Title

Self-Resistance Enzyme Directed Genome-Mining for Fungal Natural Products and Enzyme Catalyses And Cell-free In Vitro Biosynthesis of Plant Terpene Natural Products

Author

Bat-Erdene, Undramaa

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Self-Resistance Enzyme Directed Genome-Mining for Fungal Natural Products and Enzyme Catalyses And Cell-free *In Vitro* Biosynthesis of Plant Terpene Natural Products

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Chemical and Biomolecular Engineering

by

Undramaa Bat-Erdene

2021

ABSTRACT OF THE DISSERTATION

Self-Resistance Enzyme Directed Genome-Mining for

Fungal Natural Products and Enzyme Catalyses

And

One-pot *in vitro* Biosynthesis of Plant Terpene Natural Products

by

Undramaa Bat-Erdene

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2021

Professor Yi Tang, Chair

With the advances in modern science and technology, humankind leads longer and more comfortable lives than ever before. However, such progress has also given rise to worldwide challenges, one of which is the emergence of increasingly fatal drug-resistances. Since drug and herbicide resistance is inevitable due to overexposure, our need to understand and resolve this issue is an ongoing battle. In parallel, our knowledge of natural products - small molecules derived from secondary metabolism of living organisms which frequently possess crucial bioactivities - has deepened immensely. We continue to appreciate the significance of natural products and the role they play in our lives through natural product derived pharmaceuticals. New methods in discovering bioactive natural products and enzyme-based catalysts have greatly developed in recent years owing to the increasing availability of genomic data, including self-resistance enzyme directed genome-mining (SRE-DGM).

This thesis primarily focuses on understanding and utilizing the second-copy self-resistance phenomenon for the discovery of new natural products and enzyme catalysts. First, comparing the co-crystal structure of one housekeeping enzyme to the homology model of its second-copy SRE allowed us to understand part of the mechanism by which the SRE acquired its resistance toward aspterric acid, an herbicidal terpene natural product. Mutational studies on the housekeeping enzyme revealed key amino acids which increase resistance to aspterric acid. This allowed us to anticipate the inevitable naturally occurring herbicide resistance and be prepared to deploy counter measures. We also utilized SRE-DGM to find and elucidate the biosynthetic pathways of the mitochondrial complex II inhibitors, harzianopyridone and atpenin A5, both of which have been the target of investigation for more than four decades. We identified the biosynthetic gene clusters of these compounds from their fungal producers and uncovered the biosynthetic steps which include multiple iterative enzymes. In particular, a methyltransferase and a flavin-dependent monooxygenase are used iteratively to introduce the unique methoxy groups on the 2-pyridone core structure. The pathway unexpectedly requires the installation and removal of a *N*-methoxy group, which is proposed to be a directing group that tunes the reactivity of the pyridone ring. We also discovered a new type of halogenase that installs the chlorine substitutes in atpenin A5. This halogenase is able to install halogens on an aliphatic carbon and does not belong to any of the known halogenase families. These results signify the importance of SRE and its utilization in DGM for novel enzyme catalysts.

Finally, while the discovery of natural products is important, developing a robust and viable production methods of complex natural products is equally crucial. Thus, we developed an *in vitro* biosynthetic platform for producing plant terpene natural products. We established a one-pot, cell-free biosynthesis platform for nepetalactol and nepetalactone starting from the readily available geraniol. A pair of orthogonal cofactor regeneration systems permitted NAD^+-dependent geraniol oxidation followed by NADPH-dependent reductive cyclization without isolation of intermediates. The overall reaction contains 10 enzymes, four of which are biosynthetic enzymes, including a soluble P450, and five accessory and cofactor regeneration enzymes. Our *in vitro* platform yielded

~130-fold greater amount of nepetalactol and nepetalactone than the highest producing microbial platform.

The dissertation of Undramaa Bat-Erdene is approved.

Yvonne Y. Chen

Steven E. Jacobsen

Junyoung O. Park

Yi Tang, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Section 2 contains material written by Bat-Erdene,U. from the following publication: Yan, Y., Liu, Q., Zang, X., Yuan, Sh., Bat-Erdene, U., Nguyen, C., Gan, H., Zhou, J., Jacobsen, S.E. and Tang, Y. “Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action.” *Nature* 559, 415-418 (2018)

Section 3 and 4 contain material written by Bat-Erdene,U. from the following publication: Bat-Erdene. U., Kanayama, D., Tan, D., Turner, W. C., Houk, K. N., Ohashi, M. and Tang, Y. “Iterative Catalysis in the Biosynthesis of Mitochondrial Complex II Inhibitors Harzianopyridone and Atpenin B” *J. Am. Chem. Soc.* 142, 8550–8554 (2020)

Section 5 contains material written by Bat-Erdene,U. from the following publication: Bat-Erdene, U., Billingsley, J. M., Turner, W. C., Lichman, B. R., Ippoliti, F. M., Garg, N. K., O’Connor, S. E., and Tang, Y. “Cell-Free Total Biosynthesis of Plant Terpene Natural Products using an Orthogonal Cofactor Regeneration System.” *ACS. Catal.* 11, 9898–9903 (2021)

The work described in this dissertation was supported by the National Institute of Health grant R01AT010001-01 awarded to ‘Team Stricto’ and led by Yi Tang, as well as by the NIH grant R01AI141481 to Yi Tang and Kendall N. Houk.

First and foremost, I would like to express my most sincere gratitude and appreciation for Professor Yi Tang. He has been my mentor and driving force throughout my PhD journey, helping and encouraging me at each and every turn. He guided me in becoming an independent researcher, who not only focuses on one's own research, but someone who thinks about the broader implications of the work that we do. His unique take on scientific topics and vast

knowledge of science has inspired me to keep focus and dig deep into my own projects while keeping an open eye out for new opportunities, question the conventional knowledge while accepting the facts, and always push forward. He has taught me how to be a well-versed scientific writer through his many examples, which I must admit is still in progress and is something I will keep working on. He has supported me through the global pandemic we all experienced and has allowed me to enjoy my experience at UCLA even through the hard times.

My first direct interaction with PT was a memorable one in my books. It was after I got an email from PT about my possible acceptance to UCLA after a phone interview, since I was in Japan at the time. We had scheduled a phone call at 6 PM LA time and 10 AM in Tokyo time. After waiting nervously on my phone for his call at 10 AM, the call did not come. Prior to our phone call, PT had arranged for me to have an in-person interview with Prof. Kudo at Tokyo Institute of Technology. Prof. Kudo asked if I had a good chemistry knowledge, since I was going to need a lot of it. He also pointed to a polyketide compound and asked if I could propose a biosynthetic pathway for the molecule. I answered honestly that I was not completely confident in my knowledge in chemistry, and I couldn't propose a pathway for the molecule he presented. I could only hope that I was able to convey my genuine interest in the field and was willing to learn and put in as much effort as was required of me. When the phone call was an hour late, I panicked thinking he may have changed his mind after talking to Prof. Kudo. Fortunately for me, that was not the case. After I emailed PT that I was still waiting for his phone call, he explained how he had lost track of time reading papers and we had a smooth conversation over the phone, after which I was accepted into UCLA.

In my phone interview with PT, he made me feel at ease about my decision to come to the United States and gave me confidence that I am making the right decision about pursuing my PhD studies at Tang lab. However, I was feeling huge imposter syndrome when I first joined the lab, lacking confidence in my abilities and resilience. But with the support of PT and the incredible

team of scientists he assembled in the Tang lab, I was able to learn, discover and persevere through all the challenges I faced throughout my studies, for which I am eternally thankful.

I would also like to provide sincere thanks to my PhD committee: Prof. Chen, Prof. Jacobsen, and Prof. Park. Each of these individuals have provided advice and encouragement along the way, and I believe I could not have assembled a team of more brilliant scientists to guide me. I would especially like to express my gratitude for Prof. Jacobsen, for allowing me to conduct experiments in his lab and learn different sets of skills which I am hoping to utilize fully in my next endeavor.

I would like to express my gratitude to my collaborators Dr. Basudev Ghoshal and Dr. Qikun Liu from the Jacobsen lab, who has helped me to learn plant transformation techniques. Also, I would like to thank my collaborators from the Shanghai Institute of Organic Chemistry, Prof. Jiahai Zhou, Zang Xin, and Dr. Weixue Huang, for acquiring the crystal structures for my studies. My sincere gratitude also goes to Prof. Sarah O'Connor from Max Plank Institute of Chemical Ecology and her former postdoc Dr. Benjamin R. Lichman for their collaboration on our *in vitro* biosynthesis project. Thank you Prof. Neil K. Garg and Francesca M. Ippoliti for your organic synthesis expertise.

I am also forever indebted to members of the Tang lab, who have made my PhD experience richer and evermore fulfilled. I would like to thank Dr. Masao Ohashi and Dr. Yan Yan for being my mentors, teaching me and giving be invaluable advice in all my projects. I must thank Dr. John M. Billingsley for being my friend and mentor, who not only inspire me immensely for being a brilliant scientist, but also teach me how to be one's most authentic self. I will always be grateful for our late-night discussions in lab and WeHo outings (and others) for the joy that it brought me every time. Thank you for being my other unapologetic Samantha. I also need to thank the amazing undergraduate researched who worked with me during my studies William C. Turner for his dedicated work. Lastly, thank you my one and only cohort, Danielle Yee for always being the most genuine, humble and inspiring human being that you are. Thank you for listening

to my concerns across our lab bench and laughing at my awful jokes. Every member of the Tang lab has taught me valuable lessons and I have the utmost respect for each and every one of you.

I must also thank those outside of my lab who have provided me with advice and support over the past five years. I truly couldn't have done this without the emotional support of my closest friends: Undral Byambadalai, Narantsatsral Ganbold, Battuvshin Naranbat, Gunjkham Galbadrakh, Uranzaya Gankhuyag, and Narangerel Ganbaatar. Although many of you are physically away from me, you have given me courage and support every time we talked over the internet or the phone, especially during my last year of PhD which coincided with the global pandemic.

Lastly, I need to thank my family. Thank you to my mom who leaves a message for me every morning so that I don't feel alone. Thank you for understanding me, supporting me and for the many laughs we share. I truly couldn't have done this without you. Thank you to my brother, whom I can talk to at any time about anything without fear of judgement. Thank you to my sister-in-law who has not only brought my beautiful nephews into this earth but has become my confidant and true friend. Finally, thank you, to my one-and-only brilliant dad. You are not only my support but is my role-model. I know that you have my back for anything I want to achieve in life. Thank you for your wisdom that has shaped who I am and will continue to guide me. I love you all from the bottom of my heart.

VITA

2010 – 2014	Tokyo Institute of Technology B.S. in Chemical Engineering Tokyo, Japan
2014 – 2016	Tokyo Institute of Technology M.E. in Bioengineering Tokyo, Japan
2010 – 2014	Japanese Government (Monbukagakusho) Scholarship for Undergraduate Studies, Tokyo, Japan
2014 – 2016	Japanese Government (Monbukagakusho) Scholarship for Graduate Studies, Tokyo, Japan

PUBLICATIONS

Yan, Y., Liu, Q., Zang, X., Yuan, Sh., Bat-Erdene, U., Nguyen, C., Gan, H., Zhou, J., Jacobsen, S.E. and Tang, Y. "Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action." *Nature* 559, 415-418 (2018)

Hagiwara, Y., Mihara, Y., Sakagami, K., Sagara, R., Bat-Erdene, U., Yatsunami, R. and Nakamura. S. "Isolation of four xylanases capable of hydrolyzing corn fiber xylan from *Paenibacillus* sp. H2C." *Biosci Biotechnol Biochem*. 84, 640-650 (2020)

Bat-Erdene, U., Kanayama, D., Tan, D., Turner, W. C., Houk, K. N., Ohashi, M. and Tang, Y. "Iterative Catalysis in the Biosynthesis of Mitochondrial Complex II Inhibitors Harzianopyridone and Atpenin B." *J. Am. Chem. Soc.* 142, 8550–8554 (2020)

Bat-Erdene, U., Billingsley, J. M., Turner, W. C., Lichman, B. R., Ippoliti, F. M., Garg, N. K., O'Connor, S. E., and Tang, Y. "Cell-Free Total Biosynthesis of Plant Terpene Natural Products using an Orthogonal Cofactor Regeneration System." Accepted to *ACS Catal*. (2021)

PRESENTATIONS

Bat-Erdene. U., Nakamura. S., "Corn arabinoxylan degradation with GH10 xylanase from *Paenibacillus* sp. H2C." 96th Japanese Chemical Society Annual Meeting, 2016, Doshisha University, Kyoto, Japan. Presentation

Bat-Erdene. U., Yan. Y., Tang, Y., "Genome mining of a natural product herbicide with a new mode of action", Keystone Symposia; Natural Products and Synthetic Biology: Parts and Pathways, 2018, Resort at Squaw Creek, Olympic Valley, CA, USA. Poster Session

1 INTRODUCTION

1.1 Natural products and their discovery

Natural products (NPs), defined as small secondary metabolites (<10 kDa) derived from all kingdom of living organisms, have been an essential part of preserving human health and well-being since the beginning of time.[1,2] The inception of modern NP drug discovery was ignited by the serendipitous discovery of penicillin in 1928, where the contaminant fungal strain of *Penicillium* has inhibited the bacterial growth on an agar plate.[3] Since then, the discovery and utilization of NP for human health and utilities have burgeoned. According to latest reports, 33.6% of the 1394 small molecule drugs newly approved by the US Food and Drug Administration (FDA) between 1981 - 2019 is NPs or their derivatives[4]. If synthetic or semi-synthetic NP mimics are included, this number will rise to 68.8%, clearly showing the significance of NPs and their derivatives' importance for new drug discovery and application. Figure 1 shows representative bioactive NPs and their derivatives.

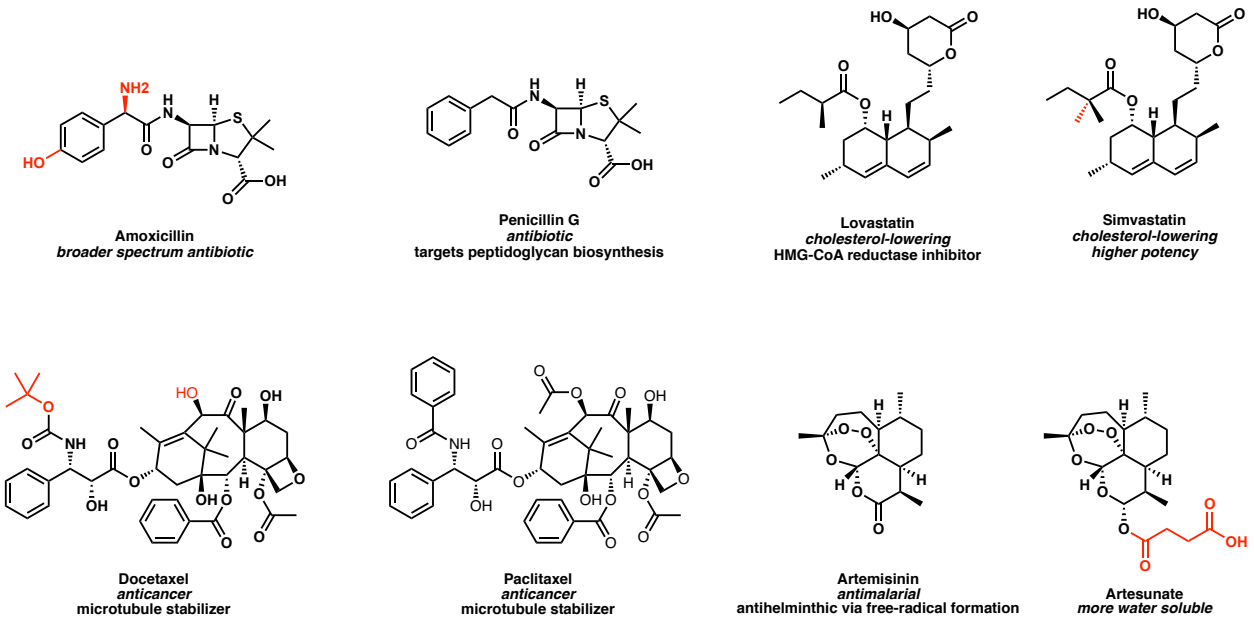


Figure 1. Representative bioactive natural products and their derivatives. Differences in structures are depicted in red

Early discovery of NPs mainly hinged on bioactivity-based screening (Figure 2. A), where collected biomass would be fractionated, and the fractions' bioactivity would be tested via different methods depending on the application. The fractionation and bioactivity testing would be repeated, until the NP responsible for the activity is isolated and the structure elucidated. Through this approach, scientists have been able to expand the bioactive NPs repertoire hugely and contribute to all endeavors of life.[5] However, since the method relied heavily on the NPs to be already present in the biomass of endogenous host at a relatively high yield, many NPs and scaffolds were left to be discovered. These include but not limited to NPs that are expressed at a lower yield, stress-responsive NPs and NPs produced by cryptic biosynthetic gene clusters. Furthermore, chances of re-discovering a known compound or scaffold increase with continuous research and discovery of NPs.

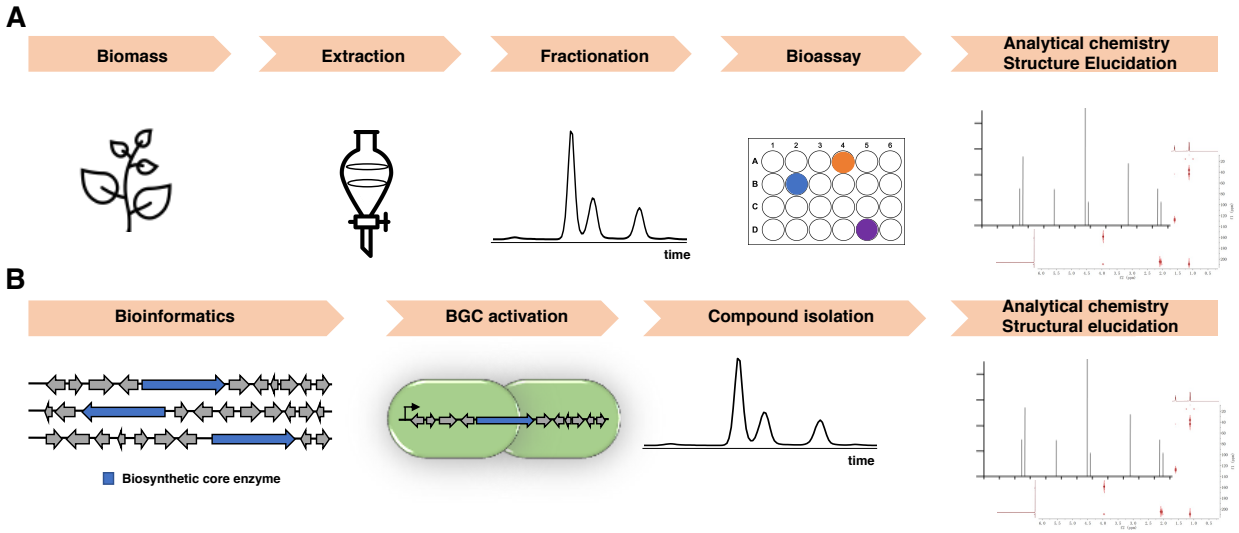


Figure 2. Conventional NP discovery methods. A. Bioactivity-based discovery, B. Genome-mining

Since the development of next-generation sequencing, the low cost (Figure S1. A) and increased availability of genomic data (Figure S1. B) have allowed scientists to develop a new method of NP discovery. In microorganisms, the enzymes responsible for the production of NPs are generally co-regulated and encoded closely on the genome as a biosynthetic gene cluster (BGC). Thus, by locating biosynthetic core genes (e.g., polyketide synthase (PKS), non-ribosomal

peptide synthetase (NRPS), terpene cyclase (TC)) and potential tailoring enzymes (e.g., cytochrome P450 monooxygenase (P450), flavin-dependent monooxygenase (FMO), methyl transferase (MT)) in the vicinity and subsequently expressing them resulted in the discovery of numerous NPs and is dubbed “genome-mining” (Figure 2. B).[6–10] Not only this approach has allowed scientists to discover previously unknown NPs and scaffolds by activating the BGC in the native host or expressing the BGC in heterologous hosts, but also allows the investigation into biosynthetic pathway of NPs and mechanistic studies of various previously unknown tailoring enzymes. Similar mining approaches can be taken for NPs derived from other kingdoms of life, such as plants, with the help of transcriptomics and metabolomics analysis.[11,12]

1.2 Demand for new bioactive natural products and enzyme chemistry to overcome drug-resistance

Drug resistance to existing pharmaceuticals has become a major global issue with the emergence of lethal multidrug-resistant superbugs[3,13,14] and major crop losses due to herbicide-resistant weeds[15,16]. Cases of resistance to all classes of anti-bacterial and anti-fungal drugs have been reported, with increasing lethality in clinical cases[17]. Continuous exposure to any bioactive compound over an extended period of time inevitably results in drug insensitivity in organisms, when organisms are evolved to defend against said bioactive compounds.[18] Thus, drug and herbicide resistance are problems that need to be addressed continuously through the introduction of new bioactive pharmaceuticals and herbicides.

In addition to introducing new bioactive compounds, another way of addressing the continuous need for new bioactive compounds is to derivatize existing bioactive compounds to acquire additional features. For instance, as depicted in Figure 1, the penicillin-G derivative amoxicillin has a broader spectrum of antibiotic activity than its precursor.[19] The cholesterol lowering NP, lovastatin was transformed to a more potent simvastatin with the addition of a methyl

group on its side-chain.[20] Although, many of such modifications can be accomplished through organic synthesis, these methods harbor drawbacks including hazardous reagent requirements, multistep protection/deprotections and poor regio- and stereoselectivity. On the other hand, enzymatic biocatalysts make precise modifications under mild conditions, thus emerging as an attractive alternative to chemical synthesis modifications. Such as the aforementioned simvastatin can be produced by an *Escherichia coli* whole-cell lysate expressing the acyltransferase LovD from its precursors monacolin J and α-dimethylbutyryl-*S*-methyl-mercaptoproprionate.[21,22] Furthermore, enzyme biocatalysts can be evolved to have more beneficial features such as heightened enzymatic activity, stability and substrate specificity via directed evolution. For example, LovD was evolved to have higher activity and stability to be used as a biocatalyst.[23] Therefore, discovering new enzymes which have activities that could be developed into biocatalysts for NP derivatization is crucial.

Additionally, our understanding of natural drug resistance needs to be deepened. If the resistance mechanism could be anticipated, we could have a strategy in advance to deploy once the inevitable resistance surfaces. Although, there are ongoing research in drug design to overcome and prevent drug resistance, the results are always delayed with minimal effects on preventing the emergence of new resistance[24,25]. Thus, in addition to discovering new bioactive compounds with new modes of action and developing new biocatalysts, it would be ideal if we could understand and develop a method to anticipate drug resistance.

1.3 Self-resistance phenomenon and second-copy self-resistance enzyme (SRE)

Self-resistance refers to a phenomenon where organisms prevent themselves from the bioactive NPs they themselves produce. Commonly adopted strategies of self-resistance include but not limited to drug efflux pumps, toxic-compound inactivating enzymes and modification of target enzyme to be tolerant to drugs (Figure 3).[26–29] Such strategies are utilized alone or in

combination by the NP producing organisms and have been found to be commonly encoded as self-resistance genes.[30–33] Some of these strategies are adopted by organisms in response to bioactive compounds produced by other organisms. A prominent example of this is the β-lactamase enzyme that deactivates β-lactam antibiotics such as penicillin. Many bacterial species have acquired this defense mechanism through horizontal gene transfer.[3]

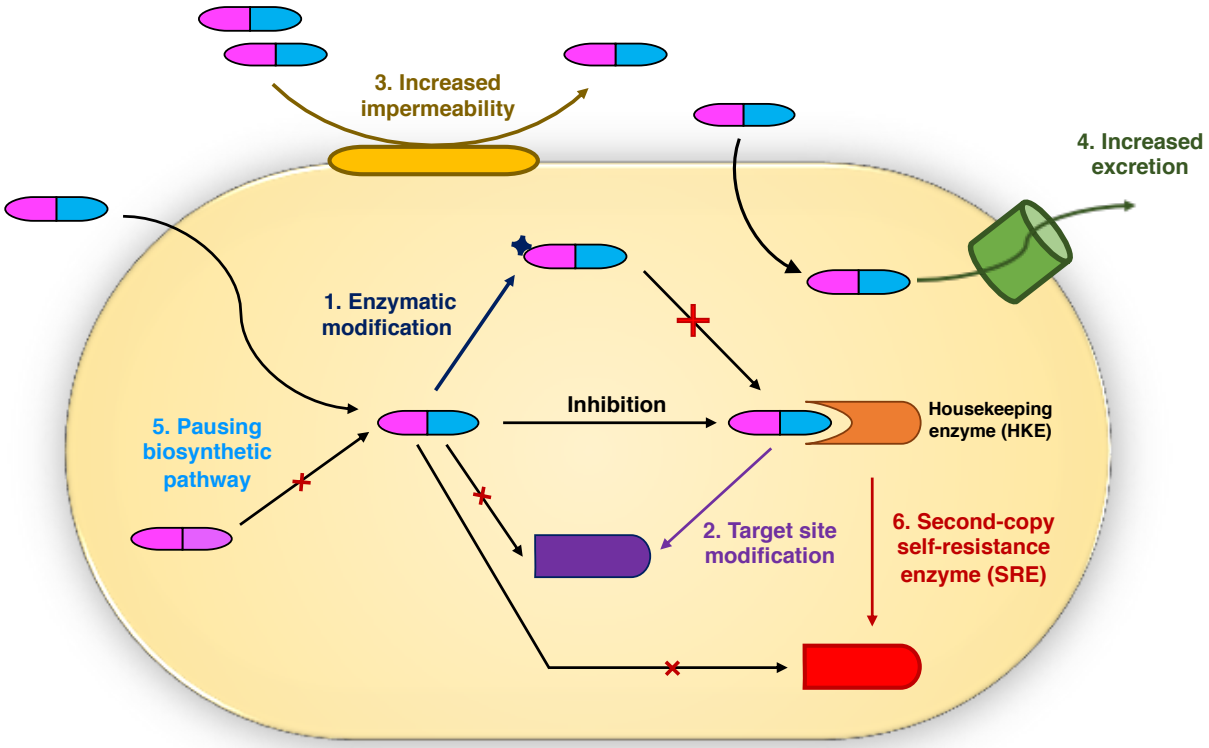


Figure 3. Self-resistance phenomenon.

A less understood self-resistance mechanism is called the second-copy self-resistance, which refers to organisms harboring a mutated and resistant copy of the enzyme (house-keeping enzyme; HKE) that is targeted by the bioactive NP.[34] This strategy differs from the target site modification strategy by the former harboring extra copy of the HKE that is mutated and resistant whereas target site modification generally occurs on the only copy of the HKE encoded in the genome. We are calling this resistant copy enzyme the second-copy self-resistance enzyme (SRE) (Figure 3).

The first bacterial second-copy self-resistance was reported in 1988 for the novobiocin-resistant DNA gyrase,[35] but it was not until 2000 that the SRE was confirmed to be clustered within the novobiocin BGC.[35,36] In 2004, C. Olano *et al*. have also identified a borrelidin-resistant second-copy threonyl tRNA synthetase (*bor*O) within the borrelidin BGC. Although they reported that *bor*O conferred borrelidin resistance to a heterologous host, the differences between *bor*O and the house-keeping threonyl tRNA synthetase that account for the resistance of *bor*O against borrelidin remains unknown.[37] In 2011, salinosporamide A (a proteasome inhibitor) second-copy SRE mechanism was investigated by the Moore group. They found that a single amino acid mutation near the binding pocket of the proteasome subunit conferred resistance to salinosporamide.[38] Intriguingly, the resistance-conferring mutation was also found in human cell lines that acquired proteasome inhibitor resistance. With the knowledge of second-copy SRE gene residing close to its toxic NP BGC, Moore group have identified another second-copy SRE via SRE-directed genome-mining (SRE-DGM) (Section 1.4), the fatty acid synthase inhibitor thiolactomycin BGC in 2015.[39] However, its mechanism for resistance has not been studied yet. Another example of second-copy self-resistance involving fatty acid synthase was published in 2014, where the BGC of platencin contained a second-copy-SRE.[40] Self-resistance of the antibiotic griselimycin-producing *Streptomyces* was revealed to be due to the second-copy SRE phenomenon, though its mechanism is not fully understood.[41]

Fungal second-copy SRE is less conspicuous and mechanistic studies are lacking. In 1999, it was revealed that lovastatin BGC contained a HMG-CoA reductase homolog within its BGC.[42] However, it has not been verified that this second-copy gene is resistant to lovastatin. The SRE-DGM was successfully used to identify mycophenolic acid (immunosuppressant targeting IMPDH) BGC in 2011, though its resistance mechanism has not been studied.[43] The authors proposed that a homolog of IMPDH may be present to confer resistance to mycophenolic acid and it was subsequently used as a probe to locate the BGC. Next cases of fungal second-copy SR were reported in 2013, 2016 and 2018 in fumagillin (anti-angiogenesis), fellutamide B

(proteasome inhibition) and echinocandin (antifungal) BGCs, respectively.[44–46] However, their resistance mechanisms are unknown, and only the self-resistance of echinocandin-producing species has been verified.[46] The first in-depth verification, utilization, and explanation for the possible mechanism of a fungal second-copy SR gene was reported in 2018 for the aspterric acid BGC, where the second-copy self-resistance gene was utilized for an herbicide system.[47]

1.4 SRE-directed genome-mining (SRE-DGM)

As exemplified by the research discussed in Section 1.3, more often than not, SRE gene is encoded within or near the BGC that is responsible for producing the bioactive NP. Although conventional genome-mining is a powerful tool to discover NPs, it lacks the ability to inform researchers of the potential bioactivity of the NP being produced. In order to bridge the two approaches, scientist have utilized the second-copy self-resistance phenomenon for genome-mining (Figure 4).[34]

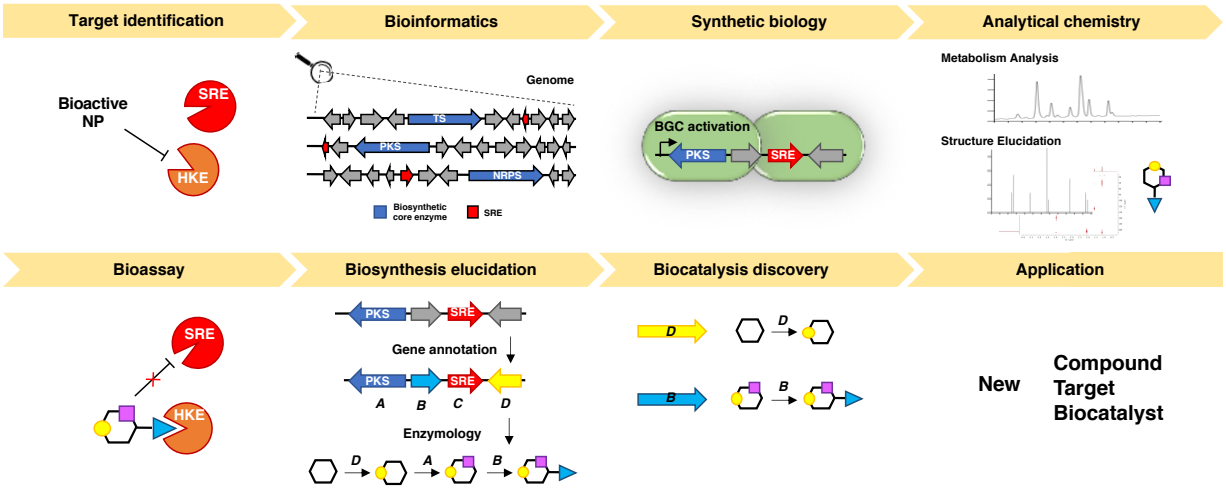


Figure 4. Self-resistance enzyme directed genome-mining (SRE-DGM)

SRE-DGM can be utilized in many ways. First, it can be used to locate the BGCs of NPs with known biomolecular targets. One such case of this application of SRE-DGM is the mycophenolic acid BGC identification discussed in Section 1.3. Another way to use SRE-DGM is

to identify the mechanism-of-action of a NP, whose biosynthetic pathway is fully or partially known. If the BGC of said NP contains a gene whose putative enzymatic role does not match the biosynthetic logic of the NP, this gene may be the SRE. The biomolecular target of the antibiotic griselymicin was identified in this manner, where the eventual target DNA polymerase sliding clamp DnaN homolog was encoded in the BGC.[41] Finally, SRE-DGM can be used for discovering NPs with desired targets, where a primary metabolism enzyme or a SRE can be used as a probe to discover NP BGCs that may produce new bioactive NPs that target the designated enzyme.

The SRE is not only invaluable for DGM, but it also serves as a doorway to understanding the various mechanisms of drug resistance. Many of the reported second-copy SREs have not yet been verified and their mechanisms for resistance are unknown. Since these SREs are the mutated copies of the biomolecular target of NPs, by elucidating the resistance mechanisms of the SRE may inform us to anticipate and thwart the devastating results of naturally evolved drug resistance.

In my research, the SRE was utilized as a probe for DGM, as well as being the subject of investigation. By exploring the differences between the SRE and HKE, we wanted to uncover the resistance mechanisms by which one of these second-copy SREs acquire their resistance (Section 2). With a clearer understanding of the various mechanisms of resistance, we can anticipate the mutations that will give rise to future drug-resistance and potentially forestall their negative effects by designing the next-generation drugs based on the SRE structure. We also used SRE-DGM in elucidating some long-unknown biosynthesis of important bioactive compounds and discover an important new biocatalyst along the way (Section 3 and 4).

Lastly, the *in vitro* biosynthesis of plant terpene natural products nepetactol and nepetalactone is discussed in Section 5. We developed orthogonal cofactor regeneration systems in order to optimize the *in vitro* biosynthesis of nepetalactol and aimed to produce high titer of this valuable terpene compound.

2 ELUCIDATING THE ASPTERRIC ACID RESISTANCE MECHANISM

Recently, aspterric acid (AA) and its dihydroxyacid dehydratase (DHAD) inhibition activity was discovered by our lab using SRE-DGM and has been proven to be an herbicide with a new mode-of-action.[47] However, the resistance mechanism behind the self-resistance of the producing fungal strain remained to be solved. In this section, the SRE mechanism has been explored through mutational studies on the house-keeping copy of the plant DHAD. The mutations explored in this section were based on the crystal structure comparison between the HKE and the SRE.

2.1 Branched chain amino acid biosynthesis and its inhibitors

Branched chain amino acid (BCAA) biosynthesis is largely conserved amongst different animal kingdoms including plants and microbes. In plant and microbial BCAA biosynthesis, *L*-Threonine (Thr) is deaminated by *L*-Threonine deaminase/dehydratase (IlvA) starting the synthesis of *L*-Isoleucine (ILe). The biosynthesis steps are then shared among Ile, *L*-Valine (Val) and *L*-Leucine (Leu), where 2-ketobutyrate or pyruvate is condensed with pyruvate by acetohydroxyacid synthase (AHAS) to give 2-acetolactate for Leu and Val synthesis or 2-aceto-2-hydroxybutyrate for Ile synthesis. These keto acids are then reduced and isomerized by keto acid isomeroreductase (KARI) to 2,3-dihydroxyisovalerate and 2,3-dihydroxy-3-methylisovalerate, respectively, which in turn will be dehydrated by dihydroxyacid dehydratase (DHAD), forming 2-ketoisovalerate and 2-keto-3-methylvalerate (Figure 5). These intermediates are futher modified to synthesize the BCAAs.[48,49]

Since the BCAAs are essential for plant growth, their widely-conserved biosynthetic pathway[48,49] has been targeted for herbicide development for more than three decades[50,51]. Synthetic small molecule inhibitors targeting AHAS have been developed and commercialized as herbicides since 1972, which belong to compound families, sulfonylurea (SU), imidazolinone (IMI),

triazolopyrimidine (TP), pyrimidinylbenzoate (PYB), and sulfonylamino-cabonyl-triazolinone (SCT). Since they have been on the market for a substantial amount of time, their inhibitory activity, cocrystal structures, in addition to many of their naturally occurring resistance in plants have been studied and mechanisms uncovered.[48,52]

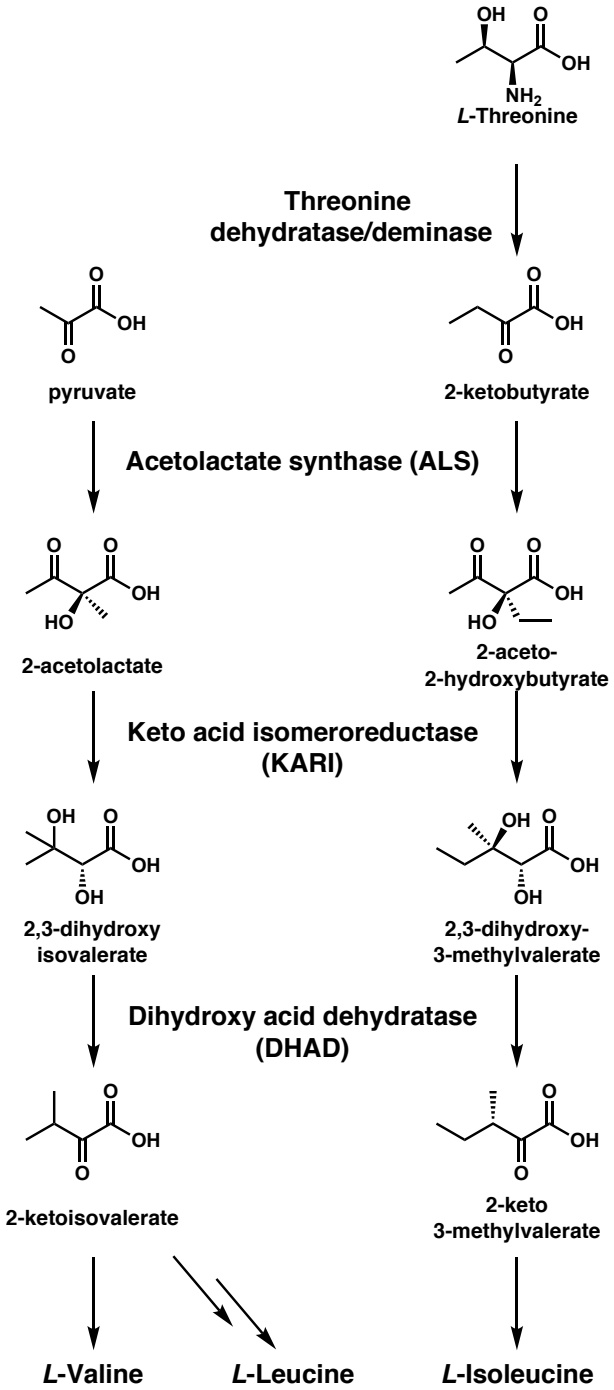


Figure 5. Partial branched-chain amino acid (BCAA) biosynthetic pathway

The second common enzyme in BCAA pathway, KARI has been targeted for synthetic antibacterial drug development, most importantly against *M. tuberculosis*[48]. Recent studies include drug leads that show inhibitory activity against *M. tuberculosis* at minimal inhibitory concentration of 0.8 μM.[48,52] The last common enzyme DHAD has been elusive for inhibition and no compounds have been developed as pharmaceuticals to target this enzyme despite its attractiveness. Although compounds inhibiting these enzymes displayed antibacterial activities, reports show that they are not superior to existing drugs[52]. However, such compounds still have the potential to be used as combinatorial treatments and need to be investigated further, given the alarming increase in the number of drug-resistant pathogens[53].

2.2 Discovery of aspterric acid and its biosynthesis using SRE-DGM

Aspterric acid (AA) and its DHAD inhibition activity was discovered by our lab using SRE-DGM and utilized as an herbicide lead with a new mode-of-action.[47] The housekeeping DHAD homolog sequence was used as a query for SRE-DGM in addition to biosynthetic core enzymes such as terpence cyclase and polyketide synthase. A potential BGC containing a terpene cyclase (AstA), two cytochrome P450 enzymes (P450s: AstB and AstC) and a putative SRE (AstD) was found in the genome of *Aspergillus terreus* (Figure 6). Heterologous expression of the biosynthetic enzymes AstA-C has confirmed the putative BGC is responsible for AA production.[47]

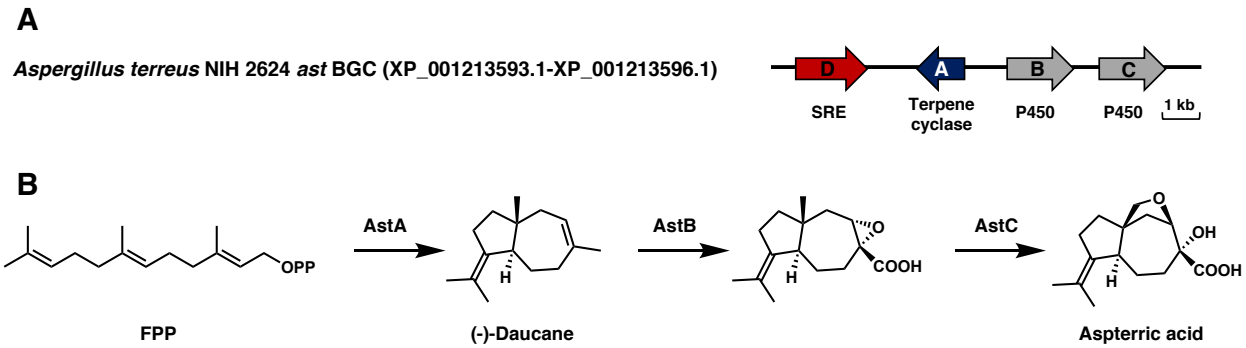


Figure 6. Aspterric acid (AA) biosynthesis. (A) AA BGC. (B) AA biosynthetic pathway

Although AA was initially isolated four decades ago,[54] and its bioactivity as a plant pollen development inhibitor was found later, its mode of action was unknown prior to this re-discovery.[55] AA did not exhibit inhibitory activity against AstD *in vitro* and *in vivo*, while its half-maximal inhibition concentration (IC_{50}) of plant DHAD (35 amino acid *N*-terminal chloroplast targeting sequence removed; pDHAD) was submicromolar at 0.5 μM *in vitro.*[47] It also exhibited strong inhibition in plant spray experiments with wild-type model plants whereas transgenic plants harboring the AstD gene were able to recover their normal growth after being sprayed with AA.[47]

2.3 Analysis of *holo*-pDHAD-AA co-crystal structure

Although, AA was proven to be a potent natural product herbicide lead, AstD's mechanism of resistance to aspterric acid remained unsolved. There are many reported cases of point mutations that underlie acquired resistance to existing drugs and herbicides.[56,57] Based on these information, in conjunction with the relatively high sequence identity between pDHAD and AstD (61%) (Figure S2), we predicted that few key amino acid changes may be responsible for the AstD's resistance to AA. However, amino acid mutations based solely on sequence alignment did increase the pDHAD resistance to AA (Data not shown). Thus, we proceeded to deduce potential amino acids that may increase the AA resistance of AstD from the structural comparison of *holo*-pDHAD-AA co-crystal structure (Figure 7) acquired by our collaborators at Shanghai Institute of Organic Chemistry and AstD homology model based on pDHAD structure.

The crystal structure of *holo*-DHAD has not been solved until our 2018 paper despite its involvement in the biosynthesis of essential amino acids,[47] due to the oxygen sensitive 2Fe-2S cluster that render it difficult for crystallization. The only other Fe-S-cluster-containing enzyme that has been crystallized to date is the *E. coli L*-arabinonate dehydratase,[58] proving the difficulty of crystallization. Although, the crystal structure of pDHAD was acquired, it was not co-crystallized

with AA in its active site. In contrast, the newly acquired structure of pDHAD was co-crystallized with AA (Figure 7. A).

2Fe-2S cluster is reported to be held in place in the active site with either four cysteine ligands or occasionally other residues such as aspartate, histidine or serine.[59] In pDHAD structure, three Cys residues (C66, C139, and C211) and an asparagine (N64) seem to be responsible for coordinating the 2Fe-2S cluster in the active site (Figure 7. B). C66, C139 and C211 are positioned near the active site with their side chain thiol groups 2.4 Å from the two irons of 2Fe-2S cluster. C211 is also 4.2 Å from the other iron, whereas N64 side chain amide group is at a 3.8 Å distance from the same iron. A fourth Cys residue (C205) appears to be 9 Å away from the iron, which is a distance not suitable to have a meaningful interaction. There appears to be no other suitable ligands near the 2Fe-2S cluster. Instead, the ether oxygen of AA is 4 Å away from the iron, making it possible to have an interaction.

The active site Mg^{2+} is positioned to interact with the carboxylic acid hydroxyl group of glutamate 98 (D98), hydroxyl group of threonine 216 (T216) and the amide of carboxamide side chain of asparagine 298 (N298). In addition, AA's carboxylic acid and hydroxyl moieties are positioned to interact with Mg^{2+}, mimicking the natural substrate (Figure 7. C).

Phenylalanine 181 (F181) and tyrosine 215 (Y215) are stacked on either side of the cyclopentane ring of tricyclic core structure of AA (Figure 7. D). In addition, proline 325 (P325) residue is stacked behind Y215 to support the hydrophobic stacking interaction. As mentioned in the previous paragraph the 2Fe-2S cluster and Mg^{2+} also has significant interactions with AA. Isoleucine 177 (I177), valine 178 (V178), serine 489 (S489) and leucine 465 (L465) residues are in interacting distance (<4 Å) with AA as well (Figure 7. D).

2.4 Potential amino acid mutations for AA resistance deduced from structure alignment

In order to deduce amino acid mutations that may lead to AA resistance in AstD, we built an AstD homology model based on pDHAD protein structure (Figure 8).

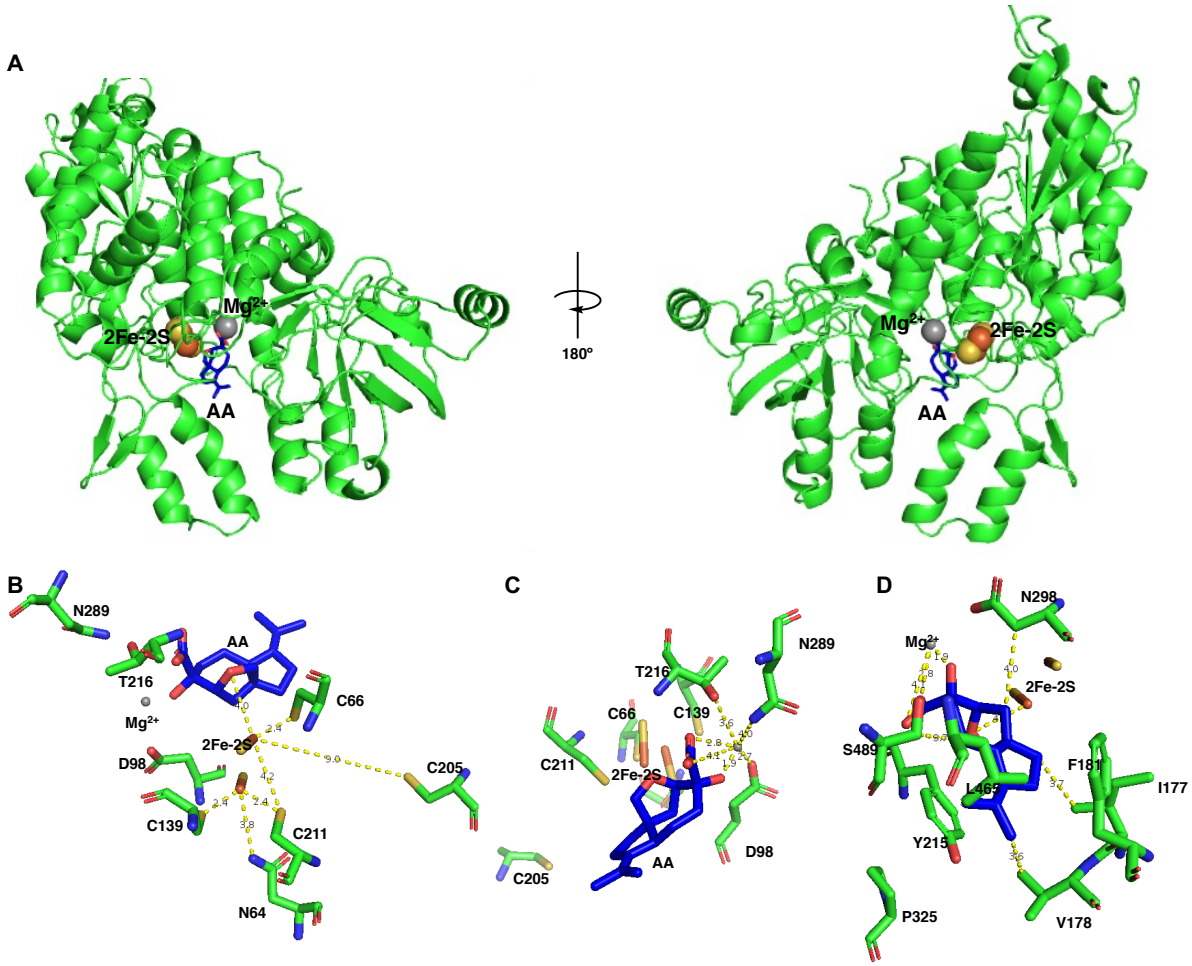


Figure 7. pDHAD crystal structure (A) Monomer of *holo*-pDHAD structure with AA complex. pDHAD exists as a homo-dimer with 2Fe-2S (orange spheres) and Mg2+ (grey sphere) in each monomer active site. AA is shown in blue sticks. (B) Closer view of pDHAD active site showing amino acids that interact with 2Fe-2S cluster (C) Closer view of pDHAD active site showing amino acids interactions with Mg^{2+}. (C) Closer view of pDHAD active site showing amino acids that interact with AA.

Comparison of the crystal structure of *holo*-pDHAD and AstD homology model revealed few key amino acid mutations that may result in increased resistance. Amino acids I177 and V496 in pDHAD corresponds to L198 and L518 in the AstD homology model, respectively (Figure S2). The larger hydrophobic side chain of L198 points toward the enzyme active site entrance, thereby

making the entrance narrower, which can prevent the bulky AA from entering the active site. Although L518 is located behind a loop, its side chain may push the loop adjacent to the active site further in, having the same entrance-narrowing effect as L198. V497 was mutated to Ile in AstD, which may also have a similar effect as V496L mutation.

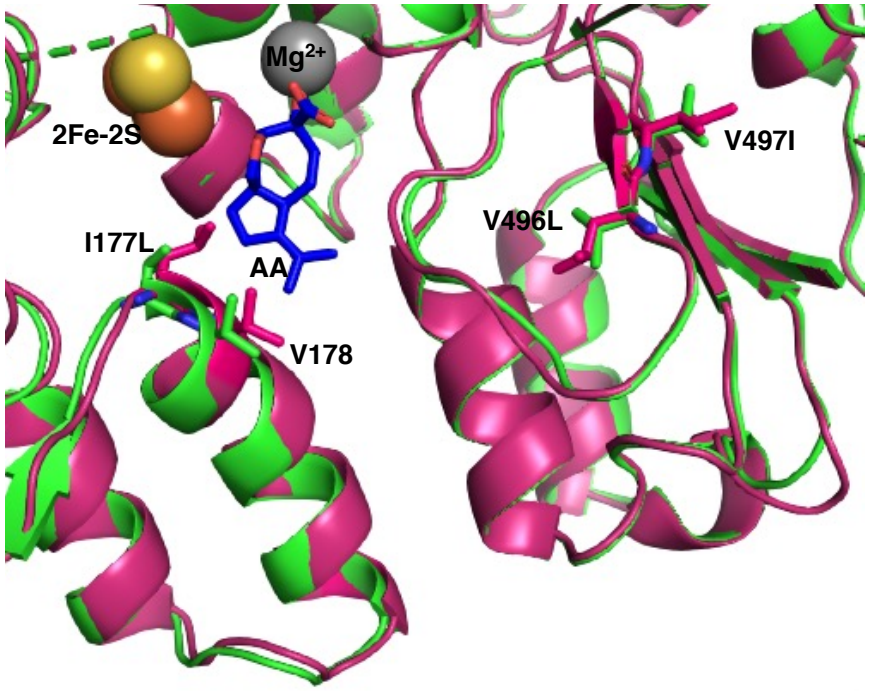


Figure 8. Active site entrance view of pDHAD crystal and AstD homology model. *holo*-pDHAD is depicted in green, AstD homology model is depicted in pink. 2Fe-2S cluster: orange spheres, Mg^{2+} ion: grey sphere; AA: blue sticks. Putative resistance conferring amino acids and their positions are indicated.

In order to confirm whether these mutations increase the resistance of pDHAD to AA, I177 and V496, V497 were mutated to amino acids with larger hydrophobic side chains, such as leucine, phenylalanine and tryptophan. In addition, V178, although not mutated in AstD was chosen to be mutated to larger hydrophobic amino acids (Table 1). Mutations were introduced through PCR with primers containing appropriate base changes for the mutations. (See Appendices for primer sequences.)

Table 1. Summary of pDHAD mutations

Positions	Mutation	Positions	Mutation
I177	Leu	V496	Leu
	Phe		Ile
	Trp		Phe
V178	Leu		Trp
	Ile	V497	Leu
	Phe		Ile
	Trp		Phe
I177 V496	Leu		Trp
I177 V497	Leu	V178 V497	Leu

2.5 Development of DHAD activity assay

In order to evaluate the effects of amino acid mutations on the catalytic activity of pDHAD, we developed an *in vitro* activity assay. Since the ketoacids produced by DHAD is not UV active, we used phenylhydrazine (PhH) to derivatize the ketoacids, making them strongly UV absorbent at λ = 340 – 345 nm (Figure 9). Both 2,3-dihydroxyisovalerate (DHI) and 2,3-dihydroxy-3-methylvalerate (DHMV) were used as the substrates and the IC_{50} was measured using DHI as substrate. The mutated pDHADs were expressed in *E. coli* BL21 (DE3) and purified using Ni^{2+}-sepharose resin. Due to the easily oxidized nature of the iron-sulfur cluster, the *E. coli* lysate was incubated with Ni-resin for only 2 hours to prevent activity loss. After 2 hours incubation at 4°C, the resin was washed then eluted with 1.5 mL buffer and used for kinetic assays directly. The enzymes could not be stored at -80°C or under any other conditions due to loss of activity, which prompted us to use freshly purified enzymes for each assay and its replicates.

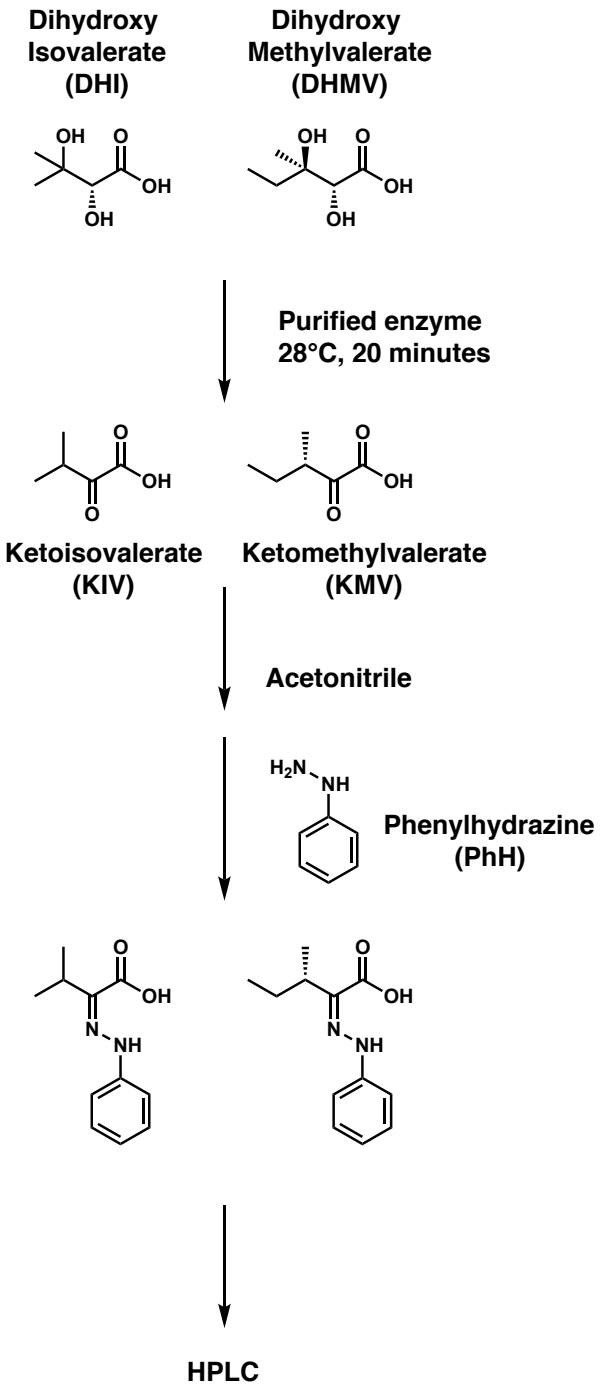


Figure 9. DHAD activity assay scheme

The SDS-PAGE of all purified enzymes are shown in Figure S3. After incubating the substrates with 0.5 μM mutant pDHADs for 20 minutes at 28°C, the reactions were stopped by adding equal volume of acetonitrile directly to the reaction. 1/25 volume of 0.1 M PhH was then added for derivatization and the mixture was incubated at room temperature for > 20 minutes.

After centrifugation the mixture was directly subjected to HPLC to monitor the enzyme activity. λ = 341 nm was used for the peak area extraction and used to calculate the product concentration. The concentrations of products were calculated by using calibration curve depicted in Figure S4.

2.6 *in vitro* evaluation of DHAD mutant activity

Parameters representing enzyme activity of wild-type pDHAD and its mutants are summarized in Table 2 and Table S1. When DHI was used as substrate, Michaelis-Menten constants k_{cat} and K_M of wild-type pDHAD were 6.85 ± 0.508 s^{-1} and 5.78 ± 0.508 mM, respectively.

Table 2. The IC_{50} and k_{cat}/K_m calculations of pDHAD and its mutants

Enzyme		IC_{50} (μM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)	
Position	**Mutation**	**DHI**	**DHI**	**DHMV**
pDHAD		0.249 ± 0.453e-3	1.19 ± 0.189	1.62 ± 0.143
I177	L	0.0502 ± 1.07e-3	0.270 ± 0.0152	0.280 ± 0.0144
	F	37.0 ± N/A	0.0230 ± 3.16e-3	0.0268 ± 7.002e-3
	W		N/A	
V178	L	0.974 ± 0.0643	0.153 ± 0.0146	0.154 ± 4.32e-3
	I	0.431 ± 0.0159	1.63 ± 0.491	1.08 ± 0.136
	F	6.00 ± 0.185	9.61e-3 ± 1.11e-3	0.0193 ± 2.60e-3
	W	2.34 ± 0.0265	8.72e-3 ± 4.05e-3	1.88e-3 ± 0. 243e-3
V496	L	0.824 ± 3.96e-3	1.90 ± 0.233	0.983 ± 0.106
	I	1.14 ± 0.0521	0.306 ± 0.0598	0.148 ± 0.0314
	F	2.94 ± 0.0118	0.623 ± 0.216	0.250 ± 0.240
	W		N/A	
V497	L	0.0590 ± 0.321e-3	1.40 ± 0.0341	1.07 ± 0.146
	I	0.0733 ± 0.988e-3	0.882 ± 0.0953	0.545 ± 0.0267
	F	1.25 ± 5.57e-3	0.0508 ± 6.23e-3	0.0356 ± 8.42e-3
	W		N/A	
I177 V496	L	0.309 ± 0.0741	1.94 ± 0.260	0.760 ± 0.0651
I177 V497	L	0.0280 ± 0.0264e-3	0.107 ± 3.13e-3	0.120 ± 4.06e-3
V178 V497	L	0.396 ± 0.0111	0.460 ± 0.0184	0.642 ± 0.0103

The $k_{cat,\ DHI}$ was slightly higher compared to the 1.2 s^{-1} that was reported in our previous paper, whereas the $K_{m,\ DHI}$ was the same. However, since the enzyme purification process was optimized for purifying enzymes with the highest possible activities, the apparent activity increase was expected. Similarly, k_{cat} and K_M were 6.50 ± 0.323 s^{-1} and 4.00 ± 0.446 mM, respectively, when DHMV was the substrate. The catalytic activity was comparable between the two substrates,

which is expected from the wild-type enzyme. (Table S1). The AA IC_{50} for pDHAD was 0.249 $\pm$ 0.453·10^{-3} μM, which was comparable to our previous report.

The IC_{50} of purified I177L and I177F mutants were 0.0502 μM and 37.0 μM, with $k_{cat,\ DHI}$ of 1.30 and 0.114 s^{-1}, respectively. The catalytic activity parameters were not obtainable for purified I177W mutant, since its activity was too low to be detected. This was true for all other Trp mutations except for V178W, for which we were able to detect extremely low activity. Although, I177 is mutated to Leu in AstD, the same mutation was not able to provide an increase in AA resistance in pDHAD. In contrast, Phe mutation was able to increase the IC_{50}, proving our hypothesis of larger hydrophobic side-chains being able to increase AA resistance. Unfortunately, the I177F mutation was also likely obstructing the natural substrates to have the correct conformation for reaction in the active site, which can explain the drop in k_{cat}.

The IC_{50} of purified V178L, V178I, V178F and V178W were 0.974, 0.431, 6.00 and 2.34 μM, respectively. These increase in IC_{50} indicates the larger hydrophobic side-chains of these mutants were able to narrow the active site entrance and prevent the AA from entering the active site. Interestingly, the V178I mutant had an increased k_{cat} of 23.4 s^{-1}, albeit the k_{cat}/K_M, the parameter indicating the overall enzymatic activity was not changed. Compared to the I177 position, the V178 side-chain protrudes into the active site entrance chamber, decreasing its chance to interfere with the catalytic activity. Thus, the relatively smaller side-chains of Leu and Ile mutations were beneficial in preventing AA to pass through the entrance chamber, compared to Val, while not obstructing the smaller natural substrates, resulting in relatively undisturbed k_{cat} numbers. However, similar to I177F, the Phe and Trp side-chains proved to be too large, decreasing the catalytic activity significantly.

The IC_{50} of purified V496L, V496I and V496F were 0.824, 1.14 and 2.94 μM, respectively, showing overall increased resistance compared to wild-type pDHAD. The k_{cat} values of V496L and V496I were not significantly compromised, indicating these mutants' side chains were large

enough to increase the resistance by pushing the entrance-adjacent loop inwards, but not too large that it compromises the ability of native substrates to enter the active site. However, the Phe mutation proves to be too drastic, and the enzyme activity is compromised.

Next, the IC_{50} of purified V497L, V497I and V497F were 0.059, 0.0733 and 1.25 μM, respectively. The Phe mutation had the same effect as all the other Phe mutations, where it increased the resistance while severely compromising the catalytic activity of the enzyme. As for Leu and Ile mutants, they were not able to increase the resistance when the mutation was at V497 position, likely due to the side-chain at this position being directed to the opposite side of the entrance-adjacent loop. Hence, the slight increase in side-chain size was not able to have the narrowing effect on the entrance chamber by pushing the loop inwards. Lastly, the Leu double mutants were not able to significantly affect the resistance.

These results suggest although the amino acids located near the active site entrance play a crucial role in resistance to aspterric acid, they alone are not sufficient in explaining the extremely high AA resistance in AstD, which is reported to be more than 8 mM. Furthermore, the enzyme activity gets compromised when the AA resistance is high. This is further confirmed by the difference in the k_{cat} of pDHAD and the reported k_{cat} of AstD at 6.85 and 0.03 s^{-1}, respectively, suggesting the AA resistance comes with compromised enzyme activity.[47] Since point mutations made to pDHAD, based on homology modeling, were not able to increase its AA resistance to the extremely high level of AstD,[47] other subtle changes in the conformation must contribute to this resistance. Furthermore, although the mutations were based on homology modeling, it is confirmed that crystal-structure based rational mutations are a formidable method for elucidating the SRE mechanism and increasing the resistance of DHAD to AA.

2.7. Crystal structures of *holo*-DHAD mutants without AA

Our collaborators were successful in obtaining the *holo*-crystal structure of two of the pDHAD mutants, V178W and V496W (Figure 10). Unfortunately, space to accommodate AA or the native substrates in the active site appears to be severely compromised, explaining why the cocrystal structure was not obtained. In these structures, we can clearly see the effects of the Trp mutations, where in the V178W mutant structure the active site entrance appears to be completely blocked by the Trp side-chain, which explains the drastic decrease in in the enzyme catalytic activity (Figure 10. B). In the V496W mutant, although, the 2Fe-2S cluster in the active site appears to be accessible, substrate's access to Mg^{2+} seems to be obstructed, which explains why we could not see enzymatic activity *in vitro* (Figure 10. C).

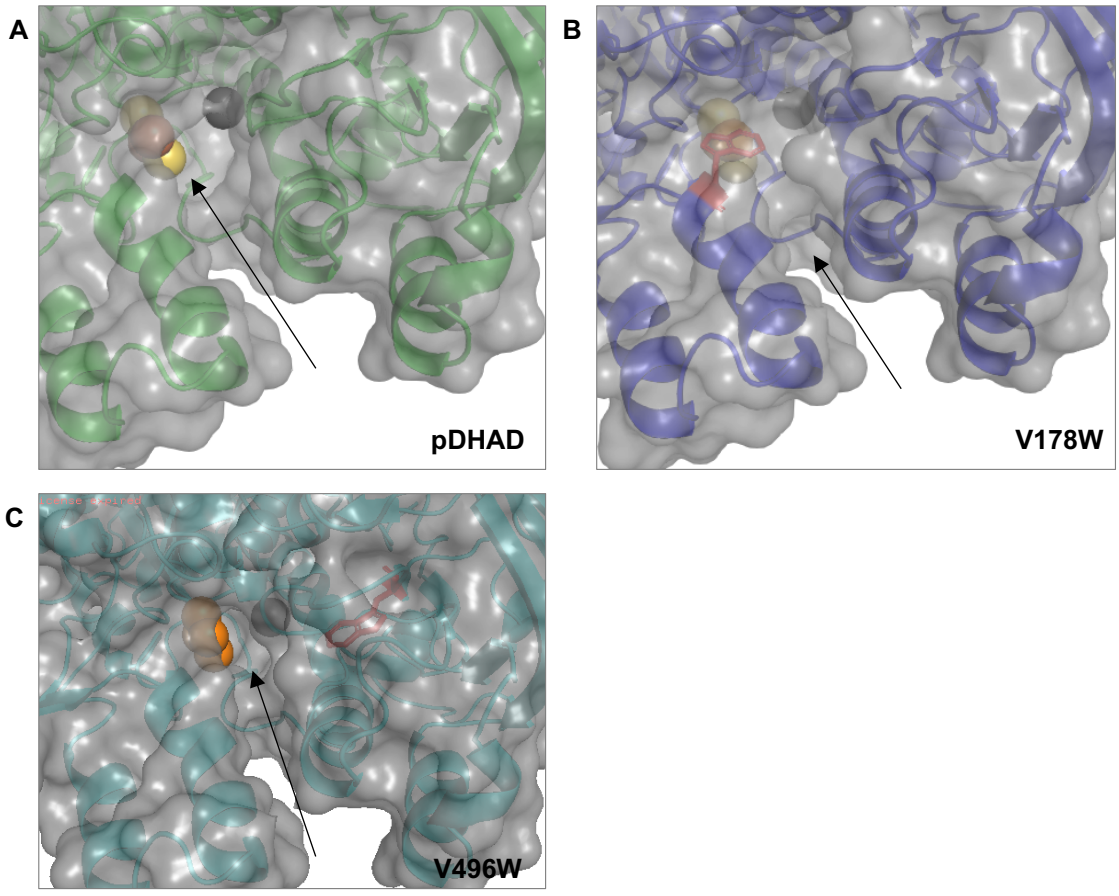


Figure 10. Protein surface of pDHAD and its mutants' active site entrance (A) pDHAD (B) V178W (C) V496W. Arrows in indicate the active site entrance

2.8 *in vivo* validation of AstD's AA resistance in yeast

A budding-yeast-based method utilizing AA inhibition of DHAD was developed in order to validate the AstD resistance to AA *in vivo* (Figure 11). The DHAD gene *ilv3* was knocked out using CRISPR/Cas9 from a *S. serevisiae* ΔURA3 strain, resulting in an Ile, Leu and Val (ILV) auxotrophic strain UB02, in addition to uracil auxotroph. Uracil was used as a selection marker for plasmid transformation. When *A. terreus* DHAD (fDHAD) or *astD* was episomally introduced to UB02, the growth on ILV-deficient media of the respective resulting strains, fDHAD and AstD, was recovered, compared to empty vector transformation (Figure 11. A). Further *in vivo* IC_{50} assays support the AA inhibition of yeast growth on ILV deficient media, with IC_{50} of 11.2 ± 0.452 μM and >1000 μM, for fDHAD and AstD respectively (Figure 11. B)

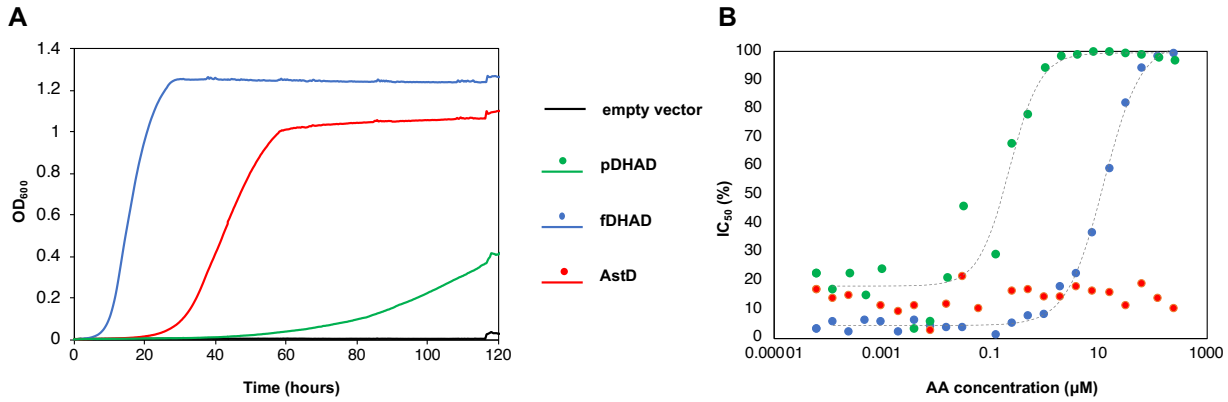


Figure 11. *in vivo* AA inhibition assays in yeast. (A) Growth curve of yeast transformed with fDHAD, AstD or pDHAD expressing plasmids. (B) *in vivo* AA growth inhibition of yeast transformed with fDHAD, AstD or pDHAD expressing plasmids.

When pDHAD was episomally introduced to UB02 for selection, the yeast growth was significantly delayed compared to fDHAD (Figure 11. A), which could have been caused by pDHAD codon incompatibility in yeast. The *in vivo* IC_{50} of pDHAD was 0.169 ± 0.0456, which was comparable to its *in vitro* IC_{50} at 0.249. This system can be further developed for pDHAD directed-evolution efforts.

2.9 Preliminary *in planta* AA resistance of mutant pDHADs

Although transgenic plants have been used in agriculture for more than four decades since the introduction of glyphosate, the issue of transgenic plants for human consumption has been met with a constant fear from the public[60–62]. Moreover, the insertion of a transgene into a food crop resulting in the expression of foreign proteins that differ significantly in structure, function, or quality from natural plant proteins is subject to more stringent regulations under the Federal Food, Drug, and Cosmetic Act (FFDCA)[63,64]. In July 2018, the EU court has declared that gene-edited plants fall under the same laws as GMOs, suggesting more restrictive regulations compared to that of traditionally bred plants. In contrast, FDA regulates genetically-modified (GM) crops for human consumption as "substantially equivalent" to non-GM crops and deems it unnecessary for pre-market approval, as long as the foreign gene does not differ significantly from natural plant proteins. Hence, it is desirable that we use gene-editing techniques rather than introducing foreign genes such as AstD into crops when developing AA into a commercial herbicide.

Homologous-recombination-directed repair of CRISPR/Cas9 mediated double-strand breaks in eukaryotic genome proceeds efficiently in lower eukaryotes. However, low-efficiency repair and poor availability of donor DNA impede the use of the method in higher eukaryotes, such as plants[65]. In recent reports, researchers have successfully edited bases by fusing base-editing enzymes to Cas9 nickases[66–68] and described different fusion enzymes being utilized for different base edits, and reporting detailed protocol for targeted mutagenesis in rice and wheat.[69] High efficiency of CRISPR/Cas9-mediated mutations were accomplished in the model plants *Arabidopsis thaliana* and rice, and a successful insertion of a selectable marker in endogenous ALS locus [70] attests to the feasibility of endogenous pDHAD editing for increasing aspterric acid resistance.

Unfortunately, since base-editing techniques are not sufficiently developed to introduce the precise base-edits needed for generating the amino acid mutations summarized in Table 1, we opted to re-introduce the mutant pDHADs using the same technique used for generating transgenic plants. However, because the gene that is being introduced is plant-derived, we rationalized that it should be comparable to direct gene-editing in plants. The mutant pDHADs (I177F, I177W, V178F, V178W, V496F, V496W, V497F and V497W) were cloned into the pEG202 based vector, which encodes for glufosinate resistance gene. Separate plasmids expressing wild-type pDHAD and AstD were used as controls. The plasmids were introduced to the model plant *A. thaliana* via *Agrobacterium*-mediated transformation, and correct transformants were selected on glufosinate-containing agar plates. We were able to acquire three plant lines for pDHAD transgene. As for the mutant pDHADs, we obtained two plant lines for each of V178F, V178W, V496F, V496W, and V497W. We also acquired three lines for V497F and 11 lines for I177F. After the transgenic plants were confirmed through sequencing, their seeds were plated on AA-containing (0, 0.1, 0.25, 0.5, 1 and 5 μM) agar plates and their growth were observed for 14 days after germination. The plant fresh-weights were measured after 14 days and their fresh-weight per plant is listed in Table S3. Figure 12 shows the average fresh-weight percentage of the mutant lines plated on different concentrations of AA, compared to their fresh-weight on agar plates not containing AA.

