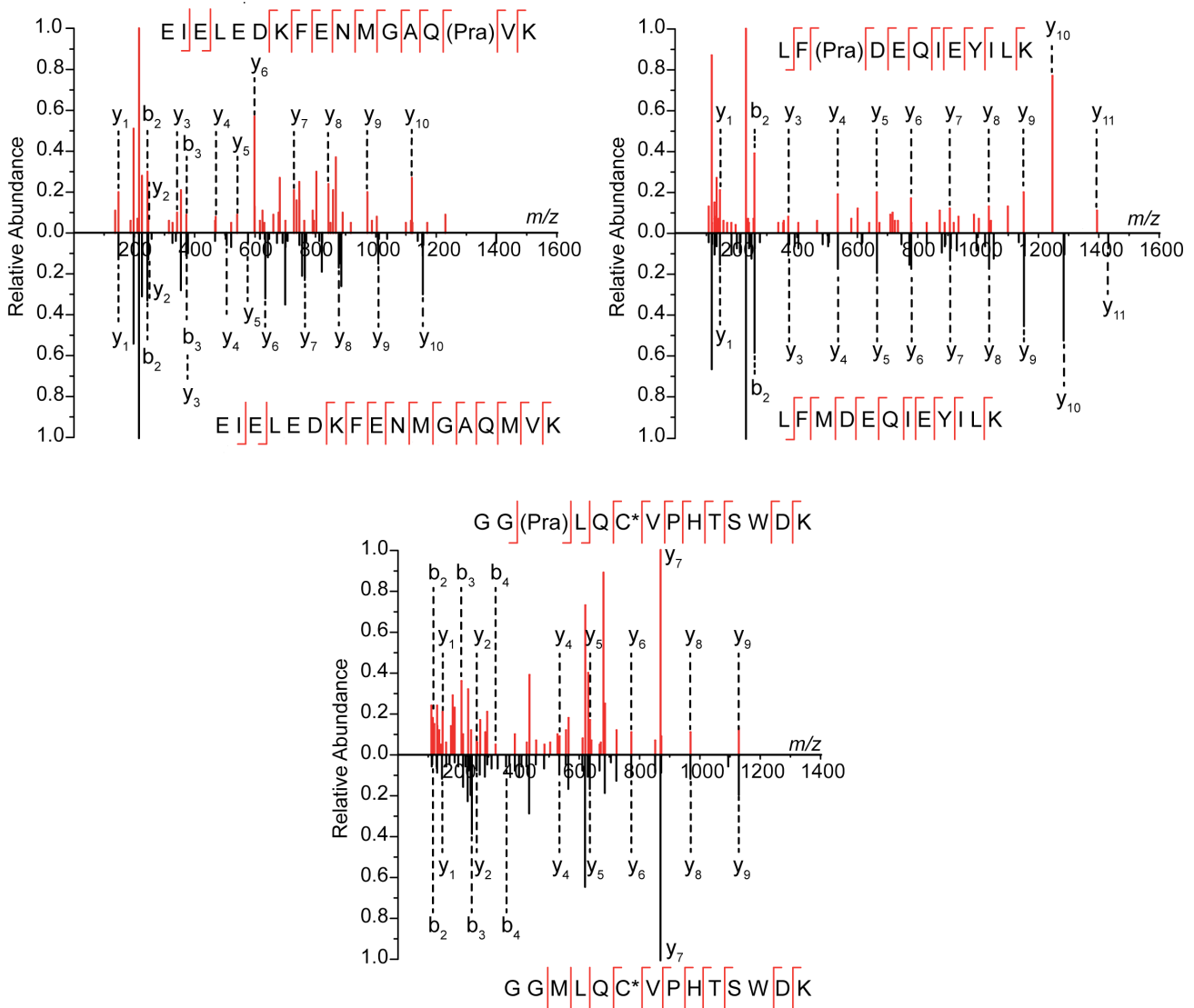
Figure 3.32. Total protein was extracted from cell pellets of *E. coli* B834(DE3) pPra PraGro. Extracted protein was subject to tryptic digest and bottom-up proteomic analysis using an using a Thermo LC-Q Exactive and spectra were analyzed using X! Tandem [31]. Representative high-resolution tandem MS/MS spectra of peptides (GroEL, showing Pra replacing Met from *E. coli* B834(DE3) pPra PraGro are shown (EIELEDKFENMGAQ(Pra)VK, GroL, WP_000729117; LF(Pra)DEQIEYILK, BesC, WP_080628534; GG(Pra)LQC*VPHTSWDK, BesD, WP_016975823). The spectrum above the axis gives Peptide-Spectrum Matches (PSMs) to sequences with Pra substituted for Met while the spectrum below the axis shows the corresponding Met-containing native peptide (C* = carbamidomethyl derivative from iodoacetamide capping). Spectra shown are representative of 2 independent biological replicates.

115
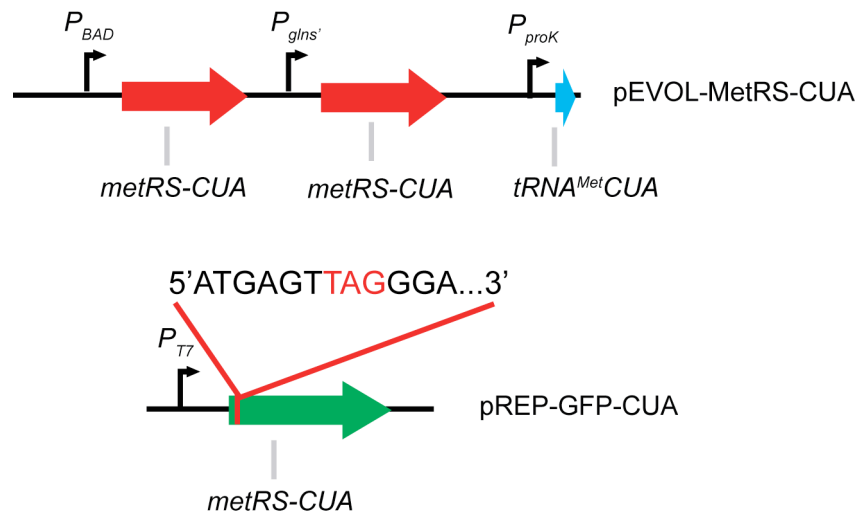
Figure 3.33. Design of pEVOL-MetRS-CUA and pREP-GFP-CUA. pEVOL-MetRS-CUA contains two copies of MetRS-CUA – one under a constituitive promoter (*glns'*) and another under an arabinose inducible promoter (P$_{BAD}$), and one copy of tRNA^{Met}CUA under constitutive promoter (*proK*). pREP-GFP-CUA contains one copy of eGFP with a TAG codon inserted at the third amino acid position.
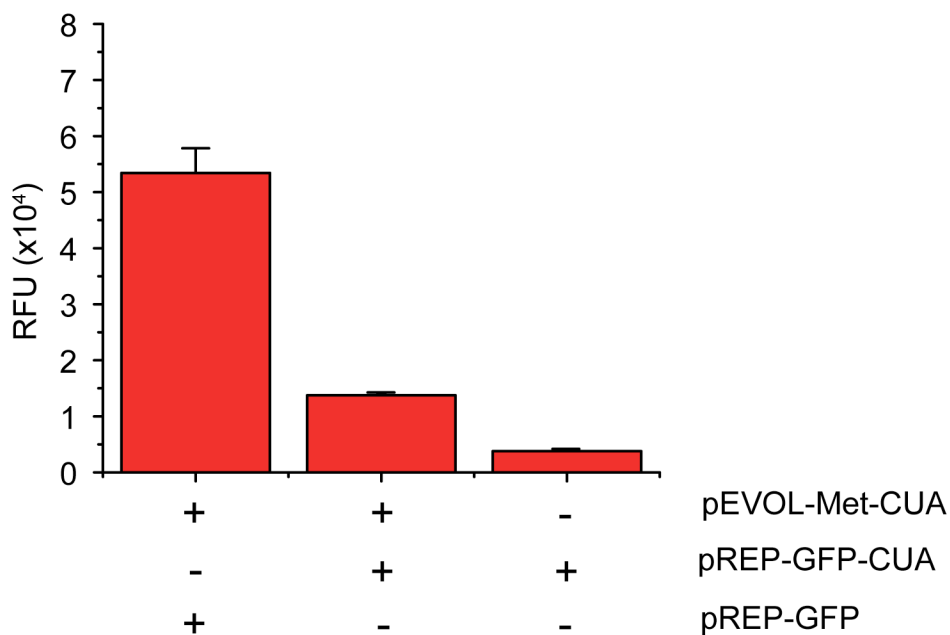
Figure 3.34. Relative fluorescence of E. coli BL21 co-transformed with pEVOL-Met-CUA pREP-GFP (positive control), pEVOL-Met-CUA pREP-GFP-CUA, or only pREP-GFP-CUA (negative control). Increase in fluorescence for pEVOL-Met-CUA pREP-GFP-CUA compared to only pREP-GFP-CUA suggest that cells are able to translate TAG codon in GFP-CUA due to presence of MetRS-CUA and tRNA^{Met}CUA. Compared to the positive control, this corresponds to approximately 30% of the RFU after background correction. Error bars shown are standard deviation (n = 6).

Schmitt *et. al.* had reported various mutations which would alter MetRS anti-codon specificity for tRNA^{Met}. In particular, they found that MetRS mutations P460R, W461A, D456K, V463A, A464M, N453I (MetRS-CUA) altered relative k_{cat}/K_M from >6×10^{-6} to 1.5 (tRNA^{Met}CUA to tRNA^{Met}CAU). Since these values were reported *in vitro*, we decided to test out the mutations *in vivo* to measure how effective the MetRS-CUA was at amber suppression. We constructed the *MetRS-CUA* gene and cloned two copies of it (one under constitutive promoter and another under arabinose inducible promoter), alongside tRNA^{Met}CUA, into pEVOL to generate pEVOL-MetCUA. To report on amber suppression, we constructed an eGFP-based reporter, pREP-GFP-CUA, with a CUA codon near the N$_T$ (eGFP-N$_T$-CUA). Successful translation of eGFP would therefore lead to an increase in fluorescent signal, and allows us to measure eGFP fluorescence (*ex: A$_{480}$, em: A$_{508}$*) as a proxy for MetRS-CUA suppressor activity *(Figure 3.33)*. Compared to the fluorescence detected when expressing wildtype eGFP, we found that cells expressing pEVOL-MetCUA showed 30% of the fluorescence (*Figure 3.34*). Evolved ARS from other groups have shown as high as 80% relative fluorescence of eGFP-N$_T$-CUA compared to wildtype eGFP when performing similar experiments [30]. However, when we introduced the PraRS mutations, which confer Pra selectivity, into our suppressor MetRS-CUA to generate PraRS-CUA, we observed no activity towards Pra, suggesting these two sets of mutations are incompatible. Future work in this area should focus on evolving an orthogonal

PraRS/tRNA^{Met}CUA pair using positive-negative selection workflows, described elsewhere [32,33].

Efforts towards Leu residue replacement by allylglycine in *E. coli.* We previously reported being able to produce 50-100 µM of allylglycine in *E. coli in vivo* by overexpressing BesC (from *P. fluorescens*) in the absence of BesD (*Figure 3.11*). Much like our goal with Pra-Met replacement, we were interested in developing a system for residue replacement of Leu with allylglycine *in vivo*. Previous literature had reported successful incorporation of allylglycine into proteins through tRNA^{Leu} using an editing-domain deficient LeuRS (LeuRS T252Y) [34,35]. While we were never able to detect clear allylglycine replacement, we did spend some time developing methods for quickly detecting allylglycine incorporation using click chemistry.

In the spectrum of reactivity with tetrazines, unactivated terminal alkenes are one of slowest reacting substrates in the reverse-electron demand Diels Alder reaction (tetrazine-ligation) [36,37]. Nonetheless, we explored using tetrazine ligation as a way to derivitized allylglycine on proteins since the attractiveness of having a reactive handle that can be modified through copper-free reaction conditions greatly expands application space. *In vitro* reactions using purified allylglycine and a Cy5-tetrazine (*Figure 3.35*). To summarize what we learned, in aqueous conditions and at room temperature we saw no reaction product after 24 h (data not shown). However, we observed formation of product after heating reaction to 90 °C for 1 h and providing a 10-fold molar excess of Cy5-tetrazine (100 µM allylglycine to 1 mM Cy5-tetrazine). It has yet to be determined if derivitization prior to SDS-PAGE gel loading, when one would normally denature proteins for 5 min at 90 °C, is an appropriate step for derivatizing proteins for visualization.

The poor reactivity at room temperature with tetrazines motivated us to look for alternative click reactions for derivatizing proteins. We found the 'tetrazole-photoclick' to be a potentially suitable alternative, as the observed (up to $k_{2 =}$ 0.52 M^{-1} s^{-1}) [38] for terminal-alkenes in this reaction is orders of magnitude higher than the reported tetrazine-ligation reactions (k_2 0.0002-0.0042 M^{-1} s^{-1}) [37,39]. We learned here that performing the reaction in water under 302 nM UV irradiation for 10 min was sufficient to see significant conversion of allylglycine to the tetrazole-ligated product (*Figure 3.36*).
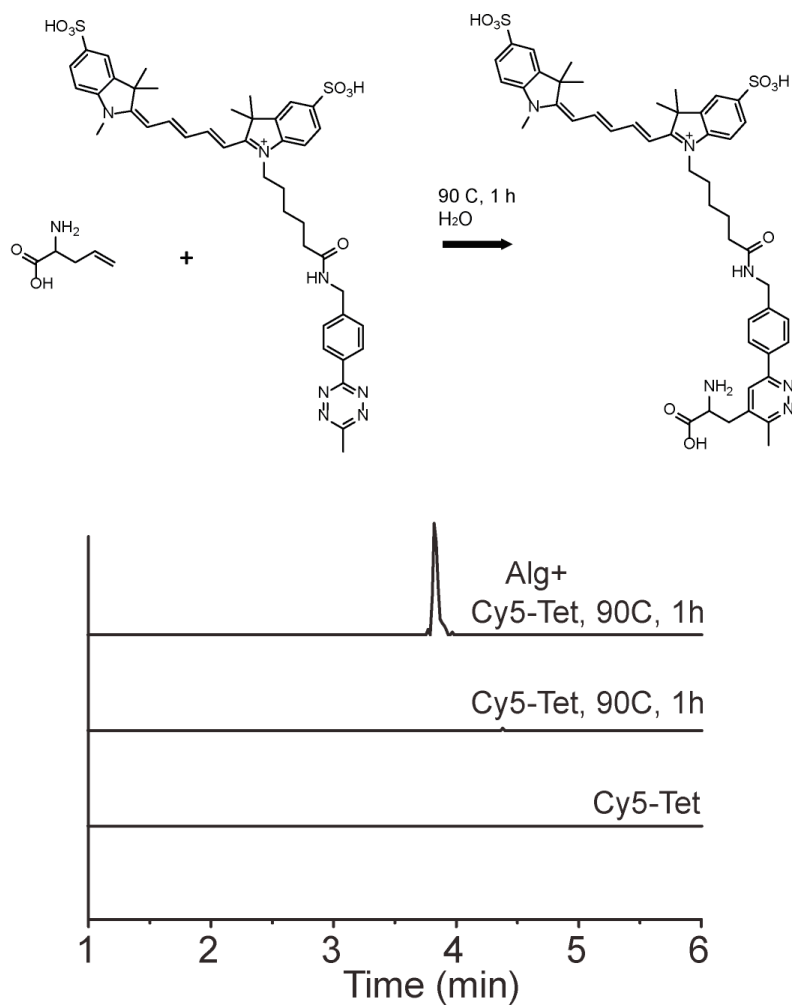
Figure 3.35. (Top) Reaction scheme for ligation of Cy5-tetrazine to allylglycine through tetrazine ligation. (Bottom) Representative extracted ion chromatogram for allylglycine-Cy5-Tetrazine (*m/z* = 911.3456) when 100 µM of L-allylglycine is reacted for 1h at 90C with 1 mM of Cy5-tetrazine. No allylglycine added, and pure Cy5-tetrazine shown as controls.

119
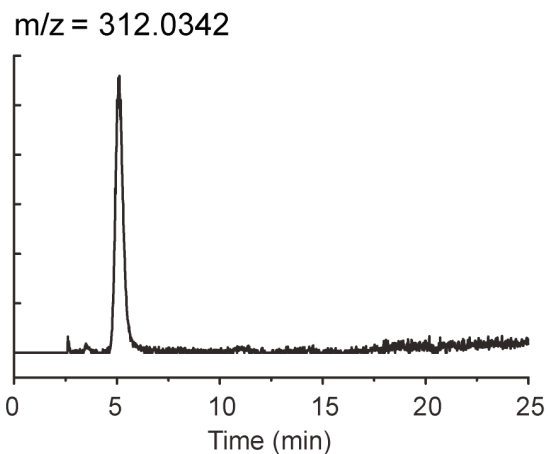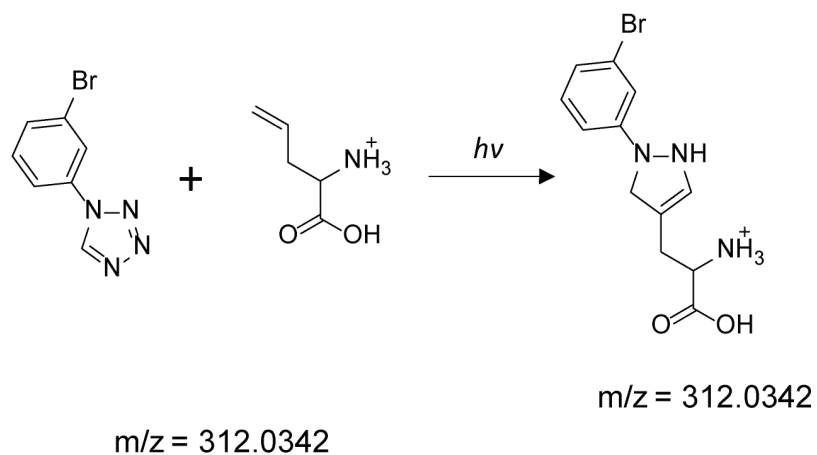
Figure 3.36. (Top) Reaction scheme for ligation of 5-Br-phenyl-tetrazole to allylglycine through tetrazole-photoclick reaction. (Bottom) Representative extracted ion chromatogram for Bromo-phenyl-tetrazole-allylglycine (*m/z* = 312.0342) when 100 µM of L-allylglycine is reacted for 10 min under UV light (λ = 302 nm) with 1 mM of 5-Br-phenyl-tetrazole (5mM).

3.4. Conclusion

The ability to expand the genetic code of *E. coli*, and other organisms, to use amino acids beyond the standard 20 has long been subject of interest. Indeed, a variety of amino-acyl tRNA synthetases have been evolved that are able to charge nsAA onto tRNAs. These amino acids can then be introduced into elongating peptides ribosomal, giving rise to proteins that contain non-canonical residues. The ability to chemically modify proteins by introducing nsAA has given rise to a variety of applications, including: immobilization of proteins onto solid supports, attachment of fluorophores onto proteins for imaging or detection, enrichment of protein through introduction of a 'pull-down' tag, solubilization of proteins through ligation of small molecules that can enhance solubility, and even production of antibody-drug conjugates as next generation therapeutics [6,14,21,27,40–46].Though versatile, previous methods for expansion of protein function through introduction of a nsAA have relied on exogenous feeding of nsAAs. In this chapter, we aimed to leverage the βes biosynthetic gene cluster for the *de novo* production of various nsAA that could ultimately be incorporated into proteins or peptides, *in vivo*.

Translating the βes for heterologous expression in *E. coli* required optimization of media conditions and enzymes. Due to possible intracellular instability of the first metabolite in the βes pathway, 4-Cl-lysine, we found that culturing cells in slightly acidic pH (pH $\leq$ 6.5) greatly increased titers of 4-Cl-lysine in culture broth. When aiming to make 4-Cl-allylglycine, we found that using homologues of BesC and BesD, from *P. fluorescens*, greatly improved titers over using the *S. cattleya* homologues. Achieving high conversion of 4-Cl-allylglycine to Pra *in vivo* proved to be challenging, possibly due to poor solubility of BesB. Through a variety of expression optimization methods for BesB including attachment of a solubility tag, expression of chaperones, codon optimization, promoter screening, strain optimization, we were able to build a three-protein pathway (pPra: containing BesD, BesC, BesB) for making Pra *in vivo*.

As a proof-of-concept, we showed that Pra made from lysine using βes pathway genes can be incorporated into proteins. Specifically, we were able to perform residue replacement of Met with Pra by expressing pPra alongside PraRS in *E. coli* B834. We believe there is potential for improving Pra incorporation by improving pathway flux. We estimate that over half of the 4-Cl-allylglycine produced by BesD and BesC remains unconsumed. One potential avenue of improvement is through increasing soluble expression of BesB, which we had noted is largely insoluble. Beyond Pra and PraRS, there is potential for residue-incorporation of other βes pathway amino acids into proteins using engineered an ARS, including allylglycine through LeuRS T252Y and 4-Cl-lysine through an engineered pyrrolysyl-RS. In addition, there is potential to incorporate Pra site-specifically into proteins using a *de novo* evolved PraRS/tRNA^{Met}CUA. Ultimately, a variety of improvements are still required to fully exploit the potential of βes pathway to be used in expanding the genetic code of engineered organisms.

3.5. References

1. Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature.* doi:10.1038/s41586-019-1020-y

2. Low, S. C. & Berry, M. J. Knowing when not to stop: selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci.* **21,** 203–208 (2002).

3. Stadtman, T. C. Selonocyseine. *Annu. Rev. Biochem.* 83–100 (1996).

4. Gaston, M. a, Zhang, L., Green-Church, K. B. & Krzycki, J. a. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **471,** 647–50 (2011).

5. Krzycki, J. A. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* **8,** 484–491 (2004).

6. Voloshchuk, N. & Montclare, J. K. Incorporation of unnatural amino acids for synthetic biology. *Mol. Biosyst.* **6,** 65–80 (2010).

7. Dieterich, D. C., Link, J. A., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9482–7 (2006).

8. Hinz, F. I., Dieterich, D. C., Tirrell, D. A. & Schuman, E. M. Noncanonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. *ACS Chem. Neurosci.* **3,** 40–49 (2012).

9. Niehues, S. *et al.* Impaired protein translation in Drosophila models for Charcot-Marie-Tooth neuropathy caused by mutant tRNA synthetases. *Nat. Commun.* **6,** 7520 (2015).

10. Erdmann, I. *et al.* Cell-selective labelling of proteomes in Drosophila melanogaster. *Nat. Commun.* **6,** 7521 (2015).

11. Yuet, K. P. & Tirrell, D. A. Chemical tools for temporally and spatially resolved mass spectrometry-based proteomics. *Ann. Biomed. Eng.* **42,** 299–311 (2014).

12. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20,** 1044–1048 (2002).

13. Chatterjee, A., Xiao, H. & Schultz, P. G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 14841–6 (2012).

14. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chemical Reviews* **114,** 4764–4806 (2014).

15. De Graaf, A. J., Kooijman, M., Hennink, W. E. & Mastrobattista, E. Nonnatural amino acids for site-specific protein conjugation. *Bioconjug. Chem.* **20,** 1281–1295 (2009).

16. Uttamapinant, C. *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc. Nat. Acad. Sci. U. S. A.* **107,** 10914–10919 (2010).

17. Vanbrunt, M. P. *et al.* Genetically encoded azide containing amino acid in mammalian cells enables site-specific antibody–drug conjugates using click cycloaddition chemistry. *Bioconjug. Chem.* **26,** 2249–2260 (2015).

18. Baruah, H., Puthenveetil, S., Choi, Y.-A., Shah, S. & Ting, A. Y. An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew. Chem. Int. Ed. Engl.* **47,** 7018–21 (2008).

19. Wang, B., Brown, K. C., Lodder, M., Craik, C. S. & Hecht, S. M. Chemically mediated site-specific proteolysis. Alteration of protein-protein interaction. *Biochemistry* **41,** 2805–2813 (2002).

20. Schmied, W. H., Elsässer, S. J., Uttamapinant, C. & Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression

and engineered eRF1. *J. Am. Chem. Soc.* **136,** 15577–15583 (2014).

21. Xie, J. & Schultz, P. G. Adding amino acids to the genetic repertoire. *Curr. Opin. Chem. Biol.* **9,** 548–554 (2005).

22. Wang, L. & Schultz, P. G. Expanding the genetic code. *Angew. Chemie Int. Ed.* **44,** 34–66 (2005).

23. Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb. Cell Fact.* **8,** 1–8 (2005).

24. Tegel, H., Ottosson, J. & Hober, S. Enhancing the protein production levels in Escherichia coli with a strong promoter. *FEBS J.* **278,** 729–739 (2011).

25. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22,** 346–353 (2004).

26. Mogk, A., Tomoyasu, T., Deuerling, E., Bukau, B. & de Marco, A. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC Biotechnol.* **7,** 32 (2007).

27. Truong, F., Yoo, T. H., Lampo, T. J. & Tirrell, D. A. Two-strain, cell-selective protein labeling in mixed bacterial cultures. *J. Am. Chem. Soc.* **134,** 8551–8556 (2012).

28. Schmitt, E., Meinnel, T., Panvert, M., Mechulam, Y. & Blanquet, S. Two acidic residues of E. coli MetARS as negative discriminants towards the binding of non-cognate tRNA anticodons. *J. Mol. Biol.* **233,** 615–628 (1993).

29. Meinnel, T. *et al.* Selection of suppressor methionyl-tRNA synthetases: mapping the tRNA anticodon binding site. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 291–5 (1991).

30. Young, T. S., Ahmad, I., Yin, J. a & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361–74 (2010).

31. Craig, R. & Beavis, R. C. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467 (2004).

32. Neumann, H., Slusarczyk, A. L. & Chin, J. W. De Novo generation of mutually orthogonal aminoacyl-TRNA synthetase/ tRNA pairs. *J. Am. Chem. Soc.* **132,** 2142–2144 (2010).

33. Jaric, J. & Budisa, N. Design of orthogonal pairs for protein translation: Selection systems for genetically encoding noncanonical amino acids in E. coli. *Hydrocarb. lipid Microbiol. Protoc.* (2017). doi:10.1007/8623

34. Tang, Y. & Tirrell, D. A. Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **41,** 10635–10645 (2002).

35. Lincecum, T. L. *et al.* Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Mol. Cell* **11,** 951–963 (2003).

36. Pidgeon, S. E. & Pires, M. M. Metabolic remodeling of bacterial surfaces via tetrazine ligations. *Chem. Commun.* **51,** 10330–10333 (2015).

37. Chen, X. & Wu, Y. W. Selective chemical labeling of proteins. *Org. Biomol. Chem.* **14,** 5417–5439 (2016).

38. Wang, Y., Song, W., Hu, W. J. & Lin, Q. Fast alkene functionalization in vivo by photoclick chemistry: HOMO lifting of nitrile imine dipoles. *Angew. Chemie - Int. Ed.* **48,** 5330–5333 (2009).

39. Rieder, U. & Luedtke, N. W. Alkene – tetrazine ligation for imagine cellular DNA. *Angew. Chemie - Int. Ed.* **53,** 9168–9172 (2014).

40. Hartman, M. C. T., Josephson, K., Lin, C.-W. & Szostak, J. W. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One* **2,** e972 (2007).

41. Xiang, Z. *et al.* Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nat. Methods* **10,** 885–8 (2013).

42. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in escherichia coli. *Biochemistry* **52,** 1828–1837 (2013).

43. Agostini, F. *et al.* Biocatalysis with unnatural amino acids: enzymology meets xenobiology. *Angew. Chemie - Int. Ed.* **56,** 9680–9703 (2017).

44. Magliery, T. Unnatural protein engineering: producing proteins with unnatural amino acids. *Med. Chem. Rev. - Online* **2,** 303–323 (2005).

45. Mahdavi, A. *et al.* Engineered Aminoacyl-tRNA Synthetase for Cell-Selective Analysis of Mammalian Protein Synthesis. *J. Am. Chem. Soc.* **138,** 4278–4281 (2016).

46. VanBrunt, M. P. *et al.* Genetically encoded azide containing amino acid in mammalian cells enables site-specific antibody–drug conjugates using click cycloaddition chemistry. *Bioconjug. Chem.* 150911072412002 (2015). doi:10.1021/acs.bioconjchem.5b00359

Chapter 4: *Translational-machinery control of β-ethynylserine by a trans-editing tRNA deacylase*

4.1. Introduction

The fidelity of protein translation is maintained in large part by ensuring that tRNAs are correctly charged with their cognate amino acid [1–3]. Though many kinetic steps follow tRNA aminoacylation (*e.g.* binding EF-Tu, docking to the ribosome, and peptide bond formation), there are few known mechanisms in the downstream steps of translation that can distinguish if the correct amino acid is linked to its cognate tRNA. Charging the wrong amino acid onto a tRNA can therefore result in mistranslation of proteins, since incorrectly charged tRNA can still be translated to generate protein with the wrong amino acid sequence. This problem exist even among the canonical set of 20 amino acids, as aminoacyl-tRNA synthetases routinely make mistakes when charging amino acids onto tRNA [2–4]. Fidelity control for charging some amino acids onto tRNA is thought to occur through a double-sieve mechanism. The first opportunity, or first sieve, of fidelity control lies in an ARS's native ability to discriminate against other non-cognate amino acids when charging tRNA. Though structurally divergent amino acids can easily be excluded through binding kinetics alone, structurally-similar amino acids face high rates of misacylation onto tRNA. For example threonine can be activated and charged by the valyl-tRNA synthetase at a rate of 1 in 350, valine is activated by the leucyl-tRNA synthetase at a rate of about 1 in 180, and serine is activated by threonyl-tRNA synthetase at a rate of 1 in 1000 [1,2,4,5].

Though misacylation of amino acids can occur at a high rate by various combinations of tRNA synthetases and non-cognate amino acids, the observed rate of mistranslated amino acid is orders of magnitude lower at 1 in 10,000 to 100,000 mistakes per residue. Not surprisingly, Nature has evolved a convergent strategy for improving tRNA fidelity by introducing an additional catalytic step for deacylating incorrectly aminoacylated tRNA, also known as the second-sieve [6–10]. Not surprisingly, many of the ARSs display issues with fidelity in aminoacylation in the "first sieve" contain an additional catalytic proofreading domain that will hydrolyze the incorrectly aminoacylated tRNA. These *cis*-acting domains are found all throughout life and have evolved to ensure translational fidelity of the 20 canonical set of amino acids onto their cognate tRNAs [11]. In some organisms (mostly bacteria and archaea), standalone *trans*-acting domains also exist to facilitate the proofreading process, especially for those tRNAs with more difficult selectivity issues [12–16].

