UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Investigating the organofluorine physiology of Streptomyces cattleya: Regulation of transcription and control of mistranslation

Permalink

https://escholarship.org/uc/item/1cf2h1qj

Author

McMurry, Jonathan Lewis

Publication Date

2017

Peer reviewed|Thesis/dissertation

Investigating the organofluorine physiology of
Streptomyces cattleya:
Regulation of transcription and control of mistranslation

by

Jonathan Lewis McMurry

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michelle C. Y. Chang, Chair
Professor Dave F. Savage
Professor Wenjun Zhang

Spring 2017

Investigating the organofluorine physiology of *Streptomyces cattleya*:
Regulation of transcription and control of mistranslation

by Jonathan Lewis McMurry

Abstract

Investigating the organofluorine physiology of *Streptomyces cattleya*:

Regulation of transcription and control of mistranslation

by

Jonathan Lewis McMurry

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Michelle C. Y. Chang, Chair

The soil-dwelling bacterium *Streptomyces cattleya* produces the antibiotics fluoroacetate and fluorothreonine, and has served as a model system for the discovery of naturally occurring fluorine-selective biochemistry. While fluoroacetate has long been known to act as an inhibitor of the TCA cycle, the fate of the amino acid fluorothreonine is still not well understood. Here, I show that while fluorothreonine is a substrate for translation, this activity is averted in *S. cattleya* by the activity of two conserved proteins. The first, SCAT_p0564, acts *in vitro* and *in vivo* as a fluorothreonyl-tRNA selective hydrolase, while the second, SCAT_p0565, is proposed to be a fluorothreonine exporter. Additionally, overexpression of SCAT_p0564 in the model strain *Streptomyces coelicolor* M1152 confers resistance to fluorothreonine, suggesting that the antibiotic activity of this compound is related in part to its ability to enter the proteome. The ability of SCAT_p0564 to selectively hydrolyze fluorothreonyl- over threonyl-tRNA is striking, given that these macromolecular substrates differ by a single atom. In order to understand the basis of this selectivity, I have solved the crystal structure of this enzyme, and characterized its ability to act on the related substrate chlorothreonyl-tRNA. I also have also begun to elucidate the regulatory architecture of the organofluorine biosynthesis locus, with the aim of understanding how this unusual process is controlled. I find that transcription of the fluorinase is driven by the master regulator FlG, while FlF is an amino acid binding transcription factor that may be required for full expression of the fluorothreonine transaldolase. These findings expand the range of known naturally occurring fluorine biochemistry, and a represent a step towards the rational discovery of new organofluorine metabolism in nature.

Table of Contents

Chapter 1: Introduction

Chapter 2: A fluorothreonine-selective aminoacyl-tRNA deacylating protein maintains translational fidelity in the organofluorine producer *Streptomyces cattleya*

Chapter 3: Structural and biochemical investigation of tRNA editing   proteins with activity towards halogenated threonine analogues

Chapter 4: Understanding the transcriptional regulation of the organofluorine biosynthesis in *Streptomyces cattleya*

Appendices

List of Figures, Schemes, and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

ARS   aminoacyl-tRNA synthetase
ATP   adenosine triphosphate
BGC   biosynthetic gene cluster
d   day
bp   base pairs
BSA   bovine serum albumin
Cl-SAD   chlorothreonine-associated second additional domain
Da   daltons
DTT   DL-dithiothreitol
$FLAG_3$ the peptide sequence DYKDHDGDYKDHDIDYKDDDDK
FThr   4-Fluorothreonine
h   hour
HEPES   4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC   high pressure liquid chromatography
kDa   kilodaltons
LB   lysogeny broth
LCMS/MS   liquid chromatography tandem mass spectrometry
min   minute
MWCO   molecular weight cutoff
NADH   nicotinamide adenine dinucleotide
NiNTA   nickel nitrilotriacetic acid
QQQ   triple quadrupole tandem mass spectrometer
Q-TOF   quadrupole-time of flight tandem mass spectrometer
RT   retention time
PMSF   phenylmethanesulfonyl fluoride
RCF   relative centrifugal force
tRNA   transfer RNA
Thr   threonine
Tris   tris(hydroxymethyl)aminomethane
TRS   threonyl aminoacyl-tRNA synthetase

Acknowledgments

Chapter 1: *Introduction*

1.1 Fluorine in biology

The use of fluorine has become a dominant feature of modern medicinal chemistry. Its unique combination of compact size and potent electronegativity lends it the ability to make sterically conservative alterations to a target compound that can effect major changes in activity, bioavailability, or pharmacokinetics [1]. As a result, 51 new chemical entities that entered phase II clinical trials between 2011 and 2016 contain this element [2]. In light of the obvious utility of this element in the design of bioactive compounds, its near-absence from biological systems is all the more striking [3]. While organochlorine and organobromine natural products are common in nature, a mere 22 biogenic organofluorines have been identified (*Figure 1.1*), likely arising from two underlying fluorination strategies [4–6]. We hope to discover new organofluorine-related biochemistry in order to better understand how enzymes have evolved to recognize fluorinated substrates, and to guide the discovery of novel biosynthetic pathways.

In 1943, it was recognized that fluoroacetate occurs in nature, and is the active component of *Dichapetalum cymosum*, a toxic plant that frequently results in livestock deaths in its native South Africa [4]. Subsequently, the production of fluoroacetate and related compounds was recognized to be a widespread defensive strategy, as it is present in plants from Africa, Australia, and South America [7]. While plant-derived organofluorines do not display a high degree of structural diversity, they are notable for their ability to accumulate in high concentrations. For example, the toxic shrub *Dichapetelum toxicarium* can accumulate ω-fluorooleic acid in its seeds equivalent to 3% of all fatty acids present [8]. It is widely believed that commensal bacterial may be the ultimate source of organofluorine compounds found in plants [9], but this hypothesis has yet to be definitively confirmed.

Nucleocidin

4-Fluorothreonine

(2R3S4S)-5-Fluoro-2,3,4-trihydroxypentanoic acid

Fluoroacetate

ω-Fluoro fatty acids

(2R,3R) 2-fluorocitrate
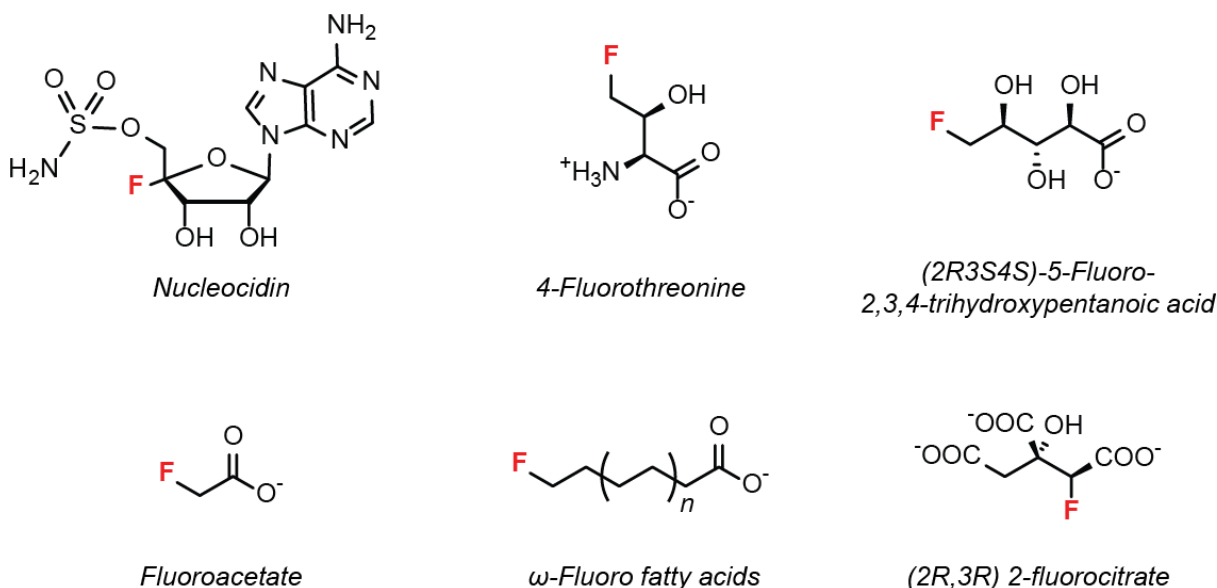
Figure 1.1. Select naturally occurring organofluorines. Products originating from bacteria are shown in the top row, while products found in plants are shown on the bottom row.

Despite the relatively early discovery of fluoroacetate, the first definitive microbial producer of this compound was not isolated until 1987. In the course of optimizing fermentation conditions for the production of the carbapenem thienamycin, it was discovered that *Streptomcyes cattleya* would produce additional antibiotics in the presence of fluoride-containing media constituents. The source of this antibiotic activity were isolated after a heroic large-scale purification, and was determined to be 4-fluorothreonine [*10*]. Fluoroacetate was also identified as a co-product by ^{19}F NMR. The antibiotic activity of fluorothreonine was found to be contingent on media conditions; while it is a potent antibiotic in minimal media it has negligible activity in rich media. Likewise, feeding with serine or threonine was found to abrogate its antibiotic activity, consistent with an antimetabolite mode of action [*11*].

Due to the novel nature of the organofluorine biosynthesis pathway *(Figure 1.2)*, the identification of the fluorinating enzyme itself proved quite recalcitrant. The key breakthrough came when the lab of David O'Hagan identified *S*-adenosyl methionine as the co-substrate for an S$_N$2-based fluorination reaction [*12*]. This recognition facilitated the activity-guided purification of the 5′-fluorodeoxyadenosine synthase, or fluorinase [*13*], which in turn enabled the discovery of the organofluorine biosynthesis locus. Subsequent studies confirmed that fluorination proceeds via an S$_N$2 reaction mechanism, and elucidated the structural basis of fluorination [*14*, *15*]. Fluoride is tightly hydrated in aqueous solution, requiring a unique hydrophobic lid to facilitate the dehydration of fluoride and avoid the spurious hydrolysis of SAM [*15*]. While the fluorinase is somewhat inflexible in terms of its substrate scope, it has proven useful for the radiochemical synthesis of ^{18}F-labeled compounds [*16*]. This utility has been expanded by exploiting the reversibility of the halogenation reaction to facilitate the fluorination of hydrolytically stable 5′-chlorodeoxyadenosine analogues [*17*]. By using a late-stage enzymatic fluorination approach with ^{18}F$^-$ as a reactant, high radiochemical yield can be achieved.
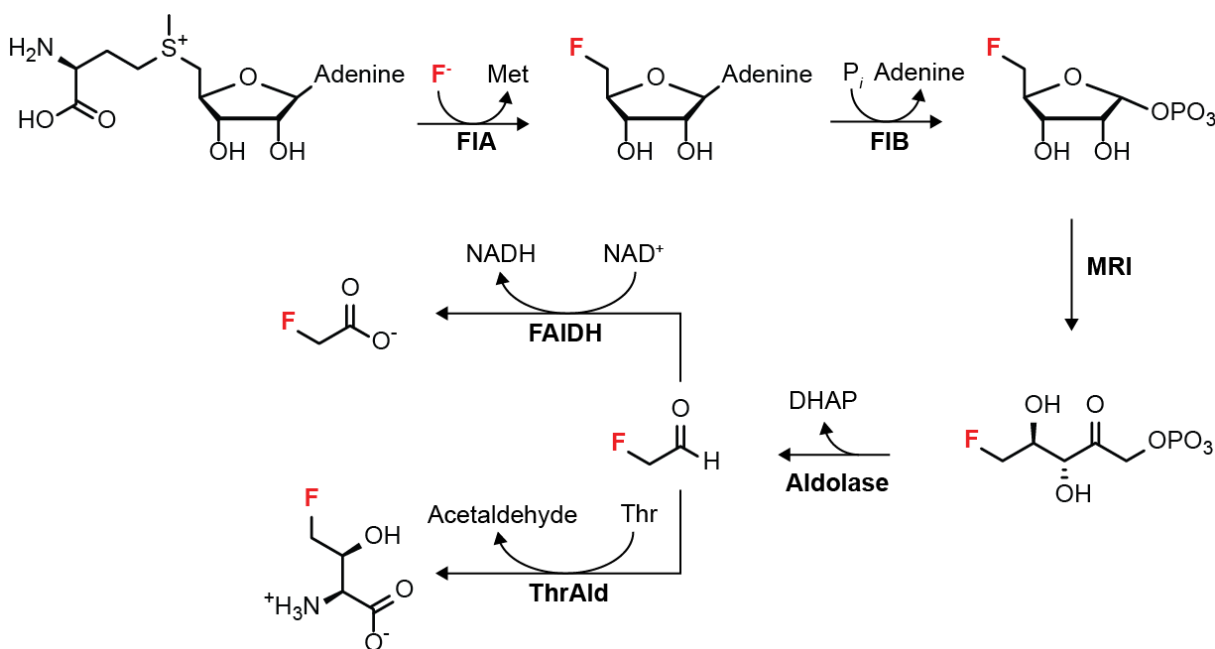
Figure 1.2. *The organofluorine metabolic pathway from S. cattleya.*

3

The discovery of the fluorinase stimulated further efforts to characterize the downstream components of the organofluorine metabolism pathway, the most striking of which is the fluorothreonine transaldolase [*18*]. This enzyme is an unusual fusion of a PLP-dependent domain with homology to serine hydroxymethyltransferases, and a class II aldolase domain. The fluorothreonine transaldolase has a strict requirement for threonine as a substrate, which stands in contrast to other transaldolases, which typically accept glycine. Unfortunately, this fascinating protein has thus far resisted detailed biochemical characterization, as functional protein cannot be expressed in high yield in either *Escherichia coli* or *Streptomcyes coelicolor* [*19*]. The prospect that the enzyme is bifunctional is highly appealing, as the enzyme responsible for the conversion of 5′-fluoro-5′-deoxy-ribulose-1-phosphate to fluoroacetaldehyde has not yet been definitively identified [*19–21*].

The new explosion of bacterial whole genome sequence data over the past decade has also helped inform bioprospecting efforts that have revealed similar organofluorine metabolism in other *Streptomyces* species. Deng and colleagues discovered *Streptomyces* sp. MA37 from the rhizosphere of *Antiaris toxicaria* in a Ghanaian botanical garden. This organism produces a variety of compounds also found in *S. cattleya,* including fluoroacetate and fluorothreonine, but additionally produces novel compound, 5-fluoro-2,3,4-trihydroxypentanoic acid [*5, 22*]. The marine species *Streptomyces xinghaensis* was also found to encode a fluorinase, and produce fluoroacetate [*23*]. This BGC is notable for possessing a truncated fluorothreonine transaldolase sequence that lacks the N-terminal PLP containing domain. In keeping with the idea that this component is required for the transaldolase reaction, *S. xinghaensis* does not produce fluorothreonine. Fluoroacetate production in this organism also seems to be dependent on the expression of the fluoroacetyl-CoA thioesterase FlK. Genome sequencing has also turned up cryptic organofluorine pathways in *Nocardia brasiliensis* and *Actinoplanes* sp. N902-109; standard culturing conditions fail to yield measurable organofluorine production in either strain. Notwithstanding, *in vitro* assays confirm that these organisms encode the fluorinase activity [*22*]. Unlike the *S. cattleya* organofluorine BGC, all newly discovered clusters encode the transaldolase and the fluorinase at a single genomic locus, along with a putative methyribulose isomerase. The proposed fluoroacetaldehyde dehydrogenase is not clustered with the fluorinase in any case, nor is any predicted riblulose-1-phopsphate aldolase. This is consistent with the observed metabolic plasticity of the pathway at steps downstream of fluorination. [*21, 24*]

Recent developments have also highlighted dramatic headway in the elucidation of the nucleocidin biosynthesis pathway in *Streptomyces calvus*. While nucleocidin was discovered in 1957 and identified as an organofluorine compound in 1969 [*25, 26*], it biosynthesis remained unexplored for years in large part due to the fact that the publicly available strain is unable to produce the compound [*27, 28*]. Genome sequencing of the strain revealed that this deficiency is due in part to a spontaneous mutation that prevents the expression of key pathway due to the lack of a tRNA^{Leu} (UUA). Comparison with the ascamycin cluster from *Streptomyces* sp. JCM9888 revealed a conserved set of proteins that appear to be involved in the formation of the unusual 5′-*O*-sulfonamide moiety, but the enzyme responsible for C-F bond formation has yet to be discovered [*6, 29*]. The nucleocidin BGC lacks a homologue of the fluorinase, and achieving fluorination at the 4′-position is likely to require an alternate mechanism of fluorination.

Close inspection also reveals that the nucleocidin gene cluster has some auxiliary features in common with the organofluorine biosynthesis cluster from *S. cattleya.* First, it contains a purine nucleoside phosphorylase with significant homology to FlB. This protein is in fact essential for

organofluorine synthesis in *S. calvus*. Furthermore, the nucleocidin cluster contains NucU, a homologue of the putative cation-proton antiporter FlH. Phylogenetic analysis indicates that these two proteins are in fact closely related, and group into clade that is distinct from the homologous protein from the *Salinispora pacifica* salinisporamide K cluster [*30*]. It is appealing to speculate that these proteins may in fact act as selective halogen importers, although deletion of *flH* in *S. cattleya* retains the ability to produce organofluorines under laboratory conditions [*20*].

The biology of the fluoride ion itself has also received increased attention, thanks to the discovery of a widely distributed fluoride-responsive riboswitch. This genetic element serves to regulate the expression of a variety of fluoride-sensitive enzymes in bacteria. Its most highly conserved function, however, is to activate the transcription of the hypothetical protein CrcB, which in turn confers resistance to fluoride [*31*]. The mystery of how fluoride can be recognized selectively by a highly anionic macromolecule was largely resolved by the crystal structure of the fluoride riboswitch; the RNA backbone does not interact directly, but rather forms a cage that mediates interactions via three tightly-coordinated Mg^{2+} ions [*32*]. Detailed biochemical investigation of CrcB-like transporters from diverse bacteria has shown that these proteins are in fact highly selective for the export of fluoride over chloride [*33*]. This selectivity is may be mediated by an unusually narrow transmembrane pore that is lined with hydrogen bond donors, as well as unusual edge-on interactions with a set of highly conserved phenylalanine residues [*34, 35*]. These developments have helped to highlight the pervasive environmental influence of fluorine on bacterial evolution.

1.2 Control of translational fidelity by the ARS

A major question in the organofluorine physiology of *S. cattleya* is understanding more about the downstream fate of fluorothreonine, especially whether it is inserted into ribosomally-synthesized products. Translation is an extremely demanding biochemical process, requiring polypeptide synthesis at a rate of approximately 10 amino acids s^{-1}, and an error rate approaching 1×10^{-3} [*36*]. As a consequence, numerous steps in the translation have evolved mechanisms to derail the usage of erroneous intermediates [*37, 38*]. Meeting these accuracy criteria is a particularly difficult task for the aminoacyl-tRNA synthetases (ARS), which must distinguish between a set of similar small molecule substrates. This challenge is largely tackled at the activation and aminoacylation steps, where synthetases employ an array of strategies to select the correct amino acid/tRNA pair. These suffice for about half of the ARS, but from the earliest days of molecular biology, it was recognized that not all discrimination problems could be solved in a single step [*39*].

As many amino acids differ from one another by relatively minor structural substitutions, achieving sufficient selectivity to support the demands of translation requires a multi-part process, frequently referred to as the "double sieve" model [*40*]. The metaphorical second sieve is a hydrolytic editing function, which can be implemented at the level of the aminoacyl-adenylate ("pre-transfer") or at the level of the aminoacyl-tRNA ("post-transfer"). Such activity is commonly referred to as aminoacyl-tRNA editing, as it involves the correction of a mistake before it can be propagated to an error in at the protein level. Pre-transfer editing activity is frequently encoded in the same active site that is responsible for aminoacylation [*41, 42*]. Post-transfer activity, in contrast, requires a separate active site, typically on the same polypeptide. Both

activities are widespread in class I and class II synthetases, with ten of the canonical ARS displaying some form of editing activity [43].

Post-transfer editing activities have been found to mediated by several distinct protein families. The most extensively studied are the CP1 domains found in class I synthetases. These domains assist in discriminating between the extremely similar aliphatic amino acids, as well as non-proteinogenic amino acids that can accumulate in specific environmental conditions [44]. Class II synthetases feature their own diverse cast of post-transfer editing domains. Among these, the ability of the second additional domain of the threonyl aminoacyl tRNA synthetase to reject Ser-tRNA^{Thr} has been thoroughly characterized [45, 46]. Likewise, the prolyl-tRNA synthetase contains a structurally unrelated inserted domain that hydrolyzes Ala-tRNA^{Pro} [47]. Strikingly, both of these class II editing domains have free-standing homologues that exert their activity in trans [48]. The most widespread *trans*-editing function, however, is the enforcement of homochirality in translation by the D-tyrosyl deacylase. This protein selectively hydrolyzes aminoacyl-tRNAs with D-stereochemistry, and homologues are present across all three domains of life, a testament to its primordial importance [49, 50].

Recent developments in the field have made clear that low fidelity translation is not universally maladaptive. Some organisms may exploit condition-specific expression of low-fidelity aminoacyl-tRNA synthetases to expand the range of protein functions that can be encoded by a single nucleic acid sequence [51, 52]. Further investigation of this phenomenon is likely to yield a variety of new results, as role of selective mistranslation has been under-explored for technical and conceptual reasons.

Xenobiotic amino acids that can bypass the translational quality control systems provide another set of exceptions to the dominant paradigm of high-fidelity translation [53]. The plant metabolite, canavanine, is an outstanding example. This subtle mimic of arginine can be activated and translated into protein, but disrupts protein folding as a consequence of the dramatically altered side chain pK$_a$, rendering it a valuable tool for the study of protein homeostasis [54]. Interestingly, insects that prey upon canavanine producers have been shown not to translate this amino acid into protein [55]. Whether this additional selectivity filter is implemented at the level of amino acid activation of at the post-transfer stage has yet to be elucidated. Numerous other compounds with similar properties have also been discovered [56].

The extremely compact nature of the carbon-fluorine bond has enabled the creative exploitation of the native translational machinery to accommodate a variety of fluorinated amino acids *(Figure 1.3)*. Probably the most widely used example, and one that highlights the varied properties of fluorinated amino acids, is 4-fluorotryptophan. This amino acid can be incorporated into protein with a high degree of efficiency by tryptophan auxotrophs of *E. coli* [57]. Many fluorotryptophan-substituted proteins are fully functional, but the compound is also quite toxic, and cannot support native growth rates, even in highly evolved strains [57, 58]. In a similar vein, trifluoromethylmethionine, fluorophenylalanine, and *m*-fluorotyrosine can all be charged by the endogenous ARS with a reasonable degree of efficiency, and have found applications in protein-observed ^{19}F NMR studies [57, 59–61]. Fluoroproline and β-fluoroasparagine have also been identified as substrates for translation, and have been used as biochemical probes thanks to subtle differences in stereochemical and electronic properties with respect to their naturally occurring counterparts [62–64]. The systematic investigation of noncanonical amino acids for the generation of genetically encoded peptide libraries has also added γ-difluoroglutamate, and trifluoroleucine to the list of fluorinated amino acids that can be used for protein biosynthesis [65, 66].

Engineered systems have been shown to accommodate an even wider array of fluorinated substrates. Merely overexpressing the valyl- or leucyl-tRNA synthetases can facilitate the efficient incorporation of hexafluorinated derivatives of these amino acids [*67*]. By using an evolved *Methanocaldococcus janaschii* tyrosyl-tRNA synthetase, a variety of fluorinated tyrosine

Figure 1.3. *Synthetic fluorine-containing amino acids that can be incorporated into protein by the native translational machinery.*

analogues can be incorporated in a site-specific fashion, facilitating the interrogation of tyrosyl radical intermediates in the ribonucleotide reductase [*68*]. One common thread across many of these analogues is that the effects of their incorporation can be unpredictable-some proteins can retain full activity upon substitution with fluorinated amino acids, while others display a significant loss of activity. This is consistent with the ability of fluorine to exert a significant chemical influence in the absence of major steric perturbations. As a consequence, fluorinated amino acids present opportunities to test the limits of enzymatic selectivity, and the consequences of subtle changes on protein structure and function.

1.3 Regulation of secondary metabolism in *Streptomyces*

Control of natural product biosynthesis in *Streptomyces* is a sophisticated process that integrates information about environmental conditions and the complex developmental cycle of these organisms [*69, 70*]. Typically, secondary metabolism begins after cells have reached the stationary phase and aerial mycelium formation has begun. This coordination is mediated in part by the rare tRNA^Leu (UUA), which is required for both of these processes [*71*]. Another set of developmental signals comes from diverse, potentially species-specific quorum sensing molecules [*69*]. Global regulators of secondary metabolism have also been shown to exert positive or negative effects in response to the availability of a diverse set of nutrients [*72–74*].

Natural product production has also been shown to be regulated by the presence of foreign antibiotics, either by direct recognition or mechanisms that have yet to be elucidated [*75, 76*]. In a striking finding, Dorrestein and coworkers have shown that secondary metabolism in *Streptomyces* spp. varies through time and space in response to challenges from competing microorganisms [*77*]. These findings are potentially more reflective of the conditions that these organisms face in the wild, and contravene the longstanding observation that limited metabolite profiles are observed under standard laboratory culture conditions.

Apart from control exerted by global regulatory process, many of BGCs also possess at least one "cluster situated regulator" that is required for the expression [*69*]. Reverse-engineering of signaling pathways has also found applications in stimulating the production of transcriptionally

7

silent antibiotics, even in situations where the underlying regulatory architecture is not well understood. One such example is provided by the stambomycins, novel macrocycles from *Streptomyces ambofaciens*, which were discovered by overexpression of a large ATP-binding LuxR-type regulator in conjunction with comparative LCMS/MS [*78*]. In other cases the deletion of predicted repressors, frequently members of the TetR family, has proven to be equally useful [*79*].

Recent studies have also begun to move beyond engineering to elucidate some of the governing principles underlying pathway specific regulation. Cluster-situated regulators have been found to respond directly to the concentrations of pathway intermediates [*80*] and products [*81*], facilitating sophisticated responses with a minimal set of components. Motivation for achieving a mechanistic understanding of the role of small molecule-regulator interactions can be found by considering the case of riboswitches. These conserved signaling elements have been used to link proteins of unknown function to known small molecules of interest [*33*, *82*]. Achieving a similar level of understanding with respect to the interactions of transcription factor-small molecule interactions would open up new options for the targeted discovery of natural products.

1.4   Thesis organization

The goal of this work is to describe new organofluorine biochemistry from nature. By applying a comparative genomic approach, I have been able to identify a fluorothreonine resistance module from *Streptomyces cattleya*. This genetic element encodes a fluorothreonine exporter, SCAT_p0565, and a putative tRNA editing protein, SCAT_p0564. The latter protein was shown to act *in vitro* and *in vivo* as a fluorothreonyl-tRNA selective hydrolase, which is the central finding presented in *Chapter 2*. These findings were facilitated by the development of LCMS/MS strategies for the biochemical and organismal characterization of processes relating to the maintenance of translational fidelity. In *Chapter 3*, I expand on these findings by reporting the crystal structure of SCAT_p0564, and propose strategies to further elucidate the mechanism of action of this protein. *Chapter 4* reports the characterization of the regulatory components of the organofluorine biosynthesis locus.

1.5   References

1.   Müller K, Faeh C, Diederich F (2007) Fluorine in pharmaceuticals: Looking beyond intuition. *Science* 317(5846):1881–1886.

2.   Zhou Y, et al. (2016) Next generation of fluorine-containing pharmaceuticals, compounds currently in phase II-III clinical trials of major pharmaceutical companies: New structural trends and therapeutic areas. *Chem Rev* 116(2):422–518.

3.   Gribble GW (2003) The diversity of naturally produced organohalogens. *Chemosphere* 52(2):289–297.

4.   Gribble GW (2002) Naturally occurring organofluorines. *Organofluorines*, The Handbook of Environmental Chemistry, ed Neilson AH (Springer Berlin Heidelberg), pp 121–136.

5. Ma L, et al. (2015) Identification of a fluorometabolite from *Streptomyces* sp. MA37: (2R3S4S)-5-fluoro-2,3,4-trihydroxypentanoic acid. *Chem Sci* 6(2):1414–1419.

6. Zhu XM, et al. (2015) Biosynthesis of the fluorinated natural product nucleocidin in *Streptomyces calvus* is dependent on the *bldA*-specified Leu-tRNA(UUA) molecule. *Chembiochem* 16(17):2498–2506.

7. Hall RJ (1972) The distribution of organic fluorine in some toxic tropical plants. *New Phytologist* 71(5):855–871.

8. Harper DB, O'Hagan D (1994) The fluorinated natural products. *Nat Prod Rep* 11(2):123–133.

9. Hendriks CBS (2013) The role of endophytes in the metabolism of fluorinated compounds in the South African *Dichapetalaceae*. Dissertation (University of Pretoria). Available at: http://repository.up.ac.za/handle/2263/24657 [Accessed April 25, 2017].

10. Sanada M, et al. (1986) Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya*. *J Antibiot* 39(2):259–265.

11. Rabinovitz M, Finkleman A, Reagan RL, Breitman TR (1969) Amino acid antagonist death in *Escherichia coli*. *J Bacteriol* 99(1):336–338.

12. O'Hagan D, Schaffrath C, Cobb SL, Hamilton JTG, Murphy CD (2002) Biochemistry: biosynthesis of an organofluorine molecule. *Nature* 416(6878):279.

13. Schaffrath C, Deng H, O'Hagan D (2003) Isolation and characterisation of 5′-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Lett* 547(1–3):111–114.

14. Cadicamo CD, Courtieu J, Deng H, Meddour A, O'Hagan D (2004) Enzymatic fluorination in *Streptomyces cattleya* takes place with an inversion of configuration consistent with an S_N2 reaction mechanism. *Chembiochem* 5(5):685–690.

15. Dong C, et al. (2004) Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* 427(6974):561–565.

16. Onega M, et al. (2010) An enzymatic route to 5-deoxy-5-[18F]fluoro-d-ribose, a [18F]-fluorinated sugar for PET imaging. *Chem Commun* 46(1):139–141.

17. Thompson S, et al. (2015) A two-step fluorinase enzyme mediated 18F labelling of an RGD peptide for positron emission tomography. *Chem Commun* 51(70):13542–13545.

18. Murphy CD, O'Hagan D, Schaffrath C (2001) Identification of a PLP-dependent threonine transaldolase: A novel enzyme involved in 4-fluorothreonine biosynthesis in *Streptomyces cattleya*. *Angew Chem Int Ed Engl* 40(23):4479–4481.

19. Deng H, Cross SM, McGlinchey RP, Hamilton JTG, O'Hagan D (2008) *In vitro* reconstituted biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase. *Chem Biol* 15(12):1268–1276.

20. Zhao C, et al. (2012) Insights into fluorometabolite biosynthesis in *Streptomyces cattleya* DSM46488 through genome sequence and knockout mutants. *Bioorg Chem* 44:1–7.

21. Walker MC, Wen M, Weeks AM, Chang MCY (2012) Temporal and fluoride control of secondary metabolism regulates cellular organofluorine biosynthesis. *ACS Chem Biol* 7(9):1576–1585.

22. Deng H, et al. (2014) Identification of fluorinases from *Streptomyces* sp. MA37, *Norcardia brasiliensis*, and *Actinoplanes* sp. N902-109 by genome mining. *ChemBioChem* 15(3):364–368.

23. Huang S, et al. (2014) Fluoroacetate biosynthesis from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B-24674. *Org Biomol Chem* 12(27):4828.

24. Weeks AM (2013) Molecular insights into fluorine chemistry in living systems. Dissertation (UC Berkeley, Berkeley, CA). Available at: http://escholarship.org/uc/item/3sc8k37r [Accessed April 13, 2017].

25. Thomas SO, et al. (1956) Nucleocidin, a new antibiotic with activity against Trypanosomes. *Antibiot Annu*:716–721.

26. Morton GO, Lancaster JE, Van Lear GE, Fulmor W, Meyer WE (1969) The structure of nucleocidin. 3. (A new structure). *J Am Chem Soc* 91(6):1535–1537.

27. Fukuda K, Tamura T, Segawa Y, Mutaguchi Y, Inagaki K (2009) Enhanced production of the fluorinated nucleoside antibiotic nucleocidin by a *rif*^{R}-resistant mutant of *Streptomyces calvus* IFO13200. *Actinomycetologica* 23(2):51–55.

28. Bartholomé A, Janso JE, Reilly U, O'Hagan D (2016) Fluorometabolite biosynthesis: isotopically labelled glycerol incorporations into the antibiotic nucleocidin in *Streptomyces calvus*. *Org Biomol Chem* 15(1):61–64.

29. Zhao C, et al. (2014) Characterization of biosynthetic genes of ascamycin/dealanylascamycin featuring a 5′-O-sulfonamide moiety in *Streptomyces* sp. JCM9888. *PLOS ONE* 9(12):e114722.

30. Eustáquio AS, et al. (2011) The discovery of salinosporamide K from the marine bacterium "*Salinispora pacifica*" by genome mining gives insight into pathway evolution. *Chembiochem* 12(1):61–64.

31. Baker JL, et al. (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* 335(6065):233–235.

32. Ren A, Rajashankar KR, Patel DJ (2012) Fluoride ion encapsulation by Mg^{2+} ions and phosphates in a fluoride riboswitch. *Nature* 486(7401):85–89.

33. Stockbridge RB, Robertson JL, Kolmakova-Partensky L, Miller C (2013) A family of fluoride-specific ion channels with dual-topology architecture. *eLife* 2:e01084.

34. Stockbridge RB, et al. (2015) Crystal structures of a double-barrelled fluoride ion channel. *Nature* 525(7570):548–551.

35. Last NB, Kolmakova-Partensky L, Shane T, Miller C (2016) Mechanistic signs of double-barreled structure in a fluoride ion channel. *eLife* 5:e18767.

36. Gromadski KB, Rodnina MV (2004) Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell* 13(2):191–200.

37. Zaher HS, Green R (2009) Quality control by the ribosome following peptide bond formation. *Nature* 457(7226):161–166.

38. Asahara H, Uhlenbeck OC (2005) Predicting the binding affinities of misacylated tRNAs for *Thermus thermophilus* EF-Tu·GTP. *Biochemistry* 44(33):11254–11261.

39. Pauling L (1958) *Festschrift fuer Prof. Dr. Arthur Stoll* (Birkhauser Verlag, Basel, Switzerland).

40. Fersht AR, Kaethner MM (1976) Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry* 15(15):3342–3346.

41. Boniecki MT, Vu MT, Betha AK, Martinis SA (2008) CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase. *Proc Natl Acad Sci USA* 105(49):19223–19228.

42. Dulic M, Cvetesic N, Perona JJ, Gruic-Sovulj I (2010) Partitioning of tRNA-dependent editing between pre- and post-transfer pathways in class I aminoacyl-tRNA synthetases. *J Biol Chem* 285(31):23799–23809.

43. Perona JJ, Gruic-Sovulj I (2013) Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. *Aminoacyl-tRNA Synthetases in Biology and Medicine*, Topics in Current Chemistry, ed Kim S (Springer Netherlands), pp 1–41.

44. Cvetesic N, Palencia A, Halasz I, Cusack S, Gruic-Sovulj I (2014) The physiological target for LeuRS translational quality control is norvaline. *The EMBO Journal*:e201488199.

45. Dock-Bregeon A-C, et al. (2004) Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution. *Mol Cell* 16(3):375–386.

46. Waas WF, Schimmel P (2007) Evidence that tRNA synthetase-directed proton transfer stops mistranslation. *Biochemistry* 46(43):12062–12070.

47. Beuning PJ, Musier-Forsyth K (2000) Hydrolytic editing by a class II aminoacyl-tRNA synthetase. *Proc Natl Acad Sci USA* 97(16):8916–8920.

48. Ahel I, Korencic D, Ibba M, Söll D (2003) Trans-editing of mischarged tRNAs. *Proc Natl Acad Sci USA* 100(26):15422–15427.

49. Calendar R, Berg P (1967) D-Tyrosyl RNA: Formation, hydrolysis and utilization for protein synthesis. *J Mol Biol* 26(1):39–54.

50. Ahmad S, et al. (2013) Mechanism of chiral proofreading during translation of the genetic code. *Elife* 2:e01519.

51. Schwartz MH, Pan T (2017) tRNA misacylation with methionine in the mouse gut microbiome in situ. *Microb Ecol*. doi:10.1007/s00248-016-0928-0.

52. Schwartz MH, Pan T (2016) Temperature dependent mistranslation in a hyperthermophile adapts proteins to lower temperatures. *Nucleic Acids Res* 44(1):294–303.

53. Bullwinkle T, Lazazzera B, Ibba M (2014) Quality control and infiltration of translation by amino acids outside of the genetic code. *Annu Rev Genet* 48(1):149–166.

54. Boyar A, Marsh RE (1982) L-Canavanine, a paradigm for the structures of substituted guanidines. *J Am Chem Soc* 104(7):1995–1998.

55. Rosenthal GA, Dahlman DL, Janzen DH (1976) A novel means for dealing with L-canavanine, a toxic metabolite. *Science* 192(4236):256–258.

56. Fowden L, Richmond MH (1963) Replacement of proline by azetidine-2-carboxylic acid during biosynthesis of protein. *Biochim Biophys Acta* 71:459–461.

57. Pratt E, Ho C (1975) Incorporation of fluorotryptophans into proteins of *Escherichia coli*. *Biochemistry* 14(13):3035–3040.

58. Bacher JM, Ellington AD (2001) Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. *J Bacteriol* 183(18):5414–5425.

59. Hull WE, Sykes BD (1974) Fluorotyrosine alkaline phosphatase. Fluorine-19 nuclear magnetic resonance relaxation times and molecular motion of the individual fluorotyrosines. *Biochemistry* 13(17):3431–3437.

60. Duewel H, Daub E, Robinson V, Honek JF (1997) Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein ^{19}F NMR. *Biochemistry* 36(11):3404–3416.

61. Arntson KE, Pomerantz WCK (2016) Protein-observed fluorine NMR: a bioorthogonal approach for small molecule discovery. *J Med Chem* 59(11):5158–5171.

62. Renner C, et al. (2001) Fluoroprolines as tools for protein design and engineering. *Angew Chem Int Ed Engl* 40(5):923–925.

63. Hortin G, Stern AM, Miller B, Abeles RH, Boime I (1983) DL-threo-beta-fluoroasparagine inhibits asparagine-linked glycosylation in cell-free lysates. *J Biol Chem* 258(7):4047–4050.

64. Phillips MA, Stern AM, Abeles RH, Tashjian AH (1983) Inhibition of asparagine-linked glycosylation of pro-opiomelanocortin in mouse pituitary cells by DL-threo-beta-fluoroasparagine. *J Pharmacol Exp Ther* 226(1):276–281.

65. Hartman MCT, Josephson K, Szostak JW (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci USA* 103(12):4356–4361.

66. Hartman MCT, Josephson K, Lin C-W, Szostak JW (2007) An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLOS ONE* 2(10):e972.

67. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci USA* 103(25):9482–9487.

68. Minnihan EC, Young DD, Schultz PG, Stubbe J (2011) Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J Am Chem Soc* 133(40):15942–15945.

69. Liu G, Chater KF, Chandra G, Niu G, Tan H (2013) Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev* 77(1):112–143.

70. McCormick JR, Flärdh K (2012) Signals and regulators that govern *Streptomyces* development. *FEMS Microbiol Rev* 36(1):206–231.

71. Lawlor EJ, Baylis HA, Chater KF (1987) Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in Streptomyces coelicolor A3(2). *Genes Dev* 1(10):1305–1310.

72. Santos-Beneit F, et al. (2011) The RNA polymerase omega factor RpoZ is regulated by PhoP and has an important role in antibiotic biosynthesis and morphological differentiation in *Streptomyces coelicolor*. *Appl Environ Microbiol* 77(21):7586–7594.

73. Rigali S, et al. (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* 9(7):670–675.

74. Yu Z, et al. (2012) Differential regulation of antibiotic biosynthesis by DraR-K, a novel two-component system in *Streptomyces coelicolor*. *Mol Microbiol* 85(3):535–556.

75. Wang W, et al. (2014) Angucyclines as signals modulate the behaviors of *Streptomyces coelicolor*. *Proc Natl Acad Sci USA* 111(15):5688–5693.

76. Zarins-Tutt JS, et al. (2016) Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat Prod Rep* 33(1):54–72.

77. Traxler MF, Watrous JD, Alexandrov T, Dorrestein PC, Kolter R (2013) Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *mBio* 4(4):e00459-13.

78. Aigle B, et al. (2014) Genome mining of *Streptomyces ambofaciens*. *J Ind Microbiol Biotechnol* 41(2):251–263.

79. Ye S, et al. (2017) Identification by genome mining of a type I polyketide gene cluster from *Streptomyces argillaceus* involved in the biosynthesis of pyridine and piperidine alkaloids argimycins P. *Front Microbiol* 8. doi:10.3389/fmicb.2017.00194.

80. Li Y, et al. (2016) Coordinative Modulation of Chlorothricin Biosynthesis by Binding of the Glycosylated Intermediates and End Product to a Responsive Regulator ChlF1. *J Biol Chem* 291(10):5406–5417.

81. Sherwood EJ, Bibb MJ (2013) The antibiotic planosporicin coordinates its own production in the actinomycete *Planomonospora alba*. *Proc Natl Acad Sci USA* 110(27):E2500–E2509.

82. Weinberg Z, et al. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology* 11:R31.

Chapter 2: *A fluorothreonine-selective aminoacyl-tRNA deacylating protein maintains translational fidelity in the organofluorine producer Streptomyces cattleya*

2.1. Introduction

In adapting to a variety of ecological niches, living systems have evolved a wide range of chemical phenotypes. One striking example is the soil bacterium, *Streptomyces cattleya*, which biosynthesizes a variety of organofluorine compounds, including the antibiotics fluoroacetate and 4-fluorothreonine (FThr) [*1*]. While carbon-fluorine bonds have become ubiquitous in synthetic small molecules ranging from pharmaceuticals and imaging agents to polymers and liquid crystals based on the unique elemental properties of fluorine [*2–5*], only a handful of natural products contain fluorine, with the majority being discovered in *S. cattleya* [*6*, *7*]. As such, *S. cattleya* is one of the few known genetic hosts for fluorine biology and serves as a platform for exploring how the enzymatic utilization of new elements can be acquired by living organisms.

The products of organofluorine metabolism in *S. cattleya* are simple, but they pose a challenging enzymatic selectivity problem due to their close structural resemblance to key central metabolites. For example, fluoroacetate (FAc) has been show to manifest its toxicity through potent mechanism-based inhibition of the tricarboxylic acid (TCA) cycle [*8*, *9*]. In order to avoid toxicity due to TCA cycle shutdown, *S. cattleya* employs both regulatory and enzymatic detoxification strategies [*10*, *11*]. In comparison, the fate of FThr remains relatively unknown, though it provides a possible substrate for both amino acid metabolism and protein biosynthesis thanks to its structural similarity to threonine [*12*, *13*].

In this work, we examine the downstream metabolism of FThr in *S. cattleya*, with a focus on understanding its interaction with translation. The low error rate of protein synthesis ($\sim 10^{-3}$-10^{-4}) is controlled largely by the ability of aminoacyl-tRNA synthetases (ARS) to correctly charge tRNA substrates with their cognate amino acid and discriminate against other near-cognate amino acids found in the cell [*14*]. In addition to displaying selectivity at the level of amino acid activation, many of the ARSs are also capable of hydrolyzing misacylated tRNAs in order to further reduce errors in translation [*15*, *16*]. While this "double sieve" approach is able to limit the mischarging of commonly occurring amino acids, some xenobiotic amino acids can bypass these filters, enabling the *in vitro* and *in vivo* translation of polypeptides that contain a variety of noncanonical amino acids [*17–21*]. In particular, the threonyl aminoacyl-tRNA synthetase (ThrRS) has been observed to display permissiveness towards γ-substituted analogues similar to FThr [*13*, *17*].

We have used bioinformatic analysis to identify a conserved gene, SCAT_p0564, in the fluorothreonine biosynthesis gene cluster of *S. cattleya*. This protein shares homology with freestanding aminoacyl-tRNA deacylases of the YbaK-like superfamily, which along with members of the AlaX and D-Tyr deacylase families have been shown to provide hydrolytic editing activities in *trans* [*22–24*]. Biochemical characterization of SCAT_p0564 shows that this enzyme hydrolyzes fluorothreonyl-tRNA with 670-fold selectivity over threonyl-tRNA. This is lower than the >6,000-fold difference in the catalytic proficiency of the D-tyrosyl deacylase towards D- vs. L-tyrosyl-tRNA [*25*], but comparable to the selectivity exhibited by other members of the YbaK superfamily, and is consistent with a role in excluding fluorothreonine from the proteome. Indeed, the deletion of SCAT_p0564 from the *S. cattleya* genome leads to significant charging of tRNAs with FThr (~18% of total tRNA^{Thr} charging) and corresponding incorporation of fluorothreonine (~11.5% replacement of Thr) into the proteome, supporting a physiological role of preventing mistranslation of FThr. Although the knockout strain does not show any apparent growth defect, the deletion of a transporter (SCAT_p0565) in the same gene cluster increases both intracellular FThr levels and sensitivity towards exogenous FThr. In contrast, heterologous expression of SCAT_p0564 in a non-fluorothreonine-producing streptomycete does mitigate FThr toxicity.

Taken together, these results are consistent with a model where mischarging of tRNA with FThr by ThrRS can contribute to toxicity and that *S. cattleya* has evolved cellular machinery consisting of a fluorothreonyl-tRNA deacylase and FThr-selective transporter to handle its unique fluorometabolism.

2.2   Materials and methods

Commercial materials.   Amylose resin, Hiscribe T7 transcription kit, Phusion DNA polymerase, Q5 DNA polymerase, deoxynucleotide triphosphates (dNTPs), and restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Asparagine, apramycin, benzoyl chloride, β-mercaptoethanol, β-Nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt hydrate, boric acid, Dowex 50WX8, ethyl fluoroacetate, iodoacetamide, inorganic pyrophosphatase, lithium diisopropylamine, myokinase, naladixic acid, *o*-phthaldialdehyde, phthaloyl chloride, protamine sulfate from salmon sperm, pyruvate kinase/lactic acid dehydrogenase from rabbit muscle, sodium borohydride, and (R)-(−)-2-(tert-Butyl)-3-methyl-4-imidazolidinone trifluoroacetic acid were purchased from Sigma-Aldrich (St. Louis, MO). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators, Milli-Q Gradient water purification system, chloramphenicol, malt extract and yeast extract were purchased from EMD Millipore (Billerica, MA). Adenosine triphosphate disodium salt hydrate, ammonium chloride, carbenicillin, 3 kDa MWCO dialysis tubing, dithiothreitol, ethylenediaminetetraacetic acid disodium salt, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), hydrochloric acid, glucose, kanamycin sulfate, magnesium chloride, magnesium sulfate, phenylmethylsulfonyl chloride (PMSF), potassium phosphate dibasic, sodium acetate, sodium hydroxide, and tris base were purchased from Fisher Scientific (Waltham, MA). Deuterium oxide was purchased from Cambridge Isotopes (Tewksbury, MA). Fluka brand LCMS grade ammonium formate, ammonium bicarbonate, formic acid, methanol and acetonitrile were purchased from Sigma Aldrich (St. Louis, MO). Phenol chloroform (pH 5.2), sodium chloride, and sodium glutamate were purchased from MP Biosciences. Inositol was purchased from the Nutritional Biochemicals Corporation (Cleveland, OH).   Tris(2-carboxyethyl)phosphine (TCEP) was purchased from Biosynth Corporation (Itasca, IL). Pierce brand analytical grade 6 N HCl and trypsin were purchased from Thermo Fisher (Waltham, MA). Ni-NTA resin, Quia-quick PCR cleanup kit, and Quia-spin miniprep kit were purchased from Quiagen USA (Valencia, CA). ATP-^{13}C was purchased from Santa Cruz Biotechnology (Dallas, TX). Bacto^{TM} Agar was purchased from BD (Sparks, Maryland) Soy flour was purchased from Berkeley Bowl (Berkeley, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a concentration of 100 µM, and stored at 4°C. Ammonium persulfate, acrylamide/bisacrylamide (37.5:1 and 19:1), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), and sodium dodecyl sulfate (SDS) were purchased from Bio-Rad Laboratories (Hercules, CA). Lysogeny Broth (LB), Lysogeny Broth Agar (LBA), and Terrific Broth (TB) were purchased from VWR International (Radnor, PA). Ultrayield baffled flasks were purchased from Thompson Instrument Company (Oceanside, CA).