Although the fresh-weight percentage of the mutant pDHAD lines were higher than that of the wild-type pDHAD at all AA concentrations, they are still inhibited by AA i*n planta* at 5 μM AA. whereas the AstD was not inhibited (Figure 12. A). This result is consistent with the *in vitro* and *in vivo* inhibitions observed in Sections 2.6 and 2.8. The highest AA resistance was observed with V496F lines, where it retained 93% of its fresh-weight at 1 μM AA, compared to pDHAD which retained 52% of its fresh-weight. We start observing plant inhibition at concentrations as low as 0.25 μM (Figure 12. D), indicating the high concentration of 50 μM used in our previous study may not be necessary. However, in order to draw conclusion, we need more mutant lines.

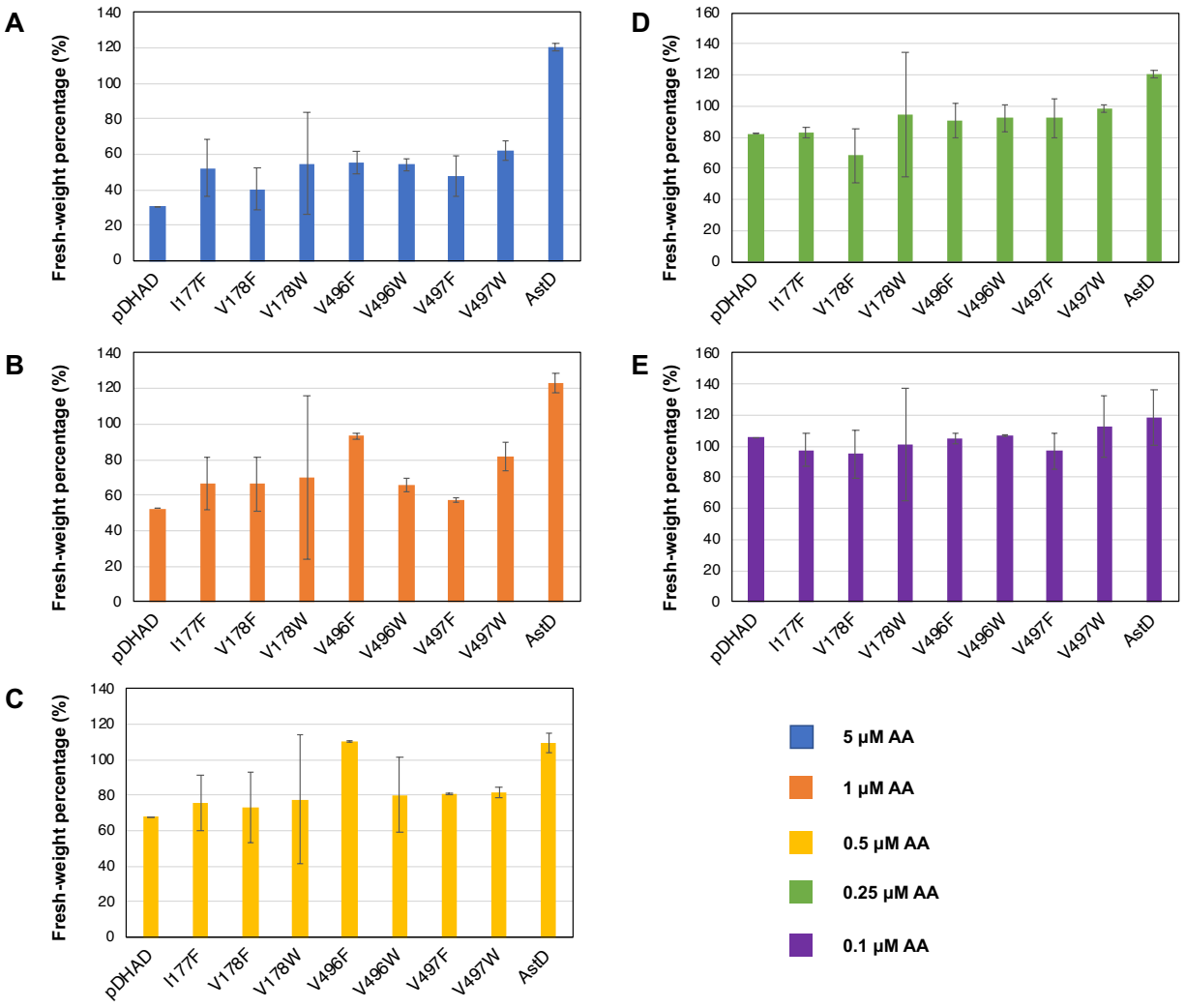


Figure 12. Fresh-weight percentage of mutant pDHAD transgene harboring *A. thaliana*

2.10 Conclusions and future directions

In Section 2, we aimed to elucidate the AA resistance mechanism of AstD. We obtained the co-crystal structure of pDHAD-AA complex, which allowed us to build a homology model of AstD. As a result, were able to deduce amino acids that may be responsible for AstD AA resistance. Our kinetic analysis of the mutant pDHADs *in vitro* confirmed our hypothesis that larger hydrophobic amino acids at locations surrounding the pDHAD active-site entrance increase the AA resistance of pDHAD. However, it also showed us that resistance comes with the

compromise of catalytic activity, exemplified by the decrease in k_{cat}. While analyzing the data we acquired, we also found another position which may potentially contribute to the AA resistance in AA, which is pDHAD V188 to Ile mutation located on one of the α-helices that make the active site entrance. Thus, this amino acid location should be examined via point mutations. In addition, the mutations which increased the AA resistance without compromising catalytic activity (V187L, V178I, V496L and V496I) should be combined in order to increase AA resistance of pDHAD. Furthermore, our *in vitro* and i*n vivo* analysis showed that there are other contributing factors to AstD's extremely high AA resistance other than the point mutations observed in AstD compared to pDHAD. Thus, our results signify how Nature is brilliant at solving NP resistance without fully compromising enzyme activity.

In order for us to fully understand what other mutations may be contributing to the resistance, we may need to pursue directed evolution methods, based on our yeast platform. Directed evolution is a well-established method used for increasing enzymatic activity, selectivity and stability for myriad of applications, spanning from industrial to purely scientific[71]. By generating pDHAD random mutagenesis library through error-prone PCR (EP-PCR), and episomally introducing them to UB02, we can select for mutants with increased activity and AA resistance.

3 HARZIANOPYRIDONE BIOSYNTHESIS

Since its first isolation in 1989, the biosynthesis of harzianopyridone, a potent mitochondrial complex II inhibitor, has remained elusive. In this section, we utilized our SRE-DGM and general retro-biosynthetic analysis to identify the BGC and harzianopyridone biosynthetic pathway. In doing so, we hoped to identify enzymes that are responsible for the unique formation of harzianopyridone's penta-substituted 2-pyridone core structure.

3.1 Mitochondrial complex II enzymes and its inhibitors

Oxidative phosphorylation (OXPHOS) is an energy generating process that takes place in the inner membrane of prokaryotic and the mitochondria of eukaryotic organisms. Due to its similarity and prevalence, it has attracted researchers to target the complexes in the system for antifungal and anticancer therapies.[72] The OXPHOS consists of five complexes, NADH-ubiquinone oxidoreductase (Complex I), succinate-ubiquinone oxidoreductase (Complex II), ubiquinol–cytochrome-C reductase, cytochrome bc_1 complex (Complex III), cytochrome-c oxidase (Complex IV) and ATP synthase (Complex V), which generates ATP via oxidative phosphorylation.[73] Although many NPs have been identified to inhibit complex I, III and V, only a handful have been identified to inhibit complex II and IV.[74]

Mitochondrial complex II (CII), a.k.a., succinate-ubiquinone oxidoreductase (SQR), succinate dehydrogenase (SDH) participates in both the OXPHOS and TCA cycle and is a homodimer of four subunit tetramer complex (Figure 13. A). SDHA, a flavoprotein, contains a covalently bound FAD and the succinate binding site where succinate gets oxidized to fumarate with the help of FAD. SDHB contains three iron-sulfur clusters: [2Fe-2S], [4Fe-4S], and [3Fe-4S], where the electrons from $FADH_2$ is eventually transferred to ubiquinone. SDHC and SDHD are hydrophobic subunits that form the transmembrane portion of CII, with heme b and ubiquinone

binding sites. In these subunits the electron from $FADH_2$ is transferred to ubiquinone forming ubiquinol as the end product.

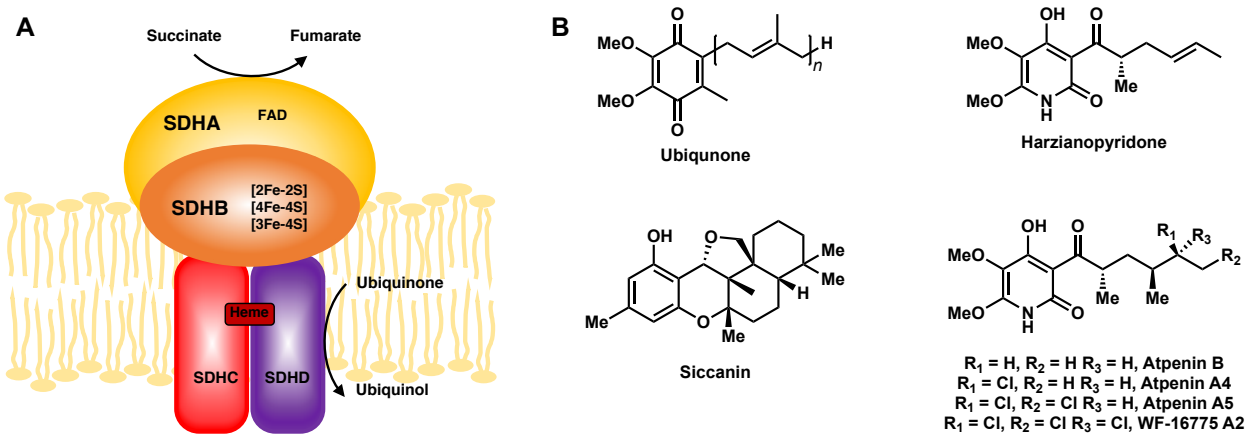


Figure 13. Mitochondrial complex II and its fungal inhibitors. (A) CII subunits; SDHA:flavoprotein, SDHB: iron-sulfur cluster containing subunit, SDHC and SDHD: hydrophobic transmembrane subunits (B) CII electron acceptor ubiquinone and fungal CII inhibitor NPs

So far, three fungal NPs have been identified to specifically inhibit mitochondrial complex II (CII), namely harzianopyridone, atpenin-related compounds (atpenin A4 and A5) and siccanin (Figure 13. B). The structurally related 2-pyridones, harzianopyridone and atpenin A5[75] show potent inhibitory activity against mammalian CII at nanomolar concentration.[76] Structures of these compounds closely match that of the electron carrier ubiquinone with their penta-substituted pyridine cores.[77] The C5 and C6 methoxy groups in harizianopyridone and atpenin-related compounds match to those present in ubiquinone, and are not found in any other 2-pyridone natural products.[78] The CII co-crystal structure have been solved with ubiquinone, where atpenin A5 docking revealed that it occupies the same physical space as the natural electron acceptor ubiquinone, confirming its competitive inhibition.[77]

Because of the exceptionally potent activities of harzianopyridone and atpenin A5, these compounds have been the subjects of numerous total synthesis efforts,[79–82] as well as structure-activity-relationship studies.[83,84] The synthetic strategy relies on the C3-C7 coupling between a fully functionalized organometallic pyridine nucleophile with the corresponding acyl aldehyde, both requiring lengthy synthetic steps to prepare. Additionally, the aliphatic chlorines of atpenin

A5 is proposed to be important for the activity since its potency increases with the level of chlorination. Another two compounds that share the same chromophore are WF-16775 A1 and A2. Although they have not been tested for their complex II inhibitory activities, they display potent angiogenesis inhibition.[85]

3.2 Identifying atpenin-related compounds and harzianopyridone BGCs with SRE-DGM

Combining the knowledge of general 2-pyridone biosynthesis and CII protein sequences, we were able to identify BGCs that are potentially responsible for producing harzianopyridone and atpenin-related compounds. Based on the biosynthesis of other 2-pyridone natural products such as tenellin,[86] as well as isotope feeding studies of harzianopyridone,[87] both harziaopyridone and atpenin-related compounds are proposed to derive from a polyketide-amino acid containing tetramic acid, which is produced by a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS). In the pathways of tenellin[88] and leporin,[89] a ring expansion P450 ($P450_{RE}$) catalyzes a radical-mediated ring expansion to afford a 4-hydroxy-3-acyl-2-pyridone intermediate. Subsequent hydroxylations and methylations catalyzed by flavoenzymes and methyltransferases, respectively, can be proposed to install the C5- and C6-methoxy groups.

We searched for BGCs that contain the potential biosynthetic enzymes in addition to CII enzymes, in the genomes of *Trichoderma harzianium* and *Penicillium oxalicum*, which are known producers of harzianopyridone and atpenin A5, respectively. With this approach, we were able to identify *apn* BGC from *P. oxalicum* and other homologous BGCs from public database and privately sequenced genome (Figure 14). In addition, we were also able to identify a potential BGC from *T. harzianium* that contains most of the genes in the *apn* BGC, potentially responsible for producing harzianopyridone and renamed it *har* BGC.

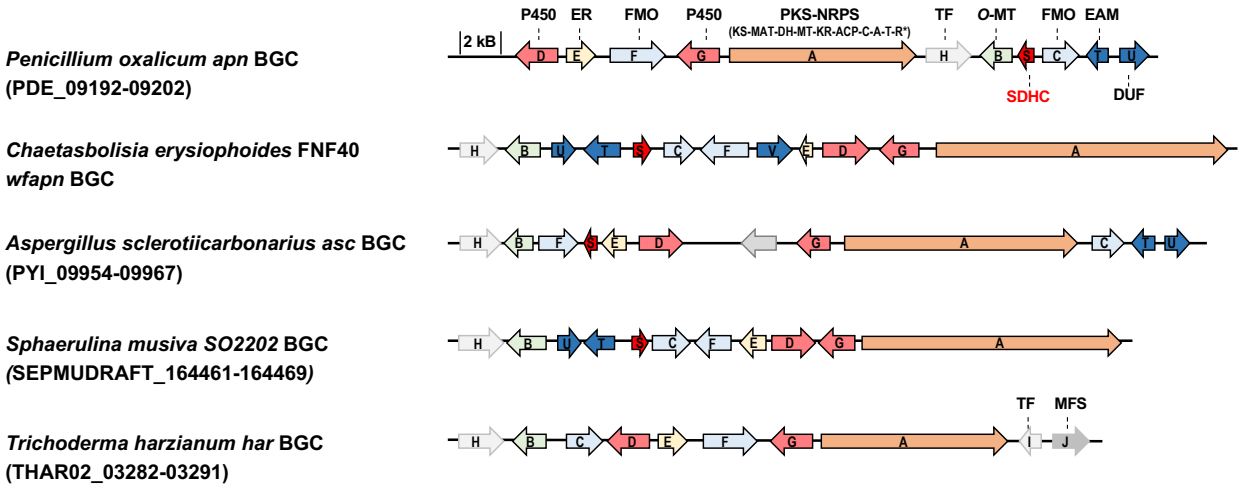


Figure 14. *har* and *apn* BGCs and homologous BGCs. Abbreviations: KS: ketosynthase; MAT: malonyl-CoA:ACP acyltransferase; DH: dehydratase; MT: methyltransferase; KR: ketoreductase; ACP: acyl-carrier protein; C: condensation; A: adenylation; T: thiolation; and R*: Dieckmann cyclization; TF: transcriptional factor; MFS: major facilitator superfamily transporter

3.3 Heterologous expression of *har* BGC in *Aspergillus nidulans*

The *har* BGC contained homologs of both PKS-NRPS (HarA) and $P450_{RE}$ (HarG) (Figure 14 and Table 3). Also encoded in the gene cluster are the trans-enoylreductase (ER) (HarE) that partners with PKS-NRPS, two flavin-dependent monooxygenases (FMOs) (HarC and HarF), an additional P450 (HarD), and an *O*-methyltransferase (*O*-MT) (HarB). HarD is predicted to be a *N*-hydroxylase based on homology to characterized enzymes. [90] The functional predictions of the genes in the cluster are therefore consistent with the structural features in harzianopyridone 1, with the exception of the putative *N*-hydroxylase, HarD. Detailed bioinformatic analysis is found in Table S3.

We investigated the metabolites produced by *har* BGC using *Aspergillus nidulans* A1145 ΔEM[91] as a heterologous expression host (Figure 15). When HarA-F were coexpressed, we detected biosynthesis of several new metabolites compared to *A. nidulans* harboring empty vectors (Figure 14. A, i and vii). We confirmed that **1** is biosynthesized at 0.5 mg/L titer (Figures S31-S35, Table S4).

Table 3. Putative functions of genes encoded in *har* BGC

Gene name	Proposed protein function
harA	Polyketide synthase – nonribosomal peptide synthetase (PKS−NRPS)
harB	*O*-Methyltransferase (OMT)
harC	Flavin-dependent monooxygenase (FMO)
harD	*N*-hydroxylation Cytochrome P450 (P450)
harE	trans-Enoyl reductase (ER)
harF	Flavin-dependent monooxygenase (FMO)
harG	Cytochrome P450 (P450)
harH	Transcription factor (TF)
harI	Transcription factor (TF)
harJ	Major facilitator superfamily transporter (MFS)

Also produced are 5-phenyl-2-pyridone **3** at 5 mg/L (Figures S41-S45, Table S6), and two minor metabolites **6** and **7** that are *N*1-methoxylated 2-pyridones. Whereas **6** is not substituted at C5 and C6 (2.5 mg/L, Figures S56-S60, Table S9), **7** contains a C5-methoxy group (3 mg/mL, Figures S61-S65, Table S10). The presence of *N*-methoxy in **6** and **7** was surprising considering **1** does not contain this modification, although this agrees with the role of the predicted *N*-hydroxylase HarD.

Different combinations of *har* BGC genes were then expressed in *A. nidulans* to probe individual enzyme functions (Figures 15 and S5). When HarA and HarE were coexpressed, the tetramic acid **2** was produced at 7.5 mg/L (Figure 15. A, ii, Figures S36-S40, Table S5), consistent with the functional assignment of these two enzymes. This shows L-tyrosine is incorporated by the NRPS module of HarA, and a phenol cleavage step[92] is required to give **1**. When HarG, the proposed $P450_{RE}$ was coexpressed with HarA and HarE, **3** and **4** were formed (Figure 15. A, iii).

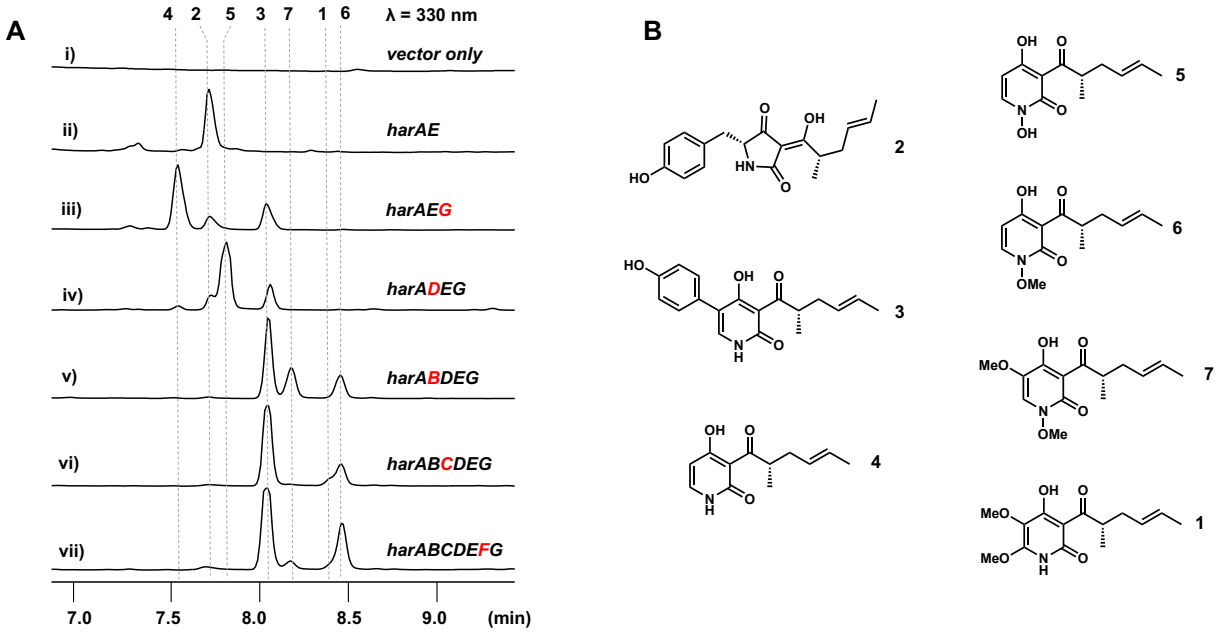


Figure 15. Heterologous expression of har BGC genes in *A. nidulans*. (A) Liquid chromatogram (LC) analysis of different combinations of har BGC gene expression. (B) Structures of compounds detected by the LC.

Compound **4** was structurally verified to be the dephenylated 2-pyridone (7.5 mg/L, Figures S46-S50, Table S7). The co-emergence of **3** and **4** suggests that HarG is responsible for both ring phenyl-cleavage steps. We verified that **3** cannot undergo phenyl cleavage when supplied to *A. nidulans* expressing HarG, which indicates this compound is a shunt product. Cox and coworkers proposed a radical mechanism for the ring expansion of tetramic acid in tenellin biosynthesis, during which the 6-hydroxy-dihydropyridone such as 8 is a proposed intermediate.[92] Subsequent dehydration of **8** can produce **3**. We proposed that if the $P450_{RE}$ heme-iron can be reduced prior to dehydration of 8, the enzyme can further catalyze oxidation of the phenyl ring of **8**, which can lead to loss of quinone and give **4** (Figure S6).

We next determined which remaining enzymes install the methoxy groups in **4** to give **1**. Only coexpression of the P450 HarD with HarAEG led to transformation of **4** to the *N*-hydroxy pyridone **5** (0.5 mg/L, Figures 15. A, iv and S51-S55, Table S8). Further coexpression of *O*-MT HarB led to the emergence of **6** and **7** (Figure 15. A, v). While methylation of **5** to **6** fits with the

predicted function of HarB, methoxylation of C5 in **6** to give **7** is not expected. Based on yeast biotransformation and *in vitro* assays described in Sections 3.4 and 3.5, we conclude formation of **7** here may be due to crosstalk with *A. nidulans* endogenous enzymes. Finally, when the FMO HarC was expressed together, formation of **1** was observed (Figure 15. A, vi). The other FMO HarF is not essential in the reconstitution of **1**, which is in agreement with the observed ancillary role of the homologs in other pyridone pathways.[92] Therefore, HarABCDEG represents the minimal set of enzymes needed to biosynthesize 1 (Figure 15. A).

3.4 Feeding studies in *Saccharomyces cerevisiae*

In order to examine whether **6** and **7** are pathway intermediates, we performed biotransformations using *Saccharomyces cerevisiae* (Table 5, Figures 15-18). 33 μM of compounds **4**-**7** were individually fed to yeast expressing combinations of HarB, HarC and HarD, extracted after 12 hours and analyzed by LC/MS. All four compounds are shown to be biosynthetic intermediates, as each can be biotransformed into **1** (Table 4. entries 3, 7, 11 and 14).

Table 4. Yeast biotransformation of biosynthetic intermediates[a]

entry	substrate	HarD	HarB	HarC	product(s)
1	4	+	-	-	5
2	4	+	+	-	6
3	4	+	+	+	**1** (>95%), **7**
4	4	-	+	+	No conversion
5	5	-	+	-	6
6	5	+	+	-	6
7	5	-	+	+	**1** (>95%), **7**
8	5	-	-	+	No conversion
9	6	-	-	+	10
10	6	-	+	-	No conversion
11	6	-	+	+	1 (>90%), 7, 10
12	7	-	-	+	No conversion[b]
13	7	-	+	-	No conversion
14	7	-	+	+	1

[a]All substrates fed at 33 μM to *S. cerevisiae* expressing the indicated combination of enzymes. [b]the substrate was consumed, but no products were detectable by LCMS.

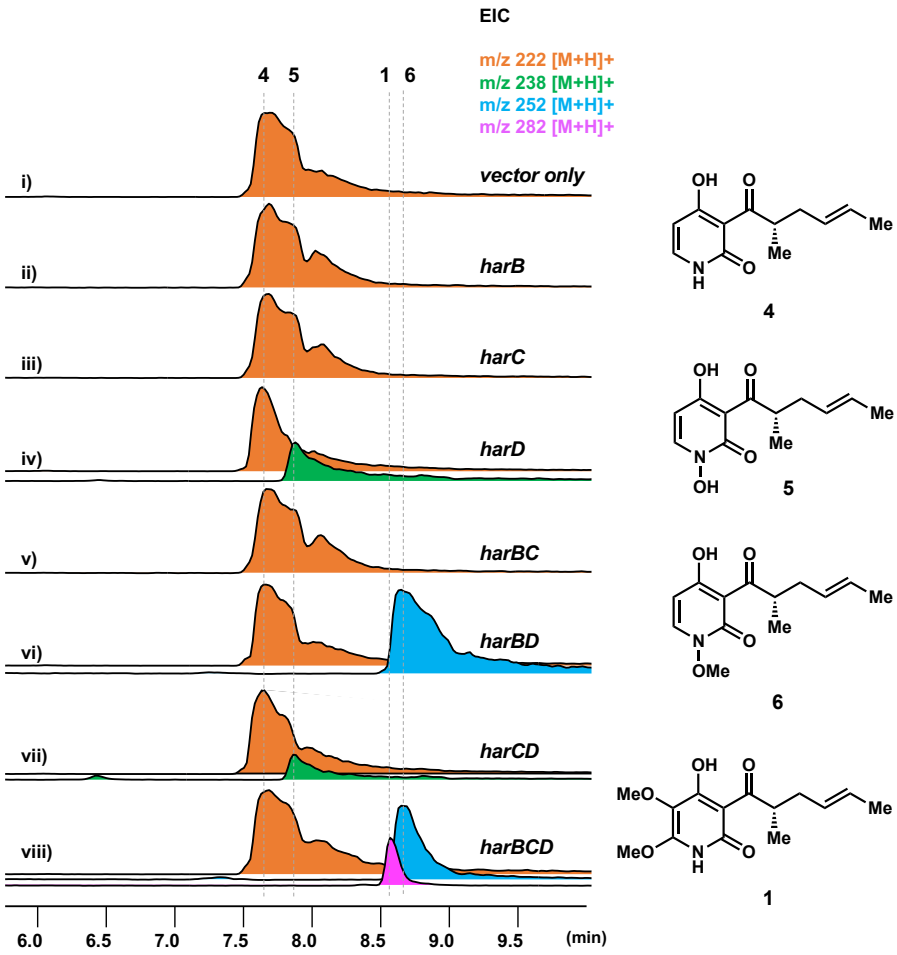


Figure 16. LC/MS analysis of extracts from feeding compound 4 to *S. cerevisiae* expressing different combinations of *har* genes. Yeast cultures fed with 33 μM compound 4 was extracted after 12 hours and concentrated.

Compound **4** was exclusively oxidized to **5** by HarD (Table 4. entry 1, Figure 16. trace iv), which was methylated by HarB to form **6** (Table 4. entries 2 and 5, Figure 16. trace vi). *N*-methoxylation is required for the formation of **1**, as **4** was not consumed by HarB and HarC, or their combination without HarD (Table 4. entry 4, Figure 16. traces ii, iii, v, and vii), nor was **5** oxidized in the presence of HarC only (Table 4. entry 8, Figure 16. trace iii).

The biotransformation of **5** into **1** by HarB and HarC demonstrates that HarB is responsible for all *O*-methylation reactions (Figure 17. traces ii and v), while HarC is required for both C5 and C6 hydroxylations (Figure 17. trace v). HarC was required for the transformation of **6** to **7** and **1** (Figure 17. Traces v and vi). This suggests the emergence of **7** in *A. nidulans* expressing ABDEG

(Figure 15. trace v) was likely due to endogenous monooxygenases. We also observed a new compound **10** with m/z (+) 286 emerges with **6**, **7** and **1**, when **5** was fed to yeast expressing HarB and HarC (Figure 17. trace v). Although, 10 was proposed to be a C5, C6 di-hydroxylated dihydropyridone (Figure 20), with feeding **5** to yeast, we did not know from which intermediate **10** was derived from, **5** or **6**. We were also unable to characterize **10** since it was readily degraded during purification.

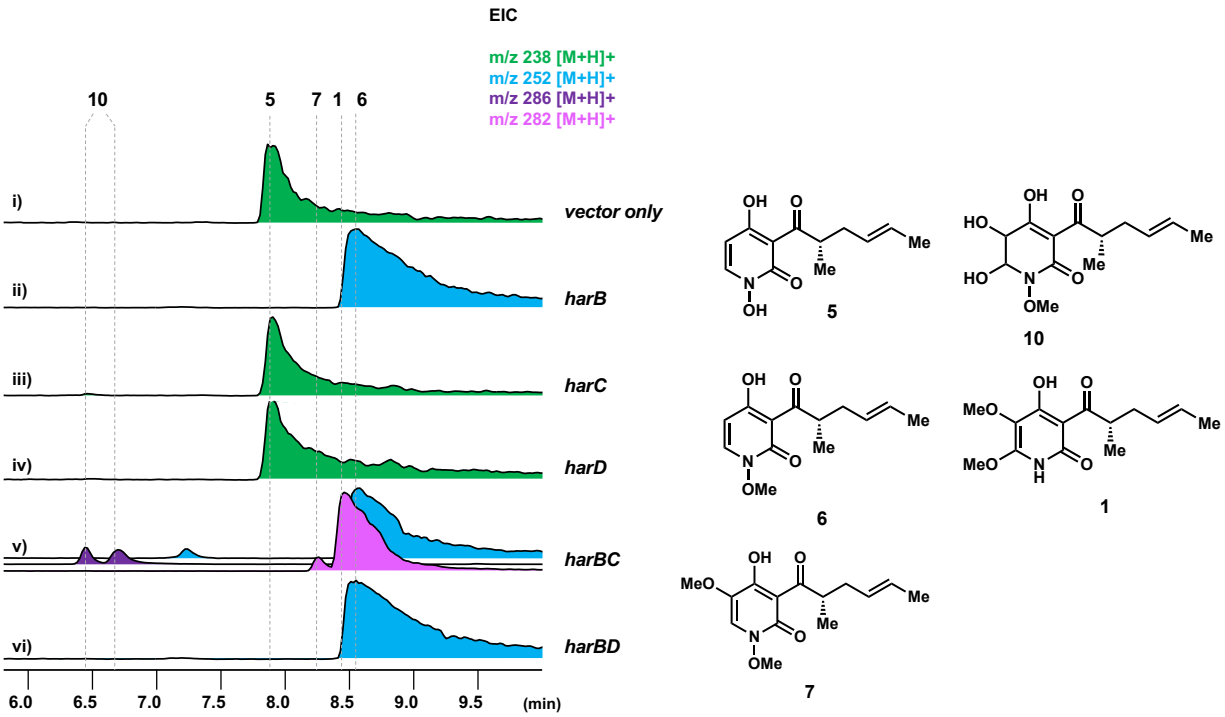


Figure 17. LC/MS analysis of extracts from feeding compound 5 to *S. cerevisiae* expressing different combinations of *har* genes. Yeast cultures fed with 33 μM compound 5 was extracted after 12 hours and concentrated.

We were able to determine **10** was the product of HarC acting on compound **6** (Figure 18. trace iii). Since **10** does not get consumed with the inclusion of HarB, we assume it is a shunt product (Figure 18, trace v). Lastly when **7** was fed to yeast expressing only HarC, no product could be identified although **7** was consumed (Table 4. entry 12, Figure 19. trace iii), indicating formation of an intermediate that degrades in the absence of HarB.

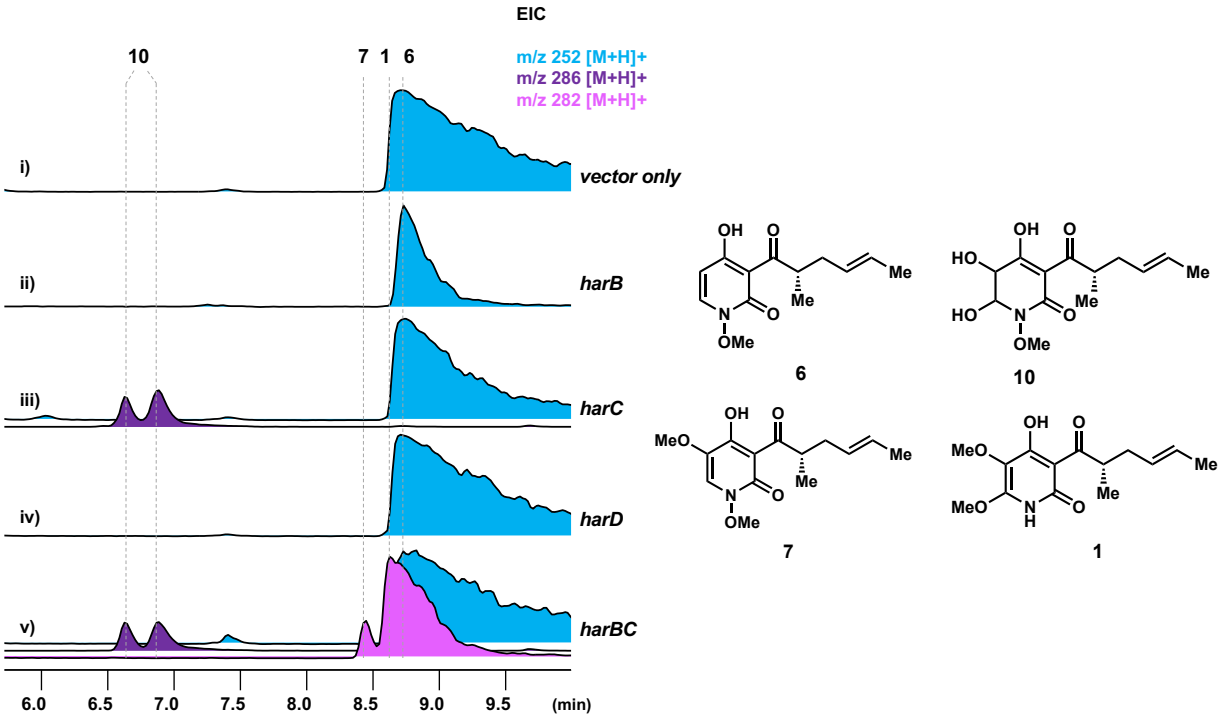


Figure 18. LC/MS analysis of extracts from feeding compound 6 to *S. cerevisiae* expressing different combinations of *har* genes. Yeast cultures fed with 33 μM compound 6 was extracted after 12 hours and concentrated.

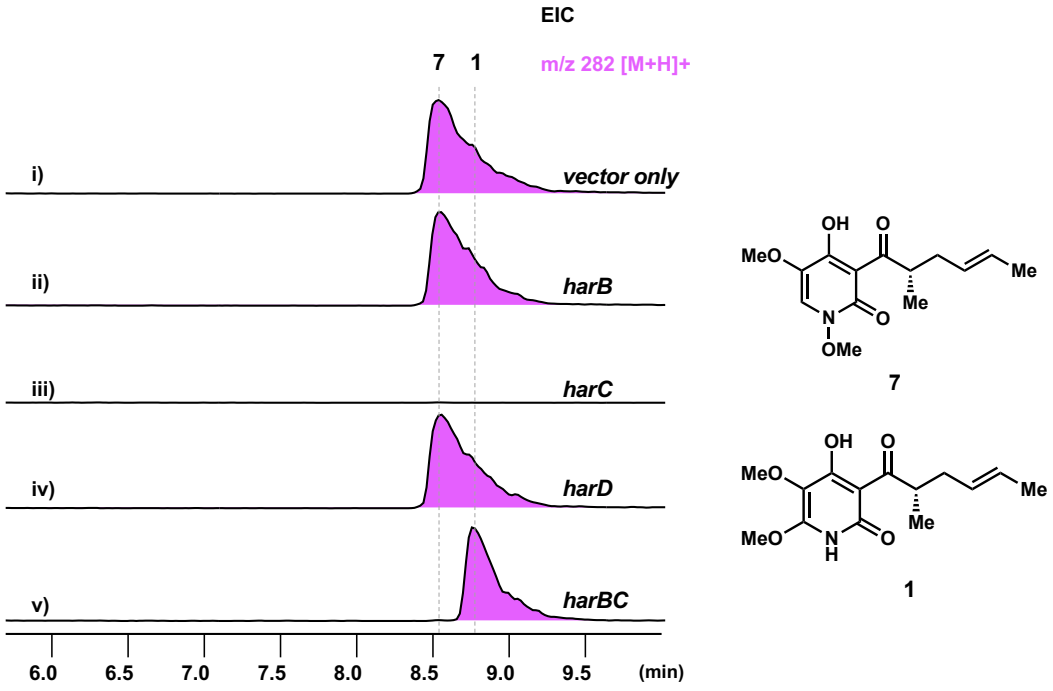


Figure 19. LC/MS analysis of extracts from feeding compound 7 to *S. cerevisiae* expressing different combinations of *har* genes. Yeast cultures fed with 33 μM compound 7 was extracted after 12 hours and concentrated.

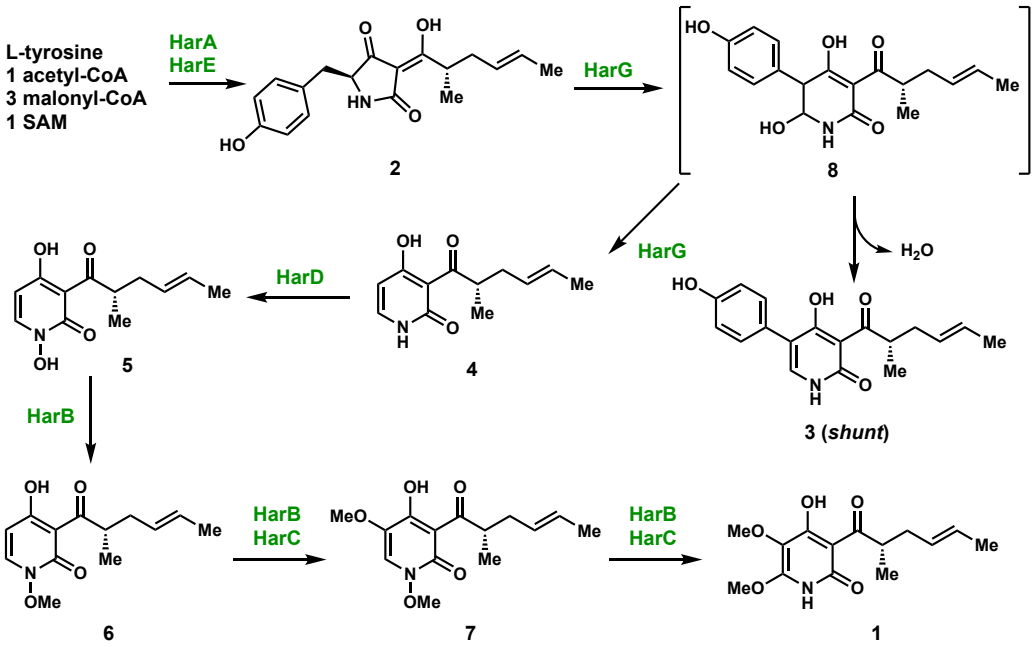


Figure 20. Proposed biosynthetic pathway of harzianopyridone.

Thus, combined with the *A. nidulans* heterologous expression, we were able to propose the biosynthetic pathway of harzianopyridone. The biosynthesis starts with the HarA and HarE producing the tetramic acid **2**, which then undergoes a ring-expansion, dephenylation reaction catalyzed by HarG. The dephenylated intermediate **4** gets *N*-hydroxylated for the biosynthesis to proceed, which is an unusual transformation, since harzianopyridone pyridone nitrogen is not substituted. The hydroxylated intermediate **5** then goes under an iterative biocatalysis performed by HarB and HarC to give the final product harzianopyridone **1** (Figure 20).

3.5 Mechanistic studies of unusual iterative catalysis in harzianopyridone biosynthesis

In order to investigate the iterative activities of HarB and HarC, we performed *in vitro* assays using purified enzymes (Figure S7). Starting with **6**, HarB and HarC together were able to produce **1**, with a small amount of **7** detected (Figure 21). The reaction requires both NADPH and SAM (Figure S8). The NADPH is required for regeneration of reduced flavin after the hydroxylation reaction. Adding HarC alone to **6** led to decrease in **6** and formation of **10**.

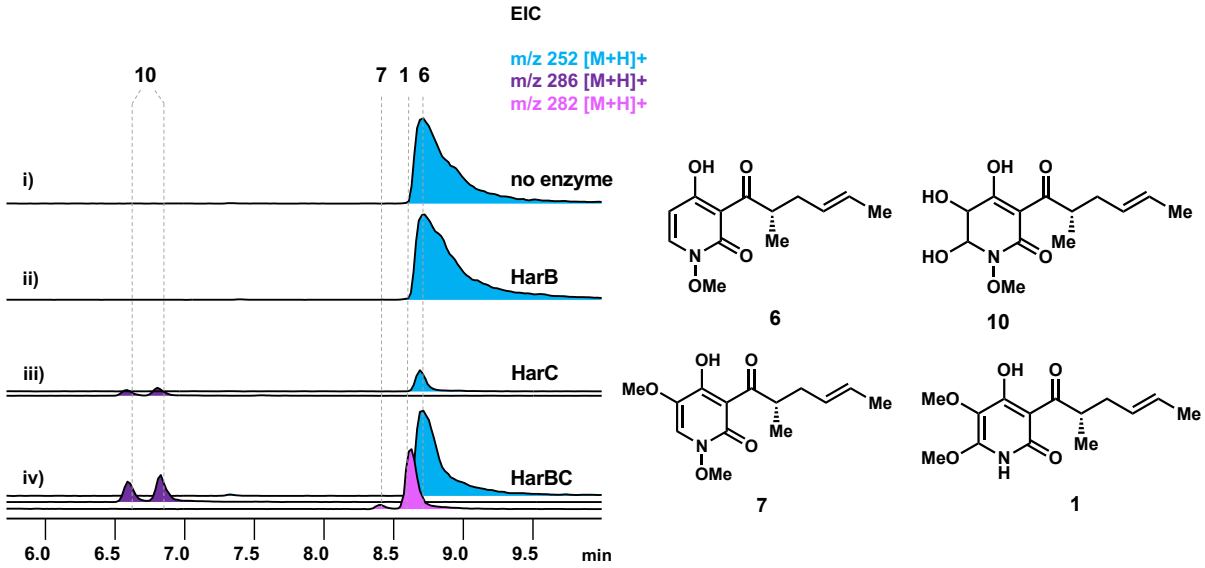


Figure 21. *in vitro* HarB and HarC enzyme requirements to produce 1 with compound 6 as substrate. The reaction mix contained 10 μM HarB and/or 5 μM HarC, 50 μM FAD, 1 mM NADPH, 100 μM SAM and ~100 μM compound 6. The reactions were incubated at 30°C for 2 hours, quenched with equal amount of MeCN to terminate reaction.

In vitro conversion of **7** to **1** by HarB and HarC was complete within 20 minutes (~20 turnovers) as shown in Figure 22, and also requires both NADPH and SAM (Figure S9). **7** was nearly completely consumed by HarC, with no detectable product, indicating the formation of an unstable intermediate. The enzyme-catalyzed transformations of **6** and **7** to **1** are oxygen dependent, as reactions performed under anaerobic conditions did not lead to product formation (Figures S10-S11). Anaerobic reactions were performed in a glove box and the method is explained in detail in Section. 6.

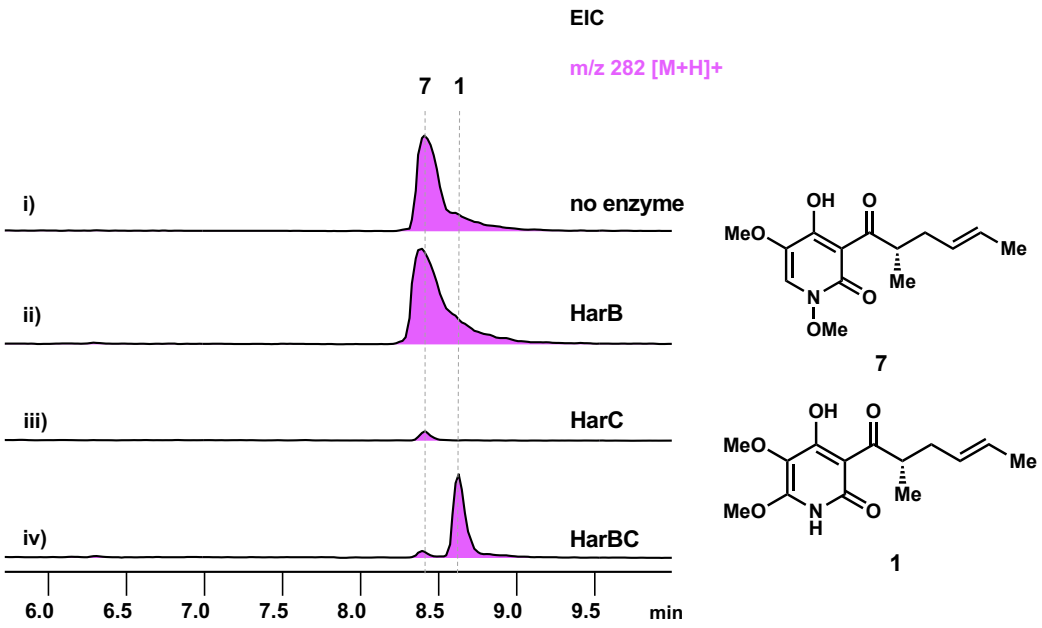


Figure 22. *in vitro* HarB and HarC enzyme requirements to produce 1 with compound 7 as substrate. The reaction mix contained 10 μM HarB and/or 5 μM HarC, 50 μM FAD, 1 mM NADPH, 100 μM SAM and ~100 μM compound 7. The reactions were incubated at 30°C for 2 hours, quenched with equal amount of MeCN to terminate reaction.

To investigate the origin of the C5 and C6 oxygen atoms in **1**, we performed the enzyme assays in the presence of $H_2{}^{18}O$. We observed the C7 ketone oxygen can exchange with $H_2{}^{18}O$ and lead to +2 MW (Figure S12. B). Therefore, after each reaction was completed in $H_2{}^{18}O$, the solvent was lyophilized, and the samples were re-equilibrated in in $H_2{}^{16}O$ before MS analysis (Figure S12. A).

From **6** to **1**, we observed a +2 increase in MW of 1, which indicates one of the C5 and C6 oxygens in **1** is derived from H_2O (Figure 22). The +2 increase was also observed in the conversion of **7** to **1** (Figure 23). These labeling results implicate that the first oxygen atom introduced into **6** to form **7** is from molecular oxygen, while the second oxygen introduced to convert **7** to **1** is from water instead of O_2 or a methoxy migration.

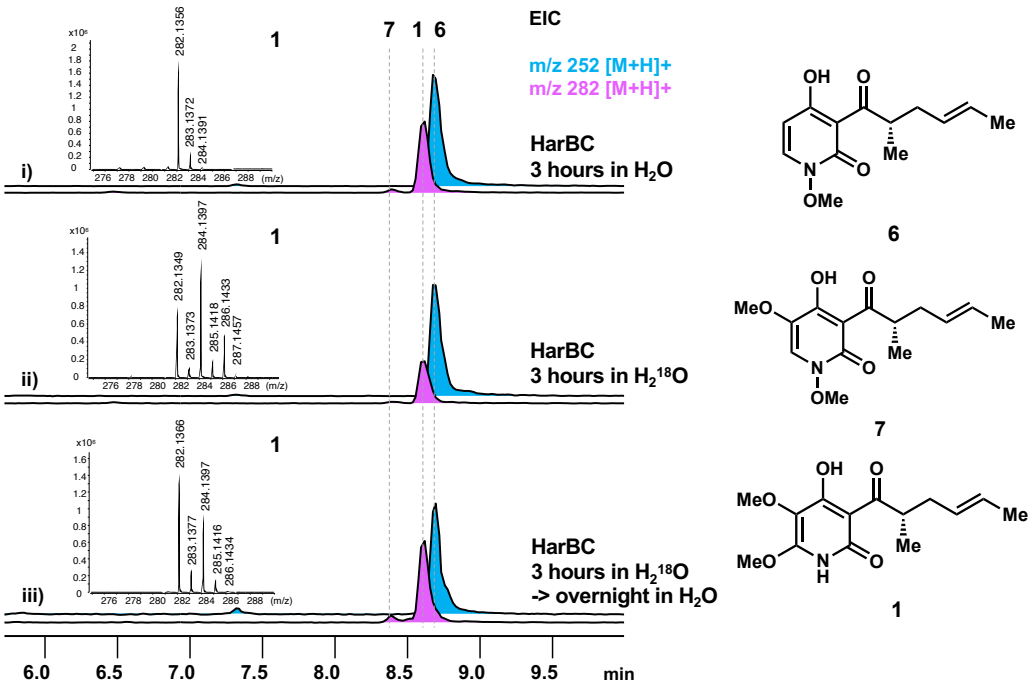


Figure 23. LC and QTOF analysis of *in vitro* HarBC reactions in $H_2{}^{18}O$ with compound 6 as substrate: i) 3 hours HarBC enzymatic reaction in $H_2{}^{16}O$; ii) 3 hours HarBC enzymatic reaction in $H_2{}^{18}O$; iii) After 3 hours-$H_2{}^{18}O$ reaction, the reaction supernatant was lyophilized and placed in $H_2{}^{16}O$ overnight.

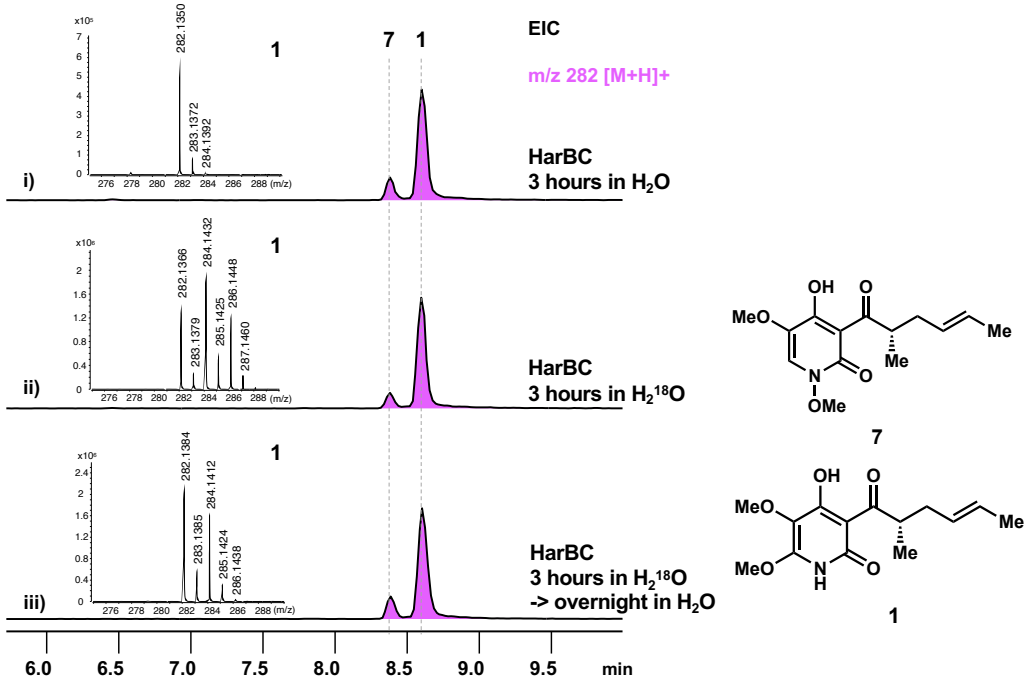


Figure 24. LC and QTOF analysis of *in vitro* HarBC reactions in $H_2{}^{18}O$ with compound 7 as substrate: i) 3 hour HarBC enzymatic reaction in $H_2{}^{16}O$; ii) 3 hour HarBC enzymatic reaction in $H_2{}^{18}O$; iii) After 3 hours-$H_2{}^{18}O$ reaction, the reaction supernatant was lyophilized and placed in $H_2{}^{16}O$ overnight.

Our analysis of the biosynthesis of **1** has revealed unexpected biosynthetic logic and raises mechanistic questions (Figure 20). For example, installation of the *N*-OMe group followed by its removal in the last step is an interesting strategy. One could propose this is a protective group to prevent *N*-methylation by the iterative HarB. However, assays in which **4** or **1** was incubated with HarB did not result in *N*-methylation (Figure 16). Alternatively, we propose *N*-methoxy group serves as a directing group that increases nucleophilic character of the nitrogen through the alpha effect.[93] The increased nucleophilicity of *N1* promotes electrophilic aromatic substitution (EAS) of **6** at C5 to give **9**, using flavin hydroperoxide as an electrophile. In the absence of HarB, **9** can form **10**, either via direct water attack at C6 or through an epoxide intermediate **11**. HarB could facilitate the aromatization of **9** to **12**, followed by C5-*O*-methylation to give **7** (Figure 25).

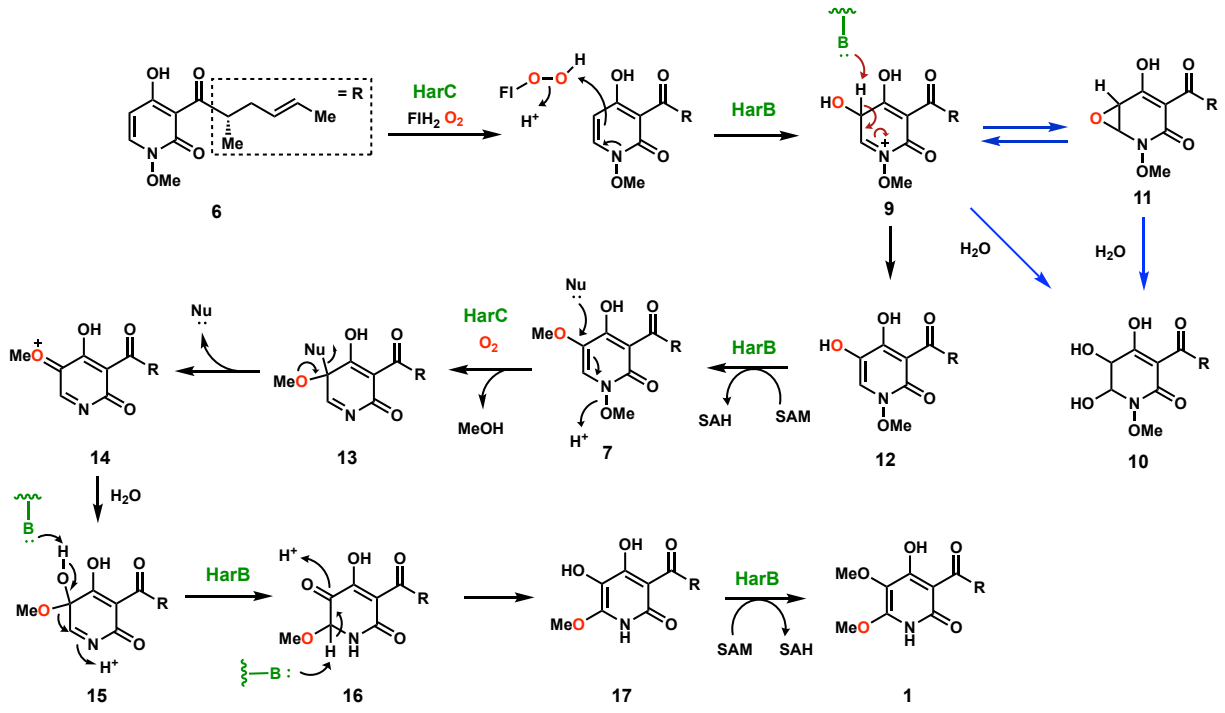


Figure 25. Proposed functions of FMO HarC and *O*-MT HarB in biosynthesis of **1**

Following EAS, the *N*-methoxy directing group can be removed as methanol via a reaction analogous to the Bamberger rearrangement.[94] It remains unresolved how the redox neutral conversion from **7** to **1** requires aerobic conditions, while the oxygenation originates from water. As shown in Figure 25, an active site nucleophile (Nu), which we proposed could be flavin peroxide (Fl-OO^-) to account for the O_2 requirement, can attack C5 to drive the ejection of methanol and yield **13**. Elimination of the Nu can form the labile cation **14**. HarB may facilitate the attack of water at C5 to form the hemiacetal **15**, followed by a semi-pinacol like rearrangement to **16**. **16** can rearomatize to **17** and be subjected to C5-*O*-methylation by HarB to give **1**. It may also be possible for a direct water attack at C6 of **14**, followed by aromatization and C6-methylation to give **1** (Figure 25).

3.6 CONCLUSION

In summary, our work revealed the enzymatic steps in the biosynthesis of harzianopyridone. The BGC of harzianopyridone was found with the help of SRE-DGM and enzymes HarABCDEG have been confirmed to be responsible for the biosynthesis of harzianopyridone. The biosynthesis started with the HarA and HarE producing the tetramic acid **2**, which then undergoes a ring-expansion, dephenylation reaction catalyzed by HarG. The dephenylated intermediate **4** gets *N*-hydroxylated for the biosynthesis to proceed, which is an unusual transformation, since harzianopyridone nitrogen is not substituted. The hydroxylated intermediate **5** then goes under an iterative biocatalysis performed by HarB and HarC to give the final product harzianopyridone **1.** Notably, four of the six enzymes (HarA, G, B and C) perform iterative catalysis, which underscores the highly programmed and unpredictable functions of fungal biosynthetic enzymes.

4 ELUCIDATION OF ATPENIN A5 BIOSYNTHESIS IN SEARCH FOR A NOVEL HALOGENASE

After elucidating the biosynthetic pathway of harzianopyridone in Section 2, we shifted our focus to identifying and elucidating the biosynthesis of atpenin A5. Not only atpenin A5 is a more potent mitochondrial complex II inhibitor, but chlorination on its non-activated acyl sidechain also makes it an ideal compound to pursue for the discovery of novel halogenase enzymes. In this section, we utilized our knowledge from Section 2 in identifying the BGC of atpenin A5 and elucidated the biosynthetic pathway in search for novel halogenase enzymes that acts on a non-activated sp^3 carbon.

4.1 Halogenated natural products and the significance of halogen substitutions

Organohalogen moieties (typically F, Cl, Br, I) are present in many commercial pharmaceuticals and agrochemicals, playing an important role in their bioactivity.[95] Few of these compounds are NPs, including the antibiotics vancomycin and chloramphenicol, anti-cancerous rebeccamycin and the mycotoxin aspirochlorine (Figure 26). The significance of the halogen moieties in these compounds have been the subject to many studies, including biosynthetic studies, structure – activity relation studies and protein interaction. For example, when vancomycin is dechlorinated, its antimicrobial activity decreases significantly due to the reduced binding activity to its biomolecular target.[96]

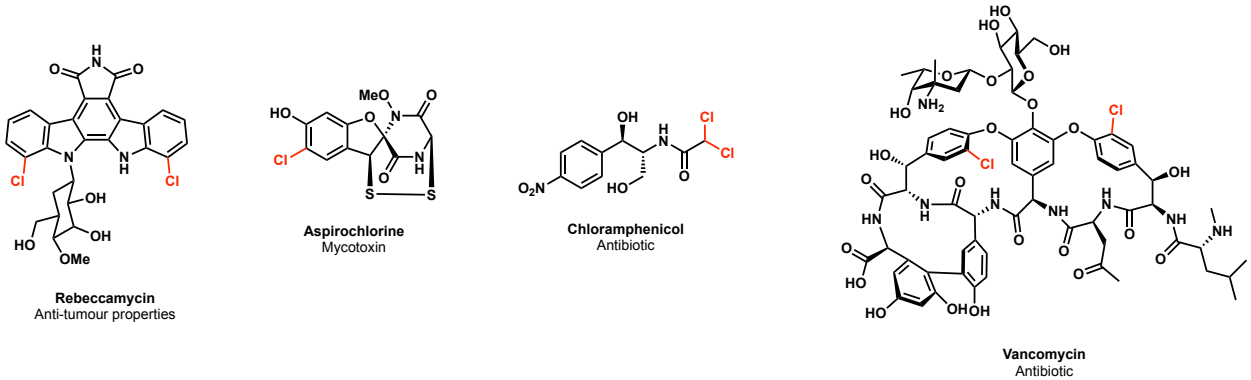


Figure 26. Representative NPs with halogen moieties. The halogen atoms are depicted in red.

The impact of halogens in the bioactivity of NPs are thought to be brought on by their modulation of lipophilicity and nonspecific hydrophobic interactions with protein targets,[97,98] in addition to the intermolecular interactions with target proteins called halogen-bonds.[97–100] This halogen-bond is explained by the electron-deficient “sigma-hole” of the halogen that allows the halogen to interact with the lone pairs of heteroatoms like N, O and S in protein residues in a manner that is similar to hydrogen bonds.[101]

4.2 Halogenation enzymes

Due to their prevalence and importance of halogens in pharmaceuticals, numerous organic synthesis strategies have been developed.[102–104] However, such methods often generate non-specific halogenation and require multiple synthetic steps and harmful agents. In the other hand, halogenation enzymes (halogenases) have been evolved to catalyze stereo- and regio-specific halogenations in mild conditions. Indeed, because of their importance the discovery and study of halogenases have been the focus of many researchers, resulting in the discovery four types of halogenation enzymes.

First is the haloperoxidase enzymes, which are divided into heme-iron-dependent haloperoxidases and vanadium-dependent haloperoxidase.[95,105,106] The general mechanism of haloperoxidases is the same, where the reactive hypohalous acid (HO-X) is generated through oxidation of the halide (H-X) and is dependent on hydrogen peroxide. In heme-iron-dependent haloperoxidase active sites, the halide ion is oxidised to ferric hypohalite by the ferryl-oxo species. This species in turn is generated through binding of hydrogen peroxide to the ferric resting state (Figure 27. B). Similarly, in vanadium-dependent haloperoxidases, hydrogen peroxide binds to vanadium which is followed by halide addition and finally the release of the hypohalous acid (Figure 27. C). The halogen atom in hypohalous acid is then attached by a nuchleophilic substrate, resulting in a halogenated product (Figure 27. A).

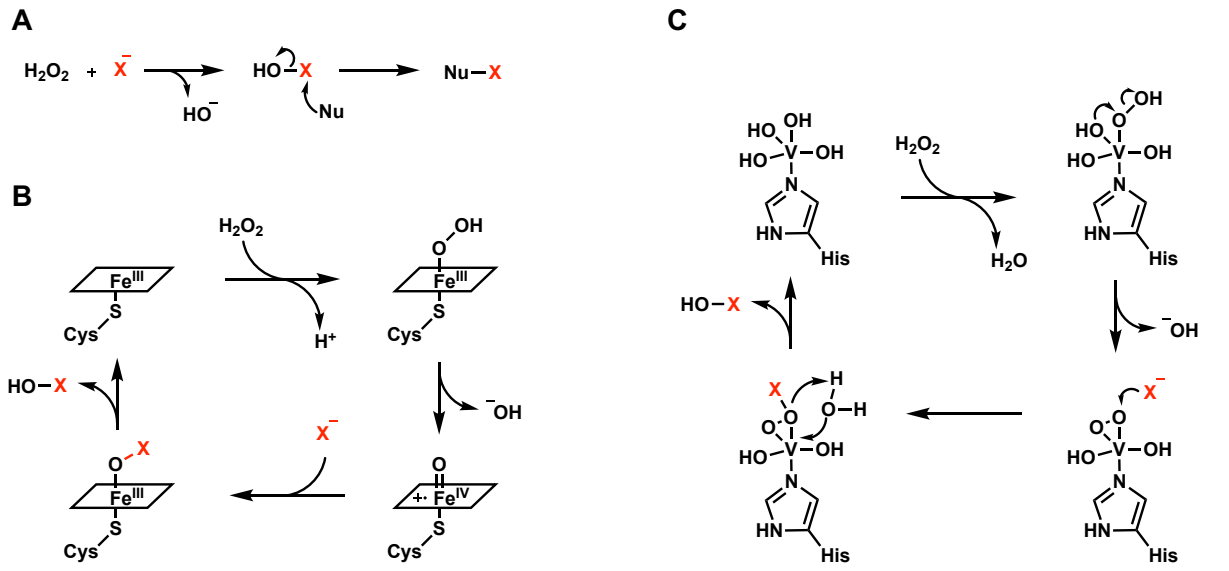


Figure 27. Haloperoxidase proposed reaction mechanisms. (A) General halogenation mechanism (B) Heme-iron-dependent halogenase catalytic mechanism (C) Vanadium-dependent halogenase catalytic mechanism

Another type of halogenase that utilizes hypohalous acid as its reactive species is the flavin-dependent-halogenase (Figure 28). The hypohalous acid is generated similarly to how the hydroxylated substrate is formed in flavin-dependent monooxygenase (FMO) reactions. In FMOs, molecular oxygen is used to generate C4a-hydroperoxy-flavin using reduced flavin ($FADH_2$), which in turn allows diverse reactions such as hydroxylation, epoxidation, Baeyer−Villager oxidation, and heteroatom oxidations.[107] After the hypohalous acid is produced, the hydroxyflavin cofactor loses a water to generate the oxidized flavin, which is reduced by a separate flavin reductase to provide the reduced flavin for the next catalytic cycle (Figure 28).

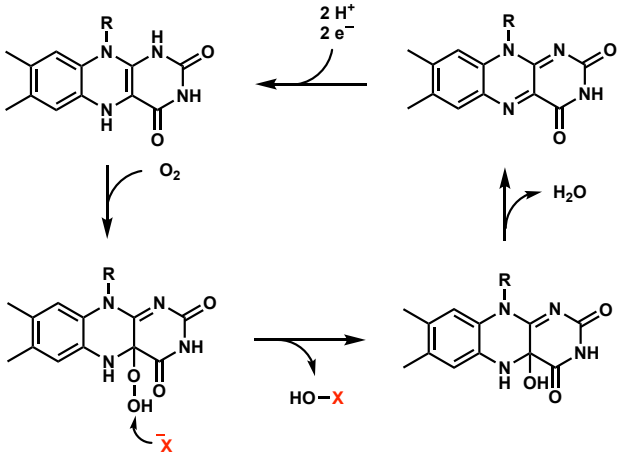


Figure 28. Flavin-dependent halogenase reaction mechanism

Next, we have the non-heme-iron-dependent halogenation enzymes, which uses α-keto-glutarate (α-KG) for its catalysis. This class of halogenase is involved in the halogenation of unactivated carbon centers on aliphatic moieties, since it uses radical species to activate the substrate (Figure 29).[105] However, this type of enzymes mainly utilize substrates tethered to acyl or peptidyl carrier proteins (PCP), rendering its use as a biocatalyst difficult.

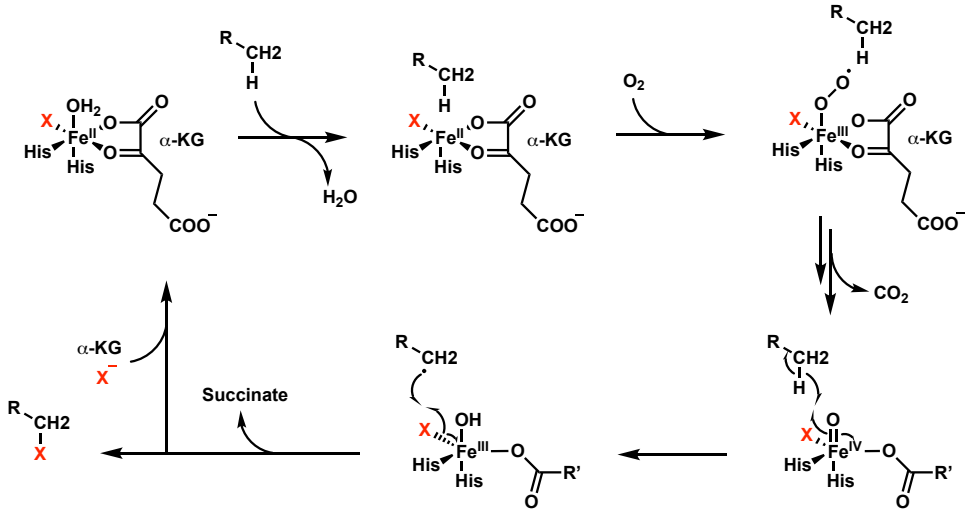


Figure 29. α-KG-dependent halogenase proposed reaction mechanism

Lastly, the nucleophilic halogenase which catalyzes a S-adenosyl-methionine (SAM) -dependent halogenation, is the only known class of enzyme to generate fluorinated NPs. A fluoride anion acts as the nucleophile attacking the C5' of SAM releasing methionine in an S_N2 fashion (Figure 30). Similar halogenation was observed in salinosporamide biosynthesis where a chlorination was the native reaction. However, in halomethane synthases, nucleophilic attack by the halide ion occurs at the S-methyl moiety as in all other methyltransferases, generating halomethanes.[106]

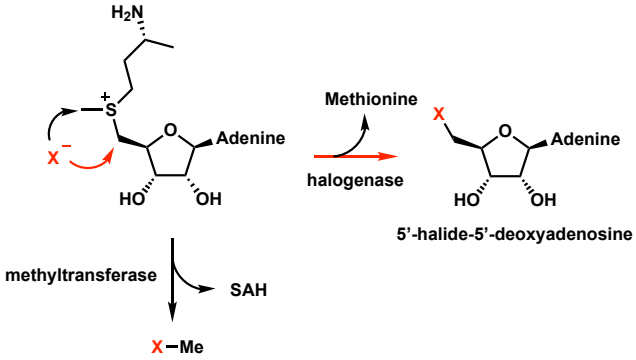


Figure 30. S-adenosyl-methionine (SAM) dependent halogenase.

4.3 Domain of Unknown Function 3328 (DUF3328) enzymes and DUF3328 involvement in fungal NP biosynthesis

Domain of Unknown Function (DUF) enzymes refer to a large set of enzyme families within the Pfam database that do not include any protein of known function.[108,109] As of 2019, in Pfam release 32.0, there are 3961 DUFs which corresponds to 22% of all entries. Many of these DUFs are distributed widely among prokaryotic and eukaryotic organisms, and their functions are getting deciphered yearly.

Although few of these DUFs have been shown to be involved in bacterial NP biosynthesis, only DUF3328 has been reported to be involved in fungal NP biosynthesis.[110–112] In the biosynthesis of fungal ribosomally synthesized and post-translationally modified peptide (RiPP) NPs ustiloxin B and asperipin-2a, DUF3328 is proposed to catalyze an oxidative cyclization.[110–113] Most recently, Jiang *et. al.*, showed DUF3328 proteins catalyze stereo- and regiospecific chlorination and hydroxylation and an intramolecular *O,N*-transacylation in the biosynthesis of mycotoxin cyclochloritine.[114]

4.4 Elucidation of the biosynthetic pathway of atpenin B

The *apn* BGC contains additional features to *har* BGC, including a homolog of SDHC (ApnS), a probable transport protein containing two EamA-like transporter family domain (ApnT) and a DUF3328 protein (ApnU) (Table 5). In addition to these 3 genes, a cytochrome B561 (WfapnV) was identified in the putative BGC from *Chaetasbolisia erysiophoides* FNF40 (Figure 14). Since this strain is known to produce compounds WF-16775 A1 and A2, of which the latter is trichlorinated,[85] we hypothesize this additional feature maybe important for chlorination. We found the homolog of WfapnV in the genome of *P. oxalicum*, albeit on a different scaffold than that of *apn* BGC.

Table 5. Putative functions of proteins encoded in *apn* BGC

Gene name	Proposed protein function
apnA	Polyketide synthase – nonribosomal peptide synthetase (PKS−NRPS)
apnB	*O*-Methyltransferase (OMT)
apnC	Flavin-dependent monooxygenase (FMO)
apnD	*N*-hydroxylation Cytochrome P450 (P450)
apnE	trans-Enoyl reductase (ER)
apnF	Flavin-dependent monooxygenase (FMO)
apnG	Cytochrome P450 (P450)
apnH	Transcription factor (TF)
apnS	Succinate dehydrogenase subunit C (SDHC)
apnT	Putative transport protein (EAM)
apnU	Domain of unknown function (DUF3328)
apnV	Cytochrome B561 (B561)

P. oxalicum contains two copies of SDHC, including ApnS, and they share 43.2% identity, supporting our initial hypothesis of ApnS may be the second-copy SRE (Figure S13). All three additional features are conserved amongst the homologous BGCs that we identified, indicating its involvement in the biosynthesis of the NP being produced (Figure 14). Especially, since DUF3328 is suggested to be involved in halogenation, ApnU is most likely to be involved in the reaction that chlorinates the side-chain of atpenin A5.

When we heterologously expressed ApnABCDEG in *A. nidulans*, the LC/MS anaylysis showed atpenin B, which has a MW of 297, was produced at 0.3 mg/L (Figure 31, Table S11). This confirmed the *apn* BGC is indeed responsible for the production of the atpenin-related compounds. We also expressed the same set of enzymes from *A. sclerotiicarbonarius* (AscABCDEG) and observed similar results (data not shown). However, with our initial heterologous expression of ApnU in addition to ApnABCDEG, we weren't able to observe any

halogenated compound (Figure 31). Additional expression of the remaining BGC genes was not able produce atpenin A5 in our initial try and LC-MS analysis (Figure S15).