Organisms that have the ability to biosynthesize non-standard amino acids for secondary metabolism risk additional exposure to translation errors, beyond the canonical pathway for misacylation through standard proteinogenic amino acids. *S. cattleya*, producer of 4-fluorothreonine (F-Thr) and β-ethynylserine (βes), has evolved its own set of *trans*-acting proofreading enzymes to prevent misacylation of these amino acids onto tRNA^{Thr} (*Figure 4.1*) [17,18]. Work by our group previously identified and characterize FthB, a F-Thr-tRNA^{Thr} deacylase which was found to have a relative $k_{cat}/K_m > 650$ for F-Thr over Thr [19]. In this chapter, we present the discovery and characterization of an analogous enzyme from a different hydrolase family, BesG, which has evolved to carry out a similar function as FthB. By deacylating βes-tRNA^{Thr}, BesG helps prevent βes translation into proteins at residues where Thr should be present.
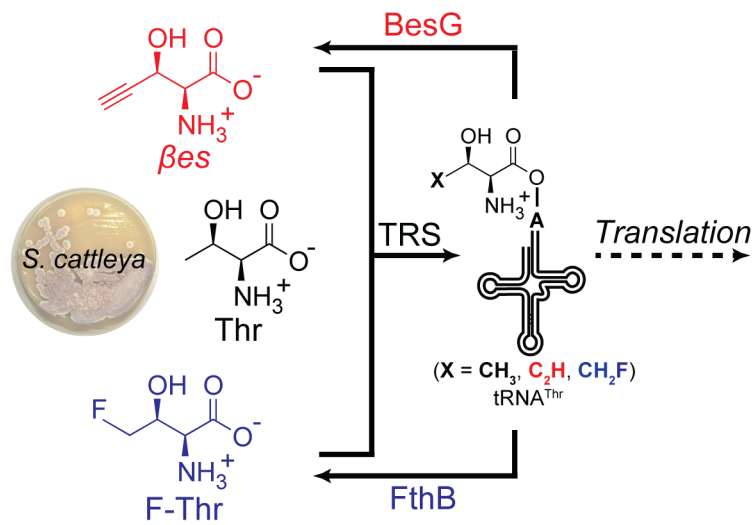
Figure 4.1. *S. cattleya* makes two threonine-like amino acids: F-Thr and βes. Both F-Thr and βes can be charged onto tRNA^{Thr} by TRS. FthB is a F-Thr-tRNA^{Thr} deacylase that selectively hydrolyzes F-Thr-tRNA^{Thr} over Thr-tRNA^{Thr}. We propose BesG is a βes-tRNA^{Thr} deacylase.

Methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), yeast extract, malt extract, and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), Nanosep 10k Omega filters, and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), adenosine monophosphate sodium salt (AMP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), pyruvate kinase, lactate dehydrogenase, isocitrate dehydrogenase, lysozyme, magnesium sulfate heptahydrate, poly(ethyleneimine) solution (PEI), β-mercaptoethanol (βME), sodium phosphate dibasic heptahydrate, D-mannitol, apramycin sulfate salt (Am), betaine, dimethylsulfoxide (DMSO), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), copper sulfate (CuSO$_4$), Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), maltose, L-propargylglycine (Pra), L-glutamate, and γ-glutamyl-transpeptidase from equine kidney (γ-GT), were purchased from Sigma-Aldrich (St. Louis, MO). Hygromycin B was purchased from Santa Cruz Biotechnology (Santa Cruz, California). 3-Azido-7-hydroxycoumarin was purchased from AK Scientific, Inc (Union City, CA). Formaldehyde (16% (*w/v*), methanol-free) and PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Tamra-Azide, 5-isomer was purchased from Lumiprobe (Hunt Valley, Maryland). Bacto^{TM} Agar was purchased from BD (Sparks, Maryland) Soy flour was purchased from Berkeley Bowl (Berkeley, CA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, T4 DNA ligase, Antarctic phosphatase, Phusion DNA polymerase, PURExpress (Δaa, tRNA) IVTT kit, T7 HiScribe Quick High Yield RNA Synthesis Kit, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 µM in 10 mM Tris-HCl, pH 8.5, and stored at either 4 °C for immediate usage or -20 °C for longer term usage. DNA purification kits, miRNeasy Mini Kit, and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). SoluLyse reagent was purchased from Genlantis (San Diego, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators and 5,000 MWCO regenerated cellulose ultrafiltration membranes were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). TEV protease was purchased from the University of California, Berkeley Macro Lab (Berkeley, CA).

Bacterial strains. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection (Manassas, VA), and *Streptomyces coelicolor* M1152 was obtained from the John Innes Centre (Norwhich, Norfolk, UK). *E. coli* DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production of BesA, BesE, BesG, and Ggt as well as for *in vivo* βes incorporation experiments. *E. coli*

GM272 harboring the non-transmissible, *oriT*-mobilizing plasmid pUZ8002 was used for conjugative plasmid transfer into *S. cattleya. E. coli* B834(DE3) was used for *in vivo* studies involving incorporation of Pra into proteins using PraRS.

Construction of plasmids for protein expression. Gibson ligation was used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase. Oligonucleotides used are listed in *Table S1*. To construct pet16b-BesA, pet16b-BesE, pet16b-BesG, pet16b-TRS-SCAT, pet16b-TRS-COLI, primer sets p16b-BesA-A/F, p16b-BesE-A/F, p16b-BesG-A/F, p16b-TRS-SCAT-A/F, p16b-TRS-COLI-A/F. Genes *besA, besE, besG,* and *thrS* (SCAT) were amplified from *S. cattleya* genomic DNA, while *thrS* (COLI) was amplified from *E. coli* genomic DNA. Corresponding PCR fragments cloned into pET16b using NdeI/BamHI cut sites. To construct pETDuet-BesAE, primer sets pETDuet-BesA-F/R and pETDuet-BesE-F/R were used to amplify DNA from *S. cattleya* which was then inserted sequentially into pETDuet-1 by Gibson ligation using HindIII/NcoI and KpnI/NdeI cut sites respectively. To clone pET28a-Ggt, primer set pET28a-Ggt-F/R was used to amplify *Ggt* from *E. coli* DH10B, which was subsequently cloned into pET28a using NdeI/EcoRI cut sites. pEAGgt was constructed by sequentially cloning PCR fragments amplified using pEAGgt-E-F/R (*S. cattleya*), pEAGgt-A-F/R (*S. cattleya*), and pEAGgt-Ggt-F/R (*E. coli*) into pet28a using Gibson ligation and cut sites NdeI/KpnI, KpnI, and KpnI, respectively. pβes was cloned by amplifying two fragments and sequentially cloning them into petDUET-1 by Gibson ligation. Fragment A was amplified from pPra using primer sets pBes-A-F/R and cloned using NcoI/HindIII cut sites while fragment B was amplified from pEAGgt using primer sets pBes-B-F/R and cloned using cut sites EcoRV. To construct pSV272-MBP-BesG, primer set pSV272-MBP-BesG-F/R was used to amplify *besG* from *S. cattleya* genomic DNA, which was then cloned into pSV272 using SfoI cut site. To construct pEAGgt, primer sets pEAGgt-E-F/R, pEAGgt-A-F/R, and pEAGgt-Ggt-F/R were used to amplify DNA from *S. cattleya* (*besE & besA*) or *E. coli* DH10B (*Ggt*), which were cloned in series into pACYCDuet-1 using: NdeI/KpnI, KpnI, and KpnI cut sites. Following plasmid construction, all cloned inserts were sequenced at Genewiz (San Francisco, CA).

Expression of His$_{10}$-tagged proteins. *E. coli* BL21 Star (DE3) was transformed with appropriate protein expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics in a 2.8 L-baffled shake flask to OD$_{600}$ = 0.05. The cultures were grown at 37°C at 200 rpm to OD$_{600}$ = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.2 mM) and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 × *g* for 7 min at 4ºC and stored at -80ºC.

Purification of His$_{10}$-BesA, His$_{10}$-BesE, His$_{10}$-BesG, and His$_{10}$-TRS-SCAT, His$_{10}$-TRS-COLI. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5) supplemented with PMSF (0.5 mM). The cell paste was homogenized and lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 9,000 psi. The lysate was then centrifuged at 13,500 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (*w/v*) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at 13,500 × *g* for 20 min at 4°C. The soluble lysate was incubated with Ni-NTA (0.5 mL resin/g of cell paste) for 45 min at 4°C, then resuspended and loaded onto a column by gravity flow. Column was washed

with wash buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5) for 15-20 column volumes. The column was then eluted with elution buffer (50 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole, 20 mM βME, 10% (*v/v*) glycerol, pH 7.5). rotein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (*v/v*) glycerol, 1 mM DTT, pH 7.5) using PD-10 desalting columns.

Final protein concentrations before storage were estimated using the $\varepsilon_{280\ nm}$ calculated by ExPASY ProtParam and measured by nanodrop, and are as follows: His_{10}-BesA: 5.6 mg/mL ($\varepsilon_{280\ nm}$ = 40,255 $M^{-1}\ cm^{-1}$), His_{10}-BesE: 2.9 mg/mL ($\varepsilon_{280\ nm}$ = 48,930 $M^{-1}\ cm^{-1}$), His_{10}-BesG: 0.45 mg/mL ($\varepsilon_{280\ nm}$ = 26,930$^{-1}\ cm^{-1}$), and His_{10}-TRS-SCAT: 5.4 mg/mL ($\varepsilon_{280\ nm}$ = 75,750 $M^{-1}\ cm^{-1}$), His_{10}-TRS-COLI: 6.0 mg/mL ($\varepsilon_{280\ nm}$ = 97,010 $M^{-1}\ cm^{-1}$).

Derivatization and detection of β-ethynylserine with 3-N$_3$-7-OH-Coumarin. 25 μL of cell culture supernatant or *in vitro* reaction was added to a solution containing 3-azido-7-hydroxy-coumarin (20 μM), THPTA (100 μM), and sodium ascorbate (10 mM) with and without $CuSO_4$ (50 μM) in 25 μL of H_2O. After incubating for 30 min, samples were quenched by addition of 5 μL of 10% formic acid (v/v). Aliquots were then analyzed using an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 μm, 2.1 × 50 mm; Agilent) using a linear gradient from 5% to 95% acetonitrile over 5 min at a flow rate of 0.6 mL/min with 0.1% (*v/v*) formic acid, pH 3.5 as the mobile phase. Mass spectra were acquired using an Agilent 6530 QTOF with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

***In vitro* assay for βes production with BesA, BesE, and Ggt.** Reactions (50 μL) contained sodium αKG (5 mM), sodium ascorbate (5 mM), and $(NH_4)_2Fe(SO_4)_2$ · $6H_2O$ (1 mM), L-glutamate (5 mM), ATP (2.5 mM), and $MgCl_2$ (5 mM) in 100 mM HEPES (pH 7.5). Reactions with Ggt contained 1 U of γ-glutamyl transpeptidase from equine kidney, while reactions with BesAE contained BesA (6.4 μM) and BesE (4.4 μM). Reactions were initiated by addition of L-propargylglycine (2 mM) and allowed to proceed for 2 h. Samples were then derivatized by CuAAC with 3-N$_3$-7-OH-coumarin analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for 3-N$_3$-7-OH-coumarin CuAAC analysis.

***In vivo* biosynthesis and incorporation of βes with BesA, BesE, and Ggt or Pra with PraRS.** Single colonies of *E. coli* BL21 Star (DE3) transformed with appropriate plasmids (Empty plasmid control: pETDUET-1/pet28a, BesAE: pGBes/pet28a, BesAE + Ggt: pGBes/pGgt) were inoculated into LB Cb Km (10 mL) and grown overnight at 37°C with shaking at 200 rpm. For Pra incorporation, *E. coli* BL21 Star (DE3) transformed with pBAD24-PraRS and cultured with appropriate antibiotics. Overnight cultures were inoculated into LB Cb Km (25 mL) in a 250 mL-baffled shake flask to OD_{600} = 0.05 and grown at 37°C and 200 rpm until the OD_{600} reached 0.6 to 0.8. Cultures were cooled on for 20 min before inducing with IPTG (0.2 mM) and then cultured for 48 h at 16°C and 200 rpm. Cells were washed twice with 0.9% (*w/v*) NaCl by pelleting and resuspension. The cell pellet was then resuspended in M9 media containing 33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM NH_4Cl, 0.4% glucose, 100 mg/L of the 20 proteinogenic amino acids, 1 mM L-propargylglycine, 1 mM $MgSO_4$, 300 μM $CaCl_2$, and 1 mg/ml thiamine to OD_{600} = 0.4 - 0.6 and cultured for 24 h at 16 °C and 200 rpm. For cultures that contained additional Thr, 2 mM was added. Supernatant samples were then derivatized by CuAAC with 3-N$_3$-7-OH-coumarin analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for 3-N$_3$-7-OH-coumarin CuAAC analysis. Pellet

samples were derivatized by CuAAC with Tamra-Azide and analyzed by fluorescent-PAGE using protocol for derivatization of E. coli proteome with Tamra-Azide.

***In vitro* assay for translation and analysis of βes-containing peptide.** Peptide containing βes was synthesized through *in vitro* transcription translation (IVTT) using PURExpress (Δaa, tRNA; NEB). Manufacturer protocol was followed with the following modifications: an amino acid solution containing 100 mg / L of 19 canonical amino acids (-Thr) was used instead of the supplied 20 amino acid mixture; reactions with Thr contained 333 μM L-Thr; reactions with **β**es contained 333 μM of chemically synthesized L-βes. 50 ng of pPeptide-ACC was used as template; reactions were ran overnight at 37 °C. Reactions were then diluted with 1 volume of 50 mM HEPES (pH 7.5) and passed through a Nanosep 10k Omega centrifugal filter by centrifuging for 5 min at $9,800 \times g$. For condition where βes-containing peptide is derivatized with 3-N_3-7-OH-Coumarin, protocol for 3-N_3-7-OH-coumarin CuAAC reaction and analysis was followed. Resulting samples were then analyzed using an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 μm, 2.1×50 mm; Agilent) using a linear gradient from 5% to 95% acetonitrile over 5 min at a flow rate of 0.6 mL/min with 0.1% (*v/v*) formic acid, pH 3.5 as the mobile phase. Mass spectra were acquired using an Agilent 6530 QTOF with the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

Derivatization of *E. coli* proteome with Tamra-Azide. 1 mL of cell culture (OD 2-3) was centrifuged at $9,800 \times g$ for 5 min and washed two times with 1 mL of 0.9 % NaCl to remove unincorporated, free amino acid. The resulting cell pellet was resuspended in SoluLyse (150 μL) and incubated for 1 h at room temperature. Samples were centrifuged at $9,800 \times g$ for 5 min and the supernatant (50 μL) was mixed with an assay solution (50 μL) containing Tamra-Azide (40 μM), THPTA (100 μM), and sodium ascorbate (10 mM) in PBS (pH 7) with or without $CuSO_4$ (50 μM). The labeling reaction was incubated for 1 h at room temperature. Laemmli buffer was added and samples were heated at 95ºC for 5 min before analyzing by SDS-PAGE (MiniProtean 8-16% TGX gels; BioRad) using a constant current (30 A) and maximum voltage (200 V) for 1.5 h. Gels were imaged for fluorescence using a ChemiDoc MP Imager (Biorad; λ_{ex} = 520-545 nm, λ_{em} = 577-613 nm, exposure = 150 s).

Proteomic analysis of *E. coli* with endogenous βes production. Cell pellets for *E. coli* BL21 Star (DE3) pEAGgt and *E. coli* BL21 Star (DE3) pET-DUET-1 were lysed by bead beating. Proteins from soluble fraction were precipitated by adding cold acetone to the lysate (10:1 (v/v) and pelleted by centrifugation at 14,000 xg for 15 min at 4 °C. Supernatant was decanted and the protein pellet was left to dry at room temperature for 1 h. The pellet was then resuspended in 100 μL in buffer (8M Urea, 75 mM NaCl, 1 mM EDTA, 50 mM Tris/HCl, pH 8.0). Disulfides were reduced by adding 5 mM DTT. Cysteines were alkylated by adding 10 mM iodoacetamide. To digest, samples were first diluted 1: 4 with 50 mM Tris/HCl (pH 8.0), then sequencing grade trypsin (Promega) was added at a ratio of 1:50 (enzyme : total protein w/w). Digestion was allowed to proceed overnight at 37 °C. Samples were acidified by adding formic acid to 5% final concentration, and subsequently centrifuged at 14,000 xg for 15 min at 4 °C to remove precipated protein. Supernatant was subsequently used for LC-MS/MS.

LC-MS/MS analysis was performed on a Thermo Q-Exactive HF. 1mg of total peptides were analyzed on an Eksigent nanoLC- 415 HPLC system (Sciex) coupled via a 25cm C18 column (inner diameter 100mm packed in-house with 2.4mm ReproSil-Pur C18-AQ medium, Dr.

Maisch GmbH). Peptides were separated at a flow rate of 200 nL / min with a linear 106min gradient from 2% to 25% solvent B (100% acetonitrile, 0.1% formic acid), followed by a linear 5min gradient from 25 to 85% solvent B. Samples were run for 170 min. Data was acquired in data dependent mode using Xcalibur 2.8 software. MS1 Spectra were measured with a resolution of 60,000, an AGC target of 3E6 and a mass range from 375 to 2000m/z. Up to 15 MS2 spectra per duty cycle were triggered at a resolution of 15,000, an AGC target of 2E5, an isolation window of 1.6 m/z and a normalized collision energy of 27.

Data analysis, was performed using Petunia web server using the following parameters: masses [peptide mass tolerance = 20, peptide mass unit = 2, mass type parent = monoisotopic, mass type fragment = monoisotopic, isotope error = C13 error]; search enzyme [number of enzyme termini = fully digested, maximum number of missed cleavages = 2]; variable modifications [Met +15.99, Thr + 9.98, Cys + 57]; fragment ions [fragment bin tolerance = 0.02, fragment bind offset = 0, use B and Y ions]; misc parameters [digest mass range = 400-5000, number of results = 100, max fragment charge = 6, max precursor charge = 6]. For filtering data, the following additional criteria were used: T-111 modification, only +1,+2,+3 charge states; min xcorr = 3; min delta scn = 0.1; max Sp rank = 2.

General procedure for high resolution HPLC/MS and HPLC/MS-MS analysis of polar metabolites. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1 × 100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra acquired on an Agilent 6530 QTOF (Agilent) used the following source and acquisition parameters: Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum. Mass spectra acquired on an Agilent 6460C QQQ with Agilent Jet Stream source used the following source and acquisition parameters: fragmentor voltage = 60 V, collision energy = 0 V; Gas temperature = 325 ºC; drying gas = 10 l/min; nebulizer = 35 psig; capillary voltage = 3500 V; sheath gas temperature = 350 ºC; sheath gas flow 12 l/min; fragmentor 60 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum. Ions were monitored by MRM using the following transitions: βes (m/z = 130.1→112.1); Thr (m/z = 120.1→102.1); Pra (m/z = 114.1→96.1). See Appendix Figure A2.3 for additional information of transitions.

General procedure for HPLC/MS-MS analysis of aminoacyl-tRNA. Samples were ran on an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 μm, 2.1 × 50 mm; Agilent) using a linear gradient from 5% to 30% acetonitrile over 40 s, followed by a linear gradient from 30% to 100% acetonitrile over 90 s at a flow rate of 0.7 mL/min with 10 mM ammonium as the mobile phase. Mass spectra were acquired using an Agilent 6460C QQQ, with Agilent Jet Stream as a source, with the following source and acquisition parameters: Gas temperature = 300 ºC; gas flow = 5 l/min; nebulizer = 45 psig; sheath gas temperature 250 ºC; sheath gas flow = 11 l/min; capillary voltage = 3500 V; fragmentor 135 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum; collision energy 15 V. Ions were monitored by MRM using the following transitions: Thr-adenylate (m/z = 369.1 → 136.1); βes-adenylate (m/z = 379.1 → 136.1).

General procedures for the synthesis of βes. All chemicals were purchased from Sigma-Aldrich and used without further purification. NMR characterization was performed on a Bruker AV-600 (600 MHz) instrument equipped with either a 5 mm broadband observe Z-gradient probe or a 5 mm triple inverse broadband Z-gradient probe. Spectra were acquired at 298 K and chemical shifts were expressed relative to tetramethylsilane. Scheme for synthesis of βes is available in Appendix Figure A2.4.

Synthesis of (*R*)-2,2-dimethyloxazolidine-3,4-dicarboxylic acid 3-tert-butyl ester 4-methyl ester (Bes-Cpd2, D-Garner's ester). To N-(tert-butoxycarbonyl)-D-serine methyl ester (10 g, 45.6 mmol) dissolved in acetone (140 mL) was added p-toluenesulfonic acid (100 mg) and 2,2-dimethoxypropane (56 mL, 0.46 mol). The mixture was heated to reflux and stirred for 4 h. Additional p-toluenesulfonic acid (100 mg) and 2,2-dimethoxypropane (28 mL, 0.23 mol) was added, following which the reflux was continued for 4 h and the resulting mixture was concentrated. The residue was dissolved in Et2O (80 mL) and washed with sat. NaHCO3 (3 x 50 mL). The combined aqueous layers were back-extracted with Et2O (3 x 40 mL). The combined organic layers were then washed with brine (50 mL), dried over Na2SO4, and concentrated. Purification by silica gel chromatography (1:1 hexanes/Et2O) provided the product as a yellow oil (9.86 g, 83%). 1H NMR (600 MHz, CDCl3) δ 4.49 (dd, J = 6.8, 2.6 Hz, 0.4H), 4.38 (dd, J = 7.1, 3.1 Hz, 0.6H), 4.17-4.10 (m, 1H), 4.07-4.01 (m, 1H), 3.76 (s, 3H), 1.67 (s, 1.7H), 1.64 (s, 1.3H), 1.53 (s, 1.8H), 1.49 (s, 5H), 1.41 (s, 5.2H). 13C NMR (151 MHz, CDCl3) δ 171.6, 171.2, 152.0, 151.1, 95.0, 94.3, 80.8, 80.2, 66.2, 65.9, 59.2, 59.1, 52.3, 52.2, 28.3, 28.2, 26.0, 25.1, 24.9, 24.3. Most peaks are doubled due to carbamate rotamers.

Synthesis of (*R*)-4-formyl-2,2-dimethyloxazolidine-3-carboxylic acid *tert*-butyl ester (Bes-Cpd3, *D*-Garner's aldehyde). In an oversized flask under nitrogen atmosphere, D-Garner's ester (5.19 g, 20 mmol) was dissolved in anhydrous toluene (40 mL). The contents were cooled to -78 ºC, then a solution of DIBAL-H (1 M in toluene, 35 mL) was added dropwise via syringe over 15 min, ensuring the reagent was trickled down the wall of the cold flask. The reaction was stirred at -78 ºC for 2 h then slowly quenched with MeOH (10 mL). The mixture was poured into 1 M Rochelle's salt (120 mL) and stirred vigorously for 1 h to obtain clear phase separation. The organic layer was set aside, and the aqueous layer was extracted with Et2O (3 x 50 mL). The combined organic layers were washed with brine (50 mL), dried over Na2SO4, and concentrated. Purification by silica gel chromatography (1:1 hexanes/Et2O) provided the product as a light yellow oil (3.66 g, 80%). 1H NMR (600 MHz, CDCl3) δ 9.61 (s, 0.4H), 9.55 (d, J = 2.6 Hz, 0.6H), 4.36-4.32 (m, 0.4H), 4.21-4.17 (m, 0.6H), 4.13-4.03 (m, 2H), 1.65 (s, 2H), 1.60 (s, 1.6H), 1.56 (s, 2.4H), 1.51 (s, 3H), 143 (s, 6H). 13C NMR (151 MHz, CDCl3) δ 199.4, 199.3, 152.6, 152.3, 95.0, 94.3, 81.3, 81.0, 64.7, 64.6, 63.9, 63.4, 28.25, 28.20, 26.6, 25.7, 24.6, 23.7. Most peaks are doubled due to carbamate rotamers.

Synthesis of (*R*)-4-((*R*)-1-hydroxy-3-(trimethylsilyl)prop-2-yn-1-yl)-2,2-dimethyloxazolidine -3-carboxylate *tert*-butyl ester (Bes-Cpd4). A flask fitted with a reflux condenser was placed under nitrogen atmosphere, charged with anhydrous THF (30 mL), and cooled to 0 ºC. Ethynyltrimethylsilane (2.2 mL, 16 mmol) was added, followed by ethylmagnesium bromide (1 M in THF, 15 mL) dropwise. The mixture was heated to reflux with stirring for 2 h and then cooled to room temperature. Meanwhile, in a second flask under nitrogen, copper(I) iodide (4.19 g, 22 mmol) was dissolved in anhydrous THF (30 mL) and dimethyl sulfide (12 mL). This solution was stirred at room temperature for 1.5 h, then cooled to -78 ºC. The Grignard solution was cannulated into the copper solution dropwise, then the mixture was warmed to -30 ºC

(acetonitrile/dry ice) and stirred for 30 min. After cooling the organocopper reagent to -78 ºC, D-Garner's aldehyde (2.29 g, 10 mmol) was added dropwise as a solution in THF (5 mL). The mixture was stirred for 15 min at -78 ºC, then warmed to room temperature and stirred for 16 h overnight. The reaction was quenched with sat. NH4Cl (100 mL) and stirred for 30 min to achieve complexation of copper salts, then the product was extracted with Et2O (3 x 80 mL). The combined organic layers were washed with 0.5 M HCl (2 x 50 mL, precipitates residual copper salts) and brine (50 mL), then dried over Na2SO4 and concentrated to ~20 mL. Precipitates were removed by filtration over Celite and rinsing the solids with Et2O (3 x 20 mL). After further concentration to ~5 mL, the final precipitates were removed by filtration through a glass pipette with a plug of filter paper. Purification by silica gel chromatography (1:1 hexanes/Et2O) provided the product as a very viscous yellow oil (2.19 g, 67%). 1H NMR (600 MHz, DMSO-d6) δ 5.75 (d, J = 5.5 Hz, 0.5H), 5.72 (d, J = 5.5 Hz, 0.5H), 4.69 (t, J = 5.1 Hz, 0.5H), 4.64 (t, J = 5.2 Hz, 0.5H), 4.01 (dd, J = 9.3, 3.4 Hz, 1H), 3.94 (q, J = 7.9 Hz, 1H), 3.82 (m, 0.5H), 3.78 (m, 0.5H), 1.51 (s, 3.5H), 1.41-1.39 (2 s, 11.5 H), 0.10 (s, 9H). 13C NMR (151 MHz, DMSO-d6) δ 152.3, 151.6, 106.7, 106.5, 94.3, 94.0, 89.3, 89.1, 80.0, 79.6, 64.5, 64.1, 62.0, 61.1, 60.5, 28.45, 28.39, 26.8, 25.9, 25.2, 23.9, 0.2. Most peaks are doubled due to carbamate rotamers.

Synthesis of (*R*)-4-((*R*)-1-acetoxy-3-(trimethylsilyl)prop-2-yn-1-yl)-2,2-dimethyloxazolidine-3-carboxylate *tert*-butyl ester (Bes-Cpd5). Bes-Cpd4 (1.15 g, 3.5 mmol) was dissolved in pyridine (10 mL), then acetic anhydride (1.0 mL, 10.5 mmol) and DMAP (21 mg, 0.18 mmol) were added. The mixture was stirred at room temperature for 1.5 h, then was diluted with Et2O (30 mL) and washed with water (3 x 20 mL). The combined aqueous layers were back-extracted with Et2O (3 x 20 mL). The combined organic layers were then washed with brine (40 mL), dried over Na2SO¬4, and concentrated. Purification by silica gel chromatography (4:1 hexanes/Et2O) provided the product as a light yellow oil that crystallized to a white solid upon cooling (1.10 g, 85%). 1H NMR (600 MHz, CDCl3) δ 5.88 (d, J = 4.9 Hz, 0.5H), 5.80 (d, J = 6.1 Hz, 0.5H), 4.21-4.11 (m, 1.5H), 4.03-3.97 (m, 1H), 3.96-3.91 (m, 0.5H), 2.08-2.06 (2 s, 3H), 1.64-1.48 (5 s, 15 H), 0.16 (s, 9H). 13C NMR (151 MHz, CDCl3) δ 169.1, 169.0, 152.4, 151.7, 100.0, 99.5, 95.0, 94.4, 92.4, 91.9, 80.6, 64.7, 64.6, 64.2, 64.1, 59.3, 58.4, 28.34, 28.28, 26.8, 25.9, 24.8, 23.6, 20.9, -0.4. Most peaks are doubled due to carbamate rotamers.

Synthesis of (3*R*,4*R*)-4-((*tert*-butoxycarbonyl)amino)-5-hydroxy-1-(trimethylsilyl)pent-1-yn-3-yl acetate (Bes-Cpd6). Bes-Cpd5 (369.5 mg, 1 mmol) was dissolved in 3:1 AcOH/H2O (20 mL) and stirred with heating at 50 ºC for 16 h overnight. The solvent was evaporated, then the residue was dissolved in EtOAc (30 mL) and washed with sat. NaHCO3 (2 x 20 mL). The combined aqueous layers were back-extracted with EtOAc (3 x 20 mL). The combined organic layers were then washed with brine (40 mL), dried over Na2SO4, and concentrated. Purification by silica gel chromatography (2:1 hexanes/EtOAc) provided the product as a white solid (Rf 0.17, 161.4 mg, 49%). 1H NMR (600 MHz, CDCl3) δ 5.59 (d, J = 5.5 Hz, 1H), 4.98 (d, J = 9.1 Hz, 1H), 3.99-3.92 (m, 1H), 3.90-3.84 (m, 1H), 3.69-3.63 (m, 1H), 2.24-2.18 (m, 1H), 2.11 (s, 3H), 1.45 (s, 9H), 0.17 (s, 9H). 13C NMR (151 MHz, CDCl3) δ 169.6, 155.6, 99.9, 92.5, 79.9, 63.5, 61.9, 54.7, 28.3, -0.4. The minor product in which the acetyl group migrated to the primary position can be recovered as a yellow oil (Rf 0.30, 40.3 mg, 12%). 1H NMR (600 MHz, CDCl3) δ 5.01 (d, J = 7.9 Hz, 1H), 4.51 (t, J = 5.4 Hz, 1H), 4.34-4.28 (m, 1H), 4.28-4.21 (m, 1H), 4.06-3.99 (m, 1H), 2.92 (d, J = 5.1 Hz, 1H), 2.11 (s, 3H), 1.48 (s, 9H), 0.19 (s, 9H). 13C NMR (151 MHz, CDCl3) δ 171.1, 156.0, 103.1, 91.2, 80.0, 63.0, 62.5, 53.8, 28.3, 20.9, -0.3.