Phylogenetic and genomic context analysis. NCBI RefSeq database was searched with BLAST using SCAT_p0564 as the query sequence. Sequences with e-values < 1E-50 (roughly corresponding to >50% sequence identity) were initially selected for phylogenetic analysis, however a subset of sequences from highly populated branches of the tree were removed in order to fit the tree on a single page. Multiple sequence alignment was performed with ClustalW [26]. The MEGA6 interface was then used to analyze the alignment by Maximum Likelihood based on

the JTT matrix-based model, and uncertainty in the topology of the resulting tree was evaluated with 500 bootstrap replicates [27, 28]. Positions with gaps were discarded, and 142 amino acid positions were used to construct the final tree. Genbank files with the pertinent genomic contexts of were obtained from NCBI RefSeq and visualized using Snapgene viewer (GSL Biotech LLC, Chicago IL). Domain architecture was assigned using the NCBI CDD server [29].

Preparation of fluorothreonine. Fluorothreonine was synthesized as documented previously [30].

Bacterial strains. Cloning was performed in *E. coli* DH10B-T1^R**,** and protein expression was performed in *E. coli* BL21(DE3) harboring the pRARE2 plasmid for the expression of rare tRNAs. *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 (Norwich, Norfolk, UK). The non-methylating strain *E. coli* GM272 harboring the plasmid PUZ8002 was used for conjugative transfer of DNA into *Streptomyces* species [31, 32].

Construction of plasmids. Gibson assembly was used to carry out plasmid construction with *E. coli* DH10B-T1^R as the cloning host [33]. PCR amplifications were carried out using Phusion polymerase (New England Biolabs, Ipswich MA) or Platinum Taq High Fidelity polymerase (Thermo Fisher, Waltham MA), and restriction digests were performed using enzymes obtained from New England Biolabs. The plasmid pET16b-His$_{10}$-ThrRS was constructed by amplification of the target gene from *S. cattleya* genomic DNA using the primers J303/J304, and insertion into NdeI/BamHI digested pET16b. PSV272.1-His$_6$-MBP-Tev-p0564 was constructed by amplification from gDNA with the primers J400/J401 and insertion into SfoI/HindII digested PSV272.1. The plasmid pIJ10702-tRNA25 was constructed by amplification from gDNA with the primers J062/J063 and insertion into BamHI digested pIJ10702. The plasmid pSET152-ermEp*-p0564 was constructed by amplification from gDNA with primers J118/J119 and insertion into NdeI/BamHI digested pSET152-ermEp*. The plasmid pET16b was obtained from Novagen, while PSV272.1 was derived from pET27b, also from Novagen. Plasmids pSET152-ermEp* and pIJ10702 were obtained from the John Innes Center [34, 35]. Descriptions of the plasmids and strains used in this study, as well as sequences of primers are detailed in *Appendix 1.1*. Constructs were sequenced by Quintara Biosciences (South San Francisco, CA)

Purification of *S. cattleya* Threonyl aminoacyl-tRNA synthetase (TRS). For purification of TRS, E. coli Bl21 (DE3) was transformed with the expression plasmid MC_1773 and pRare2. Overnight cultures in Terrific Broth (TB) were used to innoculate expression cultures to OD$_{600}$ = 0.05 in 3× 750 mL TB in UltraYield flasks. Cultures were grown at 37°C until the OD$_{600}$ reached 0.6-0.8 before being chilled on ice for 30 min, after which IPTG was then added to 0.2 mM final concentration and growth was continued overnight at 16°C. Cells were harvested by centrifugation and resuspended in 5 mL 50 mM HEPES pH 7.5, 150 mM sodium chloride, 10 mM magnesium chloride, 5 mM BME, 10% glycerol per gram cell paste before storage at -80°C.

Cells were thawed, incubated with 0.5 mg/mL lysozyme on ice for 30 min, and lysed by passage through a French pressure cell (Thermo Scientific; Waltham, MA) at 15,000 psi. DNAse I (fermentas) was added to 1 u/mL, and the lysate was clarified by centrifuging 2 × 20 min at 15,000 RCF. Lysate was then supplemented with imidazol to 20 mM, applied to 4 mL NiNTA resin and washed with 25 column volumes 25 mM HEPES, 500 mM sodium chloride, 10 mM magnesium chloride, 5 mM BME, 10% glycerol, 20 mM imidazole, pH 7.5, at which point the A$_{280}$ of the eluent dropped below 0.1. Protein was eluted with 25 mM HEPES, 20 mM sodium

chloride, 10 mM magnesium chloride, 5 mM BME, 10% (*v/v*) glycerol, pH 7.5, and applied directly to a 5 mL HiTrap Q column (GE). Separation was accomplished by a gradient from 0-500 mM sodium chloride over 20 column volumes. Fractions containing ThrRS were diluted with 25 mM HEPES, 10 mM magnesium chloride, 50% (*v/v*) glycerol, pH 7.5, to obtain a final glycerol concentration of 20%. Protein was then concentrated to ~6 mg/mL using a 30 kDa MWCO spin filter (Amicon) and flash-frozen in liquid nitrogen before storage at -80°C. The resulting protein could be applied to tRNA without reduction in RNA integrity as visualized by urea-PAGE. Concentration was 6.2 mg/mL as determined using the extinction coefficient calculated by ExPASY ProtParam ($\varepsilon_{280\ nm}$= 75,750 M^{-1}cm^{-1}) [*36*].

Purification of SCAT_p0564. For the purification of SCAT_p0564, E coli BL21(DE3) was transformed with the expression plasmid MC_2115 and pRare2. Overnight cultures in ZY505 media (non-inducing) were used to innoculate expression cultures to OD_{600} = 0.05 in 2× 1L ZY5052 media (autoinducing) in UltraYield flasks [*37*]. Cultures were grown for 10 h at 37C with shaking at 300 rpm, and were harvested by centrifugation for 10 min at 10,000 rcf before storage at -80°C. Pellets were resuspended in 5 mL 300 mM sodium chloride, 5 mM BME, 10% glycerol, 10 mM magnesium chloride, 50 mM HEPES pH 7.5 per gram cell paste and homogenized. Lysozyme was added to 0.5 mg/mL, followed by shaking for 20 min at 30°C Lysis was accomplished with a Misonix Sonicator 3000 (power = 8, 5 s on, 25 s off, 2.5 min total process time, 1/2" tip). Lysate was clarified by centrifugation at 15,000 RCF for 20 min. DNA was then precipitated from the supernatant by the addition of protamine sulfate to 0.075% followed by centrifugation at 15,000 RCF for 20 min. The supernatant was removed, and imidazole was added to 20 mM before batch binding with 4 mL Ni-NTA resin for 30 min at 4°C. Resin was washed with 30 column volumes lysis buffer supplemented with 20 mM imidazole, at which point the A_{280} of the eluent had dropped below 0.1. Protein was eluted with lysis buffer supplemented with 250 mM imidazole, and dialyzed against lysis buffer. After 2 buffer changes, TEV protease was added (1:50 mass:mass with respect to substrate) and cleavage was allowed to continue overnight at 4°C. TEV-cleaved protein was separated from MBP by sequential passage over 4 mL Ni-NTA resin and 2 mL amylose resin. Protein was then diluted 1:1 with 50% (*v/v*) glycerol, 5 mM BME, to obtain a final glycerol concentration of 30%. It was then concentrated using a 3 kDa MWCO Amicon spin concentrator and flash-frozen in liquid nitrogen before storage at -80°C. Concentration was 3.4 mg/mL as determined using the extinction coefficient calculated by ExPASY ProtParam ($\varepsilon_{280\ nm}$= 8,940 M^{-1}cm^{-1}) [*36*].

Selection and preparation of transfer RNA. The genome of *S. cattleya* encodes four tRNA^{Thr} isoacceptors, as predicted by tRNAscan SE [*38*]. Two of these (SCAT_tRNA18 and SCAT_tRNA25) are functionally identical, differing only in whether or not the 3′-terminal CCA is encoded in the genome. This sequence is also nearly identical to tRNA^{Thr} encoded by the model organism *S. coelicolor*, and is predicted to recognize the ACC codon, which is the dominant codon used for threonine in *Streptomyces*. Consequently, this tRNA was used for *in vitro* studies.

Preparation of SCAT_tRNA25 was accomplished by transcription using a Hiscribe T7 High Yield RNA Synthesis Kit (NEB). Template was prepared by amplification from plasmid MC1374 with primers J064 and J071 using Phusion polymerase. The reaction was then conducted using the manufacturer's recommended conditions for short RNAs, with the exception that inorganic pyrophosphatase (Sigma) was added to 5 u/mL u/mL and RNAse-IN (Ambion) to 0.1 u/mL. For the production of ^{13}C-labeled tRNA, ATP-^{13}C (Santa Cruz) was used instead of the supplied ATP. Transcription reactions were quenched with formamide and separated on a 30cm, 8% urea-PAGE

gel with TBE buffer system. The desired band was by visualized by UV-shadow, excised, and extracted overnight at 4 °C with 200 mM sodium chloride, 10 mM Tris pH 7.5, 1 mM EDTA. RNA was precipitated by addition of 1:10 5 M ammonium acetate (Life) and 1:1 IPA, followed by centrifugation for 10 min at 20,817 RCF. At this point it was washed three times with ice cold 70% (*v/v*) ethanol and air dried before resuspension in nuclease-free water.

For concentration measurements, RNA was hydrolyzed by addition of 1 volume 1 N NaOH and incubation at 37°C for >1hr, and the pH was restored to neutral by addition of 2 volumes 1 M sodium phosphate monobasic. The absorbance at 260 nM was used to calculate concentration based on the extinction coefficients of free NMPs and the nucleotide composition of the RNA ($\varepsilon_{260 nm}$= 823,780 $M^{-1}cm^{-1}$) [*39*].

Aminoacylation assay. Aminoacylation reactions were performed at 30°C in the presence of 50 mM HEPES pH 7.5, 10 mM magnesium chloride, 25 mM potassium chloride, 5 mM DTT, 5 mM ATP, 2 U/mL inorganic pyrophosphatase, 24 µM tRNA, 5 mM amino acid, and 50 nM ThrRS. The deacylating protein SCAT_p0564 was present at 0nM, 50nM, or 250nM (0:1, 1:1, or 5:1 with respect to ThrRS). The reaction was initiated by addition of tRNA, and at appropriate time points 2.5 µL portions of sample were removed and quenched by mixing with 5 µL of digestion solution consisting of 200 mM Ammonium acetate pH 5.2, 1.5 U/uL RNAse A, and ^{13}C-labeled aminoacyl tRNA internal standard. Protein was precipitated by addition of 52.5 µL of 5.7% (*w/v*) TCA (final concentration 5% *w/v*), and centrifugation for 1hr at 4,000 RCF.

The resulting 3' adenosyl-threonine adducts were analyzed by UPLC-QQQ. Samples were injected using an Agilent 1290 autosampler, and separated on a Sigma Titan C18 column (1.9 µm, 2.1 × 30 mm; Sigma-Aldrich) connected to an Agilent 1290 UPLC (Agilent Technologies, Santa Clara CA). Column temperature was maintained at 20°C, and a flow rate of 0.7 mL/min was used. The gradient was isocratic 4% mobile phase B until 0.8 min, followed by a linear gradient from 4-30% mobile phase B from 0.8 to 1.55 min, with 5 mM ammonium formate as mobile phase A and acetonitrile as mobile phase B. Mass spectrometry was performed on an Agilent 6460C QQQ with Agilent Jet Stream source. Drying gas flow was 11 L/min at 325°C while sheath gas flow was 12 L/min at 350°C. Quantification was accomplished using the (M+H) → 136.1 transition for analytes, and the (M+H) →141.1 transition for internal standards. Fragmentor voltage was set to 135 V for all analytes. Collision energy was set to 15 V for adenosine and 3'-adenosyl-threonine, and 20 V for 3'-adenosyl-fluorothreonine. Integration was performed using Masshunter Quantitative Analysis software (Agilent) and fitting of standard curves was performed using Microsoft Excel (Microsoft, Redmond WA).

Deacylation assay. Kinetic analysis of deacylation was performed using a reaction system that couples AMP release to NADH consumption. Reactions were performed in a Helma 105.252-QS 1.5 mm pathlength microcuvette and monitored at 340 nM using an Agilent 8453 spectrophotometer with Agilent 89090A temperature controller set to 30°C (Agilent). Aminoacylation was initiated by the addition of TRS to 2.5 µM in a solution composed of 50 mM HEPES pH 7.5, 10 mM magnesium chloride, 25 mM potassium chloride, 5 mM DTT, 5 mM ATP, 400 µM NADH 2 U/mL inorganic pyrophosphatase, 10 U/mL myokinase, 10 U/mL pyruvate kinase/lactate dehydrogenase, tRNA, and 2 mM amino acid. The aminoacylation reaction was allowed to proceed to completion (1-4 min depending on the concentration of tRNA), and then deacylation was initiated by adding SCAT_p0564 to a final concentration of 10-65 nM (fluorothreonine) or 6-19 µM (threonine). The rate of SCAT_p0564-catalyzed hydrolysis was obtained from the rate of NADH consumption as measured by A_{340}. The observed rate scaled with

the concentration of deacylating protein added. Change in A_{340} observed after equilibration, but prior to the addition of deacylating protein was not subtracted. The concentration of aminoacyl tRNA was measured by removing 2.5 µL of reaction mixture into 5 µL of quench mix. Sample processing and LCMS was performed as for the aminoacylation reaction, with the exception that particulates were removed by centrifugation for 30 min at 10,000 RCF. Nonlinear curve fitting was performed using Microcal Origin 6.0 (Microcal Software Inc., Northampton MA).

Preparation of aminoacyl-tRNA standards. Aminoacyl-tRNA standards were prepared by aminoacylation of 24uM tRNA in the presence of 50 mM HEPES pH 7.5, 10 mM magnesium chloride, 25 mM potassium chloride, 5 mM DTT, 5 mM ATP, 1 mM amino acid, 10 U/mL inorganic pyrophosphatase, 20uM tRNA, and 100nM ThrRS for 30 min. Reactions were quenched by extraction with phenol chloroform, pH 5.2, and an equimolar amount of the amino acid not present in the charging reaction was added to the aqueous phase. The quenched reaction was then diluted with 2 volumes of 20 mM sodium acetate pH 5.2, 10 mM magnesium chloride (DEAE-A). The tRNA was bound to 0.5 mL pre-equilibrated DEAE-sepharose and washed with 5 column volumes DEAE-A supplemented with 200 mM sodium chloride to remove free amino acids before elution with 3 column volumes of DEAE-A supplemented with 1 M sodium chloride. Eluent was precipitated by the addition of 1/10 volumes 5 M ammonium acetate and 1 volume IPA, then washed with $3 \times 70\%$ ethanol and air dried before resuspension in 10 mM sodium acetate pH 5.2 and storage at -80°C. For quantification of the extent of aminoacylation, aminoacyl tRNAs were diluted 1:4 and 1:16 with 100 mM borate pH 10.4 and allowed to hydrolyze for 20 min before quantification via amino acid analysis. Removal of free (non-aminoacylated) amino acid was verified by the absence of the amino acid that was added post-quench.

Construction of cosmids for gene disruption in _S. cattleya_. The _S. cattleya_ NRRL 8057 cosmid library described previously was used to prepare cosmids for the disruption of SCAT_p0564 and SCAT_p0565 [_10_]. Plates containing the cosmid library in pooled form were screened for the SCAT_p0564 coding sequence by colony PCR using primers J144 and J145. Wells resulting in initial hits were streaked out to individual colonies, which were re-screened to obtain the desired cosmid (f7.2). The cosmid was isolated using the Qiagene QIAprep Miniprep Kit, and the boundaries of the inserted fragment were identified with sequencing from T7 and T3 promoters. The Am^R/OriT cassette was then amplified using primers J239/240 (for $\Delta SCAT\_p0564$::Am^R) and J240/241 (for $\Delta SCAT\_p0565$::Am^R). The cassette was inserted into f7.2 using the REDIRECT protocol in BW25113 cells containing the helper plasmid pIJ790 [_40_]. Colonies harboring the Am^R marker were isolated on selective media, miniprepped, and sequenced with primers J235 and 236 to confirm the insertion site.

Conjugative transfer of plasmids from _E. coli_ to _S. cattleya_ and _S. coelicolor_. Conjugations were performed using a modified version of the standard procedure [_34_]. Chemically competent _E. coli_ GM272 cells harboring PUZ8002 were transformed with OriT-containing plasmids or cosmids, and selected on the appropriate antibiotics. Single colonies were picked into 5 mL LB supplemented with the appropriate antibiotics and grown overnight at 37°C. Seed cultures were used to innoculate 50 mL growths to OD_{600}=0.05 in 250 mL baffled flasks, which were then grown at 37°C until the OD_{600} reached 0.3-0.4. The _E. coli_ culture was then pelleted by centrifugation for 5 min at 9000 RCF and supernatant was discarded. The pellet was washed by three successive additions of ice cold, antibiotic-free LB and centrifugation. After the last spin, the pellet was resuspended in 500 µL water, and mixed with approximately 1×10^8 _Streptomyces_ spores that had previously been heat shocked for 5 min at 55°C. The combined cells

were then centrifuged for 5 min at 2,000 RPM in a benchtop microcentrifuge. The majority of the supernatant was discarded, and cells were resuspended and plated on mannitol-soy flour plates (MSMg, 20 g soy flour, 20 g mannitol, 20 g agar, 2.03 g magnesium chloride hexahydrate per liter, 30 mL per plate) which were incubated at 30°C. After 24 h 750 µg naladixic acid was applied in 500uL water to repress growth of the *E. coli*. Immediately afterward (for *S. coelicolor M1152*) or after another 24 h (for *S. cattleya*), apramycin (1.5 mg) was applied in 800 µL water. Plates were incubated for another 5-7 d, at which point colonies were evident on successful conjugation plates.

Colonies surviving apramycin selection were restreaked on MSMg, and propagated in 3 mL cultures of GYM (yeast extract, 4 g/L; malt extract, 10 g/L; glucose, 4 g/L; agar, 10 g/L) at 30°C. Half the biomass was plated for spores on MSMg + 1.5 mg apramycin, while the other half was pelleted and stored at -20C. Genomic DNA was prepared from pellets using the salting out procedure [34]. For *S. cattleya* knockout strains, double crossover homologous recombination was confirmed by PCR with primers J146_564middleF/J147_564middleR (for *Δp0564*::Am^R) and J144_565middleF/J145_565middleR (for *Δp0565*::Am^R). PCR from WT and single crossover gDNA gave a short band while the double crossover strains showed a single, longer band. *S. coelicolor* knockins were identified by Am^R phenotype.

S. cattleya growths. Spores of *S. cattleya* were heat shocked at 55°C for 5 min before inoculation into 50 mL seed cultures of GYM media (yeast extract, 4 g/L; malt extract, 10 g/L; glucose, 4 g/L; presterile pH 5.0) in 250 mL baffled flasks with glass beads. After 14 h of growth, cultures were re-innoculated to OD 0.05 in 100 mL GYM media in 500 mL UltraYield flasks with glass beads. Following another 12 h of growth, sodium fluoride was added to a final concentration of 2 mM. Growth was monitored by OD_{600}, and biomass was collected 6 d subsequent to fluoride addition. Culture media was removed, split into 2 mL portions, and centrifuged for 1 min at 18,000 RCF at room temperature to pellet the cells. Supernatant was removed by decanting, and cell pellets were flash frozen in liquid nitrogen and stored at -80°C. Four biological replicates of each strain (WT, *Δp0564*, and *Δp0565*) were grown, harvested, and used for downstream analysis. Statistical comparison of measured physiological parameters was performed in Microsoft Excel (Microsoft, Redmond WA) using single tailed t-tests with Welch's correction for unequal sample variance.

^{19}F NMR analysis of organofluorine production. Lyophilized supernatant was resuspended in ^{19}F NMR buffer (20% D_2O, 80% H_2O, 100 mM Tris-HCl pH 7.5, 1 mM 5-fluorouracil) at a ratio of 1 mL buffer per 10 mL culture supernatant. NMR spectra were collected on a Bruker AVQ-400 spectrometer at the College of Chemistry NMR Facility at the University of California, Berkeley (128 scans; o1p = -200ppm; d1 = 1 s). Spectra were referenced to 5-fluorouracil (-168.33 ppm vs. $CFCl_3$). Concentrations of fluorothreonine and fluoroacetate were calculated by normalizing the total integrated area of fluorothreonine, fluoroacetate, and fluoride to 2 mM.

Amino acid analysis. Amino acid analysis was performed by OPA derivatization followed by LCMS. Derivatization of samples was accomplished on an Agilent 1290 autosampler using a program that sequentially mixed 0.5 µL of sample, 1.25 µL 400 mM pH 10.4 borate, 0.25 µL 10 mg/mL OPA/mercaptopropionic acid, and 16 µL 0.4 % acetic acid before injection. OPA adducts were then separated on an Agilent Eclipse Plus RRHD C18 column (1.8 µm, 2.1 × 50 mm; Agilent) connected to an Agilent 1290 UPLC. Column temperature was maintained at 40°C, and the flow rate was 0.6 mL/min. The gradient was isocratic at 2% mobile phase B (45% methanol, 45% acetonitrile, 10% water)/98% mobile phase A (10 mM ammonium acetate pH 8.2) until 0.2 min, followed by a linear gradient from 2-31.45% B from 0.2 to 4.2 min.

Detection was performed on an Agilent 6460C QQQ with Agilent Jet Stream source. Drying gas flow was 10 L/min at 325°C while sheath gas flow was 12 L/min at 350°C. Quantification was performed using a (M+H) → (M+H-105.1) transition that was found to be characteristic of OPA adducts. Fragmentor voltage was set to 135 V and collision energy was set to 15 V for all analytes. Integration of peaks was performed using Masshunter Quantitative Analysis software (Agilent) and fitting of standard curves was performed with Microsoft Excel.

Preparation of samples for whole-protein amino acid analysis. Culture media (2 mL) was removed from shake flasks and harvested by centrifugation at 18,000 RCF for one min before supernatant was discarded. The sample was then flash frozen and stored at -80 before analysis. Pellets were thawed and washed 2 times with 1 mL lysis buffer (50 mM Tris pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1 mM PMSF). They were then lysed by sonication with Misonix 3000 with 1/8" microtip (power =5, 5 s on 25 s off for 45 s total process time). Lysate was clarified by centrifugation for 10 min at 18,000 RCF. The supernatant was diluted to 300 µg/mL in 500 µL (as determined by BCA assay) and spiked with 5 µL of 100 mM B-hydroxy norvaline. TCA was added to 10% and protein was precipitated by centrifugation for 20 min at 20,817 RCF. The supernatant was stored for analysis of free amino acids, while the pellet was washed 3 times with 100uL 10% (*w/v*) TCA and 2 times with 90% acetone 0.01% HCl. It was then air-dried and stored at -20C. Pellets were subsequently resuspended in amino acid grade 6 N HCl in anaerobic culture tubes, and the headspace was cleared of oxygen by 3 cycles of vacuum followed by backfilling with argon. Hydrolysis was performed at 100°C for 24-28 h, after which samples were evaporated to dryness by speedvac, resuspended, and analyzed by LCMS as described previously. All glassware used for hydrolysis was pre-cleaned by boiling in 1 N HCl and rinsing with milliQ water. Samples collected at 48 and 72 h indicate that threonine and fluorothreonine are both stable under the conditions employed, while an absence of B-hydroxy norvaline in the hydrolysate indicated that <1% of free amino acids remained after TCA precipitation.

Preparation of samples for proteomics. Whole-protein samples for proteomics were lysed using a Misonix Sonicator 3000 with 1/8" microtip (power =5, 5 s on 25 s off for 45 s total process time) in 50 mM Tris pH 7.5, 150 mM sodium chloride, 1 mM EDTA. Lysate was cleared by centrifugation for 10 min at 18,000 RCF at 4°C. The resulting supernatant was precipitated by addition of TCA to 10% (*w/v*) and pelleted by centrifugation for 10 min at 20,817 RCF at 4°C.

Pellets were resuspended in 8 M urea in phosphate buffered saline (PBS), and the concentration was determined using BCA protein detection reagent (Pierce) with lysozyme standards. Samples were then diluted to obtain 80 µg protein in 30 µL 8 M urea. Subsequently, 40 µL of 100 mM ammonium bicarbonate and 30 µL of 0.2% (*w/v*) Rapigest in 100 mM ammonium bicarbonate were added in succession and the sample was vortexed briefly. The sample was then reduced by addition of 10 µL of 110 mM TCEP (Bisynth, Itasca IL) followed by gentle room temperature inversion for 30 min, at which point 2.5 µL of 500 mM iodoacetamide was added and samples were incubated in the dark for 30 min. Unreacted iodoacetamide was quenched by addition of 25 µL 100 mM DTT, and the sample was diluted with 120 µL PBS and 1.2 µL 1% (*w/v*) Rapigest before addition of MS-grade trypsin (Pierce) on a 1:25 mass:mass basis. Digestion was allowed to proceed overnight at 37°C, and was quenched by the addition of formic acid to 5% final concentration. After centrifugation for 30 min at 20,817 RCF, samples were desalted by application to a pre-conditioned Sep-Pak Vac 1cc (50 mg) tc18 SPE column (Waters, Milford MA), washed with 2 × 1 mL 0.1% formic acid, and eluted with 2 × 150 µL 80% acetonitrile 0.5 % formic acid. Desalted samples were dried by speedvac. For shotgun proteomics, samples were

submitted to the UC Davis Proteomics Core at the UC Davis Genome Center for analysis by LC-Q Exactive. Targeted proteomics analysis was performed in house by LC-QQQ.

Shotgun proteomics data analysis. Raw data files were searched with MaxQuant version 1.5.7.0. The database search was performed against the *S. cattleya* proteome (Uniprot accessions FQ859185 and FQ859184) with reversed protein sequences as decoys. Precursor mass tolerance was set to 5 ppm, and fragment mass tolerance was set to 10 ppm. Up to one missed cleavage was allowed, and carbamidomethylation of Cysteine was set as a fixed modification. Oxidation of Methionine, N-terminal acetylation, and Threonine to Fluorothreonine (+17.99058 Da) substitutions were allowed as variable modifications. Peptide identifications were filtered to Posterior Error Probability values <1% for downstream analysis of fluorothreonine incorporation. For assessment of differential incorporation of fluorothreonine, data for peptides encoding single threonine residues were split into separate datasets corresponding to the fluorothreonine-containing and unmodified peptides. Intensities from the Label Free Quantitation routine of MaxQuant were input into the MSStats package to assess differential fluorothreonine incorporation. Codon usage statistics were tabulated using custom Python scripts [*41*]. Gene ontology assignments were obtained from MicroScope [*42*]. R and Python scripts used for data analysis are available from: https://github.com/mcmurryj/Fthr-proteomics. Annotated tandem MS spectra were generated using PeptideShaker [*43*].

Targeted proteomics.   Aliquots of peptide (2.5 µg) were injected via autosampler and separated on an Agilent AdvanceBio Peptide Mapping C18 column (2.7 µm, 2.1 × 250 mm; Agilent) connected to an Agilent 1290 UPLC. Column temperature was maintained at 60°C, and the flow rate was 0.4 mL/min. The gradient was isocratic at 3% mobile phase B (90% acetonitrile, 10% water, 0.1% formic acid)/ 97% mobile phase A (0.1% formic acid) until 5 min, followed by a linear gradient from 5-50% B from 5 to 65 min. Peptides were detected with an Agilent 6460C QQQ with Jet Stream source for quantification runs, while an Agilent 6530 QTOF with ESI source was used for initial survey runs. For quantification runs, drying gas flow was 10 L/min at 280°C while sheath gas flow was 12 L/min at 325°C. For survey runs, drying gas flow was 10 L/min at 300°C, and data acquisition was performed with 1 MS and 3 MS/MS spectra per second.

Two criteria were applied to select peptides detected in survey runs for downstream analysis. First, peptides from abundant organofluorine biosynthesis and housekeeping proteins were chosen in order to ensure robust detection. Second, the threonine-containing variant of each peptide was required to be detected with high confidence (Q-value < 0.01). Database searches were conducted using Morpheus [*44*].

The Skyline targeted proteomics package was then used to select an initial set of transitions, which were pared down to those exhibiting the best signal to noise for final analysis. For all fragments associated with a given fluorothreonine-containing peptide, the corresponding y- or b-ions were also considered for the equivalent unmodified (threonine-containing) peptide. Masshunter Quantitative Analysis software was used to perform integration of the resulting data. Integration was initially performed using Agile parameter-free integration, with manual correction applied in cases where the software assigned a single peak to multiple integration regions. The fluorothreonine incorporation rate *p* was then estimated as $p = \frac{F}{(F+T)}$ where F is the peak area corresponding to the fluorothreonine-containing peptide and T is the peak area corresponding to the threonine-containing peptide. For peptides with multiple threonine residues, it was sometimes possible to identify a transition that specifically measured incorporation at one site. When this

was not possible, incorporation rate was estimated as $p = \frac{F}{(F+T)*n}$ where F is the peak area corresponding to the fluorothreonine-containing peptide, T is the peak area corresponding to the threonine-containing peptide, and *n* is the number of threonine residues encompassed by the quantifier transition. Incorporation rates inferred in this fashion are consistent with those measured using single-threonine peptides. Transitions used for targeted proteomics are documented in *Appendix 2.2*.

Fluorothreonine Growth Inhibition Assays. Spores for fluorothreonine growth inhibition assays were prepared by plating from liquid media onto mannitol-soy flour plates supplemented with 10 mM magnesium chloride. Spores were harvested by scraping with a plastic sterile scraper (*S. cattleya*) or cotton swabs (*S. coelicolor*), and were washed 3 times with water and once with 20% (*v/v*) glycerol before resuspension in 20% (*v/v*) glycerol for storage at -20°C. Approximately 2.5×10^7 spores were heat shocked for 5 m at 55°C, plated on Hopwood minimal media (10 g/L agar, 10g/L sugar 0.5 g/L asparagine, 0.2 g/L magnesium sulfate heptahydrate, 0.5 g/L potassium phosphate dibasic, 10 mg/L iron sulfate heptahydrate) and allowed to dry for 1 h at room temperature. Fluorothreonine was applied to a sterilized disc of filter paper which was then placed in the center of the plate. Plates were cultured at 30°C and photographed 5 d after plating with a Nikon Coolpix P300 (Nikon USA, Melville NY).

For growth inhibition assays in liquid media, approximately 1×10^7 spores were heat shocked for 5 min at 55°C, and inoculated into 4 g/L inositol, 10g/L glycerol, 0.5 g/L potassium phosphate dibasic, 1.5 g/L ammonium chloride, 0.2 g/L magnesium sulfate heptahydrate, 10 mg/L each of zinc sulfate, iron sulfate heptahydrate, manganese chloride, and calcium chloride [*1*].

2.3   Results and discussion

Conserved genomic context of fluorothreonine biosynthesis. For many years, *S. cattleya* was the only known genetic host for production of FThr, but recently a number of bacterial genome sequences encoding organofluorine biosynthesis clusters have been published [*45–48*]. This wealth of new data has enabled the use of comparative genomic approaches to identify new processes related to fluorometabolism. Using this approach to analyze genomes encoding two genes responsible for C-F bond formation (fluorinase) and FThr production (FThr transaldolase), we identified two previously uncharacterized conserved coding sequences (*Figure 2.1*). These genes in *S. cattleya*, SCAT_p0564 and SCAT_p0565, encode an aminoacyl-tRNA deacylating protein of the YbaK-like superfamily (NCBI CDD cl00022) and a putative amino acid exporter of the EamA-like superfamily (NCBI CDD cl23754), respectively.

Given the previously described role of YbaK-like proteins in maintaining translational fidelity [*24*, *49*, *50*], we were intrigued by the possibility that SCAT_p0564 could play a role in controlling the incorporation of FThr into protein via hydrolysis of fluorothreonyl-tRNA^{Thr}. The conservation of an amino transporter in the FThr biosynthetic cluster also suggested that it might serve a complementary role by exporting FThr from the cell. Our belief that these two proteins perform fluorothreonine-related functions was further bolstered by examination of the chlorothreonine biosynthesis cluster from *Streptomyces* sp. OH-5093. While it shares no biosynthetic similarity with the fluorothreonine cluster, it contains both an EamA-type transporter and an open reading frame with homology to the aminoacyl-tRNA deacylase AlaX (*Figure 2.1*) [*51*].

25
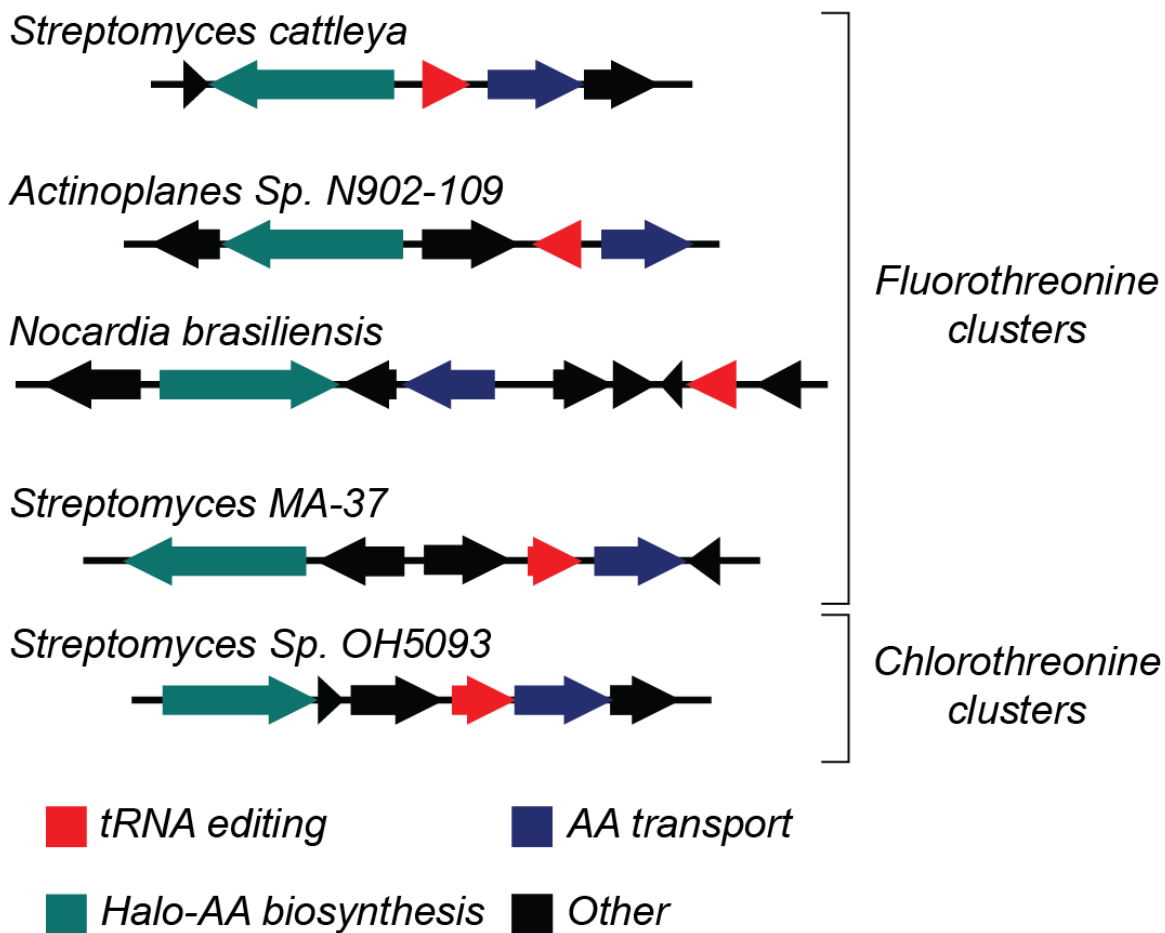
Figure 2.1. *Gene conservation of fluorothreonine-associated proteins. All sequenced organofluorine loci as identified by the presence of a fluorothreonine transaldolase CDS (green) contain homologues of both the YbaK-like protein SCAT_p0564 (red) and the EamA domain protein SCAT_p0565 (blue). In addition, the biosynthetically-unrelated chlorothreonine gene cluster also contains similar organization with respect to these two functions.*
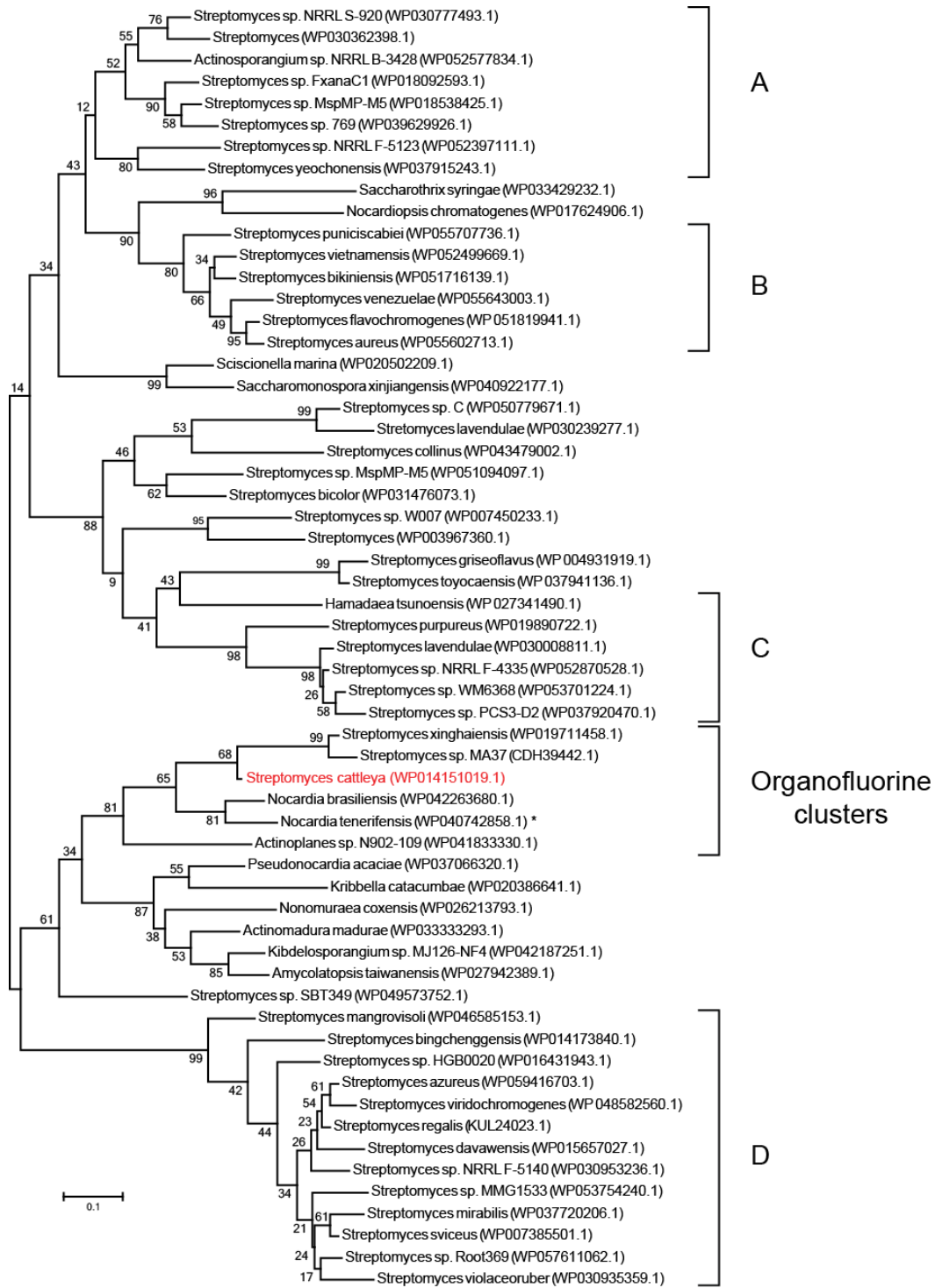
Figure 2.2. *Expanded phylogeny of SCAT_p0564 homologues with species of origin and NCBI RefSeq ID. Clades of putative editing proteins that display qualitative conservation of genomic context are denoted with brackets A-D. Example contexts are presented in figure S2. Sequence WP_040742858 from Nocardia tenerifensis, denoted with an asterisk, groups with editing proteins from known organofluorine clusters despite lacking a fluorinase. Bootstrap values are indicated at branch points.*

27
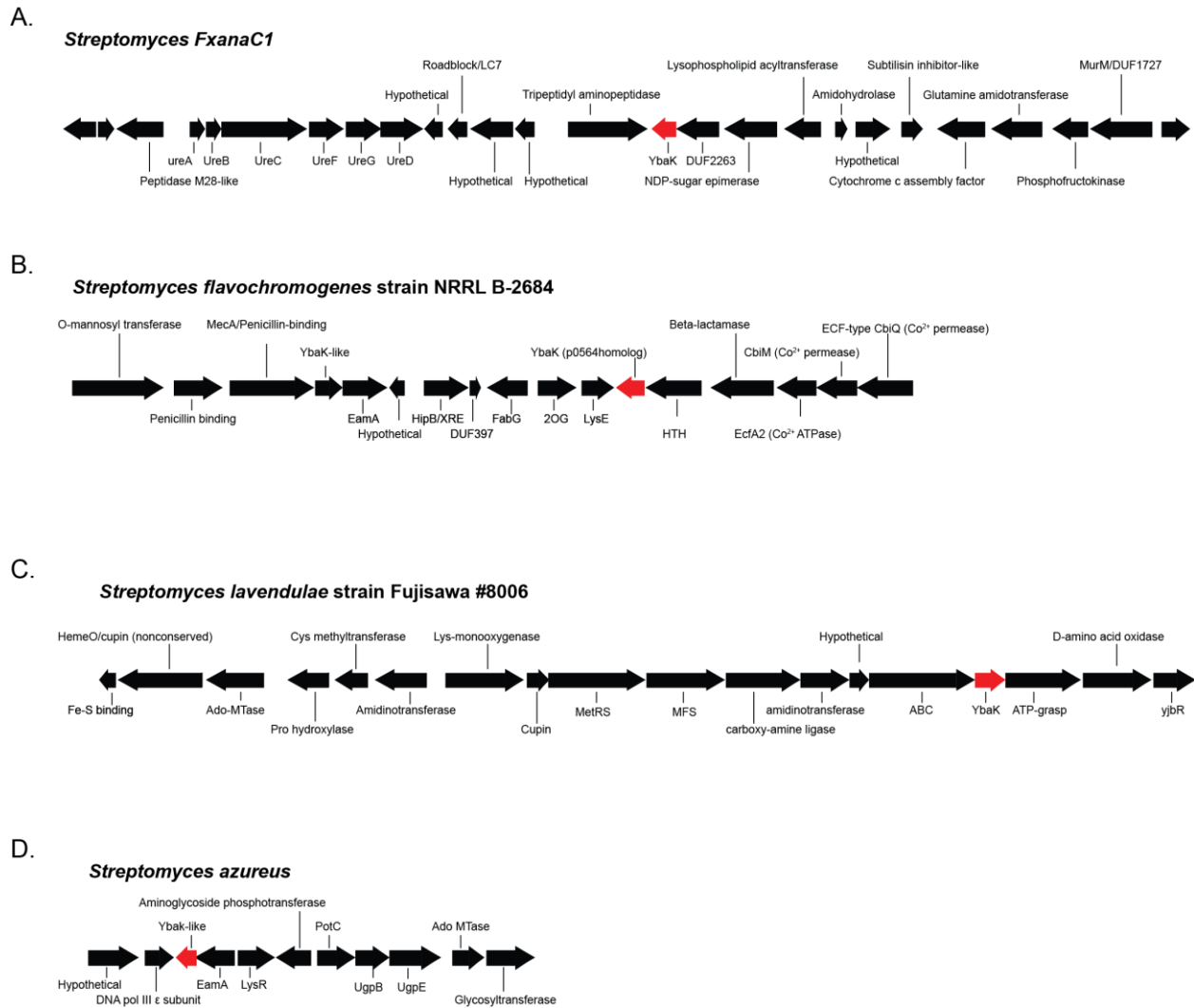
Figure 2.3. *Example genomic contexts of putative aminoacyl-tRNA editing proteins from phylogenetic clades A-D, as labeled in figure 2.1. Group C is of particular interest, given the preponderance of proteins with bioysynthetic potential, as well as an intriguing, truncated homologue of the methionyl-tRNA synthetase. Putative tRNA editing proteins are marked in red.*

This functional conservation is consistent with both tRNA-editing protein and transporter playing a role in adapting to the production of halogenated threonine analogs.

Phylogenetic analysis of close orthologs (BLAST e-values < 1E-50) further supported the assignment of SCAT_p0564 as a fluorothreonine-associated protein. The sequences from organofluorine biosynthesis clusters group together to form a clade that contains just one protein that does not originate from a FThr transaldolase-containing gene cluster. This suggests that they may all descend from an ancestral FThr biosynthetic cluster (*Figure 2.2*). The majority of the remaining proteins are found in other *Streptomyces* species, and many are found within a set of four distinct conserved genomic contexts (*Figure 2.3*). Other outliers with non-conserved genomic contexts are clustered with known or putative NRPS modules, indicating that recruitment of YbaK-like proteins may be a common adaptation to the biosynthesis of unusual amino acids.

SCAT_p0564 prevents buildup of fluorothreonyl-tRNA *in vitro*. In order to characterize the role of SCAT_p0564 in fluorothreonine metabolism, we first examined its effects on aminoacylation *in vitro*. Both SCAT_p0564 and the ThrRS from *S. cattleya* were cloned, heterologously expressed in *E. coli*, and purified to homogeneity (*Figure 2.4*). *S. cattleya* tRNA^{Thr} (GGU) was prepared by *in vitro* transcription from PCR-amplified template using the T7 polymerase, followed by urea-PAGE gel purification. To measure aminoacylation state, RNAse A was used to cleave adenosine from the 3′ end of uncharged tRNA and 3′-O-threonyl adenosine from threonyl-tRNA [52]. These compounds were then analyzed by LC-MS/MS (*Figures 2.5 and 2.6*) [53]. To enable quantification of the 3′-O-adenosyl esters, threonyl- and fluorothreonyl-tRNA standards were prepared and purified by anion exchange before measuring their concentration by amino acid analysis. *(Figures 2.7 and 2.8)*. Internal standards composed of aminoacyl tRNAs prepared with ^{13}C-labeled ATP were also employed to extended the linear range of the assay and control for matrix-dependent and temporal variation in MS response (*Figure 2.9*).

This assay was applied to investigate the biochemical activity of SCAT_p0564 (*Figure 2.10 A*). In the absence of deacylating protein, Fthr was observed to be an excellent substrate for the *S. cattleya* ThrRS, with 69% charging achieved after 30 min in the presence of 50 nM ThrRS *(Figure 2.10 B)*. When SCAT_p0564 was included in the aminoacylation reaction at a 1:1 molar ratio with respect to ThrRS, the time-dependent accumulation of fluorothreonyl-tRNA dropped dramatically to 18%. Increase of SCAT_p0564 to a 5:1 molar ratio reduced the measured level of fluorothreonyl-tRNA below the calibrated range of the assay (<2%). In contrast, 59% charging of threonine was achieved in the absence of SCAT_p0564 under the same conditions and the tRNA aminoacylation state remained unaffected by even the highest level of SCAT_p0564 *(Figure 2.10 C)*.