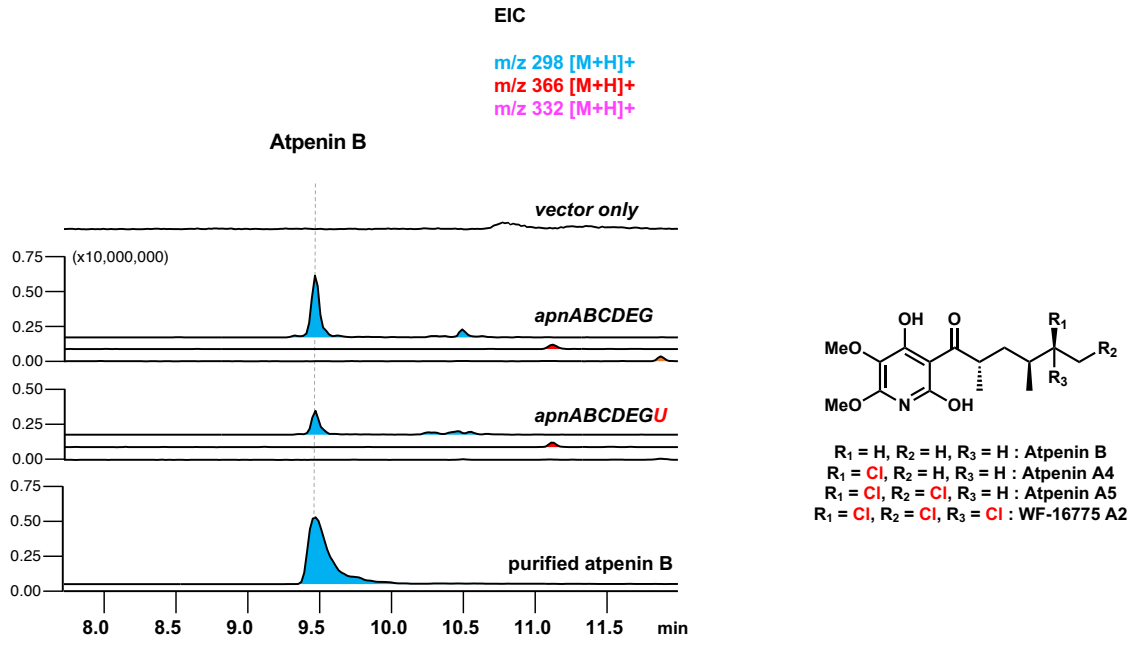


Figure 31. LC/MS analysis of *A. nidulans* expressing *apn* BGC gene combinations.

4.5 Establishing *P. oxalicum* knock-out strains

Since our initial attempt at heterologously producing atpenin A5 was not successful, we decided to initiate gene knock-out studies on the original atpenin A5 producing strain (*P. oxalicum* FO-125) to determine if ApnS-U was involved in the halogenation in atpenin A5 biosynthesis. Once we confirmed the sensitivity of *P. oxalicum* to hygromycin, plasmids designed to knock-out genes *apnS, apnT and apnU*, individually, were built (Figure S16). The plasmids were designed to have 2000 bp flanking region of the target gene on each side of hygromycin B phosphotransferase gene (hph) fused to promoter (TrpCp) and terminator (TrpCt) elements of the *A. nidulans* trpC gene. The plasmids also contain elements needed for yeast homologous recombination (YHR) and *E. coli* origin of replication (Figure S16). PCR reactions were then

performed to generate two PCR products with 500 bp homologous region, where one includes *N*-terminal 2000 bp region, TrpCp and approximately half of *hph* sequence, and the other includes the rest of *hph* gene sequence, TrpCt, and *C*-terminal 2000 bp region. When the PCR products were transformed to *P. oxalicum* protoplasts and selected on hygromycin B, they would homologously recombine and knock-out the targeted gene. We were able to generate ΔApnS, ΔApnT, ΔApnU and ΔApnV knock-out strains, in addition to ΔApnG strain to be used in later experiments, whose deletions were confirmed by agarose gel electrophoresis (Figure S17).

4.6 Analysis of *P. oxalicum* knock-out strains' metabolite profiles

We then analyzed the metabolites produced by the knock-out strains. We observed that ΔApnT and ΔApnU strains were exclusively producing atpenin B, compared to the wild-type strain producing atpenin B, A4 and A5 (Figure 32), indicating these genes are essential for the production of halogenated atpenins. We can also observe almost complete abolishment of halogenated compound production in ΔApnS. This indicates although *apnS* maybe involved in the biosynthesis of halogenated atpenins, its presence is not essential for the production. We did not observe significant change in metabolite profile in ΔApnV strain, demonstrating its involvement in the halogenation is unlikely. We can also see the complete abolishment of atpenin-related compounds production in ΔApnG strain, confirming its essential role in producing the precursors to atpenins.

4.7 Transient expression of ApnT and ApnU in knock-out strains

Since we determined *apnT* and *apnU* were essential for the chlorinated atpenins production from the knock-out strains metabolite profiles, we proceeded to confirm if ApnT and ApnU could restore the production of atpenin A5 if transiently expressed. Since the strains are

already hygromycin B resistant, we elected glufosinate/phosphinothricin as the selection marker for transient expression of ApnT and ApnU.

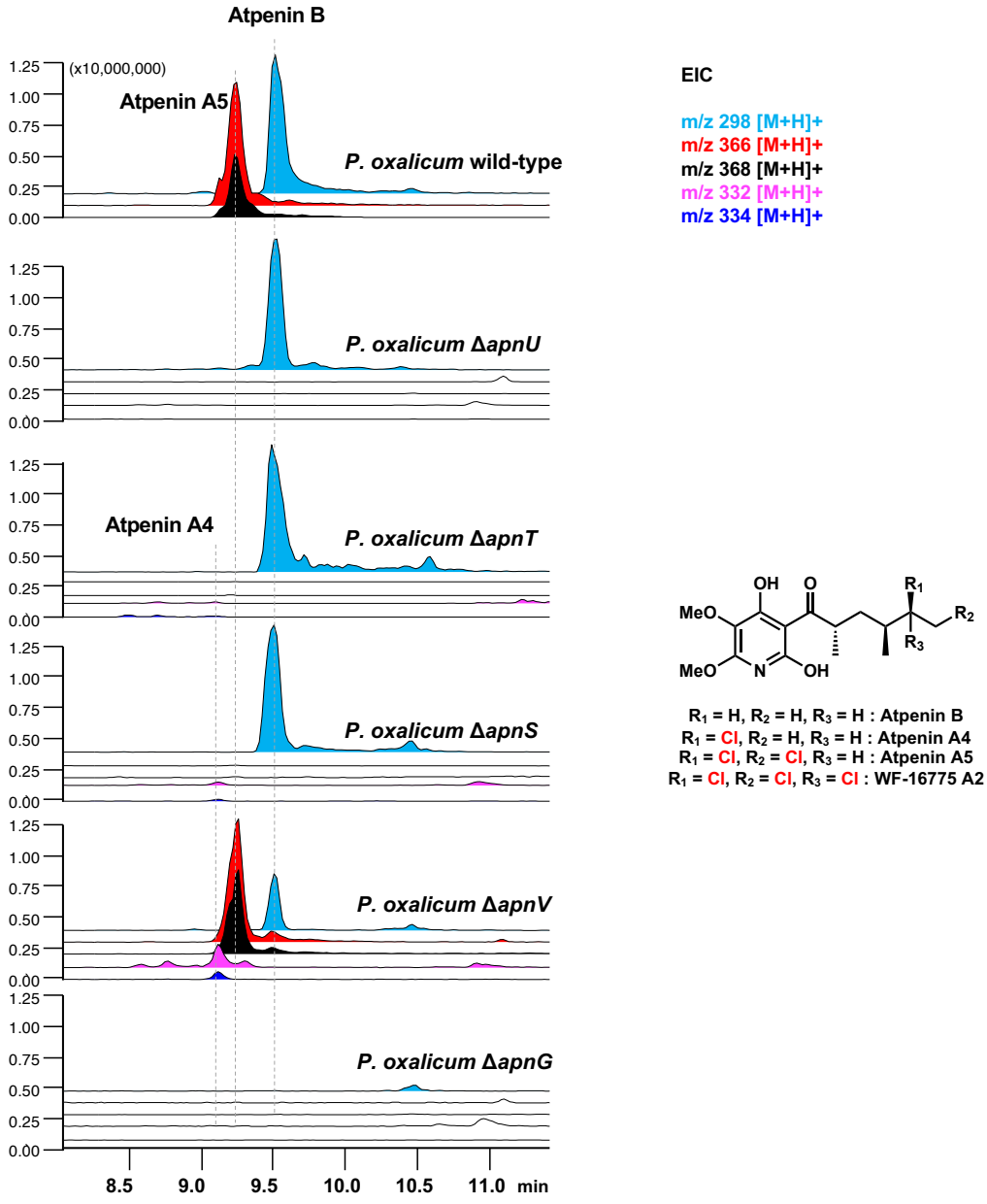


Figure 32. LC/MS analysis of *P. oxalicum* knock-out strains.

The plasmids were designed with glyceraldehyde-3-phosphate dehydrogenase promoter (gpdAp) fused to either *apnT* or *apnU* gene with its native terminator region (500 bp), phosphinothricin N-acetyltransferase (barR) gene fused to promoter (TrpCp) and terminator (TrpCt) elements of the *A. nidulans* trpC gene and *E. coli* origin of replication for cloning (Figure

S18). The resultant plasmids were linearized with restriction enzymes and transformed to ΔApnT and ΔApnU strains. When the metabolites produced by the transient expression strains were analyzed with LC-MS, we were able to observe the recovery of aptenin A5 and atpenin A4 production (Figure 34), albeit at a significantly lower level than the wild-type production. This suggests although these enzymes are necessary for recovering the production of chlorinated compounds, their transient expression level or timing is not ideal for the production of chlorinated atpenin production.

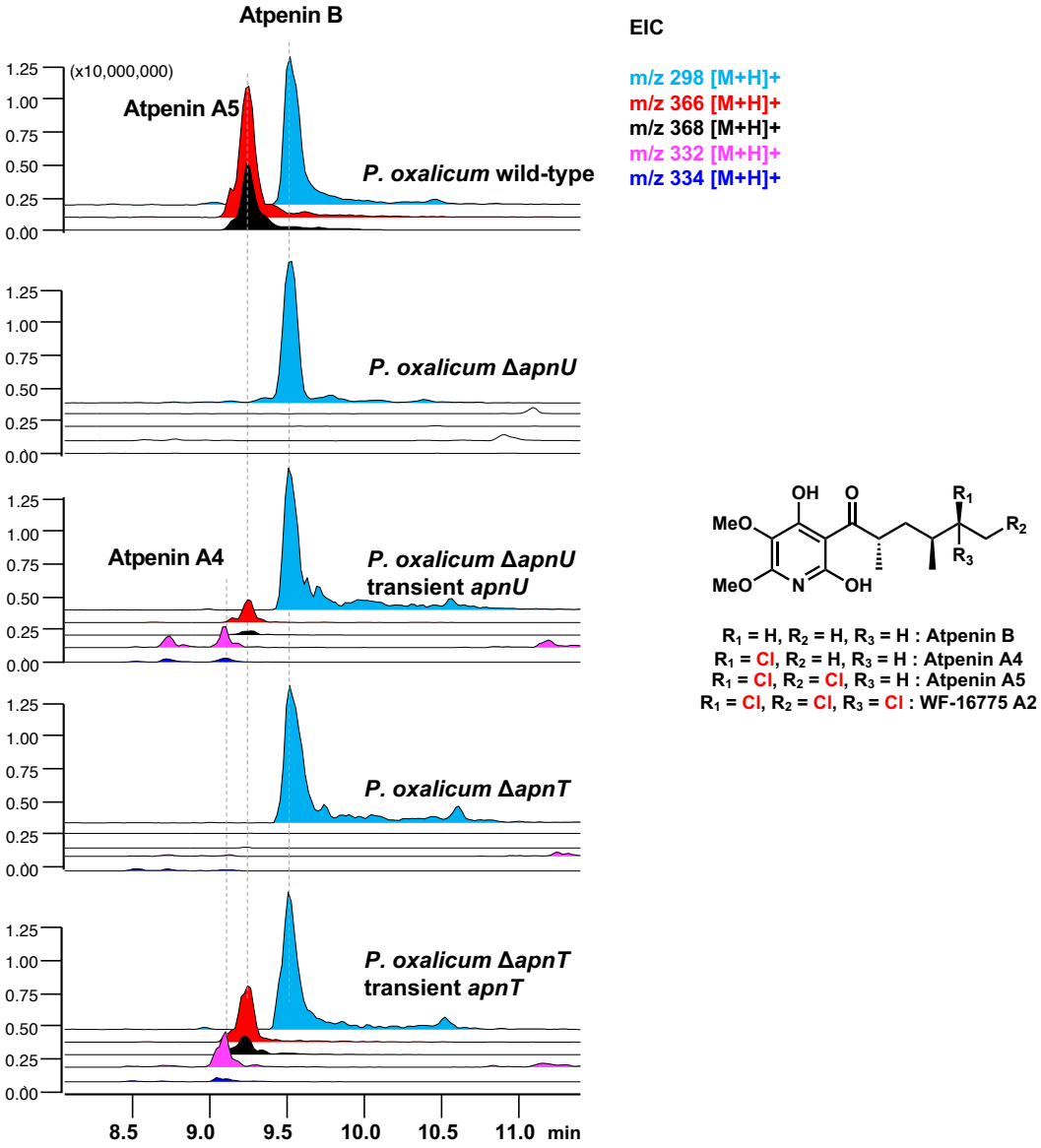


Figure 33. LC/MS analysis of *P. oxalicum* transient expression strains metabolites.

4.8 Determination of precursor to chlorinated atpenins

Although we were able to confirm the necessary enzymes involved in the production of chlorinated atpenins, the question of whether atpenin B was the precursor to these compounds remained. In order to determine this and eliminate the possibility of halogenation occurring on an earlier intermediate, we fed atpenin B to ΔApnG strain. When atpenin B was fed to ΔApnG strain after one day of growth resulted in the production of atpenin A4 and atpenin A5, confirming the chlorination occurs after atpenin B is produced (Figure 34). This result signifies the halogenation if indeed occurring on an aliphatic carbon and the halogenation enzymes involved are previously unknown type of halogenation enzyme.

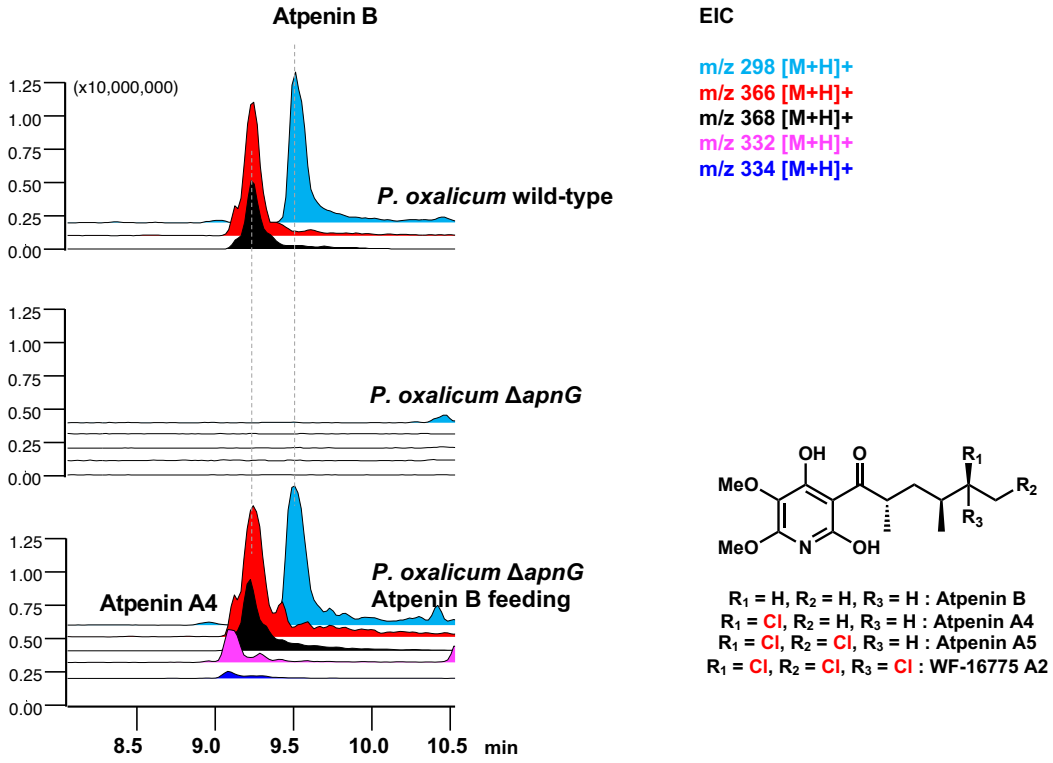


Figure 34. LC/MS analysis of metabolites from *P. oxalicum* Δ*apnG* strain and atpenin B feeding

4.9 Heterologous production of atpenin A5 in *A. nidulans*

So far, we have determined atpenin B to be the precursor to chlorinated atpenins, as well as the involvement ApnT, ApnU and ApnS in the production of chlorinated atpenins. However, in order to determine which enzymes are sufficient and necessary to produce chlorinated atpenins, we decided to try the *A. nidulans* heterologous expression again. The heterologous expression strains were grown in CD-ST media supplemented with NaCl and KCl each at 2.5 mM. We also analyzed the metabolites on HRMS-QTOF system. This time, we were able to confirm the heterologous production of atpenin A5 in *A. nidulans*, albeit at a very low concentration (Figure 35). The overall production of chlorinated atpenins was very low in *A. nidulans*, despite the production of its immediate precursor atpenin B was considerably high. This may be due to the non-ideal enzyme expression in *A. nidulans,* lack of suitable accessory enzymes and/or atpenin A5 being a potent inhibitor. Nonetheless, we were able to establish *apnABCDEGTU* to be the minimal set of genes to produce atpenin A5 heterologously.

Interestingly, the production of atpenin A5 was increased substantially when ApnV or $ApnS_2$ (ApnS with different starting position) was co-expressed. Suggesting these enzymes assist the function of ApnU and/or ApnT. Since ApnT is a putative transport protein and ApnU is suggested to be involved in oxidative cyclization or chlorination in other NPs, we assume ApnU is the halogenation enzyme with ApnT as an essential accessory enzyme. Since both ApnV and ApnS are heme-binding enzymes, the cofactors needed for the halogenation reaction maybe regenerated by these enzymes, thus the increase in production maybe explained with efficient cofactor regeneration.

We also fed atpenin B to *A. nidulans* expressing genes other than ApnA, which confirmed the conversion of atpenin B to atpenin A5 in a heterologous host (Figure 36). However, the amount of atpenin A5 was significanly lower than atpenin A4, indicating the chlronation reaction halted after the initial chrorination, further confirming the inefficiency of the heterologous host. Notably,

the bioconversion level was comparable between *apnBCDEGTU*, *apnBCDEGTUV* and *apnBCDEGSTUV* expressing strains, bringing into question whether the inclusion of ApnV or ApnS is increasing the production of chlorinated compounds (Figure 36).

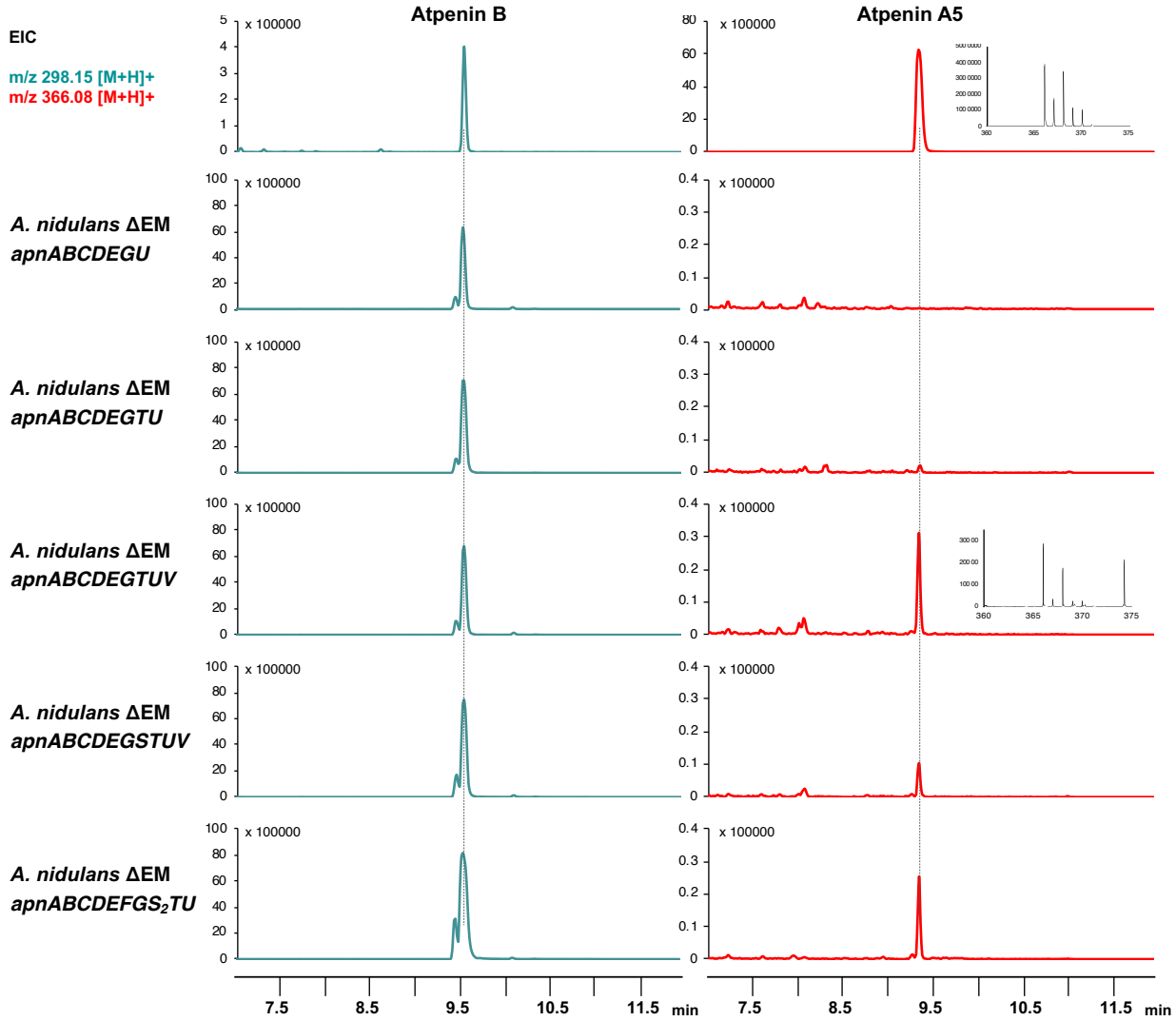


Figure 35. HRMS-QTOF analysis of A. nidulans expressing different combinations of *apn* BGC

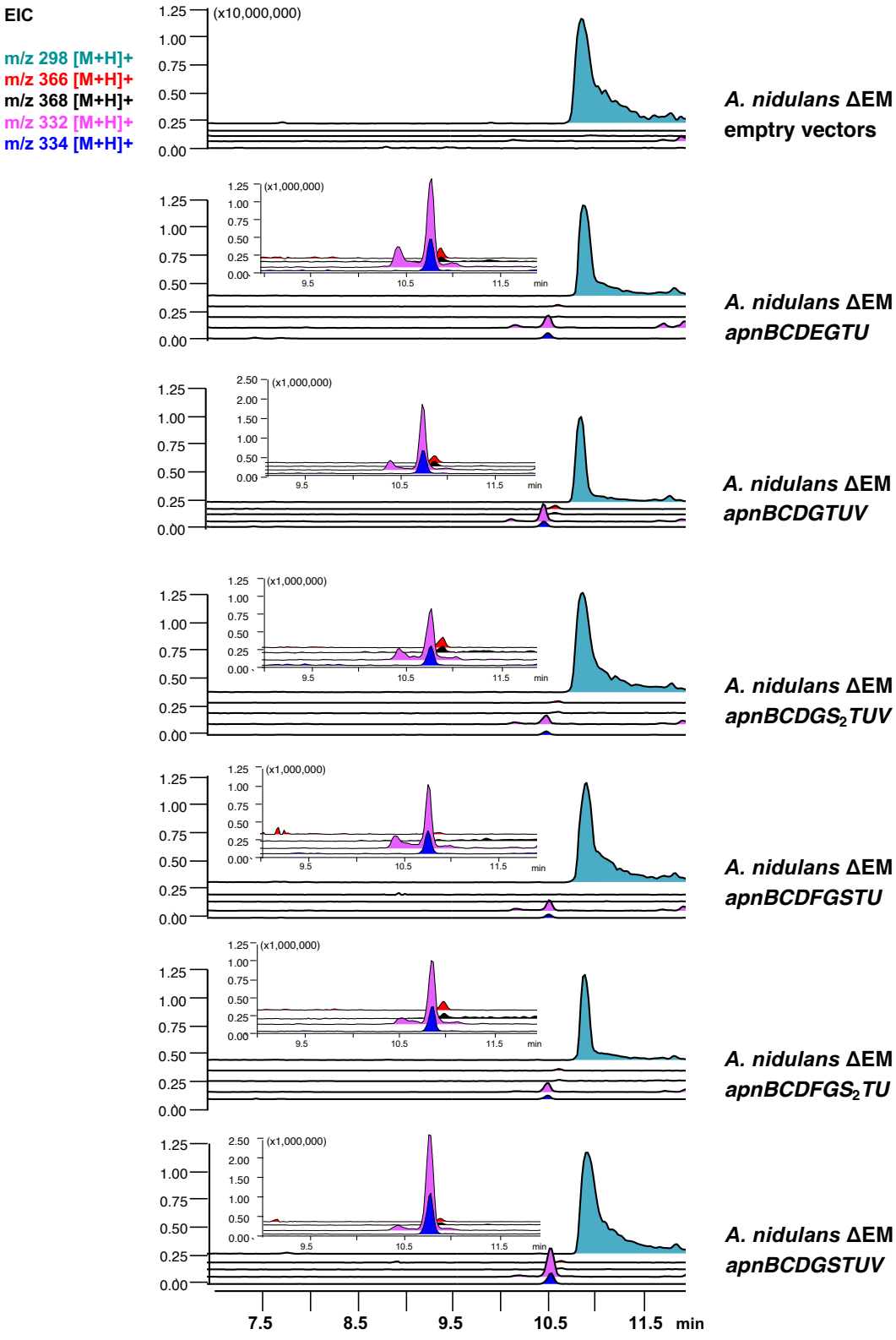


Figure 36. LC/MS analysis of feeding atpenin B to *A. nidulans* expressing different combinations of *apn* BGC

4.10 Conclusions and future direction

In this section, we were able to uncover the BGC responsible for the potent CII inhibitor atepenin A5 with the help of SRE-DGM and *har* BGC. Enzymes ApnABCDEGTU have been confirmed to be the minimal set of enzymes that produce atpenin A5 heterologously. We have also determined ApnU, a DUF3328 enzyme, is most likely responsible for the halogenation reaction, which would a new type of halogenase enzyme that activates and installs halogens on a non-activated aliphatic carbon. We also observed a putative transport protein ApnT is essential in the halogenation reaction, possibly being the partner enzyme to ApnU. We also see ApnV and ApnS, which are heme-binding enzymes, increase the production of halogenated products, suggesting these enzymes provide cofactor regeneration specifically needed for ApnU activity.

In order for us to fully understand the halogenation mechanism of ApnU, we will need to perform mechanistic studies on ApnU and ApnT. Since these enzymes are membrane bound or transmembrane proteins, microsomal preparation of the enzymes are desired to study their mechanism. This can be achieved through yeast microsomal studies on the premise these enzymes are active when expressed in yeast. Furthermore, the cofactor and metal dependence of the enzymes need to be carefully studied in order for us to understand the full catalytic potential of this novel halogenase.

5 ONE-POT *IN VITRO* BIOSYNTHESIS OF PLANT MONOTERPENE

Parallel to finding new NPs and enzyme catalyses, application of existing enzyme catalysts in order to provide reliable source of valuable chemicals is as equally important. Such uses are the motivation behind finding interesting biocatalysts in the first place. In this section, our endeavor for cell-free production of plant monoterpenes is discussed.

5.1 Monoterpene-indole alkaloids (MIA)

Amongst pharmaceutically relevant NPs, plant-derived monoterpene-indole alkaloids (MIAs) comprise more than 3000 compounds with diverse biological activities including antimalarial (quinine), antihypertensive (ajmalicine), and anticancer (vinblastine, vincristine and camptothecin) (Figure 37).[115]

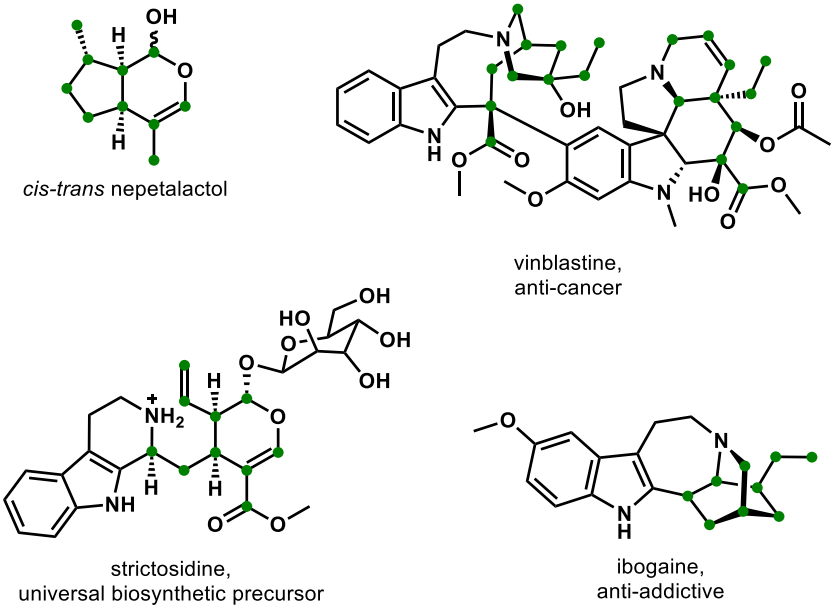


Figure 37. Structures of *cis-trans* nepetalactol and representative MIAs. Nepetalactol serves as the ten-carbon terpene core of strictosidine, the biosynthetic precursor to vinblastine, ibogaine, and ~3,000 additional monoterpene indole alkaloids.

However, some of the most valuable MIAs are produced by their native plant hosts at extremely low concentrations, resulting in a high price of the pharmaceuticals. Furthermore, the native hosts frequently produce their constituents as complex mixtures of structurally related compounds making the purification process extremely arduous. Thus, alternative production methods towards MIAs and MIA precursors, including metabolic engineering and total organic syntheses have been investigated.[116–123] The first *de novo* heterologous production of strictosidine, the universal precursor to complex MIAs, was achieved in 2015, via expression of 21 heterologous gene expressions and deletion of 3 gene in *Saccharomyces cerevisiae.* However, due to endogenous yeast metabolism and plant gene expression issues, the highest strictosidine yield achieved to date is only 0.5 mg/L[122]. Thus, new strategies to produce higher yields of strictosidine must be explored, such as total synthesis or feeding of pre-cursors to increase the titer.

Cis-trans-nepetalactol (referred to hereafter as "nepetalactol") is a precursor to all MIAs[117] (Figure 37). Furthermore, nepetalactol is the immediate precursor to nepetalactone, the active ingredient in the cat attractant catnip (*Nepeta sp.*) and a natural insect repellent [124], thus making nepetalactol an important compound in the production of all bioactive MIAs and a natural target for high-yield production. However, with no sustainable or synthetic source, nepetalactol production is dependent on plant extraction, thus facing the same issues encountered by MIAs and resulting in a high price. In an effort to establish a nepetalactol production strain, Billingsley *et al*, demonstrated a 45 mg/L titer from fed 8-hydroxygeraniol through metabolic engineering and fermentation of *S. cerevisiae*[120]. However, 8-hydroxygeraniol is not a readily available substrate for large-scale nepetalactol production. Yee et al, achieved 6 mg/L *de novo* production of nepetalactol in yeast by localizing the biosynthetic genes in the mitochondria, thus preventing yeast metabolism of the intermediates as much as possible[121]. Despite these efforts, the titer of nepetalactol is still very low for larger scale production, as nepetalactol production in microbial

hosts is hindered by endogenous enzyme interference, inefficient membrane-associated enzyme expression, and substrate-product toxicity.

5.2 Synthetic biochemistry and its benefit

In recent years, cell-free "synthetic biochemistry" has gained traction as an attractive alternative to synthetic chemistry and metabolic engineering.[125] Traditionally, synthetic chemistry strategies have afforded bulk commodities in large amounts in a cost-effective manner throughout history. However, for more structurally challenging synthetic targets such as complex bioactive secondary metabolites, expensive protective groups, hazardous chemicals and harsh conditions may be required for uncommon chemical transformations with high stereoselectivity and yield. An alternative route to producing complex molecules is to engineer microorganisms such as *E. coli* or *S. cerevisiae* to heterologously express biosynthetic genes required for production of the desired compound. However, in many cases, heterologous production of chemicals in these hosts may be derailed by bottlenecks or shunt pathways, as well as substrate and product toxicity which further lowers yields. Furthermore, this method does not alleviate the burden of multi-step purification for the final products, since in many cases the products are co-purified with endogenous metabolites.

In vitro biosynthesis allows us to combine the benefits of both strategies to produce desired compounds at high yield, using mild reaction conditions. "Cell-free" methods exploit the precise regio- and stereoselectivities of enzymes to perform chemical transformations of unprotected substrates. Moreover, this approach enables precise temporal control of the enzymatic reaction: by carefully monitoring the reaction progress, stepwise addition of appropriate enzymes to the system may be employed to further improve yields and efficiencies, which may in turn result in a streamlined purification process for high-value chemicals. While several challenges limiting synthetic biochemistry have been described, corresponding engineering solutions in the

form of enzymatic modules have been established. For example, Opgenorth and colleagues have described a molecular rheostat and synthetic purge valve, two systems which have been designed to manage excess cofactor buildup.

Cell-free biosynthesis platforms can also outperform microbial *in vivo* biosynthesis by eliminating competing metabolic pathways and overcome potential toxicity.[126,127] Notwithstanding the increasing number of examples of cell-free biosynthesis, a number of challenges exist which limit the utility. In particular, efficient and orthogonal cofactor supply and regeneration is an ever-present obstacle, especially for more complex pathways in which multiple (redox) cofactors are involved. Numerous approaches have been developed to address this obstacle, including reengineering of enzyme cofactor specificity,[128] use of chemically orthogonal unnatural cofactors[129,130] and in an impressive demonstration, the use of a molecular rheostat and synthetic purge valve to manage excess cofactor buildup.[131,132]

5.3 Establishing the *in vitro* nepetalactol biosynthesis

Biosynthesis of *cis-trans* nepetalactol **6** from geraniol **1** requires NAD^+-dependent oxidation and NADPH-dependent reductive cyclization (Figure 38). Diverging from primary metabolism, geranyl pyrophosphate is hydrolyzed by geraniol synthase to give geraniol **1**.[122] Regiospecific hydroxylation of one of the terminal methyl groups by the P450 geraniol-8-hydroxylase (G8H) provides 8-hydroxygeraniol **2**.[133] Next, tandem and reversible NAD^+-dependent oxidation of **2** by geraniol oxidoreductase (GOR) generates the dialdehyde 8-oxogeranial **5** (via either 8-hydroxygeranial **3** or 8-oxogeraniol **4**).[117] Stereoselective reduction of **5** using NADPH to an enol intermediate by iridoid synthase (ISY), followed by enzyme-assisted cyclization by a major latex protein–like enzyme (MLPL) result in **6**.[118,123] 5 can spontaneously form the ring-opened iridodials **6***, which can also derive through the ring opening of **6**. Dehydrogenation of **6** by nepetalactol-related short-chain reductase/dehydrogenase 1 (NEPS1)

forms nepetalactone **7**, which is a potent insect repellent and the active cat-attractant in catnip.[134] Nepetalactol **6** can be further modified into strictosidine (Figure 37).

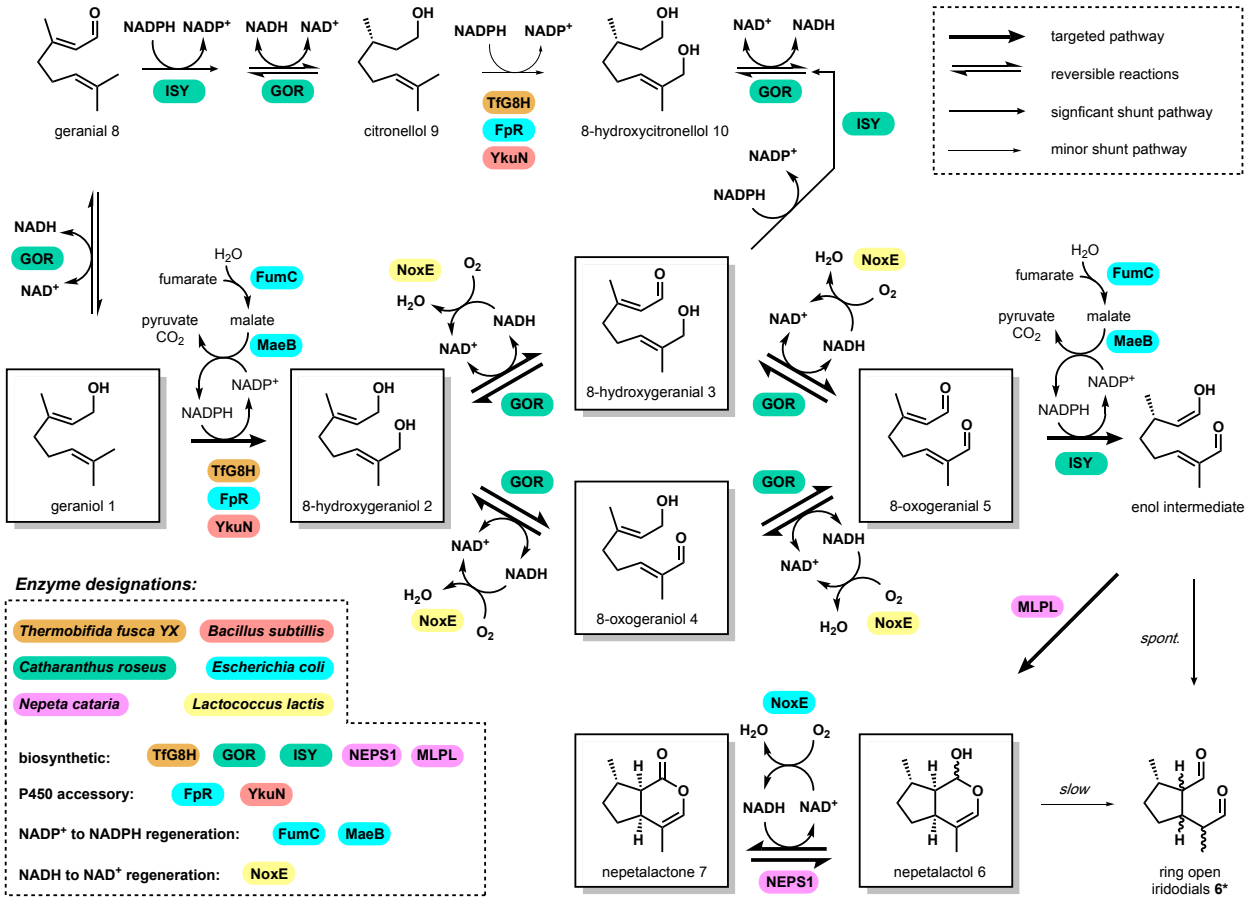


Figure 38. Biosynthesis of nepetalactol and nepetalactone along with possible shunt products. On pathway intermediates are boxed. Cofactor regeneration enzymes are only shown for main pathway reactions.

Initial efforts to recapitulate nepetalactol production using cell-free biosynthesis were prompted by the expensive cost from commercial vendors, the low titers observed in microbial hosts,[121,135–137] and difficulties in implementing synthetic routes. Reported syntheses suffer from low yields and enantioselectivities or rely on the costly synthon (−)-citronellol as an enantiopure starting material.[138–143] To perform cell free biosynthesis, geraniol 1 was selected as the starting material because of its high abundance as an essential oil and low cost. Since the plant homologues of G8H are membrane-bound and not suitable for *in vitro* biocatalysis, a functionally equivalent, soluble bacterial P450 was used. CYP154E1 (TfG8H) from *Thermobifida fusca* YX

was reported to perform the same hydroxylation as G8H and can be reductively regenerated by the NADPH-dependent cytochrome P450 flavodoxin/ferredoxin reductase (FpR from *Escherichia coli*) and flavodoxin (YkuN) from *Bacillus subtillis*.[144,145] Using TfG8H, FpR and YkuN, we were able to observe near complete hydroxylation of **1** to 8-hydroxygeraniol **2** (Figure 39).

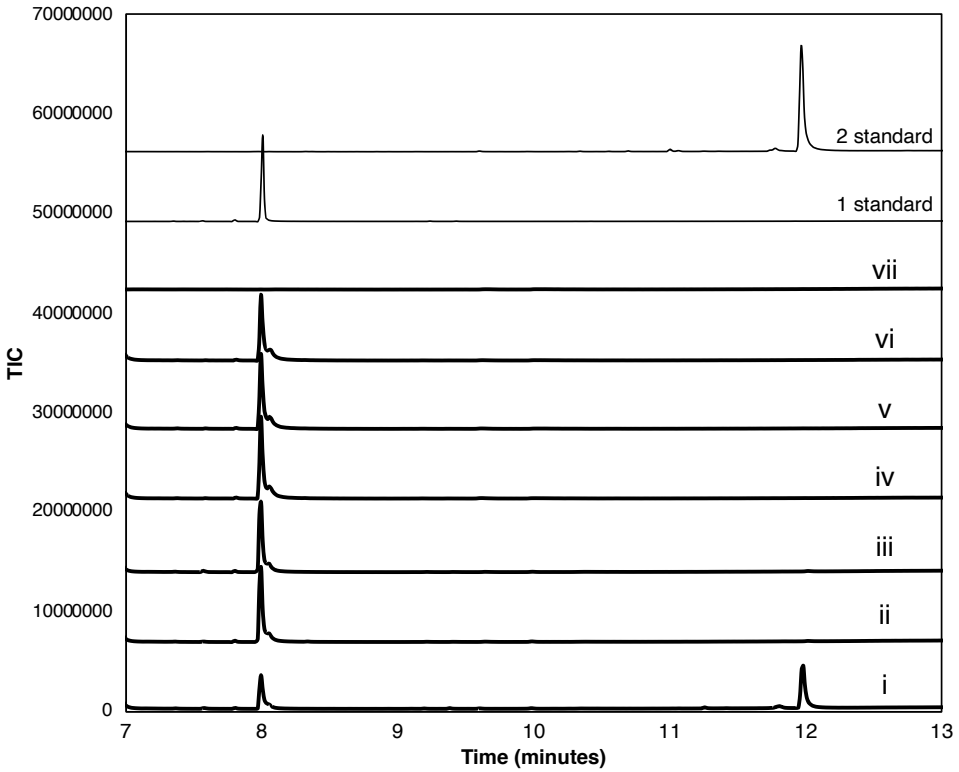


Figure 39. GC/MS analysis of geraniol hydroxylation activity of TfG8H. Each reaction contains 2 mM geraniol, 5 μM TfG8H, 10 μM FpR, 10 μM YkuN, 4 mM NADPH unless otherwise specified, i. sample reaction, ii. 100 μM NADPH, iii. no TfG8H, iv. no FpR, v. no YkuN, vi. no NADPH, vii. no geraniol.

We next confirmed full conversion of **2** (340 mg/L) to 8-oxogeranial **5** when combined with GOR and excess NAD^+ (Figure 40). Upon incubation of **5** (330 mg/L) together with ISY and MLPL, nepetalactol **6** was exclusively formed using excess NADPH. ISY could utilize both NADPH and NADH as reducing cofactors, albeit showing strong preference for NADPH. Formation of ring opened **6*** was detected when MLPL was excluded from the reaction (Figure 41), consistent with prior *in vivo* reconstitution efforts.[136]

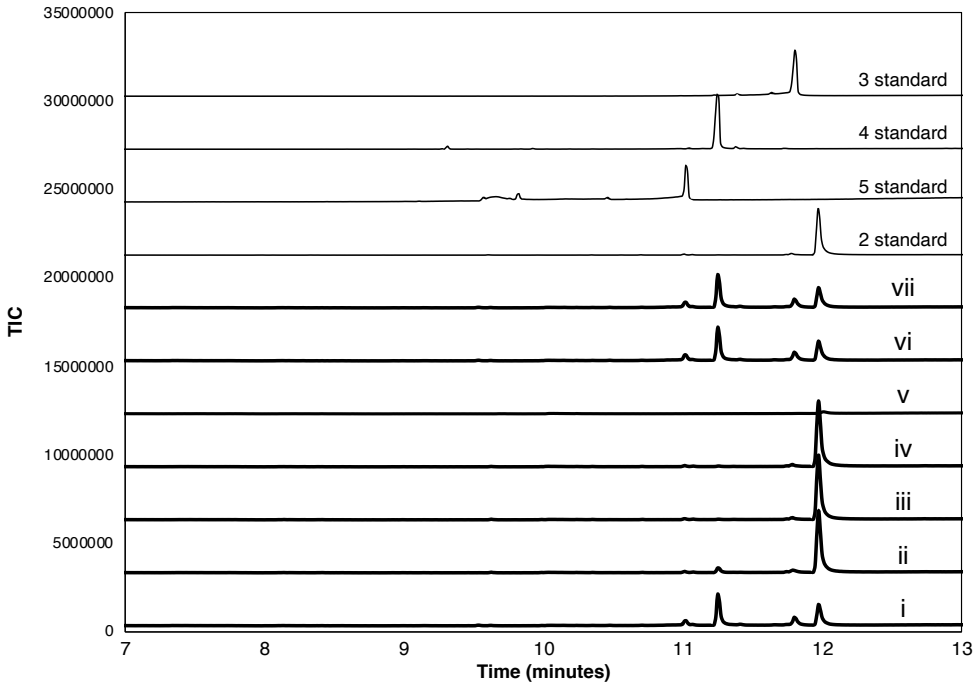


Figure 40. GC/MS analysis of 8-hydroxygeraniol oxidation by GOR. Each reaction contains 2 mM 8-hydroxygeraniol, 10 μM GOR and 4 mM NAD^+ in BTP buffer (pH 9.0) unless otherwise specified, i. sample reaction, ii. 100 μM NAD^+, iii. no GOR, iv. no NAD^+, v. no 8-hydroxygeraniol, vi. Phosphate buffer pH 7.5, vii. Tris-HCl buffer pH 8.0.

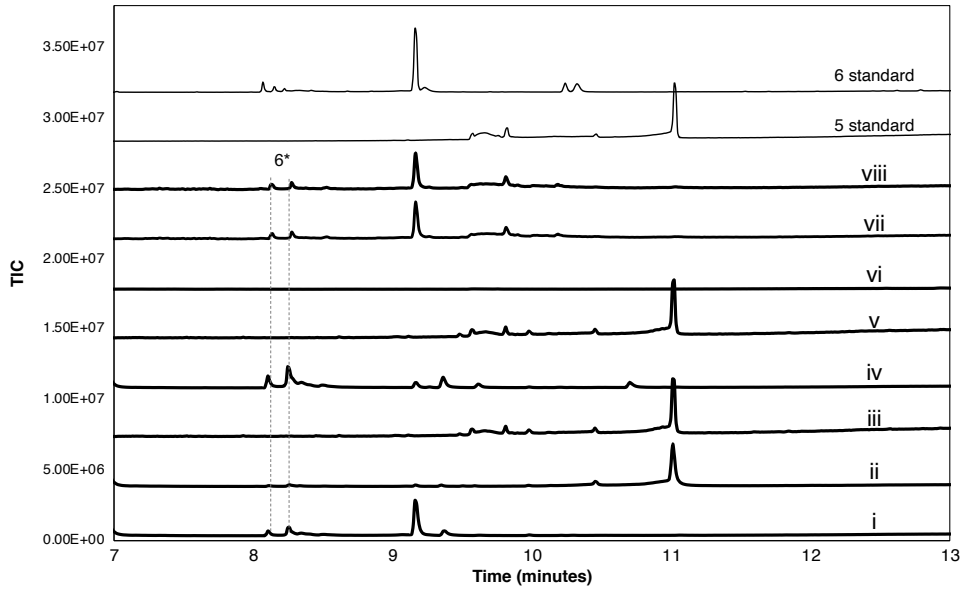


Figure 41. GC/MS analysis of 8-oxogeranial reduction and cyclization by ISY and NmMLPL. Each reaction contains 2 mM 8-oxogeranial, 1 μM ISY, 10 μM NmMLPL and 4 mM NADPH in BTP buffer (pH 9.0) unless otherwise specified, i. sample reaction, ii. 100 μM NADPH, iii. no ISY, iv. no NmMLPL, v. no NADPH, vi. no 8-oxogeranial, vii. Phosphate buffer pH 7.5, viii. Tris-HCl buffer pH 8.0.

5.4 Establishing the cofactor regeneration systems

Balancing cofactor usage is especially important in systems where both reductive and oxidative reactions are involved. When the biosynthetic pathway uses a single type of cofactor, such as NAD(H) or NADP(H), concomitant oxidation of reducing equivalents upon substrate reduction serves to regenerate oxidizing equivalents, and vice versa. During active metabolism, however, estimated ratios of NAD^+:NADH range from 200:1 to 600:1,[146] whereas estimated ratios of $NADP^+$:NADPH range from 1:30 to 1:200.[146] Thus, many biosynthetic pathway enzymes have evolved to use different types of cofactors. Indeed, natural product biosynthetic logic frequently employs both NAD^+-dependent oxidation and NADPH-dependent reduction steps.[147–150] Without an orthogonal cofactor regeneration system, combining all the enzymes in one pot will lead to futile redox cycles. Therefore, to achieve one-pot reconstitution of such pathways, it is essential to eliminate crosstalk when regenerating both cofactors. The situation is made even more complex in the event that a reaction is reversible, which is true for many of the NAD(P)H-dependent oxidoreductases.[151]

With the biosynthetic enzymes in hand and their activities confirmed, we set out to address the excessive cofactor concentrations needed for each step. Cofactors are indeed a major price contributor in cell-free systems; however the cost can be significantly reduced if an inexpensive sacrificial substrate can be used to recycle spent cofactors. Because the interconversion of 8-hydroxygeraniol to 8-oxogeranial is reversible, and ISY can accept partially oxidized intermediates as a substrate, we identified the ratio of NAD^+ to NADH as a key parameter affecting conversion. Thus, in addition to cost, the major requirement considered for robust NAD^+/NADPH regeneration was substrate orthogonality. In other words, we needed to identify an NADH oxidase that would not recognize NADPH, and an $NADP^+$ reductase that would not recognize NAD^+. In doing so, we could eliminate futile cycling of cofactors which would drive down yield and increase cost. We selected NADH-oxidase (NoxE) from *Lactococcus lactis* which catalyzes a NADH-dependent

reduction of oxygen to water as the NAD^+ regenerating enzyme since it would not require an additional substrate and has been shown to be NAD^+ specific[152]. Indeed, incubation of NoxE with NAD^+ resulted in rapid cofactor oxidation, however when incubated with NADP+, no oxidation was observed (Figure 42. A, B).

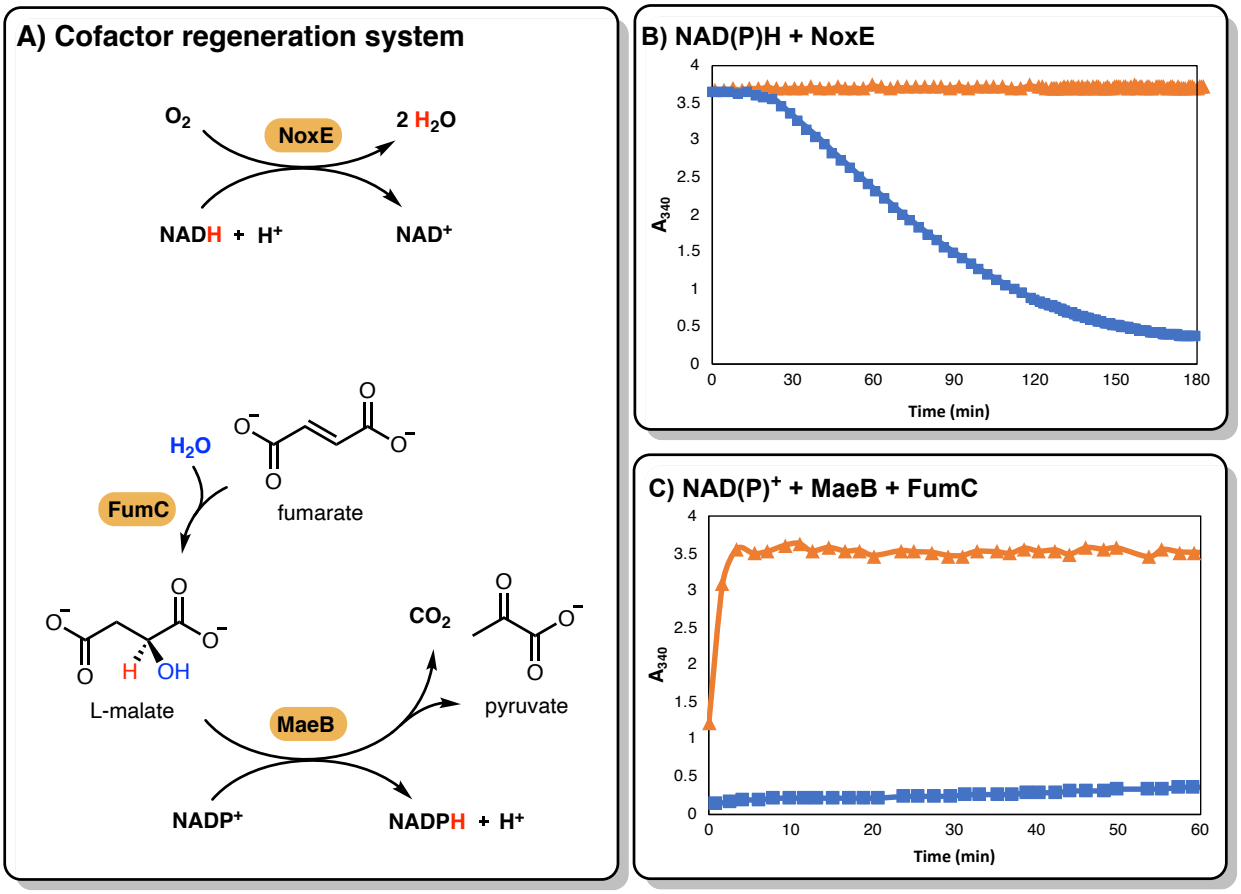


Figure 42. Oxidative and reductive cofactor regeneration system. (A) NoxE is used as a NAD+ regeneration enzyme. Substrate specificity for NADH (blue) over NADPH (orange) is confirmed; (B) the FumC/MaeB pair is used as a NADPH regeneration system. Substrate specificity for $NADP^+$ over NAD^+ is confirmed. Both assays in (A) and (B) were performed by spectroscopically at λ=340 nm.

In contrast, a cost-effective NADPH regeneration system that does not reduce NAD^+ was not readily available. The conventional glucose-6-phosphate (G6P) dehydrogenase or glyceraldehyde-3-phosphate (GAP) dehydrogenase, which convert G6P to 6-phospho-D-glucono-1,5-lactone and GAP to 1,3-bisphosphoglycerate, respectively,[153,154] use expensive substrates G6P and GAP. Other NADPH-regeneration enzymes such as glucose-1-

dehydrogenase and isocitrate dehydrogenase were also not suitable due to their non-specific cofactor usage or high substrate cost. Combining the requirements of cofactor orthogonality, ease of enzyme expression and cost effectiveness, we chose a two-enzyme system which consists of fumarate hydratase (FumC) and $NADP^+$-dependent malic enzyme (MaeB) from *Escherichia coli*. MaeB catalyzes the decarboxylation of (*S*)-malic acid to generate pyruvate in a strictly $NADP^+$-dependent manner.[155] While (*S*)-malic acid is relatively expensive, it can be readily generated from the hydration of the inexpensive fumarate. In addition to its cofactor selectivity, the decarboxylation reaction catalyzed by MaeB is irreversible and thus drives the coupled reaction forward. Cloning and characterization of MaeB and FumC confirmed that the two-enzyme system displayed excellent selectivity towards $NADP^+$ over NAD^+ in the presence of fumarate (Figure 42. A, C).

5.5 Confirming the established cofactor regeneration system compatibility with nepetalactol biosynthetic enzymes

When coupled, 10 μM GOR and 5 μM NoxE were able to fully convert 2 mM **2** to **5** in the presence of limiting 100 μM NAD^+ (Figure 43) within 1 hour (Figure S21).

To test this system in catalysis, we performed the coupled reaction of TfG8H/FpR/YkuN and FumC/MaeB with limiting concentrations of NADPH. Full conversion of 310 mg/L (2 mM) geraniol **1** to 8-hydroxygeraniol 2 with 5 μM TfG8H, 10 μM FpR, 10 μM YkuN, 1 μM FumC, 10 μM MaeB and 100 μM NADPH (Figure 44) was observed within 1.5 hours (Figure S22). This full conversion establishes a robust *in vitro* P450 biocatalytic reaction, which typically requires excess NADPH due to suboptimal electron transfer between the P450 and its partner enzymes.[156] The FumC/MaeB regeneration system was also fully compatible with ISY and MLPL and supported the full conversion of 330 mg/L (2 mM) **5** to nepetalactol **6** in the presence of 0.5 μM ISY, 5 μM MLPL, 1 μM FumC, 10 μM MaeB and 100 μM NADPH (Figure 45).

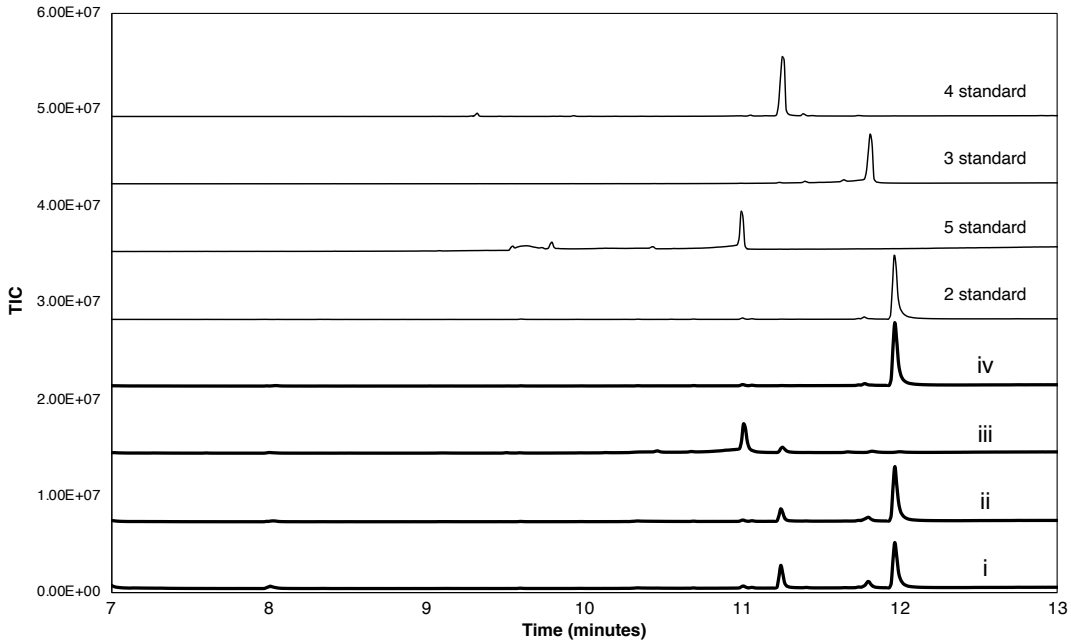


Figure 43. GC/MS analysis of 8-hydroxygeraniol oxidation with GOR with NAD^+ regeneration. Each reaction contains 2 mM 8-hydroxygeraniol, 10 μM GOR, 5 μM NoxE and 4 mM NAD^+ in BTP buffer (pH 9.0) unless otherwise specified, i. no NoxE, ii. no NoxE with 100 μM NAD^+, iii. with NoxE and 100 μM NAD^+, iv. no GOR.

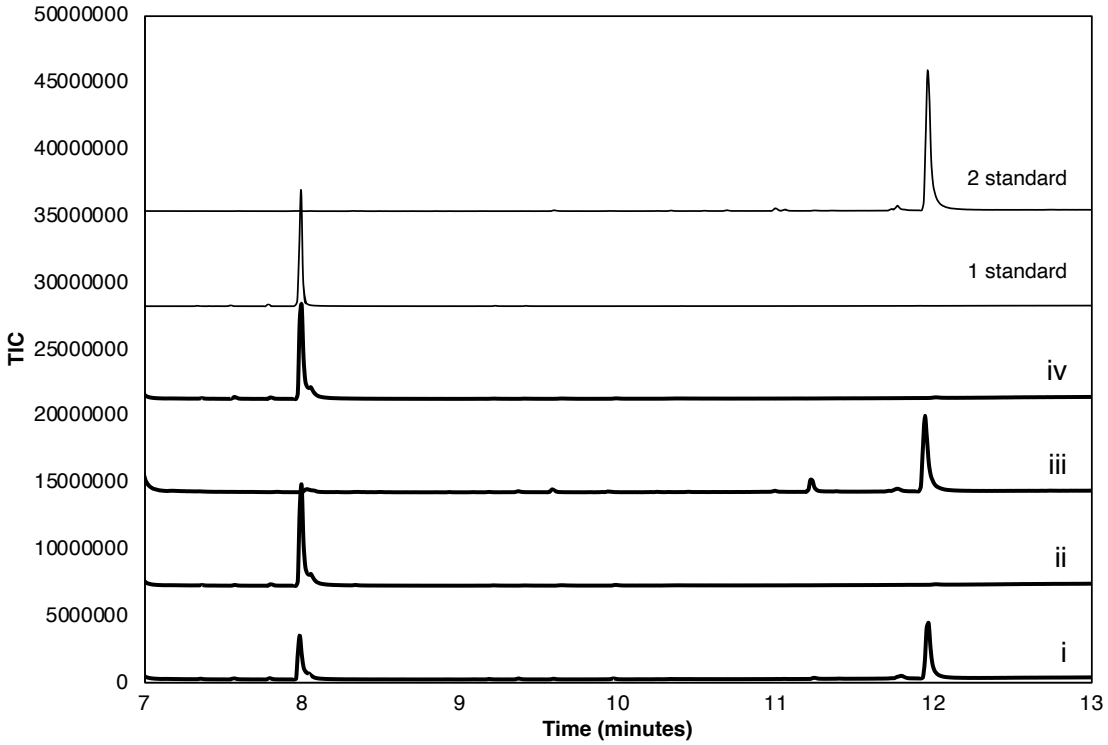


Figure 44. Geraniol hydroxylation by TfG8H with NADPH regeneration. Each reaction contains 2 mM geraniol, 6 mM fumarate, 5 μM TfG8H-full, 10 μM FpR, 10 μM YkuN, 1 μM FumC, 10 μM MaeB and 4 mM NADPH in BTP buffer (pH 9.0) unless otherwise specified, i. no FumC/MaeB, ii. no FumC/MaeB with 100 μM NADPH, iii. with FumC/MaeB and 100 μM NADPH, iv. no TfG8H.

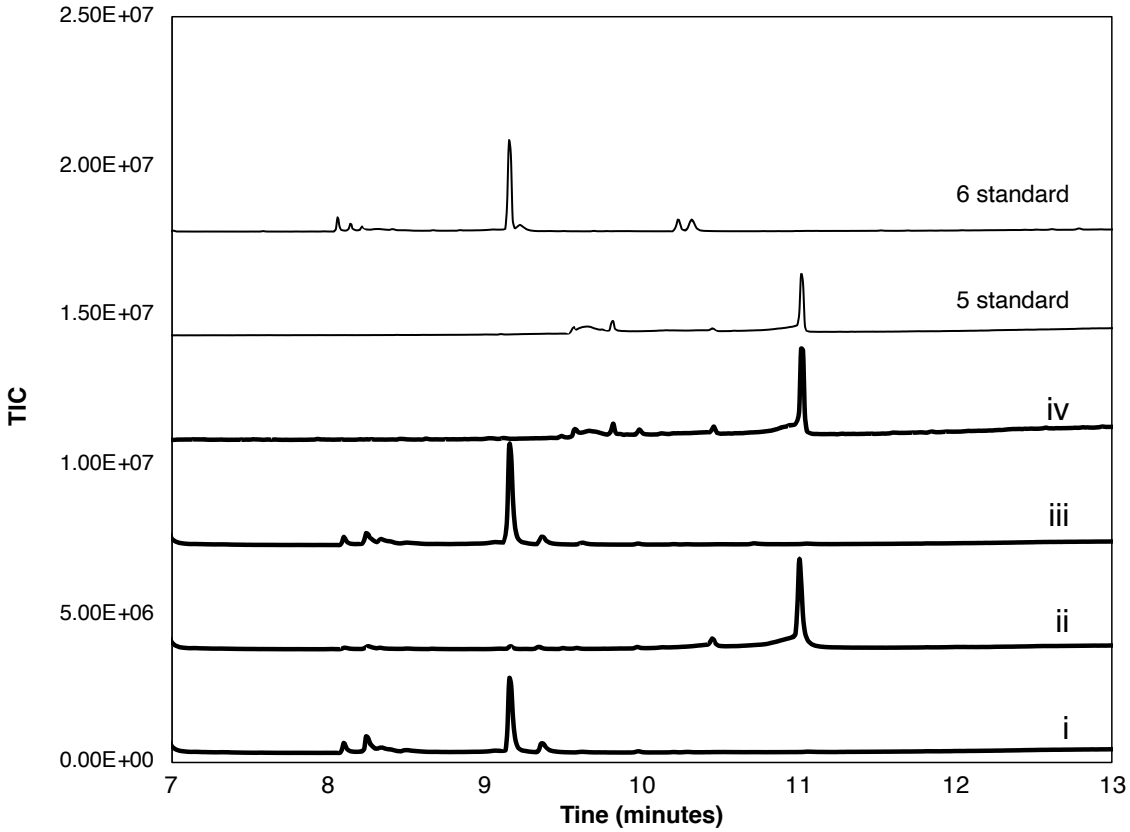


Figure 45. 8-oxogeranial reduction and cyclization by ISY/NmMLPL with NADPH regeneration. Each reaction contains 2 mM 8-oxogeranial, 0.5 μM ISY, 5 μM NmMLPL, 1 μM FumC, 10 μM MaeB and 4 mM NADPH in BTP buffer (pH 9.0) unless otherwise specified, i. no FumC/MaeB, ii. no FumC/MaeB with 100 μM NADPH, iii. with FumC/MaeB and 100 μM NADPH, iv. no ISY/NmMLPL.

5.6 Small-scale *in vitro* biosynthesis of nepetalactol

Despite demonstration of competent *in vitro* activities of the individual enzymes, one-pot synthesis of **6** from **1** as shown in Figure 38 using sub-stoichiometric amounts of cofactors is challenging. First, an orthogonal cofactor regeneration system is required to regenerate NAD^+ from NADH (for the GOR oxidation step), while not oxidizing NADPH, which is required for G8H and ISY turnover. Similarly, the cofactor regeneration system must also regenerate NADPH from $NADP^+$, while not reducing NAD^+ to NADH. Second, the enzyme activities of the regeneration systems must be carefully tuned to match the differential reactivities G8H, GOR and ISY. The

oxidation of **2** to **5** is stepwise and reversible, and can accumulate mono-aldehyde intermediates **3** and **4** (Figure 38). With **3**, ene-reduction catalyzed by ISY follow by aldehyde reduction catalyzed by GOR can give irrecoverable shunt products such as 8-hydroxycitronellol **10**. Hence, the ratio of NAD^+ to NADH available for GOR oxidation must be well-controlled.

With an orthogonal cofactor regeneration system in hand, a one-pot, one-step cell-free biosynthesis of nepetalactol **6** was attempted using all purified enzymes. As expected from the substrate promiscuity of ISY and GOR (Figure 38), incubation of all nine enzymes along with 310 mg/L **1** led to a number of shunt products. Although **5** was produced (15% yield), a considerable amount of citronellol **9** and 8-hydroxycitronellol **10** were formed (Figures 38 and 46).

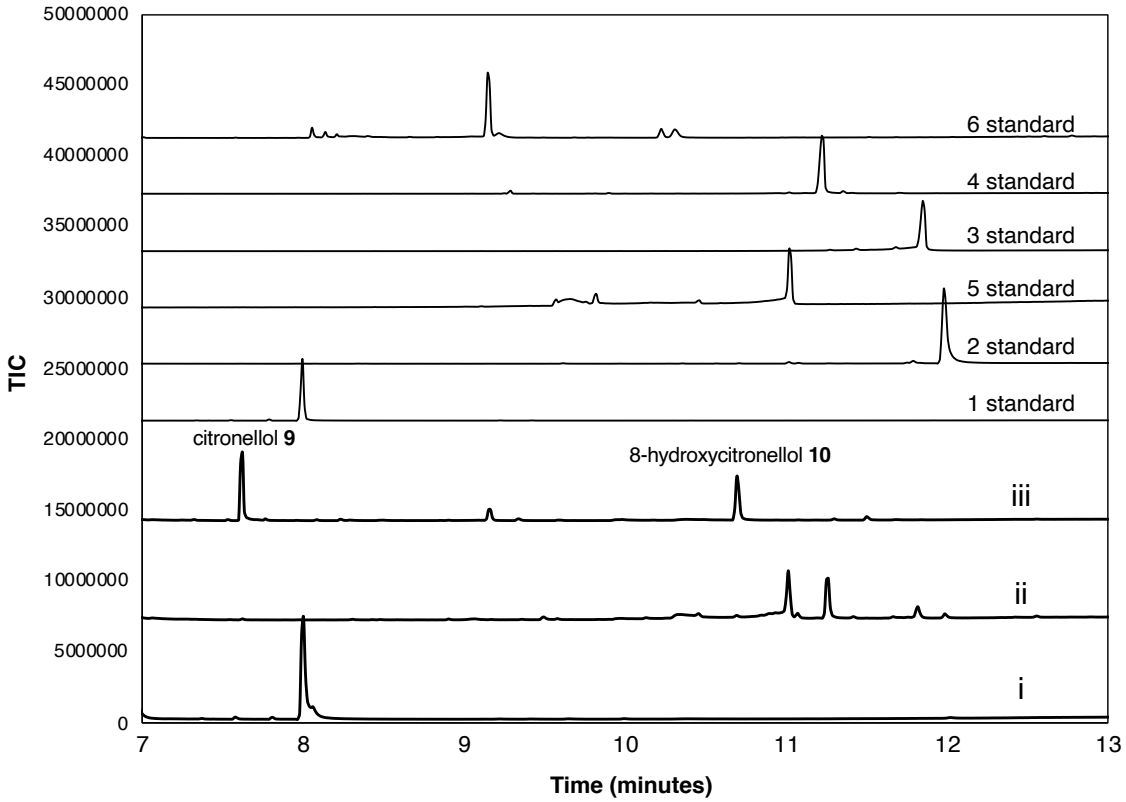


Figure 46. Small-scale one-pot one-step bioconversion of geraniol to nepetalactol. Each reaction contains 2 mM geraniol, 5 μM TfG8H-full, 10 μM FpR, 10 μM YkuN, 10 μM GOR, 0.5 μM ISY, 5 μM NmMLPL, 1 μM FumC, 10 μM MaeB, 100 μM NADPH and 100 μM NAD^+ in BTP buffer (pH 9.0) unless otherwise specified, i.0 min, ii. no ISY, iii. with ISY

The alcohol **9** is formed when **1** is oxidized to geranial **8** by GOR, which then undergoes ene-reduction by ISY, followed by a GOR-catalyzed reduction (Figure 38). To reduce the

formation of **9** which is an irrecoverable shunt product, we used a multi-step, one-pot approach in which ISY was added after G8H and GOR reactions were completed. A 200 µL-scale two-step approach was successful in producing a greater amount of nepetalactol (65% yield) (Figure S23).

5.7 10 mL-scale *in vitro* biosynthesis of nepetalactol and nepetalactone

However, the formation of **9** remained and was particularly problematic in a larger scale reaction (10 mL) (Figure S24). We attributed formation of **9** to low G8H hydroxylation activity caused by oxygen transport deficiency in the reaction vessel, which led to the O_2-independent GOR oxidation and subsequent conversion to **9**.[116,118,123] Increasing the agitation rate resulted in the aggregation of enzymes, causing the reaction to terminate.

To eliminate the formation of **9** and **10** as major shunt products, a one-pot drop-in strategy was pursued, where the biosynthetic enzymes were added sequentially after the upstream reaction was completed. As an initial test, 310 mg/L **1** was incubated with TfG8H, FpR, YkuN, FumC, MaeB along with 100 µM NADPH and 6 mM fumarate for 2 hours. Upon full conversion of **1** to **2**, GOR, NoxE and 100 µM NAD^+ were added directly to the mixture, which was incubated for two additional hours. Finally, ISY, MLPL and 6 mM fumarate were added and reacted for two hours. This scheme fully converted 3.1 mg geraniol to nepetalactol in a 10 mL reaction mixture (Figure S25), forming approximately 3.4 mg nepetalactol in the reaction (>99% conversion). We then attempted to increase the amount of substrate added to the reaction through the addition of multiple aliquots of **1** (Figure 47). The TfG8H reaction was supplemented with an additional 310 mg/L geraniol and 6 mM fumarate every 1.5 hours, and substrate hydroxylation was monitored (Figure 47, trace i-iv). Our NADPH-regeneration system supported hydroxylation of a total of 930 mg/L of **1** to **2** within 4.5 hours. Subsequent addition of GOR and reaction for two hours led to complete conversion of **2** to **3**, **4**, and **5**, and only very minor amount of **8** (Figure 47, trace v). Finally, addition of ISY and MLPL resulted in the formation of 940 mg/L of nepetalactol **6** (93%

yield) in two additional hours (Figure 47, trace vi). Overall, this one-pot mixture operating at 10 mL scale produced ~ 1 g/L of **6** after 8.5 hours (Figure 47, trace vi).