Synthesis of (2*S*,3*R*)-3-acetoxy-2-((tert-butoxycarbonyl)amino)-5-(trimethylsilyl)pent-4-ynoic acid (Bes-Cpd7). Bes-Cpd6 (108.7 mg, 0.33 mmol) was dissolved in acetone (10 mL). Jones reagent (1 M CrO3 in 10% aqueous H2SO4, 2.0 mL) was added dropwise and the mixture stirred for 16 h overnight at room temperature. Isopropanol (2 mL) was added to quench the reaction and the majority of solvent was evaporated. The residue was dissolved in EtOAc (30 mL) and washed with 0.5 M HCl (2 x 20 mL). The combined aqueous layers were back-extracted with EtOAc (3 x 20 mL). The combined organic layers were then washed with brine (40 mL), dried over Na2SO4, and concentrated. Purification by silica gel chromatography (2:1 hexanes/EtOAc + 1% AcOH) followed by concentration from toluene to remove acetic acid provided the product as a white solid (98.0 mg, 86%). 1H NMR (600 MHz, methanol-d4) δ 5.79 (br, 1H), 4.94 (br, 1H), 2.08 (s, 3H), 1.50 (s, 9H), 0.17 (s, 9H). 13C NMR (151 MHz, methanol-d4) δ 169.4, 156.5, 99.3, 91.1, 79.4, 64.4, 27.4, 19.1, -1.7. Some carbon resonances are not visible due to deuterium-induced signal broadening.

Synthesis of (2*S*,3*R*)-2-((*tert*-butoxycarbonyl)amino)-3-hydroxypent-4-ynoic acid (Bes-Cpd8) and (2S,3R)-2-Amino-3-hydroxypent-4-ynoic acid hydrochloride (Bes-HCl). Bes-Cpd7 (51.5 mg, 0.15 mmol) was dissolved in 2 M methanolic KOH (2 mL) and stirred at room temperature for 2 h. The mixture was diluted with methanol (20 mL) and then pre-washed Dowex 50W was added until the solution was neutral to water-wetted pH paper. The resin was filtered off and rinsed with methanol (2 x 20 mL). The solvent was evaporated to provide crude Bes-Cpd8 that was used directly in the next step.

Bes-Cpd8 was dissolved in 6 M HCl (2 mL) and stirred at room temperature for 2 h. The mixture was diluted with water (6 mL) and washed with EtOAc (3 x 6 mL). The aqueous layer was lyophilized to provide the product as a light brown powder (23.6 mg, 95% over 2 steps). 1H NMR (600 MHz, D2O) δ 5.00 (t, J = 2.6 Hz, 1H), 4.17 (d, J = 3.0 Hz, 1H), 2.98 (d, J = 2.2 Hz, 1H). 13C NMR (151 MHz, D2O) δ 168.9, 78.7, 77.6, 59.5, 57.6.

***In vitro* transcription and purification of tRNA and aminoacyl-tRNA.** dsDNA oligos *for in vitro* tRNA transcription (*Appendix A1.2C*) containing T7 promoter and corresponding tRNA were synthesized by IDT. PCR using Phusion polymerase with primers tRNA-Express-F and corresponding reverse primer (primer name reads acceptor stem 5' to 3', *Appendix A1.2B*) was used to amplify DNA from the synthetic oligos to obtain high-concentration (200 ng/uL) DNA used for *in vitro* transcription. The HiScribe T7 High Yield RNA Synthesis Kit (NEB) was used for *in vitro* transcription, with the recommended conditions for short RNAs. Transcription reactions contained 1 μg of DNA template and were ran overnight at 37 °C. To make aminoacyl-tRNA, TRS (1 μM), ATP (5 mM), MgCl2 (10 mM), amino acid (βes or Thr, 1 mM), HEPES (50 mM, pH 7.5) in 20 μL of nuclease free H2O was added to 30 μL of HiScribe reaction (from overnight) and allowed to incubate for 30 min at 37 °C. tRNA or aminoacylated-tRNA was then purified from these reactions directly using miRNeasy purification kit (QIAGEN) following manufacturer protocols.

Quantification of aminoacyl-tRNA. Aminoacyl-tRNA was quantified after purification by hydrolysis of a subset of βes- or Thr-tRNA followed by quantification using LC/MS-MS. To hydrolyze βes- or Thr-tRNA, 5 μL of purified tRNA was mixed with 5 μL of 0.1 M NaOH for 1 h at 37 °C. Reactions were neutralized by adding 10 μL of 0.1 M NaHPO4, then mixed 1:1 (v/v) with MeOH for analysis on polar phase LC method. Standards of βes or Thr were ran alongside hydrolyzed samples and general procedure for high resolution HPLC/MS-MS analysis of polar

metabolites was used. The MRM transitions βes (m/z = 130.1→112.1) and Thr (m/z = 120.1→102.1) were chosen for generating standard curve.

***In vitro* acylation reactions with BesG.** Reactions (25 µL) contained tRNA^{Thr} (20 µM), ATP (5 mM), MgSO$_4$ (10 mM), HEPES (100 mM, pH 7.5), TRS (either from *E. coli* or *S. cattleya*, 100 nM) with or without BesG (1 µM). Reactions were initiated by the addition of βes (0.1-1mM), L-Thr (0.1-1mM) or both (as indicated) and incubated at 37 °C. Samples (2.5 µL) were removed every 5 minutes and added to a solution of RNaseA (1 U/µL) in ammonium formate (75 mM). Protein was removed by precipitation with 5% TCA (v/v) and centrifugation at 14,000 g for 10 minutes. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QQQ using protocol for aminoacyl-tRNA analysis.

***In vitro* deacylation timecourse of various tRNA^{Thr} by BesG.** Deacylation studies of tRNA^{Thr} from *S. cattleya* and *E. coli* were performed by using tRNA^{Thr} acylated with respective TRS (either from *S. cattleya* or *E. coli*, respectively). High concentration of tRNA was prepared by IVT as described previously. Prior to tRNA purification, tRNA from overnight IVT reactions (30 µL) was charged by the addition of ATP (5 mM), MgCl$_2$ (10 mM), βes or Thr (1 mM), and TRS (1 µM), followed by incubation at 37°C for 30 min. Charged tRNA was then purified using miRNeasy kit to yield pure aminoacyl-tRNA. Immediately after, a portion of aminoacyl-tRNA was hydrolyzed for quantification, and quantified using method previously described. Reactions for deacylation timecourse contained MgCl$_2$ (10 mM), ZnCl$_2$ (1 µM), BesG (1 µM), HEPES (pH 7.5) in water. Reactions were initiated by addition of aminoacyl-tRNA (3.5 µM) and immediately incubated at 37°C. Samples (2.5 µL) were removed every 5 minutes and added to a solution of RNaseA (1 U/µL) in ammonium formate (75 mM). Protein was removed by precipitation with 5% TCA (v/v) and centrifugation at 14,000 g for 10 minutes. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QQQ using protocol for aminoacyl-tRNA analysis.

***In vitro* platereader assay of BesG.** Reactions were performed in 384 well plate using a Synergy HTX plate reader (Biotek) by monitoring at 340 nM with temperature set to 30 °C. Reactions (25 µL) contained TRS (2.5 µM), ATP (5 mM), MgCl$_2$ (10 mM), KCl (25 mM), DTT (5 mM), NADH (400 µM), inorganic pyrophosphatase (2 U/mL), myokinase (10 U/mL), pyruvate kinase/lactate dehydrogenase (10 U/mL), HEPES (50 mM, pH 7.5), Thr or βes (2 mM), and tRNA^{Thr} in nuclease-free H$_2$O. tRNA^{Thr} was allowed to amino-acylate for 5-10 min before initiation of reaction through the addition of BesG (1-10 µM). Rates for BesG hydrolysis were obtained from ΔA$_{340}$. Rates of background hydrolysis were subtracted from BesG-catalyzed rates. Since TRS was provided n excess, tRNA in each reaction was assumed to be fully charged at steady state. To measure concentration of amino-acylated tRNA, prepared tRNA was fully charged by TRS using conditions as described above (without addition of BesG) and quenched in 2 volumes of 200 mM ammonium acetate, pH 5.2, 1.5 U/uL RNAse A. Amount of amino-acyl-tRNA was quantified by comparing the LC/MS-MS response of amino-adenylate to a standard of pure amino-adenylate. Concentration of pure amino-adenylate was quantified by fully hydrolyzing amino-acyl-tRNA and quantifying free amino acid using method for polar metabolite analysis with MS-MS.

***In vitro* deacylation reactions with BesG on platereader.** Kinetic studies of BesG were performed by coupling deacylation NADH consumption using TRS, myokinase, lactate dehydrogenase, and pyruvate kinase. Reactions were performed in 384 well plate using a Synergy HTX plate reader (Biotek) by monitoring at 340 nM with temperature set to 30 °C.

Reactions (25 μL) contained TRS (2.5 μM), ATP (5 mM), MgCl$_2$ (10 mM), KCl (25 mM), DTT (5 mM), NADH (400 μM), inorganic pyrophosphatase (2 U/mL), myokinase (10 U/mL), pyruvate kinase/lactate dehydrogenase (10 U/mL), HEPES (50 mM, pH 7.5), Thr or βes (2 mM), and tRNA^{Thr} in nuclease-free H$_2$O. tRNA^{Thr} was allowed to amino-acylate for 5-10 min before initiation of reaction through the addition of BesG (1-10 μM). Rates for BesG hydrolysis were obtained from ΔA$_{340}$. Rates of background hydrolysis were subtracted from BesG-catalyzed rates. Since TRS was provided n excess, tRNA in each reaction was assumed to be fully charged at steady state. To measure concentration of amino-acylated tRNA, prepared tRNA was fully charged by TRS using conditions as described above (without addition of BesG) and quenched in 2 volumes of 200 mM ammonium acetate, pH 5.2, 1.5 U/uL RNAse A. Amount of amino-acyl-tRNA was quantified by comparing the LC/MS-MS response of amino-adenylate to a standard of pure amino-adenylate. Concentration of pure amino-adenylate was quantified by fully hydrolyzing amino-acyl-tRNA and quantifying free amino acid using method for polar metabolite analysis with MS-MS.

Generation of *E. coli* BL21 Star (DE3) *thrT*^{A73U}, *thrU*^{A73U}, *thrV*^{A73U}, and *thrW*^{A73U} mutants. A single colony of *E. coli* BL21 Star (DE3) pCas9 was picked into LB with appropriate antibiotics (10 mL) and grown overnight at 200 rpm, 30 °C. Overnight cultures were then inoculated at OD600 = 0.05 in LB with 0.2% arabinose and appropriate antibiotics, and grown at 200 rpm, 30 °C until OD600 = 0.4-0.6. Cells were then concentrated by centrifugation at 9000 g for 5 min and washed by suspension 10% ice-cold glycerol. Concentration with washing was repeated four more times. Cells were finally pelleted and resuspended in 1/100th the original volume of 10% ice-cold glycerol and frozen at -80 °C until needed for electroporation. To generate pTarget-T, pTarget-U, pTarget-V, pTarget-Cm primers containing the corresponding overlapping guides were introduced into pTarget by PCR amplification and ligated using Gibson ligation, then transformed into *E. coli* DH10B for purification and sequencing. dsDNA repair fragments were generated by PCR amplification from E. coli BL21 Star (DE3) using primer sets HA-thrT-A/B-F/R, HA-thrU-A/B-F/R, HA-thrV-A/B-F/R, HA-thrW-A/B-F/R. Corresponding A and B fragments were ligated using SOE PCR, then re-amplified to obtain high concentration repair fragments (500-800 ng/uL). To generate *E. coli* BL21 Star (DE3) tRNA^{Thr} *thrT*^{A73U}, *thrV*^{A73U}, or *thrW*^{A73U} 50 μL of electro competent *E. coli* BL21 Star (DE3) (with or without additional tRNA^{Thr} mutations) was electroporated with 200 ng of corresponding pTarget and 800 ng of the corresponding dsDNA repair fragment. Cells were then recovered in LB for 1h, before plating at 30 °C on LB agar plates with appropriate antibiotics. Colony PCR, using HA-thrX-A-F and HA-thrX-B-R primers, was performed to amplify region of DNA containing mutation. Un-purified PCR products were then submitted to Genewiz (San Francisco, CA) for sequence verification using primers thrT-Seq, thrV-Seq, and thrW-Seq. To generate *E. coli* BL21 Star (DE3) tRNA^{Thr} *thrU*^{A73U}, LR-Cm-ThrU-F/R primers were used to amplify a Cm^R cassette containing 40bp of homology to the intergenic region between *thrU* and *tyrU*. Amplified corresponding cassette (200 ng) was electroporated into *E. coli* BL21 Star (DE3) for insertion via λ-red mediated recombination. Cells were recovered in LB for 1 h, and subsequently plated in LB agar plates, containing Cm, and grown overnight at 30 °C. A single colony of *E. coli* BL21 Star (DE3) *thrU*-Cm^R-*tyrU* was then picked into LB containing Cm and Km antibiotics, and grown overnight at 30 °C, 200 rpm for the preparation of electrocompetent cells as previously described. To remove CmR and introduce the *thrU*^{A73U} mutation, 50 μL of *E. coli* BL21 Star (DE3) *thrU*-Cm^R-*tyrU* was electroporated with 200 ng of pTarget-Cm and 800 ng of dsDNA thrU repair oligo. Cells were then recovered, plated, and screened for mutation as previously described. To cure pTarget,

cells were grown in LB with Km and induced immediately upon inoculation with IPTG. Cells were plated in LB with Km, and single colonies were picked and checked for Sp resistance by culturing. Using the methods described, we generated *E. coli* BL21 Star (DE3) tRNA^{Thr} *thrT*^{A73U} *thr*U^{A73U} *thr*V^{A73U} *thr*W^{A73U} by sequentially mutating *thrT*, *thrU*, *thrW*, and *thrV*. After each round of mutation, tRNA^{Thr} regions were sequenced verified to ensure that mutations had not reverted. pCas was cured by growing a single colony in LB, without antibiotic at 37 °C and 200 rpm overnight, sequentially diluting, then plating on LB agar.

Construction of plasmids for *besG* gene disruption and complementation in *S. cattleya*. Primers for amplifying homology arms for *besG* gene disruption were designed to produce amplicons of either 2 kB homology upstream of start codon and downstream of stop codon of *besG*. *S. cattleya* genomic DNA was used a as a template for amplification with primer sets pIJ773-BesG-A-F/R and pIJ773-BesG-B-F/R (**Table S1**). Amplified DNA for pIJ773-BesG-A and pIJ773-BesG-B were cloned into pIJ773 sequentially using restriction-ligation with NotI/SpeI and HindIII/KpnI cut sites. For gene complementation, primers pSET152-BesG-F/R were designed to clone native *S. cattleya besG* gene into pSET152 (Hyg^R) under an ermEp* promoter using NdeI/BamHI cut sites. Standard molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. Following plasmid construction, all cloned inserts were sequenced at Genewiz (San Francisco, CA).

Conjugative transfer of plasmids from *E. coli* to *S. cattleya*. Plasmids pIJ773-BesG for gene disruption or pSET152-BesG for gene complementation was transferred into *S. cattleya* by conjugation with *E. coli* GM272 harboring the *oriT*-mobilizing plasmid pUZ8002 using the literature protocol [20]. Gene disruption was verified by PCR amplification of the target locus and neighboring loci using primer set 'dck', followed by sequencing to validate correct disruption, generating *S. cattleya ΔbesG*::Am^R (*Appendix A1.2B*). Complementation of *S. cattleya besG* knockout strains was carried out by introducing an expression cassette for *besG* gene driven by the ermEp* promoter into the phage attachment site. *S. cattleya ΔbesG*::Am^R was transformed with pSET152-BesG to generate *S. cattleya ΔbesG*::Am^R *attB*^{ΦC31}::*ermEp*-besG*. Complementation was verified by PCR using primers designated to amplify wild-type gene (**Table S1**).

Generation of Phyre2 homology model of BesG. Sequence of BesG from *S. cattleya* (WP_014151582) was submitted to the Phyre2 webserver for generation. BesG was modeled with 100% confidence and 93% coverage, with a crystal structure of probable alanyl-trna-synthase from *Clostridium perfringens* being the top homologous hit.

4.3. Results and discussion

Production of βes from Pra *in vitro*. In Chapter 3, we discussed our efforts at the heterologous expression of βes pathway proteins in *E. coli* to produce 4-Cl-lysine, allylglycine, 4-Cl-allylglycine, and propargylglycine. In order to produce βes, two additional steps are necessary: formation of a glutamyl-dipeptide of propargylglycine (Glu-Pra), catalyzed by BesA, followed by hydroxylation of this dipeptide to form Glu-βes, catalyzed by BesE. The set of proteins clustered within the βes pathway lacked putative hydrolases that could hydrolyze Glu-βes [21]. However in the native host, βes and not Glu-βes, is observed as a product in the supernatant (*Figure 2.7*). This observation primed us to further explore how Glu-βes might be hydrolyzed *in vivo* in order produce free βes that could be used as an amino acid for translation.

One of our top candidates for a Glu-βes-hydrolase was a promiscuous enzyme called γ-glutamyl-transpeptidase (Ggt or γGT). Ggt catalyzes the reversible transfer of a γ-glutamyl group from one γ-glutamyl-glutamyl dipeptide onto an acceptor amino acid [22–24]. Though non-essential, Ggts are practically ubiquitous and have been observed across all orders of life. Indeed, *S. cattleya* possesses a Ggt (SCAT_4848) which was found to be transcribed and expressed as early as 24 h into growth in liquid cultures (*Table 4.1*). We therefore hypothesized that Ggt can serve as a general Glu-βes hydrolase for production of βes *in vitro* or *in vivo* (*Figure 4.2*).

Transcriptomics (RNA-Seq)		Proteomics (Shotgun)	
Gene	*Reads*	*Protein*	*Spectral counts*
SCAT_4848 (*ggt*)	2890	SCAT_4848 (Ggt)	202
SCAT_p1047 (*besC*)	3549	SCAT_p1047 (BesC)	17

Table 4.1. Transcriptomic and proteomic analysis of *S. cattleya* indicates that SCAT_4848 (Ggt) is both transcribed and expressed. For comparison, data for a βes pathway gene (*besC*) and protein (BesC) are also shown.
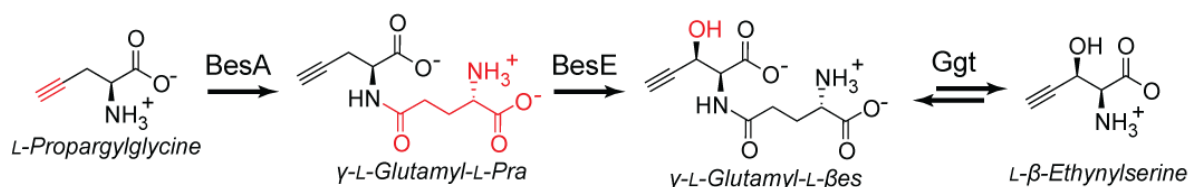
Figure 4.2. Proposed pathway for hydrolysis of Glu-βes didpeptide. Final product generated by the βes pathway cluster is Glu-βes, formed by BesE. Ggt, found in all organisms, can catalyze the reversible transfer of the γ-glutamyl moiety to another amino acid, liberating βes.

We first tested this hypothesis *in vitro* using purified BesA, BesE, and commercially available Ggt from equine kidney. To monitor formation of βes, we used CuAAC to attach a fluorophore (3-N$_3$-7-OH-Coumarin) to βes and detected product formation by LC/MS (*Figure 4.3A*). In reactions containing Glu, Pra, ATP, MgSO$_4$, and HEPES (pH 7.5) we were unable to observe formation of any free βes in the absence of Ggt, even after allowing the reaction run an extended

period of time (24 h) (*Figure 4.3B*). However, if we included 1 U/mL of Ggt, we observed formation of Cou-βes. These results suggest that neither BesA nor BesE possess any appreciable background hydrolytic activity for the dipetide and also that dipeptide formation is reasonably stable at neutral pH. In these reactions, we believe that free Glu or Pra can serve as the acceptor amino acid required for Ggt-catalyzed hydrolysis of Glu-βes.
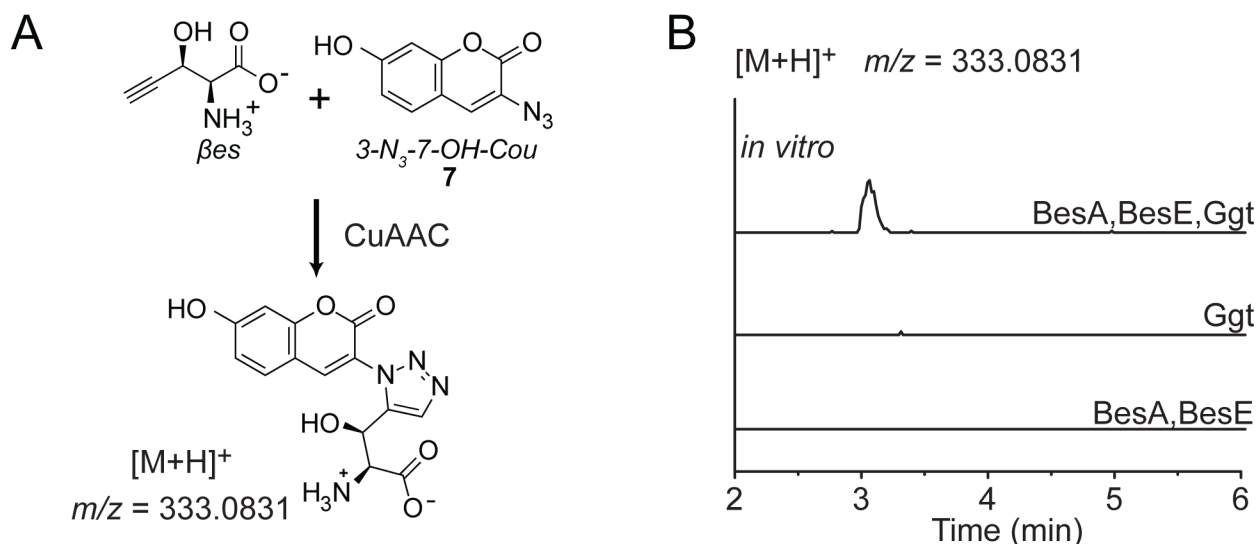
Figure 4.3. (A) CuAAC is used to derivatize free βes with 3-N$_3$-7-OH-Coumarin to form Cou-βes (*m/z* = 333.0831). Performing this reaction allows for easier detection of βes by reverse phase LC/MS compared to the free amino acid. (B) *In vitro* reactions for BesA, BesE, and Ggt with Glu, Pra, and necessary co-factors. Free βes is not observed in negative control reactions containing only BesA, BesE or Ggt. Only when all three enzymes (BesA, BesE and Ggt) are present can the Cou-βes product be detected, indicating that Glu-βes is being hydrolyzed to βes under these conditions.

Production of βes from Pra *in vivo*. Having identified that Ggt is competent to hydrolyze Glu-βes to βes *in vitro*, we wanted to examine if the same holds true *in vivo*. We decided to use *E. coli* BL21 Star (DE3) as our model host since both BesA and BesE have been shown to express relatively well in this host. Furthermore, we were able to identify a previously studied Ggt carried by *E. coli* (ECK3431) that is known to be constitutively expressed.

Cultures of *E. coli* BL21 Star (DE3) expressing BesA and BesE (pETDUET-BesA.BesE) were fed 1 mM Pra, resulting in moderate production of free βes as determined by CuAAC LC/MS (*Figure 4.4*). We used polar metabolite analysis on LC/MS to detect Glu-Pra/Glu-βes, and found that both dipeptides were present at high levels in the fermentation broth, suggesting that only a small portion of Glu-βes was being hydrolyzed (*data not shown*). To increase hydrolysis of Glu-βes, we then overexpressed Ggt (using expression vector pGgt). In *E. coli* BL21 Star (DE3) cultures expressing BesA, BesE, and Ggt (pETDUET-BesA.BesE and pGgt) we observed almost complete hydrolysis of Glu-βes dipeptide to βes (*Figure 4.4*). These results were not necessarily expected, since Ggt catalyzes a thermodynamically reversible step and one could imagine that overexpression of Ggt leads to a net-formation of Glu-βes dipeptide. Nonetheless, these experiments provided evidence that Ggt can successfully hydrolyze Glu-Bes *in vivo*, and can be used for the *in vivo* formation of βes from Pra in *E. coli* (requiring only the overexpression of BesA, BesE, and Ggt).

140
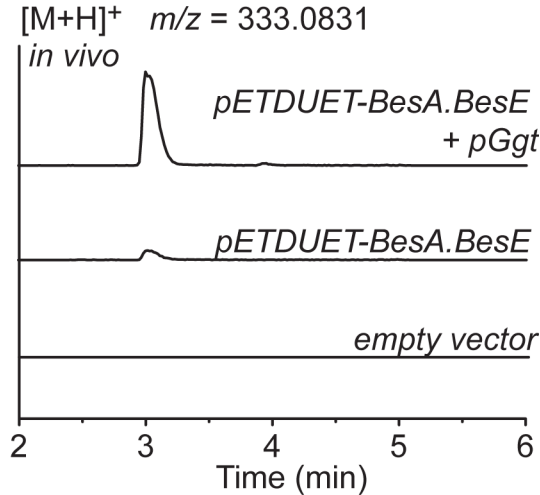
Figure 4.4. *In vivo* reactions for BesA, BesE, and Ggt with Glu, Pra, and necessary co-factors. Free βes is not observed in reactions with only BesA, BesE or Ggt. When all three of BesA, BesE and Ggt are present, we can detect Cou-βes product indicating that Glu-βes is being hydrolyzed to βes.

Full pathway reconstitution: production of βes from lysine *in vivo*. Taken together with results presented previously, we set out to reconstitute the complete biosynthesis of βes from L-lysine and construct a single plasmid, pβes containing all necessary genes for this transformation (*Figure 4.5*). When designing pβes, we attempted to make informed decisions for design considerations. pβes needed to be capable of expressing six proteins: BesA, BesB, BesC, BesD, BesE, and a Ggt. For the first three steps (BesC, BesD, BesB), we took the design from previously optimized pPra. Genes for BesC and BesD were therefore the homologues from *Pseudomonas fluorescencs*, while the BesB homologue from *S. sp*. NRRL S-1448 was chosen. Furthermore, we decided to express BesB as an MBP-fusion since we had previously observed that an N$_T$-MBP improves protein yield of BesB, likely through improvement in protein solubility. Each of these three genes is transcribed from a T7 promoter with a single T7 terminator after BesB. For BesA and BesE, we chose to use homologues from *S. cattleya*. Since we are expressing this pathway in *E. coli*, we selected the Ggt from *E. coli* as we had previously observed overexpression of *E. coli* Ggt is sufficient for Glu-βes hydrolysis. To decrease the transcriptional stress of expressing all 6 genes, *besE*, *besA*, and *ggt* were designed as an operon under the control of a single T7 promoter and T7 terminator. Ribosome binding sites (RBS) were designed computationally (RBS Calculator [25]) for *besA* and *ggt*, while the remaining RBS were derived from their parent vectors.
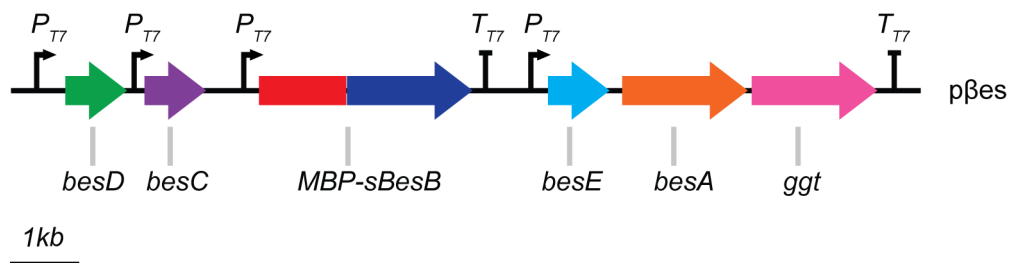
Figure 4.5. Gene organization and design of pβes. This plasmid is used for the complete biosynthesis of βes from lysine in *E. coli*.