This experiment was also run under competitive conditions at equimolar concentrations of FThr and Thr (5 mM) since both aminoacyl-tRNA-derived species could be quantified simultaneously based on their difference in mass. While the total pool of charged tRNA reached approximately the same level as with the individual amino acids (62% of total charging), FThr (43%) was able to out-compete Thr (19%) for aminoacylation in the absence of SCAT_p0564 *(Figure 2.10D)*. When SCAT_p0564 was added at a 1:1 ratio with ThrRS, threonyl-tRNA (33%) was then able to accumulate compared to fluorothreonyl-tRNA (8%) (*Figure 2.10E*). However, FThr remained inhibitory to Thr aminoacylation, which is consistent with effective competition by FThr for the charging reaction. Together, these results show that SCAT_p0564 can selectively
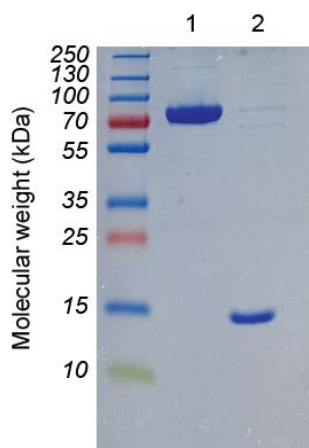
Figure 2.4. *TRS (lane 1) and SCAT_p0564 (lane 2) from S. cattleya were expressed in E. coli and purified for use in in vitro assays.*
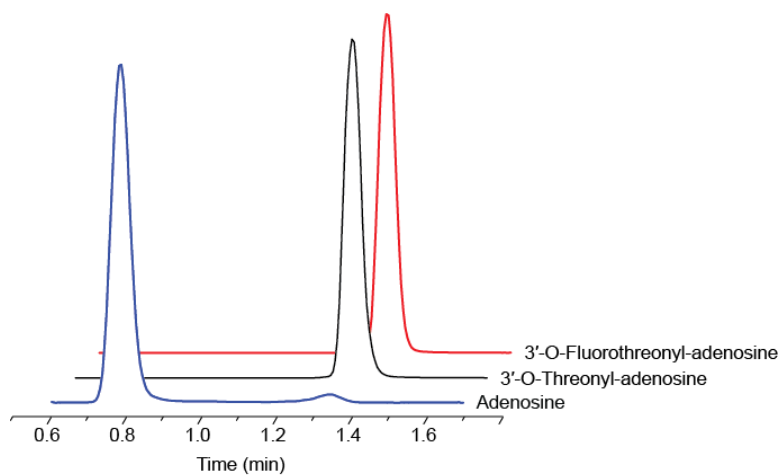
Figure 2.5. *Chromatograms from LC-MS/MS detection of adenosine and adenosine derivatives resulting from the cleavage of aminaocyl-tRNA by RNAse A. Adenonsine, 3'-O-Threonyl-adenosine, and 3'-O-Fluorothreonyl-adenosine were respectively quantified using the 268.1→136.1, 369.1→136.1, and 387.1→136.1 transitions.*
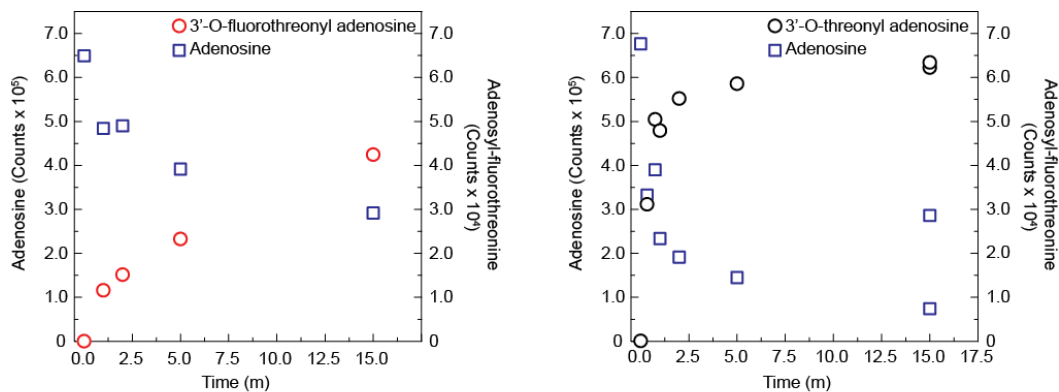
Figure 2.6. *Initial testing demonstrated that aminoacylation could be monitored by RNAse A digestion followed by LC-MS/MS. Uncharged tRNA is cleaved to release Adenosine from the 3' terminus, while charged tRNA is digested to produce a 3'-O-aminoacyl ester. As the reaction progresses, adenosine is consumed with concomitant formation of the 3'-O-aminoacyl ester.*
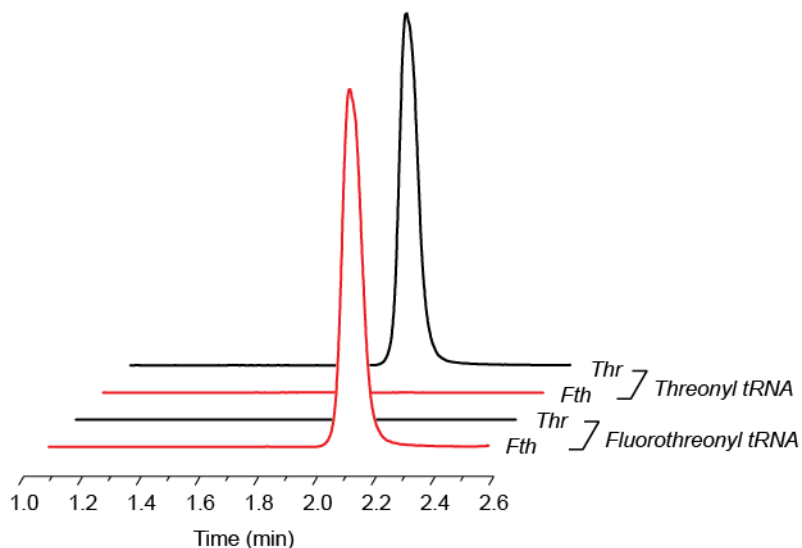
Figure 2.7.   *Aminoacyl-tRNA concentration can be measured by anion exchange followed by amino acid analysis. Aminoacyl-tRNA was prepared, and subjected to phenol-chloroform extraction. The quenched reaction was then spiked with the amino acid that was not charged on the tRNA, and free amino acids were separated from RNA by anion exchange on DEAE resin. Hydrolyzed aminoacyl-tRNA samples were derivitized with OPA and subjected to LC-MS/MS with MRM detection.   OPA-derivatized threonine was measured using the 324.1→ 219.0 transition and the equivalent fluorothreonine adduct was measured with the 342.1→237.0 transition. LCMS traces are normalized to the highest signal observed within each sample. The absence of FThr from the threonyl-tRNA sample, and Thr from the fluorothreonyl-tRNA sample, confirms that anion exchange removes free amino acids from the aminoacyl tRNA prep. This renders amino acid concentration in hydrolysate a suitable measure of aminoacyl tRNA concentration.*

31
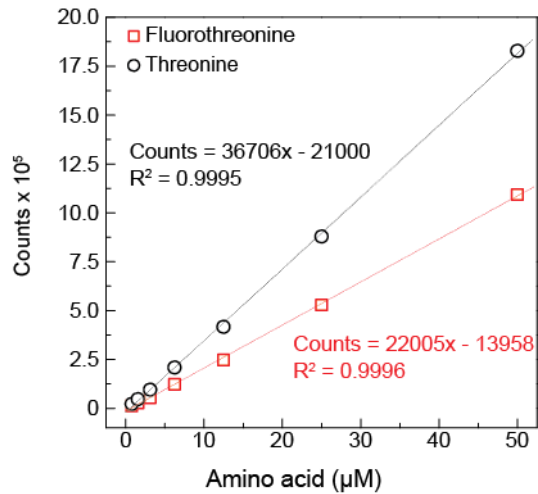
Figure 2.8. *Representative standard curves for the LCMS/MS-based quantification of amino acids derivitized with OPA. The 324.1→ 219.0 and 342.1→237.0 transitions were used for threonine and fluorothreonine, respectively.*
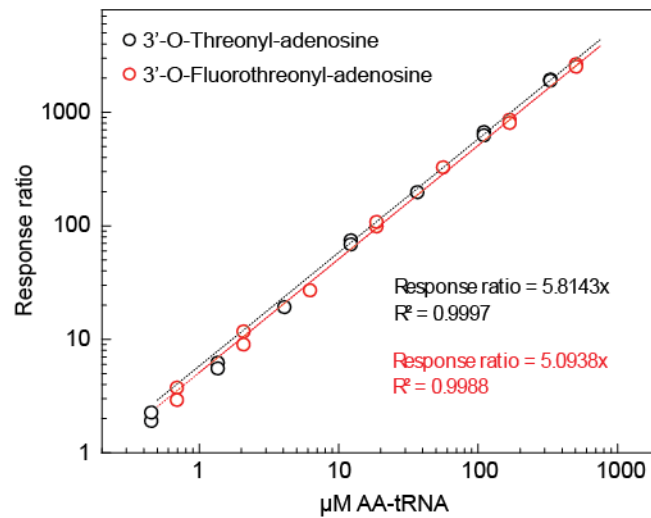
Figure 2.9. *Representative standard curves for the quantification of aminoacyl tRNA by RNAse A digestion followed by LC-MS/MS. Peak areas for analytes are normalized to ^{13}C-labeled internal standards.*
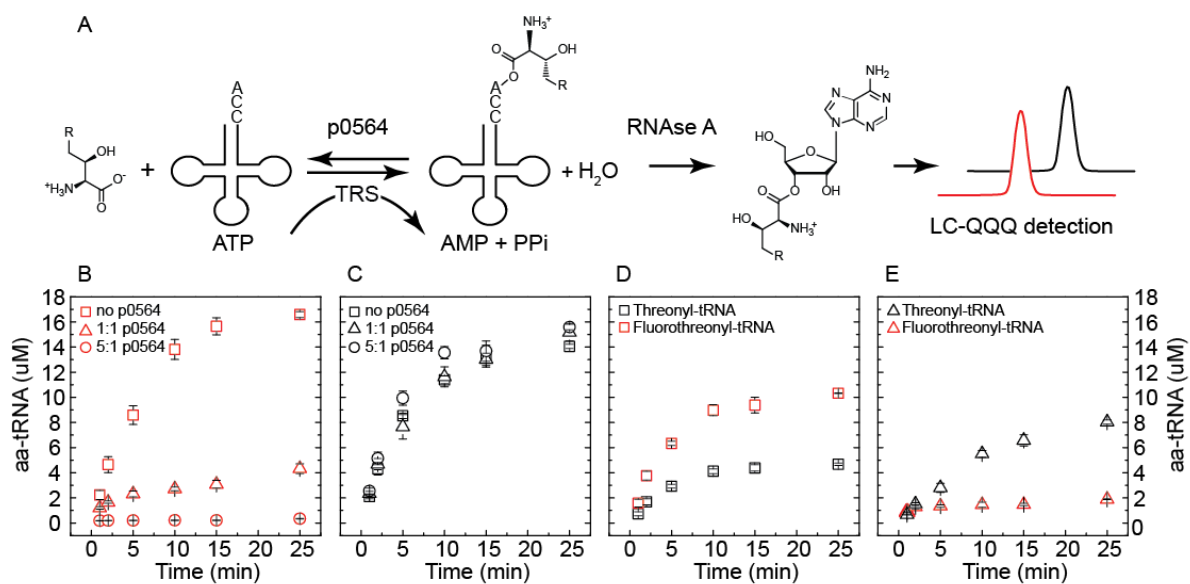
Figure 2.10 *Aminoacylation levels of tRNA^{Thr} with threonine (R=H) or fluorothreonine (R=F) as a function of SCAT_p0564. (A) The amino acid was charged by 50 nM S. cattleya ThrRS in the presence of 0, 50, or 250 nM SCAT_p0564 and monitored by LC-MS/MS following enzymatic digestion with RNAse A. In vitro transcribed tRNA^{Thr} was present at 24 µM in all reactions. (B) Fluorothreonine (5 mM). (C) Threonine (5 mM). (D) Competitive assay containing equimolar fluorothreonine (5 mM) and threonine (5 mM) in the absence of SCAT_p0564. (E) Competitive assay containing equimolar fluorothreonine (5 mM) and threonine (5 mM) in the presence of 1:1 SCAT_p0564 (50 nM). Each data point represents the mean ± s.e. (n >3).*

33

hydrolyze fluorothreonyl- over threonyl-tRNA and could potentially control the *in vivo* aminoacylation balance between these two amino acids.

The fluorothreonyl-tRNA selectivity of SCAT_p0564 is primarily mediated by k_{cat}. In order to further characterize the selectivity of SCAT_p0564 between fluorothreonyl-and threonyl-tRNA, a spectrophotometric assay employing a coupled aminoacylation/deacylation system was developed based on a similar assay reported for TyrRS [*54*]. In this assay, the steady-state rate of deacylation was measured by coupling the hydrolysis of aminoacyl-tRNA to the aminoacylation reaction that occurs upon release of the free tRNA. The subsequent production of AMP results in NADH consumption by pyruvate kinase/lactate dehydrogenase, enabling a spectrophotometric readout of the reaction *(Figure 2.11 A)*. In addition to monitoring the rate of the reaction, the concentration of the aminoacyl-tRNA substrate was validated directly by sampling from the cuvette, followed by quenching and LCMS/MS analysis as described previously.

Using this assay, we discovered that SCAT_p0564 discriminates between fluorothreonyl- and threonyl tRNA with a 670-fold difference in k_{cat}/K_m *(Figure 2.11 BC)*. The majority of selectivity is accounted for by a 130-fold increase in k_{cat} with respect to the non-fluorinated substrate, with a lesser contribution from K_m *(Figure 2.11 BC)*. Interestingly, this behavior is similar to that observed for the fluoroacetyl-CoA thioesterase (FlK), an enzyme involved in fluoroacetate detoxification from *S. cattleya* that also displays high selectivity towards a single fluorine substitution for hydrogen [*11*, *55*]. The observed K_m value for the fluorinated substrate (12.4 ± 1.7 μM) is slightly higher than the estimated *in vivo* concentration of threonyl-tRNA in bacterial cells, which could prevent its buildup and introduction into the proteome.[*56*] Relative to other characterized members of the YbaK-like superfamily, catalytic rate of SCAT_p0564 towards fluorothreonyl-tRNA ($2,200 \pm 400$ min^{-1}) appears to be high, although a direct comparison cannot be made. It has been noted that high enzyme concentrations must be employed to achieve appreciable rates of hydrolysis for many superfamily members [*49*]. For example, 1μM YbaK is required to achieve ~50% hydrolysis of 0.2 μM Cys-tRNA^{Pro} in 2.5 min, corresponding to a rate of ~0.04 min^{-1} [*24*]. Under similar conditions, 5 nM SCAT_p0564 can hydrolyze >50% of 5 μM FThr-tRNA in under 2 min. SCAT_p0564 is thus more comparable to the D-Tyr deacylase in catalytic activity, although these proteins are from different superfamilies [*57*]. It is intriguing to speculate that this difference may be the result of selective pressure. The preferred substrate of YbaK, Cys-tRNA^{Pro}, is produced slowly by ProRS, so even a relatively slow hydrolytic enzyme can keep up. In contrast, fluorothreonyl-tRNA must be hydrolyzed quickly because it is produced more quickly thanks to the competence of fluorothreonine as a substrate for ThrRS.

Elucidating the contribution of SCAT_p0564 and SCAT_p0565 to *S. cattleya* fluorothreonine physiology. With the capacity of SCAT_p0564 to preferentially hydrolyze fluorothreonyl-tRNA established *in vitro*, we sought to explore its function *in vivo*. We initiated these experiments by disrupting the gene encoding SCAT_p0564 in the *S. cattleya* genome using the REDIRECT method *(Figure 2.12)*. In addition, we disrupted the putative FThr transporter (SCAT_p0565) in order to test its contribution to *S. cattleya* fluorothreonine physiology. The resulting *Δp0564*::Am^R and *Δp0565*::Am^R strains were grown alongside the wild-type (WT) strain for 6 d in GYM media supplemented with 2 mM sodium fluoride and all three strains displayed comparable growth *(Figure 2.13)*. As organofluorine production does not begin until after the onset of stationary phase, the lack of an observable growth phenotype was not surprising [*10*]. After 6 d, cultures were harvested and the markers for FThr physiology were assessed.
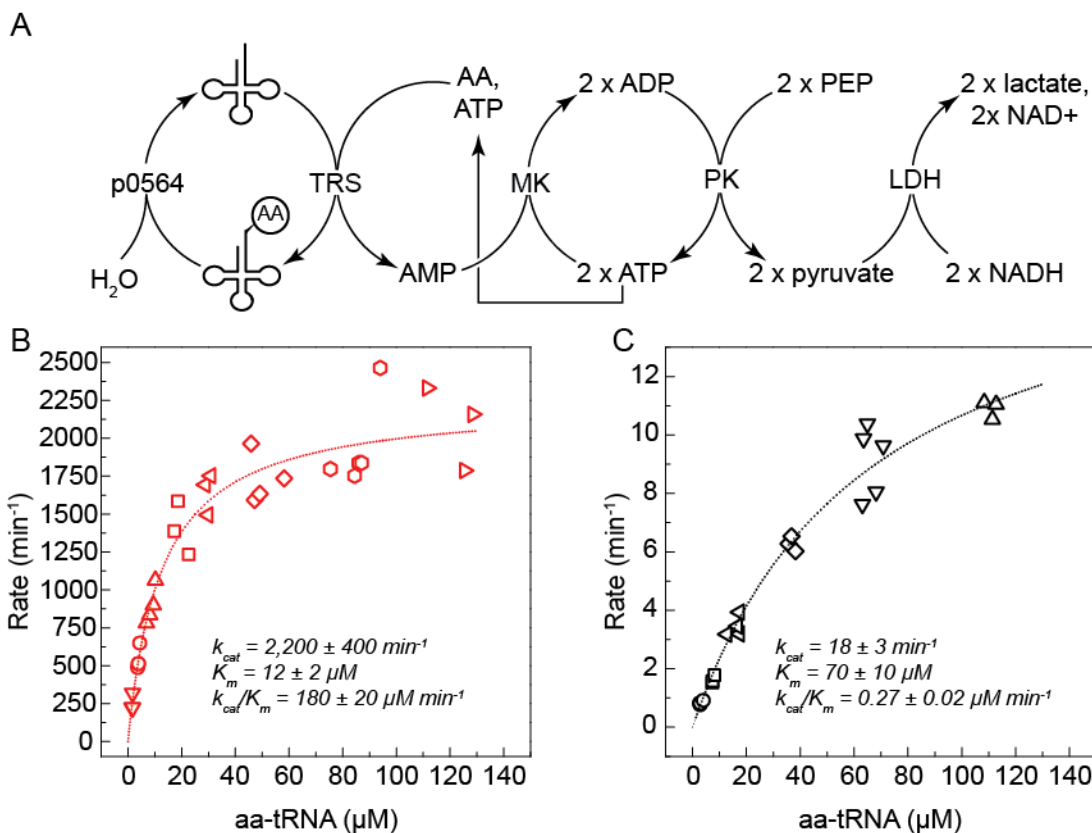
Figure 2.11. *Steady-state kinetic analysis of the deacylation of fluorothreonyl- and threonyl-tRNA by SCAT_p0564. (A). Coupled assay for monitoring aminoacyl-tRNA deacylation. Charged tRNA substrates were generated in situ by S. cattleya ThrRS. Enzymatic rates were measured spectrophotometrically by coupling AMP release to NADH consumption while substrate concentrations were measured by LCMS. (B) Fluorothreonyl-tRNA^{Thr}. (C) Threonyl-tRNA^{Thr}. (C) Symbol shape corresponds to the concentration of tRNA that was added to each reaction; at least 3 measurements were performed at each tRNA concentration.*
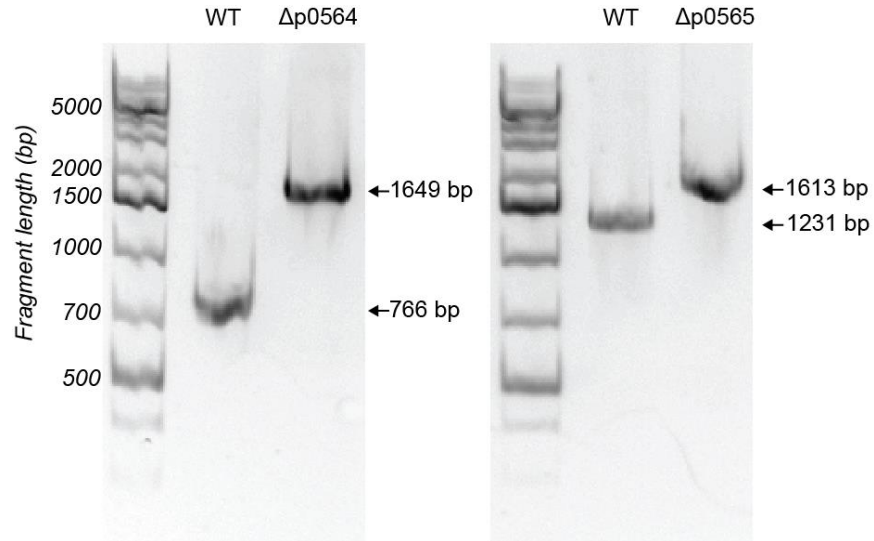
Figure 2.12. *Validation of Δp0564 (A) and Δp0565 (B) strains. Genomic DNA of WT and Δp0564 strains were amplified with primers J146/7 to yield either a 766 bp band (WT) or a 1649 bp band (Δp0564). Genomic DNA of WT and Δp0565 strains were amplified with primers J144/5 to yield either a 1231 bp band (WT) or a 1613 bp band (Δp0565).*
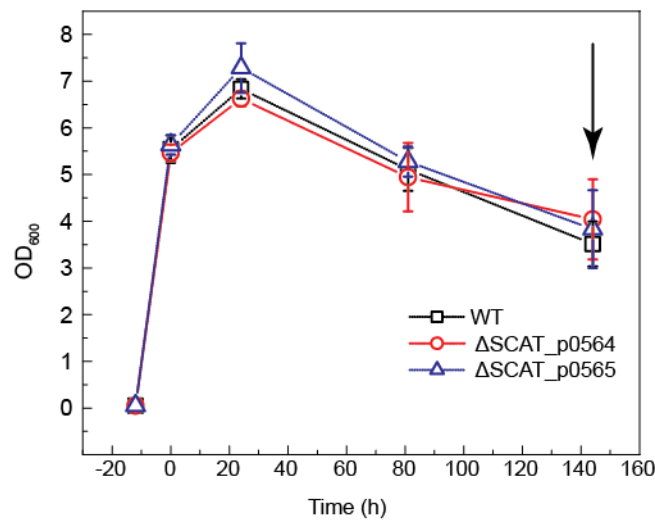
Figure 2.13. *Growth curves of S. cattleya strains in GYM + 2 mM fluoride. Fluoride was added at time zero, and samples were collected 6 d (144 h) later. Cell density was measured by OD_{600}.*

We analyzed these samples for intracellular and secreted FThr as well as its abundance in the aminoacyl-tRNA and protein pool. The concentrations of fluoroacetate and FThr in culture supernatant were monitored with ^{19}F-NMR spectroscopy (*Figure 2.14A*). The *Δp0564*::Am^R strain produced concentrations of FThr and fluoroacetate comparable to the wild-type strain, while the *Δp0565*::Am^R strain showed dramatically reduced secretion of FThr (p < 0.01). Intracellular levels of FThr and Thr were then quantified using the ratio of free amino acid to soluble protein in order to control for any differences in cell pellet recovery or lysis efficiency. These measurements showed that intracellular FThr content was comparable in WT and *Δp0564*::Am^R strains and significantly elevated (p <0.001) in the *Δp0565*::Am^R strain, with a 4-fold increase over WT (*Figure 2.14B*). This is consistent with the proposed role of SCAT_p0565 as a FThr exporter. As the FThr secretion and accumulation phenotypes of the *Δp0564*::Am^R strain are more similar to those of the wild-type, it is likely that polar effects on the expression SCAT_p0565 are minimal. In contrast, intracellular FThr levels displayed a modest but significant increase (p<0.01) in both knockout strains as compared to the WT, which may be indicative of a broader regulatory response to FThr related stress.

Whole RNA was prepared from cell pellets, digested with RNAse A, and characterized by LCMS to assess tRNA aminoacylation state. The baseline fluorothreonyl-tRNA level in WT was observed to be 4 ± 1% of the total FThr- and Thr-charged tRNA pool. This fraction increased 5 to 7-fold upon disruption of the fluorothreonyl-tRNA deacylase (18 ± 7%) and the FThr transporter (23 ± 3%) (*Figure 2.15C*). These results, like the *in vitro* aminoacylation experiments, support the hypothesis that FThr competes well for aminoacylation onto the tRNA. Furthermore, it appears that that the elevated FThr concentration resulting from knockout of the FThr transporter can overwhelm the *trans*-editing function of SCAT_0564. The ability of fluorothreonyl-tRNA to be incorporated into protein was then subsequently examined by amino acid analysis of total protein hydrolysate using LCMS/MS, as described previously (*Figures 2.7 and 2.8*). Hydrolyzed whole-protein samples contained very little fluorothreonine in WT *S. cattleya* compared to the knockout strains (*Figure 2.14D*). Strikingly, the *Δp0564*::Am^R strain exhibited the highest level of FThr incorporation (11.5 ± 0.2%) with respect to total Thr and FThr whereas *Δp0565*::Am^R showed FThr to a lower extent (4.0 ± 0.3%). As the observed FThr content in protein does not scale linearly with the charging of FThr onto tRNA, the strains examined may exhibit time-dependent differences in FThr incorporation. From these studies, we conclude that SCAT_p0564 plays a role in preventing FThr mistranslation *in vivo* by reducing the intracellular levels of fluorothreonyl-tRNA.

Targeted and shotgun proteomic experiments to examine FThr incorporation in the proteome. FThr incorporation into protein was confirmed by targeted proteomic experiments on the WT and *Δp0564*::Am^R strains. The most readily observed peptides from four extremely abundant proteins were selected for analysis (*Figure 2.15*). While the ability to make quantitative inferences about incorporation rates is limited in the absence of synthetic standards, the results were consistent with the data obtained from amino acid analysis on the protein fraction. Peak areas for transitions corresponding to Thr- and FThr-containing peptides were tabulated and used to estimate FThr incorporation, which varied between 0.10-0.32% for WT and 4.4-10.5% for the fluorothreonyl-tRNA deacylase knockout strain. Higher FThr incorporation was observed in the *Δp0564*::Am^R strain across all peptides measured (*Table 2.1A*).
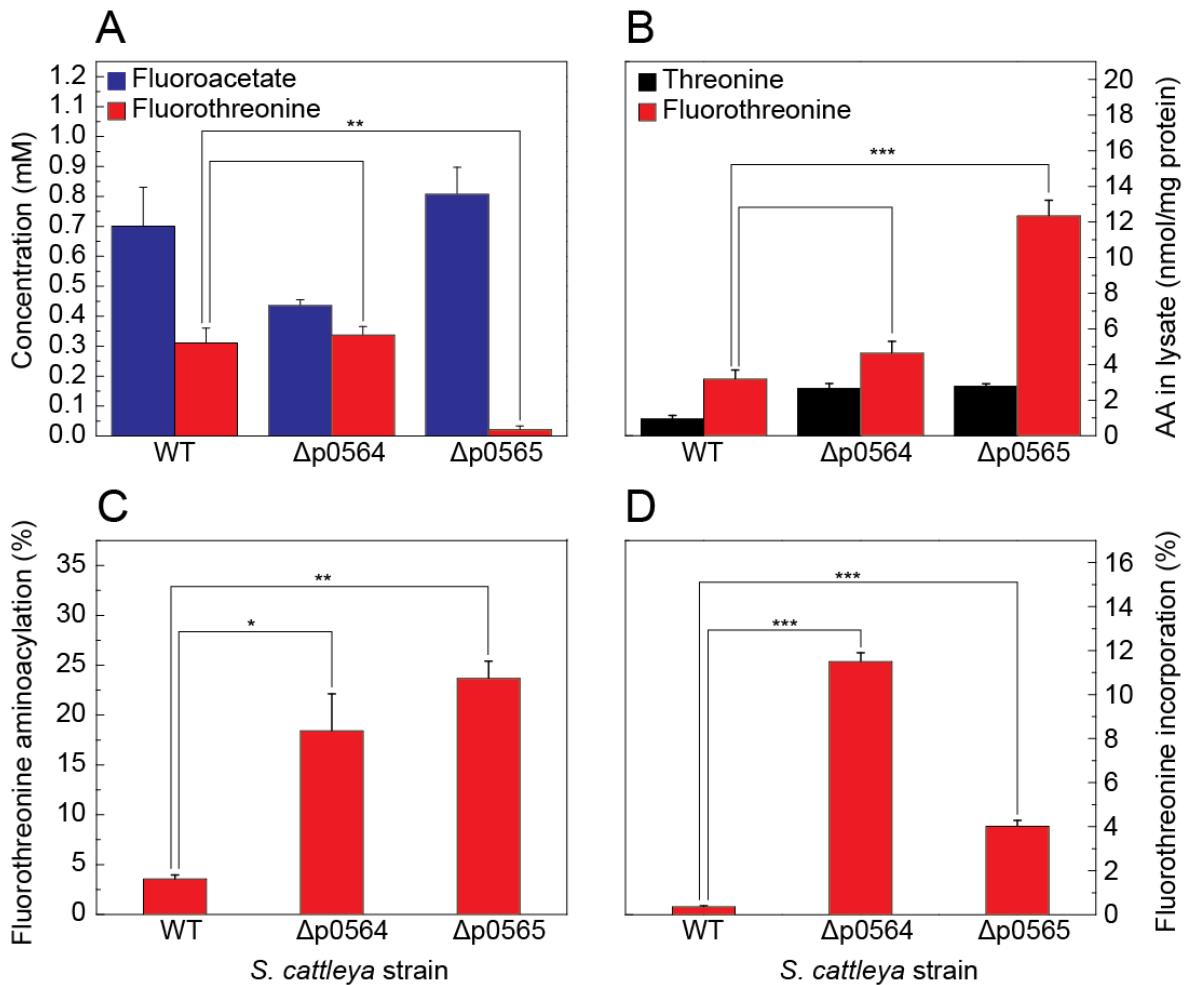
Figure 2.14. *Analysis of fluorothreonine physiological markers in wild-type, Δp0564::Am[R], and Δp0564::Am[R] S. cattleya strains after onset of organofluorine production (6 d). (A) Levels of secreted fluoroacetate and fluorothreonine in culture supernatant measured by ^{19}F NMR. (B) Intracellular amino acid levels in cell pellet lystates measured by LC-MS/MS. (C) Fluorothreonyl-tRNA abundance compared to total fluorothreonyl- and threonyl-tRNA measured by RNAse digestion followed by LC-MS/MS. (D) Incorporation of fluorothreonine into the total protein fraction measured by LC-MS/MS analysis of hydrolyzed trichloroacetic acid precipitate. Data are mean ± s.e. (n =4). Parameters that differ significantly between WT and knockout strains are marked: * (p<0.05), ** (p <0.01) or ***(p<0.001).*
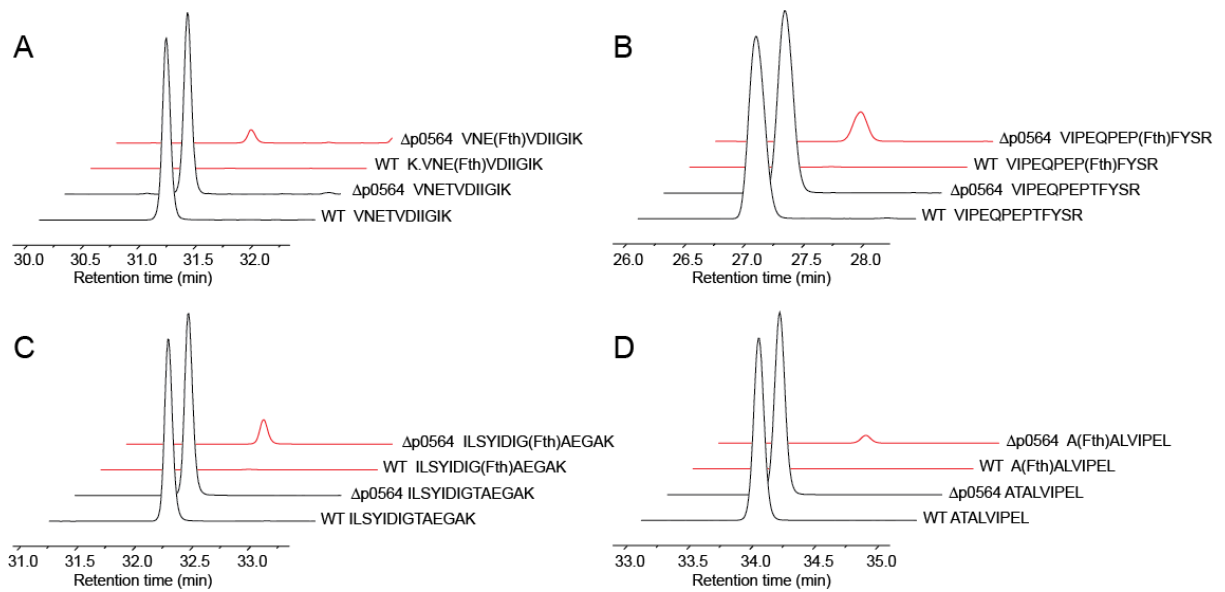
Figure 2.15. *Example MRM chromatograms for threonine- and fluorothreonine-containing peptides from WT and Δp0564 strains of S. cattleya. Chromatograms correspond to peptides from (A)TufA, (B) FAIDH, (C) FlA, and (D) GapA. Chromatograms are normalized such that the plotted intensity equals the measured signal divided by the signal corresponding to the threonine-containing peptide collected for the same sample.*

A.

Protein	Peptide	WT	Δp0564
		Estimated FThr incorporation rate	
	AAAENIIPTTTGAAK	0.07±0.02	3.4±1.1
GapA	LVDLTTFVGGR	0.03±0.01	2.0±1.1
	ATALVIPELK	0.10±0.003	4.4±0.6
	ILSYIDIGTAEGAK	0.31±0.02	10.9±0.7
	VLTGGER	0.31±0.03	10.5±0.5
FaDH	VDLGGSLSGGYYVAPTIFEGDNR	0.36±0.07	10.3±0.7
	IFQEEIFGPVVSVTR	0.31±0.11	10.7±0.7
	DLSTAYR	0.49±0.09	15.6±1.8
	TTLTAAITK	0.18±0.03	6.5±0.4
TufA	LLGLMHTIDEAIPTPQR	0.23±0.03	8.2±0.7
	VNETVDIIGIK	0.22±0.03	7.3±0.6
flA	FFPEGTVFATTTYPATGTTTR	0.42±0.04	9.5±0.2
	VIPEQPEPTFYSR	0.32±0.04	12.9±0.7

B.

Level of analysis	Fthr		Total	
	WT	Δp0564	WT	Δp0564
PSMs	20±10	1800±300	25,000±3,000	24,000±3,000
Proteins	9±2	300±20	1630±90	1600±100

Table 2.1. *Assessment of fluorothreonine incorporation in the proteomes of wild-type and Δp0564::Am[R] S. cattleya strains. (A) Incorporation of Fthr in select proteins, as measured by targeted proteomics. Transitions corresponding to a set of Fthr- and Thr-containing peptides from abundant proteins were monitored, and relative peak areas were used to estimate the rate of incorporation. (B) Incorporation of Fthr across the proteome of S. cattleya, as measured by shotgun proteomics. Fthr-containing and total hits were tabulated at the level of peptide-spectrum matches and proteins. All measurements were performed with n ≥ 3; values are reported as mean ±s.e.*

The global landscape of FThr incorporation into protein was probed in an unbiased fashion with shotgun proteomics. FThr could be readily identified in tryptic peptides prepared from *Δp0564*::Am^R as expected, and was found much less frequently in the WT samples (*Table 2.1B*). FThr-containing peptides achieved high Peptide Spectrum Match (PSM) scores corresponding to low false-discovery rates, while also yielding positive fragmentation spectra suitable to confirm peptide sequence (*Figure 2.16*).

Data from the fluorothreonine-rich *Δp0564*::Am^R samples was further scrutinized to examine possible drivers of fluorothreonine incorporation. Given the extreme GC-rich bias exhibited at the wobble position of *Streptomyces* codons, differential incorporation by codon was considered as one possibility. Examination of the codon distribution of fluorothreonine- and threonine-containing PSMs in single-threonine coding peptides found quite similar codon distributions in both cases (*Table 2.2*). Peptide- and protein-level incorporation was also estimated using precursor ion peak areas. These results suggest that the integration of fluorothreonine varies across the proteome, with mean and median values consistent with the rate of incorporation measured by amino acid analysis *(Figure 2.17)*. Despite this variation, very few proteins have significantly elevated fluorothreonine content *(Appendix 2.1)*, possibly due to low signal-to-noise ratios for peaks corresponding to fluorothreonine-containing peptides. Differential accumulation of fluorothreonine in proteins on the basis of protein function was also considered; to this end COG gene ontology assignments for threonine- and fluorothreonine-containing proteins were tabulated. Again, a similar distribution was observed for both wild-type and *Δp0564*::Am^R strains (*Figure 2.18*). These results are consistent with an underlying random incorporation of fluorothreonine into the proteome, in the context of changing intracellular fluorothreonine concentrations over time. It is likely that the complete set of factors underlying protein-level variability in fluorothreonine incorporation are quite complex.

SCAT_p0564 expression mitigates fluorothreonine toxicity in a heterologous *Streptomyces* host. The role of the fluorothreonyl-tRNA deacylase in the context of FThr toxicity was probed via expression in a model *Streptomyces* species. Consistent with prior reports [*1*], FThr produces a zone of inhibition when applied to freshly plated spores of *Strepomyces coelicolor* M1152 on media with limited amino acid availability *(Figure 2.19)*. This inhibition can be partially alleviated by the genomic integration of SCAT_p0564 under the control of the constitutive ermEp* promoter. In comparison, the negative control with a strain where only the ermEp* promoter is integrated shows no evidence of protection against FThr toxicity. The effect of SCAT_p0564 expression is even more pronounced in liquid culture, though quantification of growth is impeded by the clumpy phenotype characteristic of streptomycetes *(Figure 2.20)*. FThr also inhibits the growth of the producing strain *S. cattleya* in disc diffusion assays, despite its ability to accumulate near-mM concentrations of this compound in fermentation media. While disruption of the fluorothreonyl-tRNA deacylase shows no distinguishable phenotypic change with respect to FThr growth inhibition compared to WT, disruption of the FThr transporter causes a marked increase in the ZOI *(Figure 2.21)*. Given the lower level of FThr incorporation into the proteome in the transporter knockout strain (*Figure 2.14D*), it is possible that other targets may play a role in growth defects observed under extremely high intracellular FThr concentrations, such as inhibition of serine hydroxymethyltransferase, or tRNA-dependent incorporation into lipids and cell wall constituents. However, the ability of SCAT_p0564 to mitigate fluorothreonine
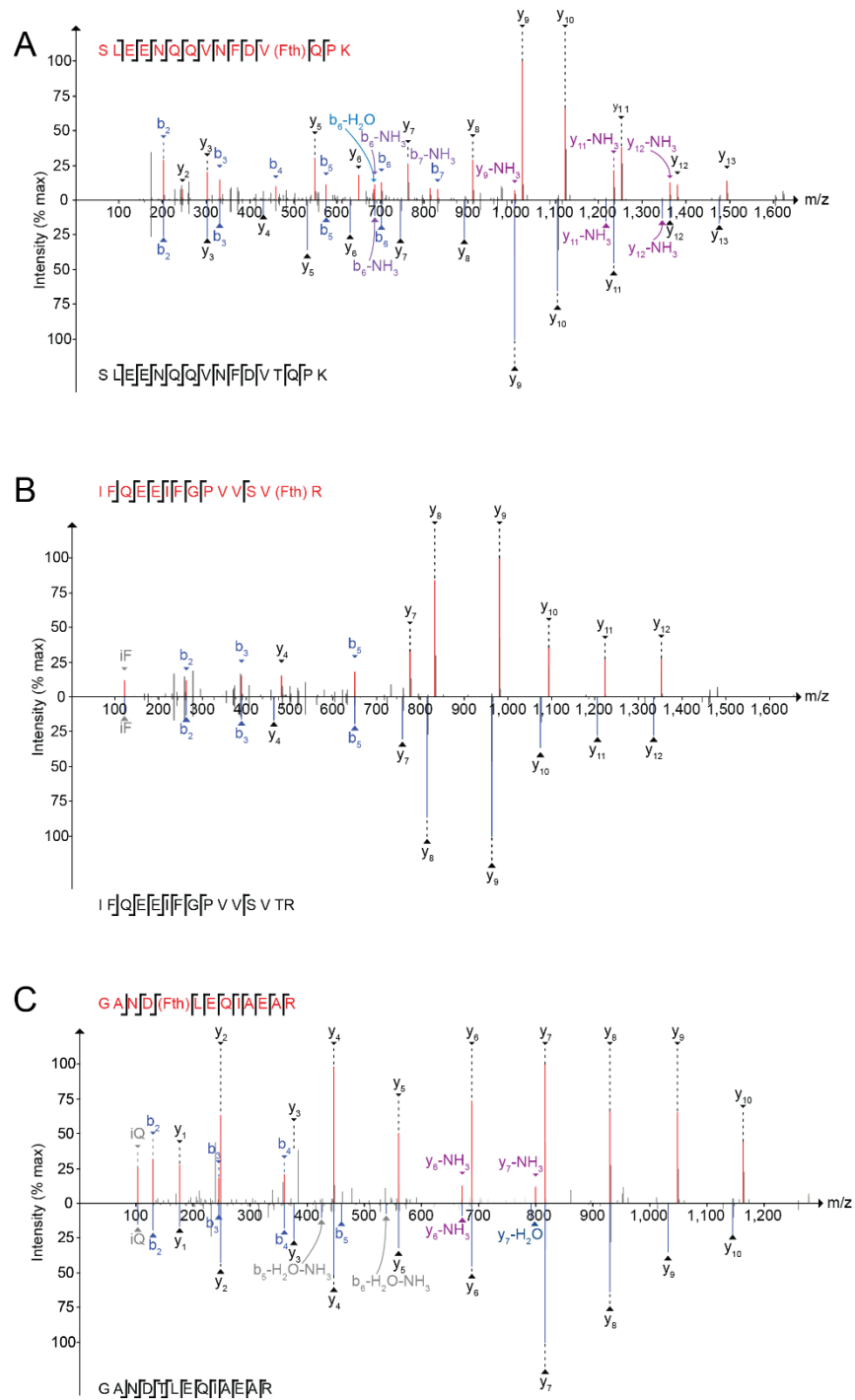
Figure 2.16. *High-resolution MS/MS spectra identifying fluorothreonine in peptides from S. cattleya. Spectra above the axis gave Peptide-Spectrum Matches (PSMs) to sequences with fluorothreonine substituted for threonine. The spectra below the axis are from the corresponding threonine-containing peptide. Peptides are from (A) the putative cold shock protein SCAT_3469, (B) the fluoroacetaldehyde dehydrogenase SCAT_0945, and (C) the putative superoxide dismutase SCAT_p0640 respectively.*

A

B
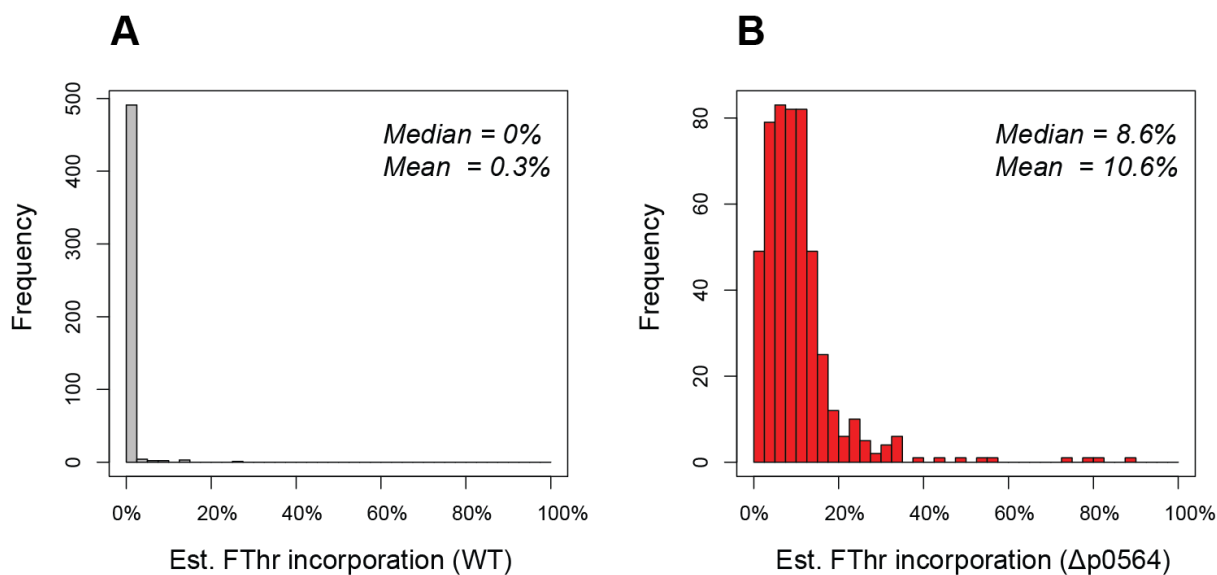
Figure 2.17. *Histograms representing estimated incorporation of fluorothreonine into peptides as measured by precursor ion intensity from shotgun proteomics datasets. PSMs corresponding to single-threonine encoding peptides with at least one FThr-containing PSM were analyzed. Data from wild-type (A) and Δp0564 (B) strains of S. cattleya are shown.*

Amino acid	ACA		ACC		ACG		ACT	
Fluorothreonine	1.7%	45	83.6%	2184	14.0%	365	0.7%	17
Threonine	1.1%	337	83.3%	25772	15.3%	4725	0.4%	121

Table 2.2. *Distribution of Peptide Spectrum Matches conditioned by threonine codon usage and presence of threonine or fluorothreonine. Frequency (left) and raw counts (right) are displayed. Only hits to peptides containing a single threonine are tabulated to eliminate ambiguity in codon usage. Peptide-spectrum matches are cumulative over four samples of S. cattleya Δp0564; the WT strain did not afford enough fluorothreonine-containing PSMs to assemble a meaningful dataset.*

		Thr-containing	Fthr-containing
■	General function prediction only	212 (14.0)	31 (10.95)
■	Amino acid transport and metabolism	146 (9.6)	29 (10.25)
■	Energy production and conversion	135 (8.9)	32 (11.31)
■	Carbohydrate transport and metabolism	124 (8.2)	26 (9.2)
■	Translation, ribosomal structure and biogenesis	115 (7.6)	28 (9.9)
■	Lipid transport and metabolism	108 (7.1)	21 (7.4)
■	Secondary metabolites biosynthesis, transport, and catabolism	100 (6.6)	16 (5.7)
■	Transcription	96 (6.3)	19 (6.7)
■	Function unknown	82 (5.4)	12 (4.2)
■	Signal transduction mechanisms	66 (4.4)	8 (2.8)
■	Post-translational modification, protein turnover, and chaperones	62 (4.1)	20 (7.1)
■	Coenzyme transport and metabolism	60 (4.0)	9 (3.2)
■	Cell wall/membrane/envelope biogenesis	48 (3.2)	11 (3.9)
■	Nucleotide transport and metabolism	48 (3.2)	8 (2.8)
■	Replication, recombination and repair	40 (2.6)	4 (1.4)
■	Inorganic ion transport and metabolism	35 (2.3)	4 (1.4)
■	Defense mechanisms	14 (0.9)	1 (0.4)
■	Intracellular trafficking, secretion, and vesicular transport	13 (0.9)	1 (0.4)
■	Cell cycle control, cell division, chromosome partitioning	7 (0.5)	1 (0.4)
■	Cell motility	4 (0.3)	1 (0.4)
■	RNA processing and modification	1 (2.16)	1 (0.4)
	Total	**1516**	**283**

Figure 2.18. *Gene ontology assignments for proteins with PSMs corresponding to (A) Thr-containing and (B) FThr-containing versions of peptides that encode a single threonine residue. (C) Legend and tabular data. Gene ontology assignments were obtained from the Microscope platform (http://www.genoscope.cns.fr/agc/microscope).*

S. coelicolor M1152

S. coelicolor M1152 + ermEp*-blank

S. coelicolor M1152 + ermEp*-SCAT_p0564

50 ± 1mm

53 ± 1mm

13 ± 2mm
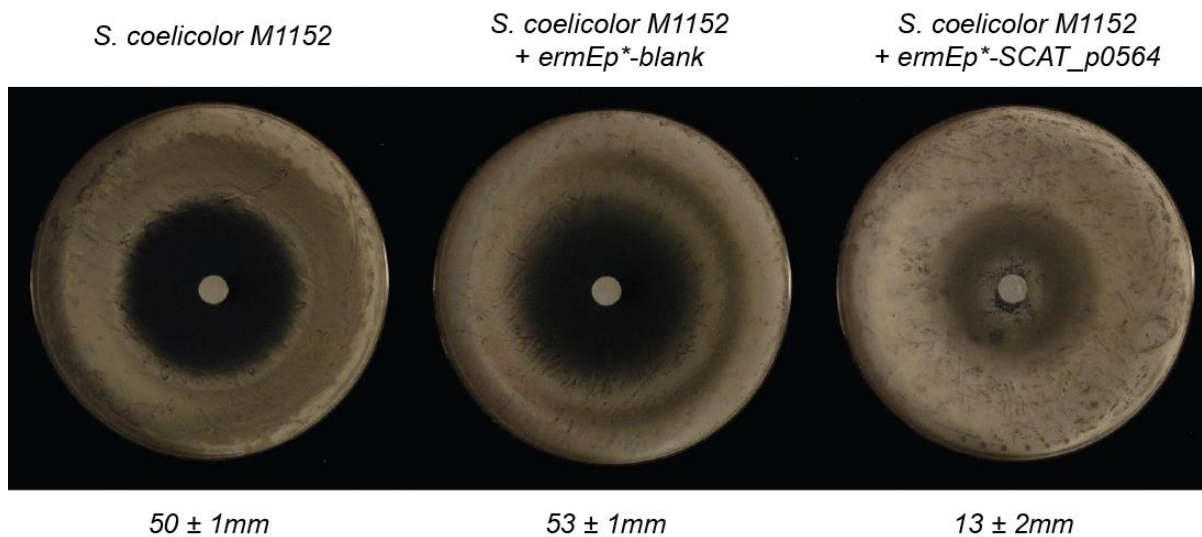
Figure 2.19. *Growth inhibition of S. coelicolor strains by fluorothreonine. Spores of S. coelicolor M1152, S. coelicolor M1152 + ermEp*-blank, and S. coelicolor M1152 + ermEp*-SCAT_p0564 were plated on Hopwood media, and 200 µg fluorothreonine was applied via filter paper. No-growth zones of inhibition are listed below the photographs. Plates were incubated 5 d at 30°C prior to observation.*

No fluorothreonine

10 µg/mL fluorothreonine
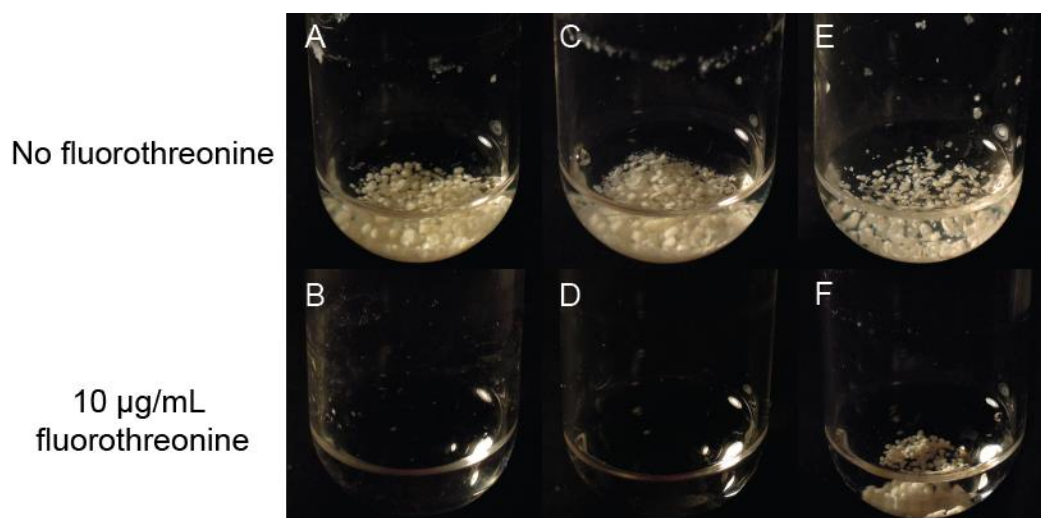
Figure 2.20. *Photographs of representative S. coelicolor M1152 cultures from fluorothreonine inhibitions assays. Strains are S. coelicolor M1152 in the absence and presence of fluorothreonine (A, B), S. coelicolor M1152 + ermEp*-blank in the absence and presence of fluorothreonine (C, D), and S. coelicolor M1152 + ermEp*-SCAT_p0564 in the absence and presence of fluorothreonine (E,F).*

S. cattleya WT   S. cattleya Δp0564::Am^R S. cattleya Δp0565::Am^R

49 ± 1mm   51 ± 1mm   66 ± 1mm
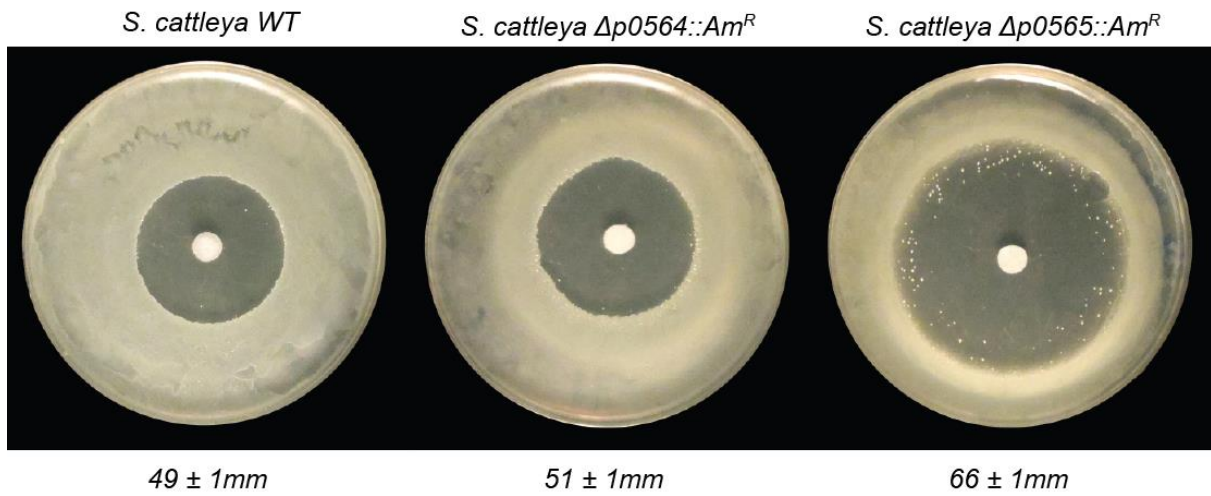
Figure 2.21. *Growth inhibition of S. cattleya strains by fluorothreonine. Spores of WT, Δp0564::AmR, and Δp0565::AmR were plated on Hopwood media, and 200 μg fluorothreonine was applied via filter paper. Plates were incubated 5 d at 30°C prior to observation.*

toxicity in *S. coelicolor* M1152 in both disc diffusion assays and liquid culture indicates that it is mediated in part by mistranslation.