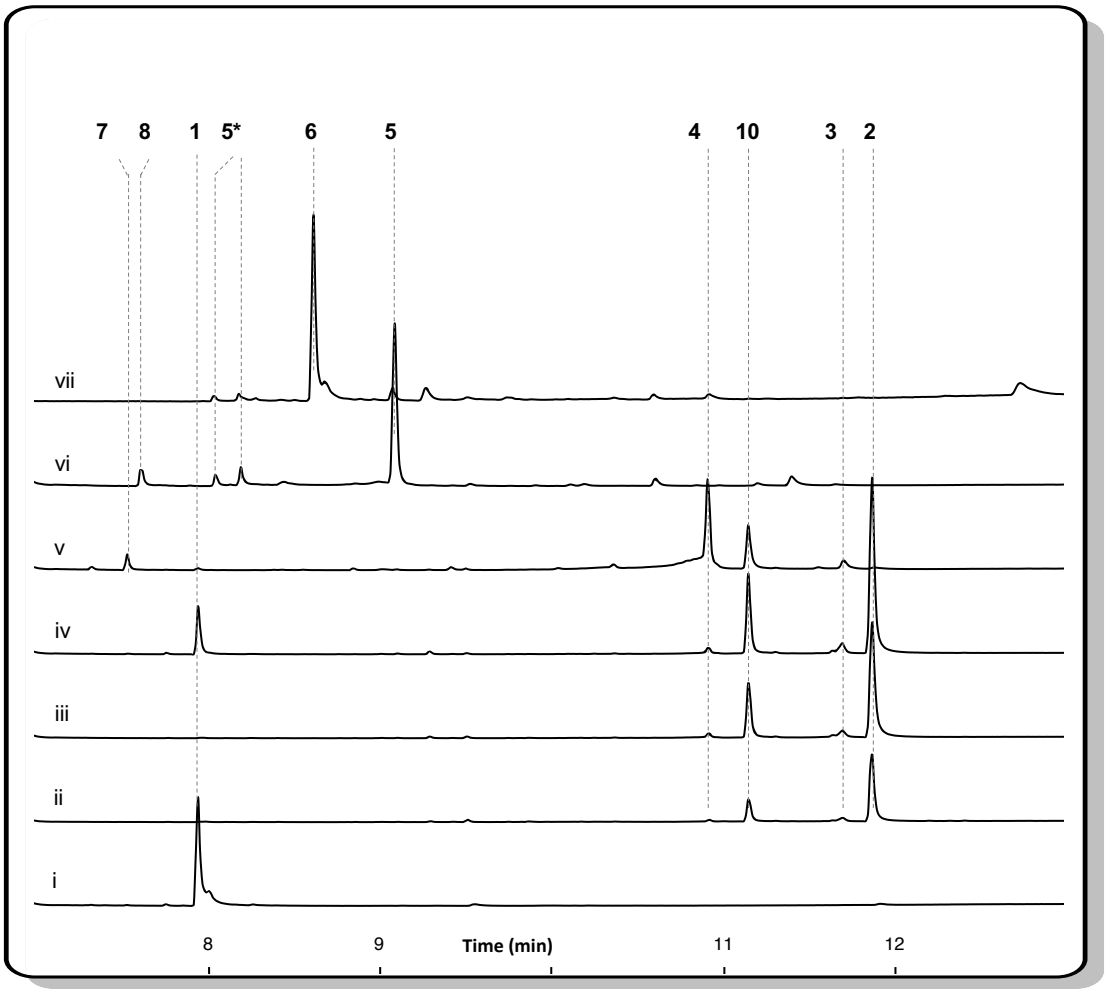


Figure 47. GC-MS chromatograms for 10 mL-scale one-pot conversion of 6 mM geraniol to nepetalactol and nepetalactone. Final reaction contained 6 mM geraniol 1, 5 μM TfG8H, 10 μM FpR, 10 μM YkuN, 10 μM GOR, 0.5 μM ISY, 5 μM MLPL, 1 μM FumC, 10 μM MaeB, 100 μM NADPH, 100 μM NAD^+ and 18 mM fumarate in BTP buffer (pH 9.0) unless otherwise specified. (i) starting material, 6 mM 1, (ii) 1.5-hour reaction with TfG8H, (iii) additional 1.5-hour reaction with TfG8H and 2 mM 1 added, (iv) additional 1.5-hour reaction with TfG8H and 2 mM 1 added, (v) 2-hour reaction after GOR was added to (iv), (vi) 2-hour reaction after ISY/MLPL were added to (v), (vii) 2-hour reaction after ISY/MLPL and NEPS1 were added to (v). Peak identities were deduced from GC-MS and by comparison to authentic standards

Encouraged by the result in Figure 47, we performed the 10 mL reaction with starting batch concentration of **1** at 957 mg/L (6.2 mM) (Figure 48), with no additional aliquots. The concentrations of **1**, **2**, **5** and **6** were measured and plotted as a function of time in Figure 48B.

GOR and NoxE was added after three hours when all of **1** were converted to **2**; while ISY and MLPL were added after five hours when all of **2** were converted to **5**. After two additional hours, all of the **5** were converted to **6** with a final concentration of ~ 1 g/L.

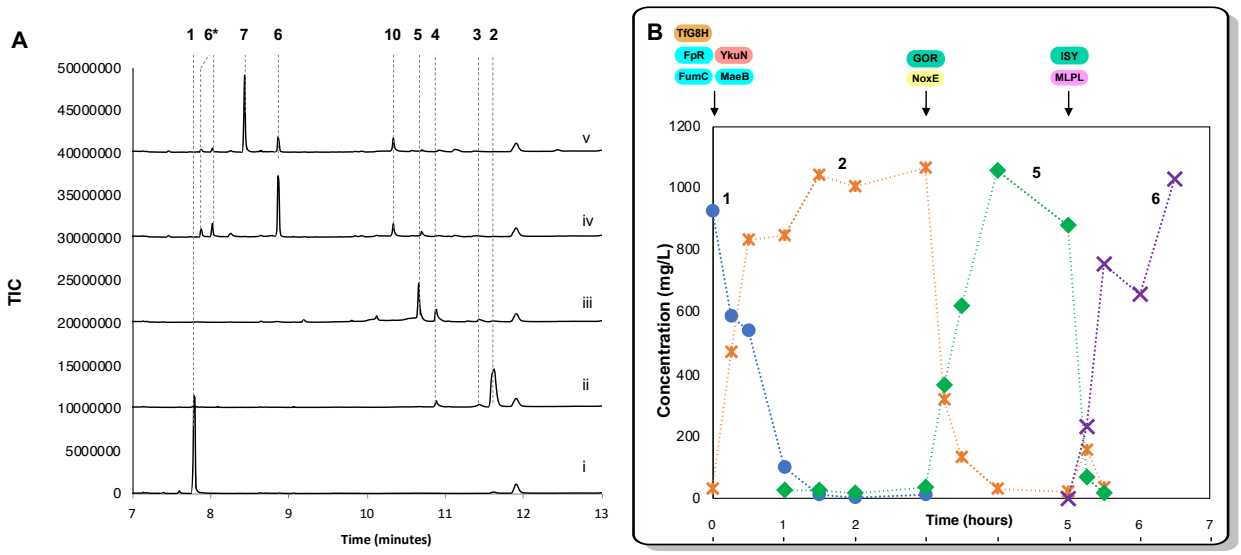


Figure 48. 10 mL-scale one-pot conversion of 6.2 mM geraniol **1** to nepetalactol and nepetalactone. (**A**) GC-MS chromatograms. Final reaction contained 957 mg/L (6.2 mM) geraniol **1**, 5 μM TfG8H, 10 μM FpR, 10 μM YkuN, 10 μM GOR, 0.5 μM ISY, 5 μM MLPL, 1 μM FumC, 10 μM MaeB, 100 μM NADPH, 100 μM NAD^+ and 18 mM fumarate in BTP buffer (pH 9.0). (i) starting material 6.2 mM gearaniol **1** (ii) 3-hour reaction with TfG8H, (iii) 2-hour reaction after GOR was added to (ii), (iv) 2-hour reaction after ISY and MLPL were added to (v), (vii) 2-hour reaction after ISY/MLPL and NEPS1 were added to (v). Peak identities were deduced from GC-MS and by comparison to authentic standards (see Supporting Information). (**B**) Substrate (**1**, blue circle) and products' (**2**, orange star; **5**, green diamond; **6**, purple cross) concentrations measured over time, with timing of added enzymes indicated with arrows atop. The reaction condition is as specified above with starting **1** concentration of 957 mg/L (6.2 mM).

To probe the compatibility of the one-pot reaction with downstream biosynthetic enzymes that act on 6, the nepetalactol-related short-chain reductase/dehydrogenase (NEPS1) was introduced to form 7. Since NEPS1 utilizes NAD^+ to convert nepetalactol to nepetalactone 7, it was added to the reaction mixture at the same time with ISY and MLPL without any additional cofactors or coenzymes. The near complete conversion of 1 to 7 (930 mg/L) was observed after 8.5 hours (Figure 47, trace vii). Production of nepetalactone by recycling sub-stoichiometric concentrations of each nicotinamide cofactor corresponds to 180- and 120-fold decreases in the required molar loading of NAD^+ and NADPH, respectively.

5.8 Conclusion

In summary, we report the production of nearly 1 g/L nepetalactol **6** or nepetalactone **7** from geraniol through the use of a pair of orthogonal cofactor regeneration enzymes. The reaction requires up to five biosynthetic and five auxiliary enzymes and can be operated in a one-pot fashion. Our results highlight a major advantage permitted by cell-free systems – the precise temporal control of enzymatic action which is difficult to program via metabolic engineering.[157–160] Our system produces nepetalactol at a titer ~130-fold greater than the highest reported in a microbial platform[121]. Depending on the estimated cost of protein (Table S3), our total material cost ranges from 60 USD to 120 USD to generate 1 g of **6**, which is significantly lower than current commercial sources. Our platform establishes a cost-effective method to produce **6**, which is useful in the biosynthetic investigation and synthesis of the monoterpene indole alkaloid natural product.

6 CONCLUSION

Our research primarily focused on understanding and utilizing the second-copy self-resistance phenomenon for new natural products and enzymatic catalysts discovery. In Section 2, we aimed to elucidate the AA resistance mechanism of AstD. The co-crystal structure of pDHAD-AA complex obtained by our collaborators allowed us to build a homology model of AstD. As a result, we were able to deduce amino acids that may be responsible for AstD AA resistance. Our kinetic analysis of the mutant pDHADs *in vitro* confirmed our hypothesis that larger hydrophobic amino acids at locations surrounding the pDHAD active-site entrance increase the AA resistance of pDHAD. However, it also showed us that resistance comes with the compromise of catalytic activity, exemplified by the decrease in k_{cat}. Furthermore, our *in vitro* and i*n vivo* analysis showed that there are other contributing factors to AstD's extremely high AA resistance.

In section 3, our work revealed the enzymatic steps involved in the biosynthesis of harzianopyridone. The BGC of harzianopyridone was found with the help of SRE-DGM and enzymes HarABCDEG have been confirmed, revealing unusual, concerted catalysis by FMO and OMT enzymes to install the methoxy moeities in harzianopyridone. Notably, four of the six enzymes (HarA, G, B and C) perform iterative catalysis, which underscores the highly programmed and unpredictable functions of fungal biosynthetic enzymes. In section 4, our efforts in elucidating the biosynthesis of atpenin A5 has led us to a novel halogenase enzyme which installs a halogen onto a non-activated aliphatic carbon.

Finally, in section 5, we report the production of nearly 1 g/L nepetalactol **6** or nepetalactone **7** from geraniol through the use of a pair of orthogonal cofactor regeneration enzymes. The reaction requires up to five biosynthetic and five auxiliary enzymes and can be operated in a one-pot fashion. Our results highlight a major advantage permitted by cell-free systems – the precise temporal control of enzymatic action which is difficult to program via

metabolic engineering.[157–160] Our system produces nepetalactol at a titer ~130-fold greater than the highest reported in a microbial platform[121]. Depending on the estimated cost of protein (Table S3), our total material cost ranges from 60 USD to 120 USD to generate 1 g of **6**, which is significantly lower than current commercial sources. Our platform establishes a cost-effective method to produce **6**, which is useful in the biosynthetic investigation and synthesis of the monoterpene indole alkaloid natural products.

7 MATERIALS AND METHODS

7.1 Strains and general culture conditions

Aspergillus terreus NIH 2624 was grown at 28°C for 14 days in stationary liquid PDB medium for the production aspterric acid or for genomic DNA or mRNA extractions.

Trichoderma harzianum ATCC® 64870™ was grown on PDA (Potato Dextrose Agar: 2.4% PDB medium, 2% Agar) media at 28°C for 7 days for cell proliferation or in liquid PDB medium for isolation of genomic DNA.

*Aspergillus nidulan*s A1145 ΔEM[91] was grown at 28°C in CD media (0.1% Glucose, 0.5 v/v% 20×Nitrate salts, 0.01 v/v% Trace elements, pH 6.5, and 2% agar for solid media) for sporulation or in CD-ST media (2% starch, 2% Casamino acids, 5 v/v% 20×Nitrate salts, 0.1 v/v% Trace elements, pH 6.5) for heterologous expression of biosynthetic gene cluster, compound production and mRNA extraction. 20×Nitrate salts was prepared by dissolving 120 g $NaNO_3$, 10.4 g KCl, 10.4 g $MgSO_4 \cdot 7H_2O$, 30.4 g KH_2PO_4 in 1 L distilled water. The Trace elements solution contained 2.20 g $ZnSO_4 \cdot 7H_2O$, 1.10 g H_3BO_3, 0.50 g $MnCl_2 \cdot 4H_2O$, 0.16 g $FeSO_4 \cdot 7H_2O$, 0.16 g $CoCl_2 \cdot 5H_2O$, 0.16 g $CuSO_4 \cdot 5H_2O$ and 0.11g $(NH_4)_6\ Mo_7O_{24} \cdot 4H_2O$ in 100 mL distilled water and the pH adjusted to pH 6.5[161].

Penicillium oxalicum FO125 strain was grown on PDA media at 28°C for 3 days for cell proliferation or in liquid PDB medium for isolation of genomic DNA. *P. oxalicum* and the subsequent knock-out strains were grown in Atpenin-producing media (1.0% Glucose, 0.5% Tryptone, 0.3% yeast extract, 0.3% malt extract, 0.1% agar, pH 6.0) at 28°C for 2-4 days for atpenin-related compound production and mRNA extraction.

Escherichia coli BL21(DE3) (Novagen) strain was used for heterologous protein expression. All *E. coli* strains were cultured in LB media at 37°C for cloning or 16°C for protein expression.

Saccharomyces cerevisiae YJB077 (*MATα his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 prb1Δ pep4Δ oye2Δ1 oye3Δ1 ari1Δ0 adh7Δ0 adh6Δ1 SAL1+CAT5(91M) MIP1 (661T) MKT1(30G) RME1(INS-308A) TAO3(1493Q) HAP1+*)[162] was used for yeast biotransformations. *S. cerevisiae* DHY ΔURA3 strain was used to generate the ILV3 knock-out strain. Yeast strains were cultured in uracil drop-out media for selection or YPD (1% yeast extract, 2% peptone, 2% glucose) media for protein expression and biotransformation at 28°C.

7.2 General DNA manipulation techniques

All DNA manipulations in this study were conducted according to manufacturers' protocols. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs, NEB). PCR was performed according to recommended protocol using Q5® High-Fidelity DNA polymerase (NEB). The primers used in this study are listed in Table S2. PCR products were confirmed by DNA sequencing and plasmid DNAs were confirmed by restriction enzyme digestion checks and sequencing. *E. coli* TOP10 and *E. coli* XL-1 strains were used for cloning. *S. cerevisiae* BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*)[163] strain was used as the yeast host for *in vivo* yeast homologous recombination (YHR) to construct the *A. nidulans* plasmids.

The RNA extractions were performed using RiboPureTM Yeast RNA Isolation Kit (Ambion) following the manufacturer's instructions. Residual genomic DNA in the extracts was digested by DNase I (2 U/μL) (Invitrogen) at 37°C for 4 hours. SuperScript III First-Strand Synthesis System (Invitrogen) was used for cDNA synthesis with Oligo-dT primers following instructions from the user manual.

The genes for TfG8H, YkuN, PdR, Pdx and NmMLPL were synthesized by IDT technologies. They were then cloned into pCR-blunt vector by ligation and transformed into *E. coli TOP10* strain for storage.

7.3 Generating mutant pDHAD plasmids

The intron free ORF of pDHAD gene was amplified using gBlock of pDHAD as a template by PCR with primers listed in Appendix B. The PCR products were then inserted into linearized pET28a via digestion/ligation, according to the manufacturer's protocol, generating pEUB10001. pEUB10001 was used as a templated for introducing point amino acid mutations using primers listed in Appendix B via PCR amplification of the whole plasmid. The resultant plasmids were sequenced and verified to have the correct mutations.

7.4 Large-scale aspterric acid isolation

For large scale isolation of aspterric acid, *A. terreus* grown for 14 days in 6 L stationary liquid PDB media and then filtered through cloth to separate the cell mass from culture liquid. The cell mass was extracted with 1 L acetone × 3 and the extracts were dried under reduced pressure by Buchi Rotavapor. The organic solutes in the culture liquid were extracted with equal volume of EA three times and dried under reduced pressure. The extracts from both cell mass and liquid was combined to yield crude extract oil which were then subjected to normal-phase silica column CombiFlash purification with hexane, followed by reversed-phase CombiFlash system, using a linear gradient of 5-95% MeOH-water with 0.1% formic acid in 35 min followed by 95% MeOH for 10 min with a flow rate of 6 mL/min (with 40 g silica column) for initial separation. Fractions containing aspterric acid were combined and used for further purification by HPLC with a semi-preparative reversed-phase column using a linear gradient of 65–100% (v/v) acetonitrile/water in 25 min, with a flow rate of 2.5 ml min^{-1}. HPLC purification was performed using a Shimadzu Prominence HPLC (Phenomenex Kinetex, 5μ, 10.0 × 250 mm, C-18 column).

7.5 Generating yeast knock-out strains for aspterric acid inhibition assays

The *URA3* gene was inserted into the *ILV3* locus of *S. cerevisiae* DHY Δ*URA3* strain to generate UB01. A 879-bp homologous-recombination donor fragment with 35–40 bp homologous regions flanking the *ILV3* ORF was amplified using primers listed in Appedix B using yeast gDNA as a template. The PCR product was gel purified and transformed into *S. cerevisiae* DHY Δ*URA3*, and selected on uracil dropout medium to give UB01. The resulting strain was subjected to verification using colony PCR and the amplified fragment was confirmed with sequencing.

The *URA3* gene inserted into the *ILV3* locus of *S. cerevisiae* DHY Δ*URA3* was deleted from UB01 using homologous recombination to generate UB02. A 150-bp homologous-recombination donor fragment with 75-bp homologous regions flanking the *ILV3* ORF was amplified using primers listed in Appendix B, gel purified, transformed into UB01, and counter-selected on 5-fluoroorotic acid (5-FoA)-containing medium to give UB02. The resulting strain was subjected to verification using colony PCR and the amplified fragment was confirmed with sequencing.

The empty plasmid pXP318 (*URA3* marker) was transformed into UB02 to generate empty vector control. Plasmid pXP318 digested with *Spe*I and *Xho*I was used as vector to introduce the gene encoding fDHAD into the plasmid UB02. The cDNA of *A. terreus* NIH 2624 served as the

template for PCR amplification. A 1.7-kb fragment obtained using primers listed in Appendix B were cloned into pXP318 using yeast homologous recombination to produce pXP318-fDHAD. Then, pXP318 - fDHAD was transformed into UB02 to generate fDHAD expressing yeast strain. fDHAD was driven by a constitutive *TEF1* promoter. Plasmid pXP318 digested with *Spe*I and *Xho*I was used as vector to introduce the *astD* gene into the yeast UB02. The cDNA isolated from *A. terreus* served as the template for PCR amplification. A 1.8-kb fragment obtained using primers listed in Appendix B was cloned into pXP318 using yeast homologous recombination to make pXP318-AstD. A Flag tag was also added to the N-terminal of AstD. AstD-pXP318 was then transformed into UB02 to generate AstD expressing yeast strain. AstD was driven by the constitutive *TEF1* promoter.

7.6 *in vivo* aspterric acid inhibition assays in yeast

S. cerevisiae UB02 transformants were grown in isoleucine, leucine and valine (ILV) dropout medium (20 g/L glucose, 0.67 g/L Difco Yeast Nitrogen Base without amino acids, 18 mg/L adenine, 76 mg/L arginine, 76 mg/L asparagine, 76 mg/L aspartic acid, 76 mg/L glutamic acid, 76 mg/L histidine, 76 mg/L lysine, 76 mg/L methionine, 76 mg/L phenylalanine, 76 mg/L serine, 76 mg/L threonine, 76 mg/L tryptophan, and 76 mg/L tyrosine) to test growth inhibition of aspterric acid on *S. cerevisiae*. Cells were incubated at 28 °C until OD_{600} nm of the control strain without aspterric acid treatment reached about 0.8. The inhibition curve was plotted as the percentage of inhibition versus AA concentrations. The OD_{600} nm was recorded for every 20 min over a total of 50 h. The growth inhibition percentage of AA on *S. cerevisiae* strains were calculated by dividing the cell density (OD_{600} nm) of the AA-treated strains to the corresponding untreated strains when OD_{600} nm reaches approximately 0.8 using the following equation, in which 0.8 is the OD_{600} nm of the untreated strain.

growth inhibition percentage = 100 − 100x(OD600 nm of **AA** treated strain/0.8)

7.7 Generating transgenic *A. thaliana* expressing mutant pDHADs

The mutant pDHADs with the chloroplast-localization signal (CLS) of 35 amino acid residues were cloned into pEG202 vector using Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific). A 3 × Flag-tag was inserted between the CLS and the pDHADs. The original CaMV 35S promoter of pEG202 was substituted by the ubiquitin-10 promoter to drive the expression of pDHADs. The construct was electrotransformed into *Agrobacterium tumefaciens* strain Agl0 and then transformed into *A. thaliana* using the standard floral dip method.[164] The *A. thaliana* Col-0 ecotype was transformed. Positive transgenic plants were selected using the glufosinate resistance marker, and were tested for survival in the presence of aspterric acid.

7.8 Growth inhibition assay of plants on agar plates

MS (2.16 g/L Murashige and Skoog basal medium, 8 g/L sucrose, 8 g/L agar) medium was used to test the growth inhibition of aspterric acid on *A. thaliana. A. thaliana* transgenic plant lines were grown under long day condition (16/8 h light/dark) using cool-white fluorescence bulbs as the light resource at 23°C. Aspterric acid was dissolved in ethanol and added to the medium

before growing plants. The medium of the control treatment contained the same amount of ethanol, but without aspterric acid.

7.9 Fungal protoplast preparation

A. nidulans A1145 ΔEM[91] was initially grown on CD agar plates containing supplements (10 mM uridine, 5 mM uracil, 0.5 µg/mL pyridoxine HCl and 2.5 µg/mL riboflavin) at 28°C for 5 days. Fresh spores of *A. nidulans* A1145 ΔEM were inoculated into 50 mL liquid CD media containing supplements in 250 mL flask and germinated at 30°C, 250 rpm for approximately 16 h. Mycelia were harvested by centrifugation at 3,500 rpm for 10 min, and washed with 10 mL osmotic buffer (1.2 M $MgSO_4$, 10 mM sodium phosphate, pH 5.8). Then the mycelia were transferred into 10 mL of osmotic buffer containing 30 mg lysing enzymes from *Trichoderma* and 20 mg Yatalase in a 125 mL flask and were shaken at 80 rpm for overnight at 30°C. Cells were collected in a 30 mL Corex tube and overlaid gently by 10 mL of trapping buffer (0.6 M sorbitol, 0.1 M Tris-HCI, pH 7.0). After centrifugation at 3,500 rpm for 15 min at 4°C, protoplasts were collected from the interface of the two buffers. The protoplasts were then transferred to a sterile 15 mL falcon tube and washed by 10 mL STC buffer (1.2 M sorbitol, 10 mM $CaCl_2$, 10 mM Tris-HCI, pH 7.5). The protoplasts were then resuspended in 1 mL STC buffer and 60 µL aliquots of the protoplasts were stored in -80°C for transformation.

P. oxalicum FO125 and knock-out strains was initially grown on PDA plates (supplemented with 300 mg/mL hygromycin in cases of knock-out strains) at 28°C for 3 days for sporulation. Fresh spores were collected and inoculated into 50 mL liquid CD media in 250 mL flask and germinated at 30°C, 250 rpm for approximately 16 h. Mycelia were harvested by centrifugation at 4.300 x g for 15 min, and washed with 10 mL osmotic buffer. The mycelia were transferred into 10 mL of osmotic buffer containing 30 mg lysing enzymes from *Trichoderma* and 30 mg Yatalase in a 125 mL flask and were shaken at 80 rpm for overnight at 30°C. Cells were collected in a 30 mL Corex tube and overlaid gently with 10 mL of trapping buffer. After centrifugation at 3,500 rpm for 15 min at 4°C, protoplasts were collected from the interface of the two buffers. The protoplasts were then transferred to a sterile 15 mL falcon tube and washed by 10 mL STC buffer. The protoplasts were then resuspended in 600 µL STC buffer and 100 µL fresh aliquots of the protoplasts were used for generating knock-out strains.

7.10 Heterologous expression of the *har* and *apn* gene clusters in *A. nidulans*

All *A. nidulans* plasmids for heterologous expression in *A. nidulans* A1145 ΔEM were built using the plasmids pYTU, pYTP, and pYTR with auxotrophic markers for uracil (*pyrG*), pyridoxine (*pyroA*), and riboflavin (*riboB*), respectively, as backbones to insert genes. gpdA promoters from *A. nidulans*, *Penicillium oxalicum* (POgpdA), *Penicillium expansum* (PEgpdA), amyB and glaA promoters were amplified by PCR.

All genes in *har* and *apn* gene clusters were amplified from genomic DNA extract from *T. harzianum* and *P. oxalicum* FO-125, respectively. *harA* gene was amplified by PCR with four pairs of primers listed in Appendix B. The overlapping DNA fragments and *Pac*I/*Swa*I-digested pYTU expression vector were co-transformed into *S. cerevisiae* BJ5464-NpgA and selected on uracil dropout media generating pAUB60001 plasmid expressing *harA* gene via *in vivo* YHR. Similarly, *harE* was amplified and ligated into vector pYTR and *harG* into pYTP, yielding plasmids pAUB60002 and pAUB60003, respectively. *harF* or *harC* was cloned into pAUB60002 separately, yielding plasmid pAUB60004 or pADK60006, respectively. *harD* was cloned into pAUB60003 to generate plasmid pAUB60005, then *harB* was inserted into pAUB60005 to get pAUB60007. pAUB60009 plasmid was constructed to express *harG* and *harB* together. Plasmids were

extracted from yeast using Zymoprep™ Yeast Plasmid Miniprep I Kit (Zymo Inc. USA), and transformed into *E. coli* TOP10 by electroporation to obtain transformants with single plasmid. After plasmid extraction from *E. coli* using Zyppy Plasmid Miniprep Kit (Zymo Research, USA), plasmids were sequenced to confirm identities. Plasmids containing genes from *apn* cluster was constructed similarly to *har* gene cluster reconstitution. For *A. nidulans* transformation, necessary plasmids were added to 60 µL *A. nidulans* A1145 protoplast suspension prepared above and the mixture was incubated on ice for 60 min. After incubating on ice, 600 µL of PEG solution (60% PEG, 50 mM calcium chloride and 50 mM Tris-HCl, pH 7.5) was added to the protoplast mixture, followed by additional incubation at room temperature for 20 min. The mixture was spread on the regeneration medium (CD solid medium with 1.2 M sorbitol and appropriate supplements including 10 mM uridine, 5 mM uracil and/or 0.5 µg/mL pyridoxine HCl and/or 2.5 µg/mL riboflavin depending on the plasmids being transformed) and incubated at 30°C for 2-3 days until single colonies appear.

apnA gene was amplified by PCR with four pairs of primers listed in Appendix B. The overlapping DNA fragments and *Pac*I/*Swa*I-digested pYTU expression vector were co-transformed into *S. cerevisiae* BJ5464-NpgA and selected on uracil dropout media generating pAUB70001 plasmid expressing *apnA* gene via *in vivo* YHR. Similarly, *apnE*, *apnD,* and *apnC* were amplified and ligated into vector pYTR and *apnG* and *apnB* into pYTP, yielding plasmids pAUB70002 and pAUB70003, respectively. *apnU* was inserted into pAUB70003 to give pAUB70004. pAUB70005 was constructed by inserting *apnE* into pAUB70001, and pAUB70006 was constructed to express *apnD*, *apnC* and *apnV*. *apnT* was inserted into pAUB70004, to generate pAUB70007, and inserting *apnS* into pAUB70006 generated pAUB70008. pAUB70007 was constructed to express *apnD*, *apnC*, *apnF* and *apnS*. The plasmids were and transformed into *E. coli* as described above and sequenced for confirmation. For *A. nidulans* transformation, necessary plasmids were transformed into 60 µL *A. nidulans* A1145 protoplast suspension as described above, spread on the regeneration medium and incubated at 30°C for 2-3 days until single colonies appear.

7.11 Generating *P. oxalicum* knock-out strains

2000 bp flanking region of *apnU* was amplified by using primers listed in Appendix B from gDNA of *P. oxaliucm* FO125 and was cloned into pHyg plasmid to be on either side of hygromycin B phosphotransferase gene (*hph*) fused to promoter (TrpCp) and terminator (TrpCt) elements. The resulting plasmid pPO70001 was used as a template to generate two PCR products with 500 bp homologous region, where one includes *N*-terminal 2000 bp region, TrpCp and approximately half of *hph* sequence, and the other includes the rest of *hph* gene sequence, TrpCt, and *C*-terminal 2000 bp region. The PCR products were gel-purified and lyophilized before getting transformed into fresh *P. oxalicum* protoplasts. The protoplast transformation was performed as described in Section 7.5, and the resultant protoplast suspension was spread on CD agar plates supplemented with 300 mg/mL hygromycin B. After 3-4 days of incubation at 30°C, the single colonies were subjected to gDNA extraction and PCR amplification to confirm the knock-out. *apnS*, *apn*T, *apnV* and *apnG* genes were knocked-out similarly.

7.12 Generating *P. oxalicum* transient expression strains

apnT or *apnU* genes were amplified using primers listed in Appendix B and cloned into pBAR plasmid. The plasmids were designed with glyceraldehyde-3-phosphate dehydrogenase promoter (gpdAp) fused to either *apnT* or *apnU* gene with its native terminator region (500 bp) and the phosphinothricin N-acetyltransferase (barR) gene is fused to promoter (TrpCp) and

terminator (TrpCt) elements of the *A. nidulans* trpC gene. The resultant plasmids pPO70007 (ApnT-expressing) and pPO70006 (ApnU-expressing) were linearized with restriction enzymes and transformed into ΔApnT and ΔApnU strain protoplasts, respectively, as described above. The resultant protoplast suspension was then spread onto PDA plates supplemented with 300 mg/mL hydromycin B and 100 mg/mL glufosinate and incubated at 30°C for 3-4 days until single colonies appeared. The transformants were subjected to gDNA extraction and PCR amplification to confirm the expression of ApnT or ApnU.

7.13 Analysis of metabolites and isolation of compounds from *A. nidulan*s transformants

For small scale *A. nidulans* transformant product analysis, the transformants were grown in 10 mL liquid CD-ST for 3-4 days under continuous shaking at 250 rpm and extracted with ethyl acetate/acetone (EA:acetone = 75:25) mixture. The organic phase was dried by SpeedVac and dissolved in methanol for analysis in LC-MS. LC-MS analyses were performed on a Shimadzu 2020 EV LC-MS with a reverse-phase column (Phenomenex Kinetex, C18, 1.7 µm, 100 Å, 2.1 × 100 mm) using positive-and negative-mode electrospray ionization with a linear gradient of 5-95% acetonitrile (MeCN)-water (H_2O) with 0.1% formic acid in 15 min followed by 95% MeCN for 3 min with a flow rate of 0.3 mL/min. HRMS-QTOF analyses of metabolites were performed on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS with a reverse-phase column (Agilent InfinityLab Poroshell 120 EC-C18, 2.7 µm, 3.0 × 50 mm) using positive electrospray ionization with 1% MeCN-H_2O with 0.1% formic acid for 2 min followed by a linear of 1-99% MeCN-H_2O with 0.1% formic acid for 9 min with a flow rate of 0.4 ml/min.

For large scale isolation of compounds, *A. nidulans* transformants were grown for 4-5 days in 4 L liquid CD-ST and then filtered through cloth to separate the cell mass from culture liquid. The cell mass was extracted with 1 L acetone × 3 and the extracts were dried under reduced pressure by Buchi Rotavapor. The organic solutes in the culture liquid were extracted with equal volume of EA three times and dried under reduced pressure. The extracts from both cell mass and liquid was combined to yield crude extracts which were then subjected to reversed-phase CombiFlash system, using a linear gradient of 5-95% MeOH-water with 0.1% formic acid in 35 min followed by 95% MeOH for 10 min with a flow rate of 6 mL/min (with 40 g silica column) for initial separation. Fractions containing target compounds were combined and used for further purification by HPLC with a semi-preparative reversed-phase column using an isocratic concentration of 45% MeCN – 55% water for 35 min followed by 100% MeCN for 10 min using the Phenomenex Kinetics, C18 column (5 µm, 100 Å, 10 × 250 mm). For elucidation of chemical structures, 1D and 2D NMR spectra were obtained on a Bruker AV500 spectrometer at the UCLA Molecular Instrumentation Center.

7.14 Analysis of metabolites and isolation of compounds from *P. oxalicum* strains

For metabolite analysis *P.oxalicum* strains, the wilt-type *P.oxalicum* or its transformants were grown in 10 mL liquid Atpenin-producing media for 3 days under continuous shaking at 250 rpm and the culture broth extracted with ethyl acetate and mycelium extracted with acetone. The organic phase was dried by SpeedVac and dissolved in methanol for analysis on LC/MS or HRMS-QTOF. LC/MS analyses were performed on a Shimadzu 2020 EV LC-MS with a reverse-phase column (Phenomenex Kinetex, C18, 1.7 µm, 100 Å, 2.1 × 100 mm) using positive-and negative-mode electrospray ionization with a linear gradient of 5-95% acetonitrile (MeCN)-water

(H_2O) with 0.1% formic acid in 15 min followed by 95% MeCN for 3 min with a flow rate of 0.3 mL/min. HRMS-QTOF analyses of metabolites were performed on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS with a reverse-phase column (Agilent InfinityLab Poroshell 120 EC-C18, 2.7 µm, 3.0 × 50 mm) using positive electrospray ionization with 1% MeCN-H_2O with 0.1% formic acid for 2 min followed by a linear of 1-99% MeCN-H_2O with 0.1% formic acid for 9 min with a flow rate of 0.4 ml/min.

7.15 Heterologous biotransformation in *S. cerevisiae* YJM077

The full ORF of *harD* was amplified from pADK60005 with primers pYUB60001 F1/R1 and cloned into 2µ-based yeast expression vector (pJB031) under ADH2-like promoters[162] using YHR in *S. cerevisiae* YJM077 strain, yielding plasmid pYUB60001. The yeast was transformed with the over-lapping PCR fragments via lithium acetate transformation method[165] and selected on uracil dropout media. *harC* and *harB* were amplified from their corresponding *A. nidulans* plasmids and cloned similarly to yield pYUB60002 and pYUB60003, respectively. *harC* and *harB* were inserted separately into pYUB60001 to generate pYUB60004 and pYUB60005, whilst pYUB60006 was cloned to express *harB* and *harC* together. pYUB60007 was built in a similar manner to express all 3 genes. The plasmids were then extracted, transformed into *E. coli* TOP10 strain to singularize the plasmids. The resulting plasmids were also sequenced for confirmation. *S. cerevisiae* YJM077 strain transformed with individual plasmids were used for biotransformation. The transformant yeast strains were selected on solid uracil drop-out media for 2-3 days, then single colonies were inoculated into 3 mL uracil drop-out media and grown for 24 hours to be used as inoculums for biotransformation in YPD media. 3 mL YPD media was inoculated with yeast transformants where optical density at 600 nm (OD_{600}) was 0.1 and grown for 24 hours before biosynthetic intermediates were added at concentrations of 33 µM each. After the addition of the substrates, the culture was grown for an additional 12 hours, before the pellet was extracted with acetone and supernatant with ethyl acetate. The extracts were dried by SpeedVac and dissolved in methanol for analysis in LC-MS as described above.

7.16 Protein expression and purification from *E. coli* BL21(DE3)

The plasmids for pDHAD and its mutants were transformed into *E. coli* BL21(DE3) individually and grown overnight in 5 mL of LB medium with 50 µg/mL kanamycin at 37°C. The overnight cultures were used as seed cultures for 1 L fresh LB media containing 50 µg/mL kanamycin and incubated at 37°C until the OD_{600} reached 0.6. The cultures were cooled on ice, before the protein overexpression was induced with 0.1 mM isopropyl-*β*-D-thiogalactopyranoside (IPTG, GoldBio, USA) and at 16°C overnight. The cultures were supplemented with 125 µM $FeSO_4$ at the same time as IPTG induction. The *E. coli* cells were harvested by centrifugation at 5300 rpm for 15 min and resuspended in 30 mL A10 buffer without NaCl (50 mM Tris-HCl, 10 mM imidazole, pH 8.0) containing 1 tablet of Pierce™ protease inhibitor (Thermo Scientific). The cell suspension was lysed on ice by sonication and the lysate was centrifuged at 17,000 *g* for 15 min at 4°C to remove the insoluble cellular debris. The recombinant *C*- terminally hexahis-tagged pDHADs were purified individually from corresponding soluble fractions by affinity chromatography with Ni-NTA agarose resin. The suspension of Ni-sepharose and lysate was incubated at 4°C for 2 hours, before washing with 50 mL A10 buffer without NaCl and 25 mL A25 buffer without NaCl. The proteins were then eluted with 1.5 mL A250 buffer without Nacl and used directly for assays. SDS-PAGE was performed to check the protein purity and Bradford Protein Assay (Bio-Rad) was used to calculate protein concentration with bovine serum albumin (BSA, Sigma-Aldrich) as standard.

The intron free ORFs of *harB* and *harC* genes were amplified using cDNA from the corresponding *A. nidulans* transformant as a template by PCR with primers listed in Appendix B. The PCR products were then inserted into linearized pET28a via Gibson assembly, according to the manufacturer's protocol. The resultant plasmids were sequenced and verified to contain no introns. The plasmids were then transformed into *E. coli* BL21(DE3) individually and the proteins were as described above, without the supplementing of $FeSO_4$. The cell suspension was lysed on ice by sonication and the lysate was centrifuged at 17,000 *g* for 15 min at 4°C to remove the insoluble cellular debris. The recombinant *N*-and *C*- terminally hexahis-tagged HarB and *C*-terminally hexahis-tagged HarC were purified individually from corresponding soluble fractions by affinity chromatography with Ni-NTA agarose resin (Qiagen) according to the manufacturer's instructions. The purified proteins were concentrated and exchanged into storage buffer (50 mM Phosphate buffer, 100 mM NaCl, 10% glycerol, pH 8.0) with Centriprep filters (Amicon). SDS-PAGE was performed to check the protein purity and Bradford Protein Assay (Bio-Rad) was used to calculate protein concentration with bovine serum albumin (BSA, Sigma-Aldrich) as standard. The proteins were aliquoted and stored at -80°C until used in *in vitro* assays.

The plasmids for nepetalactol production listed in Appendix B were transformed into *E. coli* BL21(DE3) individually, and the proteins were expressed as described above. Cultures were supplemented with 125 μM $FeSO_4$ for iron-containing enzymes and 80 g/L 5'-aminolevulinic acid for P450 expression. The *E. coli* cells were harvested by centrifugation at 5300 rpm for 15 min and the cell pellets were stored at -80°C for later use. All transformants except for transformant harboring pEUB20006 were resuspended in 30 mL A10 buffer (50 mM Tris-HCl, 100 mM NaCl, 10 mM imidazole, pH 8.0) containing 1 tablet of Pierce™ protease inhibitor (Thermo Scientific). The cell suspension was lysed on ice by sonication and the lysate was centrifuged at 17,000 *g* for 15 min at 4°C to remove the insoluble cellular debris. The recombinant *C*- or *N*- terminally hexa-His-tagged proteins were purified individually from corresponding soluble fractions by affinity chromatography with Ni-NTA agarose resin (Qiagen) according to the manufacturer's instructions. The purified proteins were concentrated and exchanged into storage buffer (50 mM Phosphate buffer, 100 mM NaCl, 10% glycerol, pH 8.0) with Centriprep filters (Amicon). SDS-PAGE was performed to check the protein purity and Bradford Protein Assay (Bio-Rad) was used to calculate protein concentration with bovine serum albumin (BSA, Sigma-Aldrich) as standard. The proteins were aliquoted and stored at -80°C until used in *in vitro* assays.

The cell pellet harboring pEUB20006 was resuspended in 30 mL MOPS10 buffer (50 mM MOPS, 100 mM NaCl, 10 mM imidazole, 1% Triton-X100, pH 7.5) containing 1 tablet of Pierce™ protease inhibitor (Thermo Scientific). The cell suspension was lysed on ice by sonication and the lysate was centrifuged at 17,000 *g* for 15 min at 4°C to remove the insoluble cellular debris. The recombinant *C*- terminally hexa-His-tagged protein was purified from corresponding soluble fractions by affinity chromatography with Ni-NTA agarose resin (Qiagen) according to the manufacturer's instructions. The purified protein was buffer exchanged into MOPS10 buffer with 10% glycerol with cellulose membrane. SDS-PAGE was performed to check the protein purity and Bradford Protein Assay (Bio-Rad) was used to calculate protein concentration with bovine serum albumin (BSA, Sigma-Aldrich) as standard. The protein was aliquoted and stored at -80°C until used in *in vitro* assays.

7.17 *in vitro* bioactivity assays of pDHAD and mutants

In vitro activity assays were carried out in 25 μL reaction mixture containing A10 buffer without NaCl, 10 mM (±)-sodium-α,β-dihydroxyisovalerate hydrate or 10 mM (±)-sodium-α,β-dihydroxymethylvalerate hydrate and 0.5 μM of purified DHAD enzyme. The reaction was initiated

by adding the enzyme. After 20 minute incubation at 30 °C, the reactions were stopped by adding an equal volume of acetonitrile. Approximately 0.04 volumes of 100 mM phenylhydrazine (PhH) was added to derivatize the product 2-keto isovalerate or 2-keto-3-methyl valerate at room temperature for 30 min. 10 µL of the reaction mixture was used for the HPLC analysis. The area of the HPLC peak with UV absorption at 341 nm was used to quantify the amount of product formed. The inhibition percentage of aspterric acid on pDHADs was determined using *in vitro* biochemical assays and calculated with following equation:

inhibition percentage = 100− 100 x (initial reaction rate with aspterric acid/initial reaction rate without aspterric acid)

7.18 *in vitro* characterization of HarC and HarD

50 µL-scale *in vitro* assays of HarC and HarD were performed at 30°C for 2 h (unless indicated otherwise), in 100 mM Phosphate buffer (pH 7.5), generally containing 5 µM HarC, 10 µM HarB, 50 µM flavin adenine dinucleotide (FAD, Sigma Aldrich), 1 mM nicotinamide dinucleotide phosphate hydrogen (NADPH, Cayman Chemicals), and 100 µM *S*-adenosylmethionine (SAM, Sigma Aldrich), and ~100 µM compound 6 or 7. The reaction was quenched with equal volume of MeCN, and centrifuged at 17,000 *g* for 5 min before LC-MS analysis.

7.18.1 *Anaerobic assays*

100 mM Phosphate buffer (pH 7.5), 10 mM FAD, 60 mM NADPH, 100 mM SAM and ~10 mM compound 6 and 7 were left in anaerobic chamber supplied with mix gas (N_2 : H_2 : CO_2 = 85 : 5 : 10) overnight for the dissolved oxygen to escape the solutions. HarB and HarC protein stocks were placed in anaerobic chamber for 3 hours on ice before setting the reactions up. The anaerobic *in vitro* reactions contained same concentrations of cofactors, proteins and substrate as specified above and left in the anaerobic chamber at room temperature. Similarly prepared reaction samples were taken out of the anaerobic chamber for control reaction in the presence of oxygen. After 2 hours of reaction at room temperature, all reactions were quenched with equal volume of MeCN in their respective environments, centrifuged at 17,000 *g* for 5 min before LC-MS analysis.

7.18.2 *in vitro* assays with $H_2{}^{18}O$

^{18}O-labeled water (97 atom-% $H_2{}^{18}O$, Sigma-Aldrich) was used in place of $H_2{}^{16}O$ water. 50 µL reactions with 100 mM Phosphate buffer (pH 7.5) prepared with either $H_2{}^{16}O$ or $H_2{}^{18}O$ were set up to contain cofactors and enzymes as specified above and incubated at 30°C for 3 hours. The enzymatic reactions were quenched with equal amount of MeCN, centrifuged at 17,000 *g* for 5 min before LC-MS analysis. The supernatants of the reactions prepared with $H_2{}^{18}O$ was then lyophilized overnight to evaporate the water. The precipitate was dissolved in $H_2{}^{16}O$ overnight for non-enzymatic water-exchange and subjected to LC/MS and QTOF analysis. Compound 1 standard was placed in $H_2{}^{18}O$ for 3 hours for non-enzymatic water exchange.

7.19 Small-scale *in vitro* enzymatic reactions for nepetalactol production

7.19.1 TfG8H *in vitro* assays

200 µL-scale *in vitro* assays were performed at 30°C for 2 hours in 100 mM Bis-Tris-Propane buffer (pH 9.0, unless otherwise indicated) generally containing 2 mM geraniol, 5 µM TfG8H, 10 µM FpR, 10 µM YkuN and 4 mM NADPH, unless otherwise indicated. Reactions with NADPH regeneration contain an additional 6 mM fumarate, 1 µM FumC and 10 µM MaeB. The reaction was stopped by extracting directly with 100 µL ethyl acetate (EtoAc) and centrifuged at 17,000 *g* for 5 min before the organic layer was subjected to GC/MS analysis on an Agilent Technologies GC-MS 6890/5973 equipped with a DB-FFAP column. An inlet temperature of 220 °C and constant pressure of 4.2 psi were used. The oven temperature was held at 60 °C for 5 min and then ramped at 60°C/min for 1.5 min, followed by a ramp of 15 °C/min for 16 min and a hold for 10 min.

7.19.2 GOR *in vitro* assays

200 µL-scale *in vitro* assays were performed at 30°C for 2 hours in 100 mM Bis-Tris-Propane buffer (pH 9.0) containing 2 mM 8-hydroxygeraniol, 10 µM GOR and 4 mM nicotinamide dinucleotide (NAD^+, Cayman Chemicals) unless otherwise specified. Reactions with NAD^+ regeneration contain 5 µM NoxE. The reactions were stopped by extracting directly with 100 µL ethyl acetate (EtoAc) and centrifuged at 17,000 *g* for 5 min before the organic layer was subjected to GC/MS analysis.

7.19.3 ISY and NmMLPL *in vitro* assays

200 µL-scale *in vitro* assays were performed at 30°C for 2 hours in 100 mM Bis-Tris-Propane buffer (pH 9.0) containing 2 mM 8-oxogeranial, 1 µM ISY, 10 µM NmMLPL and 4 mM NADPH unless otherwise specified. Reactions with NADPH regeneration contain an additional 6 mM fumarate, 1 µM FumC and 10 µM MaeB. The reactions were stopped by extracting directly with 100 µL ethyl acetate (EtoAc) and centrifuged at 17,000 *g* for 5 min before the organic layer was subjected to GC/MS analysis as described previously.

7.19.4 NoxE and FumC/MaeB cofactor selectivity assay

For NoxE cofactor selectivity, 200 µL-scale *in vitro* assays were performed in a 96-well plate (Corning® 96 Well plate, Sigma-Aldrich) at 30°C in 100 mM Phosphate buffer (pH 7.5) containing 1 µM NoxE, 400 µM NAD^+ or $NADP^+$. The absorbance at 340 nm was measured continuously for 2 hours.

For MaeB cofactor selectivity, 200 µL-scale *in vitro* assays were performed in a 96-well plate (Corning® 96 Well plate, Sigma-Aldrich) at 30°C in 100 mM Phosphate buffer (pH 7.5) containing 4 mM fumarate, 1 µM FumC, 10 µM MaeB, 4 mM NADH or NADPH. The absorbance at 340 nm was measured continuously for 2 hours.

7.19.5 Small-scale one-step or two-step one-pot reactions

200 µL-scale one-pot one-step *in vitro* assays were performed at 30°C for 2 hours in 100 mM Bis-Tris-Propane buffer (pH 9.0, unless otherwise indicated) containing 2 mM geraniol, 12 mM fumarate, 5 µM TfG8H, 10 µM FpR, 10 µM YkuN, 10 µM GOR, 0.5 µM ISY, 5 µM NmMLPL, 1 µM FumC, 10 µM MaeB, 5 µM NoxE, 100 µM NADPH and 100 µM NAD^+. The one-pot two-step *in vitro* assays were performed by directly adding 0.5 µM ISY, 5 µM NmMLPL and 6 mM fumarate after the conclusion of 2 hours reaction with 2 mM geraniol, 6 mM fumarate, 5 µM TfG8H-full, 10 µM FpR, 10 µM YkuN, 10 µM GOR, 1 µM FumC, 10 µM MaeB, 5 µM NoxE, 100 µM NADPH and 100 µM NAD^+. The reactions were then stopped by extracting directly with 100 µL EtoAc and centrifuged at 17,000 *g* for 5 min before the organic layer was subjected to GC/MS analysis as described above.

7.20 10 mL-scale nepetalactol and nepetalactone production and purification

10 mL-scale one-pot two-step *in vitro* assays were performed at 28°C, 250 rpm in a 50 mL falcon tube by directly adding 0.5 µM ISY, 5 µM NmMLPL and 6 mM fumarate after the conclusion of 2 hours reaction with 2 mM geraniol, 6 mM fumarate, 5 µM TfG8H, 10 µM FpR, 10 µM YkuN, 10 µM GOR, 1 µM FumC, 10 µM MaeB, 100 µM NADPH and 100 µM NAD^+. The 10 mL-scale one-pot drop-in *in vitro* assays were performed by first directly adding 10 µM GOR, 5 µM NoxE and 100 µM NAD^+ after the conclusion of 2 hours reaction with 2 mM geraniol, 6 mM fumarate, 5 µM TfG8H, 10 µM FpR, 10 µM YkuN, 1 µM FumC, 10 µM MaeB and 100 µM NADPH. Then 0.5 µM ISY, 5 µM NmMLPL and 6 mM fumarate was added after the conclusion of 2 hours reaction with GOR. Aliquots of 2 mM geraniol and 6 mM fumarate was added to 5 µM TfG8H, 10 µM FpR, 10 µM YkuN, 1 µM FumC, 10 µM MaeB and 100 µM NADPH and the reactions were run for 90 minutes each before GC/MS analysis in order to assess the *in-vitro* system's efficiency and robustness. For nepetalactone production 10 µM NEPS1 was added to the system at the same time as ISY and NmMLPL is added. All reactions were stopped by extracting directly with 5 mL EtoAc and centrifuged at 17,000 *g* for 5 min before the organic layer was subjected to GC/MS analysis as described above. The organic layer was dried, leaving nepetalactol oil.

Supplementary tables

Table S1. k_{cat} and K_m of pDHAD and its mutants

Enzymes		DHI		DHMV	
Position	**Mutation**	k_{cat} (s^{-1})	K_M (mM)	k_{cat} (s^{-1})	K_M (mM)
pDHAD		6.85 ± 0.508	5.78 ± 1.06	6.50 ± 0.323	4.00 ± 0.446
I177	L	1.30 ± 0.0302	4.82 ± 0.309	0.503 ± 8.90e-3	1.80 ± 0.0968
	F	0.114 ± 5.93e-3	4.95 ± 0.727	0.0171 ± 1.20e-3	0.638 ± 0.156
	W	N/A			
V178	L	1.24 ± 0.0681	8.14 ± 0.990	0.616 ± 6.92e-3	3.99 ± 0.124
	I	23.2 ± 1.14	14.3 ± 1.30	8.38 ± 0.585	7.78 ± 1.22
	F	0.0105 ± 0.296e-3	1.09 ± 0.119	0.0204 ± 0.660e-3	1.06 ± 0.134
	W	9.46e-3 ± 0.923e-3	1.08 ± 0.411	6.19e-3 ± 0.291e-3	3.30 ± 0.452
V496	L	12.7 ± 0.792	6.68 ± 0.986	1.94 ± 0.0625	1.98 ± 0.212
	I	2.00 ± 0.190	6.53 ± 1.48	0.593 ± 0.0481	4.01 ± 0.896
	F	0.141 ± 5.76e-3	0.226 ± 0.0608	0.0331 ± 2.04e-3	0.133 ± 0.0719
	W	N/A			
V497	L	8.52 ± 0.102	6.09 ± 0.178	5.13 ± 0.300	4.80 ± 0.146
	I	7.62 ± 0.486	8.64 ± 1.198	6.76 ± 0.249	12.4 ± 0.885
	F	0.0412 ± 1.11e-3	0.811 ± 0.0912	0.0218 ± 0.994e-3	0.615 ± 0.127
	W				
I177 V496	L	7.11 ± 0.362	3.67 ± 0.528	4.10 ± 0.161	5.39 ± 0.535
I177 V497	L	0.766 ± 0.0120	7.14 ± 0.259	0.271 ± 2.90e-3	2.25 ± 0.0779
V178 V496	L	3.53 ± 0.0792	7.68 ± 0.389	1.61 ± 8.52e-3	2.50 ± 0.0417

Table S2. Fresh-weight per plant grown on agar plate containing different concentrations of AA

Transgenic line	Fresh-weight per plant (mg)					
	5 μM AA	1 μM AA	0.5 μM AA	0.25 μM AA	0.1 μM AA	0 μM AA
B2-pDHAD-46	0.786	1.43	1.82	2	2.79	2.43
B2-pDHAD-50	1	1.36	1.73	1.73	2.36	2.18
B2-pDHAD-55	0.333	0.833	1.17	2	2.2	2.33
B1-I177F-60	1	1.7	1.7	1	2.6	2.5
B1-I177F-89	1.09	1.45	2.09	2	2.55	2.64
B1-I177F-67	1.21	2.17	1.75	2.14	2.93	2.86
B1-I177F-65	1	2	1.67	2.2	2.5	3
B1-I177F-30	1.5	1.5	1.43	2	2.33	2.38
B1-I177F-61	1	1.38	1.67	2.09	1.67	2.08
B1-I177F-26	1.67	1	1.5	2.67	2	3
B1-I177F-55	1.31	1.38	1.883	1.85	2.36	2.17
B2-I177F-19	1.1	1.5	1.83	1.56	1.6	1.57
B2-I177F-26	0.889	1.09	1.33	1.33	1.63	1.36
B1-I177F-69	1.5	1.67	2.05	2.42	3	2.67
B2-V178F-65	1.06	1.73	1.73	1.69	2.11	1.75
B2-V178F-71	1	1.13	1.57	1.33	2.3	3.33
B2-V178W-1	1.05	1.43	1.61	2.13	2.06	2.08
B2-V178W-2	1.14	1.36	1.5	1.67	2	1.93
B2-V496F-61	1.15	1.92	2.5	1.7	2.11	2
B2-V496F-60	0.991	1.7	1.78	1.82	1.96	1.88
B2-V496W-91	1.23	1.31	1.58	2	2.28	1.98
B2-V496W-97	0.921	1.28	1.6	1.66	1.95	1.99
B2-V497F-88	1.04	1.3	1.59	1.93	1.61	2.03
B2-V497F-81	0.973	1.15	1.61	1.73	1.78	1.93
B1-V497F-97	1.44	1.66	2.79	3.16	4.06	3.43
B2-V497W-14	1.11	1.43	1.43	1.84	1.84	1.84
B2-V497W-16	0.967	1.3	1.3	1.47	1.9	1.52
AstD-2423	2.45	2.55	2.25	2.35	2.2	1.8
AstD-2424	2.82	2.82	2.55	3	3.09	2.7

Table S3. *har* and *apn* gene cluster bioinformatics analysis[1]

Gene name	Size (aa)	Proposed protein function	Homolog (identity %)	Organism	*apn* cluster homologs (identity %)[2]
harA	3954	Polyketide synthase – nonribosomal peptide synthetase (PKS−NRPS)	ApdA (50)	*Aspergillus nidulans*	*apnA* (50)
harB	401	*O*-Methyltransferase (OMT)	AsqD (38.1)	*Aspergillus nidulans*	*apnB* (60.2)
harC	432	Flavin-dependent monooxygenase (FMO)	DbaB (37.5)	*Aspergillus nidulans*	*apnC* (67.9)
harD	499	*N*-hydroxylation Cytochrome P450 (P450)	ApdB (48.6)	*Aspergillus nidulans*	*apnD* (50.9)
harE	353	*trans*-Enoyl reductase (ER)	ApdC (51.7)	*Aspergillus nidulans*	*apnE* (54)
harF	573	Flavin-dependent monooxygenase (FMO)	ApdD (56.3)	*Aspergillus nidulans*	*apnF* (56.1)
harG	504	Cytochrome P450 (P450)	ApdE (71.7)	*Aspergillus nidulans*	*apnG* (74.2)
harH	634	Transcription factor (TF)	LepE (31.6)	*Aspergillus flavus*	*apnH* (38.7)
harI	332	Transcription factor (TF)	SorR2 (21)	*Penicillium chrysogenum*	N/A[3]
harJ	409	Major facilitator superfamily transporter (MFS)	MCH2 (30.2)	*Saccharomyces cerevisiae*	N/A

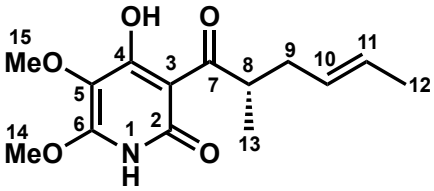
Additional genes in *apn* BGC					Accession numbers
apnI	203	Succinate dehydrogenase cytochrome B subunit (SDHC)	SDH3 (34.2)	*Schizosaccharomyces pombe*	PDE_09192
apnII	390	Probable transport protein	YPL264C (25.9)	*Saccharomyces cerevisiae*	PDE_09193
apnIII	320	Protein with Unknown function	CctP (33.2)	*Talaromyces islandicus*	PDE_09195

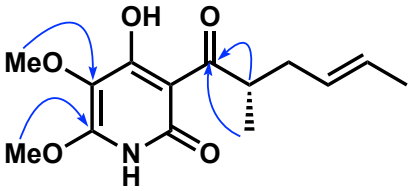
[1]: Clusters deposited in MIBiG. *har* BGC: BGC0002066; *apn* BGC: BGC0002067.

[2]: % identity with *har* homolog.

[3]: homolog not found in *apn* BGC.

Table S4. Spectroscopic data of compound 1





Key HMBC correlations

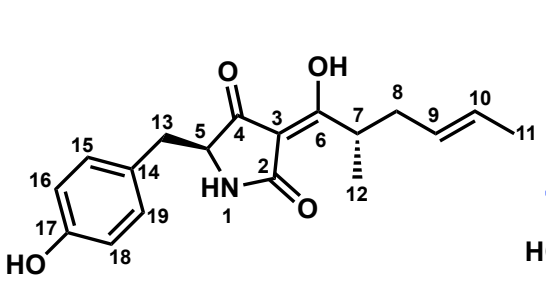
position	δ_H, mult (J in Hz)	δc	COSY
1			
2		161.8	
3		100.1	
4		165.6	
5		159.6	
6		124.4	
7		210.1	
8	3.97, overlap	44.3	H9, H13
9	2.46, m 2.01, m	37.0	H8, H10
10	5.43, overlap	129.6	H9
11	5.47, overlap	127.3	H12
12	1.60, d (5.0)	17.9	H9, H11
13	1.12, d (6.5)	16.7	H8
14-Me	3.96, s	54.9	
15-Me	3.70, s	60.7	

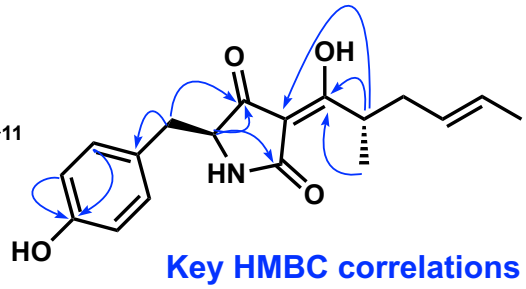
^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 1 in Acetone-d_6.

HRMS (ESI, MH$^+$) calcd for $C_{14}H_{20}NO_5$ 282.1336, found 282.1339.

$[\alpha]_D^{21}$ -5.0 (c = 0.20, MeOH), reported $[\alpha]_D^{27}$ -8.4 (c = 0.10, MeOH[166]) .

Table S5. Spectroscopic data of compound 2



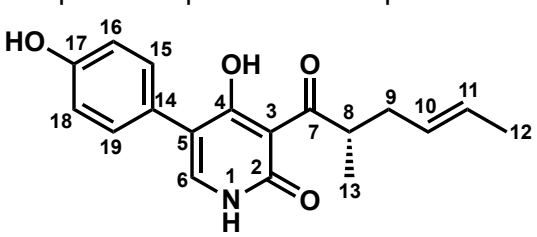


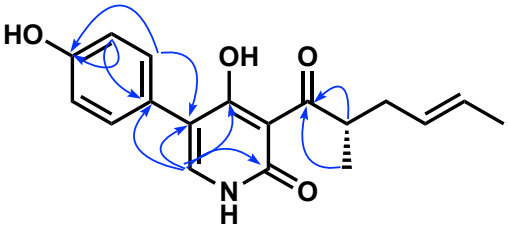
position	δ_H, mult (J in Hz)	δc	COSY
1			
2		175.6	
3		100.3	
4		194.8	
5	3.95, dd (9.0, 4.0 Hz)	63.7	H13
6		193.4	
7	3.59, m	36.8	H8, H12
8	2.31, m 2.11, m	36.4	H9, H13
9	5.31, m	127.3	H8, H11
10	5.43, m	127.5	H11
11	1.59, d (6.0)	17.9	H10
12	1.11, d (7.0)	16.4	H7
13	3.09, dd (9.0, 4.0 Hz) 2.66, dd (14.0, 8.5 Hz)	37.2	H5
14		127.9	
15/19	6.96, d (8.5)	130.5	H16/H18
17		155.4	
16/18	6.68, d (8.0)	115.7	H15/H19

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 2 in $CDCl_3$.

HRMS (ESI, MH^+) calcd for $C_{18}H_{22}NO_4$ 316.1543, found 316.1544.

Table S6. Spectroscopic data of compound 3





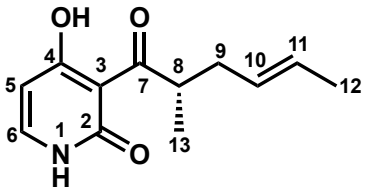
Key HMBC correlations

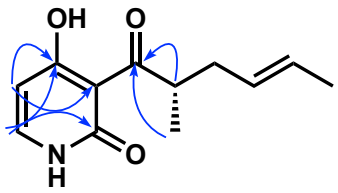
position	δ_H, mult (J in Hz)	δc	COSY
1			
2		161.4	
3		105.5	
4		175.9	
5		112.6	
6	7.53, s	140.7	
7		211.2	
8	4.17, m	42.2	H9, H13
9	2.35, m 1.97, m	35.7	H8, H10
10	5.36, overlapped	128.9	H9
11	5.38, overlapped	126.4	H12
12	1.57, d (5.0)	17.8	H9, H11
13	1.03, d (6.5)	16.3	H8
14		123.5	
15/19	7.22, d (8.5)	130.2	H16/H18
17		156.8	
16/18	6.75, d (8.5)	115.0	H15/H19

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 3 in DMSO-d_6.

HRMS (ESI, MH$^+$) calcd for $C_{18}H_{20}NO_4$ 314.1387, found 314.1391

Table S7. Spectroscopic data of compound 4





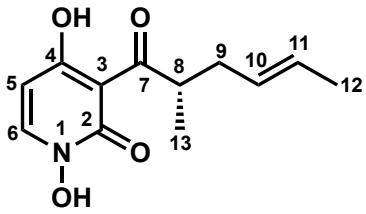
Key HMBC correlations

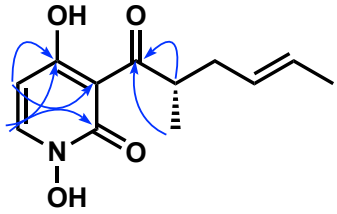
position	δ_H, mult (J in Hz)	δc	COSY
1	12.10		H6
2		164.7	
3		106.9	
4		178.5	
5	6.03, s	101.7	H6
6	7.35, s	140.2	H1 (NH), H5
7		211.8	
8	4.13, m	43.5	H9, H13
9	2.46, m 2.03, m	36.2	H8, H10
10	5.39, overlap	128.8	H9
11	5.45, overlap	127.1	H12
12	1.63, d (6.0)	18.1	H9, H11
13	1.14, d (7.0)	16.1	H8

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 4 in $CDCl_3$.

HRMS (ESI, MH^+) calcd for $C_{12}H_{16}NO_3$ 222.1125, found 222.1129.

Table S8. Spectroscopic data of compound 5





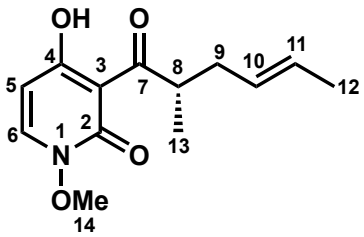
Key HMBC correlations

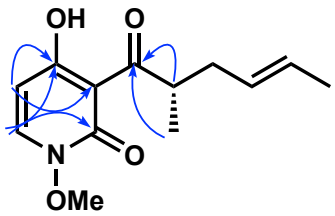
position	δ_H, mult (J in Hz)	δc	COSY
1			
2		157.3	
3		105.6	
4		174	
5	6.03, s	98.6	H6
6	7.79, s	135	H5
7		211.2	
8	4.10, m	43.9	H9, H13
9	2.46, m 2.05, m	36.1	H8, H10
10	5.44, overlap	128.5	H9
11	5.49, overlap	127.5	H12
12	1.64, d (6.0)	18.1	H9, H11
13	1.16, d (7.0)	16.0	H8

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 5 in $CDCl_3$.

HRMS (ESI, MH^+) calcd for $C_{12}H_{16}NO_4$ 238.1074, found 238.1079.

Table S9. Spectroscopic data of compound 6





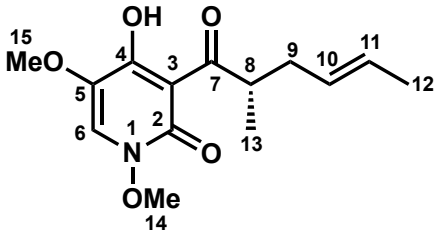
Key HMBC correlations

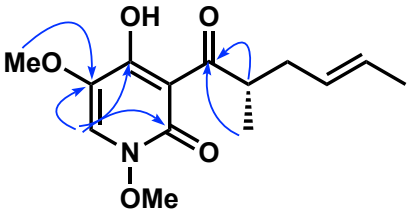
position	δ_H, mult (J in Hz)	δc	COSY
1			
2		158.0	
3		107.7	
4		176.5	
5	5.91, d (8.0)	99.7	H6
6	7.61, d (8.0)	140.6	H5
7		211.9	
8	4.12, m	43.7	H9, H13
9	2.44, m 2.03, m	36.2	H8, H10
10	5.42, overlap	128.7	H9
11	5.46, overlap	127.3	H12
12	1.62, d (5.0)	18.1	H9, H11
13	1.14, d (7.0)	16.1	H8
14	4.04, s	65.2	

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 6 in $CDCl_3$.

HRMS (ESI, MH^+) calcd for $C_{13}H_{18}NO_4$ 252.1230, found 252.1238.

Table S10. Spectroscopic data of compound 7





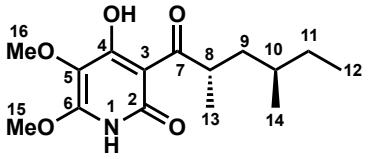
Key HMBC correlations

position	δ_H, mult (J in Hz)	δc	COSY
1			
2		156.3	
3		107.0	
4		170.7	
5		134.0	
6	7.29, s	124.5	
7		212.6	
8	4.20, m	43.6	H9, H13
9	2.46, m 2.05, m	36.2	H8, H10
10	5.41, overlap	128.6	H9
11	5.46, overlap	127.4	H12
12	1.63, d (5.0)	18.1	H9, H11
13	1.15, d (6.0)	16.1	H8
14-Me	4.05, s	64.8	
15-Me	3.80, s	58.4	

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of compound 7 in $CDCl_3$.

HRMS (ESI, MH^+) calcd for $C_{14}H_{20}NO_5$ 282.1336, found 282.1340.

Table S11. Spectroscopic data of atpenin B



reported			this study		
position	δ_H, mult (J in Hz)	δ_c	position	δ_H, mult (J in Hz)	δ_c
1			1		
2		162.6	2		162.5
3		100.6	3		100.6
4		165.8	4		165.7
5		124.9	5		124.8
6		159.9	6		159.8
7		211.7	7		211.6
8	4.30, tq (6.8, 6.6)	41.9	8	4.30, m	41.9
9	1.71, ddd (12.5, 5.9, 5.3) 1.33, m	41.1	9	1.71, m 1.33, m	41.0
10	1.38, m	30.6	10	1.38, m	30.5
11	1.21, m 1.06, m	21.0	11	1.22, m 1.06, m	20.0
12	0.71, t (7.3)	11.7	12	0.71, t (7.5)	11.7
13	1.19, d (6.6)	17.1	13	1.19, d (6.5)	17.0
14	0.86, d (6.4)	19.1	14	0.86, d (6.5)	19.0
15	3.78, s	54.3	15	3.78, s	54.2
16	3.71. s	60.6	16	3.72, s	60.5

^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of atpenin B in Pyridine-d_5[167].

HRMS (ESI, MH^+) calcd for $C_{15}H_{24}NO_5$ 298.1654, found 298.1657.

$[\alpha]_D^{22}$ -10.0 (c = 0.10, EtOH), reported $[\alpha]_D^{27}$ -14.5 (c = 0.10, EtOH)[166].

Table S12. Cost analysis for 1 g nepetalactol production

component	mass per 1 g nepetalactol (g)	component cost ($/g)	cost per 1 g nepetalactol ($)	commercial source
geraniol	0.92	$0.54	$0.53	Fisher Scientific
enzyme	2.28	$2.50 (low) $25.00 (high)	$6.13 (low) $61.29 (high)	*
NADPH	0.076	$450.00	$36.77	Calzyme
NAD^+	0.068	$14.00	$1.02	Calzyme
fumarate	0.68	$0.04	$0.03	Fisher Scientific
BSA	0.123	$31.70	$4.19	Fisher Scientific
bis-tris propane	28.4	$0.60	$18.32	Fisher Scientific
Total material cost to make 1 g nepetalactol			**$67.01 (low) $122.17 (high)**	

* Enzyme cost at scale is estimated to be between $250 and $2500 per kilogram based on communication with industry. For our material cost estimate, we have taken the upper bound ($2.50/g) as our low price point and a 10-fold increase ($25.00/g) as our high price point to account for the absence of economies of scale.