In our production run with pβes we sought to assess if chaperone co-expression would improve βes production given the solubility issues with BesB in *E. coli*. Based on results discussed in Chapter 3, we utilized a plasmid, pGro7, which contains two *E. coli* chaperonins, GroEL and GroES. We also developed an alternative method for detecting βes by LC/QQQ by CuAAC to attach N_3-PEG$_3$-Biotin to βes. The resulting PEG$_3$-Biotin products exhibited characteristic product ion (*m/z* = [Parent]→227.1), that enabled detection of Pra-PEG$_3$-Biotin and βes-PEG$_3$-Biotin by parent ion scanning (*Figure 4.6*). Using these methods, we could monitor production of βes (*Figure 4.7*) when expressing pβes + pGro7 in *E. coli* BL21 Star (DE3) grown in M9 media supplemented only with standard set of 20 amino acids (≈100 mg/L).
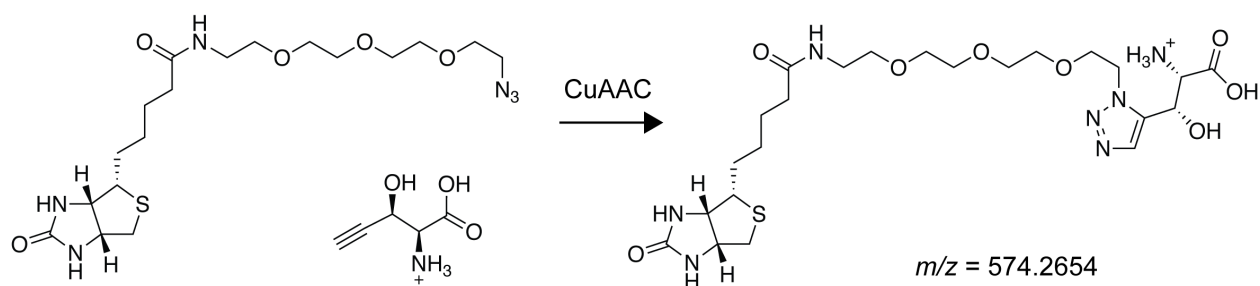
Figure 4.6. Derivatization of βes with N_3-PEG$_3$-Biotin to form βes-PEG$_3$-Biotin using CuAAC. PEG$_3$-Biotin products can be detected by LC/MS-MS by parent ion monitoring for *m/z* = [Parent]→222.1 transition.
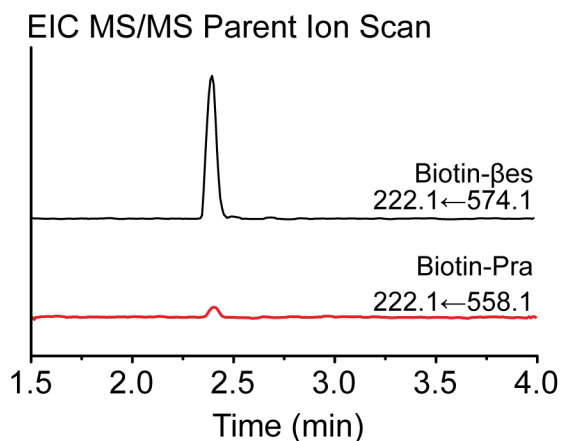
Figure 4.7. Biosynthesis of βes from lysine in *E. coli* BL21 Star (DE3) pβes pGro7. Mass spectra shown are parent ion monitoring at the *m/z* = [Parent] →222.1 transition which gives two signatures. Pra-PEG$_3$-Biotin appears as a *m/z* = 558.1→222.1 transition while *m/z* = βes-PEG$_3$-Biotin appears as the 574.1 → 222.1 transition.

Incorporation of βes into proteins *in vitro*. *S. cattleya* produces two threonine-like amino acids as secondary metabolites: F-Thr and βes. Previously, our lab investigated F-Thr aminoacylation and translation, and found that the native threonyl-tRNA synthetases (TRS) from *S. cattleya* appear to charge both Thr and F-Thr with similar efficiency onto tRNA^{Thr} [19]. Furthermore, *S. coelicolor* and *S. cattleya ΔfthB* fed F-Thr were observed to incorporate F-Thr into their proteins, presumably through mistranslation caused by the presence of FThr-tRNA^{Thr} in the cell. To test if βes could be charged and translated through TRS/tRNA^{Thr}, we performed *in vitro* transcription/translation (IVTT) reactions of a short peptide with different mixtures of amino acids (*Figure 4.8*). When only the standard 20 amino acids were present, we could detect formation of METRSKML (shown as the deformylated species) by LC/MS QTOF (*Figure 4.9*). When βes was included and Thr excluded, we could detect formation of a new peptide MEβRSKML (β=βes residue), but failed to observe formation of the Thr containing peptide (*Figure 4.10*). CuAAC was used to derivatize the MEβRSKML peptide with 3-N$_3$-7-OH-Coumarin resulting in formation of ME(β-Cou)RSKML, which we were able to detect LC/QTOF (*Figure 4.11*). Taken together, these experiments conclusively show that aminoacylation and translation of βes is possible through TRS/tRNA^{Thr}. Next, we investigated translation of βes *in vivo*.
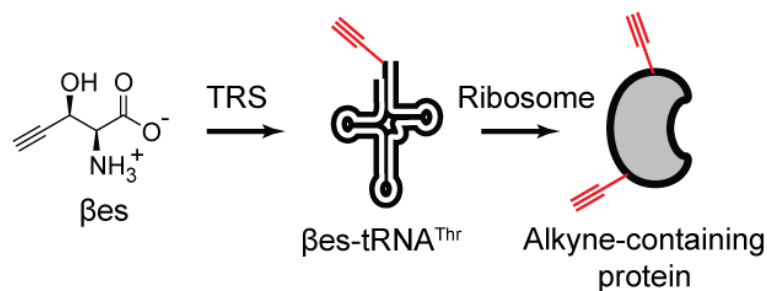
Figure 4.8. Proposed route of βes incorporation into proteins. βes is charged onto tRNA^{Thr} by TRS. βes-tRNA^{Thr} can be used in translation, incorporating βes in place of a Thr residue.

Incorporation of βes into proteins *in vivo*. The ability to direct biosynthesis of βes in *E. coli* allowed us to test a variety of hypotheses regarding βes behavior *in vivo*. *E. coli* BL21 Star (DE3) harboring pEAGgt or pETDUET-BesABesE + pGgt can convert >50% of fed Pra to βes, leaving behind only low amounts of unhydrolyzed Glu-βes dipeptide. Since βes is not a commercially available amino acid, *in vivo* experiments described henceforth utilize pathways that overexpresses BesA, BesE, and Ggt for βes generation *in situ* (*Figure 4.2, Figure 4.12*).

Our first experiment aimed to test if βes is incorporated into proteins *in vivo* and whether any observed incorporation likely occurs through translation with βes-tRNA^{Thr}. *E. coli* BL21 Star (DE3) harboring pETDUET-BesA.BesE, either with or without pGgt, was grown in M9 media containing 1 mM Pra. Since we hypothesized that βes incorporation likely occurs through tRNA^{Thr}, we then tested the impact of feeding these strains additional Thr (1 mM). If βes is incorporated by mistranslation of Thr, the addition of Thr should compete with βes incorporation, resulting in lower levels of mistranslation. To detect βes incorporation, we used a strategy previously described in Chapter 3, whereby lysates of cells are derivatized with a fluorescent azide (Tamra-Azide) by CuAAC, ran on SDS-PAGE gel, and then imaged by fluorescence. As a positive control for proteome alkyne fluorescent labeling, we grew *E. coli* B834 (DE3) cells harboring pBAD24-PraRS in the presence of 0.2 mM Pra. As a negative control for labeling, Cu was omitted from the CuAAC reaction since it is required for catalysis.

METRSKNML

EIC + m/z = 555.2757

-βes|+Thr

-βes|-Thr

+βes|-Thr

Time (min)
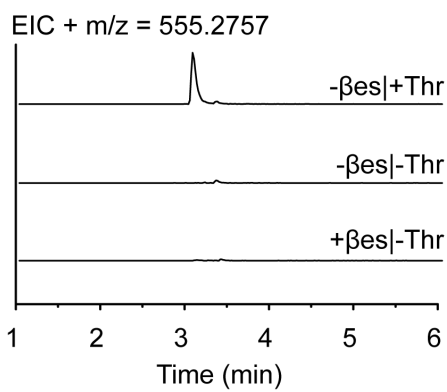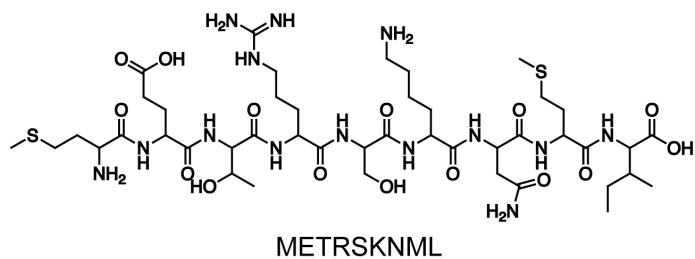
Figure 4.9. EICs for IVTT reactions of the METRSKNML peptide. The METRSKNML peptide product ([M+2H]$^{2+}$ *m/z* = 555.2757 *deformylated*) is only observed when Thr is present.

MEβRSKNML

EIC + m/z = 560.2679

-βes|+Thr

-βes|-Thr

+βes|-Thr

Time (min)
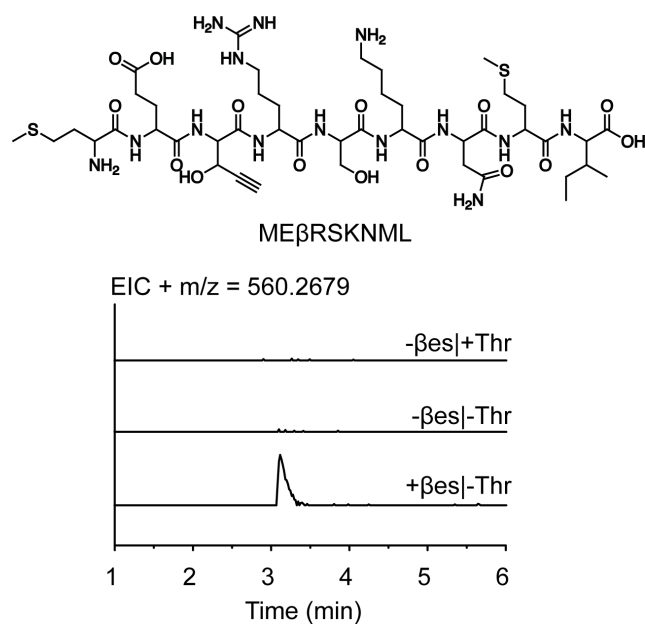
Figure 4.10. EICs for IVTT reactions of the METRSKNML peptide. (A) The MEβRSKNML peptide product ([M+2H]$^{2+}$ *m/z* = 560.2679 *deformylated*, β = βes) is only observed when βes is present.
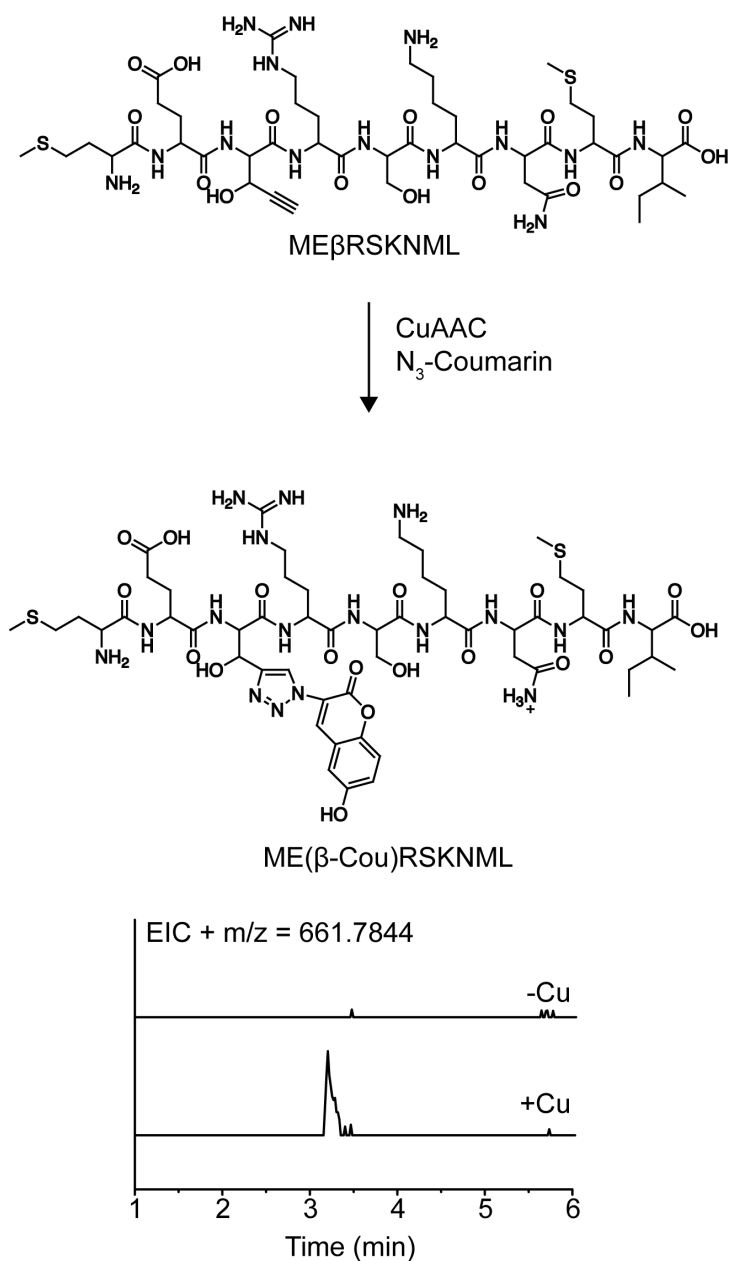
Figure 4.11. EIC for the MEβRSKNML peptide after derivation by CuAAC. The MEβRSKNML peptide product of the IVTT reaction containing βes was derivatized by CuAAc with 3-N$_3$-7-OH-Coumarin. The EIC for the resulting peptide is shown (ME(β-Cou)RSKNML; [M+2H]$^{2+}$ m/z = 661.7844 *deformylated*, β-Cou = βes residue derivatized as shown). The -Cu condition shown as negative control, since Cu is required for CuAAC.
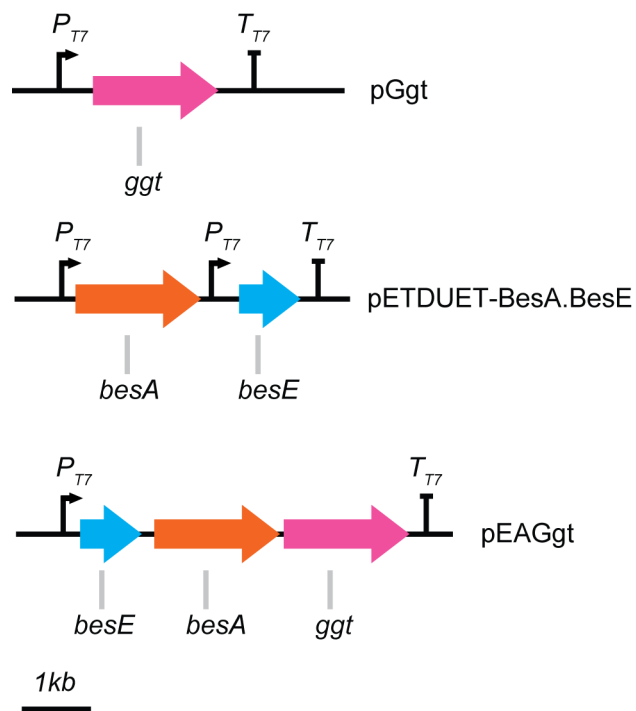
Figure 4.12. Gene, promoter, and terminator organization for pathways used in conversion of Pra to βes in *E. coli*. pGgt and pETDUET-BesA.BesE are on two different antibiotic selection markers (Km, Cb, respectively), and are therefore used concurrently. pEAGgt was designed to be express BesE, BesA, and Ggt as a single operon.

147

In all conditions, alkyne incorporation into the proteome was observed to some degree (*Figure 4.13*). The highest level of labeling occurred in cultures of *E. coli* BL21 Star (DE3) pETDUET-BesA.BesE pGg in the absence of Thr (*lane 5*). When Thr was added (lane 7), fluorescent labeling decreased suggesting that the addition of Thr does indeed lower the incorporation of βes into the proteome. In the strain lacking the glutamyl-transpeptidase (pGgt), lower labeling was also observed, which may indicate that lower levels of βes are available in this strain. Taken together, these results support that hypothesis that Ggt promotes incorporation *in vivo*, likely by increasing hydrolysis of Glu-βes and availability of free βes, as well as βes mistranslation occurring through tRNA[Thr]. However, we wanted to confirm the identity of the alkyne incorporated in these studies as Pra, Glu-Pra, or Glu-βes are all present and could potentially be incorporated into proteins.
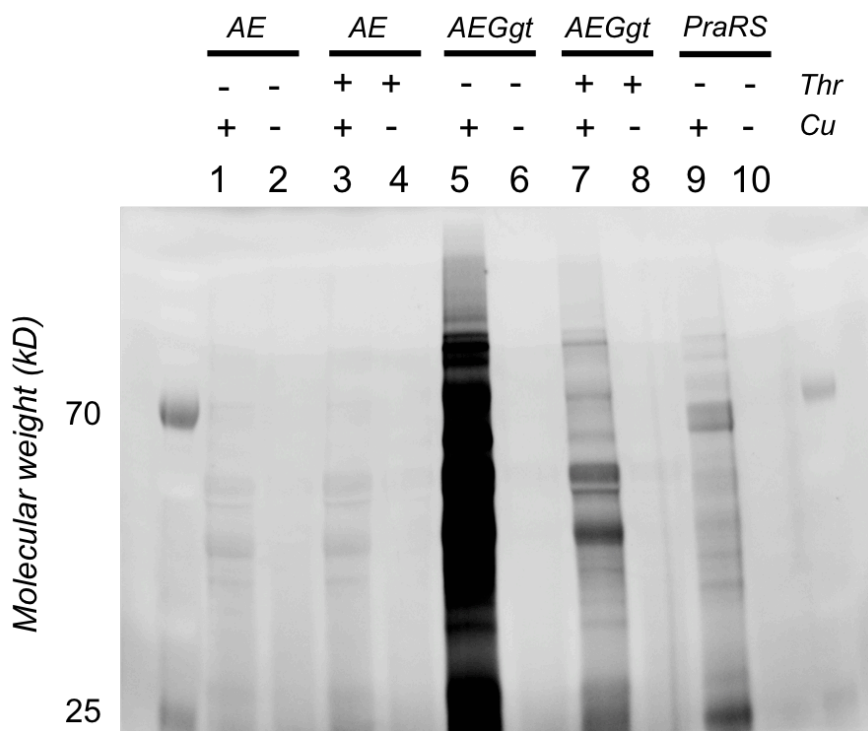
Figure 4.13. Incorporation of βes into proteome of *E. coli* visualized by fluorescence SDS-PAGE gel. *E. coli* BL21 Star (DE3) with either pETDUET-BesA.BesE (AE) or pETDUET-BesA.BesE and pGgt (AEGgt) were grown in media containing 1 mM Pra for 48 h. Lysates of cultures were derivatized with Tamra-Azide by CuAAC and visualized using fluorescence imaging (λ_{Ex} = 546 nm, λ_{Em} = 565 nm). Only faint incorporation is observed in AE conditions. Addition of Ggt in the AEGgt condition shows significant amount of βes incorporation. Less incorporation is seen in AEGgt + Thr condition, likely to competition with Thr for translational machinery. PraRS condition is shown as a positive control for labeling, which we previously validated as a method for incorporation of Pra into proteome (Met replacement). The -Cu conditions are negative controls for CuAAC, since Cu is required for CuAAC catalysis.

To directly detect which amino acid was incorporated into the proteome, we turned to top-down proteomics. Spectral hits of lysate samples derived from *E. coli* BL21 Star (DE3) pETDUET-BesA.BesE pGgt fed with Pra, but not Thr, were compared to the spectra of the same strain that was not fed with Pra. In our database search, we included a post-translational modification of Thr + 10 which corresponds to the difference in mass between Thr and βes. One peptide fragmentation spectrum is shown for peroxiredoxin, which had high degree of coverage sequence coverage for both βes and Thr (*Figure 4.14 & 4.15*). Not surprisingly, one of the most abundant βes-containing peptides was derived from Ggt, which was overexpressed in this experiment (*Table 4.2*). Having established that βes is indeed incorporated into proteins *in vivo* in the Thr site, we turned our attention to elucidating the potential role of BesG in modulating βes mistranslation in the host.
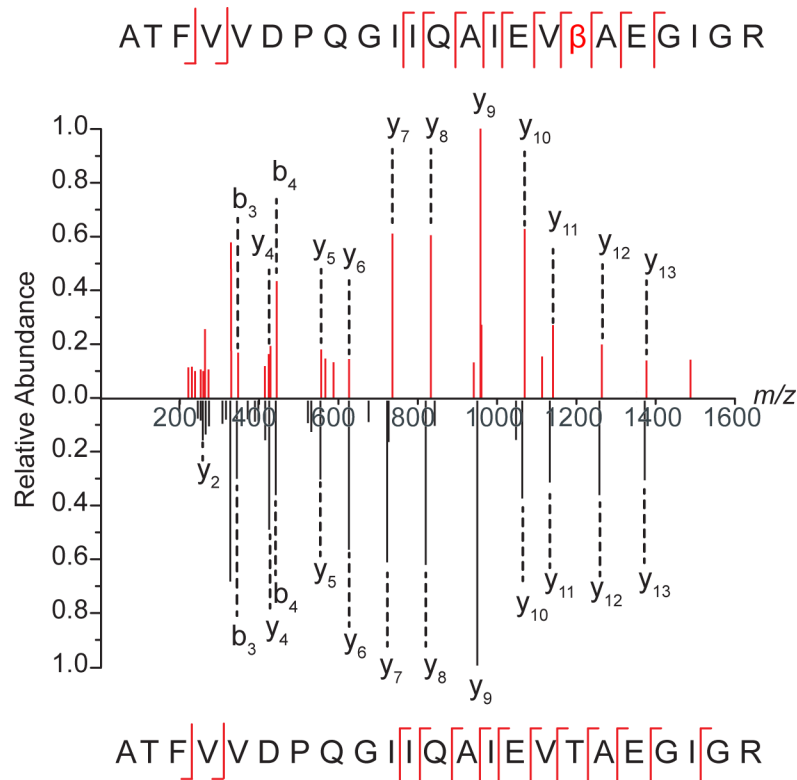
Figure 4.14. Representative high-resolution tandem MS/MS spectra of peptide for *E. coli* peroxiredoxin (ECBD_3047) with or without βes residue. The spectrum above the axis gives Peptide-Spectrum Matches (PSMs) to sequences with βes substituted for Thr while the spectrum below the axis shows the corresponding Thr-containing native peptide. Sequence above and below demark observed y-ions and b-ions observed. The βes-containing spectra and ions were not observed in sequencing of negative control sample.

A-T-F-V-V-D-P-Q-G-I-I-Q-A-I-E-V-β-A-E-G-I-G-R

y-ion/b-ion

175.1190/2220.1546
232.1404/2163.1332
345.2245/2050.0491
402.2459/1993.0276
531.2285/1863.9850
602.3257/1792.9479
713.3576/1681.9159
812.4260/1582.8475
941.4686/1453.8049
1054.5527/1340.7209
1125.5898/1269.6838
1253.6484/1141.6252
1366.7325/1028.5411
1479.8165/915.4571
1536.8380/858.4356
1664.8966/730.3770
1761.9493/633.3243
1876.9763/518.2973
1976.0447/419.2289
2075.1131/320.1605
2222.1815/173.0921
2323.2291/72.0444

y-ion/b-ion

175.1190/2210.1703
232.1404/2153.1489
345.2245/2040.0648
402.2459/1983.0433
531.2285/1854.0007
602.3257/1782.9636
703.3733/1681.9159
802.4417/1582.8475
931.4843/1453.8049
1044.5684/1340.7209
1115.6055/1269.6838
1243.6641/1141.6252
1356.7482/1028.5411
1469.8322/915.4571
1526.8537/858.4356
1654.9123/730.3770
1751.9650/633.3243
1866.9220/518.2973
1966.0604/419.2289
2065.1288/320.1605
2212.1972/173.0921
2313.2449/72.0444

A-T-F-V-V-D-P-Q-G-I-I-Q-A-I-E-V-T-A-E-G-I-G-R

Figure 4.15. Ion table for representative high-resolution tandem MS/MS spectra of peptide for *E. coli* peroxiredoxin (ECBD_3047) with (top) or without (bottom) βes residue. Both y-ions and b-ions shown. Ion masses highlighted in red were observed, and show a clear mass shift corresponding to a Thr → βes substitution.

150

Table 4.2. Table of top peptide hits for peptides containing T+10 modification (βes residue) with protein description and gene ID shown. Potential residues where βes could be incorporated is highlighted in red (**β**). Spectra were filtered using COMET search settings: charge states > +3 excluded; min xcorr = 2.5; min delta correlation value = 0.1; max Sp rank = 1; must contain T+10 modification; p-value < 0.05. Using these settings resulted in 0 matches in negative control and 0 decoys.

	Protein annotation	Gene ID
K.SETIA**β**LLAEDET**β**LDENELGEILHAIEK.N	Polyribonucleotide Nucleotidyltransferase	*pnp*
K.NASLGEMITNLSGMGVSVPNGFA**ββ**ADAFNQFLDQ SGVNQR.L	Phosphoenolpyruvate synthase	*ECBD_1944*
K.VINDNFGIIEGLMTTVHATTA**β**QK.T	G3P dehydrogenase	*ECBD_1865*
K.SLEMIAENGPDEFYKG**β**IAEQIAQEMQK.N	γ-Glutamyl-transpeptidase	*ggt*
R.NYGVVYDALGY**β**DMLPEFGGD**β**AYSDDFFVGR.V	Porin (gram negative)	*ECBD_2666*
R.ATFVVDPQGIIQAIEV**β**AEGIGR.D	Peroxiredoxin	*ECBD_3047*
R.LVNSVPGLSDV**β**SSFAMEQIKYTTSLPIE	Transcriptional regulator	*ECBD_3208*
R.AVVVIDENDNVIFSQLVDEI**ββ**EPDYEAALAVLKA	Thiol peroxidase	*tpx*
K.GMN**β**AVGDEGGYAPNLGSNAEALAVIAEAVK.A	Enolase	*eno*
K.GFGFI**β**PDDGSK.D	DNA-binding domain protein	*ECBD_0180*
K.AVPA**β**KKNIEFFEAR.R	50S ribosomal protein	*rplI*

Discovery of BesG. In our initial discovery work on the βes pathway, we had identified only three species of streptomycetes (*S. cattleya*, *S. sp*. NRRL S-1448, and *S. alboverticillatus*) that possessed the metabolic genes required for Glu-βes biosynthesis (*besA-E*, *Table 4.3*). Among the gene clusters found in these organisms, additional auxiliary genes related to βes control and regulation were identified. One of these genes, *besF*, encodes for a putative amino acid efflux pump, and was not found in streptomycetes that only harbored *besA-D* (Pra biosynthesis). Interestingly, *S. sp*. NRRL S-1448 and *S. alboverticillatus* contained an additional gene found in *bes* cluster, *besG*, which encodes a putative threonyl/alanyl tRNA-synthetase *trans*-editing domain. Searching for instances of homologues of *besG* in *S. cattleya*, we were able to identify a gene, SCAT_p1133, that was a close match to the *besG* found in *S. sp*. NRRL S-1448 and *S. alboverticillatus bes* gene clusters (*Figure 4.16*). Though prokaryotes tend to cluster genes related to a common biosynthetic function, *besG* in *S. cattleya* (SCAT_p1133) is 84 ORFs downstream from *besE* (SCAT_p1049). Nonetheless, we decided to biochemically characterize *besG* from *S. cattleya* since the clustering in other hosts suggested a function in βes metabolism. We propose that BesG is a βes-tRNA^{Thr} deacylase, with function analogous to FthB in the FThr cluster, which was found to assist in preventing FThr mistranslation [19]. This function would be consistent with the function of similar proteins found in gene clusters that produce amino acids (F-Thr and Cl-Thr) that are structurally similar to threonine. Interestingly in *S. cattleya*, FthB and other genes involved in fluorothreonine metabolism are clustered separately from the related fluoroacetate biosynthetic gene cluster despite existing as a full fluoroacetate/fluorothreonine gene cluster in other streptomycetes.

Unlike FthB, a YbaK-like *trans*-editing enzyme, BesG is predicted to be a AlaX family *trans*-editing enzyme. Within the AlaX family, BesG is most similar to AlaX-M (M = medium) since it lacks an oligomerization domain seen in AlaX-L (L = large) but contains the N_T region not found in AlaX-S (S = small) proteins (*Figure 4.17*) [14,26–28]. A homology model of BesG generated using Phyre2 reveals conservation of characteristic motifs found in other AlaX-M proteins (*Figures 4.18 & 4.19*). Among them, a glycine-rich loop (G44-46) is present and is thought to

Table 4.3. *Streptomyces* spp. that contain βes clusters with *besE* (required to make βes) also possess the auxiliary genes *besF* and *besG*.