2.4   Conclusions

The unusual ability of *S. cattleya* to utilize fluorine places a high demand on the cellular machinery to evolve new enzymes to handle the challenging selectivity problem of distinguishing a single substitution of fluorine for hydrogen. Similar to the fluoroacetate poison made by *S. cattleya*, FThr also causes toxicity that must be controlled. In this work, we explore the mechanisms evolved by *S. cattleya* to alleviate this toxicity and describe the discovery of a fluorine-selective aminoacyl-tRNA editing protein (SCAT_p0564) as well as a FThr transporter (SCAT_p0565) that are involved in this process.

FThr represents a natural candidate for mistranslation by ThrRS, and we find indeed that it is efficiently aminoacylated *in vitro* and *in vivo*. Biochemical characterization of SCAT_p0564 shows that it can hydrolyze fluorothreonyl-tRNA with 670-fold selectively in k_{cat}/K_M compared to threonyl-tRNA, suggesting that its physiological function is to protect against misincorporation of FThr into the host proteome. Given that members of the YbaK-like superfamily have not yet been observed to display selectivity with respect to the tRNA [*58*, *59*], it seems likely that the aminoacyl group remains the primary determinant of substrate specificity. In this regard, the high selectivity of SCAT_p0564 with respective to the single γ-fluorine substitution is quite striking. In comparison, other naturally-occurring fluorine-selective enzymes target the C-F bond itself in the case of the haloacetate dehalogenase, or hydrolyze a bond vicinal to the fluorine substituent in the case of the fluoroacetyl-CoA thioesterase FlK [*55*, *60*]. In contrast, the scissile bond of fluorothreonyl-tRNA is located quite distantly from the fluorine atom yet still enables a 10^2-fold effect on k_{cat}. While the catalytic basis for recognizing this subtle structural change remains unknown, it is possible that a difference in the pK_a of the β-hydroxyl group between threonyl- and fluorothreonyl-tRNA could be exploited to direct substrate positioning or to directly facilitate hydrolysis.

Genetic and physiological studies show that disruption of the gene encoding the fluorothreonyl-tRNA deacylase leads to an increase in the levels of tRNA charged with FThr as well as FThr incorporation into the proteome. In the absence of this activity, these two markers rise to quite high levels of ~18% and ~11.5% respectively. Further analysis demonstrates that FThr is indeed introduced at Thr sites across the proteome. At this time, we have not identified any significant bias in codon usage or protein identity that clearly indicate a physiological function for introduction of FThr into ribosomally-synthesized proteins and peptides. We have also assigned the function of SCAT_p0565 as a FThr transporter. Disruption of SCAT_p0565 leads to a significant increase in intracellular FThr concentration and increased toxicity associated with the addition of exogenous FThr. Both SCAT_p0564 and SCAT_p0565 alleviate toxicity, which may arise from mistranslation as well as other mechanisms that have yet to be determined. Taken together, the data suggest a model where the FThr transporter is required to avoid accumulation of critically toxic intracellular levels of FThr and that the fluorothreonyl-tRNA deacylase protects against growth defects resulting from the misincorporation of FThr into the proteome.

These two functions of transport and tRNA editing are conserved across other FThr and chlorothreonine biosynthetic gene clusters and thus appear to facilitate the management of halogenated amino acids that are closely related to a proteinogenic amino acid. Under normal

conditions, the ARS serve as the major filter for translational fidelity and are able to differentiate between cognate and noncognate substrates at the level of aminoacylation, or by hydrolytic editing domains encoded on the same polypeptide. However, *trans*-acting hydrolytic editing proteins have been found to resolve particularly difficult discrimination problems in other systems. For example, AlaX-like proteins can hydrolyze Ser-tRNA[Ala], while members of the YbaK-like superfamily have been shown to act on misacylated cysteinyl-, threonyl- and seryl-tRNAs [*23*, *24*, *50*]. In contrast to the substrates of the characterized members of these families, which are primarily tRNAs misacylated with other proteinogenic amino acids [*16*], SCAT_p0564 participates in mediating secondary metabolism. Given that homologous proteins are widespread in other *Streptomyces*, there are likely other cases where tRNA editing proteins have been recruited to secondary metabolism clusters that induce novel translational stresses.

2.5   References

1.   Sanada M, et al. (1986) Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya*. *J Antibiot* 39(2):259–265.

2.   Müller K, Faeh C, Diederich F (2007) Fluorine in pharmaceuticals: Looking beyond intuition. *Science* 317(5846):1881–1886.

3.   Le Bars D (2006) Fluorine-18 and medical imaging: Radiopharmaceuticals for positron emission tomography. *J Fluorine Chem* 127(11):1488–1493.

4.   Tirotta I, et al. (2015) ^{19}F Magnetic Resonance Imaging (MRI): From Design of Materials to Clinical Applications. *Chem Rev* 115(2):1106–1129.

5.   Berger R, Resnati G, Metrangolo P, Weber E, Hulliger J (2011) Organic fluorine compounds: a great opportunity for enhanced materials properties. *Chem Soc Rev* 40(7):3496–3508.

6.   Schaffrath C, Deng H, O'Hagan D (2003) Isolation and characterisation of 5′-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Lett* 547(1–3):111–114.

7.   Gribble GW (2003) The diversity of naturally produced organohalogens. *Chemosphere* 52(2):289–297.

8.   Clarke DD (1991) Fluoroacetate and fluorocitrate: Mechanism of action. *Neurochem Res* 16(9):1055–1058.

9.   Lauble H, Kennedy MC, Emptage MH, Beinert H, Stout CD (1996) The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *Proc Natl Acad Sci USA* 93(24):13699–13703.

10. Walker MC, Wen M, Weeks AM, Chang MCY (2012) Temporal and fluoride control of secondary metabolism regulates cellular organofluorine biosynthesis. *ACS Chem Biol* 7(9):1576–1585.

11. Weeks AM, Coyle SM, Jinek M, Doudna JA, Chang MCY (2010) Structural and biochemical studies of a fluoroacetyl-CoA-specific thioesterase reveal a molecular basis for fluorine selectivity. *Biochemistry* 49(43):9269–9279.

12. Webb HK, Matthews RG (1995) 4-Chlorothreonine is substrate, mechanistic probe, and mechanism-based inactivator of serine hydroxymethyltransferase. *J Biol Chem* 270(29):17204–17209.

13. Minajigi A, Deng B, Francklyn CS (2011) Fidelity escape by the unnatural amino acid β-hydroxynorvaline: an efficient substrate for *Escherichia coli* threonyl-tRNA synthetase with toxic effects on growth. *Biochemistry* 50(6):1101–1109.

14. Ogle JM, Ramakrishnan V (2005) Structural insights into translational fidelity. *Annu Rev Biochem* 74:129–177.

15. Fersht AR, Kaethner MM (1976) Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry* 15(15):3342–3346.

16. Bullwinkle T, Lazazzera B, Ibba M (2014) Quality control and infiltration of translation by amino acids outside of the genetic code. *Annu Rev Genet* 48(1):149–166.

17. Hartman MCT, Josephson K, Szostak JW (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci USA* 103(12):4356–4361.

18. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci USA* 103(25):9482–9487.

19. Melo Czekster C, Robertson WE, Walker AS, Söll D, Schepartz A (2016) In Vivo Biosynthesis of a β-Amino Acid-Containing Protein. *J Am Chem Soc* 138(16):5194–5197.

20. Rosenthal GA, Dahlman DL, Janzen DH (1976) A novel means for dealing with L-canavanine, a toxic metabolite. *Science* 192(4236):256–258.

21. Fowden L, Richmond MH (1963) Replacement of proline by azetidine-2-carboxylic acid during biosynthesis of protein. *Biochim Biophys Acta* 71:459–461.

22. Calendar R, Berg P (1967) D-Tyrosyl RNA: Formation, hydrolysis and utilization for protein synthesis. *J Mol Biol* 26(1):39–54.

23. Ahel I, Korencic D, Ibba M, Söll D (2003) Trans-editing of mischarged tRNAs. *Proc Natl Acad Sci USA* 100(26):15422–15427.

24. An S, Musier-Forsyth K (2004) Trans-editing of Cys-tRNA[Pro] by Haemophilus influenzae YbaK Protein. *J Biol Chem* 279(41):42359–42362.

25. Soutourina J, Plateau P, Delort F, Peirotes A, Blanquet S (1999) Functional Characterization of the D-Tyr-tRNA^{Tyr} Deacylase from *Escherichia coli*. *J Biol Chem* 274(27):19109–19114.

26. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.

27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30(12):2725–2729.

28. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275–282.

29. Marchler-Bauer A, et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):D222-226.

30. Amin MR (1996) Biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. Doctoral (Durham University). Available at: http://etheses.dur.ac.uk/5428/ [Accessed October 10, 2016].

31. Palmer BR, Marinus MG (1994) The dam and dcm strains of Escherichia coli--a review. *Gene* 143(1):1–12.

32. Paget MSB, Chamberlin L, Atrih A, Foster SJ, Buttner MJ (1999) Evidence that the Extracytoplasmic Function Sigma Factor ςE Is Required for Normal Cell Wall Structure in Streptomyces coelicolor A3(2). *J Bacteriol* 181(1):204–211.

33. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.

34. Kieser T (2000) *Practical Streptomyces Genetics* (John Innes Foundation).

35. Foulston LC, Bibb MJ (2010) Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proc Natl Acad Sci U S A* 107(30):13461–13466.

36. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, ed Walker J (Humana Press), pp 571–607.

37. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif* 41(1):207–234.

38. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955–964.

39. Cavaluzzi MJ, Borer PN (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res* 32(1):e13.

40. Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* 100(4):1541–1546.

41. Cock PJA, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423.

42. Vallenet D, et al. (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* 2009. doi:10.1093/database/bap021.

43. Vaudel M, et al. (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotechnol* 33(1):22–24.

44. Wenger CD, Coon JJ (2013) A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. *J Proteome Res* 12(3):1377–1386.

45. Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Perez-Maya AA, Ocampo-Candiani J (2012) Complete genome sequence of *Nocardia brasiliensis* HUJEG-1. *J Bacteriol* 194(10):2761–2762.

46. Deng H, et al. (2014) Identification of fluorinases from *Streptomyces* sp. MA37, *Norcardia brasiliensis*, and *Actinoplanes* sp. N902-109 by genome mining. *ChemBioChem* 15(3):364–368.

47. Huang S, et al. (2014) Fluoroacetate biosynthesis from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B-24674. *Org Biomol Chem* 12(27):4828.

48. Huang H, Ren S-X, Yang S, Hu H-F (2015) Comparative analysis of rapamycin biosynthesis clusters between *Actinoplanes* sp. N902-109 and *Streptomyces hygroscopicus* ATCC29253. *Chinese Journal of Natural Medicines* 13(2):90–98.

49. Ruan B, Söll D (2005) The bacterial YbaK protein is a Cys-tRNA[Pro] and Cys-tRNA [Cys] deacylase. *J Biol Chem* 280(27):25887–25891.

50. Liu Z, et al. (2015) Homologous trans-editing factors with broad tRNA specificity prevent mistranslation caused by serine/threonine misactivation. *Proc Natl Acad Sci USA* 112(19):6027–6032.

51. Fullone MR, et al. (2012) Insight into the structure–function relationship of the nonheme iron halogenases involved in the biosynthesis of 4-chlorothreonine – Thr3 from *Streptomyces* sp. OH-5093 and SyrB2 from *Pseudomonas syringae pv. syringae* B301DR. *FEBS J* 279(23):4269–4282.

52. Stepanov VG, Moor NA, Ankilova VN, Lavrik OI (1992) Phenylalanyl-tRNA synthetase from *Thermus thermophilus* can attach two molecules of phenylalanine to tRNA[Phe]. *FEBS Lett* 311(3):192–194.

53. Chen YG, Kowtoniuk WE, Agarwal I, Shen Y, Liu DR (2009) LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat Chem Biol* 5(12):879–881.

54. Richardson CJ, First EA (2015) A continuous tyrosyl-tRNA synthetase assay that regenerates the tRNA substrate. *Analytical Biochemistry* 486:86–95.

55. Weeks AM, Chang MCY (2012) Catalytic control of enzymatic fluorine specificity. *Proc Natl Acad Sci USA* 109(48):19667–19672.

56. Jakubowski H, Goldman E (1984) Quantities of individual aminoacyl-tRNA families and their turnover in *Escherichia coli*. *J Bacteriol* 158(3):769–776.

57. Ahmad S, et al. (2013) Mechanism of chiral proofreading during translation of the genetic code. *Elife* 2:e01519.

58. Das M, Vargas-Rodriguez O, Goto Y, Suga H, Musier-Forsyth K (2014) Distinct tRNA recognition strategies used by a homologous family of editing domains prevent mistranslation. *Nucleic Acids Res* 42(6):3943–3953.

59. Ruan L-L, Zhou X-L, Tan M, Wang E-D (2013) Human cytoplasmic ProX edits mischarged tRNA[Pro] with amino acid but not tRNA specificity. *Biochem J* 450(1):243–252.

60. Liu J-Q, Kurihara T, Miyagi M, Esaki N, Soda K (1995) Reaction Mechanism of L-2-Haloacid Dehalogenase of *Pseudomonas* sp. YL. Identification of Asp10 as the active site nucleophile by [18]O incorporation experiments. *J Biol Chem* 270(31):18309–18312.

Chapter 3: *Structural and biochemical investigation of tRNA editing proteins with activity towards halogenated threonine analogues*

3.1   Introduction

In response to the stringent fidelity requirements of translation, the aminoacyl-tRNA synthetases have evolved a variety of strategies for the post-transfer hydrolysis of mischarged tRNAs. The YbaK-like and AlaX-like families serve diverse functions in the editing of type II tRNAs, and can be found fused to the ARS itself or as an independent polypeptide [1]. Strikingly, the need to enforce homochirality in translation has driven the evolution of three distinct protein families dedicated to the hydrolysis of D-aminoacyl-tRNAs [2–4]. Many type I ARS also possess posttransfer editing domains of the CP1 family [5]. While these strategies have generally been understood to apply to primary metabolism, they have also been adapted as resistance mechanisms in secondary metabolism *(Chapter 2)*. Further investigation of these enzymatic activities could help to understand how enzymes can achieve selectivity towards biologically rare functional groups.

Despite the panoply of tRNA editing proteins that have been characterized, it has been challenging to establish a detailed understanding the basis of substrate selectivity. The macromolecular nature of the substrates, heavy reliance on substrate functional groups in catalysis, and limited options for detailed rate measurements have conspired to hamper mechanistic characterization of aminoacyl-tRNA hydrolysis. In fact, the best studied analogous system may not be an editing protein at all. Like post-transfer editing proteins, the peptidyl transferase center of the ribosome activates a nucleophile to attack the carbonyl of the aminoacyl moiety. Both kinetic and computational studies support a concerted reaction mechanism in which amide bond formation occurs in conjunction with the transfer of a proton to the 3'-oxygen via the 2'-hydroxyl and a bridging water [6–8]. Analytical challenges aside, studies on diverse aminoacyl-tRNA editing proteins have converged on a set of similar principles. As in the case of the peptidyltransferase center, the substrate is key, with the 2'-hydroxyl having been proposed to participate in catalysis in a diverse set of structurally unrelated editing proteins based on experiments with Cys-tRNA^Pro editing protein YbaK and the D-Tyr deacylase [9–11]. Aminoacyl side chain groups may also play a role in the case of YbaK, which has been proposed to hydrolyze its substrate via a nucleophilic attack with the substrate sulfhydryl [9, 12]. In contrast to the importance of the substrate, tRNA editing proteins are often agnostic to mutations in active site residues, and interactions between backbone functional groups and the substrate may dominate over sidechain-based interactions, with some exceptions such as in ProRS and ThrRS [12–15].

Here I describe strategies to begin characterizing the basis of the selective hydrolysis of γ-halogenated tRNAs by post-transfer tRNA editing proteins from amino acid BGCs. The crystal structure of the fluorothreonyl-tRNA editing protein SCAT_p0564 was solved, and used to identify conserved features that could contribute to the observed substrate selectivity. While structure-guided mutagenesis failed to identify residues that play a dramatic role in dictating selectivity, it was observed that both chloro- as well as fluorothreonine was an excellent substrate for SCAT_p0564. This finding implicates β-hydroxyl pK_a as a potential determinant of the substrate specificity of SCAT_p0564. Preliminary characterization of an AlaX-like putative editing protein from the chlorothreonine biosynthesis cluster suggests that *Streptomyces* species have evolved distinct strategies for discriminating against unnatural threonine analogues. As substrate analogues have helped to gain insight into the activity of diverse post-transfer editing proteins [10, 13], I also present a set of strategies oriented towards the biosynthesis of nonhydrolyzable substrate analogues for further characterization of SCAT_p0564.

3.2   Materials and methods

Commercial materials.   Amylose resin, Hiscribe T7 transcription kit, Phusion DNA polymerase, Q5 DNA polymerase, deoxynucleotide triphosphates (dNTPs), and restriction enzymes were obtained from New England Biolabs (Ipswich, MA). β-mercaptoethanol, β-Nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt hydrate, boric acid, 3-chloroperbenzoic acid, copper(II) chloride, Dowex 50WX8, lead(II) acetate, lithium chloride, inorganic pyrophosphatase, myokinase, protamine sulfate from salmon sperm, and pyruvate kinase/lactic acid dehydrogenase from rabbit muscle were purchased from Sigma-Aldrich (St. Louis, MO). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators, Milli-Q Gradient water purification system, and chloramphenicol were purchased from EMD Millipore (Billerica, MA).   Adenosine triphosphate disodium salt hydrate, ammonium chloride, carbenicillin, 3 kDa MWCO dialysis tubing, dithiothreitol, ethylenediaminetetraacetic acid disodium salt, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), hydrochloric acid, glucose, kanamycin sulfate, magnesium chloride, magnesium sulfate, phenylmethylsulfonyl chloride (PMSF), potassium phosphate dibasic, sodium acetate, sodium hydroxide, and tris base were purchased from Fisher Scientific (Waltham, MA). Deuterium oxide was purchased from Cambridge Isotopes (Tewksbury, MA). Fluka brand LCMS grade ammonium formate, ammonium bicarbonate, formic acid, methanol and acetonitrile were purchased from Sigma Aldrich (St. Louis, MO). Phenol chloroform (pH 5.2), and sodium chloride glutamate were purchased from MP Biosciences. 1-Benzyl N-Carbobenzoxy-L-glutamate was purchased from TCI America (Portland, OR). Tris(2-carboxyethyl)phosphine (TCEP) was purchased from Biosynth Corporation (Itasca, IL). Ni-NTA resin, Qia-quick PCR cleanup kit, and Qia-spin miniprep kit were purchased from Qiagen USA (Valencia, CA). Bacto^{TM} Agar was purchased from BD (Sparks, Maryland). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a concentration of 100 µM, and stored at 4°C. Ammonium persulfate, acrylamide/bisacrylamide (37.5:1 and 19:1), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), and sodium dodecyl sulfate (SDS) were purchased from Bio-Rad Laboratories (Hercules, CA). Lysogeny Broth (LB), Lysogeny Broth Agar (LBA), and Terrific Broth (TB) were purchased from VWR International (Radnor, PA). Ultrayield baffled flasks were purchased from Thompson Instrument Company (Oceanside, CA).

X-ray crystallography of SCAT_p0564. Protein was prepared with slight modifications from the protocol described in *Chapter 2*. BL21(DE3) cells bearing the pRare2 plasmid were transformed with the expression vector MBP-p0564. Overnight cultures grown in Terrific Broth were inoculated to OD_{600} into Terrific Broth supplemented with 17 mM sodium phosphate monobasic, 82 mM sodium phosphate dibasic. Cultures were grown to an OD_{600} of 0.6, and were induced by the addition of 0.2 mM IPTG before growth overnight at 30°C. Cells were harvested by centrifugation, and resuspended in 300 mM sodium chloride, 50 mM HEPES pH 8, 10% (*v/v*) glycerol, 1 mM TCEP, 500 µM PMSF, 1 mg/mL lysozyme before being stored at -80°C. Cell slurry was thawed, lysed by sonication, and clarified by centrifugation for 1 h at 15,000 RCF. DNA was then precipitated by addition of protamine sulfate to 0.075% (*w/v*) and centrifugation at 15,000 RCF for 30 min. The clarified lysate was applied to 12.5 mL NiNTA resin and washed with 40 column volumes wash buffer (300 mM sodium chloride, 25 mM HEPES pH 8, 10% (*v/v*) glycerol, 1 mM TCEP, 20 mM imidazole). Protein was eluted with wash buffer supplemented with 250 mM imidazole to afford ~110 mg of MBP-fusion. After 1 h of dialysis in storage buffer (100 mM sodium chloride, 25 mM HEPES pH 8, 10% (*v/v*) glycerol, 1 mM TCEP)

TEV protease was added at a 1:50 ratio with the MBP fusion and dialysis was continued overnight. The cleaved protein was passed over 5 mL amylose resin, then nickel resin, before concentration using a 10 kDa cutoff Amicon spin concentrator. The protein was then applied to a Superdex 75 10/300 GL column (GE Healthcare Life Sciences, Pittsburg PA); fractions corresponding to monomeric SCAT_p0564 were collected. The desired fractions were concentrated to ~15 mg/mL, frozen in liquid nitrogen, and stored at -80°C.

For SeMet labeling, expression was performed in M9 minimal media. Overnight cultures were grown in 2×YT, were used to inoculate 50 mL seed cultures in 2×YT. Seed cultures were grown to an OD_{600} of 0.6, and were pelleted by centrifugation for 5 min at 8,000 RCF. Pellets were washed by addition of 50 mL M9 media followed by another round of centrifugation. The washed pellets were resuspended in 50 mL M9, and 10 mL of this seed solution were used to inoculate 1 L of M9 media, which was grown at 37°C. Cultures were grown to an OD_{600} of 0.4, at which point a 10 mL of a SeMet-containing amino acid mix (threonine, lysine, and phenylalanine at 5 mg/mL; isoleucine, leucine, and valine at 10 mg/mL, and SeMet at 7.5 mg/mL) was added. After 25 min, IPTG was added to 0.2 mM and growth was continued overnight at 30°C. Cells were harvested by centrifugation and resuspended in 300 mM sodium chloride, 50 mM HEPES pH 8, 10% (*v/v*) glycerol, 5 mM TCEP, 500 µM PMSF, and 1 mg/mL lysozyme before storage at -80°C. The high reducing agent concentration was intended to prevent oxidation of SeMet residues. Affinity purification on NiNTA was carried out as described for native crystallography preps, with the exception that TCEP was used at 2 mM instead of 1 mM. Final cleanup was carried out by 1 mL Mono Q column (GE life sciences), using a linear gradient from 0 to 500 mM sodium chloride over 36 column volumes. Fractions containing SCAT_p0564 were concentrated using a 30 kDa Amicon spin concentrator, diluted with 100 mM sodium chloride, 50 mM HEPES pH 8, 10% (*v/v*) glycerol, 2 mM TCEP, re-concentrated, flash frozen in liquid nitrogen, and stored at -80°C. Selenomethionine labeling was confirmed by whole-protein mass spectrometry.

Initial crystallography screens were carried out by in sitting drop format. Protein (200 nL) and buffer (200 nL) were mixed in 96 × 3 well Intelliplates using a Mosquito liquid handler, and plates were incubated at 10°C. The JCSG+, AmSO₄, PEGs I, and PEGs II screens (Hampton research, Aliso Viejo CA) were initially used. Low quality hits were obtained in 1.8 M sodium citrate, pH 8 and 2.4 M sodium malonate, pH 7. While screening similar conditions and the use of additive screens failed to substantially improve crystal quality, streak seeding in hanging drop format was quite successful. Seeded drops would develop visible crystals within 12 h, progressing to a high density of flake-like crystals within 24 h. To obtain SeMet labeled crystals, protein (6 mg/mL) was mixed with 1.6 M sodium citrate on a glass cover slip, and a human hair mounted on a plastic spreader was used to swipe crushed crystals through the drops. By sequentially seeding a series of 4 drops, a crude serial dilution of seed crystals could be obtained. Crystals were looped and frozen directly in liquid nitrogen, relying on the high concentration of citrate as a cryoprotectant [*16*].

X-ray diffraction data was collected at BL8.3.1 at Lawrence Berkeley National Lab. Data was collected from a 360° rotation of the crystal in 0.1° oscillation angle with an exposure time of 0.1 s, at X-ray energies of 13,000 eV and 12,656.1 eV. The resulting datasets were merged and scaled with XDS and XSCALE [*17*]. Phasing was accomplished by molecular replacement in conjunction with single-wavelength anomalous dispersion (MR-SAD) as implemented in Phenix [*18*].The dataset collected at 12,656.1 eV was first used to obtain a marginal molecular replacement solution (TFZ = 9.4, LLG = 83) with an ensemble constructed from PDB files 3OP6,

2CX5, 2DXA,1VJF, 1VKI, 1WDV, and 1DBU using Phenix.Ensembler. The resulting phases was improved by the application of MR-SAD using Phenix.Phaser-EP. The resulting model was then re-built with Phenix.Autobuild [*19*], and used as a template for molecular replacement with the data collected at 13,000 eV. Repeated cycles of automated refinement with Phenix.Refine and manual refinement with Coot gave the final model [*20, 21*]. Phenix.Refine was operated with XYZ coordinates, real-space, individual B-factors, and occupancies as refinement strategies. The resolution cutoff for final refinement was set to 1.7Å, as this was the highest resolution shell with an average I/σ greater than 1. The CC* scores for this shell were also greater than 0.5, supporting the inclusion of data to at least this resolution [*22*]. Due to high B-factors observed for many solvent-exposed sidechains, residues with poorly defined sidechain density were built with alternate sidechain conformations.

Cloning. All constructs were prepared using Gibson ligation. Point mutants of SCAT_p0564 were constructed using 3-piece Gibson ligations with PCR products. The C49A mutant was amplified using primers J118/J503 and J119/J504; the C49S mutant using J118/J505 and J119/J506; the A104S mutant using J118/J507 and J119/J508; the T32A mutant using J118/J511 and J119/J512; the R31A mutant using J118/J513 and J119/J514; and the K48A mutant using J118/489 and J119/J490. The coding sequence for the Cl-SAD was ordered as the gBlock J411 from IDT (Coralville, IA), and was ligated into SfoI/HindIII digested PSV272.1. The CCA-adding enzyme was cloned by amplification from *E. coli* gDNA with primers J517/J518 and ligation into NdeI/BamHI digested pet16b-IMDH. The construct pET16b-tRNA25 was constructed by amplification from *S. cattleya* gDNA with primers J425F/J425R and insertion into BamHI/XbaI digested pET16b. The primers J425F/J425R flank the tRNA with sequences from the 3′- and 5′- sides of *E. coli* tRNA^{Ala} in order to facilitate efficient processing of the tRNA.

Multiple sequence alignments. Multiple sequence alignments of SCAT_p0564 homologues were performed as described in *Chapter 2*.

Synthesis of fluorothreonine and chlorothreonine. Fluorothreonine and chlorothreonine were synthesized as reported by Amin [*23*] and Webb [*24*], respectively.

Purification of the Cl-SAD. *Escherichia coli* BL21(DE3) cells bearing pRare2 were transformed with the expression plasmid MBP-Cl-SAD. Overnight cultures in Terrific Broth were used to innoculate expression cultures to OD$_{600}$ = 0.05 in 500 mL TB in an UltraYield flask.[12] Cultures were grown at 37°C until the OD$_{600}$ reached 0.7, at which point they were chilled on ice for 30 min before induction with 200uM IPTG and overnight growth at 16C. Cells were harvested by centrifugation and frozen at -80°C. The pellet was thawed and resuspended in 5 mL/g cell paste of lysis buffer (50 mM HEPES pH 7.5, 300 mM sodium chloride, 10 mM magnesium chloride, 5 mM BME, 10% (*v/v*) glycerol, 500 µM PMSF, 1 mg/mL lysozyme) and incubated on ice for 30 min before passage through a French press. Lysate was cleared by centrifugation for 20 m at 15,000 RCF, and DNA was precipitated by the addition of protamine sulfate to 0.075% (*w/v*) followed by 40 m of centrifugation at 15,000 RCF. Cleared lysate was loaded onto a 5 mL amylose column and washed with 20 column volumes lysis buffer, then 10 mL of low-salt wash buffer (50 mM HEPES pH 7.5, 50 mM sodium chloride, 10 mM magnesium chloride, 5 mM BME, 10% (*v/v*) glycerol). MBP-fusion was eluted with low-salt wash buffer supplemented with 10 mM maltose. The resulting eluent (~10 mg fusion protein) was diluted 2-fold, applied to a 1 mL Mono Q column (GE life sciences) and eluted with a 20 column-volume gradient from 0-1 M sodium chloride. Two distinct peaks were evident at 150 and 200 mM sodium chloride; both fractions were concentrated with a 10 kDa cutoff Amicon spin concentrator, flash frozen, and stored at -

80°C. For rate measurements, protein was thawed and diluted with into 25 mM HEPES pH 8, 100 mM sodium chloride, 1 mM TCEP, 10% (*v/v*) glycerol. The concentration of the Cl-SAD was calculated using an extinction coefficient of ε_{280} = 79,425 M^{-1} cm^{-1}, as calculated by ExPasy ProtParam [*25*] and assuming the protein was entirely in the MBP-fusion form.

Whole protein mass spectrometry. Whole protein MS was performed using an Agilent 1290 HPLC connected to a 2.1 mm × 50 mm Agilent RRHD SB-C3 300Å 1.8 μm column connected to an Agilent 6530 QTOF. Mobile phase A was comprised of 0.1% formic acid, and mobile phase B was comprised of acetonitrile. Approximately 1 μg of protein was injected per run, and chromatography was performed using a gradient of 0% mobile phase B for 1 min followed by a linear gradient to 25% B over 8 min. Drying gas was applied at 13 L/min and 325°C, and a fragmentor voltage of 175 V was used. The mass of the parent ion was deconvoluted manually by picking a series of peaks corresponding to consecutive charge states, and selecting the charge state that yielded the least variation in inferred parent ion mass.

Purification of the CCA-adding enzyme. *E. coli* BL21 (DE3) cells were transformed with pET16b-CCA. Overnight cultures were used to inoculate ZY5052 media to an initial OD$_{600}$ of 0.05. Cells were grown overnight at 30°C, harvested by centrifugation, resuspended in 50 mM HEPES pH 8, 300 mM sodium chloride, 10% (*v/v*) glycerol, 1 mM TCEP, 1 mg/mL lysozyme, and 500 μM PMSF before storage at -80°C. Frozen cell slurry was thawed and sonicated using a Misonix 3000 with pulses of 5 s at maximum power followed by 25 s rest, for a total process time of 45 seconds. Lysate was clarified by centrifugation for 50 m at 11,000 RCF, at which point polyethyleneimine was added to 0.01% final concentration to precipitate DNA. Centrifugation was continued for another 30 m at 11,000 RCF, and the supernatant was loaded onto 3 mL of NiNTA resin. The column was washed with 25 mM HEPES pH 8, 300 mM sodium chloride, 10% (*v/v*) glycerol, 1 mM TCEP, 20 mM imidazole until the A$_{280}$ of the wash dropped below 0.1. Protein was eluted with wash buffer supplemented with 250 mM imidazole. Prescission protease was added, and the protein was dialyzed against 25 mM hepes pH 8, 100 mM sodium chloride, 10% (*v/v*) glycerol, 1 mM TCEP overnight at 4°C. The protein was then run 3 times over an NiNTA column, and flowthrough was concentrated with a 10 kDa MWCO Amicon spin filter. The concentration of protein was calculated using an extinction coefficient of ε_{280} = 46,701 M^{-1} cm^{-1}, as calculated by ExPasy ProtParam [*25*].

Preparation of 3′-deoxy tRNA. Nucleotide exchange reactions were run in 30 mM HEPES pH 7.6, 30 mM potassium chloride, 6 mM magnesium chloride, 5 mM dATP, 2 mM DTT, 1 mg/mL BSA with a tRNA concentration of 20 μM and a CCA-adding enzyme concentration of 10 μM. The reaction was allowed to proceed for 2 h at room temperature before quenching into phenol-chloroform, pH 5.2.

Aminoacylation and deacylation assays. Aminoacylation and 3′-end exchange reactions were monitored using RNAse A digestion and LCMS/MS, generally as described in *Chapter 2*. The substitution of 3′ deoxyadenosine was monitored using the 254→136 transition, while the charging of 3′ deoxyadenosyl tRNA with fluorothreonine was monitored with the 371→136 transition. Charging of chlorothreonine onto standard tRNA was monitored with the 403→136 and 405→136 transitions. Standards were not prepared, so the chromatograms presented in this chapter are qualitative in nature. Hydrolysis measurements were performed using the coupled assay described in *Chapter 2*. The tRNA substrate was SCAT_tRNA25 prepared from *E. coli* expression, present at a concentration of 5 μM. SCAT_p0564 mutants were prepared as described in for protein crystallography, with the omission of any polishing step.

Preparation of *S. cattleya* tRNA^{Thr} from *E. coli*. *E. coli* BL21(DE3) cells were transformed with the plasmid pET16b-tRNA25. Overnight cultures were inoculated to an OD_{600} of 0.05 into 1 L Terrific Broth and grown to an OD_{600} of 1.1 before induction with 1 mM IPTG.   Shortly after induction growth ceased, presumably due to the toxicity of tRNA overexpression. After 2.5 h of incubation at 37°C, cells were harvested by centrifugation. The pellet was resuspended in 10 mM Tris pH 8, 20 mM BME, 6 M guanidinium hydrochloride, and 4 volumes phenol chloroform was added and the pH was adjusted to 7.5 before stirring for 1 h at room temperature. The resulting mixture was pelleted by centrifugation for 20 min at 9,000 RCF. The aqueous layer was then extracted again with 1 volume phenol chloroform before precipitation with 1 volume isopropyl alcohol. The resulting crude nucleic acid was resuspended in 20 mM sodium acetate, 10 mM magnesium chloride, pH 5.2, and applied to a 20 mL column of DEAE sepharose [*26*]. The column was then washed with 50 mL of 300 mM sodium chloride, 20 mM sodium acetate, 10 mM magnesium chloride, pH 5.2, and eluted with 700 mM sodium chloride, 20 mM sodium acetate, 10 mM magnesium chloride, pH 5.2. Eluent was precipitated with 1 volume isopropanol, washed with 70% (*v/v*) ice-cold ethanol, and resuspended in nuclease-free water before storage at -80°C. Approximately 2 µmol of tRNA could be produced from 1 L of culture, and the material obtained had approximately 60% the amino acid accepting activity of *in vitro* transcribed tRNA. Concentration of tRNA was measured as described in *Chapter 2*, assuming the sample was comprised entirely of SCAT_tRNA25.

For improved fractionation, a modified protocol was employed with a gradient separation. The pellet was resuspended in 20 mL of 100 mM sodium acetate, pH 5. The cell suspension was then lysed and extracted with an equal volume of phenol-chloroform pH 5.2 before phase separation by centrifugation for 5 min at 9,000 RCF. The organic layer was then back-extracted with an equal volume of 100 mM sodium acetate, pH 5. Pooled aqueous fractions were further extracted with 10 mL phenol chloroform, then with 10 mL chloroform, before temporary storage at -20°C. The tRNA^{Thr} was enriched using a 50 mL DEAE sepharose column connected to an Akta FPLC. The system was initially prepared by washing in 1 M sodium hydroxide to remove and denature residual protein, and was subsequently washed with 20 mM sodium acetate pH 5.2, 10 mM magnesium chloride, 1000 mM sodium chloride until the pH dropped below neutral. The crude RNA preparation was then loaded onto the column. After an initial isocratic wash with 50 mL of 100 mM sodium chloride and 10 mM magnesium chloride in 20 mM sodium acetate pH 5.2, a linear gradient to 1 M sodium chloride over 300 mL was applied. The tRNA elutes between 500 and 750 mM sodium chloride, while ribosomal RNA binds essentially irreversibly. Fractions were screened for fluorothreonine accepting capacity, with aminoacylation reactions allowed to proceed for 30 min. Fractions containing the highest fluorothreonyl-tRNA accepting activity were precipitated with 1 volume isopropanol, washed with ice-cold 70% (*v/v*) ethanol, and resuspended in nuclease-free water.

3.3 Results and discussion

Structural characterization of SCAT_p0564. The structure of SCAT_p0564 was initially solved by MR-SAD, using an ensemble of homologous proteins as an initial search template. The structure was solved to a resolution of 1.7Å, which is more than suitable to determine the overall architecture of the protein. The numbering scheme used refers to the untagged sequence of SCAT_p0564. Two monomers are found per asymmetric unit, although it is not believed that the interface is physiologically relevant. The overall fold is reminiscent of the previously crystallized *H. influenzae* YbaK [*27*]. The dominant feature is a 4-stranded, discontiguous β-sheet that runs along the long axis of the protein, the middle two strands of which are extended by an additional two residues in SCAT_p0564 with respect to the YbaK. The invariant lysine at position 48, which is conserved across the entire YbaK-like superfamily [*28*], lies on the central strand of the β-sheet and protrudes into the putative substrate binding pocket. It was noted that the adjacent C49 is conserved amongst actinobacterial homologues of SCAT_p0564. Also flanking this proposed substrate binding pocket are the conserved GXXXP loop, which is proposed to act as an oxyanion hole [*1*, *12*, *29*], and an additional, discontiguous flap containing the residue R31, which is conserved across homologues from organofluorine biosynthesis clusters. Visualization of residue-level B-factors reveals that the model quality varies over the surface of the model. While the proposed substrate binding pocket is quite ordered, several loop regions on the opposite side of the protein have quite high temperature factors, with sidechain density being largely ill-defined *(Figure 3.2C).*
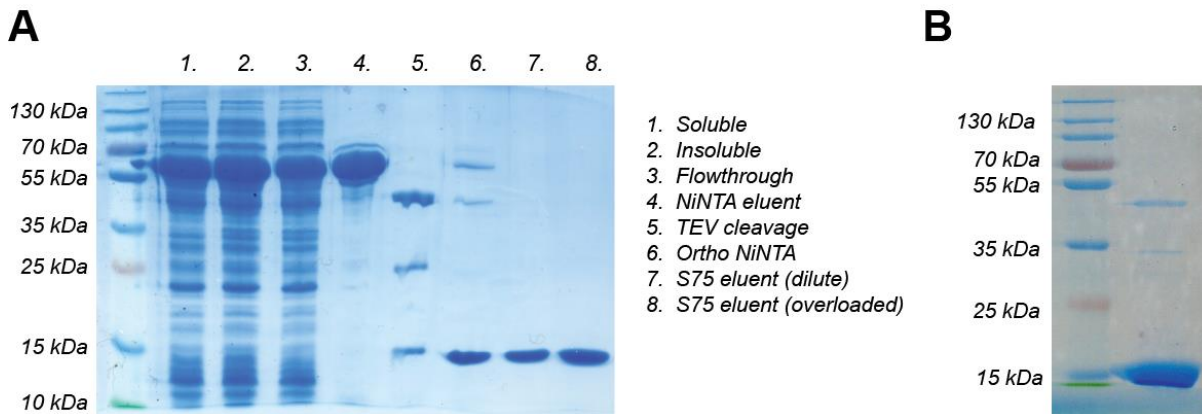
Figure 3.1. *Modified protein preps used for crystallography of SCAT_p0564. Size exclusion chromatography (A) or anion exchange (B) were added as polishing steps to improve the purity of the protein before crystallization.*

Data collection	
Wavelength	0.95368 Å
Resolution range [a]	59.41 - 1.7 (1.761 - 1.7)
Space group	C 1 2 1
Unit cell	
a, b, c (Å)	118.96, 38.11, 99.00
α, β, γ (deg)	90, 124.108, 90
Total reflections [a]	272463 (25982)
Unique [a]	40806 (4030)
Multiplicity [a]	6.7 (6.4)
Completeness (%) [a]	1.00 (0.99)
Mean I/σ(I) [a]	14.90 (1.23)
Wilson B-factor (Å^2)	24.35
R-merge [a]	0.08753 (1.352)
R-meas [a]	0.09493 (1.471)
CC1/2 [a]	0.999 (0.533)
CC* [a]	1 (0.834)
Refinement	
Reflections used in refinement [a]	40801 (4030)
Reflections used for R-free [a]	1405 (133)
R-work [a]	0.1864 (0.3354)
R-free [a]	0.2167 (0.3521)
CC(work) [a]	0.962 (0.764)
CC(free) [a]	0.946 (0.482)
Number of non-hydrogen atoms	2793
Macromolecules	2629
Protein residues	320
RMS(bonds)	0.01
RMS(angles)	0.99
Ramachandran favored (%)	98
Ramachandran allowed (%)	2.4
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.4
Clashscore	3.73
Average B-factor	29.9
macromolecules	29.46
solvent	36.97

[a] Statistics for the highest-resolution shell are shown in parentheses.