Supplementary figures

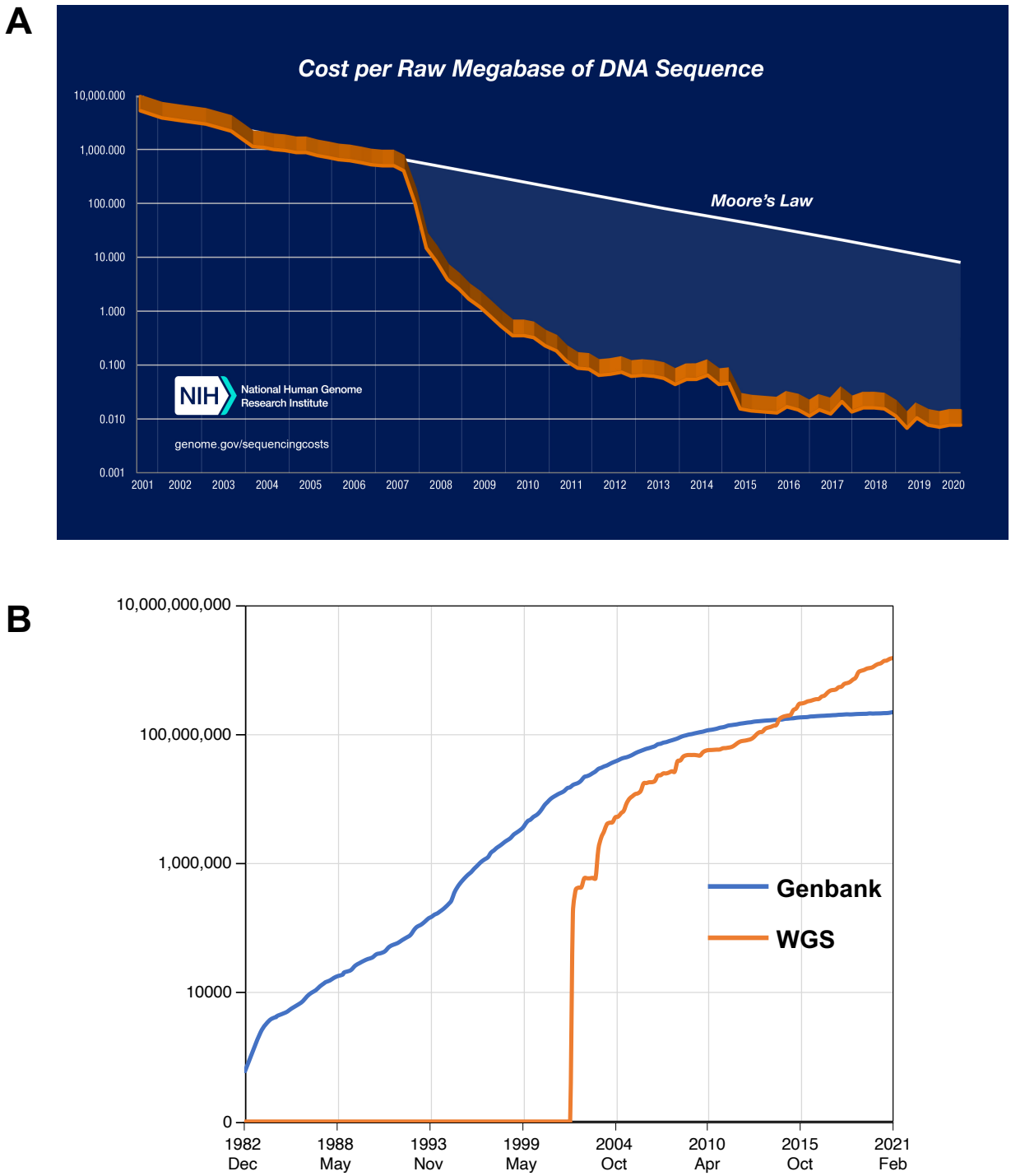


Figure S1. Availability of whole genomic DNA (A) Trend in cost of whole genomic DNA sequencing. (B) Number of whole genome sequences deposited in Genbank or WGS database. The data was adapted from publicly accessible data from NIH website

Figure S2. Sequence alignment of pDHAD and AstD. Amino acids forming α-helices that are part of the active-site entrance are depicted in cyan. Amino acids forming the loop that is the remaining part of the active-site entrance is depicted in green. Depicted by yellow are the amino acids forming the β-sheet behind the entrance-forming loop.

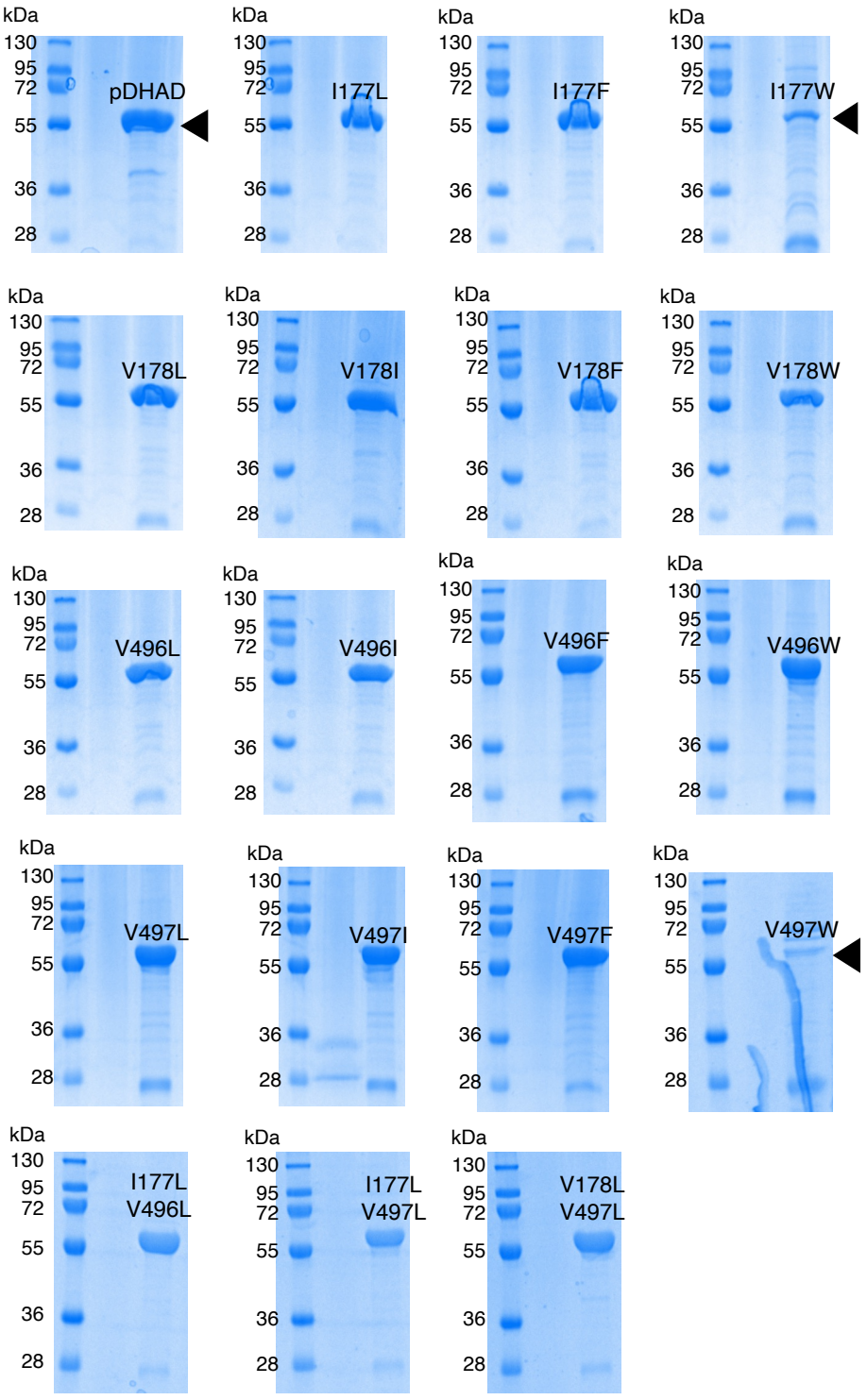


Figure S3. SDS-PAGE of purified pDHAD and its mutants

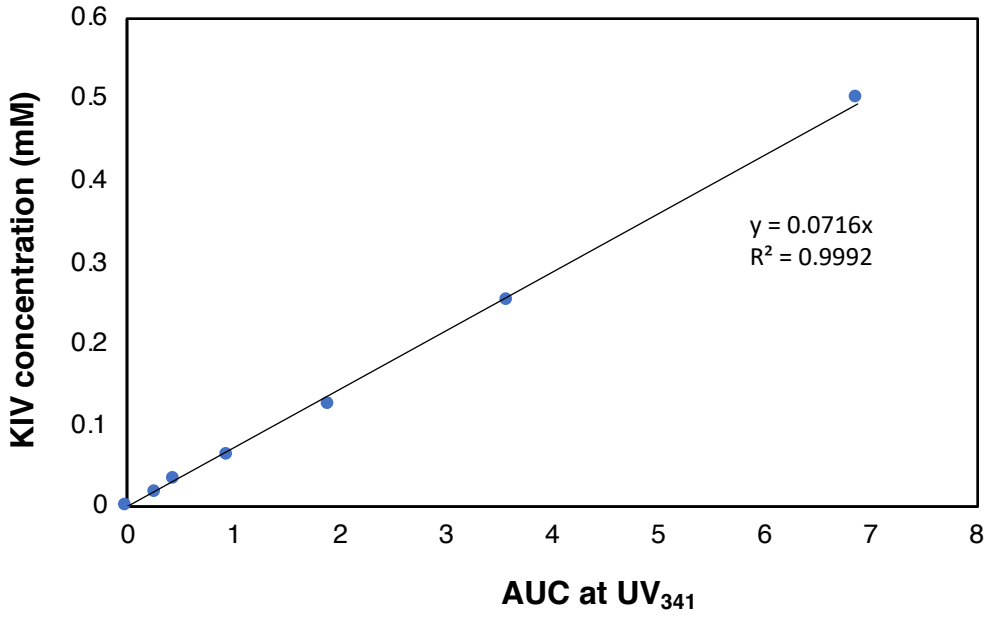


Figure S4. Calibration curve used for calculations of DHAD product concentration

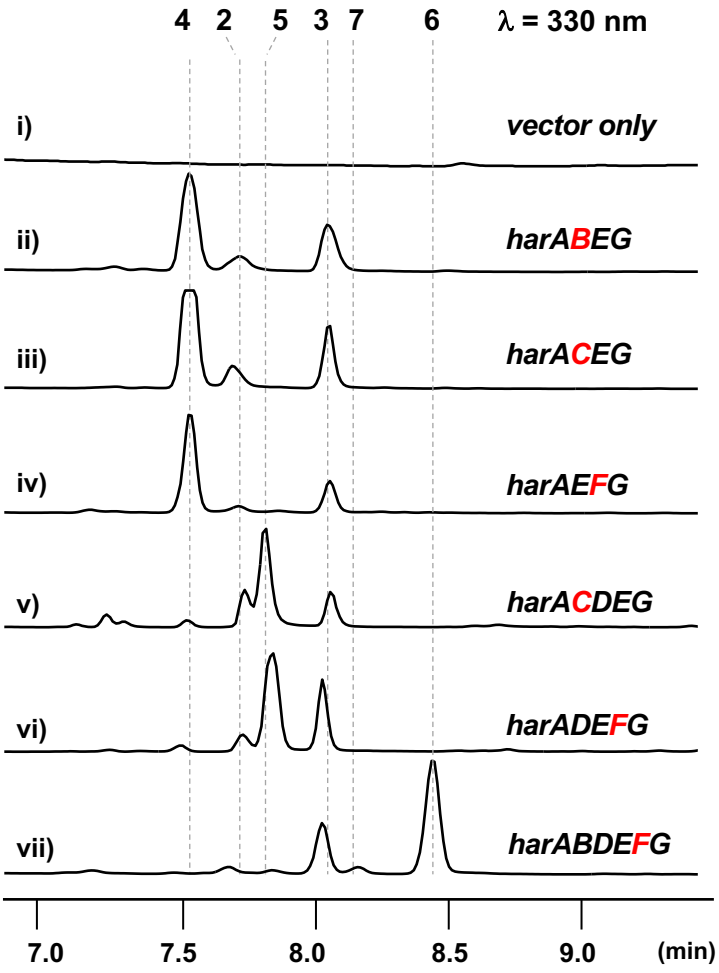


Figure S5. LC/MS analysis of extracts from *A. nidulans* expressing different combinations of *har* cluster that did not produce new metabolites compared to those expressed in Figure 15.

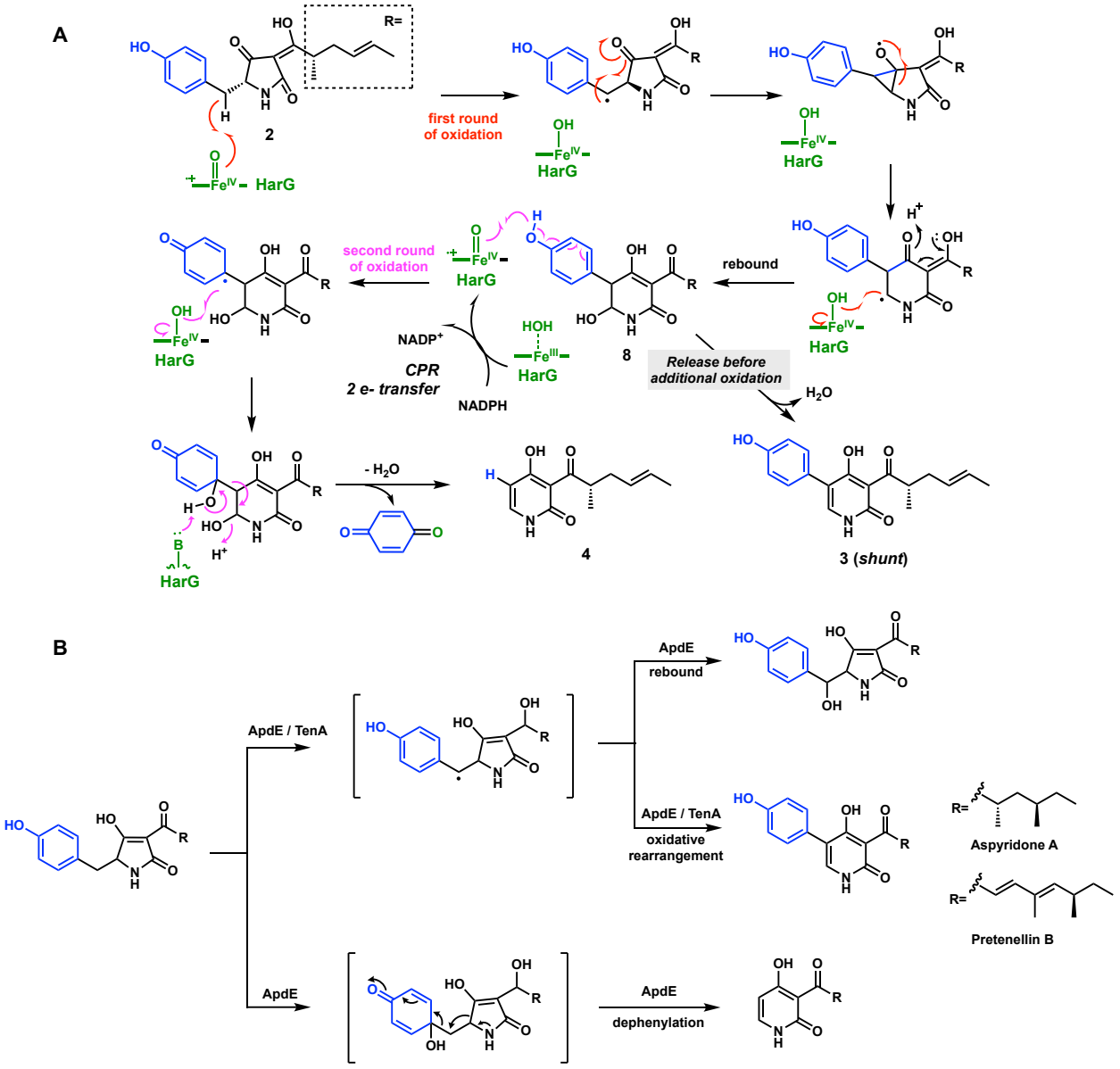


Figure S6. Mechanism of ring expansion and phenyl cleavage by HarG. (A) Proposed enzymatic mechanism of ring-expansion P450 HarG. The ring expansion mechanism is the same as proposed by Cox et al.[90] We suggest here that another round of phenyl oxidation prior to dehydration of 8 can lead to phenyl cleavage to give 4. (B) Previously proposed routes of ring expansion *or* dephenylation by Cox and coworkers for tenellin and aspyridone.[92]

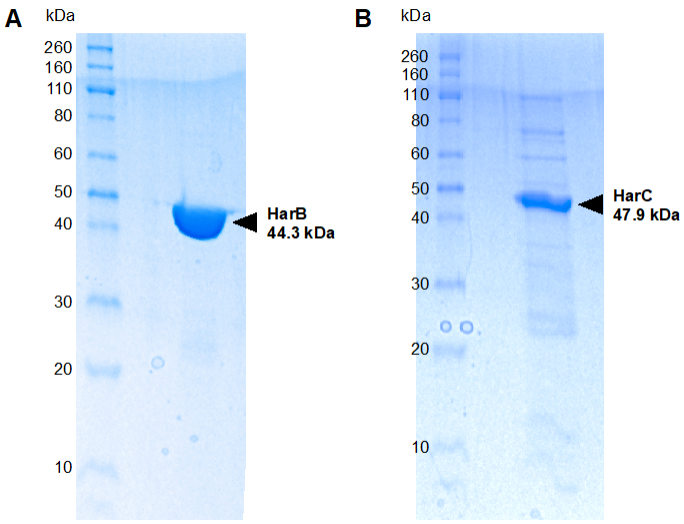


Figure S7. SDS-PAGE of purified proteins. (A) HarB. (B) HarC.

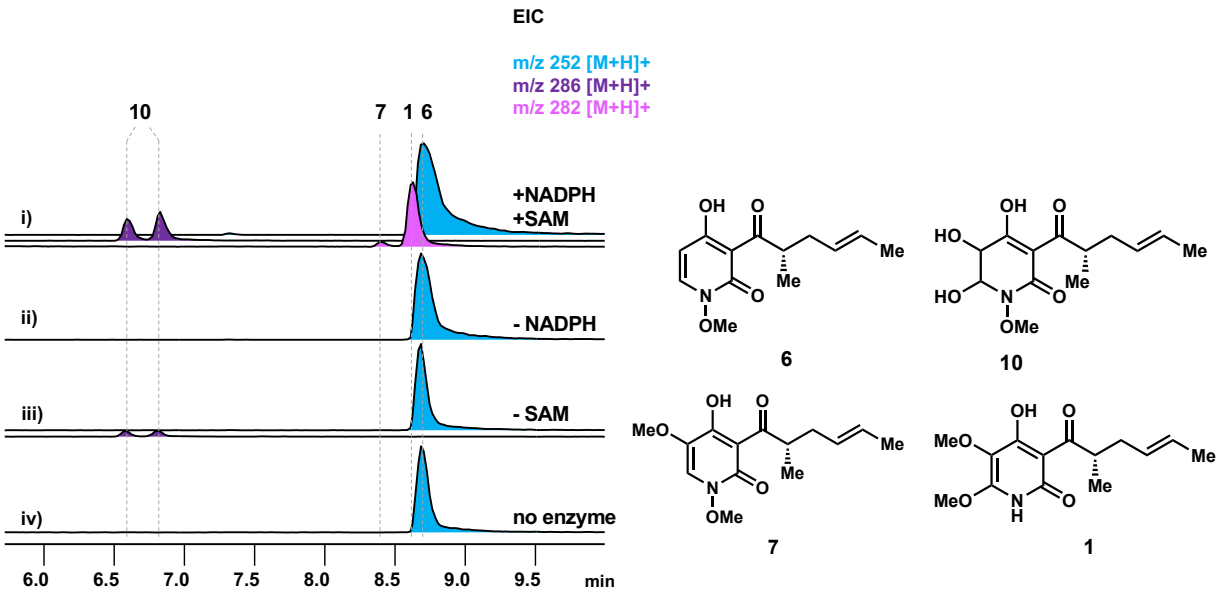


Figure S8. *in vitro* HarBC reaction cofactor requirements with compound 6 as substrate. The reaction mix contained 10 μM HarB, 5 μM HarC, 50 μM FAD, 1 mM NADPH and/or 100 μM SAM and ~100 μM compound 6. The reactions were incubated at 30°C for 2 hours, quenched with equal amount of MeCN to terminate reaction.

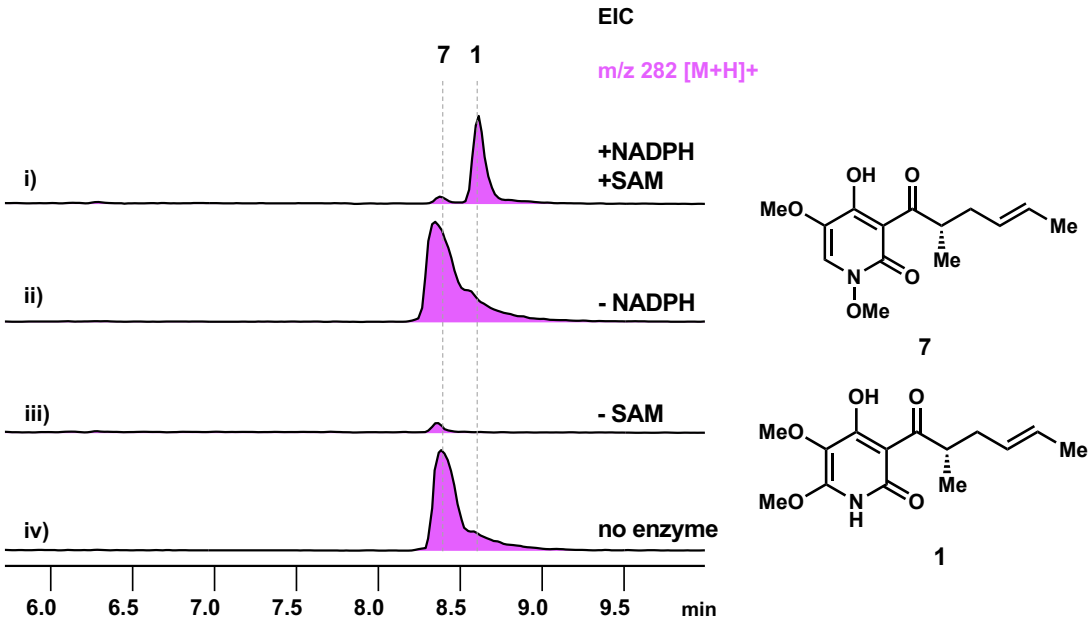


Figure S9. *in vitro* HarBC reaction cofactor requirements with compound 7 as substrate. The reaction mix contained 10 μM HarB, 5 μM HarC, 50 μM FAD, 1 mM NADPH and/or 100 μM SAM and ~100 μM compound 7. The reactions were incubated at 30°C for 2 hours, quenched with equal amount of MeCN to terminate reaction.

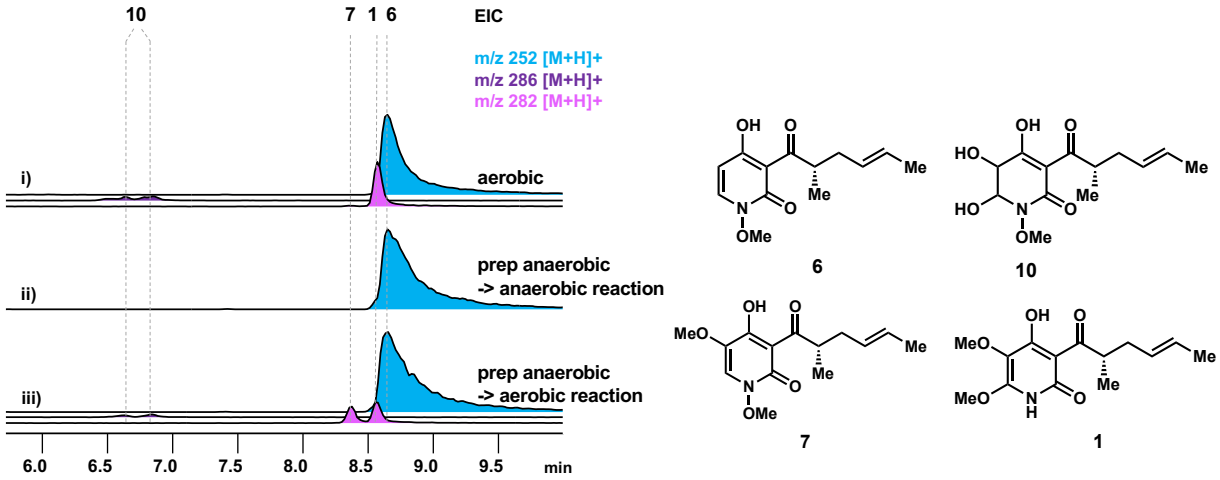


Figure S10. Anaerobic *in vitro* reactions with compound 6 as substrate. The reaction mixture contained 10 μM HarB, 5 μM HarC, 50 μM FAD, 1 mM NADPH, 100 μM SAM and ~100 μM compound 6. i) The reaction mixture was prepared aerobically and incubated under aerobic conditions. ii) The reaction mixture was prepared anaerobically and incubated under anaerobic conditions. iii) The reaction mixture was prepared anaerobically and transferred to aerobic conditions. After 2 hours of incubation at room temperature under their respective conditions, the reaction was terminated by adding equal amounts of MeCN.

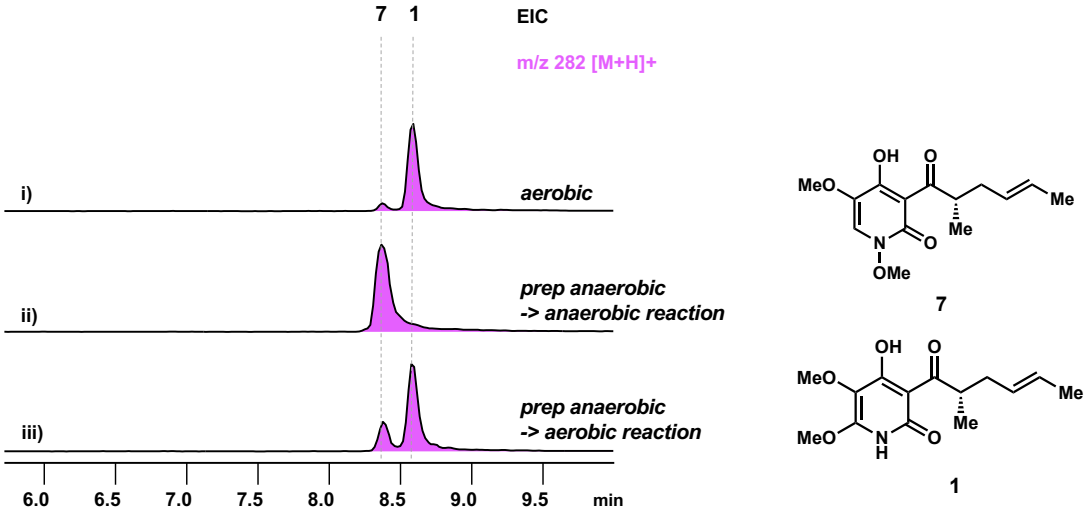


Figure S11. Anaerobic *in vitro* reactions with compound 7 as substrate. The reaction mixture contained 10 μM HarB, 5 μM HarC, 50 μM FAD, 1 mM NADPH, 100 μM SAM and ~100 μM compound 7. i) The reaction mixture was prepared aerobically and incubated under aerobic conditions. ii) The reaction mixture was prepared anaerobically and incubated under anaerobic conditions. iii) The reaction mixture was prepared anaerobically and transferred to aerobic conditions. After 2 hours of incubation at room temperature under their respective conditions, the reaction was terminated by adding equal amounts of MeCN.

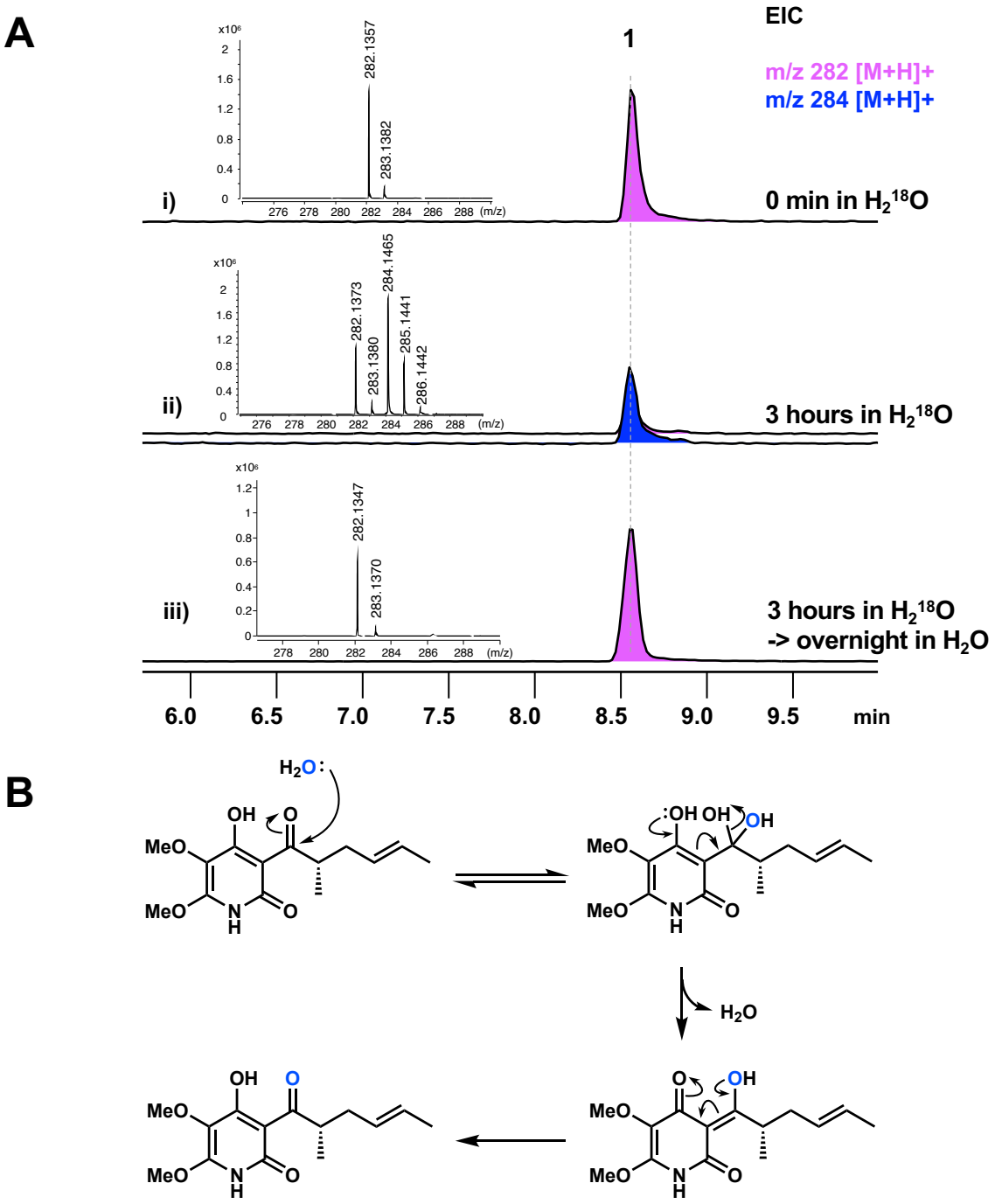


Figure S12. Non-enzymatic water exchange with compound 1. A) LC/MS and QTOF analysis of non-enzymatically water exchanged compound 1 samples i) After ~0 min of compound 1 in $H_2{}^{18}O$; ii) After 3 hours of compound 1 in $H_2{}^{18}O$; iii) After 3 hours-$H_2{}^{18}O$-exchanged 1 was lyophilized and placed in $H_2{}^{16}O$ overnight. B) Proposed mechanism of non-enzymatic water exchange in compound 1.

```
PoSDHC-housekeeping      ---------mfsqk-laqqsmrrlavqqplaiqsm--------vr---aapaaatlglqk    39
PoApnS                   msgsvqknvmlaskastmlglrravvaspslirptmglvwqqqvrlqttkpdtiqlk---    57
                                 *::.*  :  .:** .* .*  *:          **   : * :  *

PoSDHC-housekeeping      rpmtstsaeetsllakqrmnrpvsphlaiyrpqigwiasglnritgvalsgslylwatay    99
PoApnS                   ---vnkpedglrelerqrlqrpvaphlsiykwqihsvssamerntglllsgglylfatsy    114
                            ...  :    * :**::***:***:**: **  ::*.::* **: ***.***:**:*

PoSDHC-housekeeping      laspalgwhlesasmvaamgalpfaakvllkatmalpftfhsfnglrhlmwdlgrglsn-    158
PoApnS                   laapwlgwdlssatlaasfgalpvaakagikfflawpftfhlfngiryiassagytltsk    174
                         **:* ***.*.**::.*::****.***. :*  :* ***** ***:*::  . *  *:.

PoSDHC-housekeeping      nviiksgwtvvgvsvasalalafv-         182
PoApnS                   pqivkiawgvvgssalaaiglvawv          199
                           *:* .* *** *. :*:.*.
```

Figure S13. Protein sequence alignment of *P. oxalicum* house-keeping SDHC and PoApnS.

```
apnV         MSSVAQHSREQEPLLSSTSETG--------------------REPED--------YNIAWNL        34
wfapnV       MAGEQQNTIEEEPMLEELLIDEQTDANLEQKKPLYHDIVLGSGIIAQAGAWMLVAIVWGS        60
             *:.  *:: *:**:*..                                        *.*.

apnV         LSD----------TASVAQIGIWVVQAALILQPTVTPNQKQRGTRIHFTLQLLSILCFVAA        85
wfapnV       VFSHDLILFSAHPLLNSTAMLFLVQGILILQPTHTANQKKRGTYAHAATNSVAFITAIAG        120
             : .              :   :::**. ****** * ***:***  * : : ::::  :*.

apnV         FIVIELNKGTHP-HFTSPHGVLGLITIILVVLQALVGVVQYFFPITVLGSVDAGKRIYKY        144
wfapnV       LIIVEINKGNHKFQLQSPHAVLGLATYIMLALQIPLGITQFFVP-TLYGSADNAKAMYKY        179
             :*::*:***.*  :: ***.**** * *::.**  :*:.*:*.* *: **.* .* :***

apnV         HRWTGYVLLLLEVATVLAATQTGFNLNAIHIPTWGVVVAVALVVVGVGARVKARKLGLGD        204
wfapnV       HRVFGYLTLLVMLTTICAATRTGYNVNVIGMPLWIVTIASLLVVAGIAPRIKLHKFGWMA        239
             **  **: **: ::*: ***:**:*:*.* :* * *.:*  ***.*:. *:* :*:*

apnV         A----------      205
wfapnV       GRHRKAPWSYGV     251
             .
```

Figure S14. Protein sequence alignment of *P. oxalicum* ApnV and *Chaetasbolisia erysiophoides* FNF40 WfapnV. Percent identity is 39.7%.

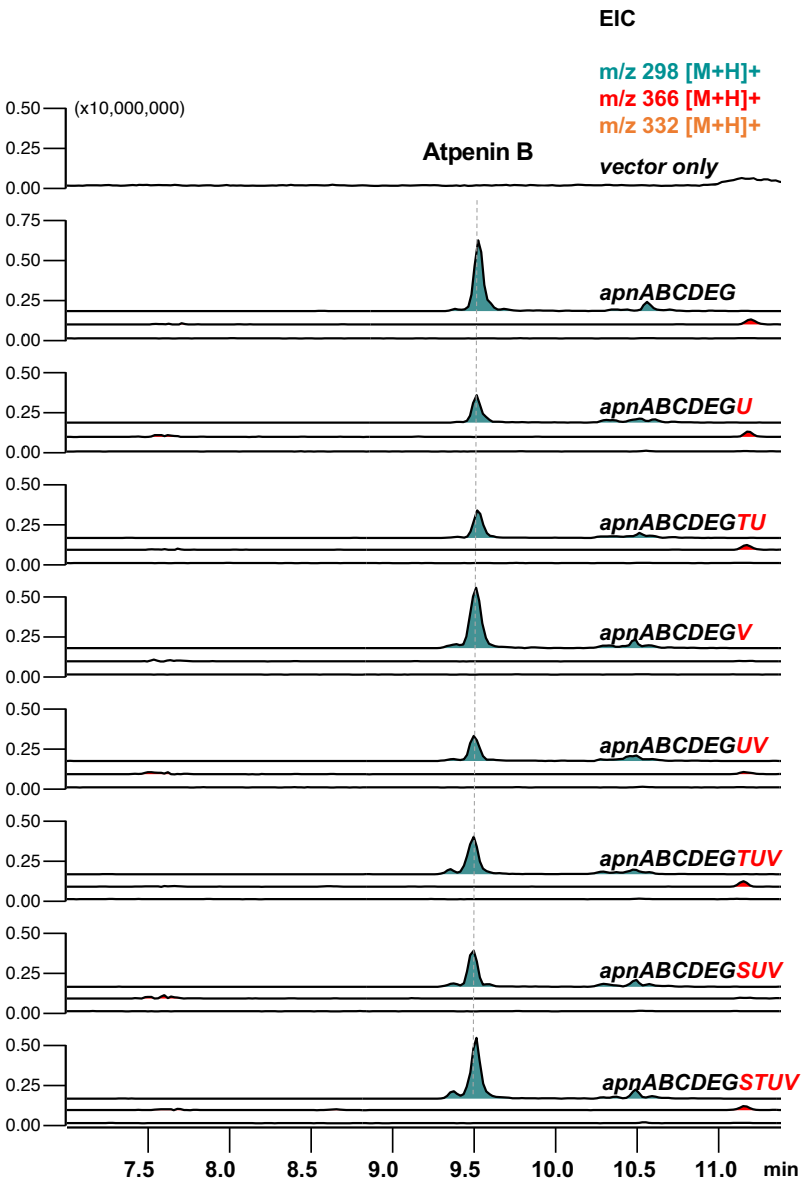


Figure S15. LC/MS analysis of *A. nidulans* expressing different combinations of *apn* BGC genes.

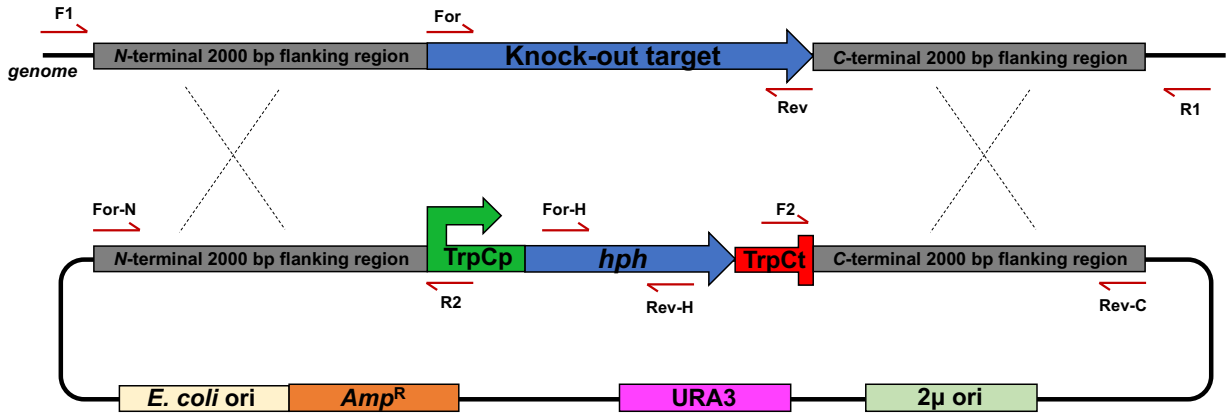


Figure S16. General plasmid structure for *P. oxalicum* gene knock-outs. Primers and their annealing locations used for gene knock-outs are indicated by red arrows.

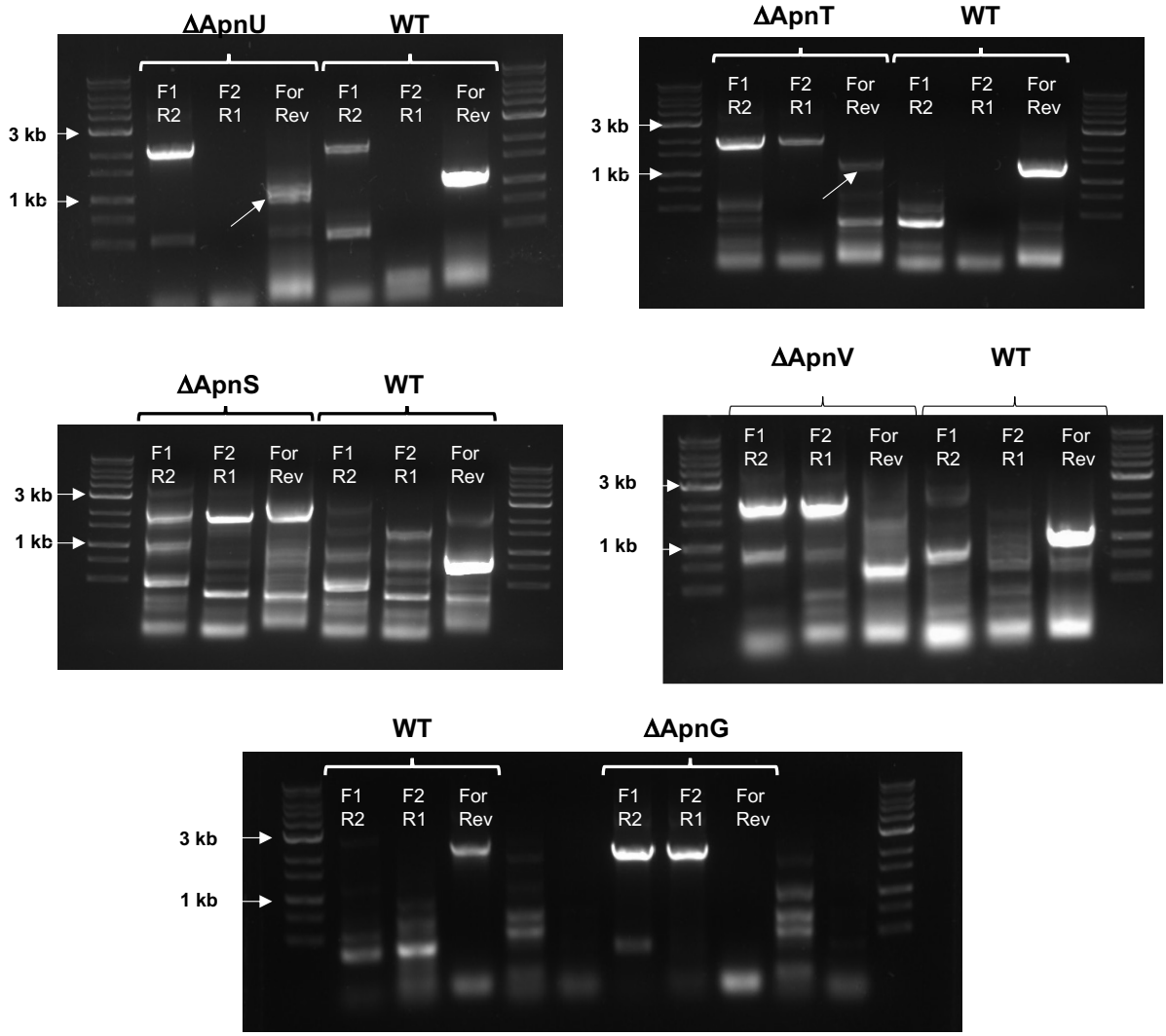


Figure S17. Agarose gel electrophoresis analysis of *P. oxalicum* knock-out strain genomes. The bands indicated with white arrows are from non-specific PCR products elsewhere on the genome

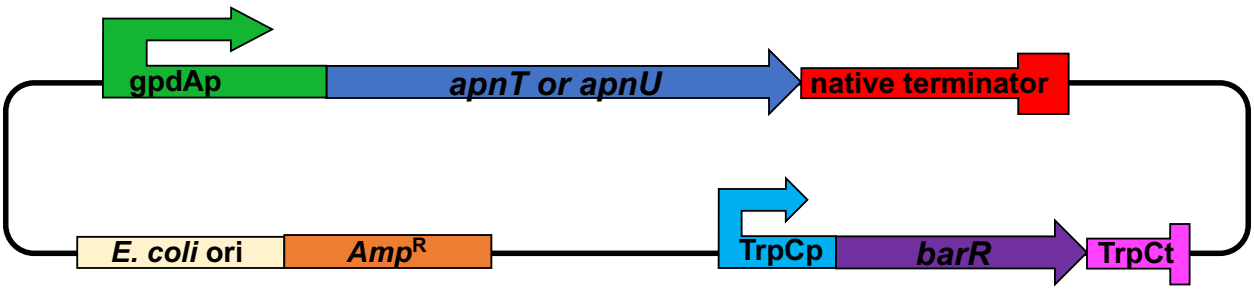


Figure S18. General plasmid structure for *P. oxalicum* transient expression of *apnU* and *apnT.*

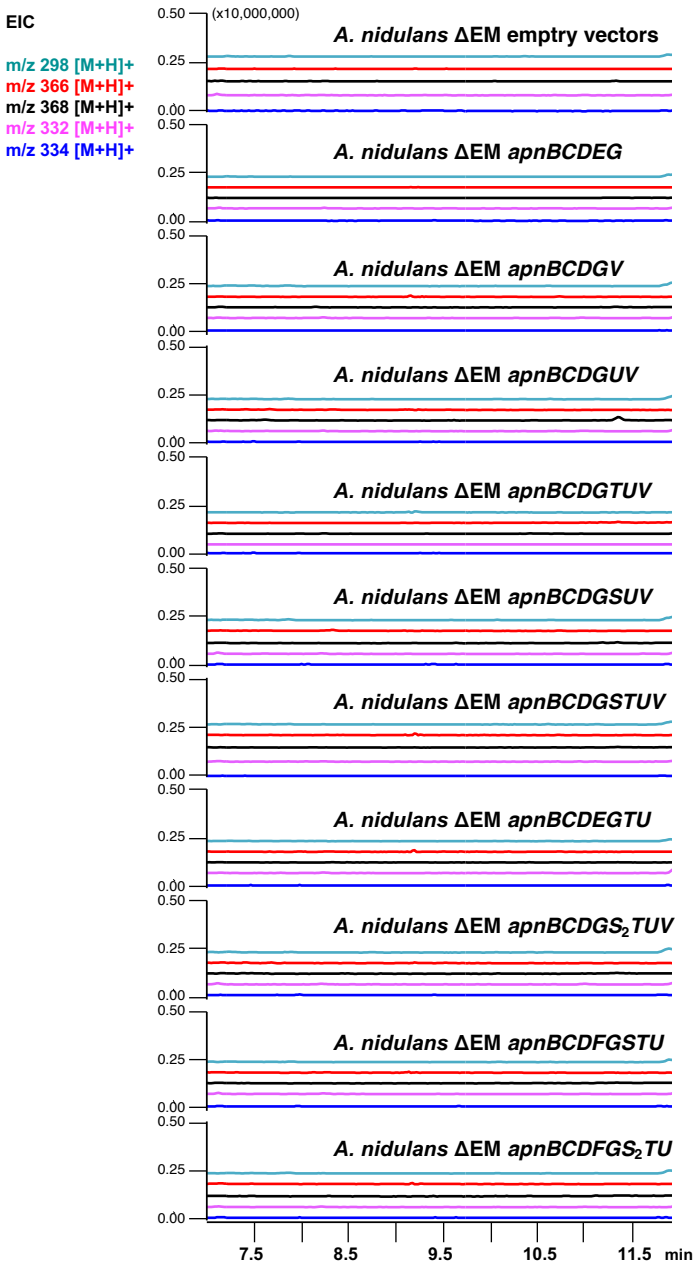


Figure S19. LC/MS analysis of extracts from DMSO-fed *A. nidulans* expressing different combinations of *apn* cluster

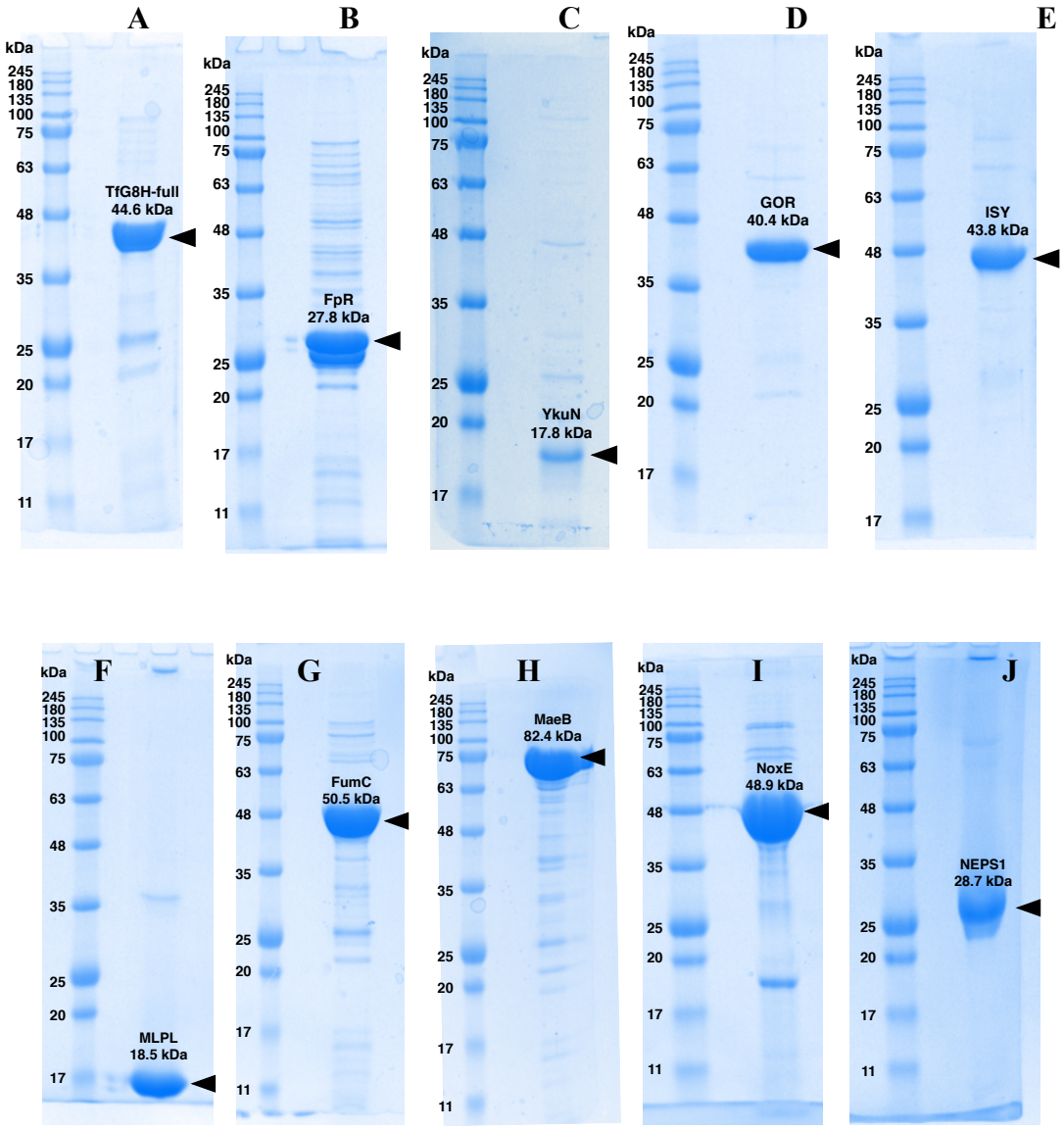


Figure S20. SDS-PAGE of purified proteins A. TfG8H-full, B. FpR, C. YkuN, D. GOR, E. ISY, F. NmMLPL, G. FumC, H. MaeB, I. NoxE. J. NEPS1

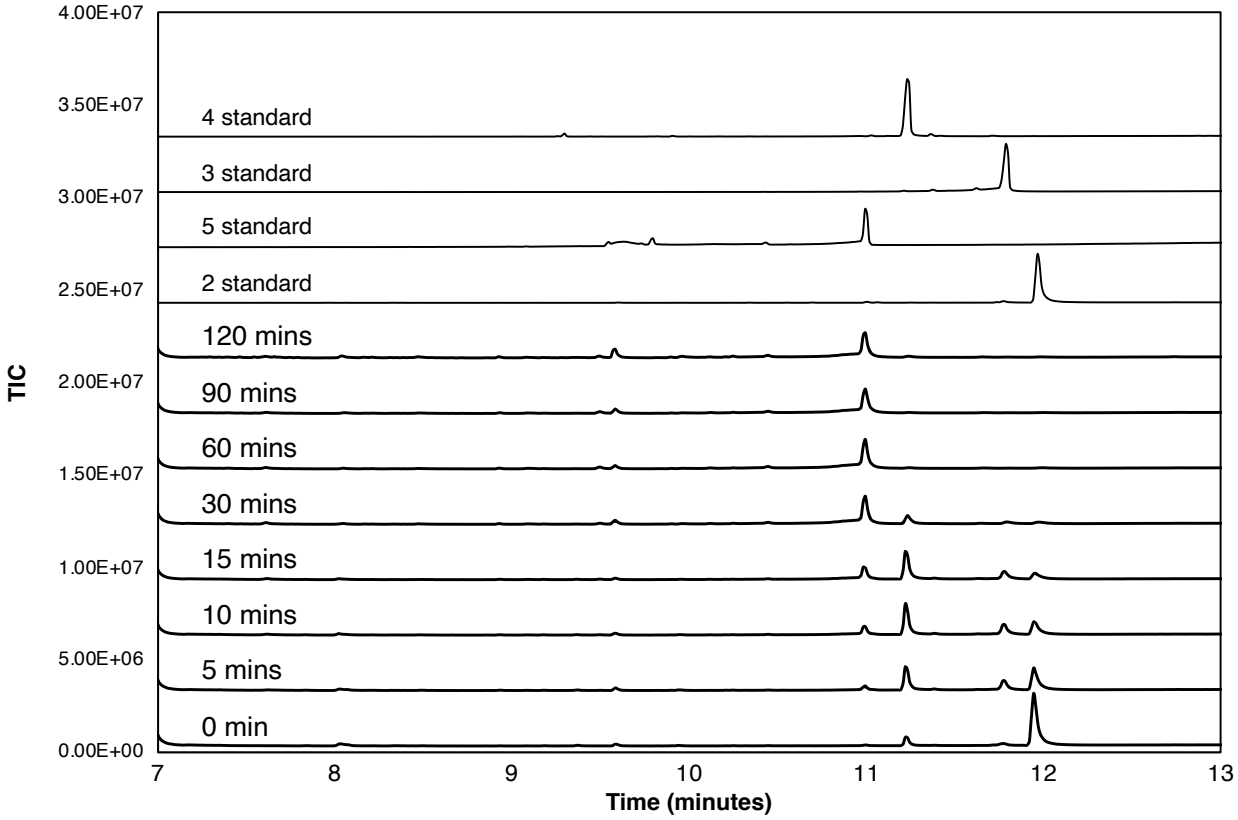


Figure S21. Time-course of 8-hydroxygeraniol oxidation with GOR with NAD^+ regeneration. Each reaction contains 2 mM 8-hydroxygeraniol, 10 μM GOR, 5 μM NoxE and 100 μM NAD^+ in BTP buffer (pH 9.0).

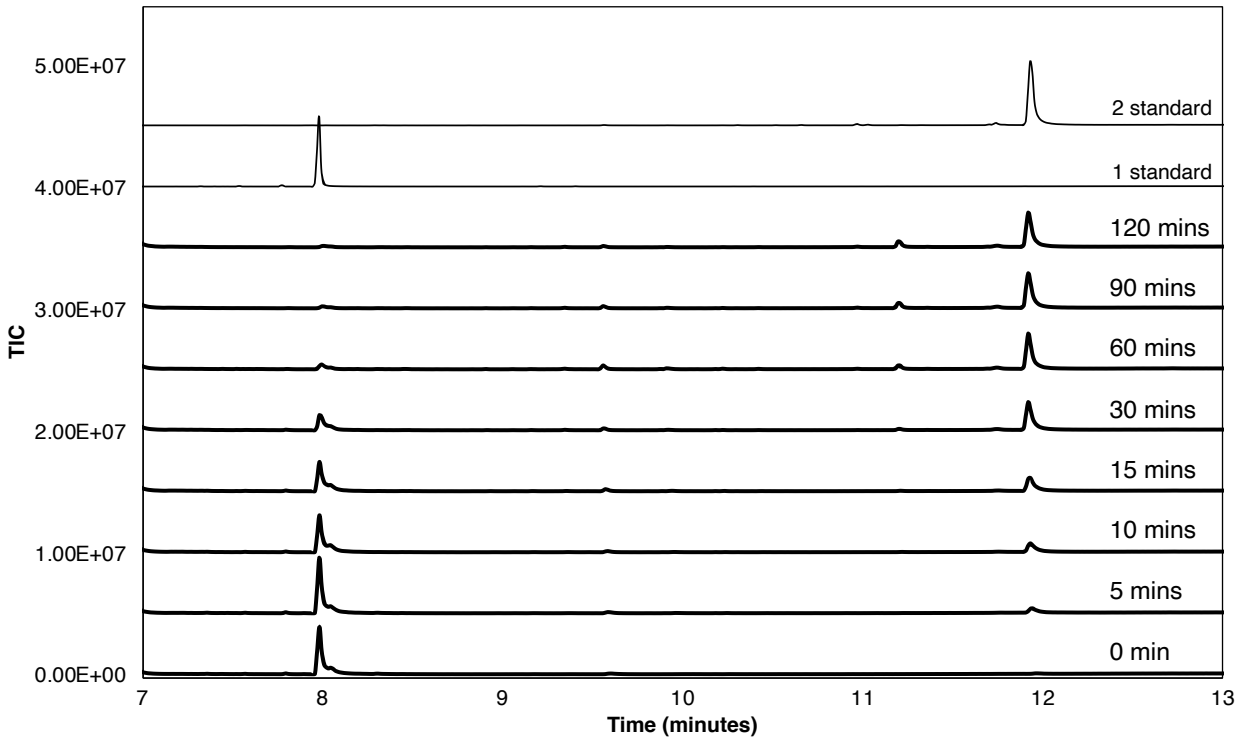


Figure S22. Time-course of geraniol hydroxylation by TfG8H with NADPH regeneration. Each reaction contains 2 mM geraniol, 5 μM TfG8H, 10 μM FpR, 10 μM YkuN, 1 μM FumC, 10 μM MaeB and 100 μM NADPH in BTP buffer (pH 9.0).

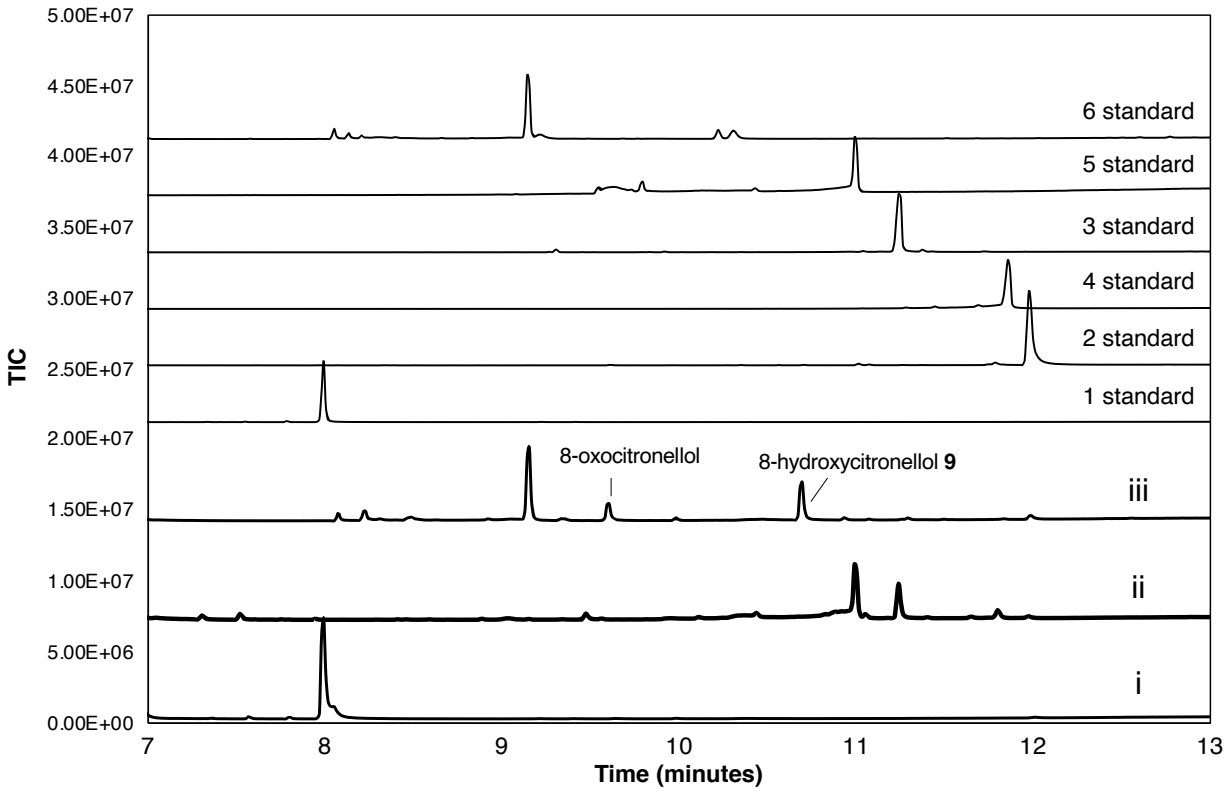


Figure S23. Small-scale one-pot two-step geraniol to nepetalactol conversion. Each reaction contains 2 mM geraniol, 5 μM TfG8H-full, 10 μM FpR, 10 μM YkuN, 10 μM GOR, 0.5 μM ISY, 5 μM NmMLPL, 1 μM FumC, 10 μM MaeB, 100 μM NADPH, 100 μM NAD^+ and 6 mM fumarate in BTP buffer (pH 9.0) unless otherwise specified, i.0 min, ii. no ISY, iii. ISY/NmMLPL added after 2 hours of TfG8H-full/FpR/YkuN and GOR reaction.

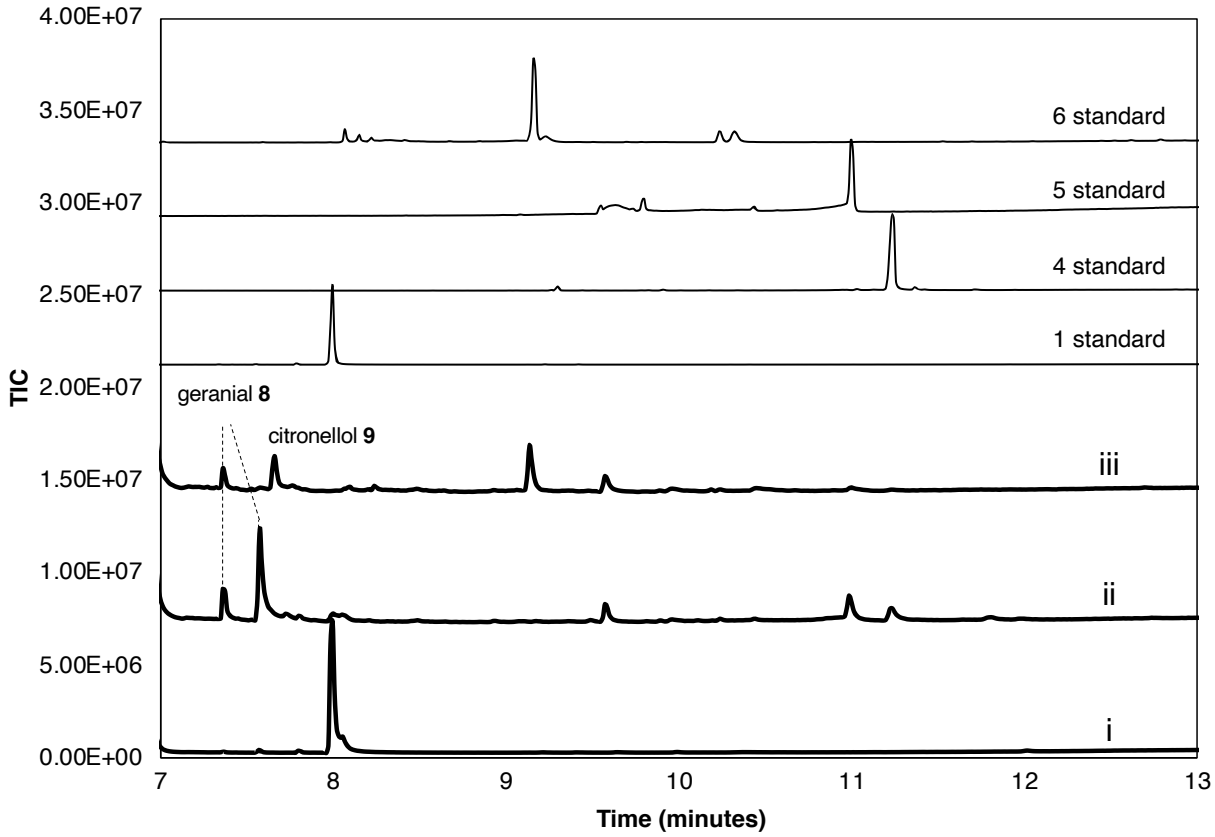


Figure S24. 10 mL-scale one-pot two-step geraniol to nepetalactol conversion. Each reaction contains 2 mM geraniol, 5 μM TfG8H, 10 μM FpR, 10 μM YkuN, 10 μM GOR, 0.5 μM ISY, 5 μM NmMLPL, 1 μM FumC, 10 μM MaeB, 100 μM NADPH, 100 μM NAD^+ and 6 mM fumarate in BTP buffer (pH 9.0) unless otherwise specified, i.0 min, ii. no ISY/NmMLPL, iii. ISY/NmMLPL added after 2 hours of TfG8H-full/FpR/YkuN and GOR reaction.

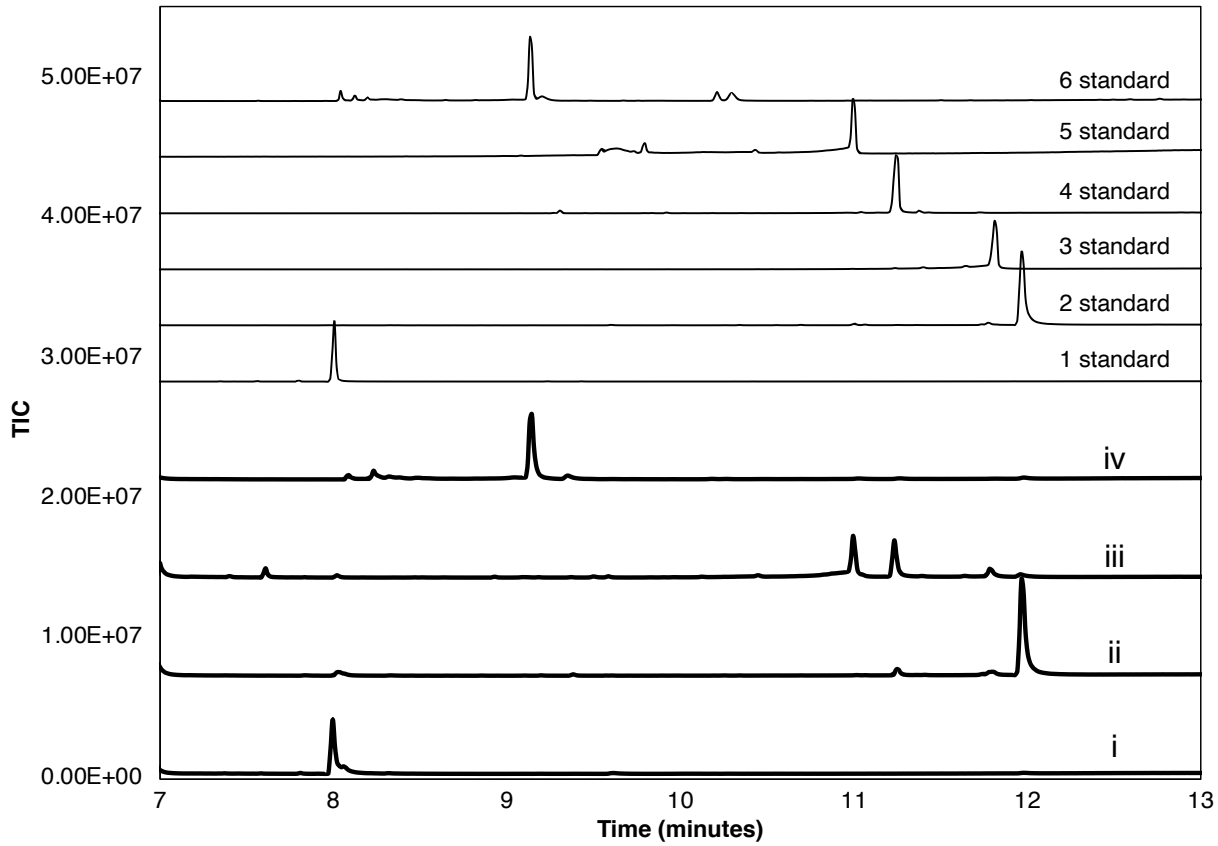


Figure S25. 10 mL-scale one-pot drop-in geraniol to nepetalactol conversion. Reaction contains 2 mM geraniol, 5 μM TfG8H-full, 10 μM FpR, 10 μM YkuN, 10 μM GOR, 0.5 μM ISY, 5 μM NmMLPL, 1 μM FumC, 10 μM MaeB, 100 μM NADPH, 100 μM NAD^+ and 12 mM fumarate in BTP buffer (pH 9.0) unless otherwise specified, i.0 min, ii. 2 hours reaction with TfG8H-full, iii. 2 hours reaction with GOR after TfG8Hl, iv. 2 hours reaction with ISY/NmMLPL after TfG8H-full and GOR

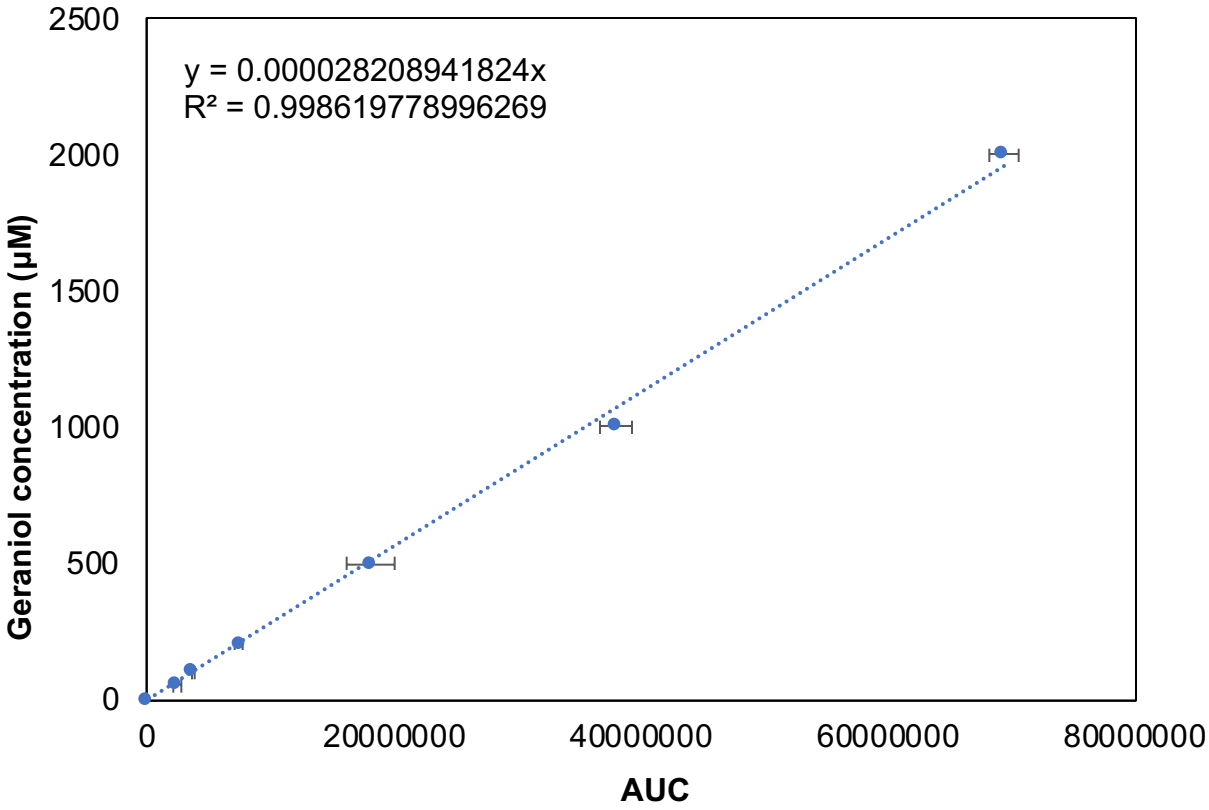


Figure S26. Geraniol calibration curve.

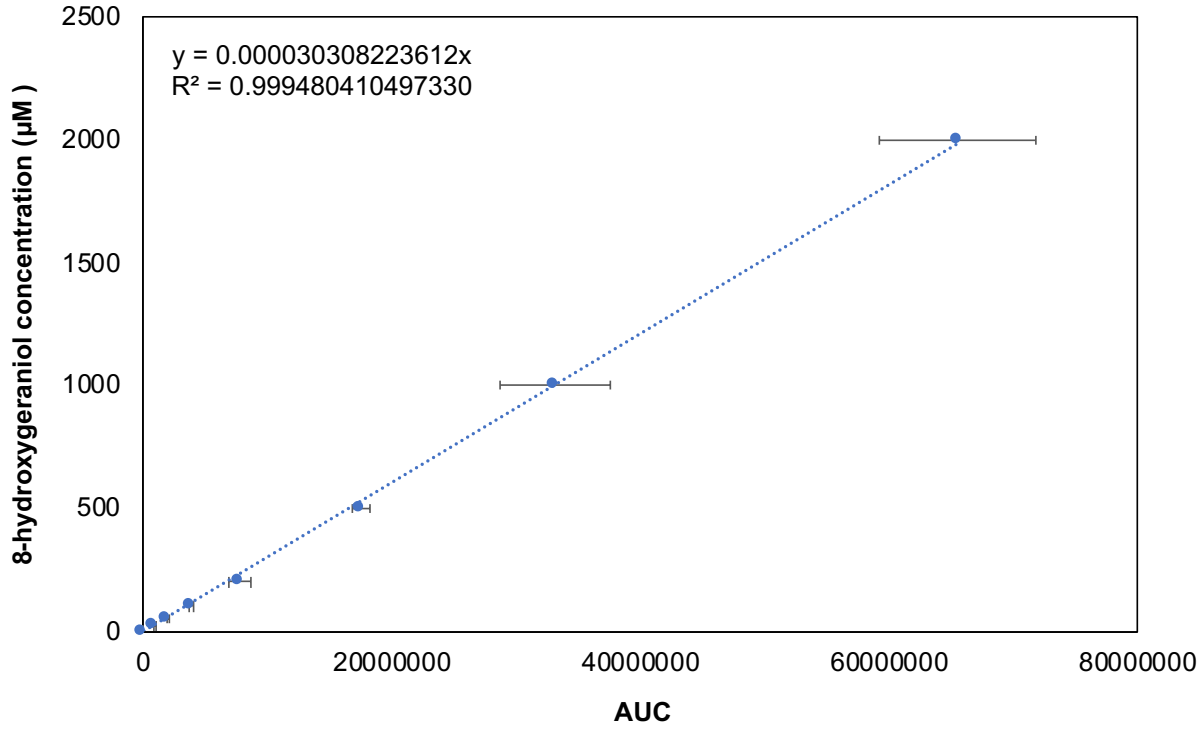


Figure S27. 8-hydroxygeraniol calibration curve.