Host	besG	besE	besF	besACBD
S. cattleya	+	+	+	+
S. sp. NRRL S-1448	+	+	+	+
S. alboverticillatus	+	+	+	+
S. catenulae	-	-	-	+
S. achromogenes	-	-	-	+
S. sp. NRRL S-31	-	-	-	+
S. lavenduligriseus	-	-	-	+
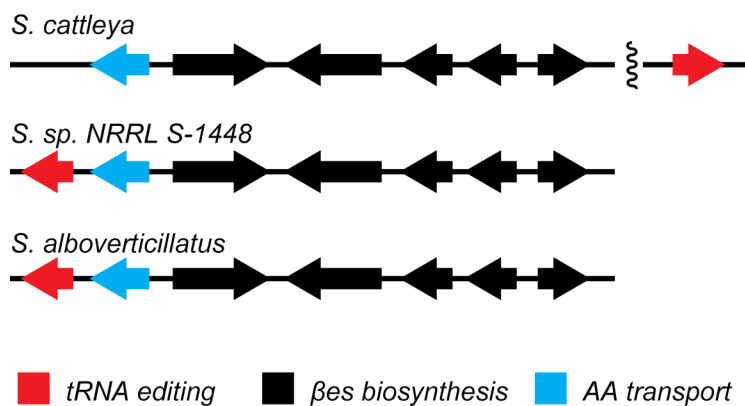
Figure 4.16. *S. sp. NRRL S-1448* and *S. alboverticillatus* have *besG* (red, tRNA editing protein) clustered with *besF* (blue, amino acid transporter) and the *bes* gene cluster (black, *besA-E*). *S. cattleya* possesses a homologue of *besG* that is 84 genetic loci downstream from the *bes* gene cluster.

```
AlaX-S     ------------------------------------------------------------ 0
BesG       -MTQQRQIYLTDTYAYEASTRVVSAERGPDTVRIALADNIFHPQGGGQPDDRGWVDQVEV 56
AlaX-M     MINMTRKLYYEDAYLKEAKGRVLEIRDN----AILLDQTIFYPTGGGQPHDRGTINGVEV 56
AlaX-L     ---MVERVYYKDPYLREIDAKIVDVRKEKDRVEVVLDRTIFYPEGGGQPGDRGVIKGNDF 57

AlaX-S     ------------------------------------------MYSIEVRTHSALHVVKG 17
BesG       RPVRDTG----PGLVHLTCPADAAWPREAVAVGTTVTSRIDPALRRLHAALHTAGHLVDA 115
AlaX-M     LDVYK---DEEGNVWHVV-----KEP-EKFKVGDEVELKIDWDYRYKLMRIHTGLHLLEH 107
AlaX-L     EIMVEDTIEKNGEIFHIG-----KLRGEIPKEGEKVKLYLDWEWRYGNMQMHTGQHILSA 112

AlaX-S     AVVKVLGSEAKW--TYSTYVKGNKGVLIVKFDRKPSDEEIREIERLANEKVKENAP--IK 73
BesG       LIDEW-GYRHVGSNHFPGQARVE-------YD--------------------LGGLDC 145
AlaX-M     VLNEVL-----------------GEGNWQLVGS------GMSVEKGRYDIAYPENLN 141
AlaX-L     VLKKLYDLDTTGFNIFRDYAKIEVN-GEVNWDMIER------AELEVNKIIINDLPVVIE 165

AlaX-S     IYELPREEAEKMFGEDMYDLFPV-PEDVRILKVVVIEDWNVNACNKEHTKTTGEIGPIKI 132
BesG       DKDALAEGLTERLGKPLAEALPVIPGERDGRRTITIEGYATEFCAGTHVPDLSLLTGVAI 205
AlaX-M     KYK---EQIISLFNKYVDEGGEVKIWWEGDRRYTQIRDFEVIPCGGTHVKDIKEIGHIKK 198
AlaX-L     EYDKLPEEIALKLRKNV-------TKVKEKIRIVKIGDVDVTPCGGTHVKSTREVGIIKV 218

AlaX-S     RKVRFRKSKGLLEIHFELLELENPS----------------------------------- 157
BesG       RS--VKVKGGRLKVGYTAEHVPVDG----------------------------------- 228
AlaX-M     LKR--SSIGRGKQRLEMWLE--------------------------------------- 216
AlaX-L     LRF-YKKSKNLWRIEFTCGYRAISKMNEILRDYWGSLDLMPNKNPPLIERINEVLRTVNS 277

AlaX-S     ----------------------------------------------------------- 157
BesG       ----------------------------------------------------------- 228
AlaX-M     ----------------------------------------------------------- 216
AlaX-L     LENRIEELRREIWEWKGKALLKEGVKVGNYTVITHVENWNMKDAQAFAIDFVKKNSNTIL 337

AlaX-S     ----------------------------------------------------------- 157
BesG       ----------------------------------------------------------- 228
AlaX-M     ----------------------------------------------------------- 216
AlaX-L     LLANEKYVLFAKNEGVPVSMRELLKEVIDELGGKGGGTDNLARGRVEAKPEEIFDVALEK 397

AlaX-S     ------- 157
BesG       ------- 228
AlaX-M     ------- 216
AlaX-L     LRSHLQV 404
```

Figure 4.17. Protein sequence alignment of BesG (from *S. cattleya*) compared to AlaX-S (O58307.1), AlaX-M (O57848.1), and AlaX-L (O57734.1) from *Pyrococcus horikoshii*. Glycine-rich loop is highlighted in cyan, while putative zinc-binding residues are highlighted in yellow.
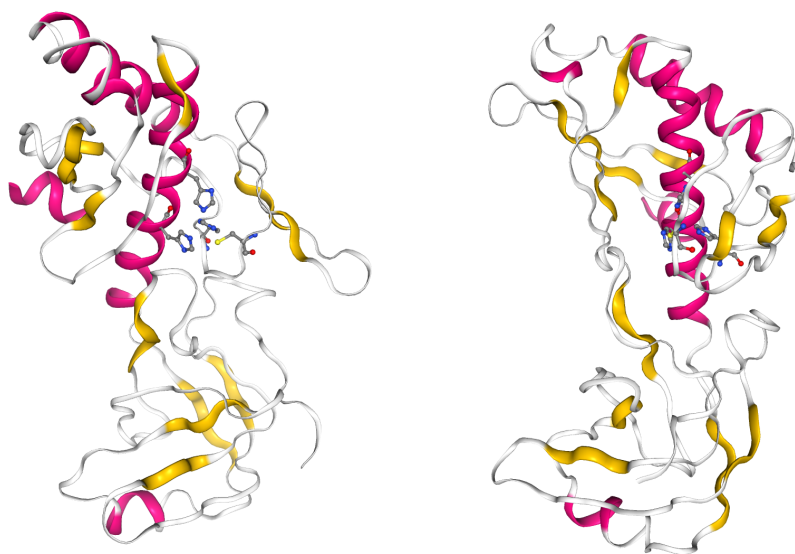
Figure 4.18. Homology model of BesG shown at two angles (90 degree roration). Colors highlight secondary structure (pink = α-helix; yellow = β-sheet). Residues thought to be involved with zinc coordination shown as ball and stick (H107, H111, C189, H193).
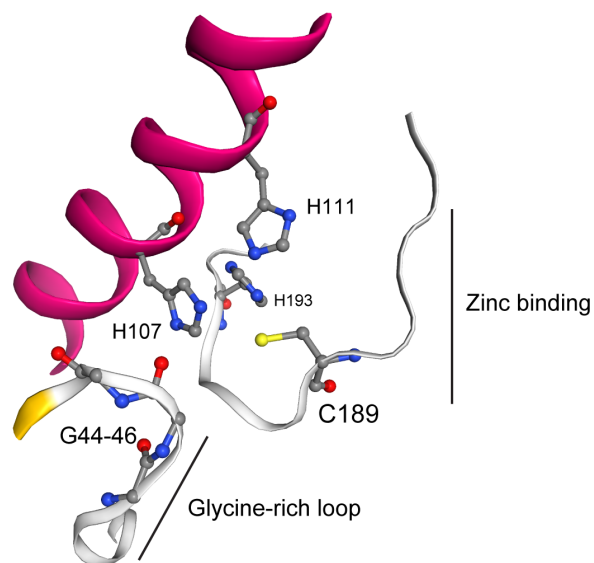
Figure 4.19. Zoom in of BesG homology model showing conserved structural features of AlaX-M protein. Glycine-rich loop (G44-46) is spatially positioned adjacent to putative active site and zinc binding residues (H107, H111, C189, H193).

be involved in substrate recognition [26]. In addition, the zinc-binding motif is also present (H107, H111, C189, H193). From structural studies on the AlaX-M from *Pyrococcus horikoshii* with computationally docked misacylated-tRNA substrate, it has been inferred that tRNA recognition might be involved in substrate recognition [26]. However, no structures are available to date of AlaX-M protein with both cognate and non-cognate aminoacyl-tRNAs.

Chemical synthesis of β-ethynylserine. Before performing *in vitro* work with BesG, we required a synthetic standard of βes. An asymmetric synthesis of a βes derivative was first reported by Zhang and van der Donk [29]. In this synthesis, Garner's aldehyde [30,31] was used as a chiral starting material, with the alkyne moiety being installed with the requisite *threo* selectivity using a copper-mediated addition of ethynyltrimethylsilane developed previously by Herold *et al.* [32]. In seven additional steps, the Herold adduct was converted to an Fmoc/*t*-butyl protected βes for use in solid phase peptide synthesis. The van der Donk strategy was adopted by Chan and O'Hagan [33], who sought to obtain free βes hydrochloride. To do so, the secondary alcohol of the Herold adduct was initially protected as phenoxyacetyl ester. Due to issues with ester migration during TFA deprotection of the N,O-acetonide, an acid-stable TBDPS ether was installed instead, necessitating two additional steps to synthesize the non-commercially available TBDPS-OTf. Subsequent TFA deprotection resulted in cleavage of both the acetonide and the Boc group, so the Boc group was reattached in the next step. Jones oxidation of the primary alcohol followed by sequential basic and acidic deprotections provided ß-ethynylserine hydrochloride in 8 steps (including the preparation of TBDPS-OTf) at 60% yield from the Herold adduct (*Appendix Figure A2.4*).

We used the strategy of Chan and O'Hagan with slight modifications, aimed towards optimizing the selective deprotection of the N,O-acetonide. Commercially available N-Boc-D-serine methyl ester was converted sequentially to Garner's ester, Garner's aldehyde, and the *threo*-Herold adduct by known methods [29–33]. The secondary alcohol was then easily protected as the acetyl ester, and various conditions for acetonide deprotection were screened. Standard conditions with *p*-toluenesulfonic acid or Amberlyst-15 resin resulted only in complex mixtures, while mild Lewis acid catalysts such as $CuCl_2$ or $Zn(NO_3)_2$ resulted in no reaction. The condition of heating in 3:1 acetic acid/water at 50 °C was finally found to provide the desired product in 49% yield. Similar to what was reported before, the migration of the acetyl group to the primary position was observed, and this minor product was also isolated and characterized. Hydrolysis at a higher temperature than 50 °C leads to substantial ester migration. Despite the modest yield, this sequence is operationally simple and leaves the Boc group intact. Jones oxidation followed by basic and acidic deprotections provided ß-ethynylserine hydrochloride in 5 steps and 34% yield from the Herold adduct.

BesG is a βes-tRNA^{Thr} hydrolase. We set about determining if BesG can deacylate βes-tRNA^{Thr} *in vitro* (*Figure 4.20A*). We used *in vitro* translation to generate high concentration of tRNA^{Thr} from *S. cattleya*. Though four isotypes of tRNA^{Thr} are available, we chose a single isotype (tRNA-25) for characterization. We then charged translated-tRNA with βes or Thr *in vitro* using TRS from *S. cattleya*, to generate βes-tRNA^{Thr} and Thr-tRNA^{Thr}. To monitor deacylation of aminoacyl-tRNA^{Thr}, we adapted an assay previously developed by our group. Briefly, aminoacyl-tRNA is fully hydrolyzed by addition of RNAse, releasing the 3′-aminoacyl adenylate which can be directly detected by LC/MS-MS. We can then compare the initial amount of aminoacyl-tRNA and the amount remaining after incubation with BesG. 2 μM of
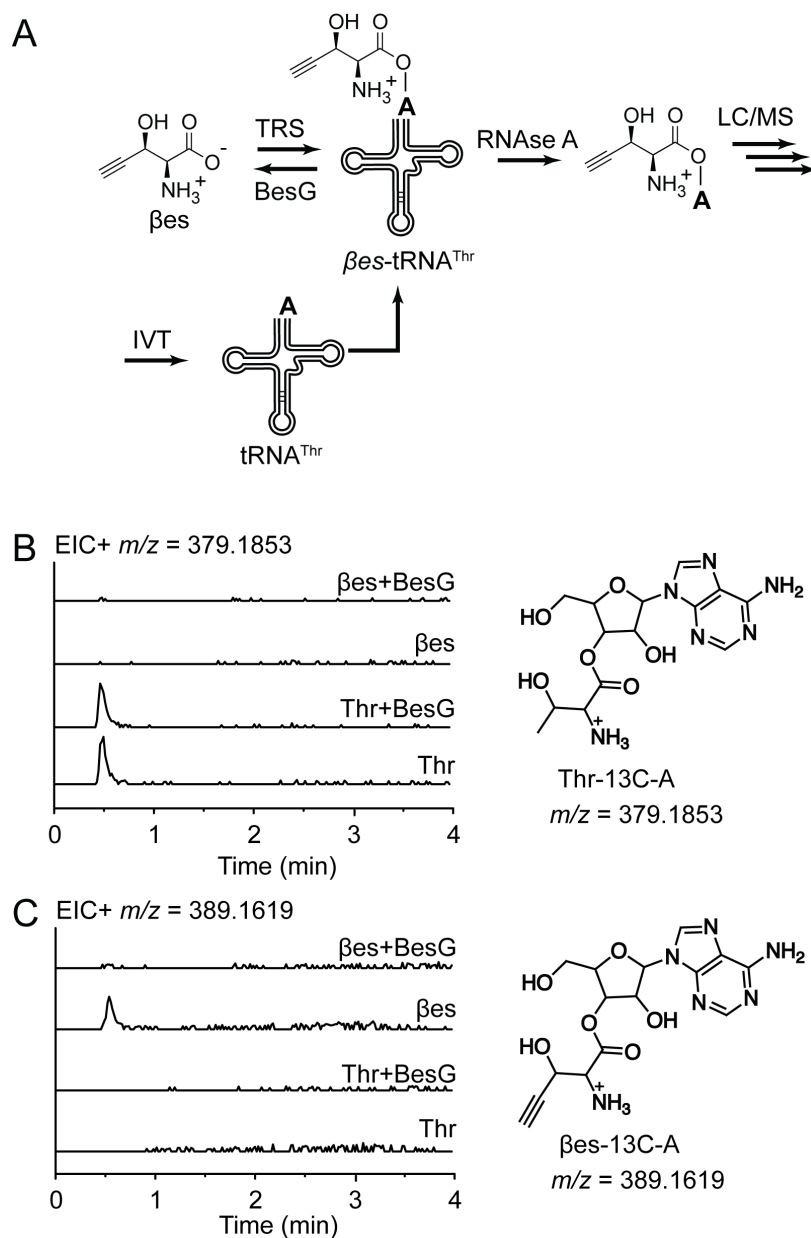
Figure 4.20. (A) Assay for detection of amino-adenylates. IVT is used to generate tRNA[Thr] which can then be charged by TRS to generate βes-tRNA[Thr] and Thr-tRNA[Thr]. Example for generating βes-tRNA[Thr] shown. Deacylation is initiated by adding BesG (50:1 BesG to TRS). Reactions are quenched by addition of 1 U/mL of RNAseA, which releases amino-adenylate product that can then be detected directly by LC/MS-MS. (B) Extract ion chromatogram monitoring for formation of Thr-[13C]-adenosine (Thr-A; [M+H]$^+$ m/z = 379.1853) shows that Thr-A only forms when Thr is present, and amount of Thr-A does not significantly decrease over course of reaction. (C) Extract ion chromatogram monitoring for formation of βes-[13C]-adenosine (βes-A; [M+H]$^+$ m/z = 389.1619) shows that βes-A only forms when βes is present, and is completely depleted after addition of BesG. These results suggest BesG can effectively hydrolyze βes-tRNA[Thr] but not Thr-tRNA[Thr].

BesG was able to mostly hydrolyze approximately 10 µM of βes-tRNA^{Thr} (>95%) after incubation at 37° C for 30 min, while the same amount of BesG hydrolyzed <5% of Thr-tRNA^{Thr} in the during the same amount of time (*Figure 4.20BC*).

In order to improve the accuracy of our measurements and to obtain time-dependent information, we performed similar experiments in time course and developed a workflow for aminoacyl-tRNA quantification. tRNA was acylated, purified, then quantified by measuring the amount of free amino acid in sample after hydrolysis using a base (0.1M NaOH) (*Figure 4.21AB*)**.** Knowing the extent of aminoacylation of a standard allows us to quantitatively measure remaining aminoacyl-adenylate of arbitrary amounts (*Figure 4.21CD*). We performed a time course experiment using physiological concentrations of tRNA^{Thr} (<10 µM) and observed that 1 µM of BesG is able to deacylate approximately 3 µM of βes-tRNA^{Thr} in 15 min (*Figure 4.21E*). Little to no hydrolysis, above background rate of hydrolysis, was observed in reactions where 1 µM of BesG was incubated with 3.5 µM of Thr-tRNA^{Thr} , further confirming that BesG is a βes-tRNA^{Thr}-specific hydrolase (*Figure 4.21F*).

In a separate set of experiments, we measured the extent of acylation for BesG in competition with TRS using a discontinuous assay. Reactions contained TRS and BesG in a 1:20 molar ratio. Reaction progress was monitored in 5 min intervals by quenching samples of reaction in RNAse, as described previously. In reactions without BesG, acylation of tRNA^{Thr} increased throughout the duration of the assay (*Figure 4.22*). However, in reactions with BesG, βes acylation appeared to taper off at approximately 10% of the final observed amount observed in the no-BesG condition. These results suggest that BesG has the ability to competitively deacylate βes-tRNA^{Thr}, even in the presence of TRS. However, we note here the apparent differences in rates between these two sets of enzymes. Steady state levels of βes-tRNA^{Thr} were achieved using 1:20 ratio of TRS to BesG, suggesting BesG is much slower at catalyzing hydrolysis than TRS is at acylating. This observation is consistent with the low *in vitro* enzymatic activity measured for other Ala-X family enzymes [26,34] and may suggest that this family may not have evolved to clear a high level of misacylated tRNA *in vivo*. In addition to these *in vitro* experiments, we would like to assess the physiological function of BesG in the native host. Toward that end, the *S. cattleya* Δ*besG*::Ap^R knockout strain has been generated.

Heterologous expression of BesG for *in vivo* tRNA editing. We next turned our attention to exploring the use of BesG as an in *in vivo* tool for controlling incorporation of βes into proteins, using *E. coli* BL21 Star (DE3) as a model host. In contrast to methods that are based on expression of engineered ARS that are permissive to incorporation of non-standard amino acids, we seek to employ BesG as a tool to perform selective incorporation of βes through editing (*Scheme 4.1*) [35–38]. Unlike many other non-standard amino acids used for this purpose, βes is readily incorporated into proteins in a residue-specific manner without the requirement of any additional engineered components. We propose using BesG as a genetically encodable protein that can prevent incorporation of βes.

Towards this goal, we built a new pathway to consolidate vectors used. pEAGgt is a single plasmid that contains an operon for expressing BesE (from *S. cattleya*), BesA (from *S. cattleya*), and Ggt (from *E. coli*). We validated that this plasmid alone was sufficient to convert Pra to βes *in vivo (Figure 4.12 & Figure 4.23)*. Similar qualitative results in protein labeling experiments to the two vector system (pET-DUET-BesA.BesE and pGgt) were observed. Initial efforts indicated that BesG had no effect on reducing βes incorporation when co-expressed in this system (*Figure 4.24*)
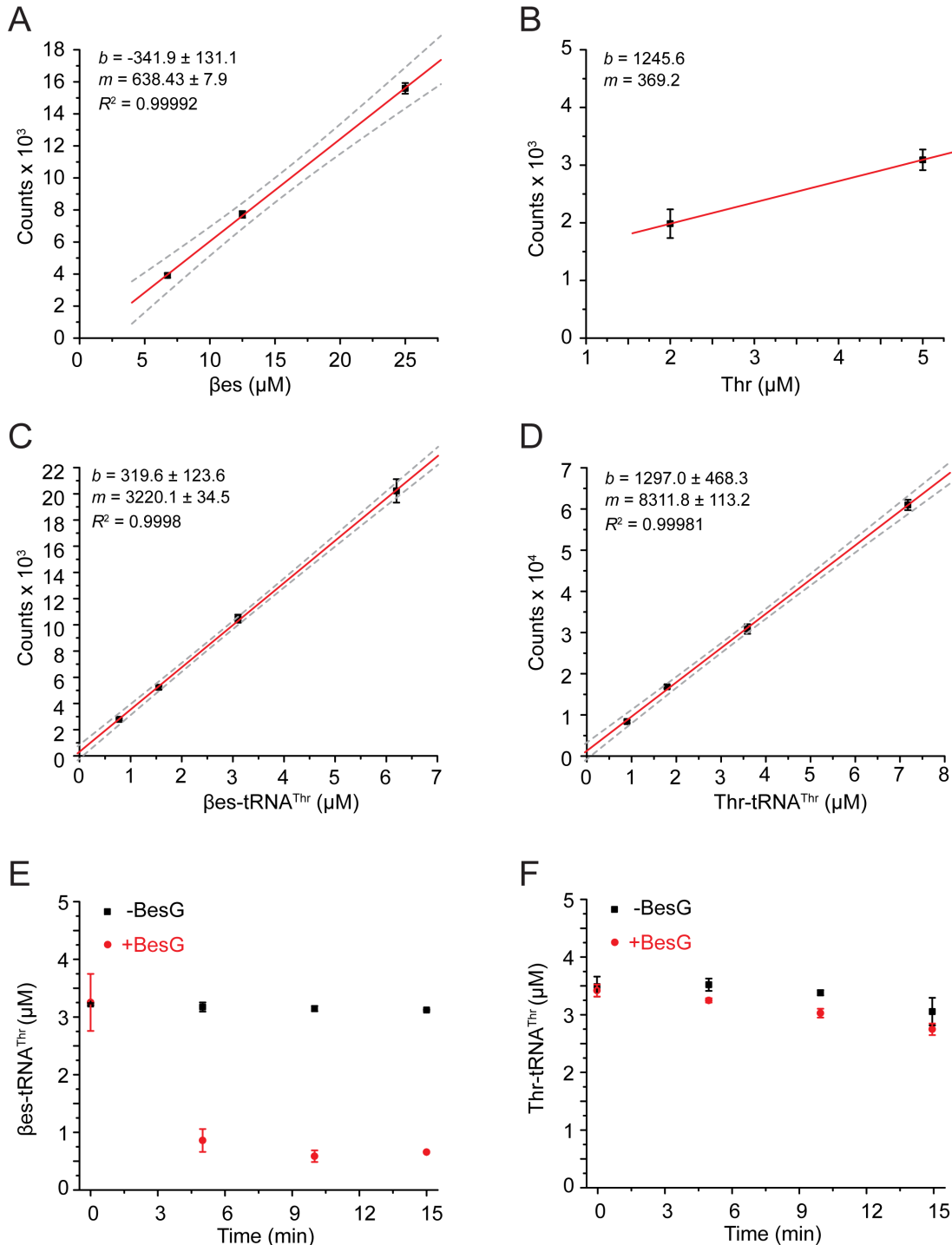
Figure 4.21. (A) & (B) Standard curves used for quantification of βes-tRNA^Thr and Thr-tRNA^Thr. Error bars are standard deviation (n=3), with 95% confidence interval shown in linear fit. (C) & (D) Standard curves of βes-adenylate and Thr-adenylate used in the quantification of βes-tRNA^Thr and Thr-tRNA^Thr for timecourse experiments. Error bars are standard deviation (n=3) with 95% confidence interval shown for linear fit. (E) & (F) Timecourse for the deacylation of βes-tRNA^Thr and Thr-tRNA^Thr in the presence and absence of 1μM BesG. Error bars are standard deviation (n=3).
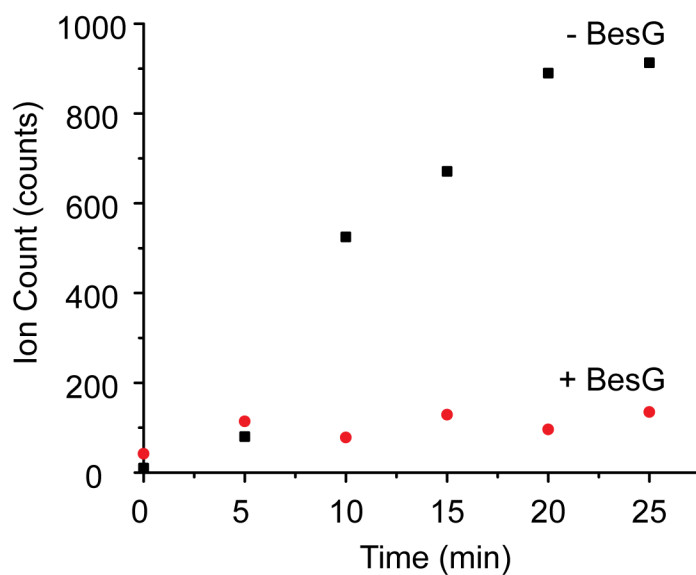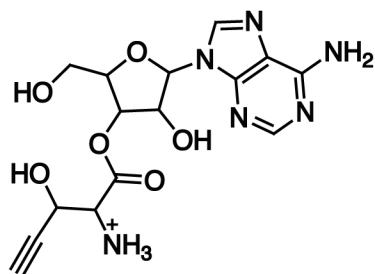
Figure 4.22. Time course of competitive acylation of TRS in the presence of BesG. Reactions containing TRS, βes, ATP, and tRNA[Thr] were monitored for presence of tRNA[Thr] (monitored as βes-adenylate, structure shown) using method described previously. Reactions containing BesG saw less formation of βes-tRNA[Thr], suggesting that sufficiently high concentration of BesG can effectively decrease βes-tRNA[Thr] even in the presence of TRS.
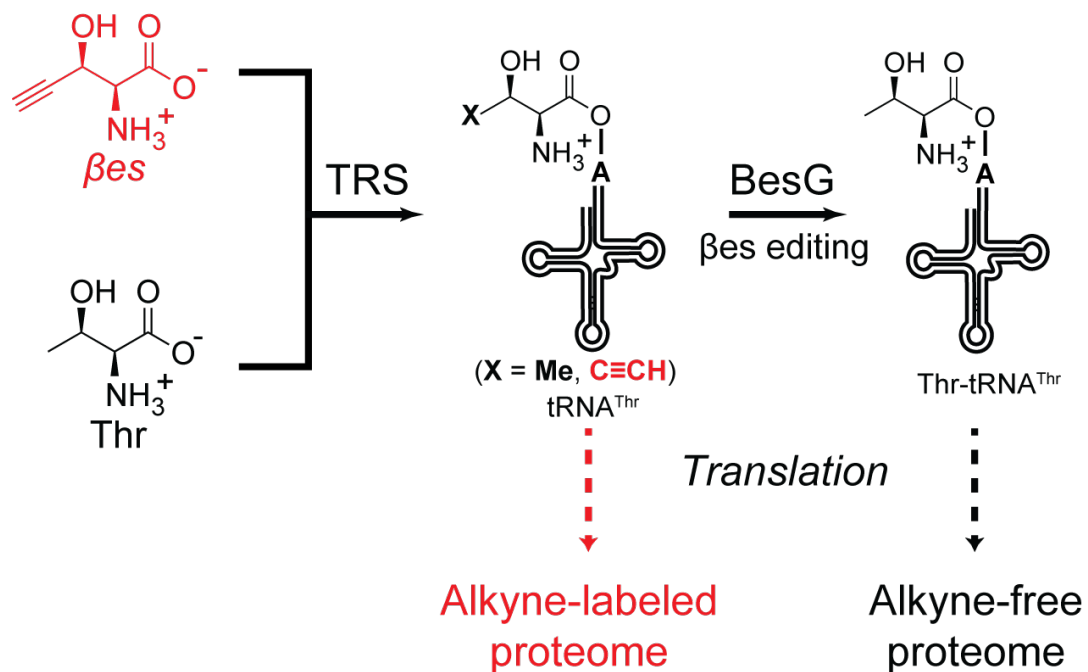
Scheme 4.1. Using BesG expression to control incorporation of βes into proteome. When BesG is not expressed, βes can incorporate into proteins (Thr residue replacement) through TRS and tRNA^{Thr}. We propose expressing BesG to stop βes incorporation into proteins.
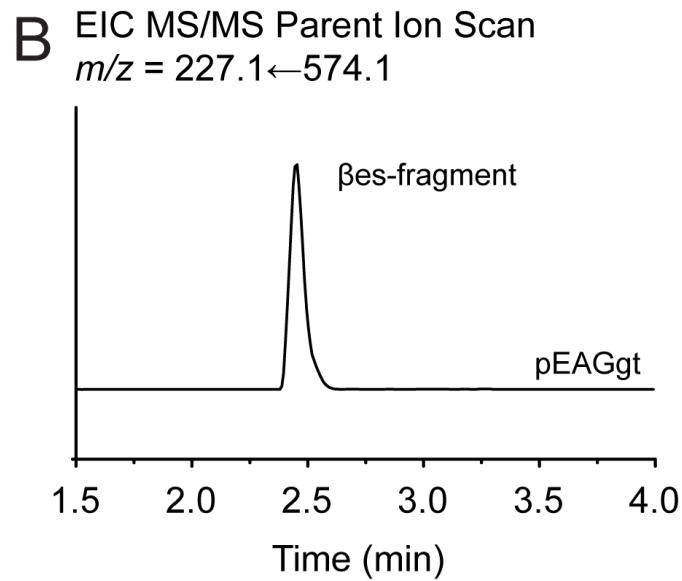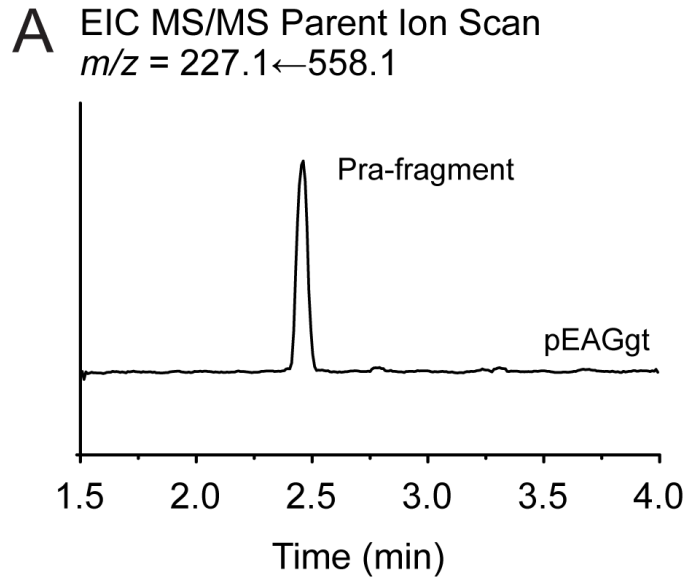
Figure 4.23 Validation of *E. coli* BL21 Star (DE3)pEAGgt vector for producing βes *in vivo*. *E.* Mass spectra shown are parent ion monitoring at the *m/z* = [Parent] →222.1 transition which gives two signatures. Cells were grown overnight after induction, and fed 1 mM Pra. (A) Pra-PEG$_3$-Biotin and (B) βes-PEG$_3$-Biotin peaks shown, indicating the single operon in pEAGgt is sufficient for converting Pra to βes.
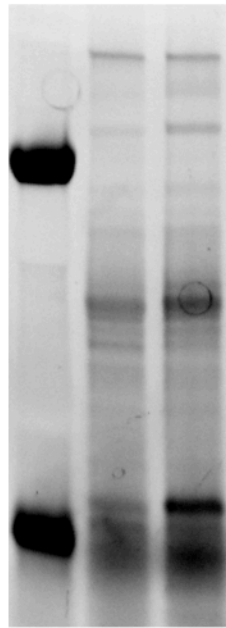
Figure 4.24. Preliminary *in vivo* incorporation of βes into proteins for E. coli BL21 Star (DE3), with or without pBesG, show little difference with respect to observed labeling of protein.

Interestingly, it has been noted that Ala-X family editing proteins may also demonstrate some recognition of the tRNA acceptor stem [26,34]. Upon examining the acceptor stem sequences of the tRNA^{Thr} from *S. cattleya* compared to *E. coli*, we found that all copies of *S. cattleya* tRNA^{Thr} have 5´-UCCA-3´ acceptor stems, while all copies of *E. coli* tRNA^{Thr} have 5´-ACCA-3´ acceptor stems (*Figure 4.25 & Figure 4.26*). Preliminary *in vitro* results indicated that the mutation of the 5´-ACCA-3´ motif to 5´-UCCA-3´ in *E. coli* tRNA^{Thr} appears to enable recognition by BesG (*Figure 4.27*). Meanwhile, mutating 5´-UCCA-3´ motif to 5´-ACCA-3´ in *S. cattleya* tRNA^{Thr} seems to abolish BesG recognition.