Table 3.1. *Data collection and refinement statistics for SeMet labeled SCAT_p0564.*
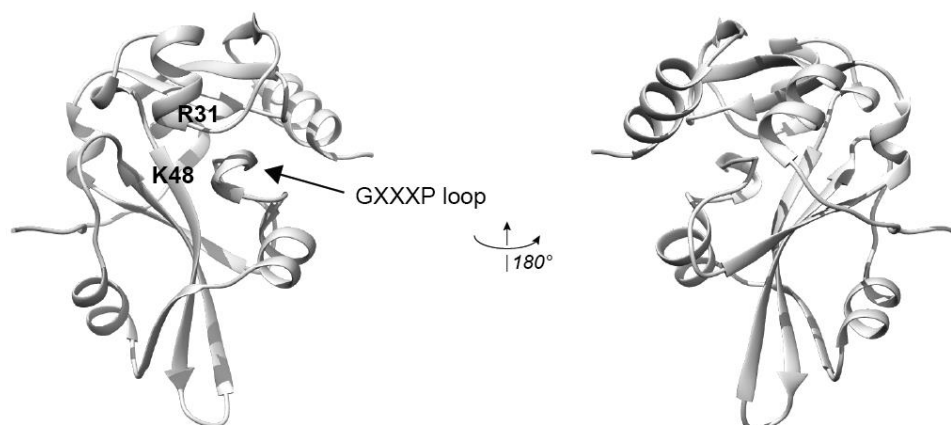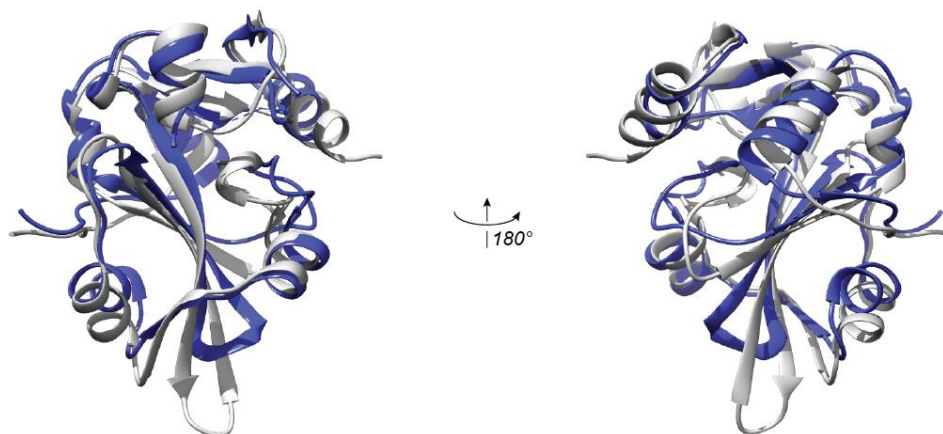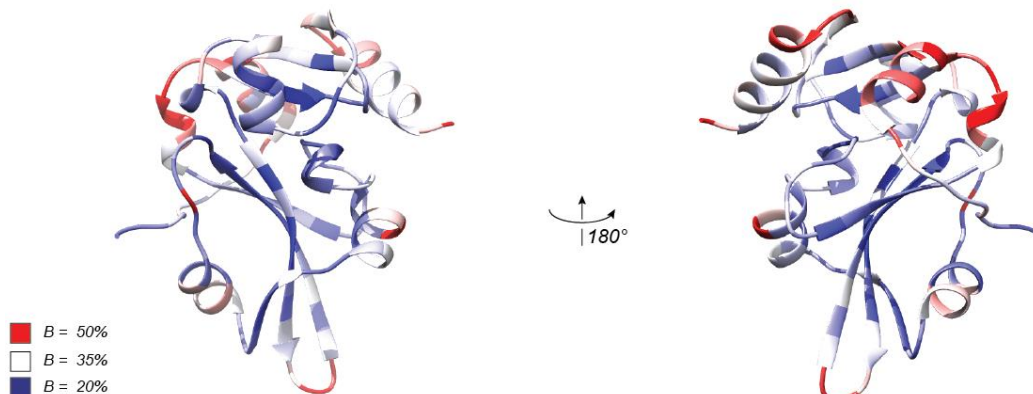
Figure 3.2. *(A) X-ray crystal structure of SCAT_p0564. The invariant lysine position K48 and the putative oxyanion hole comprised of the GXXP loop define the proposed substrate binding pocket of SCAT_p0564. The loop containing position R31 is also adjacent to the active site and includes the conserved position T32. (B) Structural alignment with YbaK from H. influenza reveals a similar fold, albeit with a greatly elongated central β-sheet. (C) Local quality assessment of the model. Model quality is high in the region around the proposed active site, but several loops on the opposite face of the protein have elevated local B-factors and poorly defined density for surface-exposed sidechains.*

62

```
E. coli P64483.1                      MTEMAKG---SVTHQRLIALLSQEGADFRVVTHEAVGKCEAVSEIRGTA
S. cattleya WP_014151019.1            MET-------ESTYDKLIDFLDSRGAQYRLIDHPPEGRTEIVSPMRGND
S. xinghaiensis WP_019711458.1        MET------GLDLYEKLVELLDHHGAEYRLIDHEPEGRTEIVSPMRGND
S. sp. MA37 CDH39442.1                METTESTEITENMYEKLVELLDRHGAEYRLIDHEPEGRTEIVSPMRGND
N. brasiliensis WP_042263680.1        MTQ-------SDPYQRLIEFLDAGSAQYRLIDHEPEGRTDVVSPMRGND
A. sp._N902-109 WP_041833330.1        MPY-------EETYDRLIALLDTNGASYRLIDHPEEGRTEIVSGYRGHE
S. exfoliatus WP_024754821.1          MSQET--------YEKLVALLDERGAAYRVIEHAPEGATEAVSALRGHE
S. venezuelae WP_055643003.1          MSQET--------YEKLVALLDERGAAYRVIEHAPEGATEAVSALRGHE
S. lavendulae WP_030008811.1          ------------MYQRLIGLLDGNQARYRLIDHVAEGRTDLASVLRGHP
S. sp. PCS3-D2 WP_037920470.1         ------------MYRTLLDLLDLHQARYRLIDHPAEGRTDLASVLRGHP
S. olivochromogenes KUN39668.1        MPGNG-------TYERLISLLDSTATAYELIDHEPEGTTETVSALRGHP
S. canus_WP_020118879.1               MAGDD-------TYDRLISLLDASAVDYELIDHAPEGTTDAVCALRGHP
S. sp._769 WP_039629926.1             MSEHET---EYETYRKLIGLLDERGARYRIYEHPAEGATEAVSALRGNA
S. sp._FxanaC1 WP_018092593.1         MTDHG-------TYTKLIGLLDERGARYRIYEHAEEGATEAVSALRGNP

E. coli P64483.1                      LGQGAKALVCKVK-GNGVNQHVLAILAADQQADLSQLASHIGGLRASLA
S. cattleya WP_014151019.1            VSEAAKCIVVMVKIGKKVTKYVLAVVPGDAKVDLNGIKSLFGGTYVSFA
S. xinghaiensis WP_019711458.1        VADAAKCIVVMVKLGKKVTKYVLAVVPGDARVDLNGIKRLLGGTYVSFA
S. sp. MA37 CDH39442.1                VADAAKCIVVMIKLGKKVTKYVLAVVPGDARVDLNGIKQLLGGTYVSFA
N. brasiliensis WP_042263680.1        LHDAAKCMVVMVKIGKKTKRYVLAVVPGDAKVDLTAVRALFEGTYVSFA
A. sp._N902-109 WP_041833330.1        VASAAKCMIVMVKLGKKVTKYVLAVVPGDARVDLQAIKGLLGGTYISFA
S. exfoliatus WP_024754821.1          LAQAAKCIIAMVKIGKKEKRFALVVVPGDRRIDLAAVKALYGGTYVSFA
S. venezuelae WP_055643003.1          LAQAAKCIIAMVKIGKKEKRFALVVVPGDKRIDLAAVKALYGGTYVSFA
S. lavendulae WP_030008811.1          LEQAAKCIVVRVSITKRVGKYVLAVVPGDRQVDLEAVGALFGGGRTAFA
S. sp. PCS3-D2 WP_037920470.1         LEQAAKCIIVRVSITKRVGKYVLAVVPGDRQVDLAAVAALFGGGRTAFA
S. olivochromogenes KUN39668.1        VSEAAKCLVLMVKIDRRVTRHVLAVVPGDRRVDLDAIRTLFAARYVGFC
S. canus_WP_020118879.1               ASEAAKCIVLRVKVDRRTTRHVLAVVPGDRRVDLDAVRALFAARYVGFS
S. sp._769 WP_039629926.1             VEQAAKCIVVMVKLDKKTKRYVLAVVPGDRRVDLGAVKALLGGSYAGFA
S. sp._FxanaC1 WP_018092593.1         VEQAAKCLVVMVKLSKKTKRYVLAVVPGDRRVDLGAVKALLGGSYAGFA

E. coli P64483.1                      SPAEVDELTGCVFGAIPPFSFHPKLKLVADPLLFERFDEIAFNAGMLDK
S. cattleya WP_014151019.1            TPEIAERLAGSVAGTILPFSFHPDLELIVDPSLLE-KEEIYFNAARLDR
S. xinghaiensis WP_019711458.1        NAETAERLAGSVAGAILPFSFHPDLELIVDPSLLE-KEEIFFNAARLDR
S. sp. MA37 CDH39442.1                NTEAAERLAGSVAGTILPFTFHPDLELIVDPSLVE-KEEIFFNAARLDR
N. brasiliensis WP_042263680.1        TPEIAENLAGSVAGTILPFSFHPDLELIVDPGVLD-IPELYFNAARLDR
A. sp._N902-109 WP_041833330.1        NPEIAERLSGCVAGTILPFSFSDELELIADPGIKA-FPELFFNAARLDR
S. exfoliatus WP_024754821.1          SPEIAEELAGSESGTILPFSFDERLELLVDPDLLT-HEEFYFNAARLDR
S. venezuelae WP_055643003.1          SPEIAEELAGSESGTILPFSFDERLELLVDPDLLT-HEEFYFNAARLDR
S. lavendulae WP_030008811.1          TPEIAERLAGSVCGTVMPLSFHPDLHLVVDEGLIL-TEEIYFNAARLDR
S. sp. PCS3-D2 WP_037920470.1         TPEIAERLAGSVCGTVMPLSFHPDLHLVVDEGLAL-TEEIYFNAARLDR
S. olivochromogenes KUN39668.1        DAATAERLARAVPGTVLPFSFDPALELVADPDVVA-RPCLYFNAARLDR
S. canus_WP_020118879.1               DAETAERLARAIPGTVLPFSFDPELEVVADPEVVK-QPRLYFNAARLDR
S. sp._769 WP_039629926.1             NTEVAERLARSVSGTILPFSFDEELTLIVDPALLA-QPEIYFNAARLDR
S. sp._FxanaC1 WP_018092593.1         TPEVAERLAGSVSGSILPFSFDEELALVVDPALLE-QPEIFFNAARLDR

E. coli P64483.1                      SVILKTADYLRIAQPELVNFRRTA--------------------
S. cattleya WP_014151019.1            SMALRTKDYLTLTDPRTEPIAATG--------------------
S. xinghaiensis WP_019711458.1        SMALRVKDYAAITSPRLEPIAAPAV-------------------
S. sp. MA37 CDH39442.1                SMALRVKDYVAITSPRLERIAAPANGTPSGPGRRTPYRCPWCSV
N. brasiliensis WP_042263680.1        SMALRTGDYVRLAHPRTEKIAQYEAAMGEAMT------------
A. sp._N902-109 WP_041833330.1        SMALRTEDYLRIAEPRFEPIAASNTTAG----------------
S. exfoliatus WP_024754821.1          SIALSGADYRSIAEPRVERVSSTD--------------------
S. venezuelae WP_055643003.1          SIALAGADYRSIAEPRVERVS-TD--------------------
S. lavendulae WP_030008811.1          SVALSTPDYLAIAQPQLAAIATAGDRVLTHSAR-----------
S. sp. PCS3-D2 WP_037920470.1         SVALSTSDYLAIAKPVLAAIATAGDRVLTNSAR-----------
S. olivochromogenes KUN39668.1        SLAMSGADYARLAEPRVERIAGP-RSD-----------------
S. canus_WP_020118879.1               SLCMSGADNERLAKPRVEPVASPPLSDALPDALPDALPG-----
S. sp._769 WP_039629926.1             SLALATEDYRAIAEPRVEPIAG----------------------
S. sp._FxanaC1 WP_018092593.1         SLALATADYVAIAEPRVEPVAG----------------------
```

Figure 3.3. *Multiple sequence alignment of SCAT_p0564, actinobacterial homologues, and the E. coli protein YeaK. Positions conserved across all sequences in the alignment are marked in* cyan*; positions conserved across all actinobacterial homologues are marked in* magenta*; positions specific to homologues from organofluorine BGCs are marked in* green*.*

Preliminary characterization of SCAT_p0564 mutants. SCAT_p0564 was aligned with a variety of homologues from other actinomycetes, as well as the *E. coli* protein YeaK. The resulting alignment was scrutinized for positions that are universally conserved, conserved across all *Streptomyces* homologues, and conserved only amongst homologues from organofluorine biosynthesis clusters *(Figure 3.3)*. Upon examination of these positions in the SCAT_p0564 crystal structure, it was noted that many clustered near the universally conserved K48 position, or around the GXXXP motif that comprises the putative oxyanion hole. Positions from amongst the last two categories were selected for mutation, and constructs were cloned and purified *(Figure 3.4A)*. The identities of all purified mutants were confirmed by whole-protein mass spectrometry. Initial characterization of the resulting proteins revealed minimal effects on catalysis, with all displaying near-WT efficiency towards fluorothreonyl-tRNA. The A104S mutation may merit some further attention, as this displayed slightly higher activity towards threonyl-tRNA, but the observed effect was modest (*Figure 3.4B*).

Activity of SCAT_p0564 on chlorothreonine. Exploration of alternate substrates was considered as another possibility to gain insight into the function of SCAT_p0564. To this end, 4-chloro-threonine (ClThr) was synthesized and assayed as a substrate for SCAT_p0564. As expected the *S. cattleya* TRS accepted this amino acid as a substrate. Aminoacylation assays were performed as described in *Chapter 2*, and confirmed that ClThr could be charged onto threonyl-tRNA *(Figure 3.5A)*. As noted by Webb and Matthews [*24*], ClThr was observed to be unstable in solution, forming a visible precipitate after 30 min of incubation in buffer. This behavior could be mitigated by using TCEP instead of DTT as a reducing agent. Somewhat surprisingly, chlorothreonyl-tRNA was found to be just as good of a substrate for SCAT_p0564 as fluorothreonyl-tRNA *(Figure 3.5 B)*. This result is consistent with sidechain pK_a playing an important role in catalysis, and is also consistent with the importance of side chain chemistry that has been observed for the previously characterized proteins YbaK and YeaK.

Characterization of a γ-chlorothreonyl-tRNA hydrolase. The presence of a conserved AlaX homologue (NCBI accession CCF23458; Cl-SAD) in the chlorothreonine biosynthesis cluster of *Streptomyces* sp. OH-5093 strongly suggested that aminoacyl-tRNA editing proteins may be a common feature of biosynthetic gene clusters encoding the production of amino acid antimetabolites [*1*, *30*]. Members of the AlaX family are structurally unrelated to SCAT_p0564, and feature a conserved zinc binding site. While the Zn ion appears dispensible for catalysis in the N-terminal editing domain of TRS, it may function in mediating substrate selectivity [*31*]. The Cl-SAD coding sequence was gene synthesized, cloned, and expressed as an MBP-fusion in *E. coli*. The cleaved editing protein was not chromatographically separated from the MBP-fusion, and as a result preliminary characterization was performed on the resulting mixture of MBP-fusion and free editing protein *(Figure 3.6A)*. The Cl-SAD was assayed on Thr, FThr, and ClThr, and was found to display progressively higher activity towards threonyl-, fluorothreonyl-, and chlorothreonyl-tRNA *(Figure 3.6B)*. Because the activity of the protein towards threonyl-tRNA is close to the background rate of AMP release observed with the coupled assay, these measurements may underestimate the ability of the Cl-SAD to select against threonyl-tRNA. Notably, the Cl-SAD displayed the highest degree of selectivity towards ClThr, and intermediate selectivity towards FThr, in contrast to SCAT_p0564.
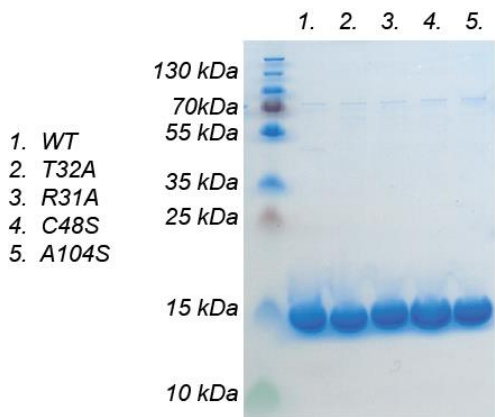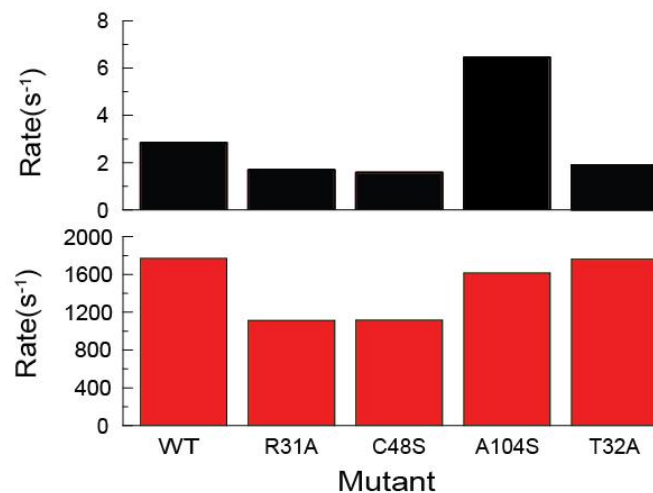
Figure 3.4. *Preliminary screening of active site mutants of SCAT_p0564. Mutants were purified (A) and screened against threonyl- and fluorothronyl-tRNA (B).*
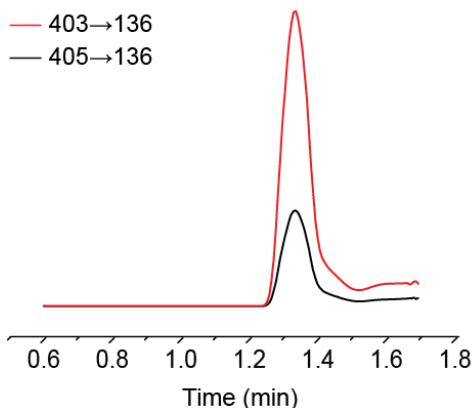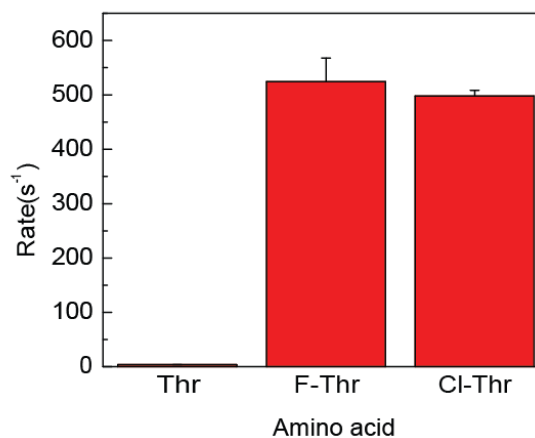
Figure 3.5. *Activity of SCAT_p0564 on threonyl-, fluorothreonyl-, and chlorothreonyl-tRNA. LCMS analysis of RNAs A digested aminoacylation reactions run in the presence of chlorothreonine and TRS, with MRM traces corresponding to the ^{35}Cl and ^{37}Cl-containing isotopomers of chlorothreonyl-adenosine (A). Continuous hydrolysis assays run in the presence of different threonine analogues in the presence of 5 μM tRNA.*
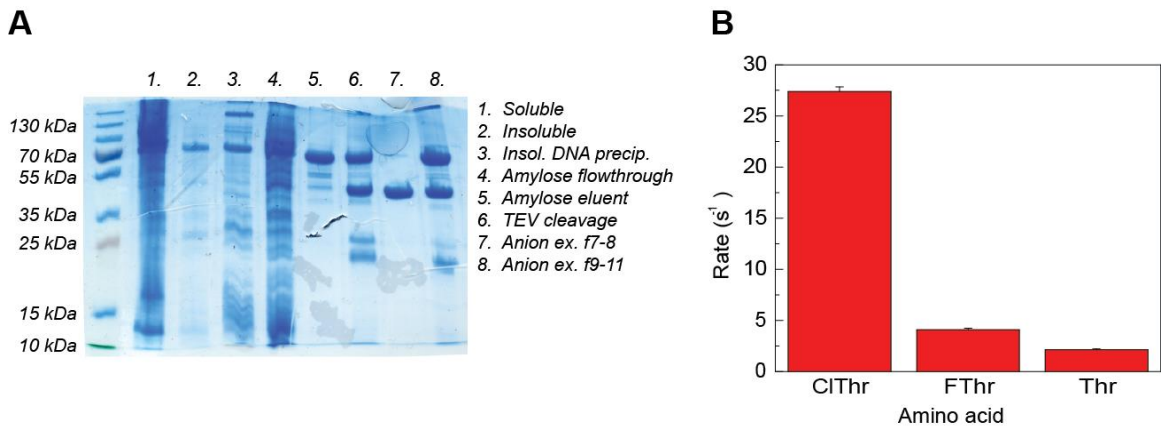
Figure 3.6. *Initial characterization of a chlorothreonyl-tRNA specific second additional domain protein (Cl-SAD) from the chlorothreonine producer Streptomyces sp. OH-5093. Preparation and partial purification of the Cl-SAD (A). Activity of the Cl-SAD towards chlorothreonyl-, fluorothreonyl-, and threonyl-tRNA.*
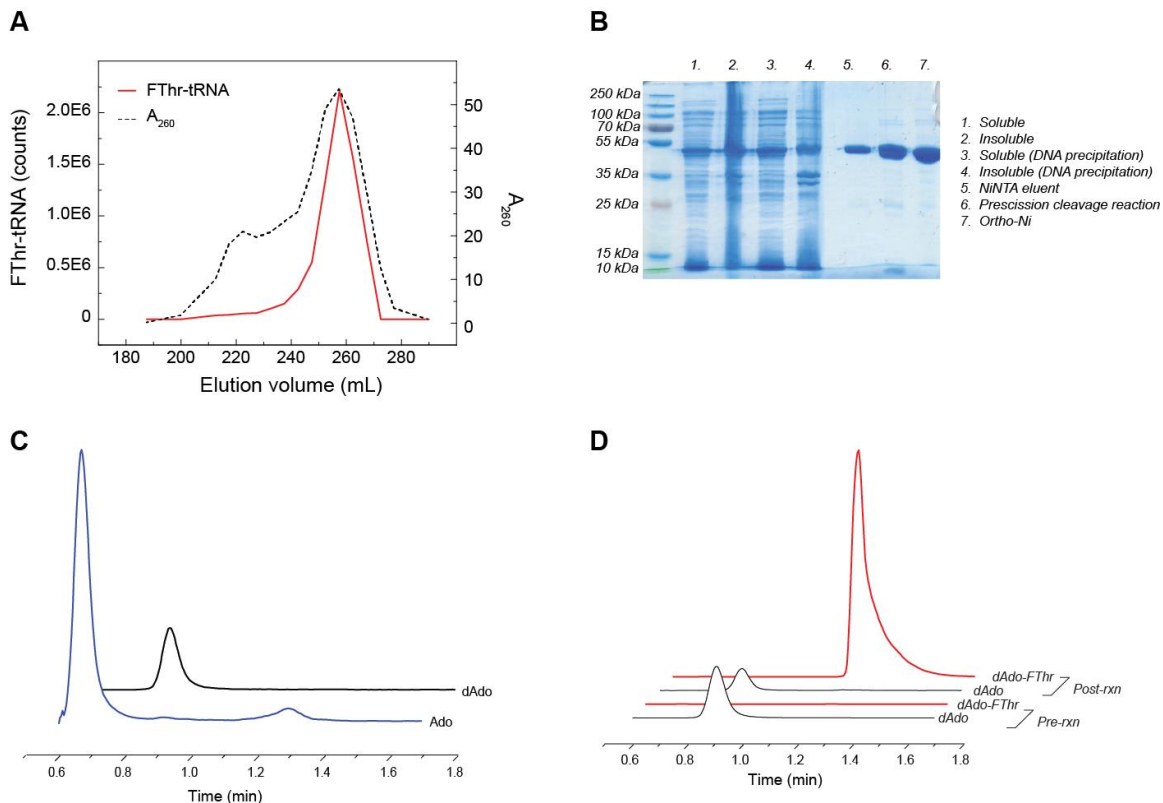
Figure 3.7. *Strategies for the production of nonhydrolyzable aminoacyl-tRNA analogues. In vitro production of tRNA combined with gradient anion exchange purification facilitates the preparation of large quantities of active tRNA (A). Purification of the CCA-adding enzyme from E. coli (B). 3'-end exchange facilitated by the CCA-adding enzyme (C). Aminoacylation of deoxyadenosine-terminal tRNA (D).*

Attempts to synthesize a nonhydrolyzable substrate analogue. Given the limited knowledge about the catalytic mechanism of SCAT_p0564, the synthesis of a non-hydrolyzable substrate analogue was a key goal for further biochemical and structural studies. Previous reports have found that both YbaK and the INS domain of ProRS rely on the free 2′ hydroxyl of the tRNA to catalyze hydrolysis; therefore the production of an aminoacyl-tRNA with a 2′-terminal deoxyadenosine was targeted. In the unlikely event that 2′-deoxy fluorothreonyl-tRNA remains a substrate for SCAT_p0564, or if the efficiency of TRS on 2′-deoxy is inadequate [*32*], the methods described would be equally applicable to the synthesis of 3′-aminoadenosine-terminal tRNA [*33*]. Initial attempts towards the production of such a compound focused on a biosynthetic approach. In order to enhance the quantity of tRNA that could be obtained, *in vivo* T7 transcription was used to drive the large-scale overexpression of *S. cattleya* tRNA25 [*34*]. The resulting material was inhomogeneous, but consistent with prior reports purity could be improved by the use of a sodium chloride gradient separation on a DEAE column *(Figure 3.7B)*. It was determined that the *E. coli* CCA-adding enzyme was competent to exchange the 3′-terminal adenosine residue of *S. cattleya* tRNA25 for deoxyadenosine, and that the resulting tRNA could be aminoacylated with FThr, albeit not to completion *(Figure 3.7C)*. Both of these steps proceeded with limited efficiency, precluding the use of this strategy for the production of a homogenous substrate analogue. Optimization of buffer conditions, reaction time, and enzyme loading would be necessary to synthesize a usable substrate, as options for the purification of aminoacyl tRNA from free tRNA are limited. The use of a microhelix tRNA substrate was also explored in light of the superior options for synthesis and purification, but aminoacylation efficiency was very low.

3.4 Conclusions

Here, we report the crystal structure of the fluorothreonyl-tRNA hydrolase SCAT_p0564. The overall fold is quite similar to previously characterized members of the YbaK-like superfamily, but with a pronounced extension of the central β-sheet. Mutagenesis of several conserved positions in the proposed substrate binding pocket failed to identify any positions that confer specificity for fluorothreonyl-tRNA. As detailed experimental information on the substrate binding interface remains lacking. Future work in this area should exploit known chemical [*35*] or biochemical [*32*, *33*] strategies to synthesize a nonhydrolyzable substrate suitable for further structural and biochemical investigation.

More fruitful insights were gleaned by examining the activity of SCAT_p0564 on an alternate substrate, chlorothreonine. Surprisingly, its activity is quite similar on the two halogenated substrates, both of which are expected to have a similar pK_a, slightly below that of threonine. This is consistent with the idea that the substrate sidechain may play an important role in catalysis. In contrast, the Cl-SAD, an AlaX-like protein from the chlorothreonine biosynthesis cluster displays the highest activity towards chlorothreonyl-tRNA. Interestingly, while these two enzymes both operate on halongenated Thr-tRNA substrates, they belong to different protein families. Probing SCAT_p0564 and the Cl-SAD with a wider array of substrates could provide even more insight into how these enzymes work, but would require tackling some tricky synthetic and bioanalytical challenges. The approaches employed herein all rely on efficient aminoacylation of the noncanonical amino acid in question; any bulkier substrates are unlikely to be charged efficiently by TRS, requiring the use of flexizymes [*36*] or chemical aminoacylation [*37*] in conjunction with a discontinuous hydrolysis assay [*38*].

The discovery of post-transfer editing activities against noncanonical amino acids may also have applications for the development of engineered systems. Recently, a series of publications have leveraged biochemical insights from the rich literature in the aminoacyl-tRNA field to develop enhanced strategies for global and site-specific incorporation of non-canonical amino acids [*39*, *40*]. In particular, the use of novel editing activities may be particularly beneficial in cases where minimal differences between the desired substrate and naturally occurring analogues make achieving a high degree of selectivity at the aminoacylation step difficult. By expanding the substrate scope of known post-transfer editing proteins to include halogenated threonine analogues, we open the door to creative strategies for the selective incorporation of these noncanonical amino acids into protein.

3.5   References

1.   Ahel I, Korencic D, Ibba M, Söll D (2003) Trans-editing of mischarged tRNAs. *Proc Natl Acad Sci USA* 100(26):15422–15427.

2.   Calendar R, Berg P (1967) D-Tyrosyl RNA: Formation, hydrolysis and utilization for protein synthesis. *J Mol Biol* 26(1):39–54.

3.   Wydau S, van der Rest G, Aubard C, Plateau P, Blanquet S (2009) Widespread distribution of cell defense against D-aminoacyl-tRNAs. *J Biol Chem* 284(21):14096–14104.

4.   Ferri-Fioni M-L, et al. (2006) Identification in archaea of a novel D-Tyr-tRNA[Tyr] deacylase. *J Biol Chem* 281(37):27575–27585.

5.   Starzyk RM, Webster TA, Schimmel P (1987) Evidence for dispensable sequences inserted into a nucleotide fold. *Science* 237(4822):1614–1618.

6.   Weinger JS, Parnell KM, Dorner S, Green R, Strobel SA (2004) Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nat Struct Mol Biol* 11(11):1101–1106.

7.   Kuhlenkoetter S, Wintermeyer W, Rodnina MV (2011) Different substrate-dependent transition states in the active site of the ribosome. *Nature* 476(7360):351–354.

8.   Świderek K, Marti S, Tuñón I, Moliner V, Bertran J (2015) Peptide bond formation mechanism catalyzed by ribosome. *J Am Chem Soc* 137(37):12024–12034.

9.   So BR, et al. (2011) Substrate-mediated fidelity mechanism ensures accurate decoding of proline codons. *J Biol Chem* 286(36):31810–31820.

10.   Dock-Bregeon A-C, et al. (2004) Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution. *Mol Cell* 16(3):375–386.

11.   Ahmad S, et al. (2015) Specificity and catalysis hardwired at the RNA-protein interface in a translational proofreading enzyme. *Nat Commun* 6:7552.

12. Kumar S, Das M, Hadad CM, Musier-Forsyth K (2012) Substrate and enzyme functional groups contribute to translational quality control by bacterial prolyl-tRNA synthetase. *J Phys Chem B* 116(23):6991–6999.

13. Ahmad S, et al. (2013) Mechanism of chiral proofreading during translation of the genetic code. *Elife* 2:e01519.

14. Kumar S, Das M, Hadad CM, Musier-Forsyth K (2012) Substrate specificity of bacterial prolyl-tRNA synthetase editing domain is controlled by a tunable hydrophobic pocket. *J Biol Chem* 287(5):3175–3184.

15. Waas WF, Schimmel P (2007) Evidence that tRNA synthetase-directed proton transfer stops mistranslation. *Biochemistry* 46(43):12062–12070.

16. Bujacz G, Wrzesniewska B, Bujacz A (2010) Cryoprotection properties of salts of organic acids: a case study for a tetragonal crystal of HEW lysozyme. *Acta Crystallogr D Biol Crystallogr* 66(7):789–796.

17. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.

18. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213.

19. Terwilliger TC, et al. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):61.

20. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352–367.

21. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Arch Biochem Biophys* 66(Pt 4):486.

22. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336(6084):1030–1033.

23. Amin MR (1996) Biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. Doctoral (Durham University). Available at: http://etheses.dur.ac.uk/5428/ [Accessed October 10, 2016].

24. Webb HK, Matthews RG (1995) 4-Chlorothreonine is substrate, mechanistic probe, and mechanism-based inactivator of serine hydroxymethyltransferase. *J Biol Chem* 270(29):17204–17209.

25. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, ed Walker J (Humana Press), pp 571–607.

26. Wittwer AJ, Stadtman TC (1986) Biosynthesis of 5-methylaminomethyl-2-selenouridine, a naturally occurring nucleoside in Escherichia coli tRNA. *Archives of Biochemistry and Biophysics* 248(2):540–550.

27. Zhang H, et al. (2000) Crystal structure of YbaK protein from *Haemophilus influenzae* (HI1434) at 1.8 A resolution: functional implications. *Proteins* 40(1):86–97.

28. Bartholow TG, et al. (2014) Strictly conserved lysine of prolyl-tRNA synthetase editing domain facilitates binding and positioning of misacylated tRNA[Pro]. *Biochemistry* 53(6):1059–1068.

29. Kumar S, Das M, Hadad CM, Musier-Forsyth K (2013) Aminoacyl-tRNA Substrate and Enzyme Backbone Atoms Contribute to Translational Quality Control by YbaK. *J Phys Chem B* 117(16):4521–4527.

30. Fullone MR, et al. (2012) Insight into the structure–function relationship of the nonheme iron halogenases involved in the biosynthesis of 4-chlorothreonine – Thr3 from *Streptomyces* sp. OH-5093 and SyrB2 from *Pseudomonas syringae pv. syringae* B301DR. *FEBS J* 279(23):4269–4282.

31. Pasman Z, et al. (2011) Substrate specificity and catalysis by the editing active site of alanyl-tRNA synthetase from Escherichia coli. *Biochemistry* 50(9):1474–1482.

32. Minajigi A, Francklyn CS (2008) RNA-assisted catalysis in a protein enzyme: The 2'-hydroxyl of tRNA(Thr) A76 promotes aminoacylation by threonyl-tRNA synthetase. *Proc Natl Acad Sci USA* 105(46):17748–17753.

33. Fraser TH, Rich A (1973) Synthesis and aminoacylation of 3'-amino-3'-deoxy transfer RNA and its activity in ribosomal protein synthesis. *Proc Natl Acad Sci USA* 70(9):2671–2675.

34. Perona JJ, Swanson R, Steitz TA, Söll D (1988) Overproduction and purification of *Escherichia coli* tRNA[Gln2] and its use in crystallization of the glutaminyl-tRNA synthetase-tRNA[Gln] complex. *J Mol Biol* 202(1):121–126.

35. Lodder M, Wang B, Hecht SM (2005) The N-pentenoyl protecting group for aminoacyl-tRNAs. *Methods* 36(3):245–251.

36. Luo X, et al. (2016) Recombinant thiopeptides containing noncanonical amino acids. *Proc Natl Acad Sci U S A*:201602733.

37. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* 244(4901):182–188.

38. Wolfson AD, Uhlenbeck OC (2002) Modulation of tRNA[Ala] identity by inorganic pyrophosphatase. *Proc Natl Acad Sci U S A* 99(9):5965–5970.

39. Richardson CJ, First EA (2016) Altering the enantioselectivity of tyrosyl-tRNA synthetase by insertion of a stereospecific editing domain. *Biochemistry* 55(10):1541–1553.

40. Völler J-S, et al. (2016) Discovery and investigation of natural editing function against artificial amino acids in protein translation. *ACS Cent Sci*. doi:10.1021/acscentsci.6b00339.

Chapter 4: *Investigation of the regulatory architecture of the organofluorine cluster of S. cattleya*

4.1. Introduction

Previous work has demonstrated that the organofluorine biosynthesis cluster is regulated in response to fluoride [1], but the underpinning molecular mechanisms have yet to be elucidated. Biological sensing of fluoride is quite underexplored, with the first documented example being the recently discovered fluoride riboswitch [2]. The widespread distribution of this regulatory RNA indicates a pervasive influence of fluoride on bacteria in the environment. The production of organofluorine compounds by the fluorinase pathway provides an additional set of possible regulatory stimuli, and would be of particular interest given that the sensing of organofluorines in nature is unknown. Understanding the general principles by which fluoride and organofluorine compounds can be sensed would also enable the rational discovery of new fluorine-associated metabolic pathways, which a central goal of our lab.

Secondary metabolism in *Streptomyces* spp. is tightly controlled by both cluster-situated and global regulators [3, 4]. Typically, interplay between global regulators controls the expression of keys cluster-situated regulators, which in turn activate the expression of key cluster proteins [5]. Pleiotropic global regulators can repress or activate BGC expression in response to a diverse array of environmental clues, including nutrient levels (phosphate, nitrogen, *N*-acetylglucose, ppGpp) [6, 7], cell density [8], and even the presence of foreign antibiotics [9]. The situation is further complicated by global roles of cluster-situated regulators, which can in certain cases act to silence competing metabolic pathways, ensuring that metabolic resources are focused on the production of a limited set of compounds [10]. The interplay between different factors controlling global regulation of secondary metabolism is a major challenge in the field of genome mining, as gene clusters are often transcriptionally inactive. As a result, identifying the correct combination of upstream regulatory factors can be quite challenging in many cases.

Recently, there has been increasing recognition of the role that small molecules play in feedback loops in secondary metabolism [5]. These interactions can be mediated either by receptor/transcription factor pairs, or by more direct transcription factor/ligand interactions. The planosporicin cluster is an example of the first paradigm, relying on product-induced liberation of the key pathway regulator PspX as part of a feed-forward loop [11]. Transcription factors can even sense metabolites directly; in particular, so-called "Atypical Response Regulators" which lack cognate histidine kinases, have been shown to act in this way [12–14]. Such proteins play a central role in the regulation of the jadomycin BGC, where the DNA-binding affinity of the master regulator JadR1 is modulated in response to jadomycin binding [12]. Lastly, riboswitches have been shown to play a role in the regulation of secondary metabolism, as in the case of *S. davawensis*, which uses a highly selective riboswitch to avoid self-toxicity during the production of roseoflavin [15].

4.2 Materials and methods

Commercial materials. Amylose resin, Hiscribe T7 transcription kit, Phusion DNA polymerase, Q5 DNA polymerase, deoxynucleotide triphosphates (dNTPs), and restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Asparagine, apramycin (Am), β-mercaptoethanol, boric acid, Dowex 50WX8, iron sulfate heptahydrate, naladixic acid, nickel sulfate heptahydrate, *o*-phthaldialdehyde, poly(deoxyinosinic-deoxycytidylic) acid sodium salt, and protamine sulfate from salmon sperm were purchased from Sigma-Aldrich (St. Louis, MO). Zinc sulfate heptahydrate was purchased from Mallinckrodt (Paris, KY). Amicon Ultra 3,000

MWCO and 30,000 MWCO centrifugal concentrators, Milli-Q Gradient water purification system, chloramphenicol, malt extract and yeast extract were purchased from EMD Millipore (Billerica, MA). Adenosine triphosphate disodium salt hydrate, ammonium chloride, carbenicillin, 3 kDa MWCO dialysis tubing, dithiothreitol, ethylenediaminetetraacetic acid disodium salt, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), hydrochloric acid, glucose, kanamycin sulfate, magnesium chloride, magnesium sulfate, phenylmethylsulfonyl chloride (PMSF), potassium phosphate dibasic, sodium acetate, sodium hydroxide, and tris base were purchased from Fisher Scientific (Waltham, MA). Deuterium oxide was purchased from Cambridge Isotopes (Tewksbury, MA). Fluka brand LCMS grade ammonium formate, ammonium bicarbonate, formic acid, methanol and acetonitrile were purchased from Sigma Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine (TCEP) was purchased from Biosynth Corporation (Itasca, IL). Ni-NTA resin, RNEasy mini kit, Qia-quick PCR cleanup kit, and Qia-spin miniprep kit were purchased from Qiagen USA (Valencia, CA). Bacto^{TM} Agar was purchased from BD (Sparks, Maryland). Hygromycin B was purchased from Santa Cruz Biotechnology (Dallas, TX). Soy flour was purchased from Berkeley Bowl (Berkeley, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a concentration of 100 µM, and stored at 4°C. Ammonium persulfate, acrylamide/bisacrylamide (37.5:1 and 19:1), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), and sodium dodecyl sulfate (SDS) were purchased from Bio-Rad Laboratories (Hercules, CA). Chloroform, Lysogeny Broth (LB), Lysogeny Broth Agar (LBA), and Terrific Broth (TB) were purchased from VWR International (Radnor, PA). DNAse I (RNAse-free), Pierce ECL Western blotting substrate, proteinase K and Trizol RNA extraction reagent were obtained from Thermo Fisher (Waltham, MA). Ultrayield baffled flasks were purchased from Thompson Instrument Company (Oceanside, CA)

Multiple sequence alignments. Protein sequence alignments were performed using Clustal Omega, and domain boundaries were annotated using the PFAM web server [*16*, *17*].

Preparation of *Streptomyces* spp. gDNA. *Streptomyces* gDNA was prepared by a scaled-down version of the salting out procedure [*18*]. Single colonies were picked into 3 mL GYM media and grown at 30°C for 2-5 d. Between 500 µL and 2 mL of culture was then pelleted by centrifugation, and the supernatant was decanted. The pellet was resuspended in 600 µL of 50 mg/mL lysozyme in SET buffer, then incubated for 1 h at 37 °C. At that point 70 µL of 10% (*w/v*) SDS and 16 µL of 20 mg/mL proteinase K were added, and the solution was incubated at 55°C. When the pellet was dispersed (1-2 h), 240 µL 5 M sodium chloride and 600 µL chloroform were added. The solution was mixed by inversion, and phase separated by centrifugation at 10,000 RCF for 10 min. The aqueous layer was carefully removed and precipitated by the addition of one volume isopropanol and centrifugation at 20,000 RCF. The pellet was washed with ice-cold 70% (*v/v*) ethanol, air dried briefly, and resuspended in buffer EB (Qiagen).

***S. cattleya* deletion strain construction.** The *S. cattleya* Δ*flA*::Am^R deletion strain was previously constructed [*19*]. The *flJ*::Am^R , *flL*::Am^R, and *ThrAld*::Am^R strains were obtained using previously cloned constructs [*19*]. Conjugations were performed as described in *Chapter 2*. Colonies were screened for double crossover homologous recombination by PCR with primers J123/197 and J125/198 (for the transaldolase), primers J123/195 and J125/198 (for *flL*), and J213_F/R (for *flJ*). However, the plasmid-based system was inadequate to obtain a double cross for *flG*, and a plasmid for the deletion of *flF* proved challenging to construct by amplification standard cloning methods.

For the construction the *ΔflF*::Am^R and *ΔflG*::Am^R strains, a cosmid-based strategy was employed. A previously-constructed cosmid library [*1*] was screened using the primers J191 and J192, which flank the *flF* coding sequence, adjacent to the *flA* coding sequence. One cosmid, H20, was identified that contained the *flF* coding sequence; Sanger sequencing from the T3 and T7 primer sites and comparison with the reference genome revealed that the cosmid encompassed the entire organofluorine biosynthesis locus. The REDIRECT system [*20*] was used to obtain constructs for the disruption of *flF* and *flG*, as described in *Chapter 2*. Briefly, primers J237/J238 (for *flF*) and J271/J272 (for *flG*) were used to amplify the OriT/Am^R cassette from pIJ773. The resulting PCR products was electroporated into B25113 cells bearing the lambda-RED helper plasmid pIJ790 and the organofluorine cluster cosmid H20. Colonies with the Am^R phenotype were then sequenced using primers J235 and J236 to confirm the correct insertion site. The initial construct used for the construction of flF proved prone to aberrant recombination *in vivo*. The coding sequences for *flG*, *flJ*, and *flH* could not be amplified from genomic DNA from double crossover strains created using this construct, and furthermore preliminary RNAseq showed that the entire region encompassed by the cosmid and 5′ to the *flF* coding sequence was transcriptionally silent, suggesting a large-scale deletion. Consequently, a plasmid with shorter homology arms was prepared. Fragments containing the upstream and downstream homology arms and the OriT/Am^R cassette were amplified from the previously constructed *flF* disruption cosmid using primers J442/3, J444/5, and J446/7. The Cb^R marker and *E. coli* origin of replication were amplified from pIJ773 using J448/9, and the plasmid was assembled by 4-piece Gibson ligation. Conjugations were performed as described in *Chapter 2*. Colonies were then screened for successful double crossover homologous recombination by screening for Km^S phenotype, and by using primers J191/2 (for *flF*) and J200/1 for *flG*.

Similarly, the primers J185 and J186 were used to screen for a cosmid encoding the biosynthesis cluster associated with the putative lantipeptide SCAT_0531, which was previously proposed to be involved with organofluorine metabolism [*1*]. Primers J423/4 were then employed to construct the *ΔSCAT_0531*::Am^R strain using the REDIRECT method; after conjugation, primers J437F/J437R were used to screen for double crossovers. Primer sequences used for screening, construction of plasmids, and confirmation of knockouts are listed in *Appendix 1.3*.

S. cattleya complementation strain construction. Complementation strains were constructed using ΦC31 and VWB phage-based integrative vectors. To facilitate the integration of multiple resistance cassettes in the same strain, Hyg^R variants of the vectors pSET152 and pSET152-ermEp* were constructed. The Hyg^R cassette was amplified from pIJ10701 with primers J256/J257 and ligated into SacI digested pSET152 or pSET152-ermEp*. Constructs with C-terminally FLAG$_3$-tagged signaling proteins driven by the ErmEp* promoter and followed by the Fd phage terminator were all constructed by 3-piece Gibson ligation [*21*]. The gBlock J259, encoding a C-terminal FLAG$_3$ tag with codon usage suitable for S*treptomyces*, was used as a template. Primers J268/J286, J268/J287, J268/342 and J268/J339 were used to amplify this region for *flF*, *flG*, *flL*, and *flJ* constructs respectively. Primers J276/J279, J278/J280, J340/J341 and J337/J338 were used to amplify the corresponding coding sequences for *flF*, *flG*, *flL*, and *flJ* from the genomic DNA of *S. cattleya*. Fragments encoding the FLAG tag and the signaling protein were ligated with NdeI/BamHI digested pSet152-ermEp*-Hyg. The 3-piece Gibson approach was preferred over the construction of a vector containing the FLAG$_3$ tag, as it allows for the overlap regions between fragments to be ligated to contain a non-repetitive sequence. In contrast, putting the FLAG$_3$ tag on the vector results in a highly repetitive overlap region and leads to undesired cloning outcomes. Conjugations were performed as described in *Chapter 2*, with the exception

that hygromycin was applied as an additional selective marker to limit *E. coli* contamination when performing conjugations into Hyg[R] strains. Strains are listed in *Appendix 1.3*.

***Streptomyces lividans* reporter strain construction.** Reporter strains were constructed by integrating the mCherry fluorescent protein into the genome under the control of the promoter regions of the *flA*, *flH*, *flI*, and *ThrAld* genes using derivatives of pAV-1, which encodes the mCherry fluorescent protein with codon usage suitable for *Streptomyces* [*22*]. To construct reporter constructs, the primers J564/J565, J616/J617, J614/J615 and J560/J561 were used to amplify the promoter regions for *flA*, *flH*, *flI*, and *ThrAld*, respectively. These fragments were then inserted into ScaI/NdeI digested pAV-1 or pAC-1 using Gibson ligation. Additionally, a reporter construct intended to test the role of the *flL* 3′-UTR was constructed by amplification from *S. cattleya* gDNA using primers J610/J611, amplification from plasmid pAV-1 using primers J612/J613, and Gibson ligation with NdeI/BamHI digest pSET152-ermEp*. Reporter strains were constructed by sequential conjugations. First, *S. lividans* TK24 was conjugated with Hyg[R], ΦC31-based vectors expressing signaling proteins from the ermEp* promoter. The resulting strains were conjugated with Am[R], VWB-based vectors expressing mCherry from various *fl* locus promoters. Conjugations were performed as described in *Chapter 2*, with the exception that hygromycin was applied as an additional selective marker to limit *E. coli* contamination when performing conjugations into Hyg[R] strains. Strains are listed in *Appendix 1.3*.

***S. cattleya* growths.** Spores of *S. cattleya* strains were heat shocked for 5 min at 55°C, and used to inoculate 50 mL GYM media, presterile pH 5. The resulting seed cultures were grown overnight at 30°C. It is very important that seed cultures do not become overgrown, or else growth of the production cultures may lag. Seed cultures should be grown for less than 16 h, and a 2-step seed culture can be used for best reproducibility. Overnight cultures were then used to inoculate 30 mL production cultures to an OD_{600} of 0.05.   Production cultures were supplemented with 2 mM sodium fluoride, and grown at 30°C with glass beads to homogenize mycelial clumps. Cultures were sampled for biomass at 3 d, and culture supernatant was harvested at 6 d by centrifugation for 60 min at 3,000 RCF. Supernatant was concentrated using a SpeedVac and stored at -20°C before analysis.