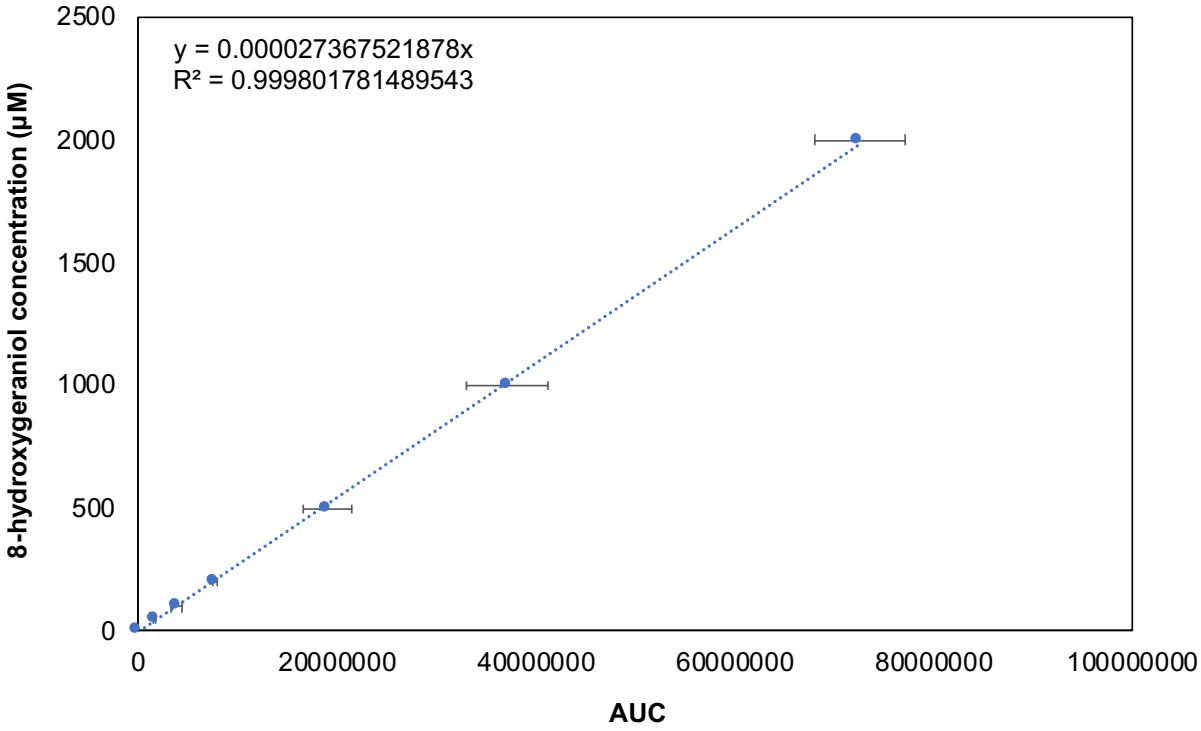


Figure S28. 8-oxogeranial calibration curve.

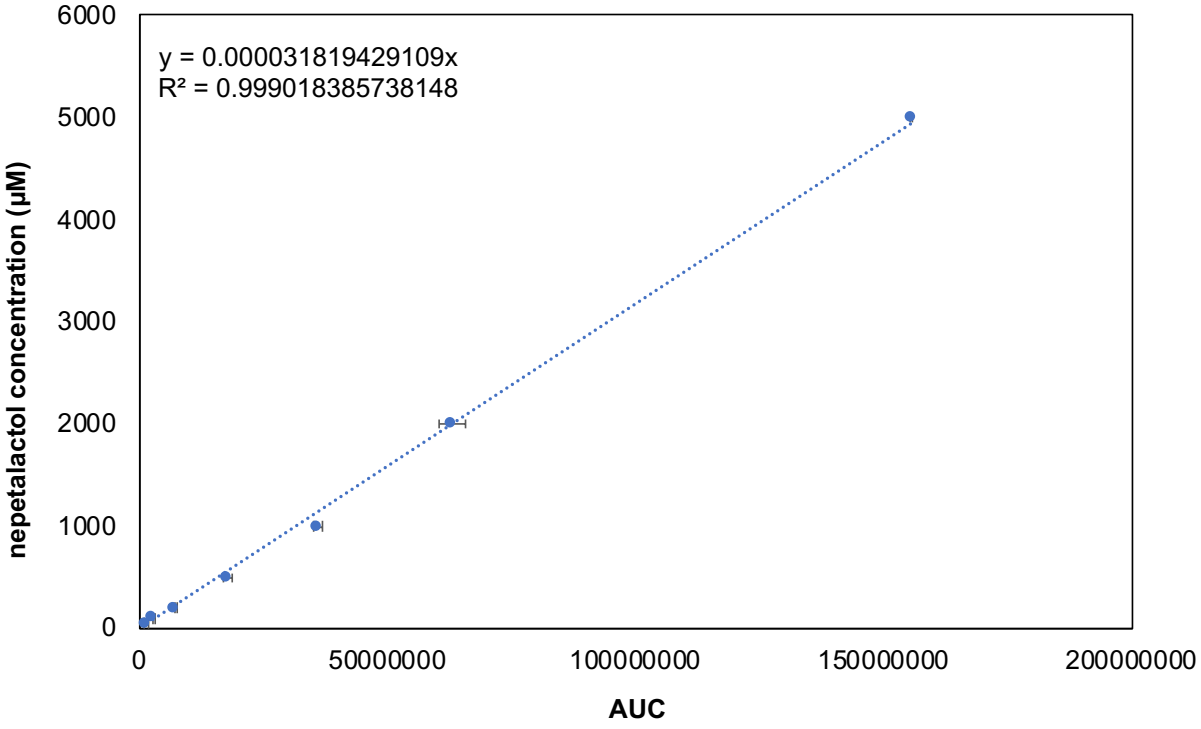


Figure S29. Nepetalactol calibration curve.

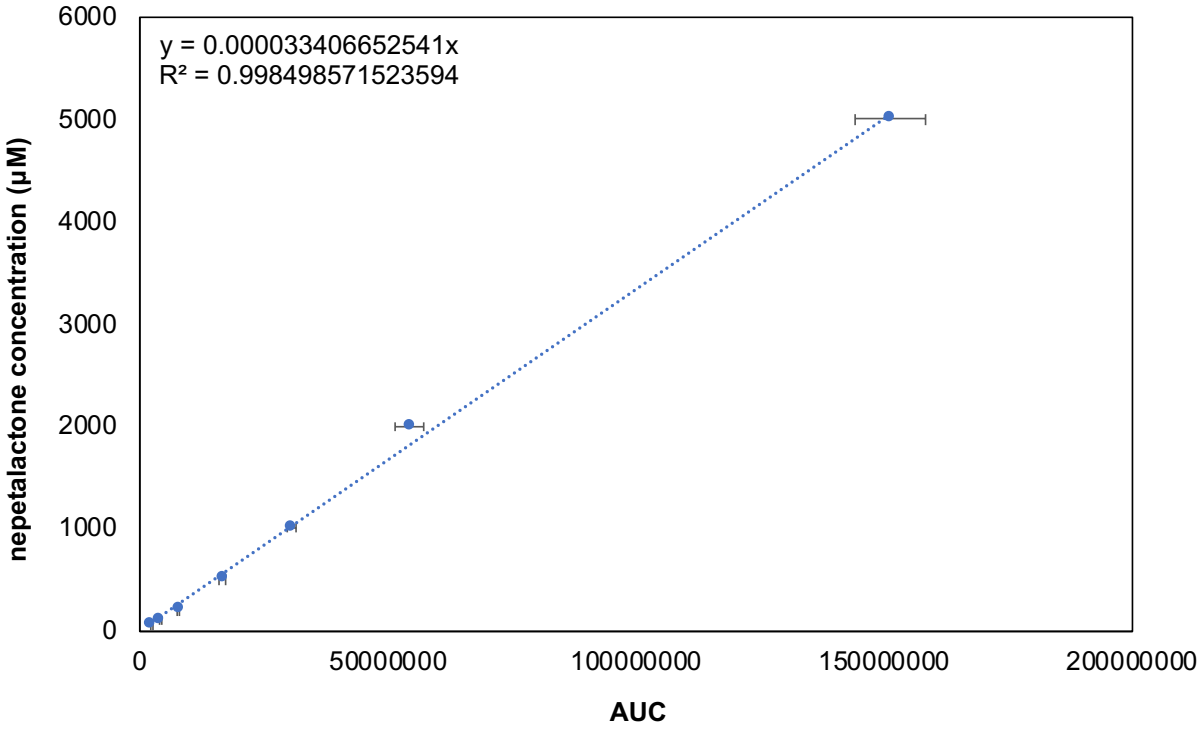


Figure S30. Nepetalactone calibration curve.

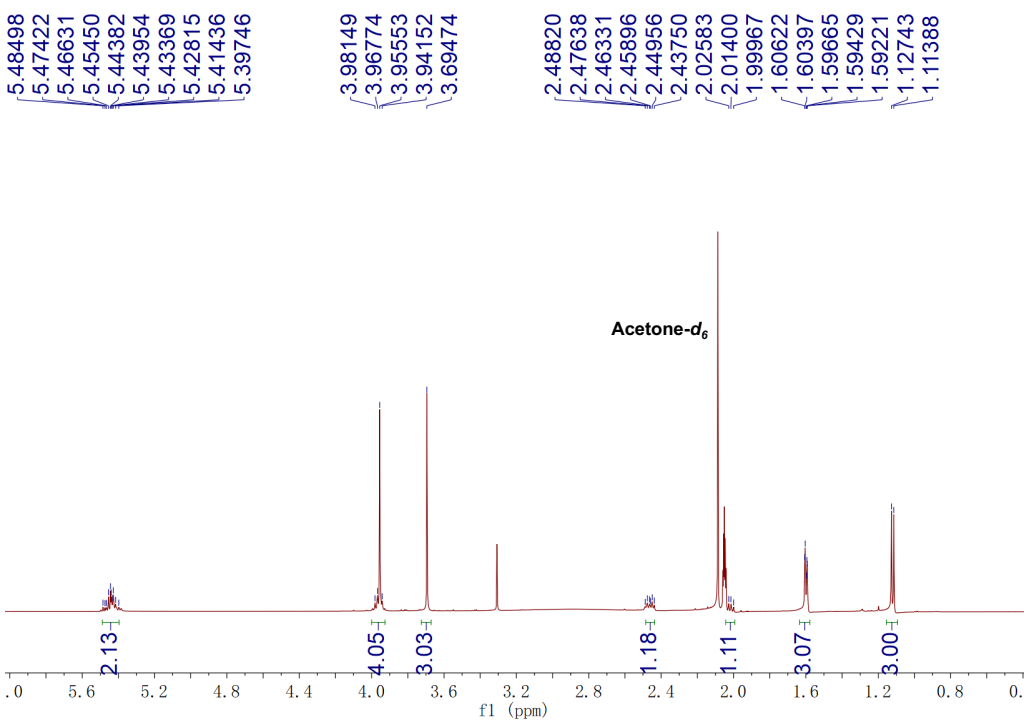


Figure S31. ^{1}H NMR (500 MHz) spectrum of compound **1** in Acetone-d_6.

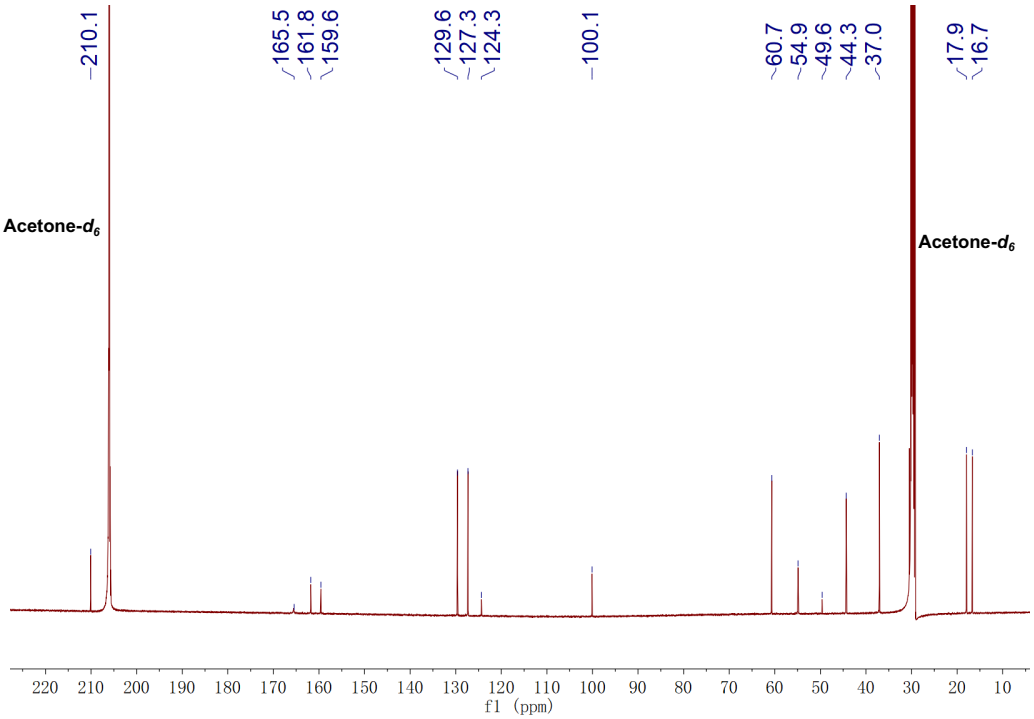


Figure S32. ^{13}C NMR (125 MHz) spectrum of compound **1** in Acetone-d_6.

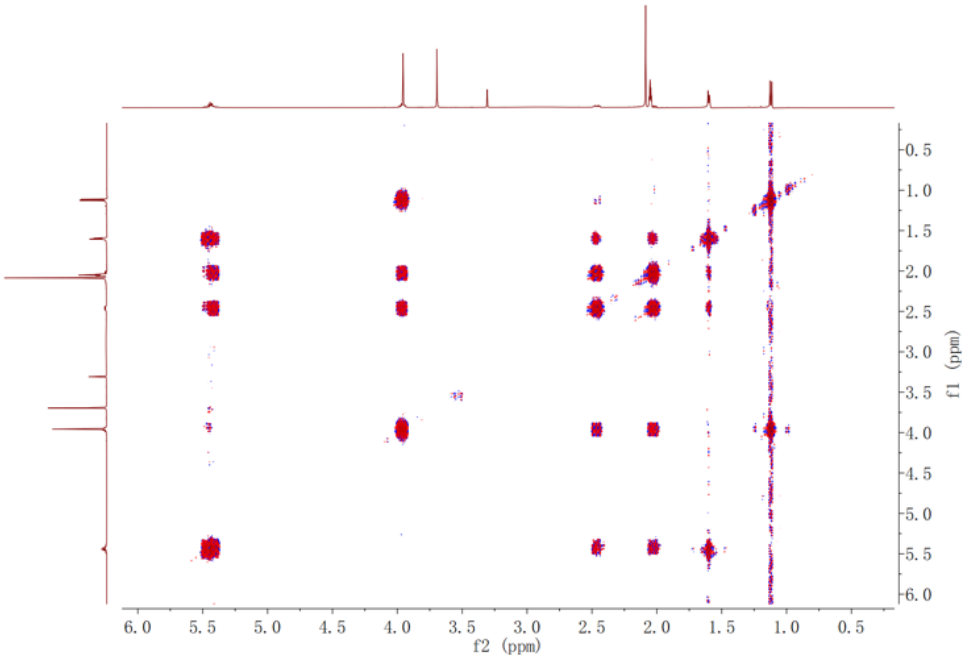


Figure S33. ^{1}H-^{1}H COSY spectrum of compound **1** in Acetone-d_6.

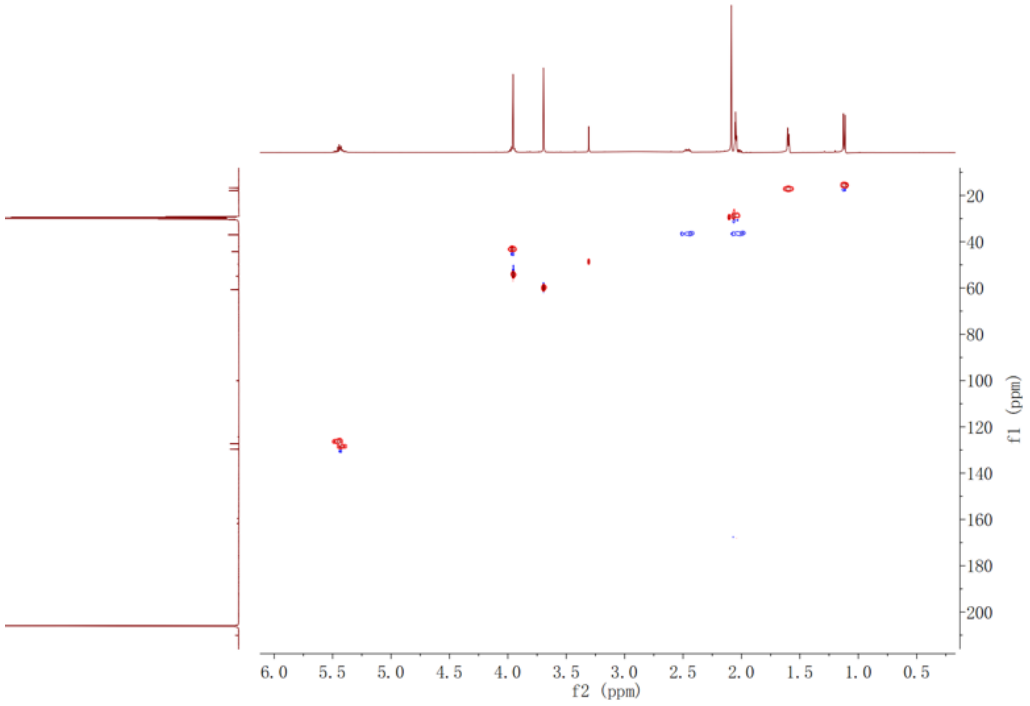


Figure S34. HSQC spectrum of compound **1** in Acetone-d_6.

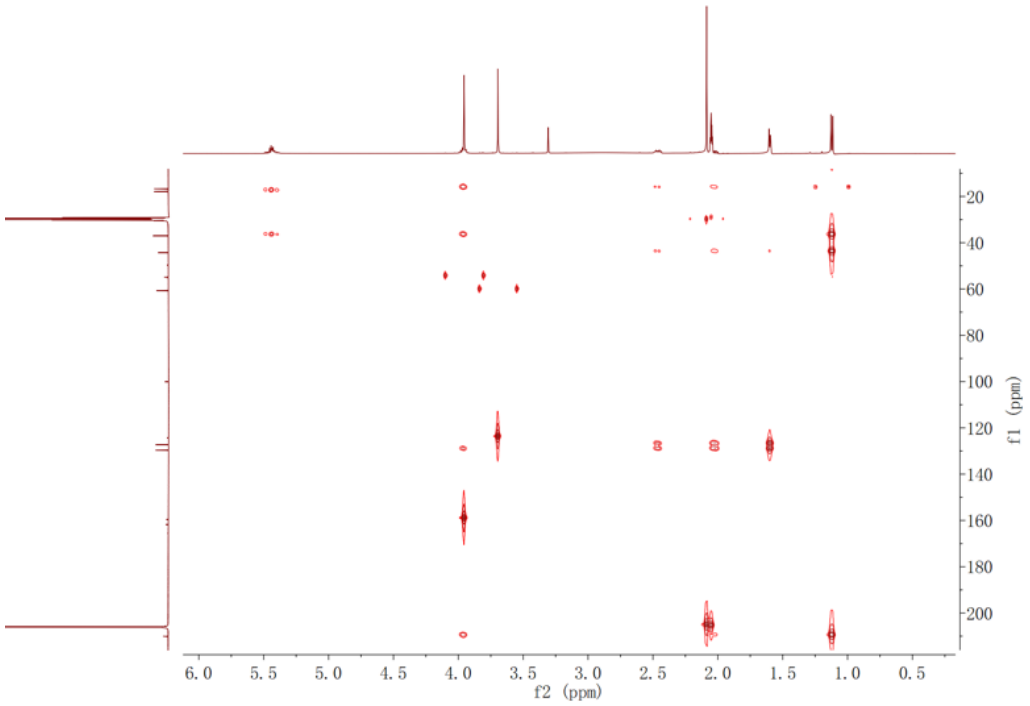


Figure S35. HMBC spectrum of compound **1** in Acetone-d_6.

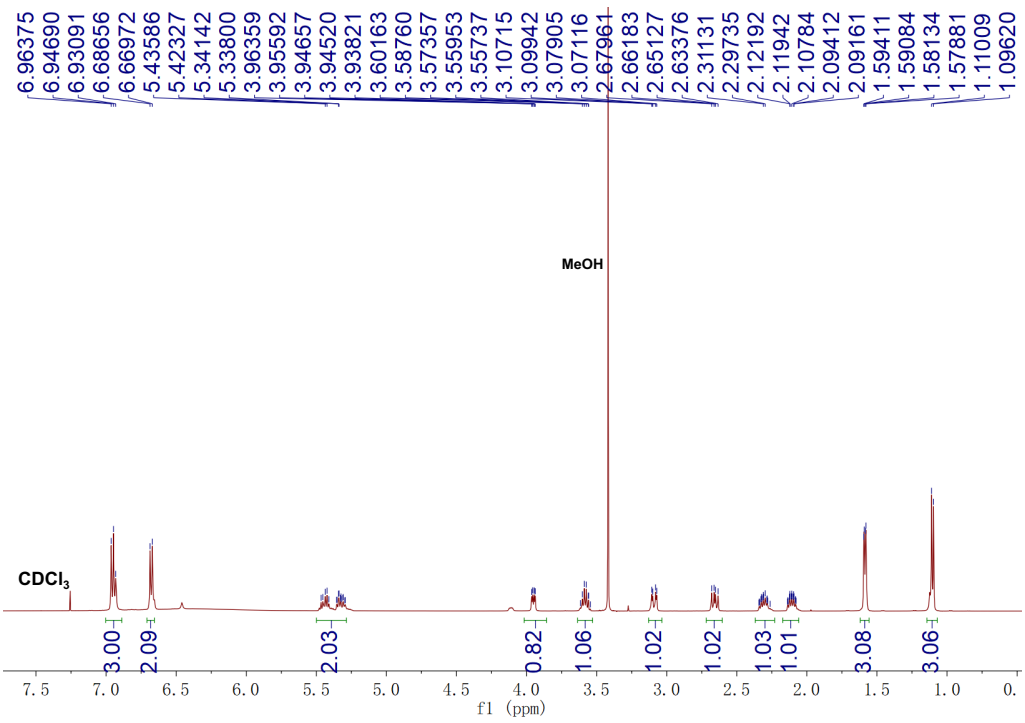


Figure S36. ^{1}H NMR (500 MHz) spectrum of compound **2** in $CDCl_3$.

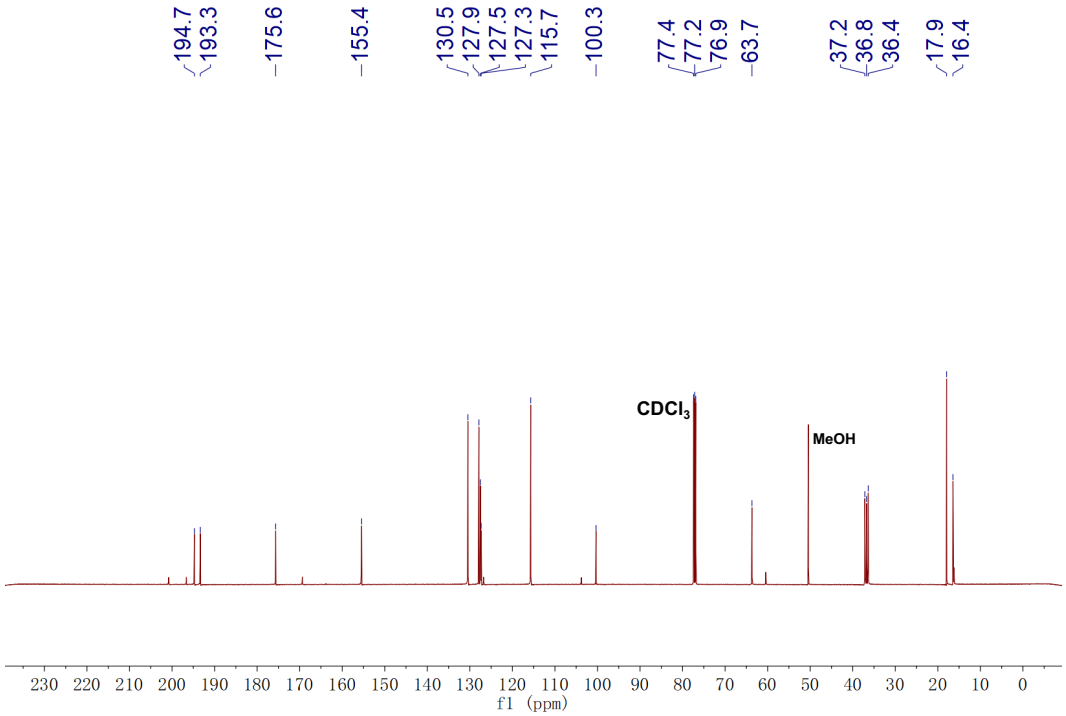


Figure S37. ^{13}C NMR (125 MHz) spectrum of compound **2** in $CDCl_3$.

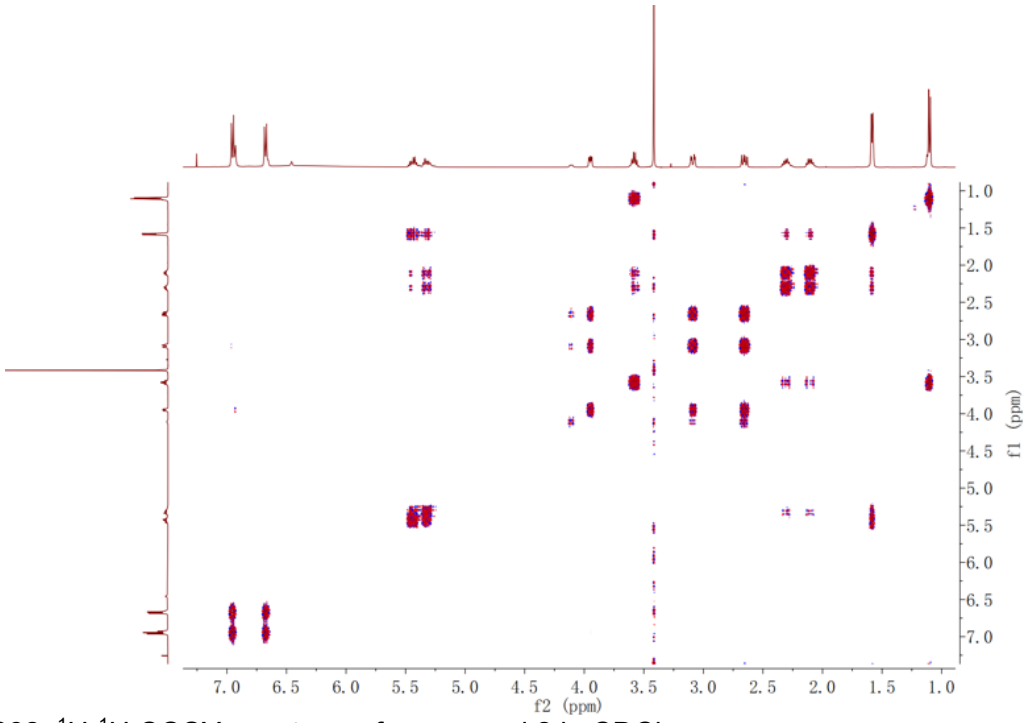


Figure S38. ^{1}H-^{1}H COSY spectrum of compound **2** in $CDCl_3$.

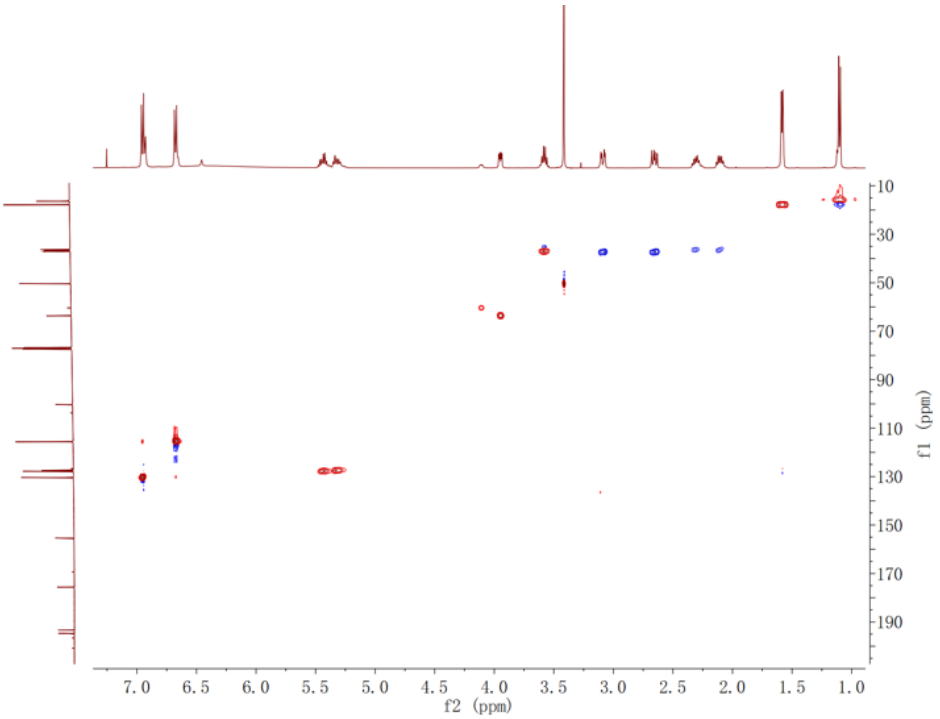


Figure S39. HSQC spectrum of compound **2** in $CDCl_3$.

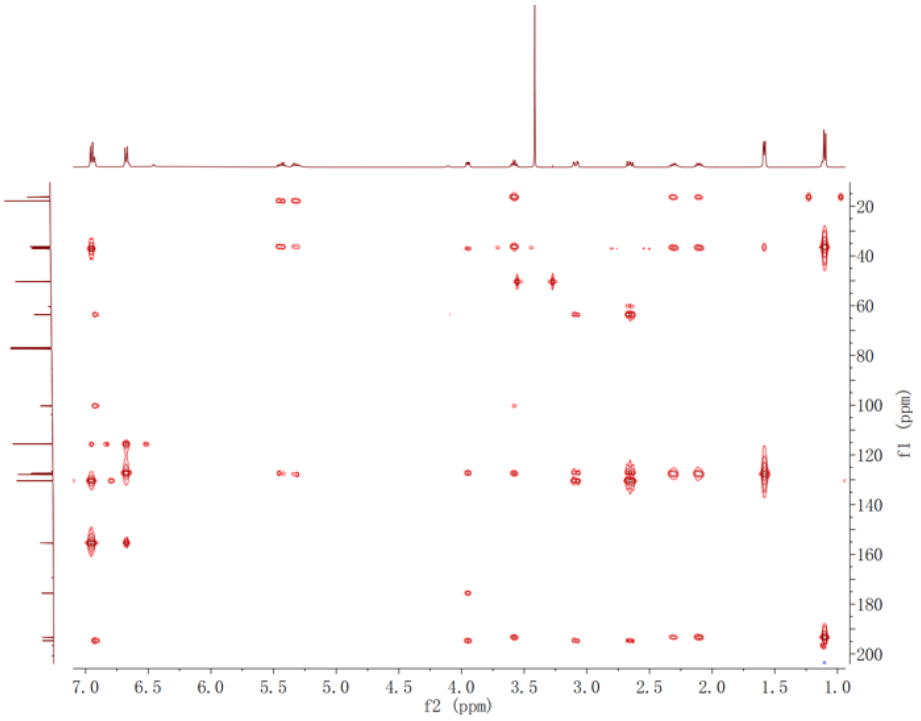


Figure S40. HMBC spectrum of compound **2** in $CDCl_3$.

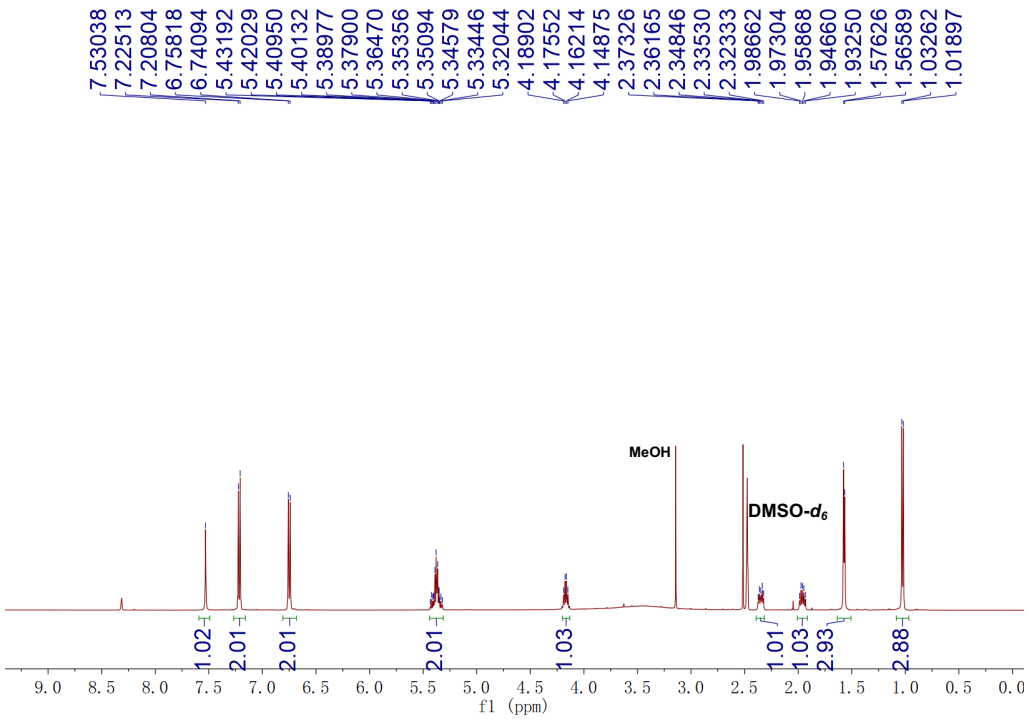


Figure S41. ^{1}H NMR (500 MHz) spectrum of compound **3** in DMSO-d_6.

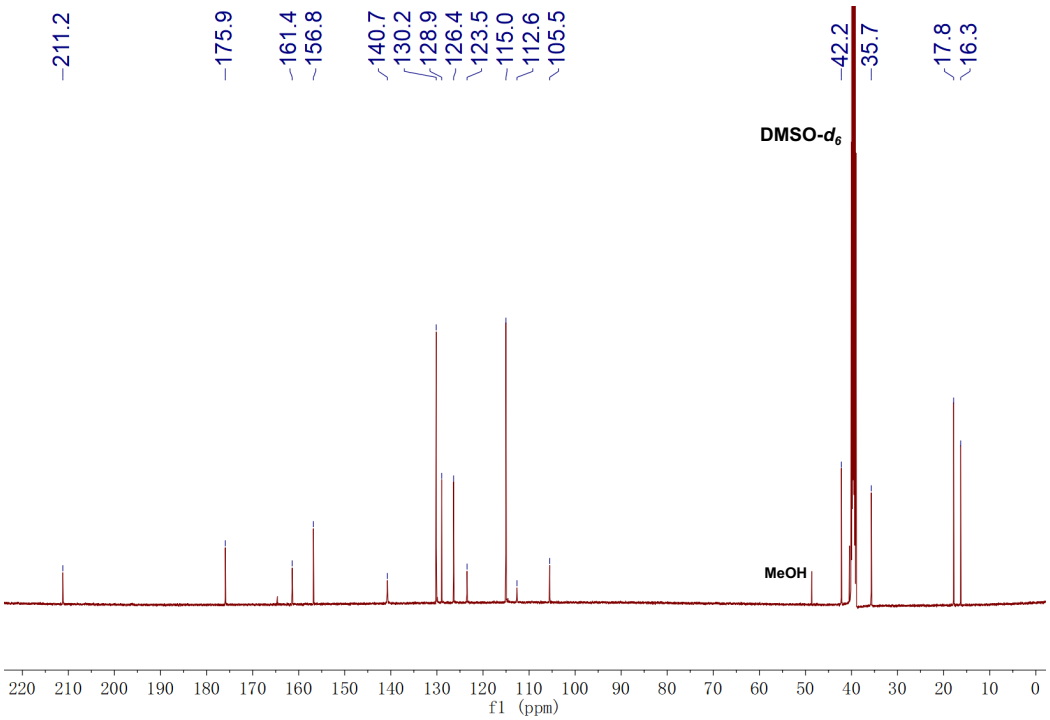


Figure S42. ^{13}C (125 MHz) NMR spectrum of compound **3** in DMSO-d_6.

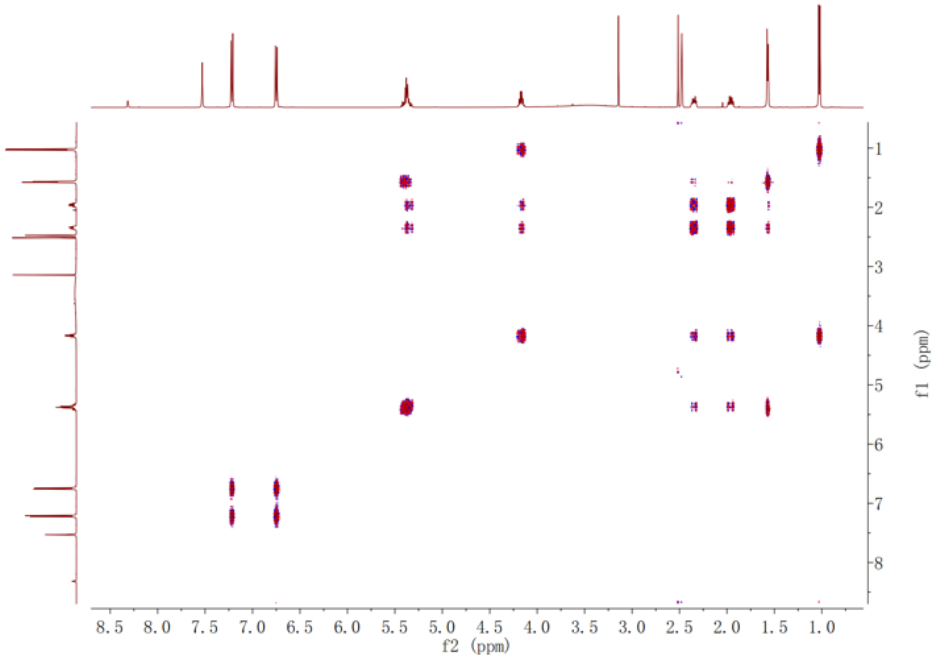


Figure S43. ^{1}H-^{1}H COSY spectrum of compound **3** in DMSO-d_6.

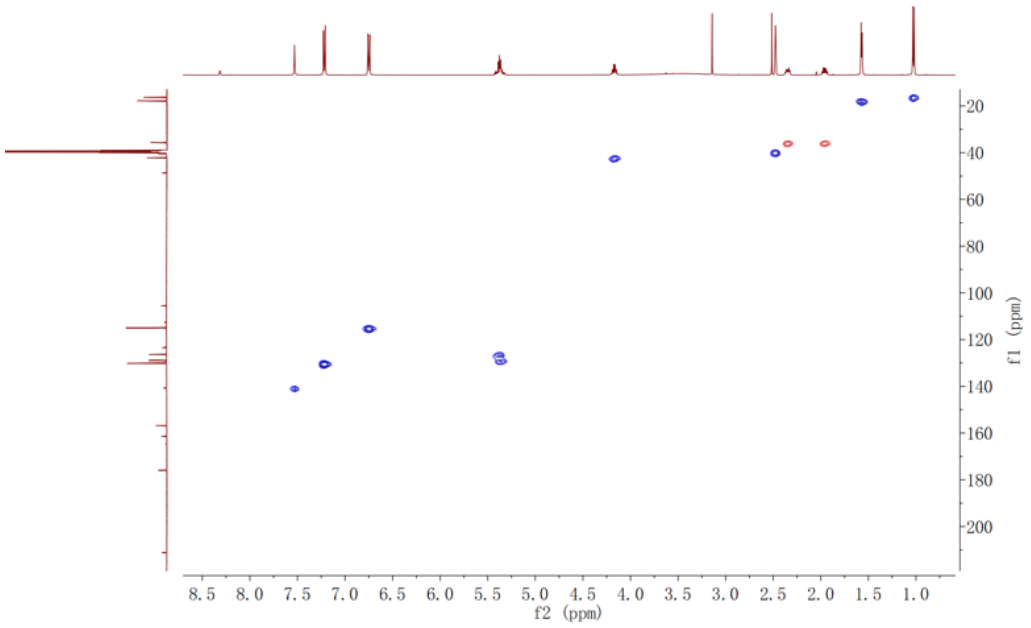


Figure S44. HSQC spectrum of compound **3** in DMSO-d_6.

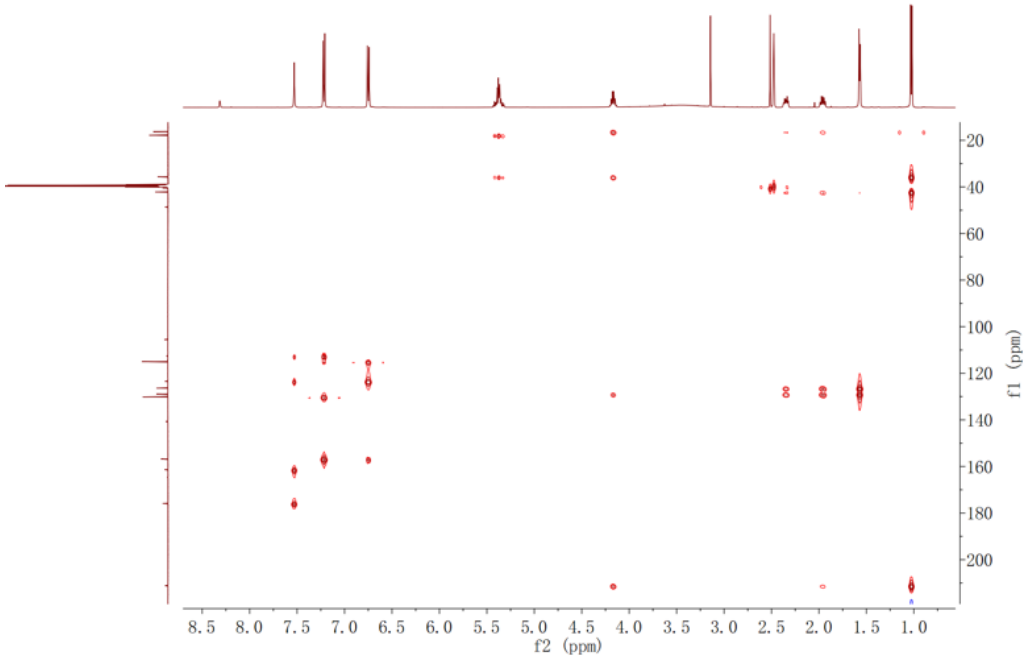


Figure S45. HMBC spectrum of compound **3** in DMSO-d_6.

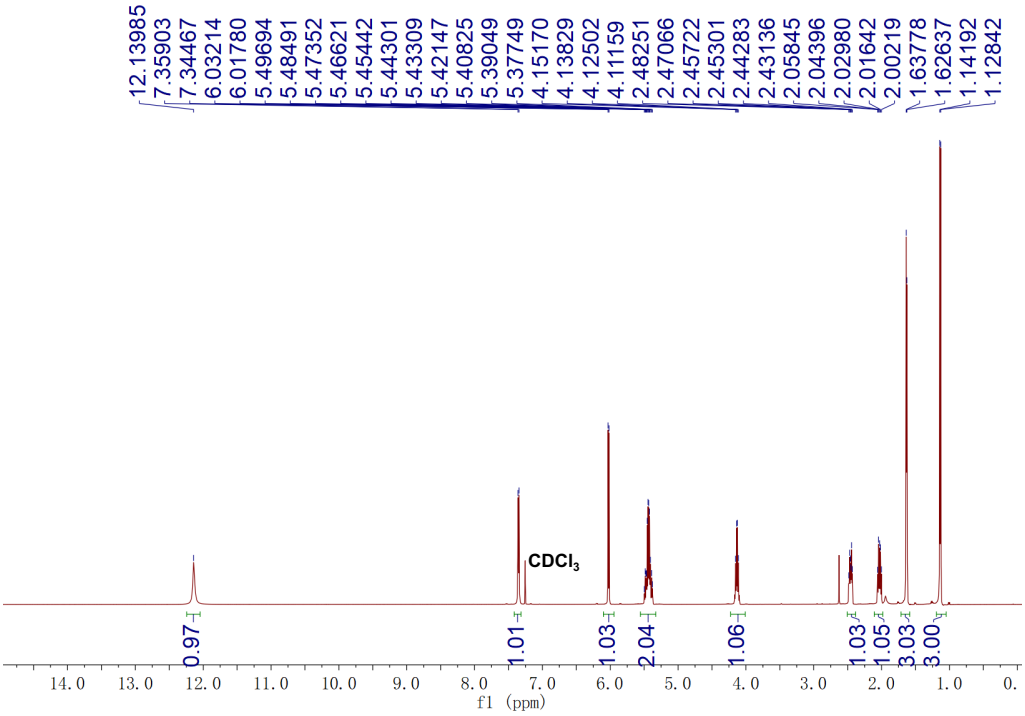


Figure S46. ^{1}H (500 MHz) NMR spectrum of compound **4** in $CDCl_3$.

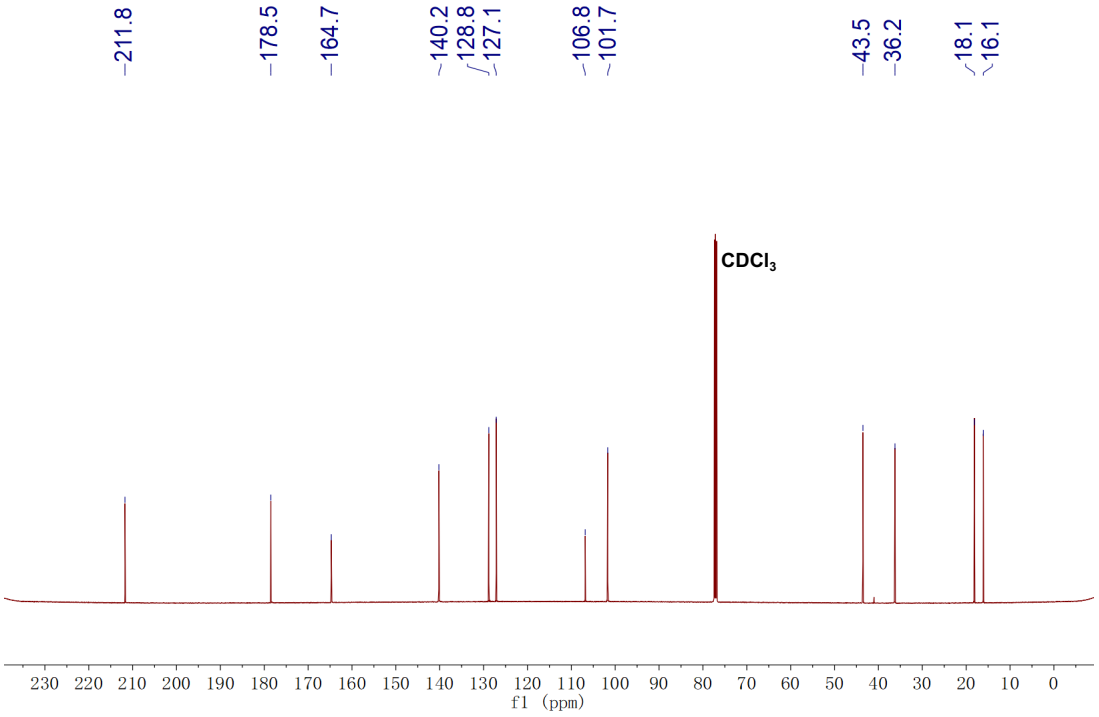


Figure S47. ^{13}C (125 MHz) NMR spectrum of compound **4** in $CDCl_3$.

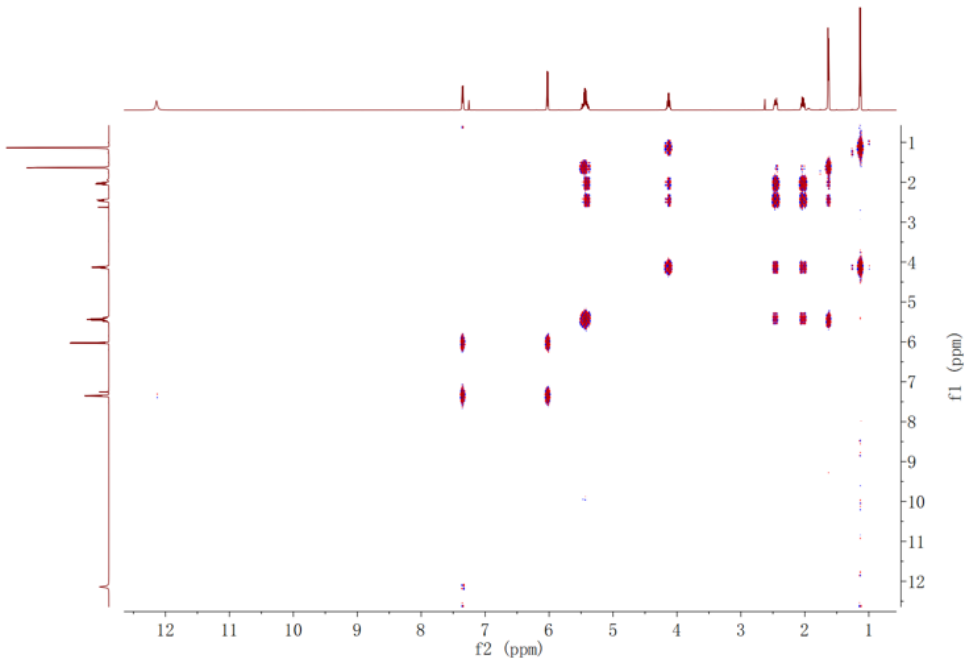


Figure S48. ^{1}H-^{1}H COSY spectrum of compound **4** in $CDCl_3$.

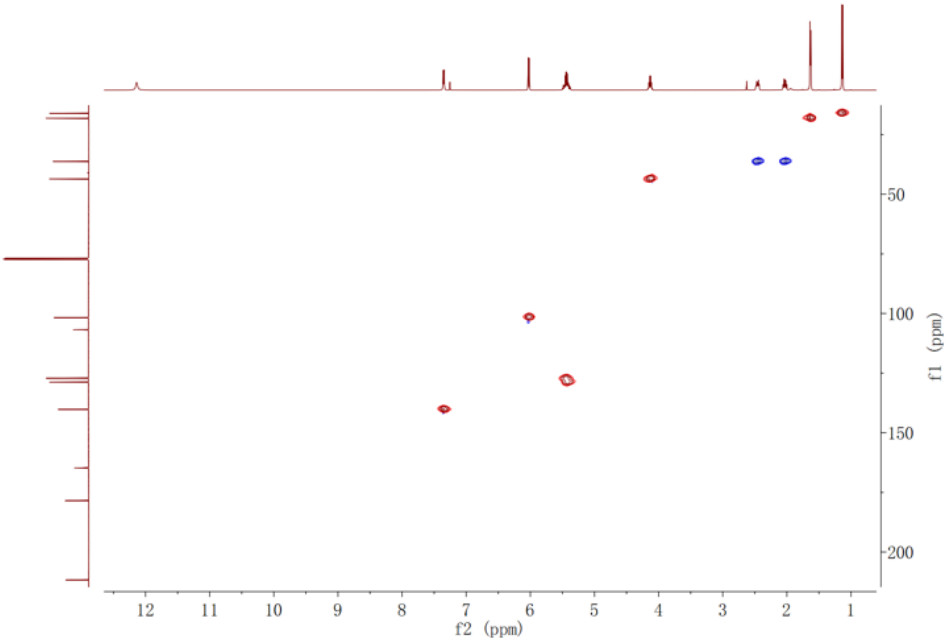


Figure S49. HSQC spectrum of compound **4** in $CDCl_3$.

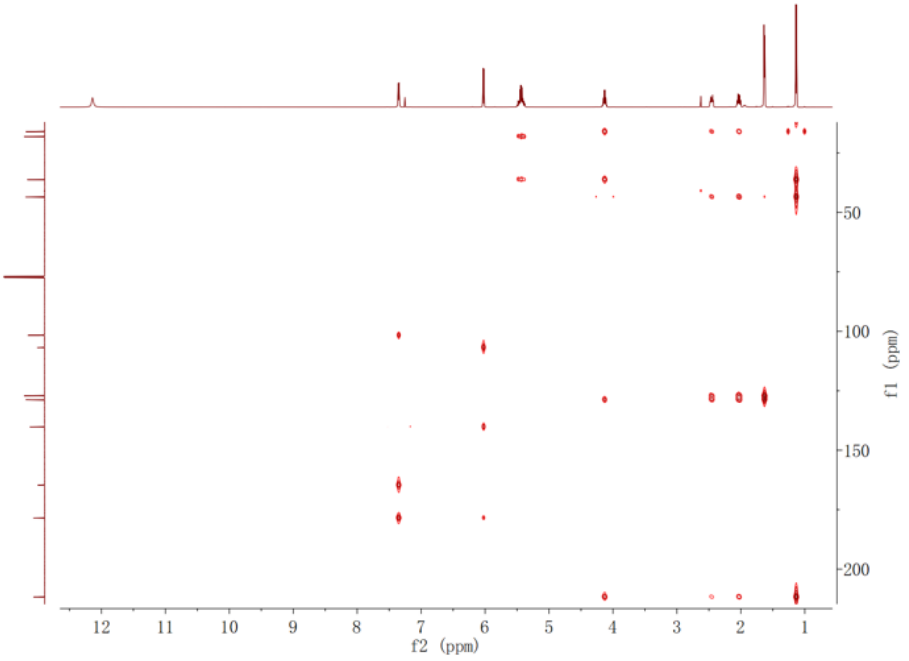


Figure S50. HMBC spectrum of compound **4** in $CDCl_3$.

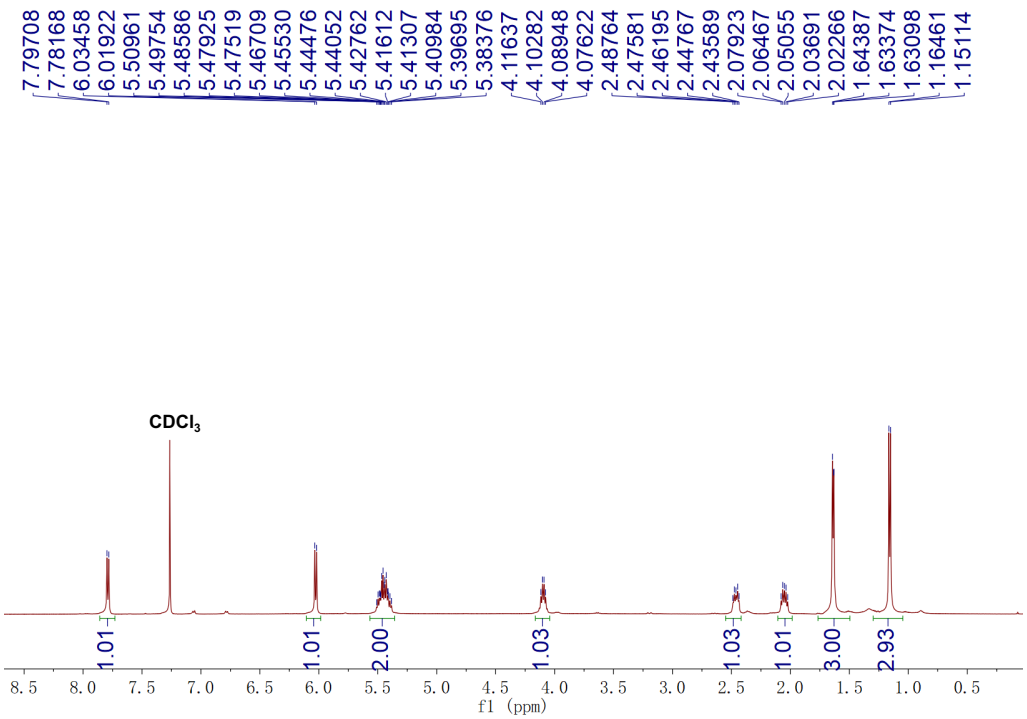


Figure S51. ^{1}H NMR (500 MHz) spectrum of compound **5** in $CDCl_3$.

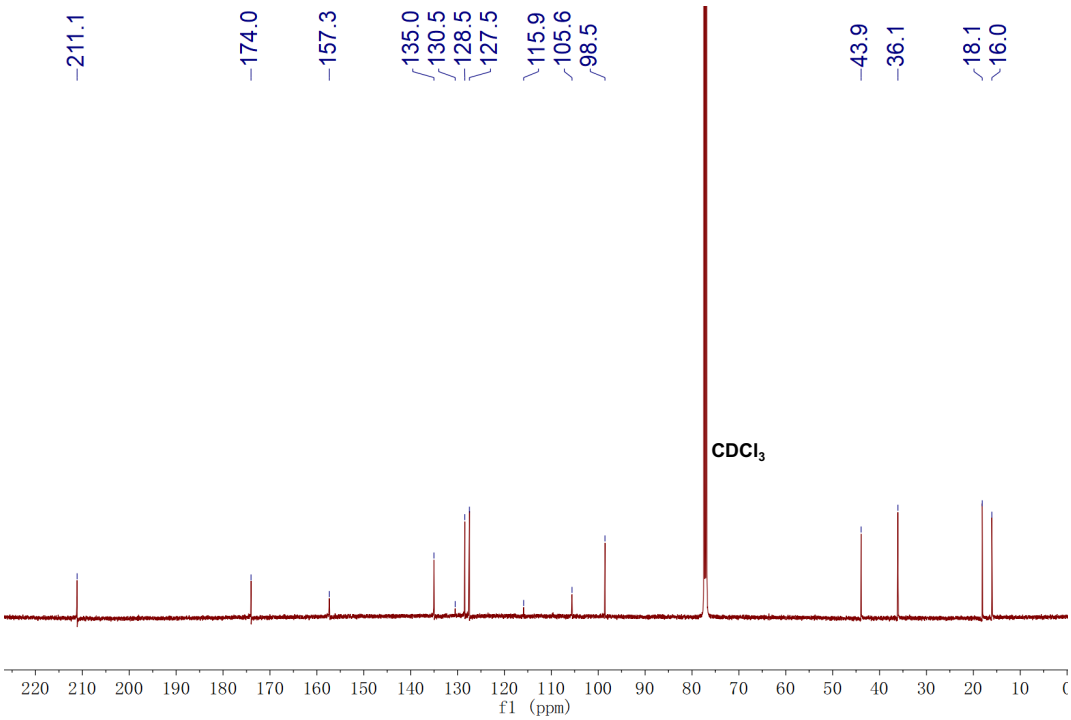


Figure S52. ^{13}C (125 MHz) NMR spectrum of compound **5** in $CDCl_3$.

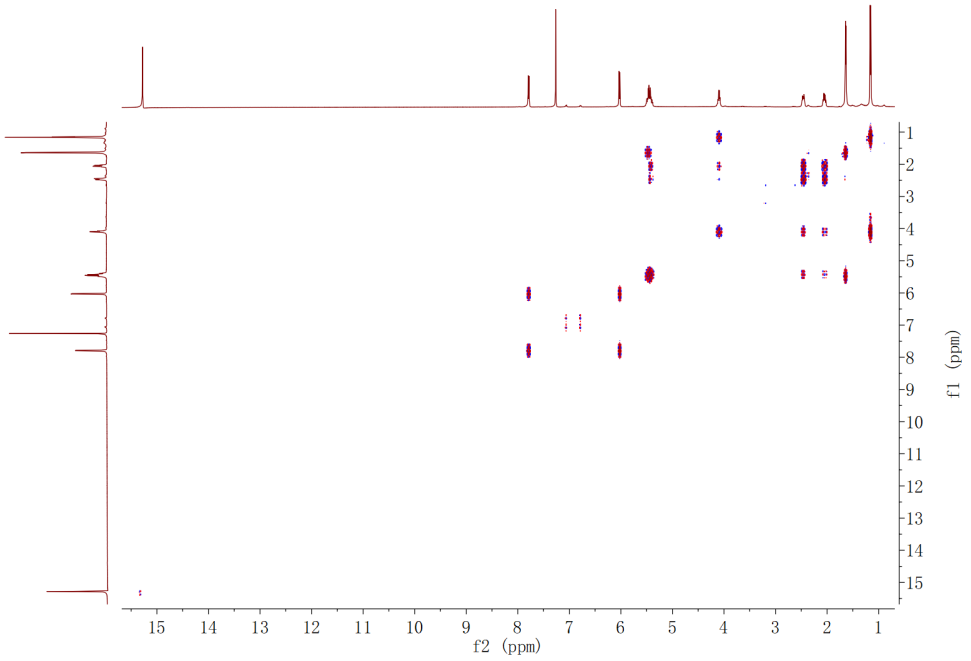


Figure S53. ^{1}H-^{1}H COSY spectrum of compound **5** in $CDCl_3$.

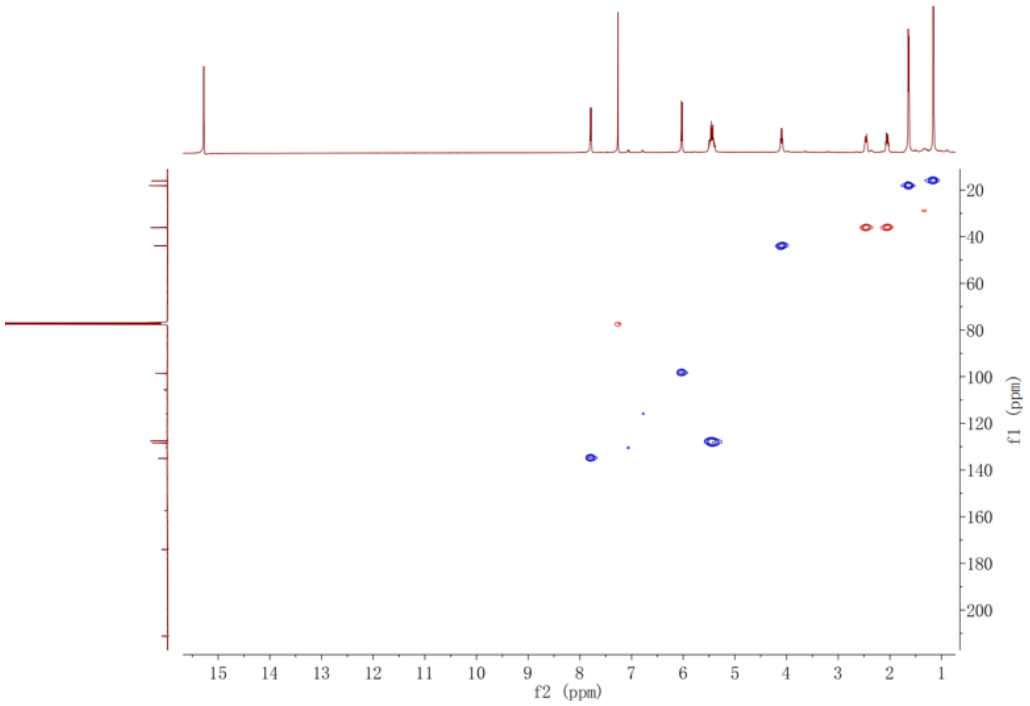


Figure S54. HSQC spectrum of compound **5** in $CDCl_3$.

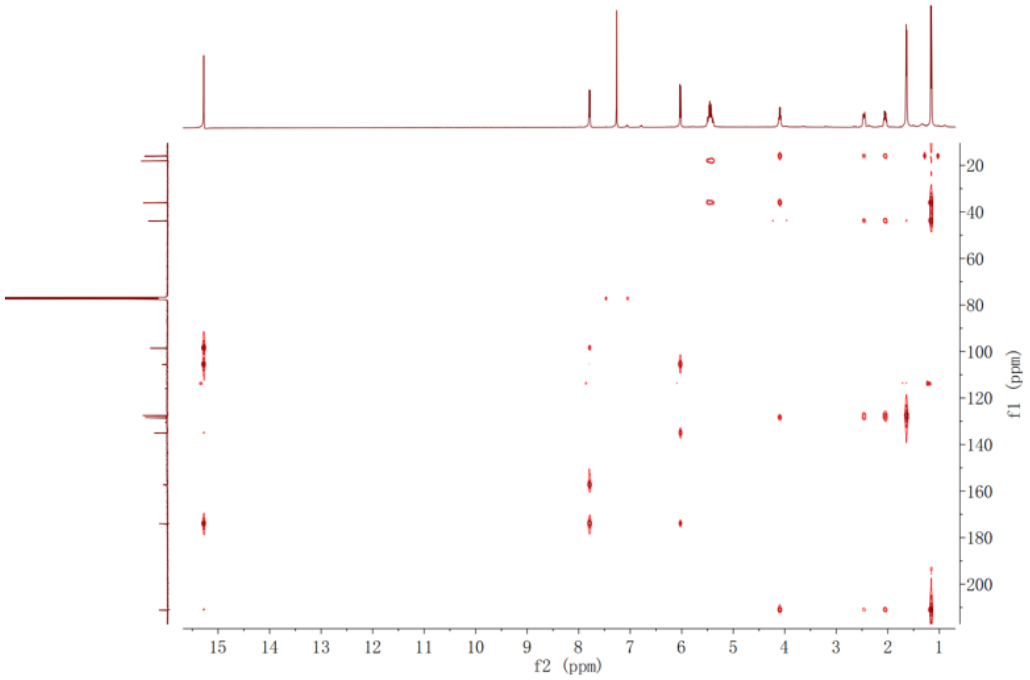


Figure S55. HMBC spectrum of compound **5** in $CDCl_3$.

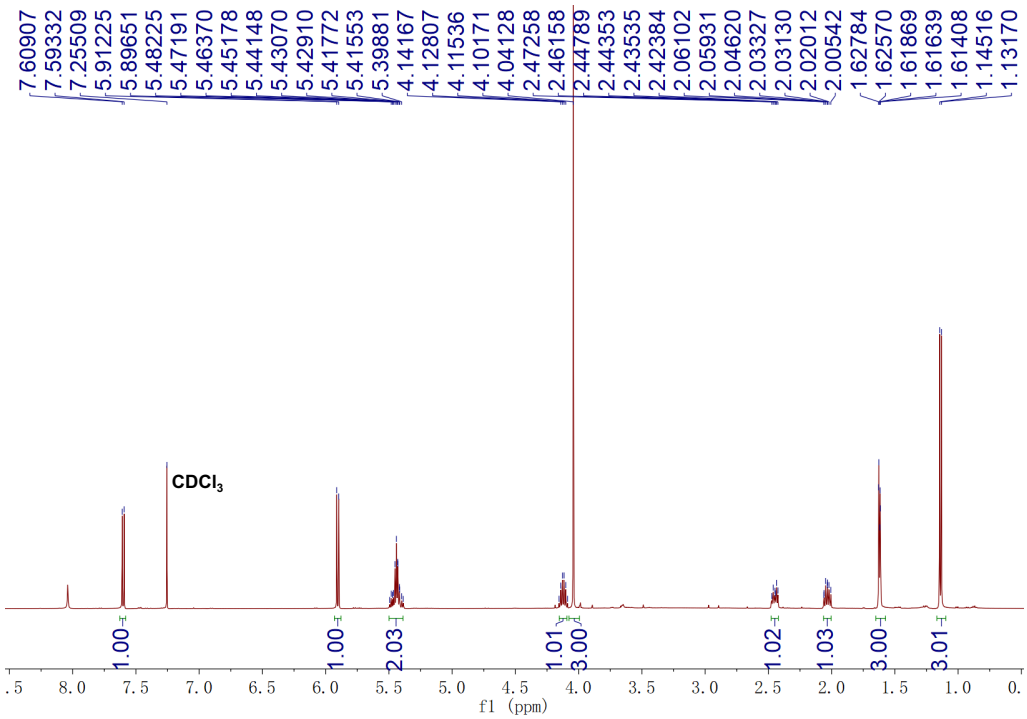


Figure S56. ^{1}H NMR (500 MHz) spectrum of compound **6** in $CDCl_3$.

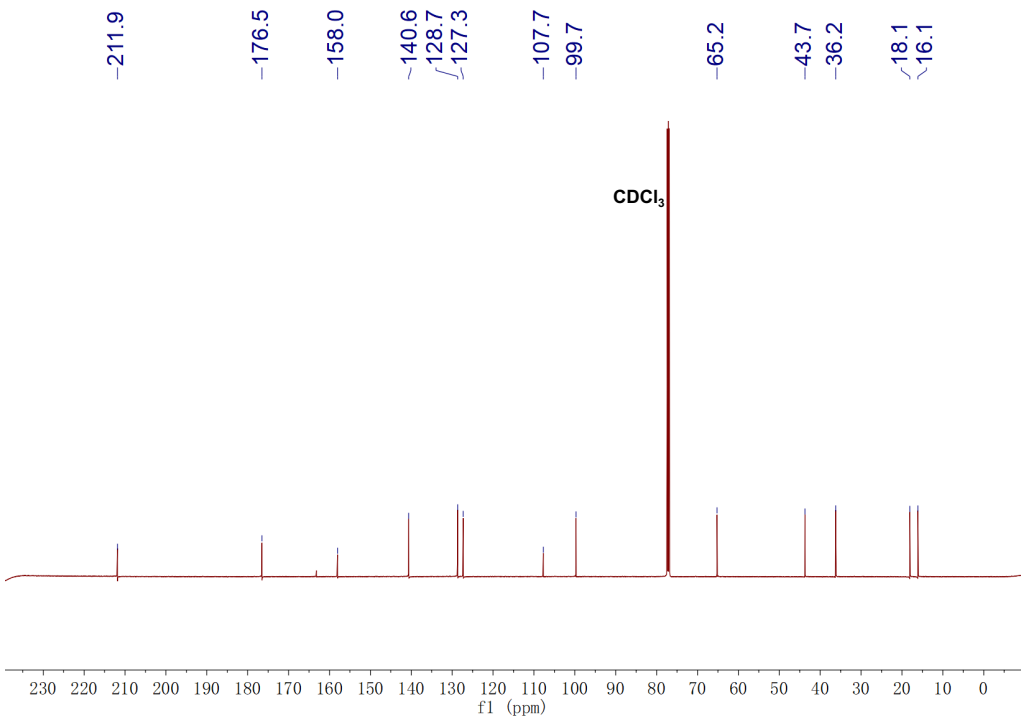


Figure S57. ^{13}C NMR (125 MHz) spectrum of compound **6** in $CDCl_3$.

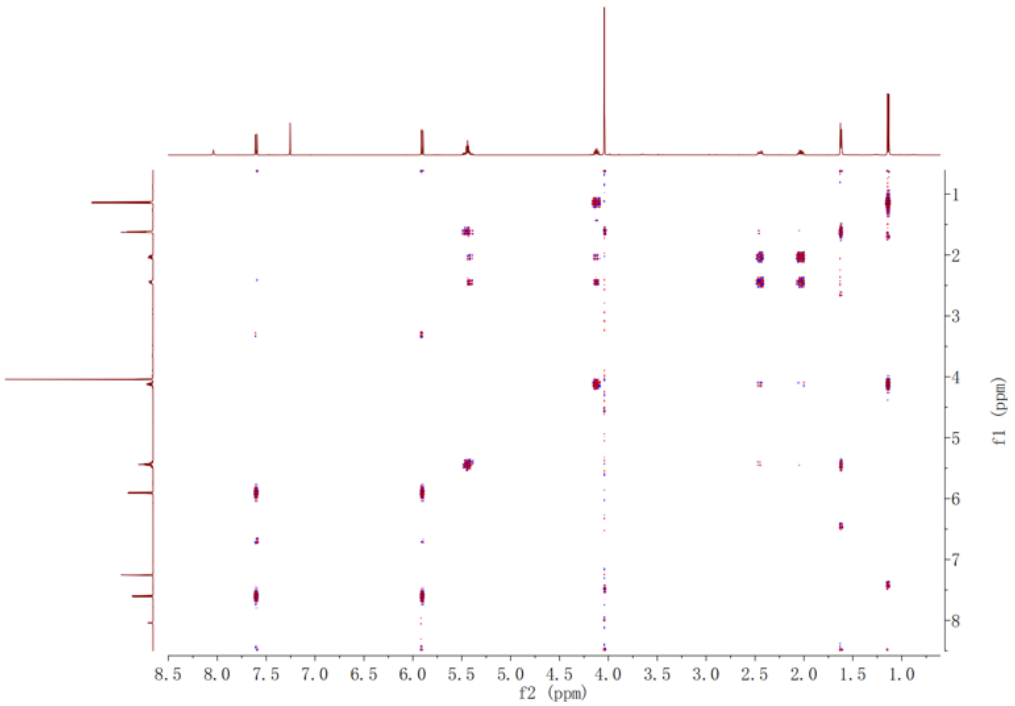


Figure S58. ^{1}H-^{1}H COSY spectrum of compound **6** in $CDCl_3$.