With this information in hand, we used Cas9 to engineer a new strain of *E. coli* BL21 Star (DE3) with all four tRNA^{Thr} acceptor stems mutated from 5′-ACCA-3′ to 5′-UCCA-3′ (strain name *E. coli* TUVW; genotype: *thrT*^{A73U} *thrU*^{A73U} *thrV*^{A73U} *thrW*^{A73U}). By mutating *E. coli* acceptor stems to have the same sequence as *S. cattleya* tRNA^{Thr} acceptor stems, we hoped to increase the *in vivo* efficacy of BesG in *E. coli*. In both *E. coli* BL21 Star (DE3) and TUVW, BesG was deployed as a codon optimized, N$_T$-MBP fusion (pBesG) since we saw moderate increases in soluble protein when BesG was expression in this manner. As a negative control, an empty vector (expressing only MBP) was used. Our test cultures contained pEAGgt + pBesG while our negative control cultures contained pEAGgt + pBesG-C189A. As a base for growth media, we used M9 with 19 amino acids (without Thr, M9aa –Thr), which allowed us precisely to control Pra and Thr present in cultures. Since BesG appears to be significantly slower at catalyzing βes-tRNA^{Thr} hydrolysis than TRS is at acylating βes onto tRNA^{Thr}, competition between Thr and βes may be needed for optimizing the desired incorporation of βes *in vivo*. Note that while in some conditions Thr can be absent entirely, *E. coli* can produce its own pool of threonine through endogenous primary metabolic pathways.

Using a similar assay as developed for Pra, we used CuAAC to attach a fluorescent dye (Tamra-Azide) to lysates of *E. coli* cultures. βes incorporation into the proteome was visualized by fluorescent imaging of SDS-PAGE gels. Our preliminary results indicate that the presence of BesG can affect proteome incorporation of βes *in vivo* (*Figure 4.28*). We see small differences between *E. coli* BL21 Star (DE3)and the *E. coli* TUVW strains, suggesting that acceptor stem identify might not be important for *in vivo* BesG activity under these conditions. However, the higher activity on the mutated tRNA stems observed *in vitro* may affect how conditions can be tuned. While these qualitative results are reproducible, additional experimentation and quantification are required to verify these observations. Toward these goals, a catalytically-dead mutant (BesG C189A), which lacks a crucial cysteine required for zinc binding, has been designed to replace the empty vector as a negative control.
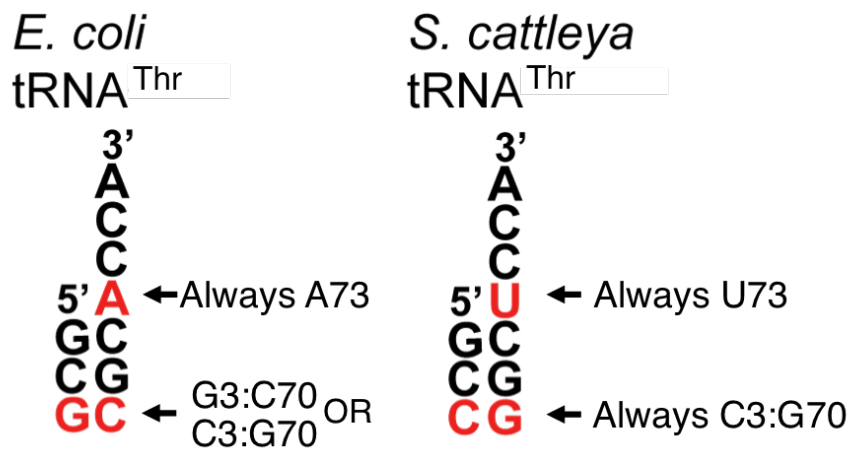
Figure 4.25. Isoforms of tRNA^{Thr} acceptor stems in *E. coli* and *S. cattleya*. *E. coli* tRNA^{Thr} acceptor stems all contain A73, whereas all *S. cattleya* tRNA^{Thr} acceptor have a U73. The 3:70 basepair varies is either C3:G70 or G3:C70 in *E. coli*, whereas it is C3:G70 in all *S. cattleya* tRNA^{Thr} isoforms.
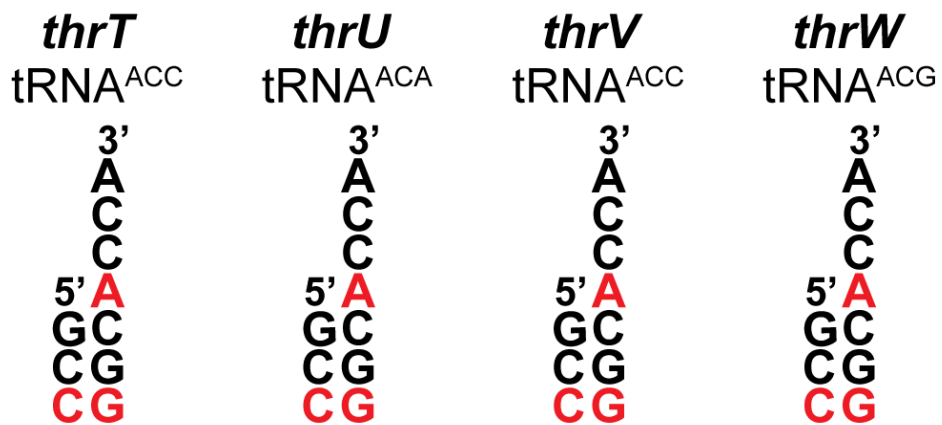
Figure 4.26. *E. coli* BL21 Star (DE3) tRNA^Thr acceptor stem isoforms. *E. coli* BL21 Star (DE3) contains four isoforms of tRNA^Thr, gene name is shown above each acceptor stem. Anticodon for each isoform is shown as superscript of tRNA.
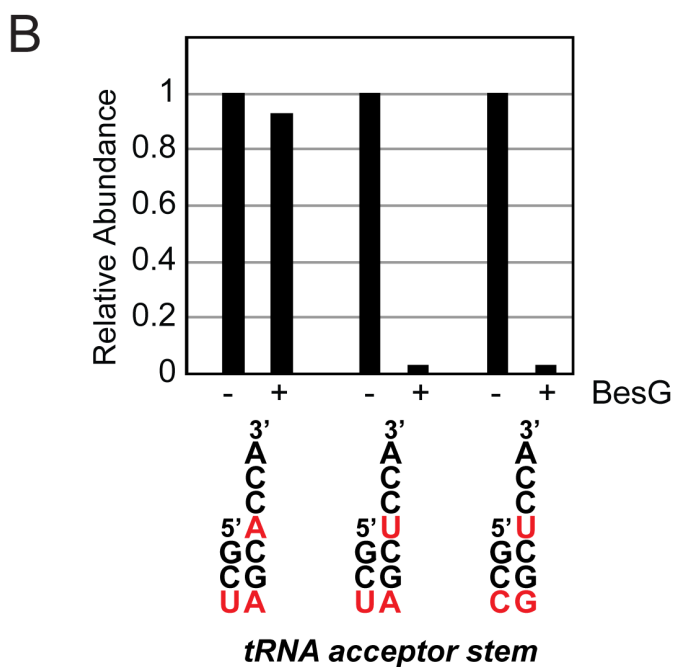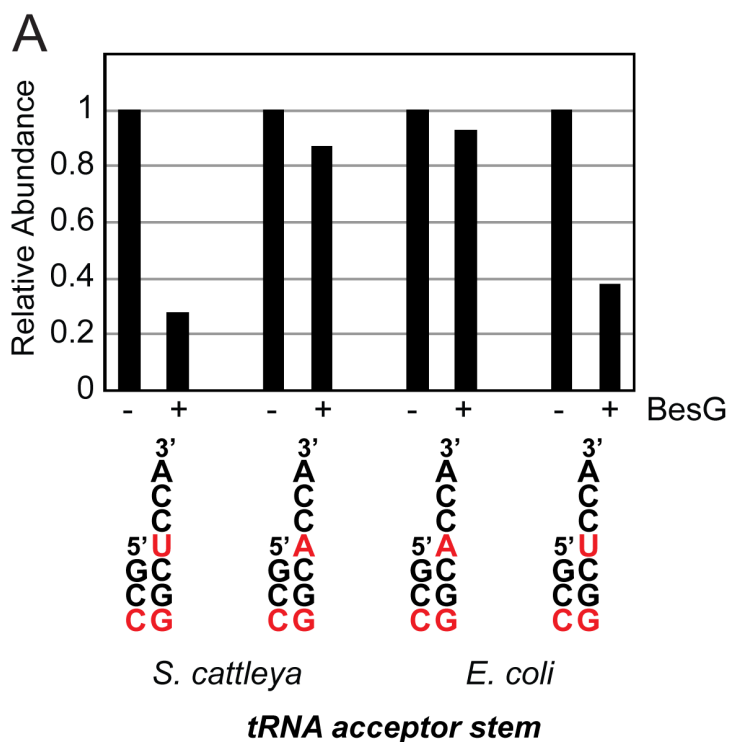
Figure 4.27. (A) Deacylation assays for tRNA^Thr from either *S. cattleya* or *E. coli*, with varying acceptor stems at the 73 position. Reactions were incubated for 20 min with or without BesG. Abundance, relative to condition without 1 μM BesG added, of βes-tRNA^Thr is shown. (B) Deacylation assays for tRNA^Thr from *E. coli*, with varying acceptor stems at the 3, 70 and 73 positions. Reactions were incubated for 20 min with or without BesG. Abundance, relative to condition without 1 μM BesG added, of βes-tRNA^Thr is shown.
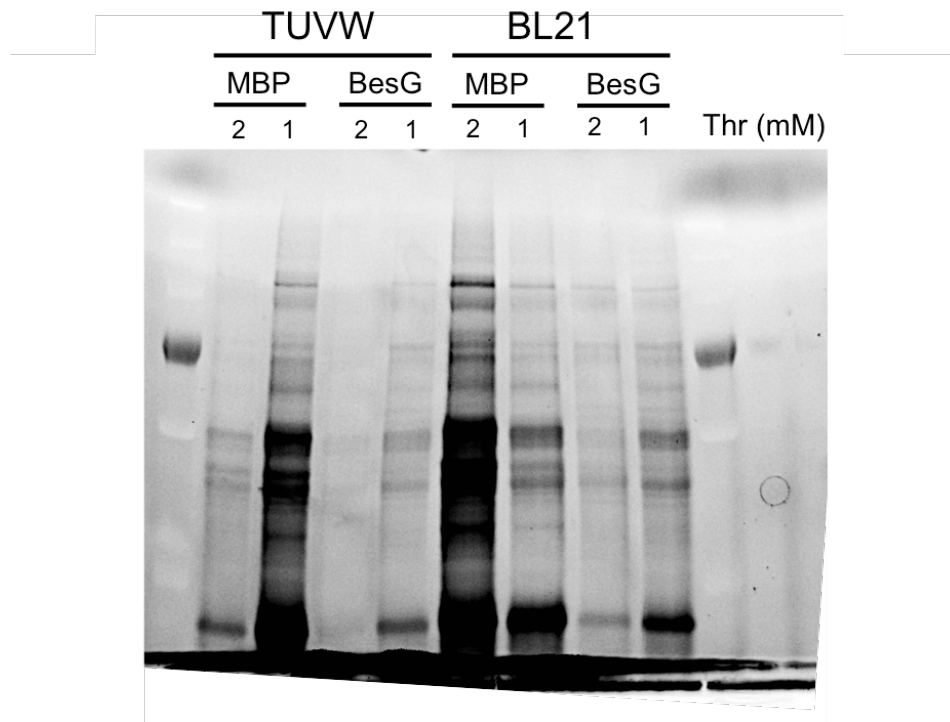
Figure 4.28. Preliminary *in vivo* incorporation of βes into proteins for *E. coli* expressing MBP-BesG or empty vector (MBP). Two cell lines were tested as hosts: either wildtype BL21 Star (DE3)or a BL21 Star (DE3)thr tRNA mutant strain (TUVW) that has all acceptor stems on Thr tRNA mutated to from 5'-ACCA-3' to 5'-UCCA-3'. Both strains harbor pEAGgt for the production of βes from Pra *in vivo*. Cells also expressed either BesG as an MBP-fusion (BesG) or empty vector (MBP). Thr was added to compete with βes incorporation. Note that results are preliminary, and additional controls are required to better understand results.

4.4. Conclusion

Protein translation requires a high degree of fidelity control to ensure that proteins synthesis is performed accurately. The observed rate of mistranslation has previously been estimated to be 1 mistake per 10,000 to 100,000 residues. The bulk of translational fidelity control machinery is occurs at the aminoacylation of tRNA. ARSs can discriminate strongly against the non-cognate standard proteinogenic amino acids when charging tRNA. If mistakes are made during aminoacylation, many ARSs also possess a secondary *cis*-editing domain that can remove incorrectly acylated amino acids from that tRNA. In addition to this general housekeeping function, organisms capable of biosynthesizing non-standard amino acids must inevitably solve a problem related to fidelity control.

In this chapter, we explored one possible route for the production of βes from Glu-βes (the product of the *bes* biosynthetic pathway). We established that ubiquitous enzyme, γ-glutamyl-transpeptidase (Ggt) is capable of using Glu-βes as a donor in transpeptidation *in vitro*, releasing free βes in the process. These results were validated *in vivo* with experiments that showed that overexpression of Ggt in *E. coli* resulted in higher levels of βes production. From this system, we constructed pβes, the first vector containing a full functional pathway for the generation of βes from lysine.

The ability to efficiently produce βes in living organisms allowed us to study the fate of βes *in vivo*. In these studies, we found that *E. coli* lacks the ability to discriminate between βes and Thr, and readily incorporates βes into its proteome. Furthermore, initial studies indicated that βes seemed to be incorporated at Thr sites, which is supported by proteomic experiments. The high level of βes mistranslation raises the question of how βes translation is prevented in the native host. Interestingly, *S. cattleya* produces two amino acids that are similar in structure to Thr: F-Thr and βes. Previous work from our lab identified a standalone *trans*-editing enzyme in *S. cattleya*, FthB, which was characterized to be a F-Thr-tRNA^{Thr} deacylase that is highly selective for F-Thr-tRNA^{Thr} over Thr-tRNA^{Thr}. Investigating *bes* genes of other streptomycetes containing hosts, we found that those found in *S. sp.* NRRL S-1448 and *S. alboverticillatus* have an additional clustered gene, *besG*, which is a member of the AlaX aminoacyl-tRNA *trans*-editing family with a similar function to FthB. While *S. cattleya* does not have a copy of *besG* within its own *bes* gene cluster, a homologue of *besG* was found 84 ORFs downstream. Characterizing BesG *in vitro*, we have found that it is competent to carry out the βes-tRNA^{Thr} deacylation reaction selectively over Thr-tRNA^{Thr}.

Based on these results, we have been developing a platform for utilizing BesG as a genetically-encodable tool for controlling translation of βes, a terminal-alkyne containing amino acid. We believe βes has potentially utility as a tool, since it is both completely genetically encodeable and does not requires additional engineered ARS for incorporation into proteins. Our initial experiments in *E. coli* have established a framework for working with BesG *in vivo*. Future work in this project area should aim to focus on improving flux through *bes* pathway in order to provide a means for genetically-encoding a terminal alkyne amino acid, increasing BesG activity *in vivo*, as well as understanding how BesG can change the balance between Thr and βes incorporation *in vivo*.

4.5. References

1.   Cochella, L. & Green, R. Aminoacylation error correction. *Curr. Biol.* **15,** R536–R540 (2005).

2.   Yadavalli, S. S. & Ibba, M. Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. in *Advances in protein chemistry and structural biology* **86,** 1–43 (Elsevier Inc., 2012).

3.   Ling, J., Reynolds, N. & Ibba, M. Aminoacyl-tRNA synthesis and translational quality control. *Annu. Rev. Microbiol.* **63,** 61–78 (2009).

4.   Zhou, X. L., Ruan, Z. R., Huang, Q., Tan, M. & Wang, E. D. Translational fidelity maintenance preventing ser mis-incorporation at Thr codon in protein from eukaryote. *Nucleic Acids Res.* **41,** 302–314 (2013).

5.   Sankaranarayanan, R. *et al.* Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* **7,** 461–465 (2000).

6.   Dock-Bregeon, A. C. *et al.* Achieving error-free translation: The mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution. *Mol. Cell* **16,** 375–386 (2004).

7.   Nureki, O. *et al.* Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science (80-.).* **578,** 578–583 (2010).

8.   Fukai, S. *et al.* Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell* **103,** 793–803 (2000).

9.   Fersht, A. R. & Dingwall, C. Evidence for the double-sieve editing mechanism in protein synthesis. Steric exclusion of isoleucine by valyl-tRNA synthetases. *Biochemistry* **18,** 2627–2631 (1979).

10.   Beuning, P. J. & Musier-Forsyth, K. Hydrolytic editing by a class II aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci.* **97,** 8916–8920 (2002).

11.   Beebe, K., Mock, M., Merriman, E. & Schimmel, P. Distinct domains of tRNA synthetase recognize the same base pair. *Nature* **451,** 90–93 (2008).

12.   Suga, H. *et al.* Quality control by trans -editing factor prevents global mistranslation of non-protein amino acid α-aminobutyrate . *RNA Biol.* **15,** 576–585 (2017).

13.   Liu, Z. *et al.* Homologous *trans* -editing factors with broad tRNA specificity prevent mistranslation caused by serine/threonine misactivation. *Proc. Natl. Acad. Sci.* **112,** 6027–6032 (2015).

14.   Sokabe, M., Okada, A., Yao, M., Nakashima, T. & Tanaka, I. Molecular basis of alanine discrimination in editing site. *Proc. Natl. Acad. Sci.* **102,** 11669–11674 (2005).

15.   Das, M., Vargas-Rodriguez, O., Goto, Y., Suga, H. & Musier-Forsyth, K. Distinct tRNA recognition strategies used by a homologous family of editing domains prevent mistranslation. *Nucleic Acids Res.* **42,** 3943–3953 (2014).

16.   Sharma, A. *et al.* Uneven spread of cis-and trans-editing aminoacyl-tRNA synthetase domains within translational compartments of P. falciparum. *Sci. Rep.* **1,** 1–11 (2011).

17.   Schaffrath, C., Murphy, C. D., Hamilton, J. T. G. & O'Hagan, D. Biosynthesis of fluoroacetate and 4-fluorothreonine in Streptomyces cattleya. Incorporation of oxygen-18 from [2-2H,2-18O]-glycerol and the role of serine metabolites in fluoroacetaldehyde biosynthesis. *J. Chem. Soc. Perkin Trans. 1* 3100–3105 (2001). doi:10.1039/b108720g

18.   Sanada, M., Miyano, T. & Iwadare, S. Ethynylserine, an antimetabolite of L-threonine from Streptomyces cattleya. *J. Antiboitics* **39,** 304–305 (1986).

19.   McMurry, J. L. & Chang, M. C. Y. Fluorothreonyl-tRNA deacylase prevents mistranslation in the organofluorine producer Streptomyces cattleya . *Proc. Natl. Acad. Sci.* **114,** 11920–11925 (2017).

20.   Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci.* **100,** 1541–1546 (2003).

21.   Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature.* doi:10.1038/s41586-019-1020-y

22.   Wickham, S., West, M. B., Cook, P. F. & Hanigan, M. H. Gamma-glutamyl compounds: Substrate specificity of gamma- glutamyl transpeptidase enzymes. *Anal. Biochem.* **4,** 208–214 (2011).

23.   Whitfield, J. B. Gamma glutamyl transferase. *Crit. Rev. Clin. Lab. Sci.* **38,** 263–355 (2001).

24.   Keillor, J. W., Castonguay, R. & Lherbet, C. Gamma-glutamyl transpeptidase substrate specificity and catalytic mechanism. *Methods Enzymol.* **401,** 449–467 (2005).

25.   Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to precisely control protein expression. *Nat. Biotechnol.* **27,** 946–950 (2010).

26.   Fukunaga, R. & Yokoyama, S. Structure of the AlaX-M trans-editing enzyme from Pyrococcus horikoshii. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63,** 390–400 (2007).

27.   Ahel, I., Korencic, D., Ibba, M. & Söll, D. Trans-editing of mischarged tRNAs. *Proc. Natl. Acad. Sci.* **100,** 15422–15427 (2003).

28.   Pasman, Z. *et al.* Substrate specificity and catalysis by the editing active site of alanyl-tRNA synthetase from Escherichia coli. *Biochemistry* **50,** 1474–1482 (2011).

29.   Zhang, X. & van der Donk, W. A. On the substrate specificity of dehydration by lacticin 481 synthetase. *J. Am. Chem. Soc.* **129,** 2212–3 (2007).

30.   Garner, P. & Park, J. M. The synthesis and configuartional stability of differentially protected beta-hydroxy-alpha-amino aldehydes. *J. Am. Chem. Soc.* **52,** 2361–2364 (1987).

31.   McKillop, A., Taylor, R. J. K., Watson, R. J. & Lewis, N. An improved procedure for the preparation of the Garner aldehyde and its use for the synthesis of N-protected 1-Halo-2-(R)-amino-3-butenes. *Synthesis (Stuttg).* 31–33 (1994).

32.   Herold, P. Synthesis of D-erythro- and D-threo-sphingosine derivatives from L-serine. *Helv. Chim. Acta* **71,** 354–362 (1988).

33.   Chan, J. K. K. Identification and enzyme studies of rare amino acid biosynthesis from Streptomyces cattleya. *PhD Diss. Univ. St. Andrews* (2013).

34.   Novoa, E. M. *et al.* Ancestral AlaX editing enzymes for control of genetic code fidelity are not tRNA-specific. *J. Biol. Chem.* **290,** 10495–503 (2015).

35.   Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 225–249 (2006).

36.   Xie, J. & Schultz, P. G. Adding amino acids to the genetic repertoire. *Curr. Opin. Chem. Biol.* **9,** 548–554 (2005).

37.   Beatty, K. E. & Tirrell, D. A. Two-color labeling of temporally defined protein populations in mammalian cells. *Bioorganic Med. Chem. Lett.* **18,** 5995–5999 (2008).

38.   Dieterich, D. C. *et al.* Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc.* **2,** 532–40 (2007).

Appendix 1: Supplementary Tables and Sequences

Table A1.1 Strains, plasmids, oligonucleotides, and sequences used in discovery and characterization of *bes* pathway. (A) Strains and plasmids used in this study. (B) Oligonucleotides used for plasmid construction, genomic library construction, and screening/sequencing. (C) Sequences of synthetic genes.

A1.1A. Strains and plasmids

Strain	Description	Source
BL21 Star (DE3)	*F-ompT hsdSB (rB-, mB-) gal dcm rne131 (DE3)*	Thermo-Fisher
GM272	*F- dam-3 dcm-6 hsdS2l metb1 lacYl or Z4 galK2 ga/T22 mtl-2 tonA2or A31 tsx-1 or -78 supE44 (thi-1)*	ATCC
B834(DE3)	F⁻ *ompT hsdS*B(r$_B^-$ m$_B^-$) *gal dcm met* (DE3)	Novagen
S. cattleya	NRRL 8057 (ATCC 35852)	ATCC
S. cattleya ΔbesA	Δ*besA*::Am^R-OriT	This study
S. cattleya ΔbesB	Δ*besB*::Am^R,-OriT	This study
S. cattleya ΔbesC	Δ*besC*::Am^R-OriT	This study
S. cattleya ΔbesD	Δ*besD*::Am^R-OriT	This study
S. cattleya ΔbesE	Δ*besE*::Am^R-OriT	This study
S. cattleya ΔbesB+besB	Δ*besB*::Am^R *attB^{ΦC31}*::*ermEp*-besB*-Hyg^R	This study
S. cattleya ΔbesC+besC	Δ*besC*::Am^R *attB^{ΦC31}*::*ermEp*-besC*-Hyg^R	This study
S. cattleya ΔbesD+besD	Δ*besD*::Am^R *attB^{ΦC31}*::*ermEp*-besD*-Hyg^R	This study

Plasmid	Description	Source
pUZ8002	*tra*, *neo*, Km^R, RP4	*Ref. 34*
pIJ773	*lacI*, Cb^r, Am^R, OriT, ColE1	JIC
pIJ773-BesA	*besA* flanking regions (2 kB US/DS)-Am^R-OriT, *lacI*, Cb^R, ColE1	This study
pIJ773-BesB	*besB* flanking regions (2 kB US/DS)-Am^R-OriT, *lacI*, Cb^R,ColE1	This study
pIJ773-BesC	*besC* flanking regions (2 kB US/DS)-Am^R-OriT, *lacI*, Cb^R, ColE1	This study
pIJ773-BesD	*besD* flanking regions (2 kB US/DS)-Am^R-OriT, *lacI*, Cb^R, ColE1	This study
pIJ773-BesE	*besE* flanking regions (2 kB US/DS)-Am^R-OriT, *lacI*, Cb^R, ColE1	This study
pSET152	Hyg^R, *int(φ31)*, OriT, pUC18	JIC
pSET152-BesB	BesB (ErmEp*), Hyg^R, *int(φ31)*, OriT, pUC18	This study
pSET152-BesC	BesC (ErmEp*), Hyg^R, *int(φ31)*, OriT, pUC18	This study
pSET152-BesD	BesD (ErmEp*), Hyg^R, *int(φ31)*, OriT, pUC18	This study
pET16b-His₁₀-BesA	His$_{10}$-BesA (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His₁₀-BesD	His$_{10}$-BesD (T7), *lacI*, Cb^R, ColE1	This study
pET16b-His₁₀-BesE	His$_{10}$-BesE (T7), *lacI*, Cb^R, ColE1	This study
pET16b-Strep-BesD	Strep-BesD (T7), *lacI*, Cb^R, ColE1	This study
pET16b-Strep-BesD G139D	Strep-BesD G139D (T7), *lacI*, Cb^R, ColE1	This study
pSV272.1-His₆MBP-BesB	His$_6$MBP-BesB (T7), *lacI*, Km^R, f1, ColE1	This study
pSV272.1-His₆MBP-BesB-opt	His$_6$MBP-BesB (T7) codon optimized, *lacI*, Km^R, f1, ColE1	This study
pSV272.1-His₆MBP-BesC	His$_6$MBP-BesC (T7), *lacI*, Km^R, f1, ColE1	This study
pETDUET-1	*lacI*, f1, Cb^R, ColE1	Novagen
pETDUET-BesC.BesD	BesD (T7) and BesC (T7) from *P. fluorescens*, *lacI*, Cb^R,	This study

172

	ColE1	
pPra	BesD (T7) and BesC (T7) from *P. fluorescens*, MBP-sBesB (T7) from *S.* sp. NRRL S-1448, *lacI*, Cb^R, ColE1	This study
pG-KJE8	DnaK-DnaJ-GrpE (AraB) GroES-GroEL (Pzt-1), Cm^R, f1, p15A	Takara Bio
pGro7	GroES-GroEL (AraB)*,* Cm^R, f1, p15A	Takara Bio
pKJE7	DnaK-DnaJ-GrpE (Pzt-1), Cm^R, f1, p15A	Takara Bio
pG-Tf2	GroES-GroEL-Tig (Pzt-1), Cm^R, f1, p15A	Takara Bio
pTf16	Tig (Ara), Cm^R, f1, p15A	Takara Bio
pET28a-PraRS	PraRS (T7)*, lacI*, Km^R, pBR322	This study
pBAD24-PraRS	PraRS (araB), Cb^R*,* f1, ColE1	This study
praGro	GroES-GroEL (AraB), PraRS (AraB), Cm^R, f1, p15A	This study
pET32a-BesB	Trx-His$_6$-BesB from S. sp. NRRL S-1448 (T7), Cb^R, ColE1	This study
pET41b-BesB	GST-His$_6$-BesB from S. sp. NRRL S-1448 (T7), Cb^R, ColE1	This study
pBAD24-BesB	BesB from S. sp. NRRL S-1448 (araB), Cb^R*,* f1, ColE1	This study
pTrc33-BesB	His$_6$-BesB from S. sp. NRRL S-1448 (trc)*,* Cb^R, ColE1	This study
pET16b-glnS-BesB	His$_6$-BesB from S. sp. NRRL S-1448 (glnS), Cb^R, ColE1	This study
pET16b-BesB*Δ5*	His$_{10}$-BesB *Δ5aa* (T7), *lacI*, Cb^R, ColE1	This study
pET16b-BesB*Δ10*	His$_{10}$-BesB *Δ10aa* (T7), *lacI*, Cb^R, ColE1	This study
pET16b-BesB*Δ15*	His$_{10}$-BesB *Δ15aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ20*	His$_{10}$-BesB *Δ20aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ25*	His$_{10}$-BesB *Δ25aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ30*	His$_{10}$-BesB *Δ30aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ35*	His$_{10}$-BesB *Δ35aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ40*	His$_{10}$-BesB *Δ40aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ45*	His$_{10}$-BesB *Δ45aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pET16b-BesB*Δ50*	His$_{10}$-BesB *Δ50aa* (T7), *lacI*, Cb^R*,* ColE1	This study
pEVOL-MetRS-CUA	MetRS (araB), MetRS (glnS*), tRNA^{Met}CUA (proK), Cm^R, p15A	This study
pEVOL-PraRS-CUA	PraRS (araB), PraRS (glnS*), tRNA^{Met}CUA (proK), Cm^R, p15A	This study
pREP-GFP	eGFP (T7), Km^R, ColE1	This study
pREP-GFP-CUA	eGFP-N$_T$-CUA (T7), Km^R, ColE1	This study