^{19}F-NMR analysis of organofluorine production. Concentrated culture supernatant was resuspended in 50 mM tris pH 7.5, 1 mM fluorouracil, 20% D_2O. at a ratio of 750 μL buffer per 12 mL culture supernatant. NMR spectra were collected on a Bruker AVQ-400 spectrometer at the College of Chemistry NMR Facility at the University of California, Berkeley (128 scans; o1p = -200ppm; d1 = 1 s). Correction factors to account for incomplete relaxation were calculated by measuring the same sample with d1 = 1s and d1 = 10s, and comparing change in the ratio of fluoride peak area to organofluorine peak areas. Let us denote C_{FAc} as the correction factor to apply to the fluoroacetate peak area, I_{FAc} as the uncorrected fluoroacetate peak area, and I_{F-} as the uncorrected fluoride peak area. Using this approach, the correction factor for the fluoroacetate peak area would be calculated as:

$$C_{FAc} = \left[\frac{I_{FAc}}{I_{F-}}\right]_{d1=10s} / \left[\frac{I_{FAc}}{I_{F-}}\right]_{d1=1s}$$

Concentrations were then estimated by normalizing the corrected integrated area of fluorothreonine, fluoroacetate, and fluoride to 2 mM.   For example:

$$[FAc] = \frac{I_{FAc} * C_{FAc}}{I_{FAc} * C_{FAc} + I_{FThr} * C_{FThr} + I_{F-}}$$

Expression and purification of FlA. *E. coli* BL21(DE3) was transformed with the plasmid pET16b-FlA [*23*], and colonies were picked for overnight cultures in LB supplemented with carbenicillin (50 μg/mL). Seed cultures were diluted to an OD_{600} of 0.05 in $6 \times 1L$ LB in 2.8L Fernbach flasks, and grown at 37°C. Once the OD_{600} reached 0.6, cultures were placed in ice/water baths for 30 min before induction with 0.2 mM IPTG and growth overnight at 16°C. Cells were harvested by centrifugation and resuspended in 200 mL of lysis buffer (10% (*v/v*) glycerol, 20 mM Tris pH 7.9). Lysis was accomplished by sonication with a Misonix 3000 operating at max power for pulses of 5 s, with 10 s between pulses, for a total process time of 1.5 min. Lysate was clarified by a hard spin for 20 min at 18,000 RCF, then batch bound for 1.5 h with 2 mL NiNTA resin. The resin was washed with 25 column volumes lysis buffer followed by 50 column volumes of lysis buffer supplemented with 20 mM imidazole. Protein was eluted with lysis buffer supplemented with 250 mM imidazole and buffer exchanged into lysis buffer using a G25 column. Concentration with an Amicon stirred pressure cell using a 10 kDA MWCO filter afforded 1.5 mL of protein with a concentration of 1 mg/mL, which was flash frozen in liquid nitrogen. Purified FlA was submitted to Prosci (San Diego, CA) for the generation of polyclonal rabbit antibodies.

Western blotting. Cultures of *S. cattleya* grown for 5 d in GYM media supplemented with 2 mM fluoride were harvested by centrifugation for 10 min at 9,000 RCF, and pellets were frozen at -80 °C for storage. Frozen cells were resuspended in 2 mL of 50 mM Tris, pH 7.5, 500 mM sodium chloride, 500 μM PMSF, 10 mM EDTA, 20% (*v/v*) glycerol. Lysis was accomplished by sonication with a Misonix 3000 set to power 2.5 in pulses of 5 s with 25 s rest, for a total process time of 1 min. Lysate was clarified by centrifugation for 10 min at 20,000 RCF and supernatant was used for Western blotting. Gels were transferred to a PVDF membrane by wet transfer for 2 h at 40V in Towbin buffer (24 mM Tris base, 19 mM glycine, 20% (*v/v*) methanol). The membrane was blocked overnight in TBST (20 mM Tris pH 7.4, 150 mM sodium chloride, 0.1% Tween 20) supplemented with 5% (*w/v*) milk, then incubated with a 1:10,000 dilution of rabbit anti-FlA antiserum in TBST. The membrane was then washed 3×5 min in TBST, then incubated for 1 h with HRP conjugated goat anti-rabbit antibody. After again washing 3×5 min with TBST, the membrane was exposed to Pierce ECL Western blotting substrate and imaged on a BioRad Geldoc.

Analysis of genomic context conservation. Conserved component analysis was performed using a series of Python scripts. The query sequence was subjected to a BLASTp search against a custom database composed of protein coding sequences from all microbial assemblies available from the NCBI, or alternatively against all *Streptomyces* assemblies. The sequences of the resulting hits were then extracted, along with the five adjacent upstream and downstream coding sequences. The resulting sequence database was subjected to all-vs.-all BLAST, and a network representation was constructed with nodes representing individual gene clusters and edge weights representing cumulative BLAST bit scores from all hits between proteins belonging to the clusters in question. Tabular output is also produced describing the domain architecture of each protein in a cluster (as assigned by RPS-BLAST), as well as the number of homologues found in other clusters. These routines employ the Biopython toolkit for interfacing with NCBI BLAST utilities and for parsing sequence files and search results. The NetworkX toolkit was used to write graph representations of the resulting data suitable for visualization with Cytoscape [*24*]. The pipeline can be obtained from: https://github.com/mcmurryj/synteny-BLAST.

Cloning of plasmids for the expression of signaling proteins: MBP-fusion constructs for the expression of FlF, FlG, FlJ, and FlL were prepared by PCR from *S. cattleya* genomic DNA

with the primers J248/J250, J251/J252, J253/J254, and J307/J308, respectively. The resulting fragments were subjected to Gibson ligation with SfoI/HindIII digested pSV272.1.

Expression and purification of FlJ. For purification of FlJ, *E. coli* Bl21 (DE3) was transformed with the expression plasmid PSV272.1-FlJ and pRare2. Overnight cultures in Terrific Broth (TB) were used to inoculate expression cultures to $OD_{600} = 0.05$ in 1L TB supplemented with 17 mM sodium phosphate monobasic, 82 mM sodium phosphate dibasic in an UltraYield flask. Cultures were grown at 37°C until the OD_{600} reached 0.6-0.8 and were induced by the addition of IPTG to 0.2 mM final concentration. Growth was continued overnight at 18°C. Cells were harvested by centrifugation and stored at -80°C. Cells were resuspended in 5 mL lysis buffer per gram cell paste. Lysis buffer was composed of HEPES pH 7.5, 5% (*v/v*) glycerol, 500 mM sodium chloride, 5 mM magnesium chloride, 5 mM BME, and 500 μM PMSF. Lysis was accomplished by sonication using a Misonix Sonicator 3000 (power = 10, 5s on, 25s off, 2.5 min total process time, 1/4" tip), after which DNAse I was added to 1 U/mL. Lysate was clarified by 2×30 min spins at 15,000 RCF, and loaded onto 25 mL amylose resin. The column was washed with 12 column volumes of high salt wash buffer (500 mM sodium chloride, 20 mM HEPES pH 7.5, 5 mM BME, 5% *v/v* glycerol), and then with 4 column volumes of low salt wash buffer (25 mM sodium chloride, 20 mM HEPES pH 7.5, 5 mM BME 5% *v/v* glycerol). The MBP-FlJ fusion was eluted in 50 mL of elution buffer (25 mM sodium chloride, 20 mM HEPES pH 7.5, 5 mM BME, 5% *v/v* glycerol, 10 mM maltose), affording ~800 A_{280}-mL of material which was then diluted two-fold with low salt wash buffer. TEV protease (8 mg) and TCEP (1 mM final) were added, and cleavage of the MBP tag was allowed to proceed overnight at 4°C. Some precipitate was observed upon returning; this was cleared by centrifuging for 20 min at 15,000 RCF. Imidazole was added to 20 mM, and His-MBP was partially removed by passage over a 15 mL plug of NiNTA resin. Cation exchange was used to further purify FlJ. The NiNTA flowthrough was loaded onto a Hitrap SP column and eluted with a gradient of 0-800 mM sodium chloride over 20 column volumes; FlJ was observed to elute at approximately 400 mM sodium chloride. The resulting protein was concentrated using a 10 kDa MWCO Amicon spin concentrator, diluted with wash buffer containing 100 mM sodium chloride, re-concentrated, and frozen in liquid nitrogen. Concentration was determined using the extinction coefficient calculated by ExPASY ProtParam ($\varepsilon_{280\ nm}$= 2,980 $M^{-1}cm^{-1}$) [25]. It is important to note that FlJ will not express solubly as a His-tagged protein, nor will the MBP-fusion express solubly if induced at 37°C. The limiting factor in purification is the TEV cleavage step; the use of a more efficient protease could improve yield.

Expression of MBP-FIG and MBP-FIL. *E. coli* BL21(DE3) cells bearing the pRare2 plasmid were transformed with the plasmids PSV272.1-FlG and PSV272.1-FlL. Overnight growths were inoculated into 500 mL of ZY5052 media in 2.5 L Ultrayield flasks, and were grown for 10 h at 37°C. Cultures were harvested by centrifugation and stored at -80°C. Cells were thawed and resuspended in 50 mM tris pH 7.5, 500 mM sodium chloride, 10% (*v/v*) glycerol, 1 mM DTT, and 1 mM PMSF. Lysis was accomplished by sonication in an ice water bath with 5 s pulses followed by 15 s of rest for a total process time of 2 min. Lysate was clarified by centrifugation for 30 min at 18,000 RCF, and the supernatant was applied to 5 mL amylose resin. The column was washed with 20 column volumes of lysis buffer, and protein was eluted in lysis buffer supplemented with 10 mM amylose. Protein was concentrated with a 30 kDa amicon spin column, flash frozen in liquid nitrogen, and stored at -80°C. Concentration was determined using the extinction coefficient calculated by ExPASY ProtParam ($\varepsilon_{280\ nm}$= 73,800 $M^{-1}cm^{-1}$ for MBP-FlG and 76,320 $M^{-1}cm^{-1}$ for MBP-FlL).

Gel shift assays: DNA corresponding to the *flA* promoter region was prepared by PCR from *S. cattleya* gDNA using the primers J296/J298 and Q5 polymerase with high GC additive. MBP-fusions of FlG and FlL were thawed and diluted to 0.1 mg/mL in elution buffer. Master mix composed of 20 mM HEPES pH 7.9, 100 mM KCl, 10 mM magnesium chloride, 1.25 mM DTT, 0.0285 ug/µL poly-dIdC, 1.25 ug/uL BSA, 625 µM PMSF, and 20% (*v/v*) glycerol was prepared. Master mix (8 µL), protein stock (1 µL), and DNA (50 ng/1 µL) were incubated for 30 min at 30°C, and then run on a 6% TBE-PAGE gel [*26*]. Gels were stained with SYBR gold and imaged by UV illumination.

Expression and purification of FlF. For purification of FlF, *E. coli* Bl21 (DE3) was transformed with the expression plasmid PSV272.1-FlF and pRare2. Overnight cultures in Terrific Broth (TB) were used to inoculate expression cultures to $OD_{600} = 0.05$ in 1L TB supplemented with 17 mM sodium phosphate monobasic, 82 mM sodium phosphate dibasic in an UltraYield flask. Cultures were grown at 37°C until the OD_{600} reached 0.6-0.8 and were induced by the addition of IPTG to 0.2 mM final concentration. Growth was continued overnight at 18°C. Cells were harvested by centrifugation and resuspended in 50 mM HEPES pH 7.5, 500 mM sodium chloride, 5 mM BME, 10% (*v/v*) glycerol, 500 µM PMSF, 0.5 mg/mL lysozyme before freezing at -80°C. Cells were thawed and lysed via sonication with power = 10, 5 s on, 25 s off for 2.5 min total process time. DNAse I was then added to 1 U/mL, and lysate was clarified by two sequential rounds of centrifugation for 20 min at 15,000 RCF. The clarified lysate was applied to 20 mL of amylose resin, and washed with 20 mM HEPES pH 7.5, 500 mM sodium chloride, 10 mM BME, 10% (*v/v*) glycerol until the A_{280} of the flow-through dropped below 0.1 (approximately 10 column volumes). Protein was eluted in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 1 mM TCEP, 10 mM maltose. For metal reconstitution, endogenous metals were stripped by addition of EDTA to 2.5 mM and incubation for 20 min at room temperature. Metal loading was then accomplished by the addition of 2 mole equivalents with respect to EDTA of 100 mM $NiSO_4$, $ZnSO_4$, or $FeSO_4$. The $FeSO_4$ solution was supplemented with 200 mM ascorbic acid to slow oxidation. Reconstituted MBP-FlF was used in for equilibrium dialysis immediately after preparation. Concentration was determined using the extinction coefficient calculated by ExPASY ProtParam ($\varepsilon_{280\ nm} = 76,320\ M^{-1}cm^{-1}$).

Equilibrium dialysis measurements. Equilibrium dialysis cassettes (5 kDA cutoff, Harvard apparatus) were prepared with protein on one side and matched buffer derived from the flow-through buffer during the final concentration step for protein purification on the other; equal concentrations of ligand were added to both sides. Samples were gently rotated overnight at room temperature, and were then recovered by inversion and centrifugation. The concentration of ligand on the buffer-only side of the membrane was taken to bee the concentration of free ligand, while the difference in ligand concentration between the two sides was taken to represent the concentration of bound ligand on the protein-containing side. Nucleotide binding by FlJ was assessed by measurement of A_{260} using an Agilent 8453 spectrophotometer with Helma 105.252-QS microcuvette, using $\varepsilon = 15,400\ M^{-1}cm^{-1}$. Measurement of amino acid binding by FlF was accomplished by OPA derivatization based amino acid analysis as follows. Protein was precipitated from samples by addition of 100% (*w/v*) TCA to 5% (*w/v*), and precipitate was cleared by centrifugation for 10 min at 20,000 RCF. Amino acid analysis was performed on the supernatant as described in *Chapter 2*, with the exception that UV integration was also employed for single amino acid samples with concentrations in excess of 50 µM. Signal was measured at 338 nm with 390 nm as a reference wavelength. UV integration offers superior reproducibility but lower sensitivity as compared to LCMS/MS.

79

RNA sequencing. Samples were harvested by centrifugation of 2 mL culture for 1 min at 20,000 RCF. Supernatant was removed by decanting, and the pellet was dislodged from the bottom of the tube by tapping before flash freezing in liquid nitrogen and storage at -80°C. Samples were removed from the freezer and homogenized with mortar and pestle under liquid nitrogen. The resulting powder was maintained at liquid nitrogen temperature until resuspension in 667 μL Trizol, with vigorous vortexing. After 5 min incubation at room temperature, 133 μL chloroform was added, and phase separation was accomplished by centrifugation for 5 min at 20,000 RCF and 4°C. 250 μL of the aqueous phase was removed and mixed with 375 μL ethanol before application to a Qiagen RNEAsy column. Sample was loaded onto the column, followed by a 700 μL wash with buffer RW1 and 2 × 500 μL buffer RPE. Sample loading and washes were accomplished by successive rounds of centrifugation for 30 s at 8,000 RCF. The column was then dried by centrifugation for 2 min at 20,000 RCF, and eluted in 40 μL RNAse-free water. Equivalent amounts of RNA from each sample (10 μg) were prepared in 32μL volume, and genomic DNA was depleted by digestion at 37°C with DNAse I (Thermo scientific). The digestion was run for 30 min, at which time additional DNAse I was added, and it was allowed to proceed for another 30 min. The DNAse was inactivated by the addition of 2.2 μL EDTA (10% molar excess with respect to $MgCl_2$) and heating for 10 min at 65°C. RNA was then precipitated by the addition of 1/10 volumes 5 M ammonium acetate, 1/50 volumes glycogen, and 2.5 volumes ethanol before storage at -80°C. Pellets were washed 3 times with ice cold 70% (*v/v*) ethanol, and were resuspended in 12μL RNAse free water before submission to the Vincent J. Coates Genomics Sequencing Laboratory for library prep and sequencing. The resulting 50bp, single-end reads were aligned to genome sequence files FQ859185 and FQ859184 using BWA, and were assigned to features using HTseq-count [*27–29*]. Corresponding data for the WT strain was obtained from Mark Walker and Miao Wen [*1*].

Reporter assays. Spores of reporter strains were prepared, and 2.5×10^8 spores were inoculated into 50 mL of GYM media in 500 mL Ultrayield flasks with glass beads. Cultures were grown for 4 d at 30°C, and mycelial clumps were homogenized by sonication for 1 s at the minimum power setting. Measurements were made in 96-well plates on a BioTek Synergy Mx platereader; optical density was recorded at 600 nm and fluorescence was recorded with excitation at 587 nm and emission at 610 nm. The gain was set to 130 for fluorescence measurements. Uncultured GYM media was used as a blank.

Untargeted metabolomics. Samples comprised of 800 μL culture were harvested by centrifugation for one min at 20,000 RCF. Supernatant was removed by pipetting and acidified with formic acid to 0.1% final concentration. The pellet was resuspended in 60% (*v/v*) methanol and vortexed, and was freeze-thaw cycled three times by alternating between -20° and -80°C. For rapid quenching and analysis of combined supernatant/pellet, samples 800 μL samples were quenched directly into 800 μL acetonitrile. All samples were cleared by 10 min centrifugation at 20,000 RCF prior to LCMS analysis. Samples were analyzed on an LCMS system comprised of an Agilent 1290 HPLC with 6530 Q-TOF. The column was an Ascentis RP-amide, 2.1 mm × 100 mm, with mobile phases comprised of 0.1% formic acid in water (A) and acetonitrile (B). The gradient consisted of 0% B for 1 min, followed by a linear gradient to 50% B over the next 8 min, and then from 50 to 100% over 1.5 m. The portion of the gradient < 50% acetonitrile gives the most reliable data, as hydrophobic mobile phase contaminants tend to elute at the end of the gradient. The Q-TOF was operated in MS-scan mode, with fragmentor voltage =150 V, and 11 L/min drying gas at 300°C, 11 L/min. The resulting data files were converted to mzML format with MSconvert, and were analyzed using the XCMS package in the R statistical computing

environment [*30*, *31*]. Peaks were assigned using the "centWave" method with a 10 ppm window, allowed peak widths between 10 and 30 s, 10-fold signal-noise cutoff, and 1.5×10^3 count noise cutoff. Peaks of interest from 2-way comparisons of WT vs. mutant strains were identified by specifying fold-change values > 10 and p-values < 0.01. The analysis pipeline used is available online at: https://github.com/mcmurryj/metab-compare.

4.3   Results and discussion

Regulatory components of the organofluorine biosynthetic cluster. The organofluorine biosynthesis cluster of *S. cattleya* contains a high density of signaling proteins, especially given the relative simplicity of the biosynthetic pathway itself. It features three transcription factors, FlF, FlG, and FlL, in addition to the DUF190 domain protein FlJ, which has distant homology to pII-like signaling proteins (*Figure 4.1A*). The response regulator FlG is the most conserved component of the cluster, with close homologues in all predicted fluorinase clusters, as well as the chlorinase-containing salinosporamide cluster of *Salinispora tropica* (*Figure 4.1B)*. Homologues of FlL are found in *Streptomyces xinghaiensis* and *Streptomyces* sp. MA-37, but appear to be absent in variants of the cluster found in other actinomycetes. While both FlG and FlL contain response regulator domains, they lack many of the critical residues involved in phosphorylation, and are unlikely to have a partner histidine kinase (*Figure 4.2*).

The HTH-cupin fusion FlF may be an amino acid binding transcription factor. Homologues are found in all organofluorine biosynthesis clusters that feature a full-length fluorothreonine transaldolase, as well as putative chlorothreonine BGCs with high levels of homology to the partially sequenced cluster from *Streptomyces sp.* OH5093 (*Figure 4.1C)* [*32*]. In addition, the *V. cholera* transcription factor VC_1968, which has the same domain architecture, is co-crystallized with D-methionine in the PDB entry 1Y9Q.

The conservation of genomic context of FlJ homologues in *Streptomyces* is also striking-proteins with the DUF190 domain are widespread in *Streptomyces*, and almost universally occur in conjunction with *CrcB*-type fluoride exporters. The DUF190 family has distant homology to pII-like proteins, which have a well-documented role in nitrogen metabolism [*33*], where they can integrate nutrient availability signals and participate in the regulation of ammonia transport [*34*] and amino acid metabolism [*35*, *36*]. Recently, the ligand-binding scope of these proteins has been expanded by the discovery of a bicarbonate-binding pII-like protein that regulates carbon fixation in cyanobacteria [*37*], suggesting that related proteins may be participate in diverse small molecule sensing pathways.

Of the 190 top homologues of FlJ in other *Streptomyces* species with genome assemblies available from NCBI refseq, 187 are found clustered with at least one CrcB-like protein. Other actinobacterial gene clusters with a DUF190 protein and no CRCB fluoride exporter are listed in *Appendix 3.1*. Similarly, CrcB fluoride exporters in have a strong association with DUF190 proteins; 233 out of 451 of the top homologues of the CrcB-type exporter from *S. coelicolor* (NP_631108) cluster with at least one DUF190 protein. This diverse complement of signaling
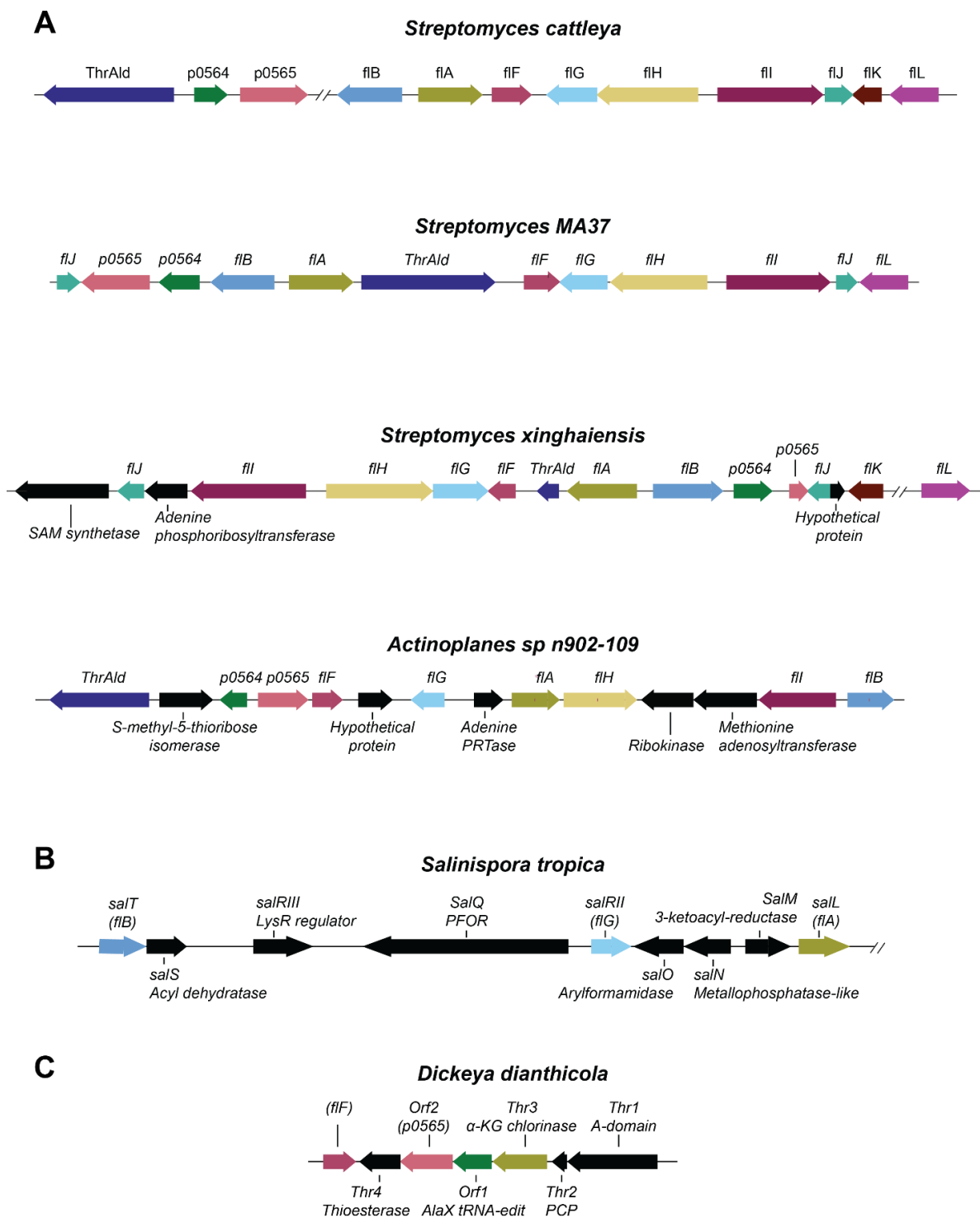
Figure 4.1. *Biosynthetic gene clusters for organofluorines and related compounds, annotated by functional and genetic conservation. (A) Organofluorine BGCs are shown with coding sequences color coded and labeled according to their orthologues in the S. cattleya cluster. Non-homologous proteins are marked in black and are labeled by annotation. (B) The salinosporamide BGC features several elements in common with the organofluorine biosynthesis locus. Genes are labeled according to Moore [38], with S. cattleya orthologues noted in parentheses. (C) The chlorothreonine biosynthesis locus, labeled according to Grgurina [32], with S. cattleya orthologues noted in parentheses.*

82

```
CitB  (E. coli)           --------------MTAPLTLLIVEDETPLAEMHAEYIRHIPGFSQILLAGNLAQARMMI   46
AfsQ1 (S. coelicolor)     ----------------MPSLLLIEDDEAIRTALEL--SLTRQGHRVATAASGEDGLKLL   41
PhoP  (S. coelicolor)     -------------MADQTHVLFVEDDDVIREATQL--ALERDGFAVTAMPDGLSGLESF   44
AbsA2 (S. coelicolor)     ----------------MIRVLLADDETIIRAGVRSILTTEPGIEVVAEASDGREAVELA   43
flG   (S. cattleya)       MTLHRGAATAAVRRQQAAIRLAVVDDNVFLRTGLAHVLAAEPALEVVAESATVREGLSAL   60
flL   (S. cattleya)       ---------------MVTTIALCDPNLLIRGGLRAILAGEPDIEVVAEVADGEQALRTV   44
                                                 ↓
citB  (E. coli)           ERFKPGLILLDNYLPDGRGINLLHELVQAHYP--GDVVFTTAASDMETVSEAVRCGVFDY  104
AfsQ1 (S. coelicolor)     REQRPDLIVLDVMLPGIDGFEVCRRIRRTDQL---PIILLTARNDDIDVVVGLESGADDY   98
PhoP  (S. coelicolor)     RADRPDIALLDVMLPGLDGVSLCRRIRDESTV---PVIMLSARADSIDVVLGLEAGADDY  101
AbsA2 (S. coelicolor)     RKHRPDVALLDIRMPEMDGLTAAGEMRTTNPD--TAVVVLTTFGEDRYIERALDQGVAGF  101
flG   (S. cattleya)       RARQPHVVLLGVASGGREPYGDVAELRRATPV--SRIVAMLAHDDPGAVRRLLAAGAQAA  118
flL   (S. cattleya)       RRLRPDLLFTALDLPGGHGFEVARALARAPETRSTGVVFLAQSFDPALAVQGLMAGGRAF  104
                             ─────────────────── Response regulator domain ───────────────
citB  (E. coli)           LIKPIAYERLGQTLTRFRQRKHMLESIDSASQKQIDEMF----------NAYARGEPKDE  154
AfsQ1 (S. coelicolor)     VVKPVQGRVLDARIRAVLRRGE-------------RESTDSASFGSLVIDRSAMTVTKNG  145
PhoP  (S. coelicolor)     VTKPFDGAVLVARIRAVLRRFGHAGGGDRTEGAGAAETGGVLTFGDLEVDTDGMEVRRAG  161
AbsA2 (S. coelicolor)     LLKASDPRDLISGVRAVASGGSCLSPLVARRL--MTELRRAPS---------PRSEVSGE  150
flG   (S. cattleya)       IAQYSSPEELIGTIRGVAGHRDRVVLSVSQR--------------------TLAGLRET  157
flL   (S. cattleya)       VSTGDPPRRLVTAARAVAAGYAVLPPAVLADGPWLTGSGGAAR---------P---GAAP  152

citB  (E. coli)           LPTGIDPLTLNAVRKLFKEPGVQHTAETVAQALT-----ISRTTARRYLE-------YCA  202
AfsQ1 (S. coelicolor)     EDLQLTPTELRLLLELSRRPGQALSRQQLLRLVWEHDYLGDSRLVDACVQRLRAKVEDVP  205
PhoP  (S. coelicolor)     RPVGLTPTEMRLLLEFSSAPGTVLSRDKLLERVWDYGWGGDTRVVDVHVQRLRTKIGQ--  219
AbsA2 (S. coelicolor)     RTTLLTKREQEVLGMLG----AGLSNAEIAQRLH-----LVEGTIKTYVSAIFTQLEVRN  201
flG   (S. cattleya)       GPHLLTARELQVVALVA----RGLRNSQIARELV-----ITEGTVKRHLSNVYAKLDAAC  208
flL   (S. cattleya)       GVPGLTGRELEVLRLMA----AGMSNAEIAGHLS-----LGEATVKSHVSRLLGKLGLRN  203
                           ─────────────────────────── HTH domain ──────────────────────
citB  (E. coli)           SRHLIIAEIVHGKVGRPQRIYHSG-----   226
AfsQ1 (S. coelicolor)     SSPTLIR---------TVRGVGYRLDPPQ   225
PhoP  (S. coelicolor)     ---DRIE---------TVRGFGYKLKA--   234
AbsA2 (S. coelicolor)     RVQAAIIAYEAGLVKDADLNR--------   222
flG   (S. cattleya)       RTEAVRKAVERGMVRTPVPITDYDLA---   234
flL   (S. cattleya)       RSQAIACAYRTGLVTVPVPGRG-------   225
```

Figure 4.2. *Alignment of flG and flL from S. cattleya with transcription factors featuring canonical response regulator domains. Residues conserved across all sequences in the alignment are marked in* magenta*; residues that participate in the phosphorylation of the response regulator domain are marked in* cyan*. The conserved, phosphorylated aspartate residue is marked with an arrow. Domain boundaries are marked below the alignment.*

proteins in the organofluorine BGC suggests that it may be subject to regulation in response to environmental cues, or to feedback from products of the organofluorine pathway.

Genetic analysis of the role of signaling proteins in organofluorine biosynthesis. The first step towards determining the regulatory architecture of the organofluorine biosynthesis cluster was to construct knockouts of all the regulatory proteins in the cluster. Incidentally, a strain containing a deletion between *flF* through *flL* (Δ(*flF-flL*)::Am^R) was also constructed during the process of trying to obtain a deletion of *flF (Appendix 4.1)*. These knockout strains were also used to generate complemented strains with FLAG$_3$-tagged signaling proteins expressed from the strong constitutive promoter ermEp* *(Figure 4.3A-D)* [39].

The resulting strains were grown for 6 d in the presence of fluoride, after which organofluorine production was assessed by ^{19}F-NMR (*Figure 4.3E*). The Δ*flL*, Δ*flG*, and Δ*flF-flL* strains are completely defective in organofluorine production. Organofluorine biosynthesis in the Δ*flL* and Δ*flG* strains can be restored by the overexpression of FLAG$_3$-tagged FlL and FlG, respectively, and the overexpression of FLAG$_3$-tagged FlG in the Δ*flL* and Δ*flF-L* backgrounds also restores organofluorine production.

In contrast to the dramatic phenotypes of the *flG* and *flL* knockouts, the Δ*flF* strain shows reduced fluorothreonine production but high-level production of fluoroacetate. It was also not possible to restore wild-type fluorothreonine production by overexpression of FLAG$_3$-tagged FlF. While it is possible that the C-terminal FLAG$_3$ tag impedes the function of FlF, previous experiments that appended a C-terminal FLAG tag to the endogenous copy of *flF* did not affect fluorothreonine production *(Appendix 4.2)*. An alternate possibility is that FlF has competing regulatory functions, such that overexpression is unable to restore the wild-type phenotype.

The role of FlJ was also unclear from these experiments. In initial trials, the Δ*flJ* strain failed to produce organofluorines, but under standardized growth conditions it shows only a mild defect in production. This may be related to a delayed onset in the expression of FlA, or to effects of the *flJ* knockout on the expression of the fluoroacetatyl-CoA thioesterase FlK, which is partially disrupted in the Δ*flJ* strain. The Δ*flJ* strain and all its derivatives did show a slight decrease in fluoroacetate biosynthesis relative to the wildtype.

The expression flG and flL in their respective knockout backgrounds restores organofluorine production, and flG alone is competent to restore organofluorine production even in the Δ*(flF-flL)* background. Other mutants were constructed but were not tested under standardized assay conditions *(Appendix 4.2)*. Consistent with the idea that regulation of FlA expression is the primary way that regulatory control of the pathway is exerted, blotting with an anti-FlA antibody reveals that strains that do not produce organofluorines also do not express the fluorinase *(Figure 4.4)*. It is important to note that these findings may be condition specific; for example, the ratio of fluoroacetate to fluorothreonine produced by WT *S. cattleya* changes substantially in chemically defined media as compared to rich media. For example, while the WT strain grown in GYM media produces a 2.3 ± 0.3 ratio of fluoroacetate:fluorothreonine, this ratio is flipped in chemically defined media, where an fluoroacetate:fluorothreonine ratio of 0.68 ± 0.05 is observed. These differences in production phenotype also raise the possibility that some of the mutants constructed may have regulatory defects that were not evident under the standard, rich media conditions employed.

The regulatory architecture of the cluster was further explored through RNAseq of the Δ*flG* and Δ*flL* strains. In comparison with previously collected wildtype data, the Δ*flG* strain shows

A

1. 2. 3.

← 1500 bp
← 1000 bp
← 700 bp
← 500 bp

1. WT
2. ΔflF
3. ΔflF + ermEp*-flF

B

1. 2. 3.

← 1500 bp
← 1000 bp
← 700 bp
← 500 bp

1. WT
2. ΔflG
3. ΔflG + ermEp*-flG

C

1. 2. 3. 4. 5.

← 3000 bp
← 1500 bp
← 1000 bp
← 700 bp
← 500 bp

1. WT
2. ΔflJ
3. ΔflJ + ermEp*-flJ
4. ΔflJ + ermEp*-flG
5. ΔflJ + ermEp*-flL

D

1. 2. 3. 4.

← 3000 bp
← 1500 bp
← 1000 bp
← 700 bp
← 500 bp

1. WT
2. ΔflL
3. ΔflL + ermEp*-flL
4. ΔflL + ermEp*-flG
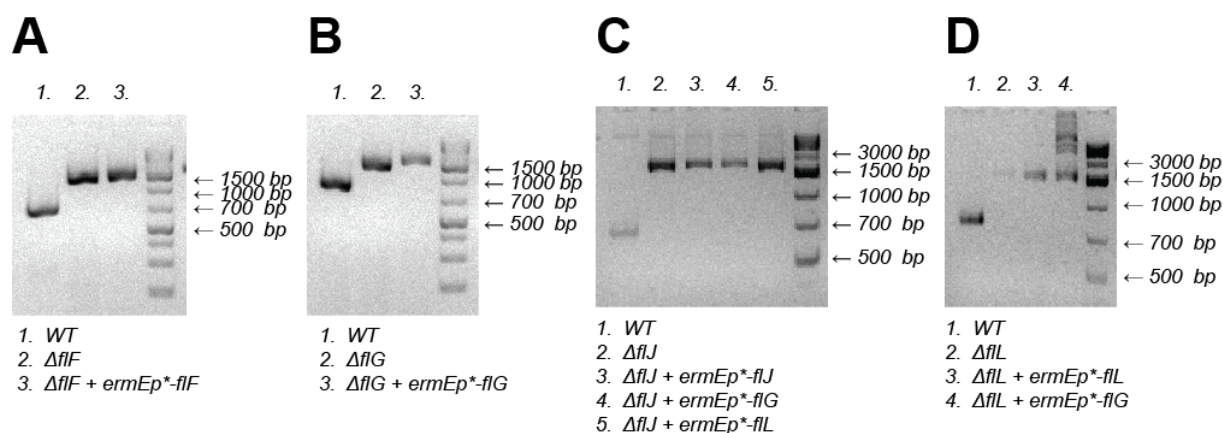
E

	FAc (mM)		FThr (mM)	
WT	0.65 ±	0.17	0.17 ±	0.04
ΔflF-L		N.D.		
ΔflF-L + ermEp-flG*	0.26 ±	0.03	0.14 ±	0.02
ΔflF	0.65 ±	0.09	0.03 ±	0.01
ΔflF + ermEp-flF*	0.85 ±	0.09	0.04 ±	0.00
ΔflG		N.D.		
ΔflG + ermEp-flG*	0.71 ±	0.06	0.35 ±	0.07
ΔflJ	0.25 ±	0.02	0.20 ±	0.02
ΔflJ + ermEp-flJ*	0.32 ±	0.07	0.29 ±	0.07
ΔflJ + ermEp-flL*	0.33 ±	0.04	0.22 ±	0.00
ΔflJ + ermEp-flG*	0.36 ±	0.01	0.21 ±	0.04
ΔflL		N.D.		
ΔflL + ermEp-flL*	1.03 ±	0.01	0.14 ±	0.03
ΔflL + ermEp-flG*	0.57 ±	0.18	0.18 ±	0.07

Figure 4.3. *Genetic confirmation and orgnaofluorine production phenotypes of S. cattleya regulatory knockouts and knock-ins. Genomic DNA from strains was amplified with primers flanking the flF, flG, flJ, and flL to confirm the knockouts. (A-D) Knockout strains and their derivatives show a ~1500 bp band corresponding to the resistance cassette, while the wild-type shows a shorter band corresponding to the native gene. (E) Strains were grown in triplicate and organofluorine production was quantified by ^{19}F NMR.*
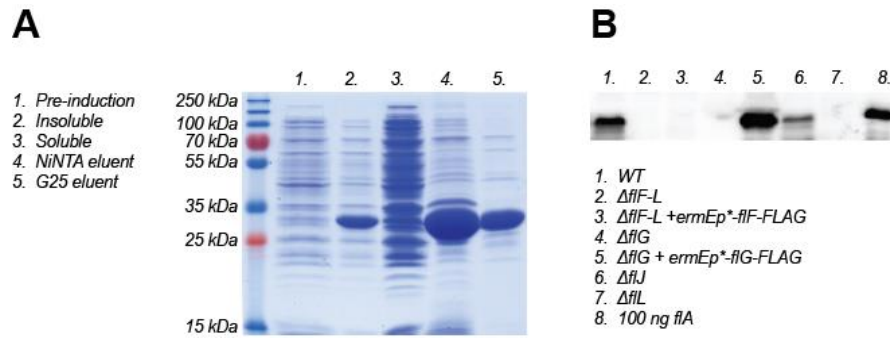
Figure 4.4. *Purification of flA and analysis of flA expression. (A) The fluorinase, flA, was expressed and purified in E. coli. (B) The resulting protein was used to generate antiserum which was used to blot lysates from S. cattleya mutants.*

	ΔflG	*ΔflL*	*WT*
flA	0.00	0.01	18.28
flB	0.01	0.00	0.55
flC	0.01	0.02	0.24
flD	0.01	0.01	0.04
flE	0.39	0.42	0.12
flF	0.01	0.01	0.84
flG	0.01	0.03	1.08
flH	1.41	0.03	0.69
flI	1.02	0.00	2.92
flJ	0.45	0.02	0.43
flK	0.39	0.09	0.79
flL	0.59	0.00	0.83
ThrAld	0.01	0.01	0.79
SCAT_p0564	3.80	3.16	1.04
SCAT_p0565	0.53	0.23	0.23

Table 4.1. *RNA-seq analysis of S. cattleya signaling knockouts. Reads per kilobase million (RPKM) for genes in the organofluorine biosynthesis locus are displayed for ΔflG and ΔflL strains are displayed, along with previously acquired data from the wild-type strain.*
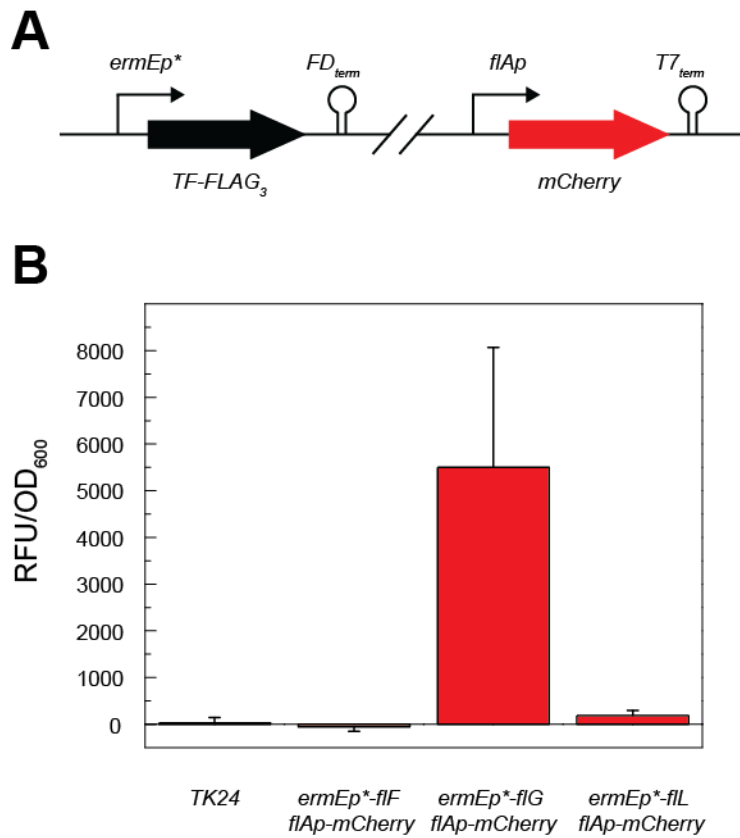
Figure 4.5. *Reporter-based approach for the analysis of transcriptional activation of the flA promoter. (A) Strains of S. lividans TK24 bearing transcription factors under the control of ermEp* and mCherry under the control of the flA promoter were constructed. (B) The resulting strains, along with the wild-type strain, were grown and assayed for mCherry fluorescence.*

dramatically reduced expression of the fluorinase (FlA), FlF, and the fluorothreonine transaldolase, while other elements of the cluster are still expressed. In contrast, the *ΔflL* strain shows a widespread reduction in transcription throughout the cluster, including *flG* (*Table 4.1*). It is unclear whether the lack of expression of the fluorothreonine transaldolase in both cases is a result of downstream effects on FlF, or if FlG itself plays an additional role in driving the expression of that protein. These preliminary results suggest that FlL drives the expression of accessory components of the cluster, while FlG is the master regulator of fluorinase expression.

This finding was further tested through the construction of reporter strains in the heterologous host *S. lividans* TK24. Reporter strains expressed FLAG$_3$-tagged versions of the transcription factors FlF, FlG and FlL in the presence of an integrated copy of an mCherry fluorescent reporter under the control of the *flA* promoter. It was determined that FlG was the only transcription factor able to drive expression from the *flA* promoter at a significant level (*Figure 4.5*). These results are consistent with RNAseq experiments, and the ability of flG overexpression to restore organofluorine production in the *Δ(flF-flL)* strain. Preliminary experiments with reporter constructs using the *flH*, *flI*, and *ThrAld* promoters did not detect fluorescence above background; it is possible that the use of a more sensitive reporter such as β-glucuronidase or bacterial luciferase would facilitate these measurements [*40*, *41*].

Amino acid binding properties of FlF. Given the bioinformatic assignment of FlF as a potential amino acid-binding transcription factor, and the defect of the *ΔflF* knockout strain in fluorothreonine production, attempts were made to measure the binding affinity of this protein towards threonine and fluorothreonine. Initial purification of the protein determined that the solubility of FlF is limited, and the protein forms a gel-like aggregate even at relatively low concentrations (~1 mg/mL). Consequently, assays were performed with the MBP-fusion, which can be prepared in high yield (*Figure 4.6A*). Equilibrium dialysis measurements revealed that the protein binds to both fluorothreonine and threonine, but fluorothreonine is bound with higher affinity (*Figure 4.6B*). Dissociation constants are reported assuming noncooperative binding with a single binding site per monomer. Selective binding to fluorothreonine was further confirmed with a competitive binding assay, where fluorothreonine achieved a higher saturation of MBP-FlF (37% vs. 14%, $n = 2$) despite being present at a lower initial concentration (190 μM vs. 1000 μM). Unfortunately, attempts to build a complete binding curve failed, with measurements becoming increasingly erratic at higher ligand concentrations. This breakdown is likely due to the sensitivity of equilibrium dialysis to small measurement errors when the concentration of ligand exceeds the concentration of receptor. Consequently, these results do not address cooperativity or ligand binding stoichiometry.

The role of transition metal loading in determining amino acid affinity was also explored, as members of the cupin family have been observed to bind a wide variety of transition metals [*42*]. ICP-MS was used to assess metal loading of purified MBP-FlF, and it was determined that the protein was predominantly loaded with iron (67%; consistent with the dirty-blue color of the concentrated protein) [*43*], but also had a substantial quantity of zinc (47%). MBP-FlF was stripped and reconstituted with different metals; preliminary results show that both Ni^{2+} and Fe^{2+} support high affinity binding, while Zn^{2+} does not (*Figure 4.6C*). The addition of Cu^{2+} in the form of $CuSO_4$ resulted in near total precipitation of MBP-flF.

Because pII-like proteins frequently bind to nucleotides in addition so other small molecule ligands, FlJ was screened for affinity to AMP, ADP, and ATP using equilibrium dialysis. Only ATP binds with high affinity (*Figure 4.7B*). Previous investigations of homologous proteins have
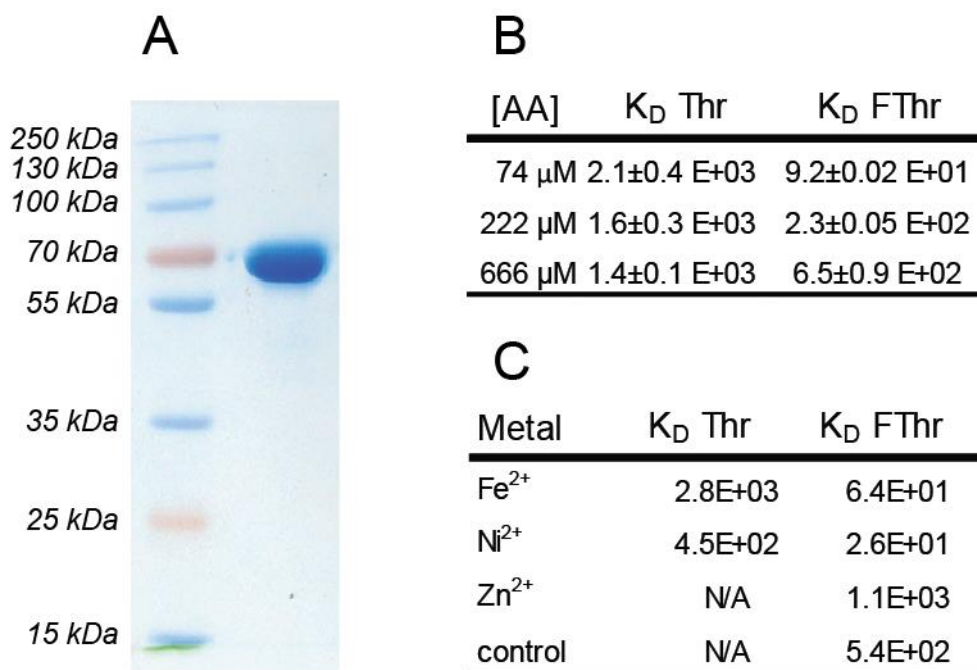
Figure 4.6. *Investigation of the amino acid binding properties of flF. (A) The protein was expressed and purified as an MBP-fusion. (B) Equilibrium dialysis was then employed at a variety of amino acid concentrations to measure bindind to fluorothreonine and threonine. (C) Preliminary investigation of the role of transition metal loading by single-replicate competitive binding assays with threonine and fluorothreonine.*
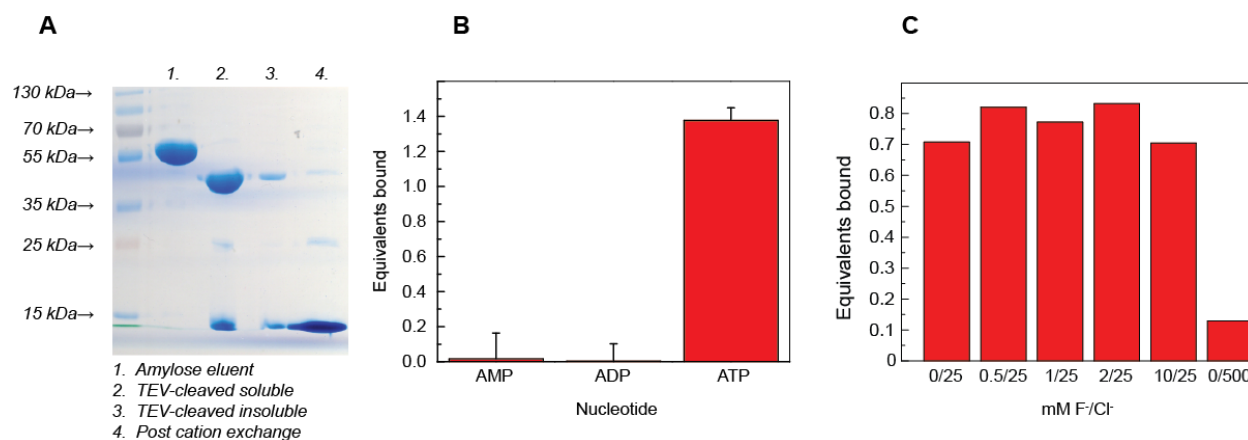
Figure 4.7. *Preliminary characterization of flJ. (A) The protein was expressed as an MBP fusion and purified by amylose affinity chromatography and cation exchange. (B) The affinity of flJ for AMP, ADP, and ATP was measured in the presence of 2.5 mM nucleotide; fractional binding of protein with nucleotide is displayed. (C) The affinity of flJ for ATP was measured in the presence of 0.25 mM ATP and variable concentrations of sodium fluoride and sodium chloride.*

89

A.