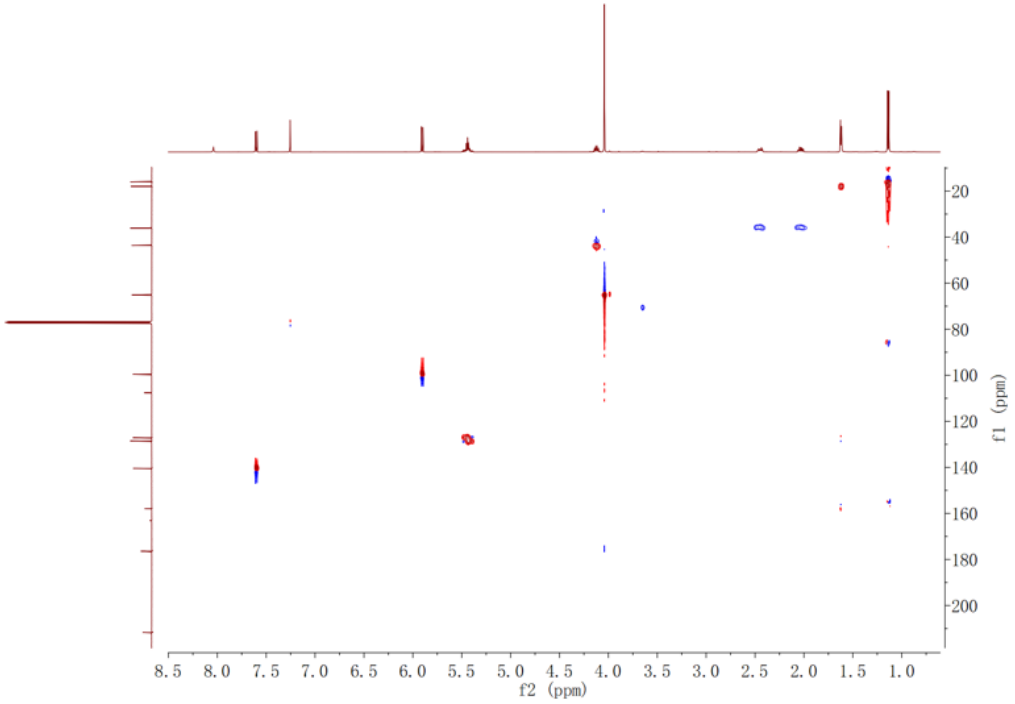


Figure S59. HSQC spectrum of compound **6** in $CDCl_3$.

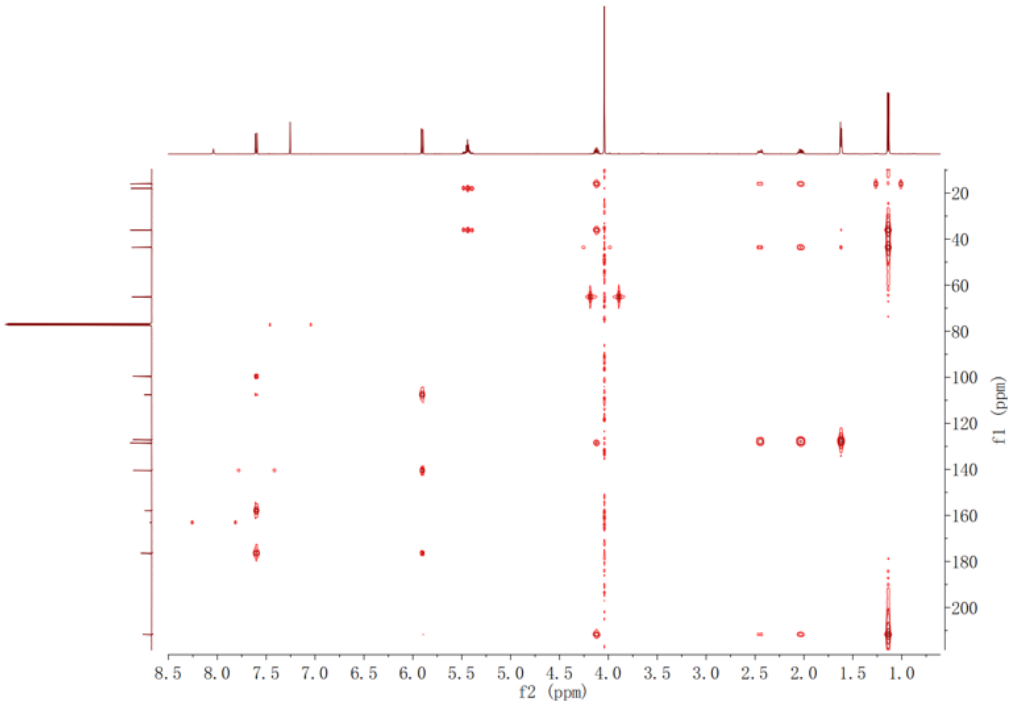


Figure S60. HMBC spectrum of compound **6** in $CDCl_3$.

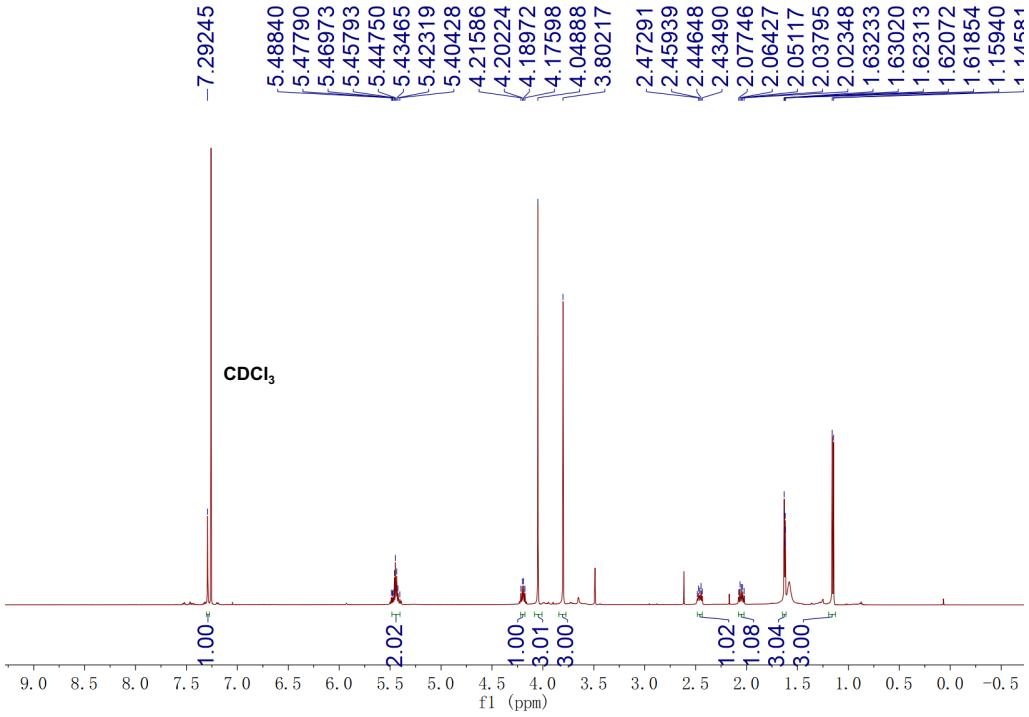


Figure S61. ^{1}H NMR (500 MHz) spectrum of compound **7** in $CDCl_3$.

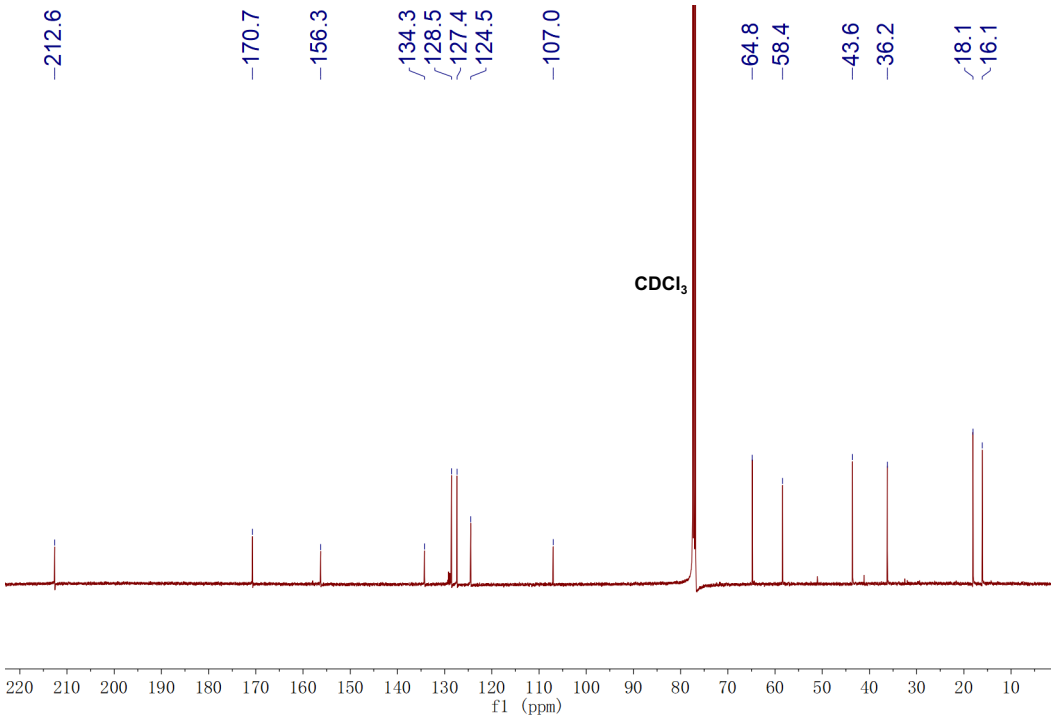


Figure S62. ^{13}C NMR (125 MHz) spectrum of compound **7** in $CDCl_3$.

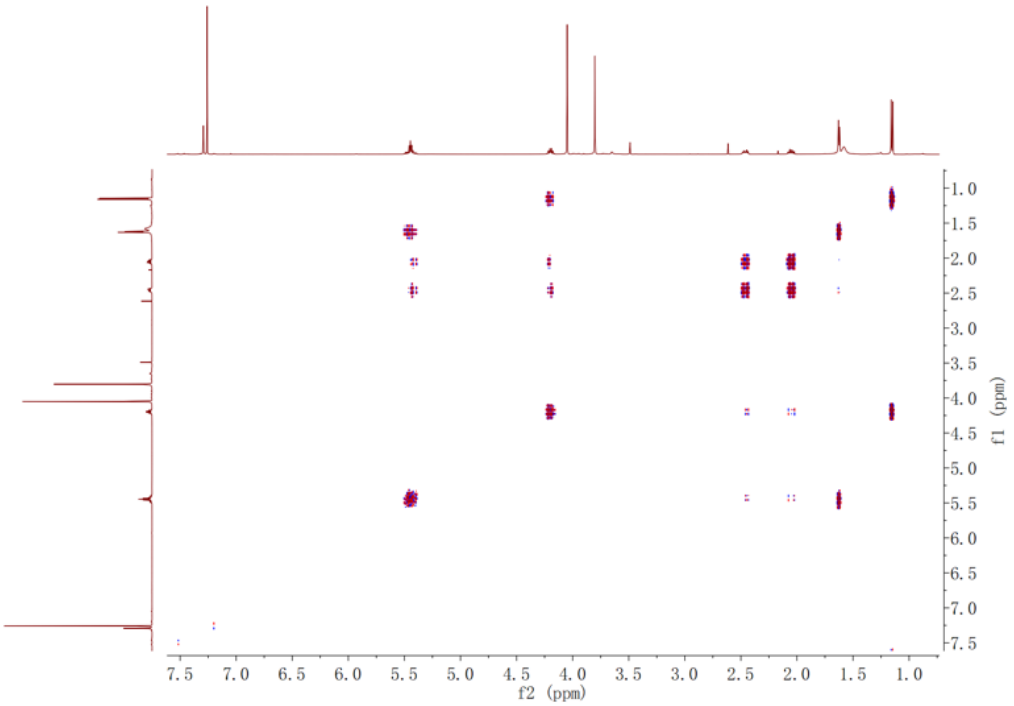


Figure S63. ^{1}H-^{1}H COSY spectrum of compound **7** in $CDCl_3$.

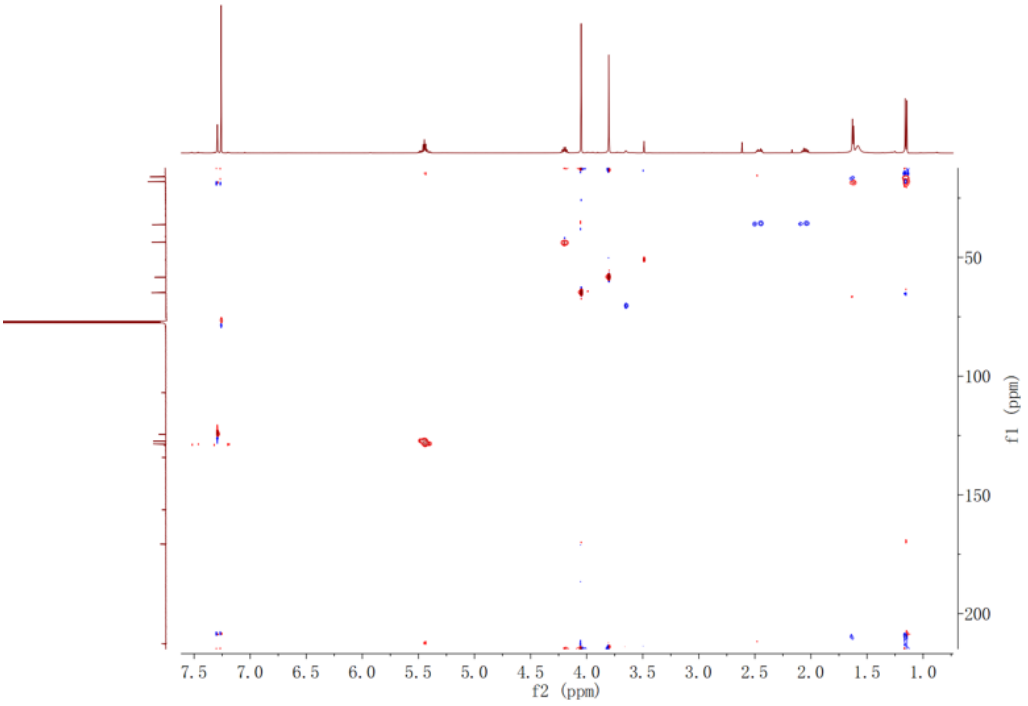


Figure S64. HSQC spectrum of compound **7** in $CDCl_3$.

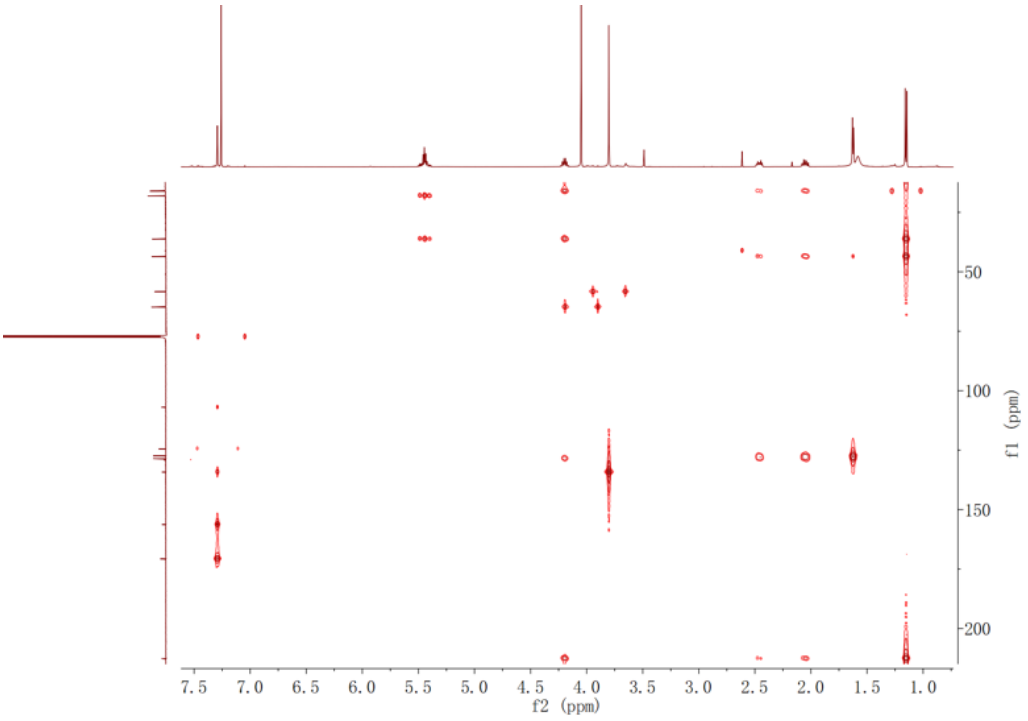


Figure S65. HMBC spectrum of compound **7** in $CDCl_3$.

8 APPENDICES

Appendix A. Gene sequences used in this study

>pDHAD (chloroplast-targetting sequence underlined)

ATGCAAGCCACCATCTTCTCTCCACGCGCCACTCTCTTCCCCTGTAAACCTCTCCTCCCTTCCCACAATGTCAACTCTCGCCGT
CCCTCAATCATCTCCTGCTCCGCCCAATCCGTAACCGCTGATCCTTCACCACCTATCACTGACACCAATAAGCTCAACAAATAC
AGCTCAAGAATCACTGAACCCAAATCCCAAGGTGGGTCTCAAGCAATTCTTCACGGTGTTGGTTTATCAGACGACGATTTGCTA
AAACCCCAGATTGGTATCTCGTCGGTTTGGTATGAAGGGAACACTTGTAACATGCATTTGCTGAAATTGTCTGAGGCTGTGAAA
GAAGGAGTTGAAAATGCAGGAATGGTTGGGTTTAGGTTTAATACTATTGGTGTGAGTGATGCTATTTCGATGGGAACGAGAGGG
ATGTGTTTTAGTTTGCAATCTAGGGATTTGATTGCTGATAGTATTGAGACTGTTATGAGTGCTCAGTGGTATGATGGTAACATA
TCTATTCCCGGTTGTGATAAAAAATATGCCTGGAACTATTATGGCTATGGGAAGGCTTAACAGACCAGGGATTATGGTTTATGGT
GGAACCATTAAGCCTGGCCATTTTCAAGACAAGACATATGACATTGTATCTGCTTTCCAGAGTTATGGAGAATTTGTTAGTGGC
TCCATAAGTGATGAACAGAGAAAGACTGTCCTCCATCATTCATGTCCTGGAGCTGGAGCCTGTGGTGGCATGTACACAGCGAAT
ACAATGGCTTCTGCTATCGAAGCAATGGGAATGTCCCTTCCTTACAGTTCTTCAATACCTGCTGAAGACCCATTGAAGTTGGAC
GAGTGTCGCTTAGCAGGAAAATATCTTCTAGAATTGCTAAAGATGGACTTAAAACCCCGTGACATTATCACTCCAAAATCACTT
CGTAATGCAATGGTTTCCGTCATGGCCCTTGGGGGATCGACAAATGCAGTATTGCATCTAATTGCTATTGCAAGGTCTGTTGGT
TTGGAGTTGACTCTTGACGATTTTCAGAAAGTTAGCGATGCGGTTCCATTTCTAGCAGATCTTAAACCCAGTGGGAAGTATGTT
ATGGAGGATATACATAAGATAGGTGGTACGCCAGCTGTTCTCCGCTATTTATTAGAGCTAGGATTGATGGATGGTGATTGTATG
ACAGTGACAGGGCAGACTTTGGCTCAAAATTTAGAAAACGTCCCAAGCTTAACAGAAGGGCAGGAGATAATTAGACCTCTGTCA
AATCCAATCAAGGAGACTGGACACATACAAATTTTGAGAGGAGACCTTGCACCAGATGGTTCTGTAGCAAAAATTACCGGCAAA
GAGGGCTTATATTTCTCTGGTCCTGCACTTGTATTTGAAGGAGAGGAATCTATGTTGGCTGCTATATCAGCAGACCCAATGAGT
TTTAAGGGAACTGTGGTTGTTATCAGAGGAGAGGGGCCTAAAGGAGGTCCAGGTATGCCTGAGATGTTGACACCGACAAGTGCA
ATAATGGGTGCCGGTCTTGGAAAGGAATGTGCATTACTGACTGATGGTAGGTTCTCTGGTGGGTCACACGGTTTTGTTGTTGGC
CACATTTGCCCTGAAGCTCAGGAGGGTGGTCCGATCGGTCTGATCAAGAACGGAGACATAATCACCATTGACATAGGAAAGAAG
AGAATAGACACACAAGTCTCACCCGAAGAAATGAATGACCGTAGAAAGAAATGGACTGCTCCTGCATACAAGGTTAACCGCGGA
GTCCTATATAAGTACATCAAGAACGTGCAATCGGCTTCAGATGGATGTGTGACTGACGAGTAA

>fDHAD

ATGCTTCTCTCTCAGACCCGGGGCCGTCTGCCCTCCCTGCGGACCCTCAAGCGGTCCATGGCGGTCCGTCCGCTCTCTACCACC
CTCCCCAAGCACAAGGCCGATGAGAAAGAGCCTGTCCTCAACAAAGTCTCCCGCCATGTCACCCAACCCATCTCCCAGGGTGCG
TCCCAGGCCATGCTCTACGCCACCGGTCTCACCGAGGCCGACATGAACAAGGCCCAGGTCGGTATCTCCTCCGTCTGGTACAAT
GGCAACCCCTGCAACATGCACCTCCTCGACCTCAACAACCGTGTGCGGGAGGGTGTCCAGAAGGCTGGCCTGGTCGGGTTCCAG
TTCAACACCGTCGGAGTCAGCGACGCGATCAGCATGGGTACCTCCGGTATGCGCTTCTCGCTGCAGAGCCGTGATCTCATTGCC
GACTCCATTGAGACTGTCATGGGTGGTCAGTGGTACGACGCCAACATCAGTATCCCTGGTTGCGACAAGAACATGCCCGGTGTC
CTGATGGCCATGGGCAGAGTCAACCGTCCCAGTCTGATGGTTTACGGTGGTTCCATTAAGCCCGGCTGCGCTTCGATGCAGAAC
AACGCCGATATCGATATCGTCTCCGCCTTCCAGGCGTACGGTCAGTTCATCAGCGGAGAGATCACCGAGCCCCAGCGTTTCGAC
ATCATCCGCAACGCCTGCCCCGGAGGTGGCGCGTGCGGTGGTATGTACACCGCCAACACCATGGCCACCGCCATCGAGGTCATG
GGTATGACCCTTCCTGGTTCCTCCTCCAACCCCGCCGAGTCGGCGGCCAAGAACATCGAGTGTCTGGCGGCCGGTGAGGCCATC
AAGAAGCTTCTGGTGGAAGACATCCGTCCCTCCGATATCCTGACTCGCCAGGCCTTCGAGAATGCCATGATTGTTGTGAACATT
ACCGGTGGTTCGACCAATGCGGTGCTGCACCTTATCGCCATTGCCGATTCGGTCGGCATCAAGCTAGATATTGAGGACTTCCAG
AAGGTCTCGGACCGTGTGCCGTTCCTGGCTGACCTCAAGCCTTCCGGCAAGTACGTCATGGCGGACCTGCACAAGATCGGTGGT
ACTCCTTCTCTGCTCAAGTTCCTGCTTAAGGAGGGCCTGATCGATGGCTCTGGTATGACCGTCACCGGTCAGACCCTGGCCAAG
AACCTGGAGAACGTGCCGGACTTCCCCGAGGACCAGAAGATCATCCGTCCTCTGTCTAACCCCATCAAGGAGACCGGTCACATC
CAGATTCTCCGTGGCTCTCTGGCCCCTGGCGGTAGTGTCGGTAAGATCACTGGTAAGGAGGGTACCGTTTTCACCGGTAAGGCT
CGCGTCTTCGACCACGAGGACGACTTCATCGCTGCGCTCGAGCGCAAGGAGATCAAGAAGGAGGAGAAGACGGTCGTTGTCATC
CGTTATACTGGACCCAAGGGTGGTCCGGGTATGCCTGAAATGCTGAAGCCTTCTTCTGCCCTCATGGGCGCCGGTCTGGGCGGC
TCCTGCGCTCTGATCACCGACGGACGTTTCTCCGGCGGCTCTCACGGATTCCTGATCGGACATATCGTCCCCGAGGCTGCTGTC
GGTGGACCGATTGGTCTGGTTAAGGACGGCGATGTGATCACCATCGACGCCGTCAAGCGTGTGCTGGATCTGGACGTTGACGAG
ACCACTCTGGCCGAGAGAAGAAAGCAGTGGGAGGCCGACAAGGAGGCTGGCCGTCTCCCTCCTACCGGTCTGACTCTGCGTGGT
ACCCTGGGCAAGTATGCTAGAAATGTCAAGGATGCCAGCCACGGCTGCATCACCGATGCTCTTGACTAA

>AstD

ATGTTCGCGTCGAGGATCCGATCGAGAGCTCTGGGGTTGCACCCACGAGCCCGGTTTGAGAATACTCGTCTTCCGGCTTCCACT
ACCGGGCGCCGTTACAAATCCGACGAGACCCTCAACCGAATCTCCTCGAAGATCACACAACCCAAATCTCAGGGTGCCTCCCAA
GCAATGCTCTACGCCACCGGCCTAACAGAGGAAGACATGTCCAAGCCGCAGGTCGGTATCTCCTCGGTGTGGTTCGAGGGTAAC
CCCTGTAACATGCACTTGCACGACCTCTCCGCAATCGTGCGTGATTCCGTGCACCGGGCGGGTCTGGTCCCAATGCGATTCAAC
TCGGTTGGAGTGTCGGATGGAATCAGTATGGGCACAAAGGGAATGAGATACAGCCTGCAGAGTCGGGAGCTGATCGCAGATGGC
ATTGAGACGGTGATGAACGCCCAGTGGTATGATGCGAATGTGTCGCTCCCGGGTTGCGATAAGAACATGCCGGGTGTGTTGATG
GCGATGGGACGCACGAATCGACCTAGTATCATGGTCTATGGCGGCAGTATCAAGCCCGGATGCAGTGCAAAGGGCCAAAAGTTG
GACCTCGTCAGCGCGTTCCAGTCGTATGGACAGTTCATCACCGGCCAGATCGACGAGAAGGAGCGGTTCGATATTATTCGCAAT
GCATGCCCCGGCAGAGGTGCCTGCGGTGGCATGTACACGGCCAATACGCTGGCCACGGCCATTGAGACCATGGGTATGACCGTT
CCGGGTAGCAGTAGTTGCCCGGCGGACGATCCCAAGAAGCTGGTCGAGTGCGAAAACATCGGCGAGGTGGTTAAGACTATGCTC
AGGGAAGATATCAAGCCCAGGGATGTCTTGACACGTCAAGCTTTTGAGAATGCAATGATTGTGGTGAATATCCTGGGTGGCAGC
ACCAACGCCGTTCTGCATCTGATTGCTATCGCCGACTCTGTCGGTATCAAGTTGACTATCGATGACTTCCAGGCCGTGTCTGAC
AAGACACCGTTCCTGGCTGATCTGAAGCCATCTGGAAAATACCTGATGAACGACCTATACAACATCGGTGGCACACCAGCCCTC
CTCAAATATCTTCTGAAGGAGGGCCTAATCGACGGCTCAGGGATTACTGTCACTGGCAAAACAATGAAGGAGAACGTAGCCTCA
TGGCCCGATTTCCCTTCCGACCAGGACATTATCCGCCCCCTCAGCAACCCTATCAAACCATCTGGCCATCTCCAGATCCTTCGC
GGGTCGCTGGCCCCGGGCGGTTCCGTGGGTAAGATTACTGGCAAGGAGGGTCTGCGCTTCGAGGGTACAGCCAAGTGCTACGAC

Appendix A. (continued) Gene sequences used in this study

TACGAAGATGCATTTATCGAGTCCCTCGAGCGGGGCGAGATCAAGAAGGGCGAGAAGACGGTCGTGATTATCCGGTATGAAGGC
CCCAAGGGTGGCCCAGGGATGCCTGAGATGCTCAAGCCCAGCGCGGCCATTATGGGTGCCGGTCTGGGCCAGGACGTTGCGCTT
CTCACGGACGGAAGATTCTCGGGTGGCAGTCACGGATTCTTGATCGGACATATCGTGCCGGAAGCCATGGAGGGCGGCCCGATC
GCGTTGGCACGGGACGGCGACCGGATCGTGATTGATGCTGAGGAAAGAGTGGTTGATCTGGAGATCCCGACTGAAGAGCTCGAG
AAGAGGAGGAAGGAGTGGAAGGCACCCCCGCTCCGATACCAGAAGGGAACTTTGAAGAAGTATTGCACGCTGGTCAGCGATGCC
AGCCATGGTTGTGTCACGGACGGACCGATCTAG

>ILV3

ATGGGCTTGTTAACGAAAGTTGCTACATCTAGACAATTCTCTACAACGAGATGCGTTGCAAAGAAGCTCAACAAGTACTCGT
ATATCATCACTGAACCTAAGGGCCAAGGTGCGTCCCAGGCCATGCTTTATGCCACCGGTTTCAAGAAGGAAGATTTCAAGAA
GCCTCAAGTCGGGGTTGGTTCCTGTTGGTGGTCCGGTAACCCATGTAACATGCATCTATTGGACTTGAATAACAGATGTTCT
CAATCCATTGAAAAAGCGGGTTTGAAAGCTATGCAGTTCAACACCATCGGTGTTTCAGACGGTATCTCTATGGGTACTAAAG
GTATGAGATACTCGTTACAAAGTAGAGAAATCATTGCAGACTCCTTTGAAACCATCATGATGGCACAACACTACGATGCTAA
CATCGCCATCCCATCATGTGACAAAAACATGCCCGGTGTCATGATGGCCATGGGTAGACATAACAGACCTTCCATCATGGTA
TATGGTGGTACTATCTTGCCCGGTCATCCAACATGTGGTTCTTCGAAGATCTCTAAAAACATCGATATCGTCTCTGCGTTCC
AATCCTACGGTGAATATATTTCCAAGCAATTCACTGAAGAAGAAAGAGAAGATGTTGTGGAACATGCATGCCCAGGTCCTGG
TTCTTGTGGTGGTATGTATACTGCCAACACAATGGCTTCTGCCGCTGAAGTGCTAGGTTTGACCATTCCAAACTCCTCTTCC
TTCCCAGCCGTTTCCAAGGAGAAGTTAGCTGAGTGTGACAACATTGGTGAATACATCAAGAAGACAATGGAATTGGGTATTT
TACCTCGTGATATCCTCACAAAAGAGGCTTTTGAAAACGCCATTACTTATGTCGTTGCAACCGGTGGGTCCACTAATGCTGT
TTTGCATTTGGTGGCTGTTGCTCACTCTGCGGGTGTCAAGTTGTCACCAGATGATTTCCAAAGAATCAGTGATACTACACCA
TTGATCGGTGACTTCAAACCTTCTGGTAAATACGTCATGGCCGATTTGATTAACGTTGGTGGTACCCAATCTGTGATTAAGT
ATCTATATGAAAACAACATGTTGCACGGTAACACAATGACTGTTACCGGTGACACTTTGGCAGAACGTGCAAAGAAAGCACC
AAGCCTACCTGAAGGACAAGAGATTATTAAGCCACTCTCCCACCCAATCAAGGCCAACGGTCACTTGCAAATTCTGTACGGT
TCATTGGCACCAGGTGGAGCTGTGGGTAAAATTACCGGTAAGGAAGGTACTTACTTCAAGGGTAGAGCACGTGTGTTCGAAG
AGGAAGGTGCCTTTATTGAAGCCTTGGAAAGAGGTGAAATCAAGAAGGGTGAAAAAACCGTTGTTGTTATCAGATATGAAGG
TCCAAGAGGTGCACCAGGTATGCCTGAAATGCTAAAGCCTTCCTCTGCTCTGATGGGTTACGGTTTGGGTAAAGATGTTGCA
TTGTTGACTGATGGTAGATTCTCTGGTGGTTCTCACGGGTTCTTAATCGGCCACATTGTTCCCGAAGCCGCTGAAGGTGGTC
CTATCGGGTTGGTCAGAGACGGCGATGAGATTATCATTGATGCTGATAATAACAAGATTGACCTATTAGTCTCTGATAAGGA
AATGGCTCAACGTAAACAAAGTTGGGTTGCACCTCCACCTCGTTACACAAGAGGTACTCTATCCAAGTATGCTAAGTTGGTT
TCCAACGCTTCCAACGGTTGTGTTTTAGATGCTTGA

>HarA

ATGGGCTTGTTAACGAAAGTTGCTACATCTAGACAATTCTCTACAACGAGATGCGTTGCAAAGAAGCTCAACAAGTACTCGTAT
ATCATCACTGAACCTAAGGGCCAAGGTGCGTCCCAGGCCATGCTTTATGCCACCGGTTTCAAGAAGGAAGATTTCAAGAAGCCT
CAAGTCGGGGTTGGTTCCTGTTGGTGGTCCGGTAACCCATGTAACATGCATCTATTGGACTTGAATAACAGATGTTCTCAATCC
ATTGAAAAAGCGGGTTTGAAAGCTATGCAGTTCAACACCATCGGTGTTTCAGACGGTATCTCTATGGGTACTAAAGGTATGAGA
TACTCGTTACAAAGTAGAGAAATCATTGCAGACTCCTTTGAAACCATCATGATGGCACAACACTACGATGCTAACATCGCCATC
CCATCATGTGACAAAAACATGCCCGGTGTCATGATGGCCATGGGTAGACATAACAGACCTTCCATCATGGTATATGGTGGTACT
ATCTTGCCCGGTCATCCAACATGTGGTTCTTCGAAGATCTCTAAAAACATCGATATCGTCTCTGCGTTCCAATCCTACGGTGAA
TATATTTCCAAGCAATTCACTGAAGAAGAAAGAGAAGATGTTGTGGAACATGCATGCCCAGGTCCTGGTTCTTGTGGTGGTATG
TATACTGCCAACACAATGGCTTCTGCCGCTGAAGTGCTAGGTTTGACCATTCCAAACTCCTCTTCCTTCCCAGCCGTTTCCAAG
GAGAAGTTAGCTGAGTGTGACAACATTGGTGAATACATCAAGAAGACAATGGAATTGGGTATTTTACCTCGTGATATCCTCACA
AAAGAGGCTTTTGAAAACGCCATTACTTATGTCGTTGCAACCGGTGGGTCCACTAATGCTGTTTTGCATTTGGTGGCTGTTGCT
CACTCTGCGGGTGTCAAGTTGTCACCAGATGATTTCCAAAGAATCAGTGATACTACACCATTGATCGGTGACTTCAAACCTTCT
GGTAAATACGTCATGGCCGATTTGATTAACGTTGGTGGTACCCAATCTGTGATTAAGTATCTATATGAAAACAACATGTTGCAC
GGTAACACAATGACTGTTACCGGTGACACTTTGGCAGAACGTGCAAAGAAAGCACCAAGCCTACCTGAAGGACAAGAGATTATT
AAGCCACTCTCCCACCCAATCAAGGCCAACGGTCACTTGCAAATTCTGTACGGTTCATTGGCACCAGGTGGAGCTGTGGGTAAA
ATTACCGGTAAGGAAGGTACTTACTTCAAGGGTAGAGCACGTGTGTTCGAAGAGGAAGGTGCCTTTATTGAAGCCTTGGAAAGA
GGTGAAATCAAGAAGGGTGAAAAAACCGTTGTTGTTATCAGATATGAAGGTCCAAGAGGTGCACCAGGTATGCCTGAAATGCTA
AAGCCTTCCTCTGCTCTGATGGGTTACGGTTTGGGTAAAGATGTTGCATTGTTGACTGATGGTAGATTCTCTGGTGGTTCTCAC
GGGTTCTTAATCGGCCACATTGTTCCCGAAGCCGCTGAAGGTGGTCCTATCGGGTTGGTCAGAGACGGCGATGAGATTATCATT
GATGCTGATAATAACAAGATTGACCTATTAGTCTCTGATAAGGAAATGGCTCAACGTAAACAAAGTTGGGTTGCACCTCCACCT
CGTTACACAAGAGGTACTCTATCCAAGTATGCTAAGTTGGTTTCCAACGCTTCCAACGGTTGTGTTTTAGATGCTTGATCTGAT
CTCACCCACACACAAATGATGACAGTGTCTAGCCTTCTTGAATCGTCACCTGAGAGTTTACCAAGTTTCTTCATTATTGACTGG
AGTGACTCGGGCATACTCAACGTAGCTCTCAAACCACTGGATGTGTCCAAAGCATTCTCGAGCAATAAGACCTATCTCTTGGTT
GGCATGACCAAAGAACTTGGTCTTTCTCTAGCCCAGTGGATGGTCAAGAATGGCGCACGTCATCTTGCGCTCACCAGCCGAAGT
GGTAGCGTAGAGGACCCTGGATGGGTTCCAGCACTGAAGGCACACGGAGCAAATGTGCAACTCTTCAGCATGGACGTCGCGGAT
CTGGAGTCAGTTACCGAAGTCGTCGAGACTATTAACAAGACCATGCCGCCTATTGCAGGAGTGTGCAACGCGGCAATGGTTCTG
TCAGATGGTTTGTTCGCAGACATGAGCTTTGATGTATTGCAAAACACGTTGAAGCCAAAGGTAGATGGAAGCCGAAACTTGGAC
CAAGTATTCTACGATACACCTCTCGACTTCTTCGTCTTGTTCTCTTCTATGGCGAGCATCATCGGAAATCCTGGACAGTCAAAC
TACCACGCCGCCAATCTCTACATGGAAGGACTTTGCGGTCAGCGCCGCAGCCGCGGACTTGCAGCCTCAGCTATGCATATAGGA
ATGGTGGCCGATATTGGCTACGTTGCCAGGAGAGGCCAATCCATGGAAGACCATCTGCGCAAGATGTTCTTCTTACCTCTGTCA
GAATCTGACGTACATCAGCTATTTGCCGAGAGCATTTTGGCAAGCCAGGACACTGTTTACGGACTGGAACCTTTCGTAGACTCG
CCTGAAGCGCCAAAGCGACCACCATGGGAAGGAAACCCTCGCTTTTCTCATTTCATGCGCAAGGAAGCATCTAGCAACGAAGTT
TCCAAGGATCGTGCCCTAGACACAGATTACAAGCAGCGCCTGCAAAGCGATGAATCGGAAGATTCTCTAATTGAAATGGTTCAG
CAAATATTCGGACGGAAGCTAGAGGCCATGATGCAGCTGTCACCCAACAGTATCAATCTCAATGTGCCTTTGATTGACTTGGGT
GTTGATTCCTTGTTGGCCGTCGAAATCAGAAGCTGGTTTTTGAAGAAGCTTGGCATGGATGTCCCAGTCTTGAAGGTGCTTTCA

Appendix A. (continued) Gene sequences used in this study

GGTGATACAACAGCGCAGTTGTGCGAGGACGCAACCAGAAGATTCCTGGGCCTCAAGCTGAAACAGAATTCAGCGTCCAAACCG
GAGCAGCAGCCAGAAGTTCTGAAAGTTAAAACAAAAGCGCCAATCTCAGAGAGCAGCAGCAAATTTCTTTCAATAGAAACACCA
GCCTTACAAGCCAGTACCGCATCCAATTCGAGCGGCCACGAAGACGATTCCGCACACTCATCGGGTGACTCTATCAGCGTCCCG
GACATTCATACTCCTAATGAAGAGTGCTCCCAAAATTCATTCAACTCGAGTTCGACCAACCTTGTTGAGTACGAGGCTGAAGTA
CTGGAATTTGGACTTAACCAGACTATTAAAGAAGCTGAGATTATGTCTCATGCTCAATCTCGACTGTACGTCGCGCAGAATTAC
GTCAATGACCCAACCACCTACAACGTCACAGTCACCTATCAAATCCGAGGAAATCTCCAGATCCCAAGATTCAAGCAGGCACTC
AACTCAGTGATGACTCACCACGAATCGTTGCGAACATGCTTCTTCAAAGATGATAATACTGGTCTTCTCAAACAGGGTGTCTTG
CCATCTCCGTCTTACAACCTCAGAATCGTGCAATCTAAGCAAGACGATGCGGTAGACCGAGAGATTGCCTTGTTCAAAGAAAAA
CACTGGGACTTGGAACATGGAAGAACGTTTGGAGTCTCTTTGATCGTTCAATCGCCGGAAGTGAGCACAGTCATCTTCAGCTAT
CACCACTTGGTAATGGATGGCGTGAGCTGGCATCTCTTCCTGCGGGATCTTAGCACAGCGTACCAGCTTCGTCCATTAGTAAAA
GCCTCGAGTAGCTACATCCATTTTGCAAAGAAACAGCAAGCCGCCATGGAAAGTGGCGATTTTGCGGGAGATTTGCTATTCTGG
AAGCAACTCCATGACCAGCCAGCAGAGCCTCTACAACTGCTGCCCATCACCAGTGTTGTTGCCCGTAAACCGCTGACCCAATAC
TCCAGCCACGTTCAATCCCGAGTCATCGATTCAGAGCTGGTGTCTCGAATCAAACAAGCTAGTAAAGATTTGAGAGTAACTCCG
TTCCACTTTCATCTTAGCATAGTTCAACTTATCCTGTCACGATTCTTGGCACTAGACGACATTTGCATTGGCGTAACTGATGCC
AACCGTTTGGATGAAGACTTTTCAGAGACTGTTGGGTTCTTTTTGAACCTTTTGCCATTGCGTTTCCAGGTGGATCAAAATAGC
ACCTTTGGTCAACTTGCCGCCAATACGTCTCGCAAAGCTCTCGAGGGCCTGGCCCATTCGGCTGTCCCCTTTGATGTTATTCTC
GATCACCTCGACGTTGCCAGATCACCAGCCTACAGCCCTCTTTTCCAAGTGGCTGTCAACTATCGGGTTGGAGCTTTGTTGAAG
ACATCCATGGGTGATTGCGAAATCGAACTCACCCAGGTCGAGGACGCTAAGAACCCCTATGATATCAGCTTTGGAGTCACTGAG
ATAACATCTGGCACCTGTCTCATTGAGCTCACATGCCAGAATGCACTCTACACTGTGGAGGATTCGCAGACGTTGATGGAGATT
TACATCCACTTACTCGAGACACTGTCCTCAAACACAGACTTGAAGGTGAAAGATTGCCCTTTGTTCGACAGTAATGCCATCTCT
CAGGCAATTGAAGTGGGCAAAGGACCAGATGTCTCTTACGACTGGCCTCCAACACTTTCTGCGAGATTCAGTGAGGTGCAGAGA
GCGTACCTCAATGACATTGCAGTCAAAGATGGATTCAGGGATGCTTCCTACTCTCAGTTATCGGACGAGGCGAATGGCATTACA
CAGCTCATTCTCGCACAAGGCGGATCTCCTGGTGCCCATATCGCAGTTCTATGCGAACCATCCGTGGATTCGGTCGCTTCCATG
CTTGCCATCCTTCAGGCGGGATGCGTCTACATCCCACTAGATCTGAGTCTTCCAGCCGCCCGTCACATATCAATCCTGAATGAT
TGTGACCCTGCGATTGTCATCGTTCACGAAGCTACCAATGAAGCCGGTATGCAATTGATTGCGGCATCGCAAACCAAGGCTCAA
TTGCTCAACGTGTCTGATATAAAAGCAAAGGATGTCCAAGTTTCGATTCAAGCACAGGCAGATGCACCCGCATTCATTCTGTAC
ACCAGTGGCTCCACGGGCAAACCAAAGGGAGTTGTCCTCTCCCAGGCCAACTTCGCCAATCACTTGGCCGTCAAGACCCAAGAG
CTTCTCATTGAGCGCGAAGTGGTCCTTCAACAGAGCTCTCTGGGCTTTGACATGTCCATCATCCAAACATTCGTAGCTCTTGCC
AACGGCGGTGAACTAGTCATCGCCCCACGCGAGAAACGTGGCGATCCTATTGCACTTTCTGAGCTGATGCAAAAGGAGGGCGTT
ACGTTCACAATTGCTACACCTACTGAATATCTCATGCTACTTCGAAATGGTGAAGAGTATCTCAAAAAATGCACTTCGTGGCAA
CAAGCGTGCATGGGAGGTGAGGTGGTGTCATCACAGCTACTCCATCAGTTCAGATCCTCTCAAATCTCAGTGAAGCAGCTTACA
AACTGCTATGGACCAACTGAAATCACCGCGGCAGCAACTTTCCAAGACTTGTCCCCAGCACTAACAGATGCCTCATTTCAAGTA
ACTGAAGGCTTGATCGGCAAGGTTTTGCCCAATTACTCCGTCAAGATTCTCGATTCCGACGGCAAAGCAGCACCACTTGGTGTG
TGTGGCGAGATCGCCATTGGTGGCCAAGGTGTCGCATCGGGATATCTTCATCTACCTGAAATGACGGCAGAGAACTTTGTACTT
GACTCAAGCAGTACTCAAAAGGATGCCAAGATGTACAGGACAGGCGACAAGGGCAGACTACGCCGCGACGGCTCACTAGAATTC
CTGGGCAGACTCGACCAAGATTCTCAAATCAAGCTTCGCGGTCTTCGAATCGAACTTGGCGAAATTGAACATGTTCTCGTCCAG
GCTGGAGAAGGCCTATTCTCCGATGTTGTGGTCACTGTCAAAGGTGATCCCGCCGTTCTAGTCGCCTACGTTGTTCTCAACAAG
GACAGGGCAGCAGATCGAGCAACGTTTCAAGCTATCCTTCCGAAATTGCCTCTTCCACAATACATGATCCCTGCTGCTATTCAA
GTTCTGGAACGGTTACCAACCAATGCCAGCGGCAAAACAGACAGGAAAGCGGTAGCGAGCCTCGACTTCGTCCTGGTAGCAGAG
GGCAGCCATTCTAACGAGCGTCTGGACTTGTCCGAGGGAGAATTACTGCTGCTCTGGCAAAAGGTTCTTCTTAATACCAACGAA
CCTCACACCTTGGACCGTAACTCGGACTTCTTCTTGCACGGCGGAAACTCGTTGCTACTCATGAAGCTTCAAGGTGCTATAAAG
GACTCCATGGATATCTCAGTCCCCATCAGCGAGCTATACCAAGTCAGCACGCTTGGAATGATGGCCGCGCATCTCCGCAACAAA
CGAGACGAACTGGCTACCGAGTTCGAACCTATTGACTGGGAGCTGGAAACTGCCATCCCTGATCACATCTTGGCTATCTCACAG
GACTCGCAACCATCAACTCCATTGAGTGCGGGGGGGCATGAGATTGTACTCACAGGATCGACGAACTTCCTCGGAGGAGCAGTC
CTGAAGGCTTTATTGAAGGAGCCCACAGTCAAGAAAGTTCACTGCATCACCGTCCCCAGTGACCAGAGCAACAAGCTTCCTGTT
CATCCTAAAGTTGTCCACTACCAAGGCACCCTTTTGTCACCCACTCTCGGACTCTCAGAGACAGAGCGCTTGAAGCTGCAAACA
TCGGCTACTTGCATCATCCATGCCGGTGCAAATGGGCACTGCTTAAACAACTACTCATCTTTACGTGTGCCCAACTTGCATTCC
ACACACTTCCTAGCCTCATTGGCTCTACCGCGAGGCATTCCCATCCACTTCATCTCATCAAATCGAGTGGCTCTCCTCTCAGGA
AGCATGCAACCGGCGGCCGAGTCCATGTCTAAATTCACTCCCCGTACAGATGGCGCAGAAGGCTTTACAGTGGCTAAGTGGGCG
AGCGAGCGGTTCCTCGAAACCCTCGCTGAGAAAGCCCATCTCCCCGTCACAATCCACAGACACTGCATCCTAACGGGTGATGAA
GCACCAAATGAAGATGCTGTGAACGGTGTCATGAAATATTCTTCATTGACTCGTACAGTTCCGCTTTTTGCGAACTTTGATGGA
TTCTTTGACCTCAAGGATGTCCATGCCGCCGCCGCTGATATTGCAGCCAATGTTCTGTCAGAAGACAGTACAGACAGAAGCGTG
CGCTTTATCCATCACTCCAGCGGCATAAAAGTCCCTATTTCAGACTTTAAGAAACATTTAGAAAAGCTCTACAGCGGAGTCTTT
GAAGAGGTAACCATGGATGAGTGGATTGGCAAAGTGCTTGAGGCGGGCATCGATCCACTGATTACCAGCTACTTGCAGGCAATG
GCTGAGAAGGAAGAGGTGATTCAATTTCCATTTTTAGGCGTTTCTTGA

>HarB
ATGGCGAACGTTAATGTGAAGGAGCTTATCGCTCAGGTCAATGCGGCCGCTCAGCAGGGCGCATCTGGTGTTGACGACAGTGAG
AGGAAGCAACTGTTGGCCGCAGTTGATAAACTGAGGGCTTCATACGAAACTCCATTTGAAACAACGATAAGAGTCGTCTTCTCA
GTATGTTAATTGTAAAAGGGTAATTTACAAACAACGAAACTAACACGTACATAAGGGTCACCAAGCCATCGCTATTCGCCTTGG
AGTTGACTCAGGACTATTCGATGCCGCCGCTAAGCAGCCCCAAGACAACATCACCGTTAAGCAACTGGCGAAAGATACGGGCGT
CGATGAGTTACTAGCGAGTAAGTGGATTTGCAATCTTCTGTGTATTTTAATGGAGAATCAGAACTAATAGTTTATAGCACGCAT
TGTTCGTTTTTTGACTGCTATGAATGTTTTCAAAGAGCCAGAGCCCGGTCGATTTGTTGCAACACCGCTCTCTGGATCCTTTGT
TTCATCATCGCCTCTTTCAGCAGCCGTTATTCACATGTAAGTCTACTCAATTGCTCATCATATTAAGACGTGCTAAACGCTGTG
TTTACAGCACACACTTTCTCGTCATTGCGTCACAGCTCCCTGAATACTTTGCAGAGAAGGGGTGGAAAAGCCCACATGATGTTT
ATGACGGCCCCTTCCAATTTGCCCTCAATACCAAAGGCACCTATTTCGACTTCTTGGGCTCTAAATCTCATTACCAGAAAGCTT
TCAACACTGTCATGACAATCTCTCATCGTAGAAAAGGTCGCAACTGGTTTACTTTCTTTCCAGTTGAAGAGAAGCTCGGGGGTG
CCAGCCCCGACCAAACTCTGCTTGTCGATGTTGGTGGAAGCCAGGGCGGAGACATCACTGCGTTCCAGAAAGCATACCCCAATC
TGCCTGGAAAGCTCATTCTGCAGGATCTCCCAGTGGTGATTGATGATATTAAGGAACTGCCCGCAGGGATTGAAGCTCAAGGCT
ATGACTTCTTCAAAGAGCAACCTGTCAAGGGCGCCAAGGGCTACTATCTACGAACCGTGCTTCATGATTGGCCCGACAAGCAAG

Appendix A. (continued) Gene sequences used in this study

CTGTTCAGATTCTGGAAAGAGTTAGAGAAGCGATGACACCGGAATCAGTTCTGCTTATCAACGAAACAGTGATTCCAGAGTCCA
ACGCCTCCCTGTCCTCTGCAACTGCGGACTTGATCATGATGGTCTCATTTGGATCACTGGAGCGCACAGAGAAGCAATTTGCAG
AGCTCTTCGCCAAAGCTCAGCTGGAACTTGTAAAGGTCTGGTTCCCCACGGATTTTGAAGTTGCTTCAGATATTTCAGAGCAGG
CTAGCTTGCTGGAGGTTAGGCTTCGAAGATAGCTTGTGGTAGTTGGAGTTCAATGCAATCAAGTGTTACATATTGAGATAAATA
CGATTCACATGCCAGGTTAAAAAGAAAGAATTGCGTATGCTCAGTAAATTCAACGAATAAATGCGATCAAACTTTTTGACGAAT
CTAGTAAATATTTGATAACTCCATCAATCCCCGGATCCTGAGATACGACCCTAGATTTTAAGCGAACCTGTAAATTCGGGATCC
GACGTAAAAAAGGATATCGAAAGATCGATGTAACCATCAAAGCTATTGTTACATATTCGTATGGTCTTAGCCTCACTCAAATAA
CATAATGGGCGCATTGACTACAGTGTCCCACAAGTCAGTTGCACTATCGTAATCGGGAAACTCGAGCCCATGGCGACCGATGGA
GATGTCGAGGCGACTGTCCTGCCCTTCCTCTCTATATCCAGCCTCTAAAACCTTATAACAAAATAATAAACTACTCCTGGCTGC
ATCCATGATGCTTTTATCAATGGGCTTTTCCTCTGAAACTGATTTTATATGTGCTAGTGAACAGGAAAACAACACCAAGGGTTC
CACCGCCGTGTCACCAGCTCGAAGACAGCTTCGTGCGGTATCAATAGCGCATTGGATAGTTTGCGTCAATTCGCTGCGTGAGAG
ACTGCTGATAAGTGGAAATGGGCTCTCCGTAATGTTGTTAATGACATCAGCACAGACATGAGCTGTTGCTTCCCATAAGAAGCT
GATAATGATGCTGGATCCAGATATAATGAGTAATGAAAAGTCATCCTCTGGCATGGCATG

>HarC

ATGGCTGTTCCATCTGGAGATGCTTCATCCCCGAAGCAAGTTCACATCCTGATCATTGGCGCAGGCCTCACGGGCCTTATCCTA
GCTCAAGCTATACGAAACTTCAATCAATCCGAAAGGAGCACTCAATCAAGTGTCAAGTACACTTATTCCATATTTGAGCGAGAC
GACCATGCATTCGCACGTGGTGGTGGGTGGAGTCTGACGATTCACTGGGCTCTGACGGATTTGCGGAACGTTTTACCGCATGAT
ATCATTTCTCGTTTCGAGGACTGCTTGGTCAATCCTGAGGCTGCAGAGAAGGGAATACCTGGTAATTTCCAATATCTGAACCTC
AAGACCGGGGAAAGGCAGGAGGCATGGCCAATACCACTCGGTGCAGCTTCGAGAGTATCAAGAGAGAAATTGTTAGCCCTTTTA
ATGGAAGGACTTGATATTAAGGTAAGGCCTTTGCCTTTTGTGATTGAGTGTTAAGTTTAACATCTTGTCAGTGGTCAAAAGCTT
TGGACACTATCAGCTACCCTTCATCAACAACAGTTACAGCCCGGTTCACCGATGGAACTAGCGAAACAGGCAACCTGATAGTTG
GCTGCGATGGTTCAAGATCTGCAGTTAGACGCAACCTCTGCCCAACTACGTATCAAAGCAACCGACTTCCAGTTCGCCTAATAG
GCATGCGTGTCGAGTTCCCGGTCGAAAGGGTCCAGGTTTGTAAAGAGATTGACGTACACTTCTTCCAAGGTGGCGATCCCGCGA
CCAATGTCTACTTTTGGTTCTCCTTTATCCACTTGCCACGACCAACGGATGATCTCAAAGTTGCTACATGTCAGATTATGCTAA
GTTGGCCGTATCGAGCCGGCTTTCTAGGCCGGGCCGAGCCAACAGACATGCCCACTACAAATGGGGAGCGTCTGGCATGGCTTC
GATCGTTGGCTAGTGATTGGGCAGAGCCATTCCGTGGGATTGCTAACGACATCCCGGATGATACGGAACTCCGGGAAATTGTAC
TAGAAGACTGGCCGCCCCAACCAAACGGCTGGGATAATCACAATGGCACGGTAACTCTCGTGGGAGACGCAGCCCACGGCATGA
CCATGTTCCGCGGCGAAGCGGCGAACCATGGCGTGATTGATGTATCTGTCCTGACCAAGCTCTTGTTCAGCGACGATGTCTGTC
AGGAAAAGGAGAATGCGTTAATTTTAGCCGTCCAGGCTTACGAAGACGAAATGATTGAGCGGACGCGCCCTGCGGTCCTCAAAT
CCAGACAGGCATGCATCGATGCAAACAACTACGAGTCTGTAAATGCGCAAAGCCCATTGATTGCAAAGCGAGTTGTCAGGGATT
AGTTTTATCATAAATTTCGGAAGCTTACGTGTTGGGACATTTCATAGCCGAGGATATTCAAAGGGAAAACGTTGTCAGGTTACA
TTCAGGGGCTAAATATGATTTAGAGATCCGGATAGCAACAGACTAAGCTTTTACTTACCTATGTACCAATGACAGTCTAGCTAC
ATTTGCTTTTCTGCTTTGCCAAGAGGAAGTAACCCTGTGATAAACCGTGTGACTAAGGGGTTGATAAAAAGATTTTATCACTAG
AGAAGATGAAGTAATATTCTTGAGAATAATAGATTATATAGAATTGGGCTAGAGTTAGGCCTTATTCATGTTTATATTATTTGA
CGAATAATATCTCAGAATACATGCAAGAGTCGGCCTGGATACATTGCTCTACCTTTTGTTTATGAACGTTTTCTCTAGTTCCCT
ATTGAAGAAAGGGAAAACAGCTCATTGAAGTGTATAGTAACTCGTATTGTCGTTGAAAGAATTGGTCATACTTCTTCATCTTCT
ACTAATTTATATAGTTTCTAATAAAATAAACTACTGACTCTTCCCTCATTTTCGACAGACTCGACAATCTCAGTTCGGACTGAA
AGTAACTGCAGCTGG

>HarD

ATGGTAAACGCCACTTGGGCTCAAGGCCTCGATGTTGCTGGCCTTTTCACTGAACAAAAATTTTCATCAAGCGTACTTCTTATT
ATTTTTGCATCTATTGTACATTTAATCTTCCTTCGACGAGTATGTTAAACATGAGACCAACTAAGAGAATGACTGCTTACTCGT
CTATAGCCTAGCGTAATTAGAGATAAAGATGGCAAACCGATCAAAGAACTTAAGGCCAATGCTCGCGTGATGAAGTTTGCGCTT
AGGTATGAGAATATTTTCCTGTTGCGCAATCGTGGTGCCAAAGATGTAGATTTAGCTAAACTGTGTACTTCACAGCGCAGAATT
ATCCAAGCAGGGTAGAGAATTGGCTGGCGATGCGCCATACATGATTCACAATGGCCAGTCCAGGGAGCTTGTCATTACGCAGCC
AGATCACTTGAAAGACTTTTACAAGAATGACACTCAACGTAGGAAACTTTCTACCGGTGTATCAGCTACAACTAATTCAGTCAT
CAGATCACCCAAAGCCAGCCAATCTCAATATGGGTGCATATTTTGGGAGGTAGGTTTGCTTCAGTATGAGATGTTGACCTTGTC
CTTGCTAATTCGCTTTTCATGAAGAATTCTGGGCCATGCTGTCGGTGTTCAAGCAGGAGAACGTTGGAAAATAATCCGAAAATA
TTTTGACCCCGAGTTTGCTTTCCGGGTTACCGTCGAGGCTATGCCACGAACGCTTGATCTCGTGAATCAATGGGCCGATCAGCT
CGTTCAACATTCAGCTCCATCATCAAAACCAGCAGCAAGCAAATCATTCACTACTGACGTGCAAGCTGTCGGCAAATTCTTGCC
TTTCAAAATTGTGTCCCAGCAGTTATATGGCCAGGTTTTCAATGAAGAGGTGAGTTTACTGTTTTTAAGCGCTACAACAGCTTC
TCAGTTCTAATACAATATTGACGCACCACTTTGATTAGTTCTTCTCGAAATTACTTGAGATGAACTCATTGCACGAGGAAATTT
TGCATGATGTGTTGCTTAATCAGCGTCTGACTTCAAAATTCTGGAACCTACTGCCAAGCAAAGCATCGAGACGTATGGACGAGT
ATAACCGAAGATGGAAGAAGTTCAACCTGGATGCAGTTGCACTGGCCAGAAAAGTACGCATCTTCTTATCTATCTTCTATCGCT
TACTCACAAGGTACATAAACAGCAAAACCTCGACTGCCCTCTTGAGCGAATTTACCGTGGAGTAGATCAAAGCTATGAGTTGGA
AGAACTAGAGGTATATGAAAACTCAAATATCCAGGTCTGATTGCTTACATATCTCAGTTTCTACACACTGTCGACGAGATTCTA
TACACAAACGTCGAAATTAGTGCTGCTGTTTTGAAGACGCTATTCACGCGCTTGGCTGAAGACCAGGCATTTCAACAGTCATTA
CGATCAGAAATCTTTCTTCAAACAGCTGCCACAGATTCCGATTTGGCCAAGTATGTCGCAAAAACGGATTCTTTGCTTAATTAT
CTTGTCATGGAGAGCATGAGACACACGCCTGCATTCTGTAAGTCCAAGATATTTTCACAACTGGCCATCTTTCTAACTATTATA
GGGTTCTCACTTCCGGAGTGCACAGCTATGGCAAAGACAATTGGCGGATATCAAGTACCAGCAAATACAGCTGTCGTGATTGAT
TCAGGAAGACTCAATAAAGAAGCAGTTACCTGGGGTGCTGATGGTCACGACTTCCGACCACAGCGGTTCAAGGAGATCCCACAA
AGCAAATGCCGATATGGATTCATGAAGTTTGGCGCTGGCGGTGCTTCCGGGAGATGTCTGGGAAAGCACGTTGCTGATATTACT
TTCAAACTGACTGTCATTGCTGTTTTGAAGAAATTCTCGTTAGAGACTCCAAAAGTTGAAATTCTGGCAGCAACACCAGCTGCA
GTTACTTTCAGTCCGAACTGAGATTGTCGAGTCTGTCGAAAATGAGGGAAGAGTCAGTAGTTTATTTTATTAGAAACTATATAA
ATTAGTAGAAGATGAAGAAGTATGACCAATTCTTTCAACGACAATACGAGTTACTATACACTTCAATGAGCTGTTTTCCCTTTC
TTCAATAGGGAACTAGAGAAAACGTTCATAAACAAAAGGTAGAGCAATGTATCCAGGCCGACTCTTGCATGTATTCTGAGATAT
TATTCGTCAAATAATATAAACATGAATAAGGCCTAACTCTAGCCCAATTCTATATAATCTATTATTCTCAAGAATATTACTTCA
TCTTCTCTAGTGATAAAATCTTTTTATCAACCCCTTAGTCACACGGTTTATCACAGGGTTACTTCCTCTTGGCAAAGCAGAAAA

GCAAATGTAGCTAGACTGTCATTGGTACATAGGTAAGTAAAAGCTTAGTCTGTTGCTATCCGGATCTCTAAATCATATTTAGCC
CCTGAATGTAACCTGAC

>HarE
ATGACCATTATCATTCCGGGCAGTCAGCGAGCACTGAAAGTTACCGGCCCTGGACAATTCCAGGTGTCATTGGCCAATGCGATC
CCCACGGTCGCTGATGATGAAGTGTTGGTGAAAGTCGTGGCTGTGGCTATCAATCCTATAGATGGGAAATCGGCCGAGCTCTCG
CCCACCGCAGGGGCAACATCAGGATGCGATTTTGCTGGCTCCGTGGTGCAGCTCGGAAGTAACATCACCAAGCCGCTGAAGTTG
GGAGACAGAGTCTGCGCTTGCATCTTTGGAAATAATCCAGAACGCCTGGACAATGGGGCATTCTCAGAGTTTGTCGCTGTGCCA
GCCGACCTGGTGCTAAAGATTCCTCATTCAATGTCTTACCAAACGGCGGCGACTCTTGGAGTAGCTGTTGCAACTGTAGGAATG
GCGCTTTACTACTCTCTCAAACTGCCATCGCCTCTATCCCTGAAGAGCGAATCCAGATATTTTCTGGTATACGGGGCCAGTACG
GCAATGGGAACCATGGCTATCCAGATGGCTGTGATGTAAGTGCTCTCCCATTTTGCTTTATCGCAACACTCTAACACCGAATTC
AAGGTCTGGGTTCTCTCCCATAGCCATATGTTCTCCTCGGAACTATGATCTAGTCAAGTCGCTCGGCGCTGTCGCTACCTTTGA
CTACCATTCACCCTCTTGCGGCAGCGAAATCCACGATTTCACAAATGACACGCTGGTGTACGCTCTCGATTGTATTGCTGACAC
ATCATCAATGACCATTTGCTACAAGGCCATTGGATCTGCAGGCGGACACTATCTGAGCTTGGATCCGTTCCCTATCCGTTGCCA
TACGCGGAGGAGTATCAAGCCAAACTGGATAAGCACGTTGACCATGTTTAACCAGGCAGTTAATTGGCAAATTGGGTATCGTAA
GGAAGCGAAACCAAAGGACCGGGCGTTTGGGCATGAATGGTTTAAAACAGCGCAGGACTTATTAGACAATGGTCATGTCAAGCC
TCATCCCTTCAGGGAGATGACTGGCGGGCTTGACGGGATTGTCGATGGAATAGTACAAGTACGAAAGTCAGAAGTTTCAGCGAT
GAAGCTGGTTTACACAATTTGAGTTTAAAACGTAGATCTGGAGCAATATATTGGAGCTGGATCATGATTAGAAGTCTTATTAAG
TTGCAATTGAGCTTTGAGCTCTAGTTCAATCATGGCACTCTCAATAAATGCTTATGCGCGCAATAACTCGTATTTTGGTCGACC
GTAAATCTGCTATTGGTCATTCGATCCGTTAAAGCATGAAATAACCATCAAAGTAGCCAAGTCGCAGATCAATATCCGGAAAGA
TAGGGTTTAGCAGGACATGGCAGAGAGACCCCATGACAATAAACAGTGTAGGTACAGTATGGGTAGTATTCTTCAAAGCCAAGA
ATGGCTCGATGGATCAAAGTAACCGAGATTTACGAAGTAGGAATGTGGTTCAGGTCCGGTGCCTGGCAAAACCCCGTCTGTGGC
CGTGCCCGGGACGGGAAATTCAGATGCTTCTCCACAACTCCTCTGCTTAAACGCGATAAATTCTGGGGTTACTCATCAACCTAT
TGACGAACCTGCAAGCTA

>HarF
ATGGACCTCACAAACTCTCGTTCCATGGAGATTTTCCGGAGCCTAGGTATGGCAGAAGAGCTTCGAAAGCTGGCCGTGCCTGCA
AAATTTCCGTTCACTTGTCTCTTCTCCACAGGCCTTAACGACGAGAAGGCGGCATCGTCGTGGGATCTACCGAGCGCAGAGAAA
CTGAACGAGCAGATTCTGGCTCAAAATGATGGTTCACTTCCCTTGGAACCCTGGATGCGTATTTCACAGGTAATTTTTGAATCA
TGGTTAAAGAGCGTGTGTGTTGAGGATCCGCTGGTGGACATTCGTTCTGGGTGGAAAGCAACAAGCGCCAAAGAAAGTGACGAC
GGTGTCGAAGTGACAGCCATTGACGCAAAAACTGGCGAAGAAACAACATTTCATAGCAGATATGCCGTTGGGTGTGACGGTGCA
AATAGCATCGTAAGGGAAAGCTTGGGAATCAAACTCGATGGCGGGCCTCTGTAAGTGAGAATGACCTCAACATCTAATCCTATG
CACTAATAGTTTAATACTTGTAGCCCTGGCCGTGCTCTTTTGGTTCATTTCAAATCCCGCGACTTGACTCGCCTACAGAAGCAG
GGACAGTTTTGGCACATCTTCTTCCCGAAGACGGTTGAGGAAGGTGGTTCGATGAAAGGTGCCATCATTGCACAAGATGAGATC
GACACTTGGACCATTCACAGATTTCTTTCTGCAGACTTTGACGACTCTCAGCTTTCTTCTGAAGATGCAGTGTACAGCACTCTT
GGTGGCGCTGGTGAGCCATTCCCAATCAAGATCGATGAAGTGCTTGTACGATCAACGTGGAATCCCAGCGTGGCAATTGCTCAA
GGTTTCATGGGCCCAAAGCAACGAATTTTCCTAGCTGGAGATGCTTGCCACCAAATGCCGCCCACTGGAGGTTATGGCATGAAC
ACAGGTATTGCTGAAGCATTTGACCTCGGCTGGAAACTCGCTGCGGCTGTTAATGGCTGGGCTGGTCCACAGCTTCTCGCTACC
TACGAGCAAGAACGCCGTTCTGTCGCACTTCTAGCATTACAAACTGCTAAAGGCCATATCGGAAGACTAATGGCAATGCCAAAG
GCAGTCGAGTTTTCCGGTAAAACACTTCAAGCTGGCGGAGAAGAGGCACGCGCCATGATTAGTCGCGTTCACGAATATATTCAA
GCAAACGACGCTCACAACAAGAGTCTGGGTGTAGAGCTGGGCTACAGATACACTTCAAGTATTTGTTTGCCAGACCAACCACAA
GATGAGCTTCCATCACCTCCGGAATTTGATCCGAGGAATTATATACCCACAACTGTCTCTGGTTACCGAGCGCCCCACGTGTTT
CTTCAAGACAAGACGCCAATTTTCGATCATTATGGTAAAGATTTTACTCTGATCGCATTCAAAAACAGCCAGCAACACCTGCAA
GCTATCGACTGGCTTAGACAGTCTGCCGAGCGGCTCAAGGTTCCGCTTCGCATCTCGTCGCTAGATGGAGAGACAACAGCGCAG
AACATCTGGTCAGTAGATCTTGTTCTGGTGCGCCCTGATGGATTTGTTTCTTGGCGCGGGAATTCGCTGGATAAGCTTGAAACT
GCCTATGAAGTTATTGCGAAAGCAATAGGACATGATTAAGGTCTGTTGATAGGTCAAAATTCGCTCCAAGCTTTTTGTAGGGTA
GTTTTGGAAGACTCGTAACTGTAGACAAATCTCTTAAAGTATTGGGGTCATTTATAGCTTCTTATAAATCTTTAAGTGATATGG
TGCTAATAAAGCTGGAATCATCGACCTCATTATTTCCTCGGTTTTACTACGTCTGTCTCACAGCTTTCTTTTATTTCCACCGTC
GATATCTTTTGCTCCCATAACCAATTGCCCCACTGGCTCACTGAGCTAATCTCACAAAACTGCATACATGTAAACATTACATTC
GTTTTATCACAACACTAATACAGCTCTAAAAAAGTTCGCATTTAGAATTTGTAATGTACAACATTTATTCGTGAAGACGCGCTG
ACATTCGCGCCGAAATATCAATTTACGCTCACAAGGGCATCATTTGCCAGTACACTACATTTTCAGCTATATTCTTAAGCCCTG
TAAAGAGCAATCTTAACGCCCGTCTTGCTTGATGG

>HarG
ATGGCTCTGCAATCATGGGGAGTTAACCTGCTCATTGGAGCGGTCGCATTTTGGCTGTTCAGAGGCATCGTCGAAAAAATTCAG
TGGGATCGGAAGAGAAGAGCGAAGCAATGCGGAGAGATTCCAAGCTATCCTCATTTGGATCCCTTCTTTGGTCTCGATGTTGTG
CTCGGAATGGCCAAATCCCTGAAAAATCATTACTTCCTGGATTGGCTTAACTGGGTACACAAGAACCGGCCAAAGACATTTCTC
GTCAACTTTGTTGGCACTCGCTTTATTTGGACTATCGAGCCGGAGAACATGAAGAGCATGTCAGCTTTCACATGGAAGAGCTTT
GGCGTGGGCCCGATGAGACGAAACAACAAGGCGACTCACCCGTTTGCCGATATTGGCGTGAATACTGTGGATGGAAAGGAATGG
GAGTTTGCCCGATTCTTGATTAAGCCGTTCTTTATGCGAGAGAAATACACCGACACGAAAAGACTAGAGAAGCACGTCGATCGC
CTGTTTTCGCTTCTTCCTGAAGAAGATGGTGTGACTGTTGATGTCCAGCCTCTAATCCAACGATGGGTCAGTATAGACGCTATA
CACGCGAATGCGAAAATACTAACCGTCAGCAGTTCCTAGATGTCGTAACTTTCCTTCTTTTCGGAGAGTCGCTCGATTCATTGG
TCTATCCGGACCGCGCTCAAATTTGCTGGGCCATGGTTGACGTCTTGAGAGGATTGCGTCTCCGCTTGCAATGGTACAAGTATC
TGTTCCTGTTCAAACACCAGAATTGGCTGGATGCTGTTTCAGTCGTACACAAATATGTCGATAGCCACATCACCAAAACCTATC
AGGAGATGGAACAGATTGATGAACTTAAATCTAAAGGCGTGGATGTCAAGGATGAGCGCACAGACTTGCTCTGGTGGATGGCTC
GCAATCTGAAAGACCGAGAGGAGCTCCGTTCACAGATGTGCCTGATCTTTGTTCCCAATAACGACACTACATCGATCTTCATTA
GCCACATTATCTGGAACTTGGCACGCCACCCAGACATCTACGCGAAATGCCGTGAGGAGGTCCTCGCTTATGGAGATGGGGAAT
TGACGTTTGAGGCTTTGAGAGGAATGAAGTATCTGAACGGTATTCTGAATGAAAGTAAGTCACTGAATGGTTAACACGATATTT