A1.1B. Oligonucleotide sequences

Name	Sequence
LR-SCAT_0184-F	acaccccccgcctcgttcaccgggcgtgacaacatcgatgatgtattccggggatccgtcgacctgcagttcg
LR-SCAT_0184-R	cgcggtgtcttcttcccgtcggcagggcgggacgcgtttgtcatgtaggctggagctgcttcaagttcctata
LR-SCAT_2136-F	cgtttccgtaggttacggttccgtagccgtagccgttacgtactccgagggacgtctatgtattccgggggatccgtcgacctgcagttcg
LR-SCAT_2136-R	acggtcctcgtcgcacccgcgccatcagcactgccgtcgttccccacgcccctcatcatgtaggctggagctgcttcaagttcctata
LR-SCAT_4823-F	gccgcggcgggcccctgcgacatcgaacggtaggagcaccattccggggatccgtcgacc
LR-SCAT_4823-R	gcctgacctcgggttcgtgagccaggcggttgtaacaatctgtaggctggagctgcttc
LR-SCAT_p1525-F	aggctgttgtcgccgtgccggggccgggagactcgacggcattccggggatccgtcgacc
LR-SCAT_p1525-R	taccggcgccggttcggaccggggacggtgtgacgcccggtgtaggctggagctgcttc
K773-BesA-UP-F	gggatccactagtatggaagcgctctgcctcgttcatcc
K773-BesA-UP-R	cgcggtggcggccgcgtcctgggcatgctgttggtccgc
K773-BesA-DOWN-F	acaaaagctggtaccggatcggcatcgacaaggaacacctcga
K773-BesA-DOWN-R	cggaatatcaagcttggcccgcgcagacgttcaggc
K773-BesB-UP-F	cggggggatccactagtcgggttgagcttgccgagggtct
K773-BesB-UP-R	caccgcggtggcggccgctgacaacggtgaggaccagcgc
K773-BesB-DOWN-F	gtatcgataagcttacctgaagtcgaagtctcagcaagagggaac
K773-BesB-DOWN-R	ggaatatcaagctttcacttgccgatgcccacgagttcc
K773-BesC-UP-F	agcccggggggatccactagtgcgcaaccaggggtacgacgac
K773-BesC-UP-R	tccaccgcggtggcggccgcccctggcggacgtacggcgg
K773-BesC-DOWN-F	agggaacaaaagctggtacccgttgcacaccgttccgccacc
K773-BesC-DOWN-R	aagcttgatattccggggatccgtctctggtttccctcacctgtcgcgc
K773-BesD-UP-F	tctagaactagtggaacggcgccaattcaccgcaca
K773-BesD-UP-R	ggatccactagtggactgccgctccaggaagccg
K773-BesD-DOWN-F	atatcaagcttgatgaatctccgatgctcgcgtatgcg
K773-BesD-DOWN-R	gaattgggtacccggcaaggccgccttcgtggg
K773-BesE-UP-F	ttatccactagtctcgcctccgatacctgaagtcgaagtctca
K773-BesE-UP-R	ggatccactagttcgaccagtccgaggaattcgaagacgc
K773-BesE-DOWN-F	atatcaagcttacgaagaacgtctgcccggtggc
K773-BesE-DOWN-R	gaattgggtacccgtggcgagttcggccgcgtc
dcK-X-US-R	atgccgctcgccagtcgattggc
dcK-X-DS-F	cgtcagccgggcaggataggtgaagta
dcK-BesA-US-F	ggcgactaccacaacgcccgtcaac
dcK-BesA-DS-R	tccccttctacatcacgggggatggcc
dcK-BesB-US-F	gtggtgtagcgggcgtgcccg
dcK-BesB-DS-R	aagaaccccggcacccggacg
dcK-BesC-US-F	gccagtactggggctgccgctt
dcK-BesC-DS-R	cggacccttcaggaccagacgacca
dcK-BesD-US-F	gcgttccaccttgcggtagttggcc
dcK-BesD-DS-R	cgaccgtgacgcatagtggccgaa
dcK-BesE-US-F	ctcacttgccgatgcccacgagtt
dcK-BesE-DS-R	aacatggcctgctcgtacgcgg
dcK- 0184-IN-F	tcatcgcaccggtcgacgtactcggg
dcK-0184-IN-R	gtgacggcggcggagtccggg
dcK-2136-IN-F	atgaccacacagccggacgtgatcacc
dcK-2136-IN-R	tcatgccgccttggcgacatcttccttacg
dcK-4823-IN-F	ttgaccgccatcgaccccaccgc
dcK-4823-IN-R	acaatctcaggcggcgacgagctcctg
dcK-p1525-IN-F	atgacgatttccaccgaggttcccctcctg
dcK-p1525-IN-R	gtcaccgccgtgcggacgccttg
dcK- 0184-OUT-F	cggcagggcgggacgcgtttg
dcK-0184-OUT-R	ctcgttcaccgggcgtgacaacatcg
dcK-2136-OUT-F	ccgtagccgttacgtactccgagggacgtct
dcK-2136-OUT-R	catcagcactgccgtcgttccccacg
dcK-4823-OUT-F	ccctgcgacatcgaacggtaggagcacc

174

dcK-4823-OUT-R	cgggttcgtgagccaggcggttgtaacaatc
dcK-p1525-OUT-F	cccggtgactggcaggctgttgtcg
dcK-p1525-OUT-R	ccggttcggaccggggacggtgtg
dcK-BesA-IN-F	tcttccggttcgacgagccggc
dcK-BesA-IN-R	cgagggtctccggttccagcagc
dcK-BesB-IN-F	aagacctcggccgatcagctccgg
dcK-BesB-IN-R	ttcgaatccggccgacccggcg
dcK-BesC-IN-F	atgaccgacctgaacaccccggaa
dcK-BesC-IN-R	cttgccgatgcccacgagttccc
dcK-BesD-IN-F	gtcatgttcaggatgatgcgggtggtg
dcK-BesD-IN-R	ccttccaccggttcggcatcgtcac
dcK-BesD-IN-F	acgacacaccaccacgcgacgt
dcK-BesD-IN-R	aggaccagacgaccagttcgccg
His$_{10}$-BesA F	catcacagcagcggccatatcgaaggtcgtcatatgactgatcccggttcgacgaagatcctgatctacgc
His$_{10}$-BesA R	ttcctttcgggctttgttagcagccggatcctcgagtcaggcgctggggtacggcccgg
MBP-BesB-SCAT-F	acctcgggatcgaggaaaacctgtattttcagggcatgtccgtctcgatccccgacgtgg
MBP-BesB-SCAT-R	agtgcggccgcaagcttgtcgacggagctcgaatttcagtcggggcgcagggcc
MBP-BesB-S31-F	acctcgggatcgaggaaaacctgtattttcagggcatgtccatccccgacgtggcctc
MBP-BesB-S31-R	agtgcggccgcaagcttgtcgacggagctcgaatttcacccggggcgcagggc
MBP-BesB-LAVE-F	acctcgggatcgaggaaaacctgtattttcagggcatgtccgtctcgatccccgacgtgg
MBP-BesB-LAVE-R	agtgcggccgcaagcttgtcgacggagctcgaatttcagtcggggcgcagggcc
MBP-BesB-S1448-F	acctcgggatcgaggaaaacctgtattttcagggcatgaccaccgaggcatccggggag
MBP-BesB-S1448-R	agtgcggccgcaagcttgtcgacggagctcgaatttcagaggggggcgcagggcc
MBP-BesB-CATE-F	acctcgggatcgaggaaaacctgtattttcagggcatgacgctcggtgcccccgagg
MBP-BesB-CATE-R	agtgcggccgcaagcttgtcgacggagctcgaatttcagtcgggacgcagggccc
MBP-BesC-F	caacctcgggatcgaggaaaacctgtattttcagggcatgaccgacctgaacacccccggaatccacc
MBP-BesC-R	gtgcggccgcaagcttgtcgacggagctcgaattctcacttgccgatgcccacgagttcccgcac
His$_{10}$-BesD-F	cagcggccatgaaaacctgtattttcagggcatgggcagtaaccggcaggaactcaaagacgtgtg
His$_{10}$-BesD-F2	catcatcatcatcatcatcacagcagcggccatgaaaacctgtattttcagggc
His$_{10}$-BesD-R	ttcctttcgggctttgttagcagccggatcctcgagtcagtccttgatggcccgtgccgccg
His$_{10}$-BesE-F	catcacagcagcggccatatcgaaggtcgtcatatgtccggcacgacacaccaccacgc
His$_{10}$-BesE-R	ttcctttcgggctttgttagcagccggatcctcgagtcaggaccagacgaccagttcgccgccg
His$_{6}$-BesF-F	acagcagcggcctggtgccgcgcggcagccatatgatgaagctctccaaagccgtctggg
His$_{6}$-BesF-R	ccgcaagcttgtcgacggagctcgaattcggatccctaccgacgtcccggagttgacgcct
His$_{6}$-γ-GT(e.c.)-F	acagcagcggcctggtgccgcgcggcagccatatgatgataaaaccgacgttttac
His$_{6}$-γ-GT(e.c.)-R	ccgcaagcttgtcgacggagctcgaattcggatccttagtaccccgccgttaaatcatc
Strep-BesD-F	agctggagccacccgcagttcgaaaagggtgcaggtatgggcagtaaccggcaggaactcaaagacgtgtg
Strep-BesD-F2	ataattttgtttaactttaagaaggagatataccatggcaagctggagccacccgcagttcgaaaag
Strep-BesD-R	gtgcggccgcaagcttgtcgacggagctcgaattctcagtccttgatggcccgtgccgccg
Strep-BesD G139D-F	ataattttgtttaactttaagaaggagatataccatggcaagctggagccacccgcagttcgaaaag
Strep-BesD G139D-R	cgcgaagctgaagtcatcccagtgccagccgtgggtgtccc
Strep-BesD G139D-F2	cggctggcactgggatgacttcagcttcgcgctgatctgggtgctcc
Strep-BesD G139D-R2	gtgcggccgcaagcttgtcgacggagctcgaattctcagtccttgatggcccgtgccgccg
MBP-BesB-Opt-F	acctcgggatcgaggaaaacctgtattttcagggcatgaccactgaggccagtggggaa
MBP-BesB-Opt-R	tttgatctcgagtgcggccgcaagcttgtcgacggagctcgaatttcacaagggacgcagagcttgctc
petDUET-A-F	aataattttgtttaactttaagaaggagatataccatgactgatcccggttcgacgaa
petDUET-A-R	cgacctgcaggcgcgccgagctcgaattcggatccggcgctggggtacggcccgg
petDUET-E-F	tagtatattagttaagtataagaaggagatatacatatgtccggcacgacacaccacca
petDUET-E-R	tttctttaccagactcgagggtaccggaccagacgaccagttcgccg
pPra-C-F	aataattttgtttaactttaagaaggagatataccatgtccatcactcaagaaacgtttc
pPra-C-R	gtcgacctgcaggcgcgccgagctcgaattcggatcctcattggcgtcctccagcgttgag
pPra-D-F	tcttagtatattagttaagtataagaaggagatatacatgaactacgtacttgatgaggct
pPra-D-R	ttcgcagcagcggtttctttaccagactcgagggtaccttaggtcaagctcttggctgcctc
pPra-B-F	cagcacgtcgaggcagccaagagcttgacctaaggtaccaagcttgcggccgcataatgcttaa
pPra-B-R	caaatttcgcagcagcggtttctttaccagactcgagtcacaagggacgcagagcttgctca
MetRS-F	atcatcatcatgagaatctctacttccagggtaccggcgcatgactcaagtcgcgaagaa
MetRS-R	ctcgagtgcggccgcaagcttgtcgacggagctcgaattttatttcacctgatgacccgg

PraRS-A-F	gtgacgtgcgcaccgccgtacgctaacggctcaatccacctcggccatatgctggagca
PraRS-A-R	catctggccaatggtgccgtccagccagacgtagaaatatttgcccggcgcgttcggaat
PraRS-B-F	tatttctacgtctggctggacggcaccattggccagatgggttctttcaagaatctgtgc
PraRS-B-R	ccttccagcatggcaggccagaacaggctaaagaagtaaacaatatctttaccgatgaag
PraRS-C-F	gtttacttctttagcctgttctggcctgccatgct
PraRS-C-R	taataaaggtgccgcgagacttggacatcttcacgccgttcaccgtcacatagccatgaa
PraRS-D-F	tgtgacggtgaacggcgtgaagatgtccaagtctcgcggcaccttta
PraRS-D-R	ggtggattgagccgttagcgtacggcggtgcgcacgtcaccagaatttttcttcgcgac
praGro-F	tgccctgcacctcgcagaaataaacaaaccccgggcatgacaaaaacgcgtaacaaaagtgt
praGro-R	taattctcatgtttgacagcttatcatcgataagctttttatttcacctgatgacccggtttagc
pSET152-BesB-F	taggatcgtctagaacaggaggccccatatggcaagctggagccacccgcagttcgaaaagggtgcaggtagcgccggctccccggcgga
pSET152-BesB-R	ccaagcttgggctgcaggtcgactctagaggatcctcagccggcgccgtcgaaggc
pSET152-BesC-F	taggatcgtctagaacaggaggccccatatggcaagctggagccacccgcagttcgaaaagggtgcaggtaccgacctgaacaccccggaatcc
pSET152-BesC-R	ccaagcttgggctgcaggtcgactctagaggatcctcacttgccgatgcccacgagttc
pSET152-BesD-F	taggatcgtctagaacaggaggccccatatggcaagctggagccacccgcagttcgaaaagggtgcaggtggcagtaaccggcaggaactcaaag
pSET152-BesD-R	ccaagcttgggctgcaggtcgactctagaggatcctcagtccttgatggcccgtgccg
p16b-BesBTrunc-5	aattttgtttaactttaagaaggagatataccatgtccgggggagctcggggcgg
p16b-BesBTrunc-10	aattttgtttaactttaagaaggagatataccatggcggcctgcagcctgcgg
p16b-BesBTrunc-15	aattttgtttaactttaagaaggagatataccatgcacaccgccgtcggccgccc
p16b-BesBTrunc-20	aattttgtttaactttaagaaggagatataccatgggccgcccccgtccccggctc
p16b-BesBTrunc-25	aattttgtttaactttaagaaggagatataccatggtgtccgtctcgatccccgacgtcg
p16b-BesBTrunc-30	aattttgtttaactttaagaaggagatataccatggtgtccgtctcgatccccgacgtcg
p16b-BesBTrunc-35	aattttgtttaactttaagaaggagatataccatgcccgacgtcgctgcggtcatcgg
p16b-BesBTrunc-40	aattttgtttaactttaagaaggagatataccatggtcatcggctacgagtcgggcgac
p16b-BesBTrunc-45	aattttgtttaactttaagaaggagatataccatgtcgggcgacgcggctacccgaa
p16b-BesBTrunc-50	aattttgtttaactttaagaaggagatataccatgacccgaagctgcatctcctgggg
p16b-BesBTrunc-R	tcctttcgggctttgttagcagccggatcctcgagtcagaggggggcgcagggcct
p32a-BesB-F	gccagcacatggacagcccagatctgggtaccatgaccaccgaggcatccggggagc
p32a-BesB-R	tggtggtggtggtgctcgagtgcggccgcaagctttcagaggggggcgcagggcctgc
p41a-BesB-F	tggctccggtgatgacgacgacaagagtcccatggaccaccgaggcatccggggagc
p41a-BesB-R	aattagtggtggtggtggtggtggtggtgctcgagtcagaggggggcgcagggcctgc
pEvol-Seq-1	aaccaaaccggtaaccccg
pEvol-Seq-2	ttttatttgatgcctggcagtt
pEvol-Seq-3	tgccgaactcagaagtgaaa
pEvol-Seq-4	gtcacactgcttccggtagt
pEvol-Seq-5	ggaggtaataattgacgatatga
pREP-CUA-F	ttttgtttaactttaagaaggagatatacatgagttagggagaagaacttttcactggag
pREP-F	ttttgtttaactttaagaaggagatatacatgagtaaaggagaagaacttttcactggagttgtccca
pREP-R	tggtggtggtggtggtgctcgagtgcggccgcaagcttcatttgtatagttcatccatgc
MetRSCUA-A-F	ctataggggaattgtgagcggataacaattcccctctatgactcaagtcgcgaagaaaat
MetRSCUA-A-R	catggccacggcgcgagcctgttctttgacatagcgaatagccagatcagccagcgccatgat
MetRSCUA-B-F	attcgctatgtcaaagaacaggctcgcgccgtggccatgaaacaggaaggccgcgatgc
MetRSCUA-B-R	gtgatgatgatgatgatgatgatgatggcccattatttcacctgatgacccggtttag
pEVOL-MetRSCUA-B-F	gccgttatacgttgtttacgctttgaggaatcccaatgactcaagtcgcgaagaaaat
pEVOL-MetRSCUA-B-R	gcgtagcgcatcaggcaatttagcgtttgaaacttatttcacctgatgacccggtttag
pEVOL-MetRSCUA-A-F	ccatacccgtttttttgggctaacaggaggaattaatgactcaagtcgcgaagaaaat
pEVOL-MetRSCUA-A-R	cgtttaaactcaatgatgatgatgatgatggtcgattatttcacctgatgacccggtttag
pB24-LeuRS-T252Y-A-F	atacccgtttttttgggctagcaggaggaattatgcaagagcaataccgcccggaaga
pB24-LeuRS-T252Y-A-R	Ctaccgccaggtaggtacaacccataaaatagtccgggcgggtagtgtaaac
pTrc33-BesB-F	gtgagcggataacaatttcacacgagctcggtaccaccaccgaggcatccggggagc
pTrc33-BesB-R	ccaaaacagccaagcttgcatgcctgcaggtcgactcagaggggggcgcagggcctgc

p16glns-BesB-F	tccggcgtagaggatcgagatctataagatcatacgccgttatacgttgtttacgctttgaggaatcccaatgaccaccgagg catccgg
p16glns-BesB-R	ggtatatctccttcttaaagttaaacaaaattatttctagtcagaggggggcgcagggcct
p16B-BesC-Fluor-F	gcggccatctagaagtgctttttcagggcccgcataatgtccatcactcaagaaacgtttcaatc
p16B-BesC-Fluor-R	cagcttcctttcgggctttgttagcagccggatccttggcgtcctccagcgttgagttt

A.1.1C. Synthetic gene sequences

BesB from *S.* sp. NRRL S-1448 (WP_030410683)

```
ATGACCACTGAGGCCAGTGGGGAATTGGGAGCCGCCTGTTCTTTGCGTCATACGGCAGTCGGTCGTCCGGT
ACCGGGTTCGCTGCACTCCGTAAGCGTCTCCATCCCAGATGTAGCAGCCGTAATTGGATACGAATCCGGGG
ACGCAGCGACCCGTTCCTGCATCTCTTGGGGGTACCCCCGCTTCCGTACTCACCCTTACGTGACGCGTGTC
GCAGGATTACTTGCGCAAGATGGTATTGGACCTGCCAATGGACTGCTGTTGACCCGTTCTGCCCGCGCTGC
GGGCGCAGCGGCGACCTATGCTGGACTTACTCCAGAAGCAGTTTTCGAGCGTAACGGTTTACACGGAGTGC
GTTTACCGGCGGACGGTCCTTCGGCCGCTCGCGCACGCGCTTATGTTCAACACACAGGAAGCCATCTTAGT
TCACGTGAGGCTGAAGACATTCTGTTGGATGCGGGACTTATTGAAAACCGTCAGGCCGAAGCCGCAGTGGA
AGAGGCTCCAGCAGAAGCTGTGCGTGCCGTCCTTGCCCAAGCCTATGGAGTTCGCGATGCTGCAGATGTTT
CCTTGCACAACTCTGGAATGAACGCTGTTGCTGCCGCGCTGGCAGCGGTCTCAGGCATTCAGCGCGAGAAT
GGCCGCCGTCGTTGGTTACAGTTAGGCTGGATCTTTTTTGACACGATGTCCTTACTTGAGAAGCGTGTGAT
TGACGTGGAGCATACGACCGTGCCAGACCCGTTCGACCTGAGTGAAATCGCTCGCGTGGCAGACGCACACG
CAGGCGAATTAGCTGGTATCATCGCGGAGGTTCCTTCCAATCCCAGCCTTCGCACACCTGATATTCAGCGC
TGCGCGAAATTGCCCACACGTGCAGGCTGTGCTTTAGTTGTAGATGCCACCATTGCAACACCATATAACGTA
GAAGTGCTGCCGTACGCCGATGTTGTCTGTGAAAGTCTTACTAAATATGCGACTGGCTCGGCAGATGTCCT
GATGGGTGCGGCTGTCGTCAATCCCGCGAGTTCCTTTGGACCTGACTTGCGCACGGAGTTGCGTCGTCACG
GTGATGAACCATACCATCGTGATACTGCACGCGTGGCCGCGCGCATCCGCGGTTATGCCGAACGTATGGAC
CGTGTAAACGCTAACGCTCTGGCCTTAGCTGCTTGTTTAGCTCGTCACGATAATGTGGTCCGTGGTGTCGG
ATGGGCATACGATGCAGCCTCGCGCGGCAATTACCGCAAGGTAGAGCGTCGCCCCGACGCCCCTGGTGGAC
TGTTGATGGTGGACTTGAAAGTGCCTCTGGAGCAAGTCTATGATCGTCTGGCAGTTGCTAAGGGCCCTTCC
TTTGGGGCCGAATTCACTATGGCGAGTCCCCAAATTTTCATTGCACACTACGACTTGCTGTCTACACCGGA
GGGTCGCGCAGCACTTCGTGCTCGCGGTTTGCACCGTGATATGCTGCGTATCTCTGTTGGAACTGAGGAGC
CCGAACGTATTGTTGAAACGTTTGAGCAAGCTCTGCGTCCCTTGTGA
```

BesB from *Opitutaceae bacterium TSB47* (WP_068769754.1)

```
GCGGCCATCTAGAAGTGCTTTTTCAGGGCcCgcatATGTCAGCCTTCAGCCCGCTTCCACTGGGCCAGCGC
ATCCCACCGTCAGTACATGGCGTCTCTTGTTCTCTGCCAACGATGCGCGATGTACGCGGTTACGAGGAGAA
GAATCCTGAAACACTGCGCCAAATGACTAGCGGTTATCCGCGTTTCGTAGTCCACCCTTTCTTGCAGCAGC
TTGCTACTCATCTGATTAAGGCAAAGAATCTGGATGGTCACACGTTGTGGCTTGCTAGTAGTGCTCGTATG
GCCGATCGCTTGGCTGCACATTTAGGAGCCGCACACGTTATCCGCATCACAGATGGTCCAATTCATGGCGT
GGCGCACCCAGAGAGTTCTGAACTTGCGTTGCGTGCAAAGAGCTTCTTACAGCACACAGGCGGGTTTTTAT
CTAGTCGCGCAGCGGAGGACCGTCTGGTTACCCACCGTGTGTTGGCAGCGGCTGCTGCAGAGCCGTTGTTT
CCCGGGGATGCGGCCGAGGAGGTTCGTCGTCAACTGGCCCCCGCCTTTCCGGGAGTCCCTGCAGAGCAACT
GCGTTTTGCAAACACAGGGATGAATGCTATTGACACGGCATTCCGTGCCGTCTCTGAGTTACAGGCCAGCC
GCGGACGTACGATCTGGATTCAGTTAGGCTGGGTTTATCTTGACACTATCGCACTGTTAAAAAAATTTACT
GCCTCCCCTGCTGACTACATTACCATCCCCGACGTCTTTGATCGTGCGGCCATCGCACGCATCTTTGCAGA
AAAGGCGGAGCGTATTGCTGGGGTGATCGCGGAGGTACCGACTAATCCACTTATCCAAACCCCTGATATTG
CAGCTCTTGCGAAGTTAGCACACCAAACGGGCGCTCGTGTAATTCTTGACCCGAGTATTAGCTCACCCTAT
TGTGTGGAAGTTCTTCCCCATGCTGACGTAGTGGTAACCTCCCTGACTAAGTATGCTGCAAATGAGGGGGA
CTTAGTAGCGGGGCTTGTAGTTGTGAATCCCGCCGGAGCCGATGCTGATCTTCTGTCGCGTGCTGTTGCTG
CTAAAATCGAACCATTATATGCGCGCGATTGGCGCGCCTTGCCGCGCAAGTCAGCCGCACCCCAGCAGTT
TTGGCACGTATCGAGGCAAGCGTACCCAAGGTTGTAGAGTTCTTGACTCGTCACCCGGCCGTCAAAGACGT
TTTTTGGGCCGAACACCCTGCTAGCCGCGACAATTATCGTCAAGTTGCACGCGCTCCGGGTGCCTGTGGTA
GCATGATTACTTTCACGCTGCGTGGTTCTGCCCCGGCTATGGAGCGCTTCTATGATCGCGTTCGTTTACCG
AAGGGTCCATCGTTCGGCATTACAACAACTTTACTGTGCCCCTTTATGTATCTGGCACATTATGACTTAGT
TACGACTGCAGCGGGTCGCGCGGAATTGGCCCGTTATGGACTTGACCCGGAATTACTGCGCCTTAGTGTAG
GATGTGAATCGGTTGATGATTTGATTGCGACGCTTGCCGAGGCGTTGGCTTAGGGATCCGGCTGCTAACAA
AGCCCGAAAGAAA
```

BesB from *Leptospira kirschneri* (WP_020763855.1)

```
TCGGGATCGAGGAAAACCTGTATTTTCAGGGCATGTTAAAAGAAATCAATGAACCACACAGAACACTTTGC
GGAGAAAGAATTCCATTTGAGAACATCCACGCGGTTTCCGTAAGTCTTCCACAACTGTCAGACGTAATTGG
TTACGAAGAAAAAGAACAGAAACTCTTTCTAGATTGAAGTCCGGTTATCCTAGGTTTGTCGCACATTCTT
ATATAGCCAGAATTTTAGACTACAACAGAGATACAAAAAAAATCAACACACCTCAGTTTATCGTTTCTTCC
AAAAAAGCTGCAAACTGGATCGTACAAAAATTCTCCATTCAAAACTTTGAATTATTGAAGACGAAGGAAT
TACAACCTTAGTCGTTTCAGATTTAAAAGATTTAGAAAAAGAAATCCTTTCTTTCATTCAACACACTGGAT
GTCTTGCTTCTTCTAGAATGGCGGAAGACTTTCTTTTAAAAAAAGGAATTCTACAAGAAATTTTTAGGGAG
AAAGTTGAAAAAGAAATCCGGTCGATAAAATTCTTTCTACTCTATCTTCTTTTTACGGAGATCTAAATCC
TGAGATACTTCTTTCCGTTTCCGGAATGAACGGCGTCTATTCTGCTTTCGAATCTTTGAGCCAGATTCAGC
GGAAAAAAGGAAAAACCATTTGGGTTCGTTTGGGTTGGTTATATGTTGATAATATTAGAATATTAGAAAAA
TATACAAAAGATTCTTATGTAATCCACAACGCCACCGATCTGGAGGAATTGGAAAGATTTTTAAAACAAAA
TTCAGAACAAGTCGCTGGAATTATCACGGAATGTCCCACCAATCCACTTCTTTTGGTTCCGGATTACAAAA
AACTCAAATCTATTGTGGATCATTATAAAATTCCTTTGGTCGTGGATATTTCCGTGGCGGGATCGGCAATA
ATTAACATTCTTCCTTACGTAGATGTAGTCGTGGAAAGCCTGACCAAATTCGCCTGTGGAAACGGCGACCT
TATGATGGGGGCCGTGGTATTGAATAAAAATTCAAGTTGGTTCTCCGAAATTCTTCCGTTATGCAAAGAAT
GGATTGAAAAACCATATTTCAGAGACTGTGAAAGACTAGCTTATGAAATCAAAGATTATGAAAAAAGGGGTT
TTTAAAATCAGTGAGAATGTTAAAAAACTTGCAGATTTTTTTTCAAAACAACCCGGTATTAAAAAAGTTTT
TTGGACAGGGTCTGCAGATTCCTTTAAAAATTTCCAAAAAATTACCAGACTTCCAAATATACATTCTGGAG
TTTTATCGATTGAACTTTCAACTCCATTAGAAAAGTTTTATGATCGACTTACTTTATTAAAAGGACCAAGT
TTTGGCACCGAATTCACTCTCAACATGCTTTATGTTTATCTAGCTCATTACGATTTAGTAATTCAAGATGA
AGGAAGAAAATTTTTAAAAGAAACGGGCTCGATCCGAATTTAATCCGAATTTCCGTTGGGACTGAAGATC
CAGAGCTTTTGATTGAGGAGTATAAAAAGGCGTTGGAAGTTTAAAATTCGAGCTCCGTCGACAAGCTTGCG
GCCGCACT
```

BesB from *Verrucomicrobia bacterium (*RFC49611.1)

```
GATCGAGGAAAACCTGTATTTTCAGGGCATGAGCGCCGGACAACACATCCCCCTAGGCCAAGCCATACCCG
ACTCCCTACACGCCGTGTCAGTGAGCATTCCGACCATGACTGACGTCGTCGGATACGAGGAGAAAAAATCG
GACACCATGGCAAAGCTGACCTCCGGCTACCCACGCTTCGTCCTACACACCTGCCTCAGAGAAATCGAAGC
CCACTGGCAAAAACTCTTCGAAACGCCCAACCATTCGATCTGGCCCACCTCCTCCGAGAGCATGGCCAAAC
GGCTCCAAAAACACCTCGCCCCCACGCCCAGCAAATTCCTCAAACATCGAGGCGTATCTGGCGTCCGGATA
CAGACTGACCCAGAACTCAACCAGCAGGCAAAGAGCTTCCTCCAACACGTAGGCGGCTTCCTCTCATCCCG
CCAAGCCGAGGACTACCTCGTCGCCGAAGGACTGCGAGGAGAGATCCAGCCGGAAAACAATTTCGAAGGCG
ACGCCCAAGCGGAAATCCTCAAAACGCTCACTCCCCTATACGGAGTCGACGAGGATTCTATCGTTCTGTCC
AACACCGGCATGAACGCCTTCTTCGCCGCCTTCGAAGCCGTCCGCTCCATACAATCCCCCAAAGGACGCGA
CTCTTGGATCAAGCTAGGCTGGCTCTACACGGACACCATGCACATCCTCGACAAGCTCAGCGGAAGCCACG
CCGCAAATGTCGAAATCCTAGACATCTTCGATCTCGACGCTCTTGAGGCAGCCCTTGACGAACGCGGCGAT
ACCATCGCAGGCATCGTCACCGAAGCCCCTACCAATCCGCTCATCCAAACGATGGACCTCGATCGTATCCA
CGAAATAGCGACAAAGCACGGAGCAATTTTTATCGTCGACCCTACTGTTGCCTCCCCAGCCAACATCGACA
TCACACCCCACGCGGACATCATCGTAAACAGCCTTACAAAATACGCCGCCAACGAAGGCGACGTCATGCTC
GGAGCAACCGCCGTGTCCCATTTCTGTCCACATCGAAACGAGCTCCTCGCCTTCCTGCGAGACGATATCGA
GCCACCTTACCCACGAGACCTCGCCCGCCTAGCGTACCAACTTCCCGGATACGGAAGCCTCGTTCAAAAAA
CGAACCAAACCACAGCCAAAGTCGTCGCCTTTCTCGAGTCGCACCCGAAAGTGAAAGCCGTCCACTGGGCA
AAAAGCGAAAGGTCCCGAGCAAACTTCGAGAAAATCGCCCGCGCCCCCGATTCCGTCGGCGGACTCATATC
CTTCGAATACGACGGTCCCCTCGCACACTTCTACGACGCCACTCCCCTTCCCAAAGGCCCCAGCTTCGGAA
TGAGCACGACCCTCCTCTGCCCCTACATCTACCTCGCCCACTACTCCATCGTCACTTCAAAATCGGGACGC
GAAACCCTCGCCCGCGCAGGCATCAGCCCCGAGCTCCTCCGCCTCGCTATCGGAGCGGAGTCAGCAGAGGA
AATCATCGCCGCTCTCAAACAAGGACTAGACGCCTAAAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCAC
TCGAGATCAAACG
```

Synthetic oligo for cloning tRNA^{Met}CUA into pEVOL (pEVOL-Met-tRNA-CUA)

```
AGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTT
CTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCG
CATTGGCTACGTAGCTCAGTTGGTTAGAGCACATCACTCTAAATGATGGGGTCACAGGTTCGAATCCCGTC
GTAGCCACCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAA
TTTGCTTTCGAATTTCTGCC
```

Table A1.2 Strains, plasmids, oligonucleotides, and sequences used in discovery and characterization of BesG. (A) Strains and plasmids used in this study of BesG. (B) Oligonucleotides used for plasmid construction, genomic library construction, and screening/sequencing. (C) Sequences of synthetic genes.