100 kDa
70 kDa
55 kDa
35 kDa
25 kDa
15 kDa

1. Post-induction
2. Soluble
3. Insoluble
4. Amylose eluent

B.

100 kDa
70 kDa
55 kDa
35 kDa
25 kDa
15 kDa

1. Post-induction
2. Soluble
3. Insoluble
4. Amylose eluent

C.

1. No protein.
2. 100 µg/mL MBP-flG

D.

poly-dIdC
flA promoter

1. No protein.
2. 100 µg/mL MBP-flL
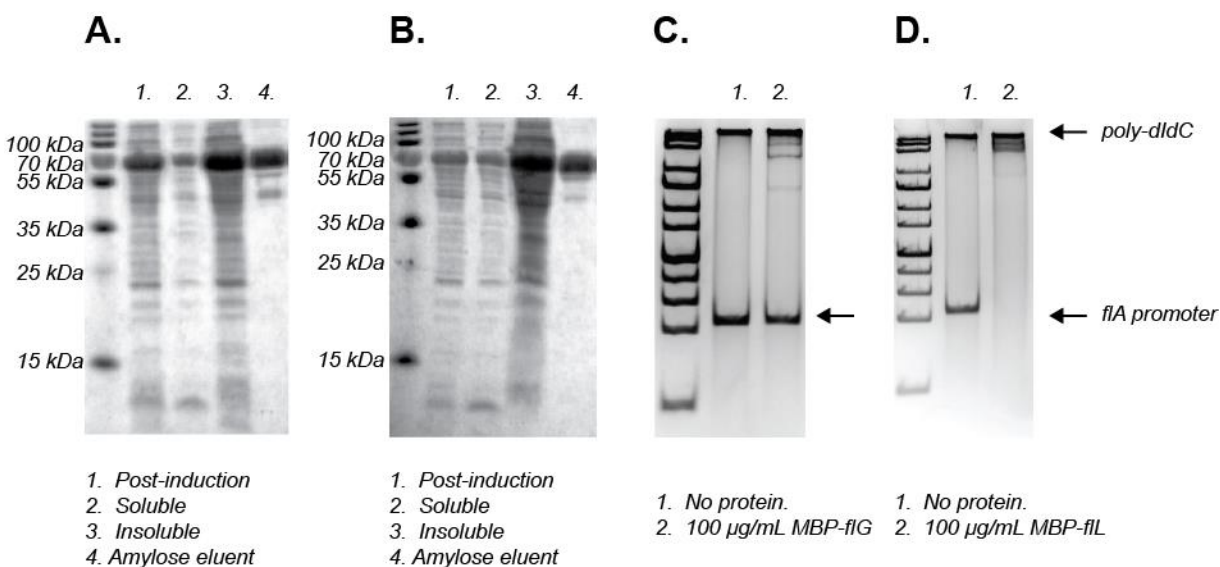
Figure 4.8. *Preliminary in vitro characterization of flG and flL. (AB) The proteins were expressed as MBP fusions and purified by amylose affinity chromatography. (CD) The affinity of the MBP fusions towards the flA promoter was measured by electrophoretic mobility shift assay.*

demonstrated that the binding of nucleotide and other ligands can demonstrate positive heterotropic cooperativity [*37*]. Consequently, the ability of fluoride to alter ATP binding was tested, but no apparent relationship was observed (*Figure 4.7C*).

Preliminary characterization of MBP-FIL and MBP-FIG. The transcription factors FlL and FlG could be expressed and purified in high yield as MBP fusions *(Figure 4.8AB).* Unfortunately, all attempts to isolate the transcription factors universally failed due to vigorous precipitation upon cleavage of the MBP fusion partner with TEV protease. The fusion proteins were used in gel shift assays with the promoter region of *flA*. Binding of the *flA* promoter was observed, especially for FlL, but the formation of multiple bands was suggestive of nonspecific binding *(Figure 4.8CD).*

Untargeted metabolomics to identify novel organofluorine-dependent compounds. The inventory of organofluorine-dependent small molecules in *S. cattleya* has not been systematically described beyond ^{19}F-NMR based approaches, which are limited in sensitivity. To this end, we undertook a systematic, untargeted metabolomics study to identify new organofluorine dependent compounds. Wild-type, *ΔThrAld*, *ΔflA*, and *Δ(flF-flL)* strains were grown in quintuplicate, harvested at 6 and 14 d, and subjected to reverse phase HPLC-MS-based untargeted metabolomics. Similar sample loading was achieved across replicates *(Figure 4.9 A).* A variety of compounds were identified as significantly upregulated in binary comparisons between WT and mutant strains, and there is considerable overlap in the compounds that appear to be differentially regulated between the three knockout strains *(Figure 4.9 B-H* and *Appendix 5).* Notably, this suggests that these compounds may be fluorothreonine-derived or the product of fluorothreonine-dependent metabolic processes. However, comparison with the *Δp0564* and *Δp0565* strains did not demonstrate any significant increase in these peaks, so their production depends strongly on the intracellular concentration of fluorothreonine or fluorothreonyl-tRNA. The comparison with the ΔflA strain also produced a variety of hits that were not replicated in the *Δ(flF-flL)*and *ΔThrAld* strains *(Figure 4.9I)*; these may be a result of the bald phenotype of the *flA* knockout, which has a spontaneous mutation rendering it deficient in sporulation.
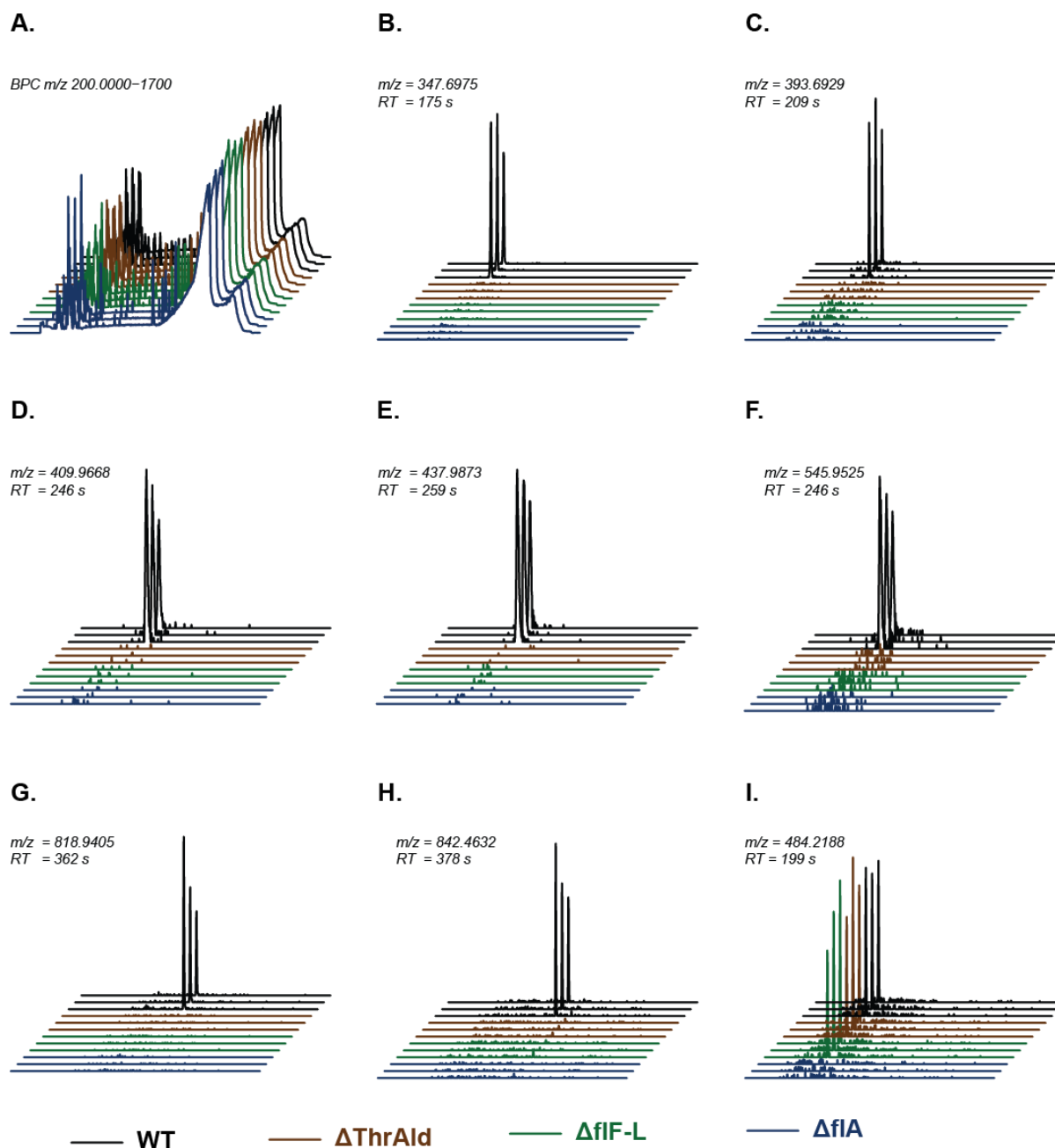
Figure 4.9. *Untargeted metabalomics of S. cattleya strains. (A) Base peak chromatograms show similar sample loading across replicates. (B-H) Several peaks were reproducibly present in the wild-type strain, but were extremely reduced or absent in organofluorine-related knockouts. (I) Several of these compounds were specific to the ΔflA strain; these may be related to morphological development rather than organofluorine metabolism.*

4.4   Conclusions

The work presented here has outlined some of the central features of the regulation of organofluorine metabolism in *S. cattleya*. The atypical response regulators FlG and FlL are the essential regulatory components of the cluster, and act in a cascade with FlL as the top-level regulator and FlG acting as the master regulator of fluorinase expression *(Scheme 4.1)*. This

conclusion is supported by the phenotypes of *S. cattleya* mutant strains, reporter assays with the *flA* promoter, and RNAseq data. The cupin-containing transcription factor FlF may play a more nuanced role, as the Δ*flF* strain is defective in fluorothreonine production, but still makes fluoroacetate. Consistent with this observation, the MBP-fusion of FlF can selectively bind fluorothreonine, indicating that this transcription factor may have the capacity to respond to fluorothreonine production. While the putative fluoride responsive signaling protein FlJ has not been shown to have a definitive role in the cluster, further work on this protein is of acute interest as it is the most promising indicator for the discovery of novel organofluorine chemistry.
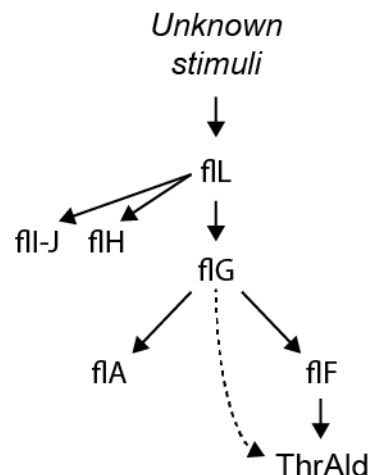
Scheme 1: *Proposed regulatory architecture of the organofluoroine biosynthesis locus.*

4.5 References

1. Walker MC, Wen M, Weeks AM, Chang MCY (2012) Temporal and fluoride control of secondary metabolism regulates cellular organofluorine biosynthesis. *ACS Chem Biol* 7(9):1576–1585.

2. Baker JL, et al. (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* 335(6065):233–235.

3. Liu G, Chater KF, Chandra G, Niu G, Tan H (2013) Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev* 77(1):112–143.

4. Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology* 8(2):208–215.

5. Chater KF (2016) Recent advances in understanding *Streptomyces*. *F1000Res* 5. doi:10.12688/f1000research.9534.1.

6. Santos-Beneit F, et al. (2011) The RNA polymerase omega factor RpoZ is regulated by PhoP and has an important role in antibiotic biosynthesis and morphological differentiation in *Streptomyces coelicolor*. *Appl Environ Microbiol* 77(21):7586–7594.

7. Nazari B, et al. (2013) Chitin-induced gene expression in secondary metabolic pathways of Streptomyces coelicolor A3(2) grown in soil. *Appl Environ Microbiol* 79(2):707–713.

8. Li X, et al. (2015) ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci Rep* 5:14831.

9. Wang W, et al. (2014) Angucyclines as signals modulate the behaviors of *Streptomyces coelicolor*. *Proc Natl Acad Sci USA* 111(15):5688–5693.

10. Anderson TB, Brian P, Champness WC (2001) Genetic and transcriptional analysis of absA, an antibiotic gene cluster-linked two-component system that regulates multiple antibiotics in *Streptomyces coelicolor*. *Mol Microbiol* 39(3):553–566.

11. Sherwood EJ, Bibb MJ (2013) The antibiotic planosporicin coordinates its own production in the actinomycete *Planomonospora alba*. *Proc Natl Acad Sci USA* 110(27):E2500–E2509.

12. Wang L, et al. (2009) Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. *Proc Natl Acad Sci USA* 106(21):8617–8622.

13. Dun J, et al. (2015) PapR6, a putative atypical response regulator, functions as a pathway-specific activator of pristinamycin II biosynthesis in *Streptomyces pristinaespiralis*. *J Bacteriol* 197(3):441–450.

14. Furuya K, Hutchinson CR (1996) The DnrN protein of *Streptomyces peucetius*, a pseudo-response regulator, is a DNA-binding protein involved in the regulation of daunorubicin biosynthesis. *J Bacteriol* 178(21):6310–6318.

15. Pedrolli DB, et al. (2012) A highly specialized flavin mononucleotide riboswitch responds differently to similar ligands and confers roseoflavin resistance to *Streptomyces davawensis*. *Nucleic Acids Res* 40(17):8662–8673.

16. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1):539.

17. Finn RD, et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279–D285.

18. Kieser T (2000) *Practical Streptomyces Genetics* (John Innes Foundation).

19. Weeks AM (2013) Molecular insights into fluorine chemistry in living systems. Dissertation (UC Berkeley, Berkeley, CA). Available at: http://escholarship.org/uc/item/3sc8k37r [Accessed April 13, 2017].

20. Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* 100(4):1541–1546.

21. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.

22. Phelan RM, et al. (2017) Development of next generation synthetic biology tools for use in *Streptomyces venezuelae*. *ACS Synth Biol* 6(1):159–166.

23. Walker MC (2013) Expanding the scope of organofluorine biochemistry through the study of natural and engineered systems. Dissertation (UC Berkeley, Berkeley, CA). Available at: http://escholarship.org/uc/item/9n89c8zv [Accessed April 26, 2017].

24. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27(3):431–432.

25. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, ed Walker J (Humana Press), pp 571–607.

26. Carey MF, Peterson CL, Smale ST (2013) Electrophoretic mobility-shift assays. *Cold Spring Harb Protoc* 2013(7):pdb.prot075861.

27. Barbe V, et al. (2011) Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites. *J Bacteriol* 193(18):5055–5056.

28. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.

29. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

30. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78(3):779–787.

31. Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24(21):2534–2536.

32. Fullone MR, et al. (2012) Insight into the structure–function relationship of the nonheme iron halogenases involved in the biosynthesis of 4-chlorothreonine – Thr3 from *Streptomyces* sp. OH-5093 and SyrB2 from *Pseudomonas syringae pv. syringae* B301DR. *FEBS J* 279(23):4269–4282.

33. Ninfa AJ, Jiang P (2005) PII signal transduction proteins: sensors of α-ketoglutarate that regulate nitrogen metabolism. *Curr Opin Microbiol* 8(2):168–173.

34. Gruswitz F, O'Connell J, Stroud RM (2007) Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GlnK, at 1.96 Å. *Proc Natl Acad Sci USA* 104(1):42–47.

35. Adler SP, Purich D, Stadtman ER (1975) Cascade control of *Escherichia coli* glutamine synthetase. Properties of the PII regulatory protein and the uridylyltransferase-uridylyl-removing enzyme. *J Biol Chem* 250(16):6264–6272.

36. Atkinson MR, Kamberov ES, Weiss RL, Ninfa AJ (1994) Reversible uridylylation of the *Escherichia coli* PII signal transduction protein regulates its ability to stimulate the dephosphorylation of the transcription factor nitrogen regulator I (NRI or NtrC). *J Biol Chem* 269(45):28288–28293.

37. Wheatley NM, et al. (2016) A PII-like protein regulated by bicarbonate: structural and biochemical studies of the carboxysome-associated CPII protein. *J Mol Biol*. doi:10.1016/j.jmb.2016.07.015.

38. Udwary DW, et al. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* 104(25):10376–10381.

39. Bibb MJ, White J, Ward JM, Janssen GR (1994) The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Mol Microbiol* 14(3):533–545.

40. Myronovskyi M, Welle E, Fedorenko V, Luzhetskyy A (2011) β-Glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl Environ Microbiol* 77(15):5370–5383.

41. Craney A, et al. (2007) A synthetic *luxCDABE* gene cluster optimized for expression in high-GC bacteria. *Nucleic Acids Res* 35(6):e46.

42. Hobbs JK, et al. (2016) KdgF, the missing link in the microbial metabolism of uronate sugars from pectin and alginate. *Proc Natl Acad Sci USA* 113(22):6188–6193.

43. Yan F, Munos JW, Liu P, Liu H (2006) Biosynthesis of fosfomycin, re-examination and re-confirmation of a unique Fe(II)- and NAD(P)H-dependent epoxidation reaction. *Biochemistry* 45(38):11473–11481.

Appendix 1: *Plasmids and oligonucleotides*

Table S1.1. Plasmid constructs (A), strains (B) and oligonucleotides (C) used in Chapter 2.

A

Name	No.	Description	Source
pet16b-TRS	1773	His$_{10}$-SCAT_TRS (T7), *lacI,* Cb^R, ColEI	This study
PSV272.1-SCAT_p0564	2115	His$_6$-MBP-TEV-SCAT_p0564 (T7), *lacI,* Km^R, ColEI	This study
pSET152-ermEp*	592	ΦC31 int, ermEp*, Am^R, pUC	John Innes Center
pSET152-ermEp*-SCAT_p0564	2829	ΦC31 int, ermEp*-SCAT_p0564, Am^R, pUC	This study
pIJ10702-SCAT_tRNA25	1374	SCAT_tRNA25 (T7)	This study
PUZ8002	406	tra, Km^R, RP4	John Innes Center
pRare2	191	CmR, p15a, tRNA^{Ile}(AUA), tRNA^{Ser}(AGG), tRNA^{Ser}(AGA), tRNA^{Asp}(CUA), tRNA^{Gly}(CCC), tRNA^{Gly}(GGA)	Novagen

B

Strain	No.	Genotype	Source
S. cattleya ATCC 35852 (NRRL 8057)	385	Wild-type	ATCC
S. cattleya Δp0564	1697	Δp0564::Am^R	This study
S. cattleya Δp0565	1700	Δp0565::Am^R	This study
S. coelicolor M1152	505	Δact Δred Δcpk Δcda rpoB(C1298T)	John Innes Center
S. coelicolor M1152 + ermEp*-p0564	2512	M1152 + pSET152-ermEp*-p0564 in ΦC31 *attB*	This study
S. coelicolor M1152 + ermEp*-empty	2511	M1152 + pSET152-ermEp*-empty in ΦC31 *attB*	This study
E. coli BL21 (de3)	1677	F– ompT gal dcm lon hsdSB(rB–mB–) [malB+]K-12(λS)	Novagen
E. coli GM272	404	dam-3 dcm-6 hsdS21 metB1 lacY1 or Z4 galK2 galT22 mtl-2 tonA2 or A31 tsx-1 or78 supE44 (thi-1)?	Yale CGSC

C

Name	Sequence
J303	Catcatcatcacagcagcggccatatcgaaggtcgtcatatgtctcagatccgtgtggtcatcaaacgcg
J304	Aactcagcttcctttcgggctttgttagcagccggatcctcacacctggacgcggcgctccaccacatca
J400	Gcgagttctggctggctagcccgtttgatctcgagtgcggccgcaagctttcagccggtcgcggcgatcggctcg
J401	Aataacaacaacctcgggatcgaggaaaacctgtattttcagggcatggaaaccgagagcacatacgacaagctc
J062	Gagaattcgcggccgcataatacgactcactatagccccaatagctcagtcggcagagcg
J063	Ggccgcaattaaccctcactaaagggatcctggagccccaatacggaatcgaaccgtaga
J064	Tggagccccaatacggaatcgaaccg
J071	Gcagagcttatcgatgataagcggtcaaacatgag
J118	Ggtaggatcgtctagaacaggaggccccatatggaaaccgagagcacatacgacaagctc
J119	Caagcttgggctgcaggtcgactctagaggatcctcagccggtcgcggcgatcggctcgg
J144	Ggccggaactgatcggtcggtcacgggaggtggcccggtatgccc
J145	Gctccgggcggtacccctggttggccacgagcgggaagcgg
J146	Cggcccggtccgaccgaccctaggtggaggtggcggtcagg

J147	Ccgggccacctcccgtgaccgaccgatcagttccggcc
J235	Agaggcgggatgcgaagaatgcgat
J236	Cggtatatccatcctttttcgca
J239	Gttccggcctaccgggacggggtgcccggggacccgcggagccgcggggggtcatgtaggctggagctgcttcgaagttcc
J240	Ctcaccgtctccgagcttcccaccgaccgccggcacggcggctgagaggtgttcatgattccggggatccgtcgacctgc
J241	Gtgcagatgggcgtcgaagatccgggaaccggtcccggccgtaccgtcggtcatgtaggctggagctgcttcgaagttcc
J242	Tgacaccatgggcatggtggtccctcccccacctcatcctccggaggacccttgtgattccggggatccgtcgacctgc

Table S1.2 Plasmid constructs (A), gBlocks (B) and oligonucleotides (C) used in Chapter 3.

A

Name	No.	Description	Source
pet16b-CCA	2830	His$_{10}$-PRE-EC_CCA (T7), *lacI,* Cb^R, ColEI	This study
PSV272.1-SCAT_p0564 R31A	2831	His$_6$-MBP-TEV-SCAT_p0564 R31A (T7), *lacI,* Km^R, ColEI	This study
PSV272.1-SCAT_p0564 T32C	2832	His$_6$-MBP-TEV-SCAT_p0564 T32C (T7), *lacI,* Km^R, ColEI	This study
PSV272.1-SCAT_p0564 T32A	2833	His$_6$-MBP-TEV-SCAT_p0564 T32A (T7), *lacI,* Km^R, ColEI	This study
PSV272.1-SCAT_p0564 C49A	2834	His$_6$-MBP-TEV-SCAT_p0564 C49A (T7), *lacI,* Km^R, ColEI	This study
PSV272.1-SCAT_p0564 C49S	2835	His$_6$-MBP-TEV-SCAT_p0564 C49S (T7), *lacI,* Km^R, ColEI	This study
PSV272.1-SCAT_p0564 K48A	2836	His$_6$-MBP-TEV-SCAT_p0564 K48A (T7), *lacI,* Km^R, ColEI	This study
PSV272.1-SCAT_p0564 A104S	2837	His$_6$-MBP-TEV-SCAT_p0564 A104S (T7), *lacI,* Km^R, ColEI	This study
pET16b-SCAT_tRNA25	2379	SCAT_tRNA25 (T7) , *lacI,* Cb^R, ColEI	This study
PSV272.1-ClSAD	2377	His$_6$-MBP-TEV-ClSAD (T7), *lacI,* Km^R, ColEI	Novagen

B

gBlock	Sequence
J411	aataacaataacaacaacctcgggatcgaggaaaacctgtattttcagggcatgcagcgccacatctacctcaaggacacgtaccggttcaccgtcaccacgcgtgtggtcgatgcgggaagcggtgagctcgggcactggctcaccttggaggacaacatcttccacccgcagggcggcggtcagccgagcgacgtgggaacggtcgacgagacggccggcgcgtcccttcaaggcgcccgggacggaagagcacgtcgtacgcctctcgtgcgagcggcccttcaccatcggcgacgaggtcaccagtgccatcgaccccgaggcacgccgccgccacgcggccctgcacacctgcggtcacgtggtggacgggttcgtacggcagctcggcttccgccaccgggtgagcaaccacttcccccgggcaggcgcgcatcgagttcgacgccggcgagggggacaagcccgacctccagcagctggcggcgacggtggaggaacgcacccggcaggccatcgaggccgatcgcaaggtgtacgcgaccgagcagggcgaccggcgcatcatcggcatcgacggcctgcaggaggaccgtgcggcggcacccacgtcagctcgctcggacagctcaccggcttctcgctgcgcgttcggtcaagatcaagggcggcgtcctgaaggtggggtacgtcgttgagcacgtctgaaagcttgcggccgcactcgagatcaaacgggctagccagcagaactcgccc

C

Name	Sequence
J425F	actatagggggaattgtgagcggataacaattcccctctagCAAGCTTTACGTGACgccccaatagctcagtcggcagagc
J425R	ccaactcagcttcctttcgggctttgttagcagccgTTTGCTGGGTGCAAAATTtggagccccaatacggaatcgaaccg
J503	Tccgaggccgcgaaagccatcgtggtcatggtcaa
J504	Ttgaccatgaccacgatggctttcgcggcctcgga
J505	Tctccgaggccgcgaaatccatcgtggtcatggtca
J506	Tgaccatgaccacgatggatttcgcggcctcggaga
J507	Aacgtctggcgggcagcgtgtccggcacgatcct
J508	Aacggcaggatcgtgccggacacgctgcccgcca
J489	Gcggcgtgcatcgtggtcatggtcaagatcgg
J490	Gcacgccgcggcctcggagacgtcgtt
J511	Atccaccggaggggcgggcggagatcgtcagcccga
J512	Ttgccgcgcatcgggctgacgatctccgcccgcccctccggt
J513	Atcgaccatccaccggagggggcgaccgagatcgtca

99

J514	Tgacgatctcggtcgcccccctccggtggatggtcgat
J515	Atccaccggaggggcggtgcgagatcgtcagcccgat
J516	Atcgggctgacgatctcgcaccgcccctccggtggat
J517	agcagcggccatCTAGAAGTGCTTTTTCAGGGCcCgcatatgaagatttatctggtcggtggtgctgttc
J518	Gcttcctttcgggctttgttagcagccggatcctcattcaggctttgggcaacgttgttc

Table S1.3. Plasmid constructs (A), strains (B) oligonucleotides (C) and (D) gBlocks used in Chapter 4.

A

Name	No.	Description	Source
pSET152-ermEp*	592	ΦC31 int, ermEp*, Am^R, pUC	John Innes Center
pSET152-Hyg-ermEp*	1645	ΦC31 int, ermEp*, Hyg^R, pUC	This study
pSET152-Hyg-ermEp*-flG-FLAG$_3$	1777	ΦC31 int, ermEp*, Hyg^R, pUC	This study
pSET152-Hyg-ermEp*-flL-FLAG$_3$	1809	ΦC31 int, ermEp*, Hyg^R, pUC	This study
pSET152-Hyg-ermEp*-flF-FLAG$_3$	1776	ΦC31 int, ermEp*, Hyg^R, pUC	This study
pSET152-Hyg-ermEp*-flJ-FLAG$_3$	2838	ΦC31 int, ermEp*, Hyg^R, pUC	This study
pAV-1	2398	VWB int, ermEp*, Am^R, pUC, mCherry	This study
pAV-flAp	2839	VWB int, ermEp*, Am^R, pUC, flAp-mCherry	This study
pAV-flHp	2840	VWB int, ermEp*, Am^R, pUC, flHp-mCherry	This study
pAV-flIp	2841	VWB int, ermEp*, Am^R, pUC, flIp-mCherry	This study
pAV-ThrAldp	2842	VWB int, ermEp*, Am^R, pUC, ThrAldp-mCherry	This study
PSV272.1-flF	1597	MBP-TEV-flF (GTG start)	This study
PSV272.1-flG	1595	MBP-TEV-flG	This study
PSV272.1-flJ	1810	MBP-TEV-flJ	This study
PSV272.1-flL	2843	MBP-TEV-flL	This study
H20	1542	Cosmid containing the organofluorine biosynthesis locus.	This study
A11.15	1543	Cosmid containing the putative lanthipeptide locus surrounding SCAT_0531.	This study
ΔflF cosmid	1564	H20 ΔflF::Am^R	This study
ΔflF plasmid	2844	Re-assembled ΔflF cosmid with 2 kb homology arms.	This study
ΔflG cosmid	1612	H20 ΔflG::Am^R	This study
PUZ8002	406	tra, Km^R, RP4	John Innes Center
pRare2	191	CmR, p15a, tRNA^{Ile}(AUA), tRNA^{Ser}(AGG), tRNA^{Ser}(AGA), tRNA^{Asp}(CUA), tRNA^{Gly}(CCC), tRNA^{Gly}(GGA)	Novagen

B

Strain	No.	Genotype	Source
S. cattleya ATCC 35852 (NRRL 8057)	385	Wild-type	ATCC
S. cattleya ΔflF	2580	*ΔflF*::Am^R	This study
S. cattleya ΔflG	1664	*ΔflG*::Am^R	This study
S. cattleya ΔflJ	1535	*ΔflJ*::Am^R	This study
S. cattleya ΔflL	1537	*ΔflL*::Am^R	This study
S. cattleya ΔflF-L	1666	*ΔflF-L*::Am^R	This study
S. cattleya ΔflG + ermEp*-flG-FLAG$_3$	1803	*ΔflG*::Am^R, pSET152-Hyg-ermEp*-flG-FLAG3 in ΦC31 attB	This study
S. cattleya ΔflJ + ermEp*-flJ-FLAG$_3$	1806	*ΔflJ*::Am^R, pSET152-Hyg-ermEp*-flJ-FLAG3 in ΦC31 attB	This study

S. cattleya ΔflJ + ermEp*-flG-FLAG₃	2845	*ΔflJ*::AmᴿR, pSET152-Hyg-ermEp*-flG-FLAG3 in ΦC31 *attB*	This study
S. cattleya ΔflJ + ermEp*-flL-FLAG₃	2846	*ΔflJ*::Amᴿ, pSET152-Hyg-ermEp*-flL-FLAG3 in ΦC31 *attB*	This study
S. cattleya ΔflL + ermEp*-flL-FLAG₃	2333	*ΔflL*::Amᴿ, pSET152-Hyg-ermEp*-flL-FLAG3 in ΦC31 *attB*	This study
S. cattleya ΔflL+ ermEp*-flG-FLAG₃	2848	*ΔflJ*::Amᴿ, pSET152-Hyg-ermEp*-flG-FLAG3 in ΦC31 *attB*	This study
S. cattleya ΔflF+ ermEp*-flF-FLAG₃	2847	*ΔflJ*::Amᴿ, pSET152-Hyg-ermEp*-flF-FLAG3 in ΦC31 *attB*	This study
S. cattleya ΔThrAld	1536	*ΔThrAld*::Amᴿ	This study
S. cattleya ΔSCAT_0531	2323	*ΔSCAT_0531*::Amᴿ	This study
S. lividans TK24	1308	Wild-type	John Innes Center
S. lividans TK24 + ermEp*-flG-FLAG₃	2849	pSET152-Hyg-ermEp*-flG-FLAG₃ in ΦC31 *attB*	This study
S. lividans TK24 + ermEp*-flL-FLAG₃	2850	pSET152-Hyg-ermEp*-flL-FLAG₃ in ΦC31 *attB*	This study
S. lividans TK24 + ermEp*-flF-FLAG₃	2851	pSET152-Hyg-ermEp*-flF-FLAG₃ in ΦC31 *attB*	This study
S. lividans TK24 + ermEp*-flG-FLAG₃ + pAV1-flAp-mCherry	2852	pSET152-Hyg-ermEp*-flG-FLAG₃ in ΦC31 *attB*, pAV1-flAp-mCherry in VWB *attB*	This study
S. lividans TK24 + ermEp*-flL-FLAG₃ + pAV1-flAp-mCherry	2853	pSET152-Hyg-ermEp*-flL-FLAG₃ in ΦC31 *attB*, pAV1-flAp-mCherry in VWB *attB*	This study
S. lividans TK24 + ermEp*-flF-FLAG₃ + pAV1-flAp-mCherry	2854	pSET152-Hyg-ermEp*-flF-FLAG₃ in ΦC31 *attB*, pAV1-flAp-mCherry in VWB *attB*	This study
E. coli BL21 (DE3)	1677	F– ompT gal dcm lon hsdSB(rB–mB–) [malB+]K-12(λS)	Novagen
E. coli GM272	404	dam-3 dcm-6 hsdS21 metB1 lacY1 or Z4 galK2 galT22 mtl-2 tonA2 or A31 tsx-1 or78 supE44 (thi-1)?	Yale CGSC

C

Name	Sequence
J123	cctgcatccgccagaggcgggatgcgaagaatgcgatgccgctcgccagtcg
J125	Cccttattcgcacctggcggtgctcaacgggaatcctgctctgcgaggctggcggg
J185	TCCGTGGTGACGTTCGCGACCTTCAATCTTGAGCAGAACGG
J186	ATCCGAACTCCGGCAGGGGACCGGACAGGATGT
J191	Aacgcgcgacgacgtaccctcgtgcggcgtgagggacg
J192	Caccgccctccgggcgggcggcgacctttcgacgaca
J193	Cggcatattcgccatggcacgccgtgtgtacaccacttggcgtacggc
J194	Gccactcctcactggtgccggggcgaagggaccgatctcg
J195	TCCCTCGGCCGGGTCGTCCAGCGCCCCCCAGTCCAGCGAA
J196	GCGCGGGGTGCCGGTGTTCGCGTGGAAGGGGGAGAGCCT
J197	Cggtaccctggttggccacgagcgggaagcgggggtcgatgatgt
J198	Aggcccaccaggaccgccagctccagcgtcgccgtcatgtca
J200	Cttcaccagcctggtggtgctctccatcgtcacctcccaggccg
J201	Tgtcgtcgaaaggtcgccgcccgcccggagggcggtg
J213F	Accgccgaactgcgctcggtcgagggacggcggctcagct
J213R	Gctccacctgtcggcactgggcgcccggctgactaagctgaccaa
J235	Agaggcgggatgcgaagaatgcgat
J236	Cggtatatccatccttttttcgca

J237	Gcgcgacgacgtaccctcgtgcggcgtgagggacggcgtgattccggggatccgtcgacctgcagt
J238	gccctccgggcgggcggcgacctttcgacgacagggcctatgtaggctggagctgcttcgaagttcctat
J248	Acaacctcgggatcgaggaaaacctgtattttcagggcatggagcgggccgagggggtcg
J249	Ctagcccgtttgatctcgagtgcggccgcaagcttctacgacgccgccggataggtcatc
J250	acaacctcgggatcgaggaaaacctgtattttcagggcgtggcggacggtgtggcgggacg
J251	Caacctcgggatcgaggaaaacctgtattttcagggcatgaccctccaccgcggcgccgc
J252	Tagcccgtttgatctcgagtgcggccgcaagcttctaggccaggtcgtagtcggtgatcg
J253	Acctcgggatcgaggaaaacctgtattttcagggcatggtgaccaccatcgcgctgtgcg
J254	ggctagcccgtttgatctcgagtgcggccgcaagctttcagccccggccggggacgggga
J256	tcaaaaaggatcttcacctagatccttttggttcatgtgcagctccatcagcaaaagggg
J257	tgcccctccaacgtcatctcgttctccgctcatgagctcaggcgccggggggcggtgtccg
J260	TTCGCCATTCAGGCTGCGCAACTGTT
J268	cggccagtgccaagcttgggctgcaggtcgactctagacgttgaaaatctccaaaaaaaaaggctccaaaagg
J271	gacgtcgcggggcagacgtcgcggggcaggcgacatgttgacgacagggcctatgtaggctggagctgcttcgaagttcc
J272	Gtgaacggggggacgccggaccccgagaccaccaggcgaaaggcggaaaccgcatgattccggggatccgtcgacctgcag
J276	ggtccttgtagtcgccgtcgtggtccttgtagtccgacgccgccggataggtcatcacga
J278	ggtccttgtagtcgccgtcgtggtccttgtagtcggccaggtcgtagtcggtgatcggta
J279	cttgtagtcgccgtcgtggtccttgtagtccgacgccgccggataggtcatcacgatcac
J280	ggttggtaggatcgtctagaacaggaggccccatatgaccctccaccgcggcgccgcgac
J286	taacaatttcacacaggaaacagctatgacatgattacgaattcgttg
J287	ccggtaccgatcaccgactacgacctggccgactacaaggaccacgacggc
J296	Gtgaattcctcctgcgatcgctcctggtgga
J298	Gcgcgcacctccggtcgtagctgatcgcgga
J307	cctcgggatcgaggaaaacctgtattttcagggcatggaagcgccccgtccccggccggc
J308	ctggctagcccgtttgatctcgagtgcggccgcatcaggcagaagacccccggccggctga
J337	cggttggtaggatcgtctagaacaggaggccccatatggaagcgccccgtccccggccgg
J338	cttgtagtcgccgtcgtggtccttgtagtcggcagaagacccccggccggctgagcggatc
J339	agccggccggggtcttctgccgactacaaggaccacgacggcgactacaaggaccacgacatcgactacaag
J340	tgccggttggtaggatcgtctagaacaggaggccccatatggtgaccaccatcgcgctgtgcgacc
J341	agtcgatgtcgtggtccttgtagtcgccgtcgtggtccttgtagtctcagccccggccggggacggggacg
J342	gtccccggccggggctgagactacaaggaccacgacggcgactacaaggaccacgacatcgac
J423	GTGACAGGAGGGTGGCGGGGCGCCGGGAGGTACCCGGCACCCCGGACCATCGCCGATCCGtgtaggctggagctgcttcgaagttcctat
J424	ACCGCGACCGCCGCCGAGGACCGGGCAACCCAGCACCACACCACCGAGTAAAGAGAGAGGTCACCattccggggatccgtcgacctgcag
J437F	AGATCCCTGAACTTCCCGGCGGACGGGTGAT
J437R	TGGCGTTGCCAAACCGTGGCCAGCCATT
J442	acgcggcctttttacggttcctggcctttttgctggctgctggtgcaaccgctggcccggc
J443	atctcaccggcgtcggcgaacgtcggggg
J444	gccgggccagcggttgcaccagcagccagcaaaaggccaggaaccgtaaaaaggccgcgt
J445	tggccggagggggggcccgcagctcttttgctccgcgacgcgccctgtagcggcgcattaa
J446	ttaatgcgccgctacagggcgcgtcgcggagcaaaagagctgcgggcccccctccggcca

103

J447	ttcggcggcgagcacgtgggccagt
J448	gacgctggacggcgtgctgccgttcga
J449	atccggctcgccgtcgtcgacgacaacgt
J560	**atggccatgttgtcctcctcgcccttggagaccatagggctgggctcctcccatggcggt**
J561	**gacactcgttcgttcacacgttgcagcagagtactgaacacctctcagccgccgtgccgg**
J562	atggccatgttgtcctcctcgcccttggagaccatgaacacctctcagccgccgtgccgg
J563	gacactcgttcgttcacacgttgcagcagagtactagggctgggctcctcccatggcggt
J564	atggccatgttgtcctcctcgcccttggagaccatgtgaattcctcctgcgatcgctcct
J565	gacactcgttcgttcacacgttgcagcagagtactgcgcgcacctccggtcgtagctgat
J610	tgtcctcctcgcccttggagaccatgcacagcgcgatggtggtcaccatg
J611	gggcacaatcgtgccggttggtaggatcgtctagcgtgtgtacaccacttggcgtacggc
J612	atggtgaccaccatcgcgctgtgcatggtctccaagggcgaggaggacaa
J613	Caagcttgggctgcaggtcgactctagagcaaaaaaccccctcaagacccgttttagaggcc
J614	atgacactcgttcgttcacacgttgcagcagagtactcgggcacctcccgaaggccgggc
J615	catgttgtcctcctcgcccttggagaccatttcatgctcggccagcgtgatttccttgcg
J616	Tcgttcgttcacacgttgcagcagagtactttcatgctcggccagcgtgatttccttgcg

D

J259	ttcgccattcaggctgcgcaactgttgggaagggcgatcggctcgttccccgatttccgtgcccgcatgcgcgcacctccggtcg tagctgatcgcggatccgcgcaatctaacaggcggcggctgcgggcgacagcgccggaattcgcaacctgtgccggggaaccatg ccgactacacccttcggtgcagggttggtgtgtgcccttgggccgatttcccttgtcgttcccggcatttatggtcgatcgggtctg atccaccaggagcgatcgcaggaggaattcacatggatatcgcgcgcggccgcgggatcctctagagactacaaggaccacgacgg cgactacaaggaccacgacatcgactacaaggacgacgacgacaagtgataaaccgatacaattaaaggctcctttttggagcctt ttttttttggagattttcaacgaattcgtaatcatgtcatagctgtttcctgtgtgaaattgtta

Table S1.4. (A) Plasmid constructs, (B) oligonucleotides and (C) gBlocks not described in Materials & Methods sections.

A

Name	No.	Primers	Restriction sites	Backbone	PCR template	Description
pET16b-p0565	2376	J415/J416	NdeI/BamHI	Pet16b-IMDH	*S. cattleya* gDNA	For expression of SCAT_p0565; not tested.
pSET152-ermEp*-p0564KKAA	2855	J438/J119; J439/J118	NdeI/BamHI	pSET152-ermEp*	*S. cattleya* gDNA	Partially defective mutant of p0564
PSV272.1-YeaK	2856	J496/7	SfoI/HindIII	PSV272.1	*E. coli* gDNA	*E. coli* tRNA editing protein YeaK
PSV272.1-FlG-DBD	2381	J398/J399	SfoI/HindIII	PSV272.1	*S. cattleya* gDNA	Helix-turn-helix domain of FlG
PSV272.1-FlF-DBD	2382	J396/J397	SfoI/HindIII	PSV272.1	*S. cattleya* gDNA	Helix-turn-helix domain of FlF
PSV272.1-FlL-DBD	2380	J394/J395	SfoI/HindIII	PSV272.1	*S. cattleya* gDNA	Helix-turn-helix domain of FlL
pET16b-PRE-Ado1	2857	J499/J500	NdeI/BamHI	pET16b-IMDH	*E. coli* gDNA	*E. coli* adenosine kinase; active/good expression

B

Name	Sequence
J496	ctcgggatcgaggaaaacctgtattttcagggcATGACTGAAATGGCTAAAGGAAGCGTG
J497	gctagcccgtttgatctcgagtgcggccgcaagcttTTACGCAGTGCGGCGGAAATTGACC
J499	gcggccatCTAGAAGTGCTTTTTCAGGGCcCgcatATGACCGCACCATTGGTAGTATTGG
J500	tcagcttcctttcgggctttgttagcagccggatcCTATTTAGAGTAAGATATTTTTTCGGAAGGGTAAG
J570	attcgcggccgcataatacgactcactatagccccaatagctcagtcggcagagcgtctc
J571	mUmGgagccccaatacggaatcgaaccgtagaccttctccttaccatggagacgctctgccg
J394	cctcgggatcgaggaaaacctgtattttcagggcgtccccgggctgaccggccgcgaact
J395	cccgtttgatctcgagtgcggccgcaagcttTCAggcgcaggcgatggcctgggagcggt
J396	acctcgggatcgaggaaaacctgtattttcagggcgtgcgccggctgcgggagttccgcg
J397	tagcccgtttgatctcgagtgcggccgcaagcttTCAcagatcgcccagcggcagccgca
J398	cctcgggatcgaggaaaacctgtattttcagggcccgcacctgctgaccgcccgcgaact
J399	ctagcccgtttgatctcgagtgcggccgcaagcttTCAcgccttgcggacggcctcggtg
J438	gtcatggtcaagatcggagcggcggtcaccaaatacgtgctggcc
J439	ggccagcacgtatttggtgaccgccgctccgatcttgaccatgac
J415	ataccatgggccatcatcatcatcatcatcatcatcacCTAGAAGTGCTTTTTCAGGGCGCCATGTCAGCGCAGCGGTCACACGTAC
J416	aactcagcttcctttcgggctttgttagcagccggatccTCATGCCACGTCCTTCTTCCGGGTGA

C

Name	Sequence
J258	ttcgccattcaggctgcgcaactgttgggaagggcgatcggctcgttccccgatttccgtgcccgcatgcgcgcacctcc ggtcgtagctgatcgcggatccgcgcaatctaacaggcggcggctgcgggcgacagcgccggaattcgcaacctgtgccg gggaaccatgccgactacacccttcggtgcagggttggtgtgtgcccttgggccgatttcccttgtcgttcccggcatttat ggtcgatcgggtctgatccaccaggagcgatcgcaggaggaattcacatggactacaaggaccacgacggcgactacaag gaccacgacatcgactacaaggacgacgacgacaaggatatcgcgcgcggccgcggatcctctagatgataaaccgatac aattaaaggctcctttttggagcctttttttttggagattttcaacgaattcgtaatcatgtcatagctgtttcctgtgtg aaattgtta

Appendix 2: *Supplemental proteomics data*

Table 1. Relative fluorothreonine levels of proteins from *S. cattleya* Δp0564. Positive $\log_2FC$ values indicate proteins with a higher-than-expected level of FThr incorporation. Negative values indicate a lower-than-expected level of incorporation.