A1.2A. Strains and plasmids

Strain	Description	Source
BL21 Star (DE3)	*F-ompT hsdSB (rB-, mB-) gal dcm rne131* (DE3)	Thermo-Fisher
BL21 Star (DE3) TUVW	*F-ompT hsdSB (rB-, mB-) gal dcm rne131* (DE3) *thrT*[A73U] *thrU*[A73U] *thrV*[A73U] *thrW*[A73U]	This study
GM272	*F- dam-3 dcm-6 hsdS2l metb1 lacYl or Z4 galK2 ga/T22 mtl-2 tonA2or A31 tsx-1 or -78 supE44 (thi-1)*	ATCC
S. cattleya	NRRL 8057 (ATCC 35852)	ATCC
S. coelicolor	M1152	JIC
S. cattleya ΔbesG	Δ*besG*::Am[R]-OriT	This study
S. cattleya ΔbesG+besG	Δ*besG*::Am[R] *attB*[ΦC31]::*ermEp**-*besG*-Hyg[R]	This study
S. coelicolor besG	*attB*[ΦC31]::*ermEp**-*besG*-Hyg[R]	This study

Plasmid	Description	Source
pUZ8002	*tra*, *neo*, Km[R], RP4	[15]
pIJ773	*lacI*, Cb[r], Am[R], OriT, ColE1	JIC
pIJ773-BesG	*besG* flanking regions (2 kB US/DS)-Am[R]-OriT, *lacI*, Cb[R], ColE1	This study
pSET152	Hyg[R], *int(φ31)*, OriT, pUC18	JIC
pSET152-BesG	BesG (ErmEp*), Hyg[R], *int(φ31)*, OriT, pUC18	This study
pET16b-His10-BesA	His10-BesA (T7), *lacI*, Cb[R], ColE1	This study
pET16b-His10-BesE	His10-BesE (T7), *lacI*, Cb[R], ColE1	This study
pET16b-His10-BesG	His10-BesG (T7), *lacI*, Cb[R], ColE1	This study
pET16b-His10-COLI-TRS	His10-ThrS (*from E. coli*, T7), *lacI*, Cb[R], ColE1	This study
pET16b-His10-SCAT-TRS	His10-ThrS (from *S. cattleya*, T7), *lacI*, Cb[R], ColE1	This study
pSV272.1-His6MBP-BesG	His6MBP-BesG (T7), *lacI*, Km[R], f1, ColE1	This study
pETDUET-1	*lacI*, f1, Cb[R], ColE1	Novagen
pETDUET-BesA.BesE	BesA (T7) and BesE (T7) from *S.* sp. NRRL-1448, *lacI*, Cb[R], ColE1	This study
pET28a-Ggt	Ggt from *E. coli* (T7), *lacI*, Km[R], ColE1	This study
pEAGgt	BesA and BesE from *S. cattleya* and Ggt *from E. coli* (T7), *lacI*, Cb[R], ColE1	
pβes	BesD (T7) and BesC (T7) from *P. fluorescens*, MBP-sBesB (T7) from *S.* sp. NRRL S-1448, BesA and BesE from *S. cattleya* and Ggt *from E. coli* (T7), *lacI*, Cb[R], ColE1 *lacI*, Cb[R], ColE1	This study
pBAD24-PraRS	PraRS (araB), Cb[R], f1, ColE1	This study

A1.2B. Oligonucleotide sequences

Name	Sequence
p16b-TRS-COLI-F	ggccatctagaagtgctttttcagggcccgcatatgcctgttataactcttcctgatggc
p16b-TRS-COLI-R	cagcttcctttcgggctttgttagcagccggatccttattcctccaattgtttaagactgcgg
p16b-TRS-SCAT-F	ggccatctagaagtgctttttcagggcccgcatatgtctcagatccgtgtggtcatca
p16b-TRS-SCAT-R	cagcttcctttcgggctttgttagcagccggatcctcacacctggacgcggcgctc
p16b-BesA-F	ggccatctagaagtgctttttcagggcccgcatatgactgatcccggttcgacgaagatcctga
p16b-BesA-R	cagcttcctttcgggctttgttagcagccggatcctcaggcgctgggggtacggcc
p16b-BesE-F	ggccatctagaagtgctttttcagggcccgcatatgtccggcacgacacaccacca
p16b-BesE-R	cagcttcctttcgggctttgttagcagccggatccggaccagacgaccagttcgccg
p16b-BesG-F	ggccatctagaagtgctttttcagggcccgcatatgacccagcagcggcagatct
p16b-BesG-R	cagcttcctttcgggctttgttagcagccggatcctcagccgtccaccgggacgtgt
pETDuet-BesA-F	aataattttgtttaactttaagaaggagatataccatgactgatcccggttcgacgaa
pETDuet-BesA-R	cgacctgcaggcgcgccgagctcgaattcggatccggcgctgggggtacggcccgg
pETDuet-BesE-F	tagtatattagttaagtataagaaggagatatacatatgtccggcacgacacaccacca
pETDuet-BesE-R	tttctttaccagactcgagggtaccggaccagacgaccagttcgccg
pEAGgt-E-F	attttgtttaactttaataaggagatataccatgtccggcacgacacaccacc
pEAGgt-E-R	gcagcagcggtttctttaccagactcgagggtacctcaggaccagacgaccagttcgc
pEAGgt-A-F	tcacggccggcggcgaactggtcgtctggtcctgagagagttcttttattaaggaggttatattatgactgatcccggttcgacgaagat
pEAGgt-A-R	gcagcagcggtttctttaccagactcgagggtacctcaggcgctgggggtacggcccg
pEAGgt-Ggt-F	aggcgttgctgaccgggccgtaccccagcgcctgatcacaacacgagaagattaaggaggtattaaatgataaaaccgacgtttttacgc
pEAGgt-Ggt-R	gcagcagcggtttctttaccagactcgagggtaccttagtaccccgccgttaaatcatccaccga
pET28a-Ggt-F	acagcagcggcctggtgccgcgcggcagccatatgatgataaaaccgacgtttttac
pET28a-Ggt-R	ccgcaagcttgtcgacggagctcgaattcggatccttagtaccccgccgttaaatcatc
pSV272-MBP-BesG-F	acctcgggatcgaggaaaaacctgtattttcagggcatgacccagcagcggcagatct
pSV272-MBP-BesG-R	tttgatctcgagtgcggccgcaagcttgtcgacggagctcgaatttcagccgtccaccgggacgtgt
pIJ773-BesG-A-F	ccgcggtggcggccgccggggcaccggcccccaccccgct
pIJ773-BesG-A-R	ccggggggatccactagtccggaggccgcctcaccggggcg
pIJ773-BesG-B-F	cccggaatatcaagcttagagttcaccccaaaagtcgcgtcacgcg
pIJ773-BEsG-B-R	gaacaaaagctggtaccagacgctgctgcggacccggtccg
pSET152-BesG-F	ggatcgtctagaacaggaggccccatatggcaagctggagccacccgcagttcgaaaagggtgcaggtatgacccagcagcggcagatct
pSET152-BesG-R	ccaagcttgggctgcaggtcgactctagaggatcctcagccgtccaccgggacgtgt
Seq-pIJ773-A-F	accatgattacgccaagctc
Seq-pIJ773-Acc-R	tgcctttgctcggttgatcc
Seq-pIJ773-OriT-F	ggggtcattatagcgatttt
Seq-pIJ773-B-R	gaagaaagcgaaaggagcgg
HA-thrT-A-F	aagtgtgctaatctgccctccgtt
HA-thrT-A-R	tgagaaaagaagtggagctgataggcagattcgaactgc
HA-thrT-B-F	tccacttcttttctcatccctgttttttccttctgtttattgcat
HA-thrT-B-R	gatgtacggaacgcctacctga
HA-thrU-A-F	agaagttcagcaaacgtgacaaagg
HA-thrU-A-R	gcgacacgatgaagaaacagcagaacggagggcagattagcac
HA-thrU-B-F	gtgctaatctgccctccgttctgctgtttcttcatcgtgtcgc
HA-thrU-B-R	gtcagcgttgttttaccatggtcaacg
HA-thrV-A-F	cggacagatttcaggtggtca
HA-thrV-A-R	tccactttttaggttaaagttcggcagattagaaaagaatttgtcttgcggcagtagcgcggtggtc
HA-thrV-B-F	aagacaaattcttttctaatctgccgaactttaacctaaaaagtggagctgatacccagagtcgaactgggg
HA-thrV-B-R	acgagaggaccggagtggacg
HA-thrW-A-F	aacactccgatgcgatcctgac
HA-thrW-A-R	tgtaactacttgatttaaatggagccgataataggagtcgaacctac
HA-thrW-B-F	tccatttaaatcaagtagttcaccatatttaaatacaccacgtttcc
HA-thrW-B-R	cacctccggggggaaataccca
LR-Cm-thrU-F	aggttcgaatccttcccccaccaccaatttcggccacgcgtgatgtccggcggtgcttttt
LR-Cm-thrU-R	tcgagccggtaagcgaacttatcgtctcgggctacgccatttacgccccgccctgccac

pTarg-F	actagtattatacctaggactgagctagctgtcaag
pTarg-T-R	tcctaggtataatactagtaacagaaggaaaaaacaggggttttagagctagaaatagc
pTarg-V-R	tcctaggtataatactagtagattagaaaagaatttgtcgttttagagctagaaatagc
pTarg-W-R	tcctaggtataatactagtacgtggtgtatttaaatatggttttagagctagaaatagc
pTarg-Cm-R	tcctaggtataatactagttttgcccatggtgaaaacgggttttagagctagaaatagc
tRNA-Express-F	ccccgaaaagtgctagtggtgc
tRNA-SCAT-UCCA-R	tggagccccaatacggaatcgaac
tRNA-SCAT-ACCA-R	tggtgccccaatacggaatcga
tRNA-COLI-AGCACCA-R	tggtgctgataggcagattcgaact
tRNA-COLI-AGCUCCA-R	tggagctgataggcagattcgaact
tRNA-COLI-GGCUCCA-R	tggagccgataggcagattcgaa
pEAGgt-E-F	attttgtttaactttaataaggagatataccatgtccggcacgacacaccacc
pEAGgt-E-R	gcagcagcggtttctttaccagactcgagggtacctcaggaccagacgaccagttcgc
pEAGgt-A-F	tcacggccggcggcgaactggtcgtctggtcctgagagagttcttttattaaggaggttatattatgactgatcccggttcgacgaagat
pEAGgt-A-R	gcagcagcggtttctttaccagactcgagggtacctcaggcgctgggggtacggcccg
pEAGgt-Ggt-F	aggcgttgctgaccgggccgtaccccagcgcctgatcacaacacgagaagattaaggaggtattaaatgataaaaccgacgtttttacgc
pEAGgt-Ggt-R	gcagcagcggtttctttaccagactcgagggtaccttagtaccccgccgttaaatcatccaccga
pBes-A-F	aataattttgtttaactttaagaaggagatataccatgaactacgtacttgatgaggctcgc
pBes-A-R	ttctgttcgacttaagcattatgcggccgcaagcttgccaatccggatatagttcctcctttca
pBes-B-F	tatattagttaagtataagaaggagatatacatatgatgtccggcacgacacaccacc
pBes-B-R	cgcagcagcggtttctttaccagactcgagggtaccttagtaccccgccgttaaatcatccac

A1.2C. Synthetic gene sequences

BesG from *S. cattleya* – codon optimized (CCB72157.1)
CAACAACCTCGGGATCGAGGAAAACCTGTATTTTCAGGGCATGACTCAACAACGTCAGA
TTTACCTTACTGATACCTATGCGTACGAGGCCTCCACTCGTGTTGTGAGCGCGGAACGC
GGACCAGACACGGTTCGCATCGCCTTAGCGGATAACATTTTCCATCCACAGGGTGGGGG
ACAACCCGACGACCGCGGATGGGTAGATCAAGTCGAGGTCCGCCCAGTACGTGACACGG
GCCCAGGGCTTGTTCATCTGACTTGTCCGGCAGATGCAGCCTGGCCTCGTGAGGCGGTG
GCAGTAGGGACCACTGTCACTAGTCGTATTGACCCCGCCCTGCGTCGCCTTCATGCCGC
ACTGCACACTGCAGGTCACCTTGTTGATGCGCTTATTGACGAGTGGGGCTACCGTCATG
TAGGTAGTAATCACTTCCCCGGACAGGCCCGCGTTGAGTATGACCTTGGGGGTCTGGAC
TGCGACAAAGATGCATTAGCGGAAGGGTTGACAGAACGCCTTGGAAAACCACTTGCGGA
AGCACTGCCAGTAATCCCAGGTGAACGTGACGGGCGTCGTACAATTACCATTGAAGGTT
ATGCCACGGAGTTCTGTGCTGGCACTCACGTCCCTGACCTGTCCCTGCTTACCGGTGTT
GCTATTCGCAGTGTGAAGGTAAAGGGTGGGCGCCTTAAAGTCGGCTACACGGCCGAGCA
TGTGCCAGTAGACGGTTGAAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGA
T

BesG C189A from *S. cattleya* – codon optimized
CAACAACCTCGGGATCGAGGAAAACCTGTATTTTCAGGGCATGACTCAACAACGTCAGA
TTTACCTTACTGATACCTATGCGTACGAGGCCTCCACTCGTGTTGTGAGCGCGGAACGC
GGACCAGACACGGTTCGCATCGCCTTAGCGGATAACATTTTCCATCCACAGGGTGGGGG
ACAACCCGACGACCGCGGATGGGTAGATCAAGTCGAGGTCCGCCCAGTACGTGACACGG
GCCCAGGGCTTGTTCATCTGACTTGTCCGGCAGATGCAGCCTGGCCTCGTGAGGCGGTG
GCAGTAGGGACCACTGTCACTAGTCGTATTGACCCCGCCCTGCGTCGCCTTCATGCCGC
ACTGCACACTGCAGGTCACCTTGTTGATGCGCTTATTGACGAGTGGGGCTACCGTCATG
TAGGTAGTAATCACTTCCCCGGACAGGCCCGCGTTGAGTATGACCTTGGGGGTCTGGAC
TGCGACAAAGATGCATTAGCGGAAGGGTTGACAGAACGCCTTGGAAAACCACTTGCGGA
AGCACTGCCAGTAATCCCAGGTGAACGTGACGGGCGTCGTACAATTACCATTGAAGGTT
ATGCCACGGAGTTCGCTGCTGGCACTCACGTCCCTGACCTGTCCCTGCTTACCGGTGTT
GCTATTCGCAGTGTGAAGGTAAAGGGTGGGCGCCTTAAAGTCGGCTACACGGCCGAGCA
TGTGCCAGTAGACGGTTGAAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGA

S. cattleya tRNAThr 5'-UCCA-3' acceptor stem
AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCTAGCC
CCGCGAAATTAATACGACTCACTATAGCCCCAATAGCTCAGTCGGCAGAGCGTCTCCAT
GGTAAGGAGAAGGTCTACGGTTCGATTCCGTATTGGGGCTCCA

S. cattleya tRNAThr 5'-ACCA-3' acceptor stem

AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCTAGCC
CCGCGAAATTAATACGACTCACTATAGCCCCAATAGCTCAGTCGGCAGAGCGTCTCCAT
GGTAAGGAGAAGGTCTACGGTTCGATTCCGTATTGGGGCACCA

E. coli tRNAThr 5'-AGCACCA-3' acceptor stem

AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCTAGCC
CCGCGAAATTAATACGACTCACTATAGCTGATATAGCTCAGTTGGTAGAGCGCACCCTT
GGTAAGGGTGAGGTCGGCAGTTCGAATCTGCCTATCAGCACCA

E. coli tRNAThr 5'-AGCUCCA-3' acceptor stem

AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCTAGCC
CCGCGAAATTAATACGACTCACTATAGCTGATATAGCTCAGTTGGTAGAGCGCACCCTT
GGTAAGGGTGAGGTCGGCAGTTCGAATCTGCCTATCAGCTCCA

E. coli tRNAThr 5'-GGCUCCA-3' acceptor stem

AAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCTAGTGGTGCTAGCC
CCGCGAAATTAATACGACTCACTATAGCCGATATAGCTCAGTTGGTAGAGCGCACCCTT
GGTAAGGGTGAGGTCGGCAGTTCGAATCTGCCTATCGGCTCCA

Table A1.3. Genbank accession numbers for *bes* biosynthetic proteins from various *Streptomyces* spp*.*

| Organism	Accession	Gene
Streptomyces cattleya	WP_014151494	*besA*
Streptomyces cattleya	WP_014151495	*besB*
Streptomyces cattleya	WP_014151496	*besC*
Streptomyces cattleya	WP_014151497	*besD*
Streptomyces cattleya	WP_014151498	*besE*
Streptomyces cattleya	WP_014151493	*besF*
Streptomyces cattleya	WP_014151582	*besG*
Streptomyces catenulae	WP_030285995	*besA*
Streptomyces catenulae	WP_030285993	*besB*
Streptomyces catenulae	WP_030285991	*besC*
Streptomyces catenulae	WP_051739772	*besD*
Streptomyces sp. NRRL S-1448	WP_030410682	*besA*
Streptomyces sp. NRRL S-1448	WP_030410683	*besB*
Streptomyces sp. NRRL S-1448	WP_030410684	*besC*
Streptomyces sp. NRRL S-1448	WP_030410685	*besD*
Streptomyces sp. NRRL S-1448	WP_030410686	*besE*
Streptomyces sp. NRRL S-1448	WP_030410681	*besF*
Streptomyces sp. NRRL S-1448	WP_051822990	*besG*
Streptomyces lavenduligriseus	WP_037703010	*besA*
Streptomyces lavenduligriseus	WP_037703017	*besB*
Streptomyces lavenduligriseus	WP_030791984	*besC*
Streptomyces lavenduligriseus	WP_030791981	*besD*
Streptomyces sp. NRRL S-31	WP_030744375	*besA*
Streptomyces sp. NRRL S-31	WP_037856819	*besB*
Streptomyces sp. NRRL S-31	WP_030744380	*besC*
Streptomyces sp. NRRL S-31	WP_030743980	*besD*
Streptomyces achromogenes	WP_030612442	*besA*
Streptomyces achromogenes	WP_037655359	*besB*
Streptomyces achromogenes	WP_030612452	*besC*
Streptomyces achromogenes	WP_030612455	*besD*
Streptomyces alboverticillatus	WP_086569404	*besA*
Streptomyces alboverticillatus	WP_086569406	*besB*
Streptomyces alboverticillatus	WP_086569408	*besC*
Streptomyces alboverticillatus	WP_086569410	*besD*
Streptomyces alboverticillatus	WP_086569412	*besE*
Streptomyces alboverticillatus	WP_086569402	*besG*

**Appendix 2: Supplementary Figures**
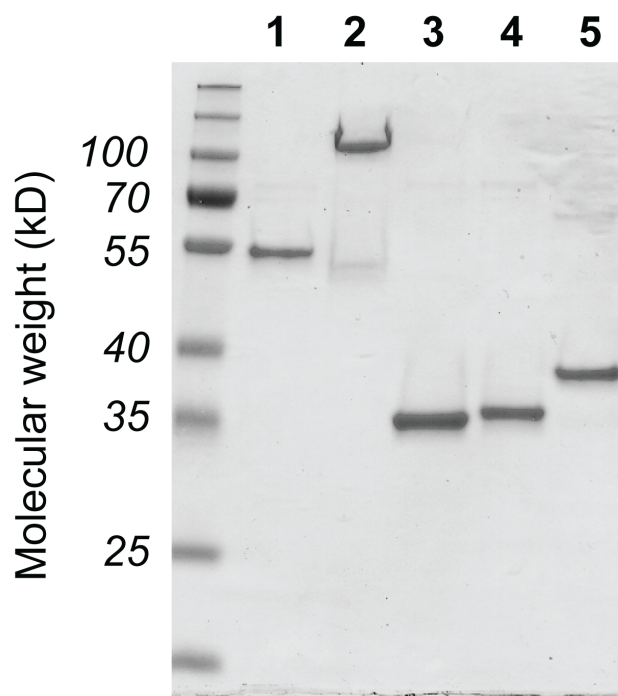
Figure A2.1. SDS-PAGE of purified enzymes used for biochemical assays and *in vitro* reconstitution of pathway. **1**, His_{10}-BesA; **2**, His_6MBP-BesB; **3**, BesC; **4**, His_{10}-BesD; **5**, His_{10}-BesE. All proteins are from *S. cattleya* except for BesB, which is from *S.* sp. NRRL S-1448.
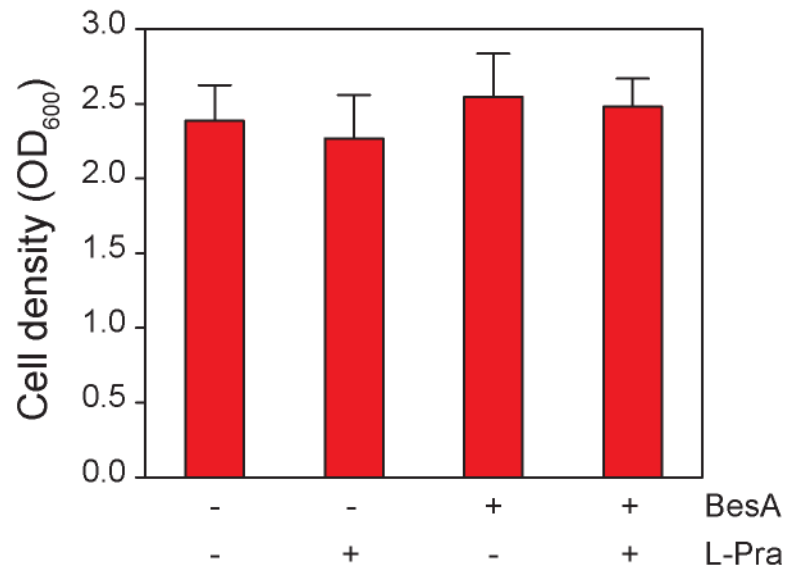
Figure A2.2. The effect of L-Pra and BesA on *E. coli* growth. *E. coli* BL21 Star (DE3) was cultured overnight in M9 medium in the presence or absence of 2 mM L-Pra. Cell growth was characterized by measuring the absorbance at 600 nm (n = 5).
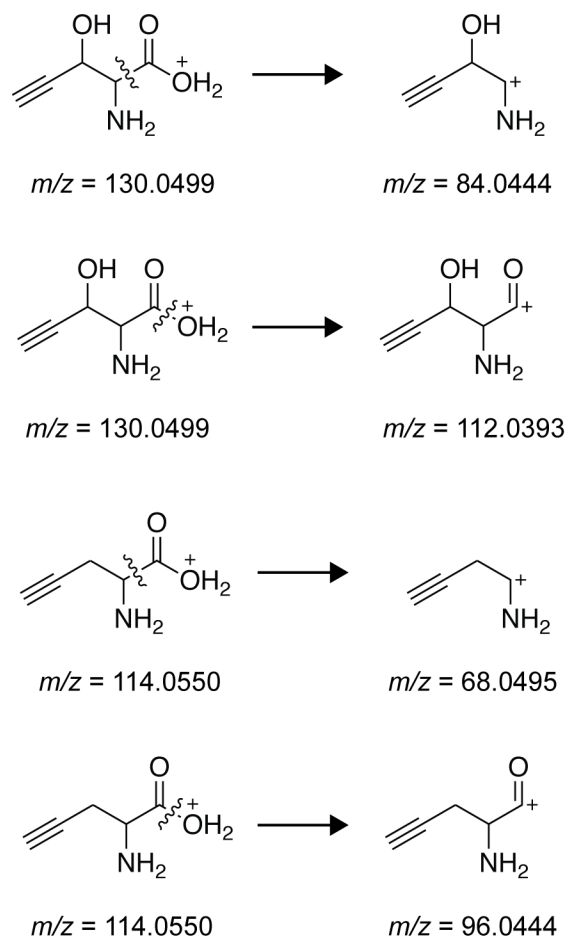
Figure A2.3. Useful MRM transitions for the detection of βes and Pra on a tandem MS-MS instrument. Multiple reaction monitoring (MRM) transitions can be used to directly detect βes and Pra if high resolution MS is not available. The transitions shown were found to be the most abundant species on QQQ when fragmentor voltage was set to 30V and collision energy set to 0V. Detailed MS settings can be found in Chapter 4 methods.
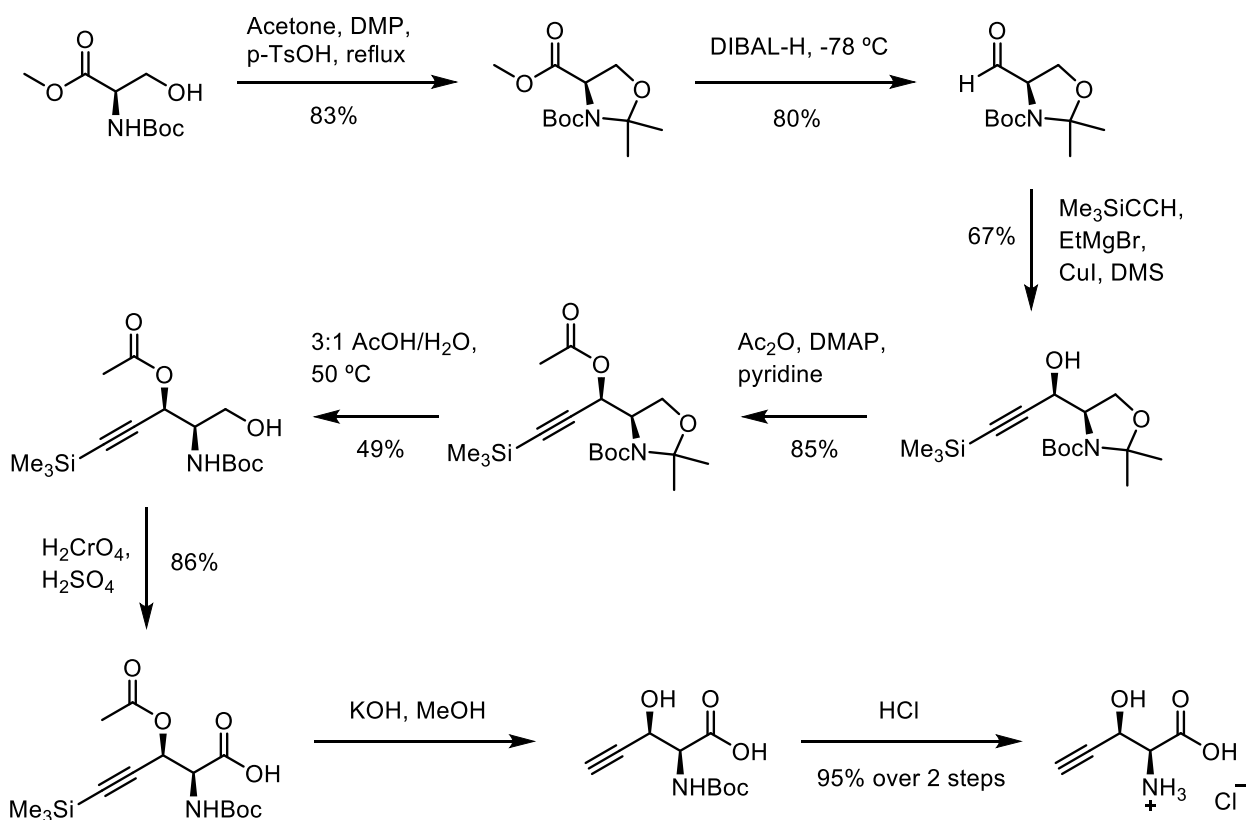
Figure A2.4. Scheme and yield for the synthesis of βes.
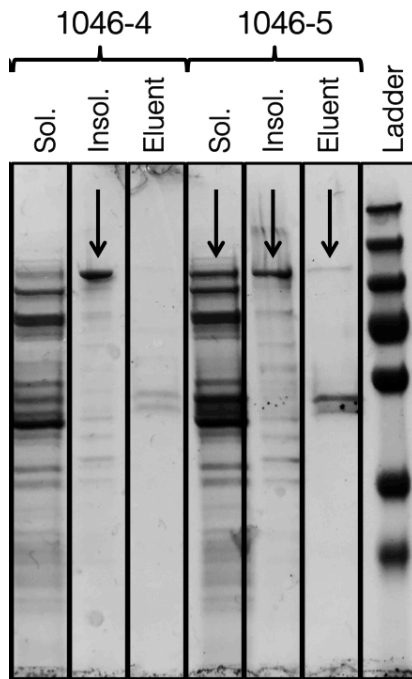
Figure A2.5. SDS-PAGE of a typical purification for MBP-BesB variants from S. sp. NRRL S-1448 (1046-4), S. catenulae (1046-5). MBP-BesB mostly found in insoluble fraction.