ID	Log$_2$FC	pvalue	adj.pvalue	Description
embl-cds:CCB72001	4.4	0.035	0.18	putative thiol peroxidase
embl-cds:CCB72272	2.3	0.11	0.29	putative D-alanyl-D-alanine carboxypeptidase
embl-cds:CCB76874	2.3	0.12	0.3	transcriptional terminator Rho (fragment)
embl-cds:CCB77194	1.9	0.057	0.23	Beta-N-acetylhexosaminidase
embl-cds:CCB73659	1.8	0.013	0.14	putative membrane protease subunit, stomatin/prohibitin homolog
embl-cds:CCB73698	1.7	0.00046	0.057	L-alanine dehydrogenase
embl-cds:CCB73807	1.7	0.068	0.24	FeS cluster formation protein
embl-cds:CCB76653	1.5	0.36	0.51	putative succinate dehydrogenase iron-sulfur protein (Complex II)
embl-cds:CCB77531	1.5	0.0033	0.11	putative aminotransferase
embl-cds:CCB71186	1.4	0.04	0.18	protein of unknown function
embl-cds:CCB77453	1.3	0.064	0.24	Acyl-CoA synthetase family member 2, mitochondrial
embl-cds:CCB78369	1.3	0.027	0.16	3'-hydroxymethylcephem-O-carbamoyltransferase
embl-cds:CCB73801	1.2	0.46	0.61	conserved protein of unknown function
embl-cds:CCB76925	1.1	0.016	0.15	putative isobutyryl-CoA mutase, chain A
embl-cds:CCB78318	1.1	0.12	0.3	Cytochrome P450 107B1
embl-cds:CCB74577	1	0.35	0.51	acyl-CoA dehydrogenase, short-chain specific
embl-cds:CCB73553	0.9	0.0033	0.11	Oxidoreductase
embl-cds:CCB74188	0.9	0.3	0.47	O-methyltransferase mdmC
embl-cds:CCB74213	0.9	0.061	0.23	Acyl carrier protein
embl-cds:CCB75156	0.9	0.036	0.18	chaperonin large subunit
embl-cds:CCB75716	0.9	0.0088	0.13	Xaa-Pro aminopeptidase 1
embl-cds:CCB75748	0.9	0.019	0.15	conserved protein of unknown function
embl-cds:CCB77988	0.9	0.096	0.26	glutamine amidotransferase for pyridoxal phosphate synthesis
embl-cds:CCB78371	0.9	0.16	0.35	N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase
embl-cds:CCB77171	0.8	0.0015	0.094	Aldehyde dehydrogenase
embl-cds:CCB77760	0.8	0.26	0.44	conserved protein of unknown function
embl-cds:CCB71397	0.7	0.037	0.18	putative amidotransferase
embl-cds:CCB71592	0.7	0.14	0.33	conserved protein of unknown function
embl-cds:CCB72069	0.7	0.061	0.23	conserved protein of unknown function
embl-cds:CCB76262	0.7	0.079	0.25	DNA-directed RNA polymerase subunit beta
embl-cds:CCB76489	0.7	0.096	0.26	putative adenosine deaminase 5
embl-cds:CCB76604	0.7	0.16	0.35	putative methyltransferase
embl-cds:CCB76610	0.7	0.029	0.16	putative fructose 1,6-bisphosphatase class II
embl-cds:CCB77173	0.7	0.065	0.24	conserved protein of unknown function
embl-cds:CCB71860	0.6	0.3	0.47	putative oxidoreductase
embl-cds:CCB71908	0.6	0.15	0.34	Ferredoxin-2
embl-cds:CCB72073	0.6	0.27	0.44	conserved protein of unknown function
embl-cds:CCB72473	0.6	0.11	0.3	Dihydrodipicolinate synthase

embl-cds:CCB74007	0.6	0.0091	0.13	conserved protein of unknown function
embl-cds:CCB74113	0.6	0.036	0.18	UfaA2 protein
embl-cds:CCB75168	0.6	0.34	0.49	putative N-acetylglucosamine-6-phosphate deacetylase
embl-cds:CCB72323	0.5	0.31	0.47	biotin synthase
embl-cds:CCB73111	0.5	0.052	0.22	conserved protein of unknown function
embl-cds:CCB73675	0.5	0.088	0.26	putative aldehyde dehydrogenase
embl-cds:CCB75834	0.5	0.22	0.4	Thioredoxin reductase
embl-cds:CCB72149	0.4	0.72	0.79	18 kDa antigen 2
embl-cds:CCB73838	0.4	0.11	0.3	Tellurium resistance protein terZ
embl-cds:CCB73850	0.4	0.22	0.41	30S ribosomal protein S1
embl-cds:CCB73923	0.4	0.26	0.44	UDP-N-acetylmuramoylalanine--D-glutamate ligase
embl-cds:CCB74160	0.4	0.27	0.44	DNA-binding protein
embl-cds:CCB75751	0.4	0.12	0.31	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex
embl-cds:CCB76632	0.4	0.46	0.61	conserved protein of unknown function
embl-cds:CCB77049	0.4	0.07	0.24	putative delta-1-pyrroline-5-carboxylate dehydrogenase
embl-cds:CCB77555	0.4	0.2	0.39	conserved protein of unknown function
embl-cds:CCB77791	0.4	0.29	0.46	putative fatty acid oxidation complex alpha-subunit
embl-cds:CCB77865	0.4	0.091	0.26	conserved protein of unknown function
embl-cds:CCB72967	0.3	0.0051	0.12	conserved exported protein of unknown function
embl-cds:CCB74298	0.3	0.53	0.67	Extracellular metalloprotease
embl-cds:CCB74582	0.3	0.27	0.44	Oligoribonuclease
embl-cds:CCB74741	0.3	0.5	0.64	Adenosylhomocysteinase
embl-cds:CCB76631	0.3	0.2	0.39	Aminoglycoside phosphotransferase
embl-cds:CCB76857	0.3	0.15	0.33	Adenosylhomocysteine hydrolase
embl-cds:CCB77092	0.3	0.31	0.47	conserved protein of unknown function
embl-cds:CCB77214	0.3	0.2	0.39	Regulatory protein
embl-cds:CCB77233	0.3	0.42	0.57	multifunctional S
embl-cds:CCB77532	0.3	0.17	0.36	D-hydantoinase
embl-cds:CCB77884	0.3	0.029	0.16	Betaine aldehyde dehydrogenase
embl-cds:CCB71707	0.2	0.72	0.79	protein of unknown function
embl-cds:CCB73420	0.2	0.5	0.64	3-oxoacyl-[acyl-carrier-protein] reductase 1
embl-cds:CCB73502	0.2	0.66	0.76	PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit dhaK
embl-cds:CCB73674	0.2	0.49	0.64	conserved protein of unknown function
embl-cds:CCB74112	0.2	0.16	0.35	putative zinc-binding dehydrogenase
embl-cds:CCB74676	0.2	0.3	0.47	4-hydroxyphenylpyruvate dioxygenase
embl-cds:CCB75548	0.2	0.33	0.49	Protein arsC
embl-cds:CCB76011	0.2	0.2	0.39	Protein grpE
embl-cds:CCB76186	0.2	0.072	0.24	NADH-quinone oxidoreductase subunit G
embl-cds:CCB76654	0.2	0.13	0.31	Succinate dehydrogenase flavoprotein subunit
embl-cds:CCB76674	0.2	0.13	0.31	putative peptide ABC transporter solute-binding protein
embl-cds:CCB76816	0.2	0.55	0.68	conserved protein of unknown function
embl-cds:CCB77212	0.2	0.61	0.73	Hydrolase of the metallo-beta-lactamase superfamily
embl-cds:CCB77423	0.2	0.42	0.57	Protein aidB

embl-cds:CCB77621	0.2	0.63	0.74	Nucleotide sugar-1-phosphate transferase
embl-cds:CCB78036	0.2	0.54	0.67	S-adenosylmethionine synthetase
embl-cds:CCB78368	0.2	0.8	0.85	Deacetoxycephalosporin C hydroxylase
embl-cds:CCB71862	0.1	0.77	0.83	Oxygenase
embl-cds:CCB73453	0.1	0.73	0.8	putative mannose-1-phosphate guanyltransferase
embl-cds:CCB73911	0.1	0.73	0.8	putative DNA-binding protein
embl-cds:CCB74462	0.1	0.58	0.7	Trigger factor
embl-cds:CCB74718	0.1	0.5	0.64	putative multiple sugar ABC transporter solute-binding protein
embl-cds:CCB76165	0.1	0.56	0.68	cold-shock protein
embl-cds:CCB76263	0.1	0.62	0.74	RNA polymerase (beta' subunit)
embl-cds:CCB76311	0.1	0.76	0.83	initiation factor IF-I
embl-cds:CCB76315	0.1	0.65	0.75	RNA polymerase (alpha subunit)
embl-cds:CCB76322	0.1	0.72	0.79	30S ribosomal protein S9
embl-cds:CCB76567	0.1	0.67	0.77	conserved protein of unknown function
embl-cds:CCB76815	0.1	0.82	0.85	conserved protein of unknown function
embl-cds:CCB77858	0.1	0.82	0.85	conserved protein of unknown function
embl-cds:CCB73758	0	0.9	0.91	putative transcriptional regulator
embl-cds:CCB75653	0	0.86	0.88	DNA binding protein
embl-cds:CCB76257	0	0.95	0.96	ribosomal protein L11 (BL11)
embl-cds:CCB76309	0	0.99	0.99	adenylate kinase
embl-cds:CCB76861	0	0.7	0.79	5'-fluoro-5'-deoxyadenosine synthase (fluorinase)
embl-cds:CCB76862	0	0.87	0.88	5'-fluoro-5'-deoxy-adenosine phosphorylase (PNPase)
embl-cds:CCB77813	0	0.86	0.88	conserved protein of unknown function
embl-cds:CCB78391	0	0.8	0.85	Short-chain alcohol dehydrogenase
embl-cds:CCB72072	-0.1	0.7	0.79	ArpA protein
embl-cds:CCB73414	-0.1	0.87	0.88	Mini-circle uncharacterized 19.1 kDa protein
embl-cds:CCB73482	-0.1	0.88	0.89	conserved protein of unknown function
embl-cds:CCB74150	-0.1	0.78	0.84	conserved protein of unknown function
embl-cds:CCB74268	-0.1	0.79	0.84	conserved protein of unknown function
embl-cds:CCB74799	-0.1	0.82	0.85	urocanase
embl-cds:CCB75075	-0.1	0.82	0.85	OsmC-like protein
embl-cds:CCB75190	-0.1	0.83	0.86	multiple sugar-binding transporter ATP-binding protein
embl-cds:CCB75769	-0.1	0.58	0.7	Dehydrogenase
embl-cds:CCB76345	-0.1	0.43	0.58	chaperonin large subunit
embl-cds:CCB76909	-0.1	0.7	0.79	Cellulose-binding protein
embl-cds:CCB77885	-0.1	0.81	0.85	3-oxoacyl-[acyl-carrier-protein] reductase
embl-cds:CCB77983	-0.1	0.71	0.79	Elongation factor G-like protein
embl-cds:CCB71405	-0.2	0.093	0.26	putative phage tail sheath protein
embl-cds:CCB71588	-0.2	0.39	0.54	4-fluorothreonine transaldolase
embl-cds:CCB71733	-0.2	0.68	0.78	conserved protein of unknown function
embl-cds:CCB72047	-0.2	0.42	0.57	Alpha-mannosidase
embl-cds:CCB74469	-0.2	0.31	0.47	Ribose-5-phosphate isomerase B
embl-cds:CCB74747	-0.2	0.02	0.15	Phosphomannomutase

embl-cds:CCB75578	-0.2	0.54	0.67	Acetyl-coenzyme A synthetase
embl-cds:CCB75911	-0.2	0.57	0.7	conserved exported protein of unknown function
embl-cds:CCB76293	-0.2	0.71	0.79	ribosomal protein L22 (BL17)
embl-cds:CCB76687	-0.2	0.41	0.57	Ferredoxin
embl-cds:CCB76819	-0.2	0.2	0.39	ATP/GTP binding protein
embl-cds:CCB77061	-0.2	0.26	0.44	Citrate lyase beta subunit
embl-cds:CCB77306	-0.2	0.66	0.76	Phosphocarrier protein HPr
embl-cds:CCB77533	-0.2	0.29	0.46	N5,N10-methylene
embl-cds:CCB78000	-0.2	0.48	0.63	conserved protein of unknown function
embl-cds:CCB71452	-0.3	0.17	0.36	Catalase-peroxidase
embl-cds:CCB72320	-0.3	0.31	0.47	Thioesterase superfamily protein
embl-cds:CCB73177	-0.3	0.26	0.44	Amidohydrolase 2
embl-cds:CCB73628	-0.3	0.5	0.64	Quinone oxidoreductase-like protein 2 homolog
embl-cds:CCB74199	-0.3	0.48	0.63	putative Peroxiredoxin
embl-cds:CCB74201	-0.3	0.21	0.39	Pyruvate dehydrogenase E1 component
embl-cds:CCB74457	-0.3	0.74	0.81	valyl-tRNA synthetase
embl-cds:CCB74766	-0.3	0.18	0.37	Carboxyvinyl-carboxyphosphonate phosphorylmutase
embl-cds:CCB76266	-0.3	0.44	0.59	elongation factor G
embl-cds:CCB76307	-0.3	0.23	0.43	ribosomal protein L15
embl-cds:CCB76697	-0.3	0.24	0.43	putative enoyl-CoA hydratase
embl-cds:CCB76891	-0.3	0.035	0.18	ATP synthase (subunit beta, component F1)
embl-cds:CCB76961	-0.3	0.083	0.26	putative enoyl-CoA hydratase echA8
embl-cds:CCB77035	-0.3	0.53	0.66	acetohydroxy-acid isomeroreductase
embl-cds:CCB77148	-0.3	0.33	0.49	ribosome recycling factor
embl-cds:CCB77344	-0.3	0.41	0.57	putative hydrolase
embl-cds:CCB77481	-0.3	0.36	0.51	putative glutathione peroxidase
embl-cds:CCB77530	-0.3	0.063	0.24	N-carbamoylputrescine amidase
embl-cds:CCB77883	-0.3	0.54	0.67	putative glutamine synthetase
embl-cds:CCB73862	-0.4	0.64	0.75	Nucleotidase
embl-cds:CCB74023	-0.4	0.27	0.44	branched-chain alpha-keto acid dehydrogenase E3 subunit (dihydrolipoamide dehydrogenase)
embl-cds:CCB75045	-0.4	0.4	0.55	DNA-binding protein
embl-cds:CCB75259	-0.4	0.63	0.75	putative thiosulfate sulfurtransferase
embl-cds:CCB75804	-0.4	0.65	0.75	putative nosiheptide resistance regulatory protein
embl-cds:CCB76010	-0.4	0.4	0.55	molecular chaperone
embl-cds:CCB76267	-0.4	0.074	0.24	elongation factor Tu
embl-cds:CCB76399	-0.4	0.25	0.44	succinyl-CoA synthetase (alpha subunit)
embl-cds:CCB76448	-0.4	0.028	0.16	ornithine aminotransferase
embl-cds:CCB76889	-0.4	0.33	0.49	ATP synthase (subunit alpha, component F1)
embl-cds:CCB78023	-0.4	0.21	0.39	putative transcriptional regulator
embl-cds:CCB73264	-0.5	0.35	0.51	conserved protein of unknown function
embl-cds:CCB73825	-0.5	0.19	0.38	glyceraldehyde-3-phosphate dehydrogenase
embl-cds:CCB74675	-0.5	0.26	0.44	Glycolate oxidase subunit glcD
embl-cds:CCB75205	-0.5	0.089	0.26	putative lipoprotein
embl-cds:CCB76258	-0.5	0.031	0.17	ribosomal protein L1 (BL1)

embl-cds:CCB76288	-0.5	0.039	0.18	ribosomal protein L3 (BL3)
embl-cds:CCB76406	-0.5	0.31	0.47	Bifunctional protein folD [Includes: Methylenetetrahydrofolate dehydrogenase ; Methenyltetrahydrofolate cyclohydrolase]
embl-cds:CCB77168	-0.5	0.61	0.73	4-aminobutyrate aminotransferase
embl-cds:CCB71406	-0.6	0.069	0.24	conserved protein of unknown function
embl-cds:CCB71666	-0.6	0.021	0.15	Superoxide dismutase [Fe-Zn] 1
embl-cds:CCB71865	-0.6	0.011	0.13	Hydrolase
embl-cds:CCB72474	-0.6	0.31	0.47	conserved protein of unknown function
embl-cds:CCB73714	-0.6	0.042	0.18	putative UPF0098 class protein
embl-cds:CCB73816	-0.6	0.23	0.42	Transketolase
embl-cds:CCB74474	-0.6	0.0034	0.11	conserved protein of unknown function
embl-cds:CCB75372	-0.6	0.26	0.44	protein of unknown function
embl-cds:CCB77215	-0.6	0.3	0.47	Sensor protein
embl-cds:CCB77595	-0.6	0.0094	0.13	exported protein of unknown function
embl-cds:CCB77861	-0.6	0.084	0.26	20S proteasome beta-subunit
embl-cds:CCB78370	-0.6	0.29	0.46	Isopenicillin N synthetase
embl-cds:CCB78452	-0.6	0.13	0.32	3-demethylubiquinone-9 3-methyltransferase
embl-cds:CCB72210	-0.7	0.069	0.24	Non-hemolytic phospholipase C
embl-cds:CCB72264	-0.7	0.15	0.34	Glycogen operon protein glgX homolog
embl-cds:CCB74224	-0.7	0.041	0.18	organic hydroperoxide resistance reductase B
embl-cds:CCB75365	-0.7	0.15	0.34	conserved protein of unknown function
embl-cds:CCB75802	-0.7	0.01	0.13	Peptidyl-prolyl cis-trans isomerase B
embl-cds:CCB75849	-0.7	0.26	0.44	30S ribosomal protein S6
embl-cds:CCB76025	-0.7	0.022	0.15	Fructose-bisphosphate aldolase
embl-cds:CCB76291	-0.7	0.072	0.24	ribosomal protein L2 (BL2)
embl-cds:CCB76316	-0.7	0.043	0.18	50S ribosomal protein L17
embl-cds:CCB76519	-0.7	0.16	0.35	putative enoyl-CoA dehydratase
embl-cds:CCB71775	-0.8	0.085	0.26	conserved protein of unknown function
embl-cds:CCB73009	-0.8	0.087	0.26	putative secreted protein
embl-cds:CCB75307	-0.8	0.045	0.19	Dihydroxy-acid dehydratase
embl-cds:CCB76913	-0.8	0.26	0.44	degradative acetoacetyl-CoA thiolase
embl-cds:CCB77525	-0.8	0.067	0.24	putative gamma-glutamyltranspeptidase
embl-cds:CCB78341	-0.8	0.14	0.33	putative ThiJ/PfpI family protein
embl-cds:CCB71078	-0.9	0.091	0.26	NAD-dependent epimerase/dehydratase
embl-cds:CCB72071	-0.9	0.12	0.31	protein of unknown function
embl-cds:CCB72162	-0.9	0.2	0.39	putative polyketide oxygenase/hydroxylase
embl-cds:CCB73408	-0.9	0.35	0.5	beta-ketoadipyl CoA thiolase
embl-cds:CCB73547	-0.9	0.021	0.15	electron transfer flavoprotein (alpha subunit)
embl-cds:CCB73649	-0.9	0.013	0.14	Homogentisate 1,2-dioxygenase
embl-cds:CCB74824	-0.9	0.02	0.15	enolase
embl-cds:CCB75589	-0.9	0.26	0.44	conserved protein of unknown function
embl-cds:CCB76477	-0.9	0.26	0.44	Lipoprotein
embl-cds:CCB77159	-0.9	0.17	0.36	putative adenosine deaminase 1
embl-cds:CCB77204	-0.9	0.056	0.23	polynucleotide phosphorylase (PNPase)

embl-cds:CCB74164	-1	0.15	0.33	putative cytochrome P450 125
embl-cds:CCB76820	-1	0.13	0.31	2-oxoglutarate decarboxylase
embl-cds:CCB77051	-1	0.19	0.38	putative branched-chain-amino-acid aminotransferase
embl-cds:CCB77457	-1	0.091	0.26	Acetyl-coenzyme A synthetase
embl-cds:CCB77809	-1	0.5	0.64	conserved protein of unknown function
embl-cds:CCB77898	-1	0.026	0.16	Phenylalanyl-tRNA synthetase beta chain
embl-cds:CCB73477	-1.1	0.026	0.16	conserved exported protein of unknown function
embl-cds:CCB74477	-1.1	0.023	0.15	Aminopeptidase N
embl-cds:CCB76410	-1.1	0.072	0.24	Malate dehydrogenase
embl-cds:CCB77359	-1.1	0.0014	0.094	aconitate hydratase (aconitase)
embl-cds:CCB77913	-1.1	0.17	0.36	glyceraldehyde-3-phosphate dehydrogenase
embl-cds:CCB71752	-1.2	0.00011	0.026	Protein korA
embl-cds:CCB72125	-1.2	0.073	0.24	Isocitrate dehydrogenase [NADP]
embl-cds:CCB72579	-1.2	0.042	0.18	YCII-related
embl-cds:CCB72036	-1.3	0.13	0.31	putative peptidase inhibitor
embl-cds:CCB72139	-1.3	0.011	0.13	conserved exported protein of unknown function
embl-cds:CCB75954	-1.3	0.021	0.15	exported protein of unknown function
embl-cds:CCB71784	-1.4	0.016	0.15	putative hydrolase
embl-cds:CCB73823	-1.4	0.011	0.13	triose phosphate isomerase
embl-cds:CCB74454	-1.4	0.019	0.15	nucleoside diphosphate kinase
embl-cds:CCB75819	-1.4	0.0065	0.12	conserved protein of unknown function
embl-cds:CCB72337	-1.5	0.26	0.44	conserved exported protein of unknown function
embl-cds:CCB73800	-1.5	0.2	0.39	2,3,4,5-tetrahydropyridine-2
embl-cds:CCB77145	-1.5	0.3	0.47	ribosomal protein S2
embl-cds:CCB72250	-1.6	0.015	0.15	Nocardamine synthetase
embl-cds:CCB75567	-1.6	0.0048	0.12	Anti-sigma-B factor antagonist (fragment)
embl-cds:CCB76511	-1.6	0.1	0.28	Acetyl-/propionyl-coenzyme A carboxylase alpha chain
embl-cds:CCB76855	-1.6	0.0057	0.12	Fluoroacetyl coenzyme A thioesterase
embl-cds:CCB77845	-1.6	0.058	0.23	Solute binding transport lipoprotein
embl-cds:CCB74055	-1.7	0.026	0.16	Extracellular solute-binding protein family 1
embl-cds:CCB71663	-1.8	0.0029	0.11	exported protein of unknown function
embl-cds:CCB77005	-1.8	0.019	0.15	putative metallopeptidase, secreted
embl-cds:CCB78006	-2	0.092	0.26	conserved protein of unknown function
embl-cds:CCB74619	-2.3	0.023	0.15	putative metallopeptidase, secreted
embl-cds:CCB77364	-2.3	0.04	0.18	Solute-binding protein
embl-cds:CCB71463	-2.4	0.13	0.32	conserved protein of unknown function
embl-cds:CCB75283	-2.7	0.011	0.13	Phosphate-binding protein pstS 3
embl-cds:CCB75991	-2.8	0.0063	0.12	putative Transcriptional regulator, TetR family
embl-cds:CCB76720	-3.1	0.028	0.16	ATP-binding protein

Table 2. Settings used for dMRM acquisition of targeted proteomics data. In every case, the fragmentor voltage was set to 130 V, while the retention time window was 2.5 minutes. Quantifier transitions are marked in bold.

Protein	Peptide	Precursor Ion	Product Ion	Ret Time (min)	Collision Energy
FaDH	VLTGGER	366.2	418.2	9.7	12.4
FaDH	**VLTGGER**	**366.2**	**519.3**	**9.7**	**12.4**
FaDH	VLTGGER	366.2	632.3	9.7	12.4
FaDH	VLT[+18.0]GGER	375.2	418.2	9.8	12.6
FaDH	**VLT[+18.0]GGER**	**375.2**	**537.2**	**9.8**	**12.6**
FaDH	VLT[+18.0]GGER	375.2	650.3	9.8	12.6
FaDH	DLSTAYR	413.2	409.2	17.1	13.8
FaDH	DLSTAYR	413.2	417.2	17.1	13.8
FaDH	DLSTAYR	413.2	488.2	17.1	13.8
FaDH	DLSTAYR	413.2	510.3	17.1	13.8
FaDH	**DLSTAYR**	**413.2**	**597.3**	**17.1**	**13.8**
FaDH	DLST[+18.0]AYR	422.2	409.2	17.3	14.1
FaDH	DLST[+18.0]AYR	422.2	435.2	17.3	14.1
FaDH	DLST[+18.0]AYR	422.2	506.2	17.3	14.1
FaDH	DLST[+18.0]AYR	422.2	528.3	17.3	14.1
FaDH	DLST[+18.0]AYR	422.2	615.3	17.3	14.1
FaDH	**DLST[+18.0]AYR**	**422.2**	**728.4**	**17.3**	**14.1**
FaDH	ILSYIDIGTAEGAK	725.9	275.2	33.2	23.5
FaDH	ILSYIDIGTAEGAK	725.9	746.4	33.2	23.5
FaDH	**ILSYIDIGTAEGAK**	**725.9**	**861.4**	**33.2**	**23.5**
FaDH	ILSYIDIGTAEGAK	725.9	974.5	33.2	23.5
FaDH	ILSYIDIGT[+18.0]AEGAK	734.9	275.2	33.4	23.8
FaDH	ILSYIDIGT[+18.0]AEGAK	734.9	764.4	33.4	23.8
FaDH	**ILSYIDIGT[+18.0]AEGAK**	**734.9**	**879.4**	**33.4**	**23.8**
FaDH	ILSYIDIGT[+18.0]AEGAK	734.9	992.5	33.4	23.8
FaDH	VDLGGSLSGGYYVAPTIFEGDNR	796.4	461.2	40.8	23.9
FaDH	VDLGGSLSGGYYVAPTIFEGDNR	796.4	850.4	40.8	23.9
FaDH	VDLGGSLSGGYYVAPTIFEGDNR	796.4	951.5	40.8	23.9
FaDH	**VDLGGSLSGGYYVAPTIFEGDNR**	**796.4**	**1048.5**	**40.8**	**23.9**
FaDH	VDLGGSLSGGYYVAPT[+18.0]IFEGDNR	802.4	461.2	41.0	24.1
FaDH	VDLGGSLSGGYYVAPT[+18.0]IFEGDNR	802.4	850.4	41.0	24.1
FaDH	VDLGGSLSGGYYVAPT[+18.0]IFEGDNR	802.4	969.4	41.0	24.1
FaDH	**VDLGGSLSGGYYVAPT[+18.0]IFEGDNR**	**802.4**	**1066.5**	**41.0**	**24.1**
FaDH	IFQEEIFGPVVSVTR	861.0	757.5	40.8	27.7
FaDH	**IFQEEIFGPVVSVTR**	**861.0**	**961.5**	**40.8**	**27.7**
FaDH	IFQEEIFGPVVSVTR	861.0	1074.6	40.8	27.7

FaDH	IFQEEIFGPVVSVTR	861.0	1203.7	40.8	27.7
FaDH	IFQEEIFGPVVSVT[+18.0]R	870.0	775.4	41.1	28.0
FaDH	**IFQEEIFGPVVSVT[+18.0]R**	**870.0**	**979.5**	**41.1**	**28.0**
FaDH	IFQEEIFGPVVSVT[+18.0]R	870.0	1092.6	41.1	28.0
FaDH	IFQEEIFGPVVSVT[+18.0]R	870.0	1221.7	41.1	28.0
flA	**FFPEGTVFATTTYPATGTTTR**	**756.0**	**804.4**	**38.3**	**22.4**
flA	FFPEGTVFATTTYPATGTTTR	756.0	967.5	38.3	22.4
flA	FFPEGTVFATTTYPATGTTTR	756.0	1068.5	38.3	22.4
flA	FFPEGT[+18.0]VFATTTYPATGTTTR	762.0	804.4	38.5	22.6
flA	**FFPEGTVFATTTYPAT[+18.0]GTTTR**	**762.0**	**822.4**	**38.5**	**22.6**
flA	FFPEGTVFATTTYPATGT[+18.0]TTR	762.0	822.4	38.5	22.6
flA	FFPEGT[+18.0]VFATTTYPATGTTTR	762.0	967.5	38.5	22.6
flA	FFPEGTVFATTTYPAT[+18.0]GTTTR	762.0	985.5	38.5	22.6
flA	FFPEGTVFATTTYPATGT[+18.0]TTR	762.0	985.5	38.5	22.6
flA	FFPEGT[+18.0]VFATTTYPATGTTTR	762.0	1068.5	38.5	22.6
flA	FFPEGTVFATTTYPAT[+18.0]GTTTR	762.0	1086.5	38.5	22.6
flA	FFPEGTVFATTTYPATGT[+18.0]TTR	762.0	1086.5	38.5	22.6
flA	VIPEQPEPTFYSR	781.9	770.4	28.0	25.2
flA	VIPEQPEPTFYSR	781.9	899.4	28.0	25.2
flA	**VIPEQPEPTFYSR**	**781.9**	**996.5**	**28.0**	**25.2**
flA	VIPEQPEPTFYSR	781.9	1124.5	28.0	25.2
flA	VIPEQPEPTFYSR	781.9	1350.6	28.0	25.2
flA	VIPEQPEPT[+18.0]FYSR	790.9	788.4	28.2	25.5
flA	**VIPEQPEPT[+18.0]FYSR**	**790.9**	**1014.5**	**28.2**	**25.5**
flA	VIPEQPEPT[+18.0]FYSR	790.9	1142.5	28.2	25.5
flA	VIPEQPEPT[+18.0]FYSR	790.9	1368.6	28.2	25.5
GapA	**ATALVIPELK**	**527.8**	**486.3**	**35.1**	**17.4**
GapA	ATALVIPELK	527.8	599.4	35.1	17.4
GapA	ATALVIPELK	527.8	698.4	35.1	17.4
GapA	ATALVIPELK	527.8	811.5	35.1	17.4
GapA	**AT[+18.0]ALVIPELK**	**536.8**	**486.3**	**35.4**	**17.6**
GapA	AT[+18.0]ALVIPELK	536.8	599.4	35.4	17.6
GapA	AT[+18.0]ALVIPELK	536.8	698.4	35.4	17.6
GapA	AT[+18.0]ALVIPELK	536.8	811.5	35.4	17.6
GapA	LVDLTTFVGGR	589.3	636.3	37.0	19.3
GapA	**LVDLTTFVGGR**	**589.3**	**737.4**	**37.0**	**19.3**
GapA	LVDLTTFVGGR	589.3	850.5	37.0	19.3
GapA	LVDLT[+18.0]TFVGGR	598.3	636.3	36.6	19.5
GapA	LVDLTT[+18.0]FVGGR	598.3	654.3	36.7	19.5
GapA	**LVDLT[+18.0]TFVGGR**	**598.3**	**755.4**	**36.6**	**19.5**
GapA	LVDLT[+18.0]TFVGGR	598.3	868.5	36.6	19.5
GapA	**AAAENIIPTTTGAAK**	**714.9**	**746.4**	**25.1**	**23.2**

GapA	AAAENIIPTTTGAAK	714.9	859.5	25.1	23.2
GapA	AAAENIIPTTTGAAK	714.9	972.6	25.1	23.2
GapA	**AAAENIIPT[+18.0]TTGAAK**	**723.9**	**764.4**	**25.3**	**23.4**
GapA	AAAENIIPT[+18.0]TTGAAK	723.9	877.5	25.3	23.4
GapA	AAAENIIPT[+18.0]TTGAAK	723.9	990.6	25.3	23.4
TufA	TTLTAAITK	460.3	503.3	22.6	15.3
TufA	TTLTAAITK	460.3	604.4	22.6	15.3
TufA	**TTLTAAITK**	**460.3**	**717.5**	**22.6**	**15.3**
TufA	TT[+18.0]LTAAITK	469.3	503.3	22.5	15.5
TufA	TTLT[+18.0]AAITK	469.3	503.3	22.8	15.5
TufA	TTLTAAIT[+18.0]K	469.3	521.3	22.8	15.5
TufA	TT[+18.0]LTAAITK	469.3	604.4	22.5	15.5
TufA	TTLT[+18.0]AAITK	469.3	622.4	22.8	15.5
TufA	TT[+18.0]LTAAITK	469.3	717.5	22.5	15.5
TufA	**TTLT[+18.0]AAITK**	**469.3**	**735.4**	**22.8**	**15.5**
TufA	VNETVDIIGIK	600.8	658.4	32.1	19.6
TufA	VNETVDIIGIK	600.8	757.5	32.1	19.6
TufA	**VNETVDIIGIK**	**600.8**	**858.5**	**32.1**	**19.6**
TufA	VNET[+18.0]VDIIGIK	609.8	658.4	32.3	19.9
TufA	VNET[+18.0]VDIIGIK	609.8	757.5	32.3	19.9
TufA	**VNET[+18.0]VDIIGIK**	**609.8**	**876.5**	**32.3**	**19.9**
TufA	LLGLMHTIDEAIPTPQR	635.7	400.2	39.1	18.1
TufA	**LLGLMHTIDEAIPTPQR**	**635.7**	**598.3**	**39.1**	**18.1**
TufA	LLGLMHTIDEAIPTPQR	635.7	711.4	39.1	18.1
TufA	LLGLMHT[+18.0]IDEAIPTPQR	641.7	400.2	38.7	18.3
TufA	**LLGLMHT[+18.0]IDEAIPTPQR**	**641.7**	**598.3**	**38.7**	**18.3**
TufA	LLGLMHTIDEAIPT[+18.0]PQR	641.7	616.3	38.7	18.3
TufA	LLGLMHT[+18.0]IDEAIPTPQR	641.7	711.4	38.7	18.3
TufA	LLGLMHTIDEAIPT[+18.0]PQR	641.7	729.4	38.7	18.3

Appendix 3: *Supplementary bioinformatic data*

Table S3.1. DUF190 proteins that do not neighbor a CRCB transporter. Asterisks (*) mark sequences from known or putative biosynthetic gene clusters. The list is extensive, but non-comprehensive.

Species	NCBI accession
Actinomadura atramentaria DSM 43919	WP_019633314.1
Actinoplanes subtropicus	WP_030437505.1
Allosalinactinospora lopnorensis	WP_046470115.1
Kitasatospora griseola	WP_052509758.1*
Nocardia abscessus NBRC 100374	WP_051169193.1
Nocardia amamiensis NBRC 102102	WP_067462198.1
Nocardia araoensis NBRC 100135	WP_051021100.1
Nocardia arthritidis NBRC 100137	WP_063048675.1
Nocardia asiatica NBRC 100129	WP_051046808.1
Nocardia beijingensis NBRC 16342	WP_067798038.1
Nocardia brasiliensis IFM 10847	WP_029892847.1
Nocardia exalbida NBRC 100660	WP_051175437.1
Nocardia gamkensis NBRC 108242	WP_062969482.1
Nocardia jiangxiensis NBRC 101359	WP_051193548.1
Nocardia mexicana NBRC 108244	WP_068027754.1
Nocardia miyunensis NBRC 108239	WP_067680631.1
Nocardia nova SH22a	WP_051494972.1
Nocardia pseudobrasiliensis	WP_062509190.1
Nocardia seriolae	WP_033090615.1
Nocardia sp. BMG51109	WP_051499416.1
Nocardia terpenica	WP_067584296.1
Nocardia transvalensis NBRC 15921	WP_051163249.1
Nocardia vaccinii NBRC 15922	WP_067896794.1
Nocardia vulneris	WP_043672616.1
Rhodococcus rhodnii	WP_037275323.1
Streptomyces antioxidans	WP_075200312.1*
Streptomyces cattleya NRRL 8057	WP_014144873.1*
Streptomyces hokutonensis	WP_019071008.1
Streptomyces iranensis	WP_044582716.1*
Streptomyces reticuli	WP_063808629.1*
Streptomyces sp. HNS054	WP_048458818.1
Streptomyces sp. NRRL F-5126	WP_030911067.1
Streptomyces sp. NRRL S-31	WP_030749466.1
Streptomyces sp. NRRL S-646	WP_030927804.1
Streptomyces sp. PsTaAH-124	WP_018569625.1
Streptomyces sp. SAT1	WP_064531258.1
Streptomyces sp. Tu 6176	WP_018569625.1
Streptomyces vietnamensis	WP_041132753.1
Streptomyces violaceusniger	WP_014057726.1*
Streptomyces yerevanensis	WP_033323677.1
Streptomyces yokosukanensis	WP_067127105.1
Sulfobacillus thermosulfidooxidans	WP_053960867.1
Thermus brockianus	WP_071677180.1
Thermus islandicus DSM 21543	WP_022797651.1
Thermus sp. 2.9	WP_039457715.1
Thermus tengchongensis YIM 77401	WP_038042456.1

Appendix 4: *Supplementary genetic characterization*

Figure S4.1. Characterization of flF-L. The unanticipated deletion was further confirmed by Sanger sequencing of a PCR fragment amplified with J402/J405, using J403 as a sequencing primer. Sequencing matched the *flF-flG* intergenic region, and then to the coding sequence for *SCAT_4133*, ~30 CDS away from the organofluorine locus.
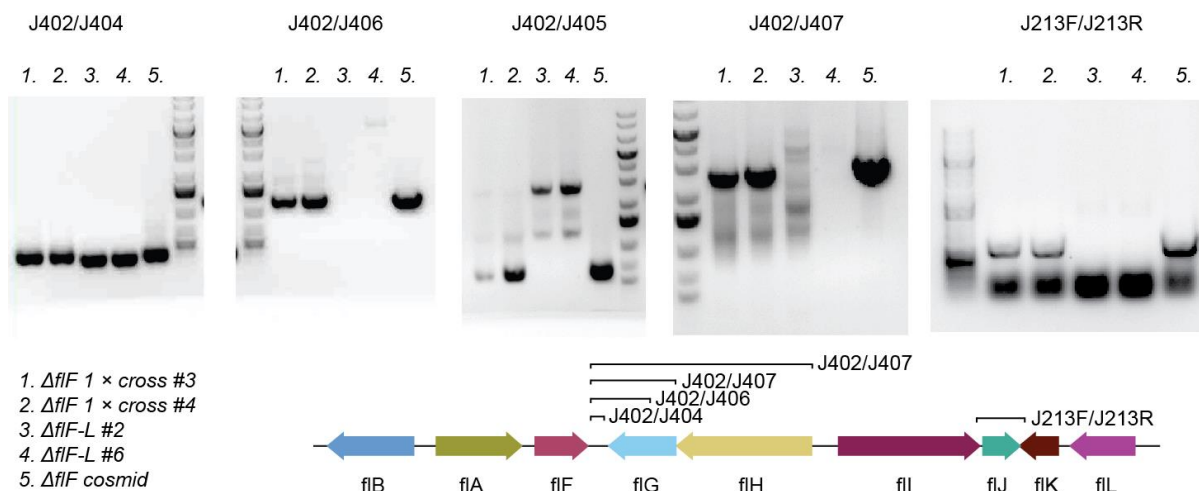
Table S4.1. *Streptomyces* strains not discussed in *Chapter 4* with notes.

Strain	No.	Notes
S. cattleya + flAp-flF-FLAG₃-fdterm	1702	Not assayed.
S. cattleya flF-FLAG	1563	Produces organofluorines. Expression too low to see by anti-FLAG western.
S. cattleya flL-FLAG	1562	Produces organofluorines. Expression too low to see by anti-FLAG western.
S. cattleya ΔflF-L + ermEp*-flF-fdterm	1829	No organofluorines.
S. cattleya ΔflF-L+ flAp-flF-FLAG₃-fdterm	1704	No organofluorines.
S. cattleya Δp0564 + ermEp*-p0564	2858	Lower ZOI for FThr in disc diffusion assay.
ΔflF12_#6+ ermEp*-flF-flagx3-fdterm	1805	No organofluorines.
S. cattleya + flAp-flG-flagx3-fdterm	1703	Not assayed.
S. cattleya + ermEp*-SCAT_0531	2581	No obvious peptide peak by LCMS.

Appendix 5: *Supplementary metabolomic data*

Table S5.1. Comparative metabolomics results reported in *Chapter 4*. The peaks reported are at least 10-fold more abundant in the WT strain as compared to the Δ*ThrAld* strain, and differ between the two strains with a p-value <0.01.

Fold change (WT/Δ*ThrAld*)	pvalue	mzmed	RT (s)	WT average	Δ*ThrAld* average
27	7.3E-05	347.6975	176	1.6E+05	5.8E+03
10	2.4E-04	438.2373	259	1.1E+05	1.1E+04
142	1.3E-04	437.9873	259	1.1E+05	7.8E+02
97	7.1E-04	818.9405	362	1.1E+05	1.1E+03
41	4.2E-04	819.4410	362	1.1E+05	2.6E+03
65	2.9E-04	842.4632	378	7.7E+04	1.2E+03
83	1.3E-03	409.9668	246	7.6E+04	9.2E+02
127	2.4E-04	842.9640	378	7.5E+04	5.9E+02
18	7.0E-05	583.6473	259	7.1E+04	3.9E+03
23	2.3E-05	583.9811	260	6.9E+04	3.0E+03
14	7.4E-04	545.9525	246	6.6E+04	4.7E+03
25	6.3E-04	819.9418	362	5.7E+04	2.2E+03
17	8.9E-05	393.6929	209	5.5E+04	3.2E+03
44	3.3E-04	438.4885	260	5.3E+04	1.2E+03
18	1.3E-04	694.3864	175	4.8E+04	2.7E+03
65	3.7E-04	843.4663	378	3.9E+04	6.0E+02
17	9.9E-03	940.7536	345	3.1E+04	1.8E+03
25	6.6E-03	941.0881	345	2.7E+04	1.1E+03
16	1.7E-03	546.2949	362	2.6E+04	1.6E+03
89	6.0E-04	546.6305	362	2.5E+04	2.9E+02
14	7.4E-03	677.3381	328	2.5E+04	1.8E+03
12	1.6E-03	735.3959	347	2.2E+04	1.8E+03
13	4.6E-03	940.4214	345	2.1E+04	1.6E+03
29	4.2E-03	886.7327	334	2.1E+04	7.2E+02
18	7.6E-04	846.9176	250	1.9E+04	1.1E+03
18	1.5E-03	735.8979	347	1.8E+04	9.9E+02
25	6.0E-03	677.8389	328	1.8E+04	7.3E+02
16	3.6E-03	886.0643	334	1.8E+04	1.2E+03
17	8.0E-05	695.3897	175	1.8E+04	1.1E+03
29	3.4E-03	941.4252	344	1.7E+04	5.9E+02
19	4.4E-04	911.9385	272	1.6E+04	8.7E+02
16	5.4E-05	940.9785	340	1.4E+04	8.7E+02
27	7.1E-05	682.0088	361	1.3E+04	5.0E+02
17	9.7E-05	912.9396	271	1.1E+04	6.1E+02

Table S5.2. Comparative metabolomics results reported in *Chapter 4*. The peaks reported are at least 10-fold more abundant in the WT strain as compared to the Δ*flF-L* strain, and differ between the two strains with a p-value <0.01.

Fold change (WT/ΔFlF-L)	pvalue	mzmed	RT (s)	WT average	ΔFlF-L average
22	7.5E-05	347.6975	175	1.6E+05	7.0E+03
71	1.4E-04	437.9873	259	1.1E+05	1.6E+03
69	7.2E-04	818.9405	362	1.1E+05	1.6E+03
32	4.3E-04	819.4410	362	1.1E+05	3.3E+03
53	2.9E-04	842.4632	378	7.7E+04	1.5E+03
31	1.2E-03	409.9667	245	7.6E+04	2.4E+03
65	2.4E-04	842.9640	378	7.5E+04	1.2E+03
15	7.1E-05	583.6473	259	7.1E+04	4.7E+03
14	1.8E-05	583.9811	259	6.9E+04	4.8E+03
12	7.9E-04	545.9525	245	6.6E+04	5.6E+03
10	7.4E-04	819.9418	362	5.7E+04	5.6E+03
15	8.8E-05	393.6928	209	5.5E+04	3.7E+03
42	3.3E-04	438.4885	259	5.3E+04	1.3E+03
13	1.3E-04	694.3864	175	4.8E+04	3.5E+03
64	3.8E-04	843.4663	378	3.9E+04	6.1E+02
17	9.9E-03	940.7536	345	3.1E+04	1.8E+03
15	7.4E-03	941.0881	345	2.7E+04	1.8E+03
19	1.6E-03	546.2949	362	2.6E+04	1.4E+03
125	6.0E-04	546.6305	362	2.5E+04	2.0E+02
13	7.5E-03	677.3381	328	2.5E+04	2.0E+03
10	5.1E-03	940.4214	345	2.1E+04	2.1E+03
21	4.4E-03	886.7327	334	2.1E+04	1.0E+03
17	1.7E-03	735.8979	347	1.8E+04	1.1E+03
11	4.2E-03	886.0643	334	1.8E+04	1.7E+03
17	6.8E-03	677.8389	328	1.8E+04	1.0E+03
11	1.3E-04	695.3897	175	1.8E+04	1.6E+03
17	3.7E-03	941.4252	344	1.7E+04	1.0E+03
17	3.4E-05	682.0088	361	1.3E+04	7.6E+02
11	4.7E-05	912.9396	271	1.1E+04	9.9E+02

Table S5.3. Comparative metabolomics results reported in *Chapter 4*. The peaks reported are at least 10-fold more abundant in the WT strain as compared to the Δ*flA* strain, and differ between the two strains with a p-value <0.01.

Fold change (WT/ΔFIA)	pvalue	mzmed	RT (s)	WT average	ΔFIA average
38	1.34E-03	466.2084	232	5.5E+05	1.5E+04
63	1.88E-04	512.2137	285	4.4E+05	7.0E+03
199	3.29E-04	512.2137	375	3.0E+05	1.5E+03
252	9.68E-04	1023.4199	375	2.3E+05	9.1E+02
234	2.53E-03	512.7139	376	1.8E+05	7.9E+02
350	4.26E-05	988.3821	411	1.8E+05	5.1E+02
27	7.26E-05	347.6975	176	1.6E+05	5.7E+03
18	1.14E-03	467.2119	233	1.5E+05	8.3E+03
106	1.34E-04	437.9873	259	1.1E+05	1.1E+03
19	1.36E-03	504.2160	287	1.1E+05	5.8E+03
70	7.21E-04	818.9405	362	1.1E+05	1.6E+03
43	4.16E-04	819.4410	362	1.1E+05	2.5E+03
97	4.30E-05	989.3852	411	1.0E+05	1.1E+03
64	5.29E-03	513.2152	375	9.5E+04	1.5E+03
57	4.21E-04	1023.4201	285	9.3E+04	1.6E+03
16	1.87E-03	409.7161	246	8.9E+04	5.5E+03
64	6.30E-04	512.7106	294	8.5E+04	1.3E+03
66	1.97E-04	1007.3967	350	7.9E+04	1.2E+03
68	2.90E-04	842.4632	378	7.7E+04	1.1E+03
58	1.29E-03	409.9668	246	7.6E+04	1.3E+03
73	2.43E-04	842.9640	378	7.5E+04	1.0E+03
196	7.49E-04	503.7027	350	7.1E+04	3.6E+02
14	7.59E-05	583.6473	259	7.1E+04	5.1E+03
14	2.23E-05	583.9811	260	6.9E+04	5.0E+03
32	6.07E-03	890.3823	413	6.9E+04	2.1E+03
458	5.55E-06	494.6951	411	6.7E+04	1.5E+02
12	7.08E-04	545.9525	246	6.6E+04	5.4E+03
57	2.13E-04	504.7180	287	6.3E+04	1.1E+03
15	6.90E-04	819.9418	362	5.7E+04	3.8E+03
87	1.41E-03	779.3392	429	5.7E+04	6.5E+02
34	3.81E-04	1024.4220	285	5.5E+04	1.6E+03
16	9.07E-05	393.6928	209	5.5E+04	3.5E+03
34	3.34E-04	438.4885	260	5.3E+04	1.6E+03
26	6.05E-04	513.2121	294	5.0E+04	1.9E+03
15	1.34E-04	694.3864	175	4.8E+04	3.1E+03
59	3.09E-03	1006.4068	387	4.6E+04	7.9E+02
178	2.22E-03	1025.4191	376	4.5E+04	2.5E+02
23	3.42E-04	504.2043	350	4.4E+04	2.0E+03
192	6.40E-04	454.7003	356	4.3E+04	2.2E+02
67	3.76E-04	843.4663	378	3.9E+04	5.8E+02
73	3.66E-05	495.1966	411	3.8E+04	5.2E+02
33	5.98E-03	891.3853	413	3.7E+04	1.1E+03

36	6.04E-05	990.3884	411	3.6E+04	9.8E+02
12	5.18E-03	613.2761	301	3.2E+04	2.6E+03
16	9.71E-03	940.7536	345	3.1E+04	1.9E+03
59	4.41E-04	1010.3643	411	2.8E+04	4.7E+02
22	6.75E-03	941.0881	345	2.7E+04	1.2E+03
16	1.73E-03	546.2949	362	2.6E+04	1.7E+03
25	2.50E-04	990.3989	358	2.6E+04	1.1E+03
137	1.43E-03	780.3425	428	2.6E+04	1.9E+02
21	3.42E-04	1008.3993	349	2.6E+04	1.2E+03
209	6.00E-04	546.6305	362	2.5E+04	1.2E+02
13	7.46E-03	677.3381	328	2.5E+04	1.9E+03
12	1.32E-03	908.3934	356	2.3E+04	1.8E+03
25	8.30E-05	455.2018	357	2.3E+04	9.2E+02
12	4.72E-03	940.4214	345	2.1E+04	1.8E+03
18	4.49E-03	886.7327	334	2.1E+04	1.2E+03
15	7.97E-04	1007.4238	286	2.0E+04	1.4E+03
Inf.	2.79E-03	445.6949	413	2.0E+04	0.0E+00
21	4.46E-04	504.2106	386	1.9E+04	9.0E+02
27	5.65E-04	1025.4264	285	1.9E+04	7.1E+02
44	1.99E-06	987.3747	411	1.9E+04	4.3E+02
11	7.60E-04	846.9176	250	1.9E+04	1.7E+03
16	1.82E-03	735.8979	347	1.8E+04	1.1E+03
10	4.21E-03	886.0643	334	1.8E+04	1.7E+03
22	6.70E-03	677.8389	328	1.8E+04	8.0E+02
18	7.40E-04	511.7091	375	1.8E+04	1.0E+03
16	9.82E-05	695.3897	175	1.8E+04	1.1E+03
20	4.72E-04	513.7131	294	1.7E+04	8.4E+02
92	4.45E-04	1011.3676	411	1.7E+04	1.8E+02
15	4.52E-04	911.9385	272	1.6E+04	1.1E+03
24	8.60E-04	504.7049	350	1.6E+04	6.4E+02
20	4.56E-05	1026.4255	375	1.5E+04	7.5E+02
11	1.65E-04	455.2007	382	1.5E+04	1.3E+03
17	1.75E-03	987.4170	354	1.3E+04	7.6E+02
136	2.12E-04	495.6989	411	1.2E+04	8.7E+01
10	7.72E-05	912.9396	271	1.1E+04	1.1E+03
26	4.99E-04	446.1972	412	1.0E+04	4.0E+02
17	6.30E-06	432.1917	333	2.2E+03	3.8E+04
15	1.07E-08	560.2352	302	1.8E+03	2.9E+04
80	3.33E-07	431.6900	333	9.8E+02	7.8E+04