Appendix A. (continued) Gene sequences used in this study

TGTAAGGGGGAGCTAACGTCAACAAGCACACCGACTGTACCCGAATGGCGTAACACAGGTTCGCAAGTGTCTCGAAGATACAGT
TCTGCCAACAGGCGGCGGTCCAGATGGGAAACAGCCCATTTACGTCCGCAAAGGTGATGTCGTGCAAGTCAACAAGAACGTGCT
TCACAGAGATCCCGAGATCTGGGGTGACGATGTGCTGGAATTCAACCCTGATCGATGGCGAGATTTACGTCCCATGTGGAACTT
TGTTCCCTTCGGTGGAGGCCCTCGAAGATGCCCAGCACAAATGTTGGTCACAGCTGAAGCCAGTTATGTGATTGCACGGATTGC
ACGGAGGTATGCTAGAATCGAATCTCGCGATCCGAGCCCTTATGTCGGAGTTATGCGAGTCGGTCCATCAAGCAAGACGGGCGT
TAAGATTGCTCTTTACAGGGCTTAAGAATATAGCTGAAAATGTAGTGTACTGGCAAATGATGCCCTTGTGAGCGTAAATTGATA
TTTCGGCGCGAATGTCAGCGCGTCTTCACGAATAAATGTTGTACATTACAAATTCTAAATGCGAACTTTTTTAGAGCTGTATTA
GTGTTGTGATAAAACGAATGTAATGTTTACATGTATGCAGTTTTGTGAGATTAGCTCAGTGAGCCAGTGGGGCAATTGGTTATG
GGAGCAAAAGATATCGACGGTGGAAATAAAAGAAAGCTGTGAGACAGACGTAGTAAAACCGAGGAAATAATGAGGTCGATGATT
CCAGCTTTATTAGCACCATATCACTTAAAGATTTATAAGAAGCTATAAATGACCCCAATACTTTAAGAGATTTGTCTACAGTTA
CGAGTCTTCCAAAACTACCCTACAAAAAGCTTGGAGCGAATTTTGACCTATCAACAGACCTTAATCATGTCCTATTGCTTTCGC
AATAACTTCATAGGCAGTTTC

>AscA

ATGGGTGAACCTATTGCCGTGGTAGGATCGGCCTGTCGCTTTCCAGGCGGAGCTACTTCAACCTCAAAGCTGTGGCAGCTGCTA
CGCCAGCCGACCGACGTACTTAGCCATTTCCCCCCAGATCGGCTAAATCTGGATAGGTTCTATCATCCGGATGGCGAACATCAT
GGAAGCACTGATGTTCAAGGCCAGTCCTATCTGCTGGCAGAGGACTGCTCTCTTTTCGATGCCGCCTTCTTCAACATCAATCCG
CTGGAGGCGGATGGCATCGACCCCCAGCAACGAATCCTGCTGGAAACGGTTTACGAGGCGGTGGAGTCTGCGGGGTGCACACTC
AAAGCGCTGCAAGGCTCATCGACATCCGTCCATGTTGGTGTCATGAATGCTGATTACTGGGATCTGCAAATTCGTGACACCGAA
ACCCTGGGAACTCACAACGCTACCGGCACAGCGCGAAGCATCCTTTCAAACCGGATATCATATTTTTTCGACCTCAAGGGACCA
TCCATGACCATCGACACAGCCTGCTCCAGTTCCTTAGTGGGGTTACATCTTGCCGTGCAGGGACTCCGAAACGGGGATGCCACA
GCAGCCATTGTCGCAGGCGCCAACCTGATCCTGGATCCCGCCATGTATATTGCCGAGTCGAGCTTGCACATGCTGTCGCCAGAC
TCACGCTGTCGCATGTGGGACAAAGACGCCAATGGGTACGCTCGCGGAGAAGGCTTTGCTGCGTTGCTGCTCAAGCCCCTGAGC
CGCGCCATTCAAGATGGAGATCAGATCGAAGGCATTATTCGAGAGACCGGGGTCAACTCAGACGGGCGGACCAAGGGTATCACG
ATGCCCAGTGCGGCAGCCCAAACCGCGCTAATCCGGCAGACCTACTGTCGGGCTGGACTCGATCCTGTTTTAGACCGATGCCAA
TATTTCGAGTGCCACGGGACAGGAACGTTAGCAGGAGACCCGGTTGAGGCTCGAGCCATTCGAGACGCCTTTTTCCCAGATGAG
CAGACGGAATTGCGGTCATCGGATCGGCTTTTCGTCGGCTCCATCAAGACAGTCATCGGCCATCTGGAGGGTTGTGCTGGTCTG
GCAGGGGTGTTGAAGGCCCTGTTGGCTATCCGATACGAGACGATTCCCCCTAATATGCACTTCAACGACCTCAATCCATCTGTT
GCTCCTTTCTACGGAAACCTGAAGGTCCCGACTGAATCGACACCCTGGCCCCTCCCGGCAGCACAGGCGCCAAGGCGAGCGAGT
GTCAATAGTTTTGGCTTTGGGGGGACCAATGCCCATGCCATTATTGAAAGCTATCATCCTCCCGCTGCCGAGTCTGGAGAAGCA
CAGCGCCCGGACGGCGAGCGATTCGTCGGGCCCCTAGTCTTCTCTGCACAAACGGGCAACTCTCTTGTGTCCATAGTCAAGAGA
TTCGAGAGGTACATTGATTCAAATCCAGAGCTCGATCTCGAACAACTGGCCTATGCTCTTCAAGCTCGGACAACATTTCCAACA
AGATCATTCTTCTCTGGCTCGACTCGCACGCGCTTGCTCCAGTTCATGAACAAATTTGTCAGCGATGCGGAAGCTGGATCAACG
TCCTCGGCAGGGACGAGAGCACAACTGATCAACGAGAAGGAAATTCCAGGCGTCCTCGGGGTATTTACTGGCCAGGGAGCCCAG
TGGGCATCCATGGGCCGTGGTCTCATCCAATCATCCCCGCTTTTTCGTGCTTCTCTCGAACGCAGCGAGCAGGTGTTGCGCCGC
CTGCCAGACGGTCCCAGCTGGTCTTTGATTCAAGAATTGACAGCCGAGGAGGGCTCGCGCTTACCAGAGGCAGAATTCGCGCAG
CCTCTGTGCACCGCTTTGCAGATTGCGATGGTGGATCTCCTGCGCGCCGCTAAGATACCACTACATGCTGTTGTTGGCCATTCC
TCTGGTGAAATTGCAGCAGTATATGCAGCCGGCATCATTACTGCCGATGCCGCGATGCAGATTGCATACTACCGTGGTTACCAT
GCAAAGCTCGCCGGCAATGCACAGGGCCAGAATGGATCAATGATGGCGGTTGGCACCTCATACGATTCAGCAATGGAATTTTGC
AGTGGACCAGATTTCAGTGGCCGGATTACCGTTGCTGCGAGCAACTCACCATTGAGTGTGACCCTATCAGGAGATACGGATGCC
ATTACGGAGGCCAAGCAACATTTTGACAACAACAAGATCTTCACTCGGATACTTCGTGTCGACAAGGCGTATCATTCTCATCAC
ATGCGCCCTTGTGCGGAGCCCTATCTCCGGTCTCTGGAGGCTTGCAACATCCAAGTACAACGGCCGTGCGCGGACTGCACGTGG
ATTTCCAGTGTTCGCGGAGATACGGATCTTCTCGATGGAGGACTCGATAGCCTAAAAGGACCATACTGGGTTGCGAATCTGCTA
CAGACCGTTCTCTTCTCGCAGGCTATCGAGACCTCCATCTGGAATGGTGGCCCATTTGATGTAGCGCTCGAGATTGGCGCCCAT
CCAGCCTTGAAAGGACCGGTAGAGCAGATCTTCAAGGCATCATTTGGGCGCGCTCCCTACTATGCTGGGCTGATGAGACGCGGC
GACGATGAAGTGGAGGCATTTTCGGGCGCTCTAGGCTTTGTCTGGTCGTACCTGGGCCCTTCGTTCATTGACTTCTGTGCGTAT
GGCATGGCATTCCAGCAGCCTGGCCAATCCATCCAAGGCACACCGAAAGTAATCAAGGATATGCCCAGTTACGCTTGGGACAAC
AAGCGGTACTGGAAAGAGTCGCGCATCTCTCGGCAGTACCGACTGGCGAATAATCGCAGCCAAGAGTTGTTGGGAAGAAGAGTC
GCGGATGATGCCGACTACGACATGCGCTGGAGGAATGTTCTTCACACGTCTGAGATTCCGTGGATCAGAGGTCATGAGTTCCAA
GGACAGGTTTTGTTCCCAGGGGCCGGATACGTGGCCATGGCCTTCGAGGCCGCTCGGGCGATTGCCTCAGGGCGTCCTGTGAAG
CTATATGAGGTCCAAAACGTCGAGCTAAGCCGTGCTATTGTCCTCCCAGATAACGGGGCCGGTGTGGAGACCGTCTTCAACGTC
AAGATCAACAACCAGAAACCCGGGGATACCTCGATCCAGGCTGAGTTTAGTTGCAGCTTCTCGACGGCGGATGGGATGAACACT
CTCACGAAAGCCTGTACAGGCACCCTTGTCATTCACCTAGGAGATTCGACCGGCCCCGAGCTGCCTCCTCGTGCACAACCTCGA
AGCCAGTTAGTTTCGGTTGACATGGATCGGTTCTACACCGTCTTCGAAAAGGTCGGGCTCAATTATCAGGGCCTCTTCCGAGGC
ATGTTGCAAGGGCAGCGTTCGTTAGGGGTAGCCTCCACGACGGCGTCGTGGCTCGAGAGTGACATGAGCAAAGAGTATCTCGTA
CATCCAGCCTTCCTGGATGTTACCTTTCAGTCTCTCTATGTCGCCTTTGCATCACCAGCCAGCGGAGAAATCTGGGCGCCGTAC
TTGCCTGTGCGCATTCATCGTTTGACAGTTGACCCAAACATCATTCGACACACCTTCAGCCCGGGAGAAATCATGATGCATGCC
GAGGCCTTCATAACCCATTCATCCTCGGCCTTGATGGAAGGTGACATCTGTTTGTATCCATCGACCGAAGACGACCGGATGGGT
TTGCAGGTAGAAGGAATCGCGATGCAGGCTGTCACAGACAGGGAGGCGTCCGACGACCGATGCATTTTTGCCACGACCAGTTGG
GGTCGGGACATTGGGTATAGCATTCCTGAATTAAGTCCCCCCATTGATGTGGATGATACCAAAGTGCTTGAAGCTCTCGAGCGC
ACGGCCTTGTATTATTATAAGGATTGCCTGAGGATGGTTCGCCCGGATGATATTTCCAGCTTCAAATGGTATCATCAGCGAATG
TTCGAAGCGGCCGAAACGCTCATCAATTTGATCCAAGATGGACACCACCCAGTGGCGAAATCGGAATGGCTCGCTGACTCTTAT
GACACCATCATGGAACTTGATAAGCTGTACGGGACGAGGGTTGATCTTAGGCTCATCCACGCCGTCGGCCGGAAACTTGTGTCT
GTGGTGCGTGGAGAAACACAACTATTGGAGGTGATGTTGGAAGACAACTTACTGAATCGCTTCTACATGGAGGGTCTGGGCTTT
TCCGTGGTCAACGATCAGATTGCGACTGTCGTCGAGCAGATCACGTTCAAACATCCTCAGGCAAACATCCTTGAAATCGGCGCT
GGGACAGGGGGCACGACTCGAAGCATTCTGGATACCATCAATGAGTCTTACGGTTCTTACACATATACCGACATCTCGAAGGGT
TTCTTCGAGGCAGCAGCGGAAAAGTTTCACGATCATCGTCATAAAATGGTCTTTAAAGCTATGGATATCGAAAAGGATGTTCTG
GAACAAGGCTTCTCGGAACAGTCATACGACATCATCATCGCGGCCAACGTTCTGCATGCCACCCGCAATCTAAACGAAACGATG
CAACAAGCCCGGTCTCTCCTTCGCCCAGGAGGCTTCCTTGTTATGATGGAGATCACAGGACTACAGATCCTTCGAACGCAATTC

Appendix A. (continued) Gene sequences used in this study

ATCATGGGTGGGCTTCCTGGTTGGTGGCTTGGAGCCGATGAAGGCCGTGTTCTGTCGCCAGCGATTTCTGCTGCACAATGGCAT
AAACTTCTTCAACGCACCGGGGTTCTCGGGAATTGATTCGATTATGCATGACATGGCTGACGAGAAAAGGCATTCTTTCTCCCTC
ATCGTCAGCCAGGCTGTGGATGATACGGTCAATCTGCTTCGTGATCCTGTCGTTGCGATCAACAATATCCTCGGCGACGGCGAT
CTTGTGATTGTGGGAGGCAACACACTCCCTGTAATGAAAATTTGTGATCAGCTCCGCACGTACCTTTCACCTTGGACAAAGGTA
ACTTTGGCAGAGAGCATCGATAAGATCTCTGTCTCCAGCCTGACCTTGACCTCGGTTATTTGTCTGGAAGAGCTGGATCGGCCA
TTGTTCGCAGAGCCTATCACTCCCGCCAAGCTTCAGGCCCTGCAGGATCTGTTCAGTCATGCTCGAAATGTCCTCTGGGTTACT
TGTGGACGACAATCAAGGGATCCATATTCAAACATGACCCTTGGTATTGCACGATCTTTATTCGCAGAGCTCCCACAGGCCCAT
ATTCAATTTTTGGATGTGGACACTCATGACAAGTCTCTCGTCAGTGTCCTTTTAAAGACTTACCTACAGCTTCACGTTGCTACG
ACCGAAGAATATGCTGCGCAGAACACCCTTTGGAAGACAGAGCCGGAAGTTATTTGGGATGATGGCCAGCTTTCAATCCCACGA
GTTGTTCCTAATAAGGCTCTCAACAACAGATACAACTCTGCCCGGAGAAAGATTACCATGGAAGTTTCCCCGAGTACAACCAGA
GTGGAATTAGCATCCACGAGCAATGGGATGGAGCTAAGAGAAGCGACCTTTGGTTCAGACTATTCACCTAGCTATCGACGGATT
ACGACCAAATATTCCATCCAGTTGCCAGGGACAGGGGCTTATCTGTGTGCGGGCATCTTGTGTGATACCGGGAGAGAGTGCTTA
GTGACCACTGGTTCCAATGCCTCTATCATTGATGCTGAGGTGGATAACATTATCCTCCTTGAGTCGCATTGCAATCCTCACTTC
TTGCGATCCGTTTCTGACCGTTTGATTGCGGCCTGTCTTCTTCAATGCATCCCTGATGGCCGCATCTTCGTATATGAACCCGAT
GTCAGTCTGCGGGACGCGCTTTCAACAAACAGAAAGTTTCTGTTTGCAACATCTGACTGTTCACGCGGACAGGGCTGGATTCAA
ATCCACCCGTATTTGTCACAACGGGCTCTGAACGCGCTGATTCCTCGGGATGTGACAGCTGTGGTGGACTTAAGCAGGGCTCTA
CCCGACCGGATTCGCTCTTTTGTGGCTCGTGATCGTGCCATTCTCCAGGATGTGCAGTCTCATAAAGTGTTGGATCCTGTTGCA
CTCCAGGATATTGTCGCCGAGATATGTGCACAAGCAGAAGAGTCGGTCGACACACCATCGGTTCACATTAAAGAGTCAGAATCC
CTGTCACCGGGCTCAATGCAGTCCGTGGTCCTCGACTGGACTGGATCTGACCTGGTTGATGCGGTGGTCCGACCTCTTCCCGTG
GCGGGCATGTTCTCTTCAACACTGACATACCTGCTCGTGGGAATGACGGGTGATTTGGGTCTCTCACTCACACGATGGATGGTG
CAGCATGGAGCTCGATCCATTGTTTTGACCAGCCGTGACCCCAAGATCAGCCCACTTTGGCTGGCCGAGATGAAGGCCCTGGGA
GCGGATATTCAAGTCCACCAGATGGACGTGTCTAGCCGTGCTTCTGTCGAAGGGGTTGTCAAGCACATTCGCGCAAACATGTTC
CCGATTGGTGGAGTCTGCAATGCTGCAATGGTGTTGTCCGATCAACTGTTTACCAAAATGAACGTCGACACTTTAACACGTCCT
CTCGGTCCCAAGGTCGACGGCACCCGTCATCTAGATGAGATCTTCAAGGACGAGCCCCTCGAATTCTTTGTTCTTTTCTCCTCC
TTGGCAACGGTCATTGGAAACGCAGGCCAGTCAAACTATCACGCCGCCAATATGTTCATGGAGGCCATCACTACCCAGCGCCGG
GCTCGGGGACTCGCAGCCTCGATTATCCACATCGGGCTCGTGACCGACGTGGGATACGTTGCTCGCCACGGTCTCGCCATGGAG
GAGCATTTGCGCAAGCTCAACTTTATGCCTTTGTCGGAGACCGATTTTCACCATTTCTTTGCAGAAGCTGTCATGGCTGGGAAA
CCAGGGTCTCGCACAGGCTCGGAATTGATTGTTGGTTTGCAACCGGTCAACGACTCCCAAGACGAGAGACCCCCTTGGGAGTCG
AACCCACGATTTTCGCACTACTTCAGCAAGGGACTGACAGCAAAGTCGTCGCAACAGCAAGATTTGGGTGAGGGTGTTGATATC
CGGCAGCAGATCGCAGAGGCAGAAAGCGTGGAGGCCCTGACGGAGATCGTCCAGGCTGCATTTTCGTTACGGCTAGAGGCAATG
ATGCAGATGGTCCCCAACAGTGTGGATATCCGGATTCCCTTGATTGATCTGGGATGCGATTCGCTGCTTGCTGTCGAAATCCGC
ACGTGGTTCTTAAGGGAGCTTGGCATGGATGTCCCGGTGCTCAAAGTGCTTTCCGGGGACTCGGTCGCTCAGATATGTGAAGAT
ACGGCTAGGAAGTTTCTAGGCTCGAAGCTGAATGTGACGCCAAAGGACTTGTCTCAGAAGACCAATGGTGAAGACAAGCACAGC
CAAGCACAATTGACACCTAGTACTTTGCCGAGATATACTGAAGACGCCGAGGAAAACGCTCAGGAAGACGCCCAGGAAGACGCC
CAGGACACTGACACTGATAGTACCTTCAATTCAGAAGAGACCGACTCACGGAGTGCTGATGATGGCATCGAAACGCCTCAATCC
TCCTCCATCGACTTGTCCCAGCAGATTCCGCTCGAGACGGAGCGCATGTCCTATTCTCAGTCCAGGCTCTGGTTCTCGTCCAAG
TCTCTCGATCCCACGACCTGCAATGTCACTGTCTCGTATCAAGTCCGAGAGCGGTTCCAGTTGTCGAGGCTTACACGTGCGTTG
GGCGCAACGATTGCTCACCATCCAGCTCTACGAACACGCTTCTTTGAACAGATCGGTACGGGACAACTGATGCAGCAAGTCATG
TTGTCTCCGCCATTTCTTCCAAAGCATGTCCACAGTGATGACAGGAGTGTGATCCAACGGGAGTTTGACCTTTTGAAAAATCAC
AAATGGGATCTCGAATCGGGGCAGACCTTTGGTGTCACCATTATCCAGCAGGGATCGGATTGCGCGACGGTTATCTTTGGTTAT
CATCATATCATTCTGGACGGCGTCAGCTGGAGCATCTTCCTCAAAGACCTGAGTCGAGCCTATCAGATGCAGGATCTATCCTGC
CCTGGAAACTATCTTCACTTCACCAGGCAAGAGTATACAGTGGCCGAGAGTGGGGACTTTTCGAGCCAACTTGAATTCTGGCGT
CAAGAGCATGAACCACTGTCCGAGGTGATGCCGCTGTTACCATTGGCCAAAGTCGCCAAACGCTCCCCACTGCAGACCTATGAT
TGCCACGTTCAACGCAAGGAAATTGACCAACATCTGGTTGCAAAGATCAGAAACGCCAGTCGGAAACTGGGCGCCACGCCATTC
CACTTCCACCTAGCCGTGATTCAGGTTGTGTTTGCTCTTTATCTCGGAATGGAAGATATGTGCATTGGCATCGCGGATGCAAAT
CGCACTGATGAGGAGTTTGCGGAGACAGTGGGCTTCTTTTTGAATCTCCTACCTCTTCGCCTTCGTCTTGAACACTCCATTCGT
TTCTCGGAGCTGGTCCGGCGGACATCCCGCAAGGTGTACGAATCGTTGTCGAACGCCCAAGTTCCGTTCAGTCTGATGCTTCAG
CACCTTGATGTCCCTCGATCTGCATCACACAGCCCTCTTTTCCAGGTCGCTGTGAACTATCGCATGGGGGCTGTCCTGCAGACG
CCCCTAGGCGACTGCCAACTGGAGTTTCTGAATGCCGAAGATACGCGAAACCCGTACGATATCAGCTTTGGTATCACCGAGACT
GCATCTCGGACTTGCCTTCTTGAGATTACCTGCCAGGATTCTCTGTACTCCACGGACGCGTCCGCGTCACTGATCGACAGCTAT
GTTTATCTTCTTGATGCTCTGTCGGATGATCCATCTGTGTCCATTCAAGATTGTTCGCTGGTGGACTCCGTGACGGCAGACCAG
GCTATCGCTTTGGGTAGAGGTCCTCGGATGACATTCGACTGGCCGACATCGATCATCGAGCGATTCGACCTGATTTCGCAGAAA
TACACAGATCATGTTGCAATTAAAGATGCGTCGGGCGCGTTTACCTACGCTCAGCTGGACGCAAGGGTCGATGAGATTATGCCA
ATAGTTTCCAACTCTACGGGTGAAAGGCGAGTTGCGACTCTCTGTGAGCCATCTGTGGATGCCATCGCCTGTATGCTGGCCATT
CTACGCGCAGGAGGTGTCTATGTTCCGTTGGACACGCGGTTTCCTGCGGCTAGACACTTGAAAATACTTCAGACTTGCGAACCG
AGTGTTGTTCTATATCATTCGGCAACGCATGATCGGTGTCTTGGACTGATGTCCAAGTGCCCGGTCGACTTTATTAACGTTTCT
CATGTGGCCAGAGCCGATCAAGATGTTCACCCGATCCGACGTTATGCTGATCCCGATTCTCCCGCGTTCTTGCTATACACCAGT
GGCTCCACGGGTGTACCCAAAGGTATCATGCTTACACAAGCCAATTTCGCTAACCACCTCGCTTTGAAAACCGAGAGGTTAGCG
ATTGGACAAGAAGTGGTCCTTCAACAAAGCTCTCTGGGATTCGACATGTCCATTGTTCAAACCTTTTGTGCCTTGGCAAACGGT
GGAGCCCTTGTCATCGCGAGTAAAGAGACTCGAGGGGATTCTGTCAAGCTGGCTCAGCTGATTCGGCGGGAGAAAGTGACATTG
ACGATTGCAACCCCAACGGAATACATGATGATGCTGCAATATGCCTCGGAAGAGTTGAAGCAGAGCCCGTTCTGGAAACAGGCA
TGCATGGGAGGCGAATCGGTGACTGACCAGCTGAAACGAGAATTCTGGCGCGTGAAATCTGGTATCACGCTGACGAATTGCTAT
GGCCCTACCGAAATCACTGCTGCTGCATCCTTCCAAACGATCTCCTCGGATAGCAGCCAAGCAAATGTGGTCGGCAAGGCATTA
CCAAATTACTCGATCTACCTCCTCGACCAGCATCGACGGCCCGTTCCAATTGGATGCACCGGTGAGATCTACGTTGGAGGTGCT
GGCGTGGCTCGAGGCTATCTAGGTCTTTCGCAAACCGAGTCTTCCTTCCTTCCGGACCCATTCGCCAGTTCGAGCGATCGTGAA
AACGGATGGACTCGCATGTACCGCACGGGAGACAGAGGACGCTTCCTTTCGGATGGGTCTTTGATTTTTCAGGGTCGCATCAAT
GGCGATAGCCAGGTCAAACTTCGCGGGCTTCGCATTGAGCTGGAAGAAGTCGAGCAGGCACTTTTACAAGCACAACCGGGTTTG
TTCGCAAACGCCGTCGTGACTACTCGCGGCGATCCGGCTTTTCTCCTGGCACATGTGGTCCTCGCGCCTGGGAAGACTCTCAGT
CACGATGAACTTCAGGTGTTTGCTCGCGCCCTGCCCTTGCCAGAGTACATGGTTCCGGCGATGATCGTCCCTGTGGACAGTCTA
CCCACGAATGTTAATGGAAAGGTCGACCGAAAAGCCGTCCAGCTACTTGCACTCCCTGATCGTCCCAATCCGGAATCCCACGCT

TCTACCCGACTAACTCTATCGGAAGGGCAGCTTGTGCTGCTCTGGAAGCGGGTTCTGCCCGAGTCCGTCCTTTCGGTGCTTCCG
CTCGAGCCCGATTCCGATTTCTTCATGCACGGCGGCAACTCTATTCTTCTCATGAAGCTCCAGGGTGCCATTAAGCAAGAGATG
GGCGTTGTTCTATCGATCAATGAACTCTATCAGGCCAGCACGGTCCGCCGGATGGCGGCCTTGATTGCAGAAGACGAACGACTG
CCCCAGCTGGGAAACATTGATTGGGAAGAGGAAACTGCACTTCCAGAAACCCTGCTGAGGCCGCGAGGAAAGGGCCGTAGTTCC
ATGGCCATACAGGGTAAGCATGAAGTCCTTCTGACCGGAGGGACAAGCTTCCTCGGCCGTGCGATGGTGGAAGCATTGGAGCGA
GACCCGACAGTGAAAACCATCCACTGCGTCGCAGTCCCGATCGAACACACTGGTCTATCGCAGTCCCGCAAGGTGGTCTGCTAC
CCAGGCAGCCTAGTCGAGCCCTCGCTGGGTCTGACGGAAGGCGAATGCGAAAAGTTACGATCCTCCATCAGCATGATCATCCAT
GCAGGGGCCAATGGCCATTGTCTGAACAACTACTTTTCTCTCCGCACCCCTAATCTCCATTCGACGCGGTTTCTGGCAGAAATG
GCTTTGCCACGATCTATCCCGGTCCATTTCATCTCCTCCAACCGAGTGGGTCTGTTGTCCGGTCAGACTGCGGTTCCGGCTATA
TCTGTATCGGACCATCCTCCGCGAACCGATGGATCCGAGGGCTTCACCGCGACGAAATGGGCCAGCGAGTGCTTCCTCGAAAAG
GTGTGTCGCCGGACCGGACTTTCGGTCACTGTCCACCGACCCTGCGCTCTGACTGGCGATCGGGCGCCGAGCGAAGACGCGTTG
AATGCCCTTTTGCGGTACTCACTCCTGACGAGCTCGGTTCCGCGATTCGACAATTTTGAGGGATATTTCGACTTTCGCGATGTG
CACCAGGTAGCTACGGAGATTGTCTCTACTGCTCTTGCTTCTCGGTTGTCTGATTCTCAGACGATCGAATTCGTCCATCACTCC
AGTGGGGTGAAAGTGCCGGTCCACCAGTTCCGGCAACATATGGAGCGCCTGCATGGGGGTGCATTCGAGGATGTCAGCGTGGCC
GAGTGGATCGAGCGGGCTCTGCGAGCGGGCATTGATCCGTTGATTACTACGTATCTGGAGGCGATGGTGCAGCGGGGAGAACGC
ATTCAGTTTCCTTACTTGGGGAACCACGGGAATTAG

>AscB

ATGGTTTCTATCAATTTCACGGAGCTAGCCAGCAATATCGAAACCGTTGCGGAGGCTACAGAGATCAAGATCGATGATGAAGCA
CGCAAGCGCCTGCTACAAGCCTGTGACAAACTGAGGACAAGCCTGGAATCCCCATTTGAGTTCACCTTACGAGTCATCTTCGCG
GTATGAGAGTGGACTATCTTTTTGTGGCTGAAAAGGCAAACATCTTCTGACACGATCCACAGGGTCATCAAGCCATGGCACTCC
GGCTAGGCATCGATATGAAACTCTTTGATGCGGTAGCTCAGCACACCAAGTCCGGCCCAGTCACGGTAACGCAATTAGCTGATG
CGAGCCAAGCCGATCCCCTCCTCGTGTGTAAGTTGCAACGCTGCTTGTGATATTAAGATCTGGACAGCATGATCACTGACGTTG
TCGTGTCCCATTTAGCGCGCATCATGCGATTTCTGGCCGCAATGGGAATATTCGATGAGGTGAATCCAGATGTCTACGTCTCCA
CTCAACTGGCAGCCGCTTACGTCTCCGGCTCACCTCTCTCTGCAGCGGTCATACATGTGTAGGGCTCGATCATGGTGATTCACC
AGATTCATTATTTTACTAACAAAACACTAGGACGCATTTCCTCATGATCCTCTCTCAGTTGCCCGCGTACTTCAAGAACAACGG
GTGGAAGAACCCCAACGATGTGTACGATGGTCCTTTTCAATATGCGATGGGAACGACGTCCCACTACTTTGACTTTTTGGCCTC
CGAGCCATATTACCAACAAGCGTTCAACACGGTCATGACTATTTCACACCGGCGACAGGGACAGAACTGGTTCAACTTTTTCCC
CGTGGAAGAGAAAATGGGAGACGCGAAGGACTCCGACGTCCTGTTAGTAGACGTGGGCGGAAGCCAGGGGGGCGACATCACCGC
CTTCCAACAAACATTTCCCCATTTACAGGGCCGACTGGTCCTCCAGGACTTGCCCATCGTGATCAACGCCATGACGGAGTTTCC
GACGGGCATTGAGTGCCAGGGTCATGACTTTTTCGATGAGCAGCCCGTGAAAGGCGCAAAGGCGTACTACCTACGGACGGTGCT
GCACGACTGGCCAGACAAACAGGCGCGACAGATCCTGGCCAAGATTCGGGAGGCCATGGCGCCGGATTCGTTGCTGCTCATCAA
TGAAACCTTGATCCCCGAGTCGAATGTGGCCTTGTCCTCGGCTCAGGCGGACTTGACCATGATGGTCTCTTTTGCGTCCCTGGA
GCGTACAAAGGCGCAGTTTGAGAACTTGCTGAATGAGTCGGGATTCGAGCTGGTGAAGGTGTGGATGCCAGAGGGGCTCACGGC
GAGTTCAGCTGAGCTTTCAAAGCAAGCCACGCTGCTGGAAGCGAGACCGAGGCCGATATAG

>AscC

ATGCCCCATTCCATCCCCCAAGAAAAACACATCCTCATCATCGGCGCCGGCCTCACCGGCCTCATCCTCGCCCAAGCCCTCCGC
CACCTCAACCACCTCCACAGTACCCAACCCCAACCCCAACCCCCCAAATACACCTACACCTACACCATCTACGAACGCGACCCC
TACGCCTTCGCCCGCGGCGCCGGCTGGAGTCTCACCATCCACTGGGCACTCACCGACCTCCGCAATATCCTCCCCCCCGATATC
CTGGCCCGGTTCCATGAGTGTCTGGTCAACCCGGGGGCCGCGGACCGGGGTATTGCGGGCAACTTTCAGTATTTGGATTTGAGG
ACGGGGAGGCCGAAGGAGAAATGGGCCATTCCGTGGGGGGCCGCGAGTCGAGTATCCAGAGAGAAACTACTAGCGTTGTTGATG
GAGGGGGTGGATATTCAGGTATATCCTATCTATTCTTATACCCTGTTATAGGGATTGGTTGATAAGTACTGGTTGCTAATATAT
ACCCCCCCCCCATAAATGGGAGGATGGGTGGATATAGTGGAATAAACACCTTACTTCCATCACCACCACCACCAACCCTACTTCC
ACCACCACCACCAGCACAATAACAGCCCACTTCATCGACCACACCACCGCAACCGGCTCCCTACTCATCGGGTGCGACGGCTCC
CGGTCGACCGTCCGGCGGCATCTAACCCCCACACCCACACTCGCTCTCCCCTCGCGTCTCCCAATCCGCCTCCTCGGGCTGCGC
GTCCCCTACCCAATCACCAAAATCGCCAAATGCGAGGCCCTGGACGTGCATTTCTTTCAGGGCGGAGATCCCCTCACCAATGTC
TATTTCTGGTTCTCGTTTATTTATCTTCCTCGGCCGGAGGATAGGTATCCGGAGGGGAAGGGGGAGACGGAGGCGGTTTGTCAG
ATGATGATGAGTTGGCCGTTTGAGAAGGGGGGGTTGCGTAGGGGTAAGAATAATGTGGGGGAGGAGGAGGAGGAGGGGGAGGGG
ATGGAGGCCCCCGGGGATAATGCGGGGCGGTTGGCGTTGATGAAGGAGTTGGCGAACGAGTGGGCGGAGCCGATGAGGGAGATG
GTGATGGATTTACCGGACGACACGGATGTTAGGGAGATTGTGCTGGAGGATTGGGTTCCTGGGGTTGGGAGATGGGGGAATCAA
GGCGGACGGGCGACGTTGGTGGGTGATGCGGCGCATGGGATGACGATGTGTGAGTTTTTTCTTCCTTTCCTTTCTTTCTTTGGG
GAGGGGGATAGGATCAGGCTAATCATTGGTCTTAGTCCGTGGTGAAGCGGCAAATCACGGCGTGATTGATGTCTCCGTGCTGGT
CAAGTTGTTGGCGGAGAGTGCGAAAGAGGGAAGTGAGTTGGAGCTGGAGGATGTGGTCAATGTCTATGAGAAGGAGATGATTCA
GCGCACTCAGCCGGCTGTGCTCAAGTCCAGACAGGCGTGCTTGGATGCTCATCGGTTTGATAGGGTGGATGGGTCGAGTCCTTT
GATTTCGAGGCGGGCGATGAAGGATTAG

>AscD

ATGACTGTCTTAAACGATGATCGACTCAATCTGGGCCGTGAATACTTAGCCGACCTTAACCTAGTTACACGGATGTCAATTAGC
CTCGTGCTGCTGGTATTCGTGCGGCAGTTGTTCTCTTACTTGACAGTGAATACCGTACGCCGCTCCTCTAAAGCTGCTTCAGCA
AGATACTGACCCTCAAAAGCGGTCCAACACCATCCGAGACTCCCAAGGCCAGCCAATCAAGGAGCTGAAGCAAGATGCTCGTCT
CCTGAAGTTTATTTCCAGGTACGCAACAATCTACCTTTGTCTCCAATACTAACACCAAGTCACAGTTTGGAACTATCAAACATG
GGCAAAGAACTGGCAGACGACAAGCCATATATCATCCAAAATGGCCGATGCCGGGAACTGGTTGTCACTCAGCCAGACCACCTC
AGGGATTTCTACGGAAAAGATACAAAAGGTAGACATGATAACTTGTCCACACTCTTGAGAATAAAAATTCTAATATGCAAACTA
ATCTAGATCATCCCAAGCCTGCGAATCTCAACATGGGCGAATACTTTGGCAGGTAATCTTCTTTTTCTACCTCTGTGACGAGTT
GCCATCTAAAACCCAGAAATTATAGGATTCTCGGTGATGCTGCGGGCGTGCAAGTCGGGGGAACGATGGAAAATCATACGAAAGT

ACTTCGATCCCGAGTTCGCTCATGGCATCAGTATGCAGGCACTCGAGAAGTTCTCATGTCAGATACACGTTTGGGCCGATGCGC
TGCAGACGACCCCAGCGGGGTTTAAAACGTCTCCTGAAAACGACGAGAATCATGGATTTATTGTGGACTTGACCAAATCCATGA
AATTCCTCCCCTTCAAGCTGGTCGCTCTTCAGATGTATGGAGAGGCTTTTACCGAGGAGGTAAATTGGATCTGACCTTCAATCA
CACTGCGATCTTTAGCTAACGGTCGAAGCTGTATGACGAGTTACTCGATATCAACAACCTCCACGTACAGATCCTCCACGACGT
CATCGGAAACAAGAAGCTGGCATCGAAACTGGGCAATTGGCTACCTAGCGCCGCTAAGAACCGAATGGATCTCTATCTGAACAG
ATGGAGGAGCTTCAACCTCGACATCATCCAGCGGGCGAGAGATGTAAACATTTTCTGTTTTTTTTTTTTTTTCTTTTTTCACC
CAAGAAACAAGAGCCTTCTGACCCTCTCTTTACCGTCACAGAACAACCTCTCCTGCCCAGTTGAGAGGATATATCGCGGAGTAG
ACCATAACCGAGAGATGAAACAAACCGAGGTATGACAGCGACACAATCCCAACCCCCCTAAAATATGTTTCTATGGCTCCTGTG
ACAGATGATTCATGATCCCAGTTCCTCCACACCCTCGACGAAATCCTCTTCGCAAACGTAGACGTCAGCTCCGCCGTCCTCAAC
ACCATGTTCAGTCAACTGGCAGCCACTCCAGCCTTCCAAGCCGCTCTTCGAACAGAAATCCTTCAATGGAAGAAAGATACCCCA
GACGACAAAAACATGACCCGGTACATCTCCAAGCAAGACACACTGCTCAATTTCGCCATCATGGAAAGCATGCGATTGACTCCT
GCCTTTTGTGAGCCTCCCCATCCTCCTCCCAAGAAAATAAAAAGGTCTGGTCTCTGACTAGTATACAACCACCGCAGGGTTTTC
CCTTCCCGAATGCACAGCCGCAGCCAAGATGATTGGCGGATACTCCATCCCTCCCAACATGCCCGTGGTGATCGACGCTCGACG
ACTCAACACCGAGTCCGCGACATGGGGAGACGATGGACGGGTGTTCCGTCCCGAGCGATTCTCGACAATATCGCAGAGTCAGTG
CCGATACAGTTTGTTGCGATTTGGCGTTGGCGGGGCGTCGGGGAAATGCATGGGGAGAAATCTTGCCGACTTGATTTTCAAACT
TGCTGTGATTGTGATCTTGCAGAGATATCAATTGCCTTTGGTGGAGGACCAGGCGAAGGAAAAAAGTGGGCAGGTGTCGACAGT
GATTCGCTTCGTTAG

>AscE

ATGTCAATTCCCAACACCCAGACCGCATTGAAAATCAAAGCCCCTGGACAATTCCAGGTGGTCTCCAGCGCGGTGCCGGAGCTC
GACCCCGACGAAATCCTGGTGCGAGTCGTCTGCATTGCGATCAACCCTGTCGACGGGAAATCAGCAGACCTATCACCCACGATA
GGTGCAACATCGGGATGCGATTTCTCGGGGGAGGTCGTCAAGCTAGGCAGTGCTGTCTCCGAAGATAGCCTCCAAGTCGGGGAA
CGAGTCTGTGGATGTGTGTTTGGCAATAACCCGGACCGGCCGGACAACGGAGCTTTCGCCCAATACGTTGCGGTTCCTGCCGAG
CTGGTTTTCAAGATCCCCCCCGGAATGTCATATGAAACCGGGGCCACATTATCCGTGGGGTTGTCCACGGTTGGCATCGCACTG
TACCACACCTGGAGGCTACCATTGCCCTCATCTTCCTCTCAGATTCCGGATGACAGTACTGCCAGCGGCTCACGGACCGTGCTG
GTGTATGGAGGAGGTACAGCGACGGGGACGTTGGCAATTCAGATGCTCCGACGGTAAGTCAAAAAGTGGTCAGTGCACCGGGAC
ATCAAGTGCTAACGCCATGGCAGATCCGGCCTGATCCCCATCACCACATGCTCGCCCCGGAATTTCAGCCGGGTCCAATCTCTG
GGGGCGGCGGCAGCGTTTGATTACGCCTCCCCGACATGTGGCCAAGATATTCGTGAATTCACCCAAGGCACGCTGAGCTTCGCG
CTCGACTGCATCGCCGACCTGGGGTCGATGAAAATTTGCTACGAGGCGATCGGCAGCGACGGCGGCCAGTATCTGAGCCTGGAG
CCATTTCCACTGCGCGGACATACACGGCGAAGTGTGCGACCGACCTGGATCATCTCATTGACCATGTTCAACAAGCCGATTCGG
TGGAAGAGGCCGTACCAGCGAGACGGGAAACCCCGGGACCGAGAATTTGCCGCGCGTTGGTTTCGCCTGGCGCAGCAGATTCTG
GATGAAGGGGAAATCAGCCTCCATCCCCATCGCGTGTCGCGCAGGGGCCTGGGCAGTGTGATTGGCGGGCTGGAGGCGGTGCAC
AAAGGCGAGGTGGCGGGCGTCAAACTAGTGTATTACATTGGCCGCTAAAGGTCGCAATCGGGTCAATCCGTGGTTTTCTGTCGG
GAGCTAGATTTGGCACTAATTAGGAATGCAATCGCCGTACCGTAACTTACAAGCCCCCTCGGCAATGCCCCCGGCCAGGTCATC
CTTCTTTGGCTGCTCTGCCTTTCTCTTCCCAAACACAATTCATTCGCTGAGAAGACCCTATCCATCATGGCGGCGACCCTCCAG
AGAAATGCTCTGCTGTTGCAGAAGGGTGAGTCTCTTGCTTAGTCCCTTATCCCTGTCCCTTAGTCCCTCATCAGTCCCCGTTGA
CTGTCATGTCATTTCGGCTGACAGCCAGACTGTCCCCAGCGTCCCTGTCAGGTCTGCGCCGCGGGGCCGTATCGTCGCCGTCTG
TCTTTCCTCTCGCAGGCGCAGCATGGCAGCAAGTCAGGTATGAATGAGATGCTTTTCGCGTA

>AscF

ATGACTATCGAAGCTTCCTTCGTCGAAGAGCTACCTTCTGACACAGTCCTCATTGTCGGAGGTGGCCCTGTCGGACTCGTTCTG
GCCACAACGCTCGCGCATCATGGAGTCAAGAGCGTCTTGATCGAGCGCAACCTCACGACGACCCGCTGGCCCAAAATGGACTTG
ACGATCGCCCGGAGCATGGAGATCTTTCGCTGCCTGGGCATCGCAAATGGGCTACGTCAACGAGGAGTACCGTCTCATTTCCCC
TTTACCTGCCTGTTCTCTAGCGGCCTTCATGCCGATAGACCTATCACCTCATGGACGCTTCCAAGTGTCGATGAGTTTGCGGCT
CAGATTGCCGCAACCAACGATGGCACGATGCCATTGGAACCCTGGCAGCGCGTCTCGCAAGAGGTTTTCGAGGCTTGGCTGAAG
GAGTTGGGTGAGGAGAACCCGATGATTGACGTCAGATTTGGATGGAAGGCCACAACCGCTCGAGAGCTCGATGATAGCGCCGAA
GTTCAGATAACGGACGCCCGCACTGGGGCGGAGAAGACAATCCGGAGTCGATATGCGGTCGGATGTGATGGTGCGAATAGCGTG
TTGCGAAACAGCTTGGGAATTGAATTGGATGGCGGGCCAATGTAAGCACCTCACCGAATGAAACCGATTAGTTAACTAACTCGG
TCGACTCATCAGCTCCGCATATGCAATCTTGGTGCACTTCAAATCCAAGGACCTTACACGTCTGCGCAAGCAGGGCCAGTTTTG
GCATGTTTTCTTCCCCAACGATGCTGCCTCGGGGGGCTCGGTCAAAGGTGCCATCATTGCCCAGGACGAAATCGATACTTGGAC
GGTGCATCGCTTTGTGCCTGTGGGGTTCGATGACTCCCAGATCACATCTGAAGAGGCAGTATATAGTGTCCTTGGTGGGATGGG
TGAGCCATATCGCATCGACATTGACGAAATCCTAGTCCGTTCTACGTGGCGGCCAAGTATCGCCGTGGCCGAGTCGTATGCAGG
ACCCAAACAACGGGTACTGCTGGCGGGAGATGCCTGTCATCAAACCGTCCCGACCGGTGGATATGGGATGAATATGGGTATCGC
TGAGGCGTTTGATCTCGGCTGGAAGCTAGCGGCCACCATTTCTGGCTGGGGAGGGCCACAACTACTTTCATCCTACGAGGCGGA
CCGACGCCCGGTTGCGCAAATGAGTGTGCAATGGTCCCAGCGGCATATGGGAAATCTTATGGTCTTGTCGAAAAAGTTGCAATT
GGACGCTGACATTGTGGATTCGGATACAGACGCAGGAAAGGACATGAGGAAAACTTGGCACGACTACTGCCAGTTGAATGATGG
ACACAACAAGAGCACCGGCGTGGAGATGGGATACCGATACCAGTCCTCTATTTGCGTTCCGAGTGAGCTGGATTGTGATGGTTC
GGCACCAGAATTTGATGCACGGAGATACATCCCTTCGACGTACCCCGGCTGTCGAGCACCACATGTTTTCCTGAAAGACGGCAC
GCCCATCTTCGATCAATACGGGCGGAAGTTTACCCTGGTGGAATTCCACGATGGAACGGCATCATCTGCGGTGGAATTGTTCCA
ACTCGCTGCTCTCAAGCGCAATGTGCCTTTGAAGACCGTTTCCTTGCACGGAGAAGACCATACGCATCAGATCTGGGGTGCAAA
ATTGGTCCTTGTCCGTCCCGATGGCTTTGTTTCCTGGCACGGAAATGAGATGCAGGACGTGAACACCGCTCTGTTGATCCTCGC
TCAGGCAGTGGGCGACATGTGA

>AscG

Appendix A. (continued) Gene sequences used in this study

ATGACTCTTGTTTGGTATTTGCCCTCGGTTGGCGTCTTCACGGCCATCTTGTGGTGGGCTATCAAGTCAGTCTTCCTGGCCGTG
GACAGCCGGATCAAAAGCAAACGCAAAGGATGCGGAACCGTTCCCAATTATCCCCATCTGGACCCGTTCCTGGGTCTGGACTTG
GTCCTTGGTATGGTAAAGTCACTTCGAGAAAACTACTTTCTGGTCTGGCTCAATCGGATTCACAAGAGCAAGCCCAAGACCTTC
CTGGTCAATTTCGTGGGCTCGCGATTCATCTATACGATTGAGCCTGAAAACATGAAGGCCATGTCGGCCTTCAACTGGCAGGAC
TTTGCAGTCGGGCCGATGCGACGCAACAACAAGGCGACCCATCCCTTTGCCGACAAAGGCGTCAACACGGTGGACGGCAAGGAA
TGGGAATTCAGCCGCTTCCTAATCAAGCCCTTCTTTATGCGTGAGGCCTTCAGCAACACCCGACGCCTTGCGCACCATGTTGAC
CGGATGATGGATCTCTTTCCGGATGACGGCGAGACCTTTGACATCCAACCTCTCATCCAGCGCTGGTTTCTCGACACCACGACC
GAGTTCCTTTTCGGGGAGTCAATGGAGTCTCTGGTCTACCAGGAACGAGCCGACATCTGCTGGACGATGGTTGACATTCTTCGG
GGACTGCGGCTCCGTCTGCAATGGTATAAATACCTGTTCCTGTTCCGACAGCAGCATTGGCTGGATGCCGTCGAAGTGGTACAC
AATTACCTGAACCGGCATATTGATCGAACCTTGGAAGAACAAGCCCAGAAAGCGCAAACCAAGACTATGTCCGAAGGTGAAGAA
AAGGAACGGACCGATCTTCTCTGGTATATGGCGTCACACCTCAAAGACAAAGAGGCTCTCCGCTCGCAGTTGTCTCTCATCTTT
GTGCCCAACAACGATACCACGTCCATCTTCATCAGTCATGTCCTCTGGAATCTCGCCCGTCATCCAGACGTGTGGGACAAGTGT
CGGGTTGAGGTTGAGGCTGTTGGAGAGGATGTCGAACTGACGTTTGAGGTTCTCCGCGGACTGAAATATCTCATCGCGGTTCTT
AATGAAAGTAAGTACACGCAATGAATCCGATGGCCTGTGATGCAGCTAAACACCTCCGTTCTTCTCTAGCACATCGACTGTTCC
CGAATGGCGTCACGCAGGTGCGCAAGTGTATCCGGGACACCACTCTCCCGGTTGGCGGCGGCCCCGACGGGAAATTACCCATCT
TCGTCCGCAAAGGCGACGTTGTCCAAGTAAACAAGAACGTTCTCCATCGGGACCCTGATATCTGGGGTGCCGATGCGGAAGAGT
TCAACCCGGACCGGTGGGAGGGGCTCCGACCCATGTGGAAGTTCGTCCCCTTCGGTGGGGGCCCACGACGGTGTCCCGCCCAGA
TGCTGGTGACCACCGAGGCCAGCTACGTGCTGGCCCGGCTGATGAGGCGCTACAAACGAATCGAGGCCCGAGATCCAAATCCTT
ATGTTGGTGTCATGCGGGTTGGTCCGTCCAACAAGTCCGGGGTGAAGGTTGCCTTGTTCACGGACTAG

>ApnA

ATGCAAGACTTGATTGCTATTGTTGGTTCTGCCTGTCGATTTCCCGGGCAATCAGATAGCCCCTCAAAATTATGGGCTCAATTA
AGGGACCCCGTCGATCTTCGCAAGACATTCTCCTCTGAAAGATTGAACCTTGCTCGATTCTACCACCCCGATGGAGAACATCAC
GGCAGCACAGATGTACAAGGCATCTCGTACTTGCTGTCAGAGGATCCCCGTCAGTTTGATGCCTCCTTCTTCAATATCAACCCC
CGCGAGGCGGAGGGAATGGATCCTCAACAAAGGCTGCTGCTGGAGACGGCCTACGAAGCCCTTGAAGCTGCTGGATATTCACTG
GAGGCCATGAACGGCTCGAAGACTTCTGTTCACGTCGGTGTGATGAACACAGATTTCAGCAACATTCAGCTTCGAGATTCAGAA
GTACTACCGACGTACAACGCCACAGGAACATCAATAAGCATTCTTTCGAATCGACTGTCATACTTTTTCAACTTAAAGGGCCCT
TCAGTGACTGTTGACACGGCTTGCTCTAGCTCATTAGTGGCCCTGCATCAAGCCGTCCAAGGCCTTCACGCCGGTGACGCAAGC
TCTGCCATTGTCGCCGGTGCCAATTTGATTTTTGATCCTGCCATGTATATCGCGGAATCAAGCTTACACATGCTATCCCCGGAC
TCATGCTCGAGGATGTGGGATAAAGATGCCAATGGATATGCTCGCGGAGAAGGTTTTGGAGTATTGGTTCTCAAGCCACTGAGC
CGAGCCATCATGGATGGGGATCATATCGAGGCTGTCATCCGCAGCACAGGCGTCAACTCCGATGGCCGTACCAAGGGCATCACG
ATGCCCAATGCTGCATCTCAAACCGAGCTCATCCGACAAACATACCGAAATGCAGGTCTCGACCCTGTCGTTGATCGATGCCAG
TATTTCGAATGCCACGGAACAGGTACCGCGGCAGGAGATCCGATAGAAGCCCGCGCGGTCCACGACGCATTCTTTCCCGCCGAA
ACAAACACGGCTTCAGATCCACCTATGCCCGACGGTAAGCTGTACGTCGGATCCGTGAAGACAATTATTGGCCATCTCGAAGGT
TGTGCGGGCATTGCCGGAGTCCTCAAAGCTGTCCTCGCCATCAAGAATCGTGTCATTCCTCCCAATTTGCATTTTCATGAGCCA
AATCCTCAGGTCCTTCCATTTTGCGCCCGTCTTGAAATTCCAACAGTCCCCATACCATGGCCGGATACTGGCAGAGCTCCCATG
CGCGCAAGTGTCAACAGCTTTGGATTTGGTGGCACCAACGCCCACGCCATCATCGAAGGATACGATGCATTGAGCAACCCAGTG
AGAGAGACAACGATATCGCCTGATGATCGATTCATCGGACCACTCTTGTTTTCTGCCCACTCCAGCACATCTCTGGTTGCGAAT
GTCAGAAATATGGCCGAGCGGATCAGATCGGACGATTCAATCGATTTGGAAAGACTGGTGTGGACACTGCATGCTCGGCGGAGC
GTCTTACCCACCAAAGTATTCTTTACAGGCGGCACAGTGCAGCGATTGCTCAATTTTATGGATCGATTTGTGACAGAGTCAGAA
GAATCCACCAGTTCGTCAGCCGGAGTCAAATATCAGCCACTGAATCCGGCCGAGGCGCCCGGAATCCTTGGTATTTTCACAGGC
CAAGGCGCGCAATGGGCTTCAATGGGCTCCGCACTGTTTCAGCAAAATATTGTATTCCGGAAAGCTATTGAACGATGCCAGACT
GCGTTGACCAGGCTAGCTGACGGCCCCGAGTGGTCCTTGGTGGACGAGCTCACCAAAGGAGCAGCTGAATCTCGCATTGGGGAA
GCAGCCTTATCACAGCCACTCTGTACCGCTCTTCAGATCGGTTTAGTGGACATGTTGAAGTCCGCAGGTATCTACCTCCATGCT
GTTGTGGGCCATTCCTCTGGCGAGATTGCCGCCGTGTACGCCGCGGAGATGATCAATGCAGAGGATGCCATCAAGATAGCGTAC
TATCGAGGATACTACGCCAAGCTCGCCGGCGGGGCAAAGGGACAAGCTGGACGTATGATGGCAACTGCTATGTCTTTTGATGAA
GCCGAGGAGTTCTGTGCCCGACCTGAATGGCGTGGTCGTCTTGTGGCTGCTGCAAGCAACTCCCCACAGAGTGTCACCCTTTCT
GGGGATATCGACGCCATCGAAGAAGCAATGCAAATGTTTGAGGCTGAGAAAAAATTTGCACGGATCCTGAGGACCGATACTGCT
TACCACTCCCACCACATGCAACCTTGTGCTGAGCCATACCTGAAGGCGTTGCAAGCCTGCCAGATTAGGGTGAAAAAGCCCCGA
AAAGACTGCGTGTGGATATCCAGCGTCCGTGGTGATGTTGAATTGCTTGAAGGAGACCTCTCAAGCTTGGCTGACCAGTACTGG
GTTGATAACATGTGCAACGCAGTTCTTTTCTCGCAAGCTGTTGAGGCCTCGATCTGGAATGGAGGTCCCTTCGATGTGGCTGTT
GAACTGGGCCCGCACCCAGCACTCAAAGGCCCAGTGGAGCAAACTATAAAGGCAGTCTACGGTCCCATACCAGCCTATGCTGGG
CTCATGCGACGTGGAGACAATGAGATTGAGGCATTTTCCGGTGGAGTTGGATTTGTTTGGTCGCGTCTGGGGCCAAAATATGTC
GACATGATTGGATATCGTAAAGGGTACCAGGGCGCTGACCTTCTGCGACCACAGGTGTTGAAAGACCTGCCACCATACTCTTGG
GATCACAGCAAGCAATACTGGAAGGAATCCCGGATCTCTCGTCAATATCGACTGCGACAAGATACTCCTCACGATTTATTAGGT
CGTCGAGTTCCCGATGACACCGATGACAGCCGCAGATGGAGAAATGTCCTGCGACTCAATGAGTTATCCTGGATCAAGGGTCAC
GTTTTCCAGGGACAGGTCCTCTTCCCTGGAGCTGGTTACGTGGCTATGGCTCTCGAAGCAGCACGAGCCCTCGCTGATGGTCGC
CCGGTCAAGCTGTTCGAAATCGAGGATGTGTCGCTGCGCCGTGCCTTGGTCATTCCAGAGCAAGGAAGTATCGAAACTGTCTTC
GCTACTAAAAGCTTAGACGCAAAAAATGGAAGAAAAGTCACCGATCGATTGGAGGCAGAGTTCTCGTGTTATTTCTGCTCCGCC
GATGGCATCGAGCCGTTGGCAAAAGCATGCACAGGTCGCCTCATCATCAACTTTGGCAATCCAGTAGCAGATGTCTTGCCGCAA
AGGACTCGACTCCCAGCAAACAACGTTCCAGTCGACATGGGGAGATTTTACGATGCCATGAGCAACGTTGGTCTGGATTATCAG
GGCATTTTCCGTGGCCTTGTTTCCGGCAAGCGATCTCTTGGCTGCTCATCAGTCAAGGCTACTTGGGGCTCGGATATGCAGATT
GACGAATATGTGATCAATCCTGGTTTCTTGGATGTTGCGTTCCAGTCTCTTTACACAGCATTCTCATCGCCTGCTAGTGGAGAA
ATTTGGGCCCCCTACTTACCCATCCACATCGAAAGACTGGCTGTGAACCCAAATGTGTCCTACCGAGCGGCAGATCTCGAGACT
GAGATGGAGGCAGACGCATTCGTCACAACTGCAAACTCCACACTCCTCAAAGGTGATATCCAGCTTTACAGATACGAATCCCAG
CATGCTAGCCTGCAGGTGGAAGGAATCTCAATGAAATCAATGTCTGAACCACAGCCAGAGAATGATAGATGCTTGTTCTCTGAG
ACGGTCTGGGGTCCTGATGTCTCTACAGGCGTCACTGAAGTGGCCAGTCGCACAGTCGAAGATGACATTGATCTGGTAGAGGCT
CTTGATCGAGTATCCTTATTCTACTGGCGAAAGCTTGTGGAAGAGGCTGGCACGGAGATTTCCCACTTCCAGTGGTTCCATCAA
CGCATGTTTGACGCCATCGGTTTTCAGATCGCCTCCGTGCGCGACGGGCAGCATCCCATCGCGAAAGCCACTTGGCTCGAGGAT

Appendix A. (continued) Gene sequences used in this study

GACTGGAGTACTATCCTCACCGTGAGCGAGCCTTACAAACATCGCACGGATATGCGACTCATCCACGCAGTTGGCGAGAATCTG
CTATCTGTCGTACGCGGTGACACACAGCTGCTGGAAGTTATGGTCCAGGATGACATGCTCAACCGCTTCTATATGGAGGGTTAT
GGCTTTTCTGTCATTAACAACGCAGTCTCAGACGCTCTGGAGCAGATTACTTTCAAATATCCGCATGCCAACATCTTGGAAATC
GGAGCTGGTACCGGTGGAACAACCCGAAGTATCCTTGATCGAATCGGCACCCGCTACGGATCTTACACATACACAGACATCTCC
CCGGCCTTCTTCAAAGCGGCAGCTGACAAATTTGAGGATGCCAGAGGGAAGATACAGTTCAGAGTTCTGGATGTGGAAAAGGAC
GTTGGTCCCCAAGGCTTCGAAGAGTGCAGTTACGACATTATTGTCGCCGCGAATGTCCTTCACGCCACACGAAAGCTTGAAGAA
ACCATGAACCATGTTCGGACTTTGCTCAAACCTGGTGGATATCTGGTTCTGATGGAAATAACTGGCCCACATGTCTTACGGACA
CAGTTCATCATGGGTGGTCTTCCAGGGTGGTGGTACGGCGTCGATGACGAACGCGTACTGTCGCCCGCCATATCAGATCAACAA
TGGAGCCGGCTTCTTCTCGATACTGGATTCTCTGGCTTGGATTGCTTGAAGCCGGACATGTTGGATGGTGACAAACACTCATTT
TCCCTAATGGTGAGCCAGGCTGTTGATGAGAAGACAAAAATGTTCCGGAATCCATTTTTGTCGACGGGCATTACTCCATCAGGA
AGTCTTTTGATTATTGGAGGTCGAACGCACAAAACTTCACAGATGATTGAGGAAGTCCGTCACTACCTTTCCGTCTGGGAGACT
AGTGTAATGGATTGGGATAGCATTCAAACGAAGCAGGAAGGCCAACTGGCTCAGTTCGAAAACATCATTTGCCTCGAGGAGCTC
GATCAGCCATTGTTCTCCACAACCATCACCGACGAGACGCTTCTGGCGCTACAAAAGGTGTTCAACGGAACAAAGAATGTTGTT
TGGGCCATTTCCAATGCTGATAATGAGAATCCCCATGTCAACATGACAGTTGGCATTGGACGAGCATTGAGAACCGAGATTCCT
GGGCTGAATCTCCAATTTGTTGATGTTGATTCTTTTGACGATGCGCCAACGTGCGCTCAGCAACTTTCACGCATGTTACTCCAG
CTGGTGATTGGGTCATTGTGCCTGACCGAGAAGGCGCTGTGGATAACTGAGCCGGAAGTTCGTGTTCAAGATGGGCGACAACTG
ATTCCCCGTGTTCTTCCTATAGCGTCCATGAACAAGACGTACAACGCAAACCGGCGTGCAATCACCGATTTCATTGATCTCGAA
ACAACGTGCGTGTCTTTGAAGGATACTGCAGGCTCAATGAGTCTGGTACGAGGAGCATCTCTGCTAGACGAGCCCGTACAGAGT
GACCATGAGCGGCTCCGAGTTCACTTTTCTCTCGCCTTGGCGGTCTCTGGCGGTTCGGCATACTATTTGTGTAGTGGAATTTTG
TGTGAAAACCAACTTCCGGCCCTGGCGCTCTCCACTAACAACGCAAACATCATCGATATTCCCAAGAACATGATCTCGATCTTG
GATTTAGATGAGCCTTGTGGTGCCGCAACGCTAGAAGCAACGGCTCTCCACCTACTGGCGAGAAACATCTGCAGTCATCTTTCG
ACATCGAGTCAAACGCTCGTCTATCAACCTAGCCCCGAGTTTGCCCAAGCCTTGCTGTCGACGGGCTGTCGATTCTTATTTGTT
ACCACGAACAAAGAGAGCCCCTCGAGCAACTGGACCTACTTCCACCCGAGGGCGTCTCGTCGTGAGATTGAGTCCCTTCTGCCA
AAGGCTCTTGACACATTCATCGACTGCACTGGCGAAGTGCCTGGGAACTTGAAATCGTGTCTTCCCACAAACTGCAACATAATA
AATACCGAGTTCCATCAGCTTGATGCCAAAAATCTGGAACCTGCGTACTCTGCAGCTGTTGCAAACTCTAGCTACTCGTGTGGT
CGAAAGAGCATTCAGATCGGAGACCTCCCCTTGCGCACACCTCCCAGCCCCGTGCCAGTGGACTGGAGAGGCGTTAAGTCTGTG
AATATTTCTCTCCAGCCACTGATCATGAACCAAATCTTCTCGCCAGACGCTACTTATCTCCTTGCTGGCATGACAGGCGACCTT
GGTCTTTCGCTATGCCGCTGGATGGTTCAAAATGGTGCCCGACACATTGCTTTGACAAGCCGGAATCCGAACGTAGACGAGGAT
TTATTGCAAAGTATGCGTCAAGGCAATAGCGATATCCGAGTCTTCCAGAATGATATCACCGACCGAGATGCTGTCCGCAATCTG
ATAGATGAAATCAGAGAGTCAATGCCACGCATTGCTGGTGTTTTCAATGCATGCATGGTCCTTCGAGATGGACTCTTTTCAGAG
ATGGACACTGACACGCTCAACAACACTCTGAAGCCAAAAGTTGATGGTTCAAAGATACTTGATGAGATATTCCAGGATGATACC
CTTGACTTTTTTGTTCTTTTCTCGTCCCTGGCAAGTATCATCGGCAATGCGGGACAGTCCAATTACCATGCTGCAAACATGTTC
CTTTCTGGTCTTGCCAACAATCGACGTAAAAAGGGTCTTGCAGCGTCTGTCTTGCATATTGGCCTGGTCACCGATGTGGGCTAT
GTTGCTCGTCGAGGCCAAGCAATGGAGGAGCGGCTCCGGAGGTTGTTCTTCCTTCCTTTGTCTGAATCGGACATCCACCACGCG
ATGGCAGAGACCATTTTGGCAAGTCCAGCTCGCAGTGCACGCCAGCACGAGATCATTCTGGGCTTGGAACCGTTTGTTGCTTCG
GCCAGTGCAACGAAAAGGCCTCCCTGGGAACACAATCCTAAGTTCTCACATTATGTGTCCCGGCCACTGCTAGACGAGAGAACG
ACGAGTGCGGCATCAGAATCGGCGACAGATGTGAAACAACTGTTGAAGACGGCGGTTTCCGCTGAGGCCGTCACTCCTGCAGTC
CAGGAAGCATTTGCCAGAAAGCTTGAATCTATGATGCAGCTCCCTATCAACAGCGTCAATCTCAACGTTCCCCTGATCGATTTG
GGATGTGACTCTCTGCTTGCAATTGAAATTCGACGCTGGTTTATCAAAGAGGTCGGCATTGACGTTCCAGTCCTTAAGGTCTTG
TCCGGTGACACCACAGCTCAAATCTGCGAAGATGCTGTTTCGCAATTCTTGGCTCTACAGCTTGAAAAGAAAGACGCGGTGCCA
TGCGACTCAGCAGAGCAACTCAAGACAGAGCCTCTGATTCTCAGCGATGCCTCTGACAAAGTTGAGAGCGCAAATGGGGACTCC
GAAACTATTAGCCAAGGCGATGATTCTCGAGAAAACTTCTCCTCGTCTTCAAGTAGTCACACAAGCCCGAGTCTCCAAGCTGTC
ACCACTAGTACCAAGTCAGGTACGCCAGCACTACTGGATGAAGACACAGAGCCTCTTGATTTGGTGTCCAAGAGGGCTATCATT
CGCGCAGGCCGGGCCTCTTTCGCACAGTCCAGACTCTGGTTCCTGACGCAGTATCTCCATGATCCGACAACGTACAATGTCACA
GTGCGTTACGATGTCAAGGGGGAAATACCAGTCTCACGAATTGTAAATGCACTGCATACAACCATCCGTCATCACCAGTCCCTT
CAAACATGCTTTTATATGGACGCAGACAAAGAAACTTTGATGCAAGCAGTCCTCTCGCCGCCTCACACTTCCGTCAAAACAGTC
CTTTCTGGAAACGAGCAGACCATCAACGACGAGTACGAATCTTTGCGTTGTCGTGTTTGGAATCTCGAAAGGGCGGAAACCTTC
GCGGCCACTATTGTCTCTGTATCGCCTGAACAGCATACCGTCATCTTCGGGTACCATCACATCATAATGGATGGAGTGAGCTGG
CATCTTTTCTTGCGAGATCTGAACCTCGCTTTTTGCCTTCGATCGTTGCCTTCGGTTGAAACAGAATACATCGACTGGTCAGAG
AAACAGCTTCAGGCAGCTCAGGGAAGAGATTTCAGCAGACAGCTTGAATACTGGCGTAAGAGGCACTGTCCTCCGCCTTCTGTC
ATGCCATTGTTGCCCATGGCTCGGACAGATTGCCGTAAGCCGCTGACCACATATGAAAGCCACGTTGTCGGCGCGGAAATTGAC
CGAGATCTTGTCTTGCACATCCGGCAGGTCAGCCAAACTTTGCGTGTGACGCCTTTCCATTTCCATCTGGCTGTGATTCAAGCC
ATCCTCAGTCGCTTGCTAGATATGGAAGATCTTTGCATTGGCGTGGCGGATGCGAACCGAACCAGCGAAGCTCATTCGGGAACT
GTTGGGTTTTTCTTGAATCTCCTCCCCGTCCGATTTAATACGAAGAGAGGCAGCACATTTCAAGAGCTCGTTTCCTCGGTCAGG
GGCACGGTTCTCGAAGCTCTCGCCAACTCAGAGGCGCCGTTCGATCTCATTCTCGAGGATCTGAAGATTGTGCGAAGCTCAGAA
TACAGTCCCTTATTCCAGGTTGCTGTCAACTACCGCATGGGGGCAATGCTCCAAGTTCCTCTCGGAGACAGCATGATGGAGATG
GTGTTTGCCGATGACGCAAAGAATCCTTACGACATCAGCTTCGGGATCACGGAGACGGCCACGGGCTCCTGTTTGCTGGAACTG
ACCAGCCAAAAAGAACTGTACACGAAGGAGTCGACTCAGCTCTTTTTGCAGATGTACATTGATGTTCTCCGCGCTGCATCTCAA
GATCCCTCGCTTCCTGTTAGCCAGCTACCGGTCACCATTAAAACGCTAAACGAAGAGAGATTCGCTGTTGCGAGAGGCCCTCGA
GTTGATCATTCTTGGCCGAAAACACTAACGGAGCGCTTTGAAGAGATGCAGAGCCTCTTCTCGGAGAGCACTGCCATCAAGAC
GGTCATTCAGAAATCAGCTACTCGCAACTGGCCAAGAGAGTTAATCAAATTGCCACAACGCTCATCGGTCAAGGTGTCGTCATG
GGAGACAACGTCGGCGTTCTCATCCATCCTTCCATCGATGCAGTTGCCTGTATGCTGGCTTTACTACAAATCGGTTGTATTTAT
ACCCCTCTTGACACGCGGCTGCCCTTGGCTCGATTGAATATTATTGTCGCCGGCTGCAAGGCACCTTTGGTTTTGCACCACACT
GCAACTGAAGACACTGCACAGGAGCTCGGAAAACTGACCAAACTGGTCAATGTTGACAATTTAGCAGAGGCGATTTGCTCAAAT
CAACCGACAGTCATCCCTGCTCCTTCTGTGGCATCTTTCTTGTTCTATACGAGCGGCAGCACGGGGGCACCAAAGGGCATTCTT
CTCACTCAAGAGAACTTCGTGAATCATCTGGCGGGGAAGACTAGCAGTCTCGCCCTCGGCAGAGAAGTCGTCTTGCAGCAGAGC
TCTCTAGGTTTCGACATGTCCGTTGTGCAAACGTTTTGCGCCCTGGGCAACGGTGGAACCTTGATCATCGCGCCGAAGGAAGCA
CGCGGAGACCCAGTTGCACTTTCCAGTCTCATGGCACGGGAACGAGTGACGTTGACGATCGCCACGCCCTCGGAGTACTCCTTG
TTGCTCCGCTTCGGTCTTGAGAATTCCCAGAAGCTTCACGCTTGGAAGCATGCCTGTATGGGTGGCGAGGTTGTGTCACGATCT
TTGATCCAGCAGTTCTGTCAATCAGATCATCCGAATTTGCAGCTGACAAACTGCTACGGTCCGACCGAAATCACTGCTGCTGTG

ACGTTCCAAAACATCTCTCTCCAATTTGAGGACAATGATATGGTGGATGGATCTCTGGTTGGAAAACCTTTGCCTAATTACTCG
GTCTACATCCTGGATGCTTCCGAATCTCCCAAGCCCGTTGGTGTCGCTGGCGAGATTTGTATTGGGGGCGCTGGAGTTTCTCCA
GGCTACTTCAGGGCTTCGGAACAAACAGAGTCAAAATTTGTTCAGGACCCCTTTGCCAGCTCAGAGGATGTTGCTCGCGGTTGG
TCAAAGATGTATAGAACCGGAGACATGGGGCGGATGCTTGAGGATGGTACTCTGGTTTTCTTGGGCCGCATGGAGGGTGATAAT
CAGGTCAAGTTGAATGGCGTGCGGATTGAGCTCGATGGAATCGCTAACAGCATTCTCACCACTGGACGTGATCTGGTTTCCGAG
GCTGTGGTCACAGTTCGCTCTGTCCCAGGTTCTGACTCATCCCTCCTTGTGGCTCACGTCGTTCCCCTTGGAGGGCAGGTTGAC
GTTTCCAGACTTCAGCAGCTCGCTAGAGATCTTCCGCTCCCGCCATACATGCTTCCGTCCATGGTCATTGCTCTGGACCGTCTG
CCAGTCAACGCCAATGGAAAGGTTGACAGAAAAGCCATTGCAATGCTACCTTTGCCAAATAATCGTGACGAGATCGCCAGAGGT
CCTGGCCTCGGGGCTACAAGACATCTCAACCTCGCAGAGGGCGAGCTGCGCCTATTATGGGAAAAAACTCTTCCGGTCTTTGGT
GGTCCTTCAAGACTGGATGCTGATTCCGATTTCTTTGTTCGAGGAGGAACATCCATGCTGCTGGTTCGACTGCAAAGTGCCATC
AAGCAGTCGATTGGGATTACAGTCCCTATCGCAGAGTTGTATCAGTTCCCGACTCTTGGTCAAATGGCGCGGCGAATCAGTCGC
CGTAAGGAACAACATGAAGCTTCCCACGCATCCGTGATTGACTGGGCTTCCGAGACCTCACTGACGCAGGACCTCCTCTCTCTG
GCTGATACTCAAGTCTCAAAAATTCGAGGGAAGGTCCAGAACGGAAAGGAGATTCTGCTGACCGGTTCCACCAGTTTCTTGGGG
AAAACAGTACTCGGATTCCTTCTCAATGACCCGTGTGTCGAGAGAGTACATTGCGTCGCTGTTCCCGTGGAGGAGACCTCTCTT
CTTCCTCCATCTGATAAGATTACCATTTACCCTGGCAACCTCCTCATGCCCAATCTAGGACTGGCACAGACCGATCTGGCAAGT
CTTCAGCGCTCACTAGACGTCATCGTTCATGCCGGAAGCACAGGGCATTGCCTGAACAACTACTCCTCACTGCGCGTTGCAAAC
GTTGGCTCAACGAGATTCCTGGCGGGACTTGCTCTGTCACGCGGAATCCCCATCCATTTCATCTCGTCAAACCGTGTCACCTTA
CTGTCTGGGAACACCAGTCTTCCCCCAGTCTCTGTTTCATCCTCTTTGCCCAACACAGATGGCTCTGAGGGCTTCACTGCCAGC
AAATGGGCATCCGAGCAGCTCCTTCAATCAGTCGCCGCGCTCCAGCCGGGACTCTCCGTGACAATTCATCGTCCATGCGCTGTC
ATAGGAGACGAGGCACCGAATGAGGACGCCCTGAATGCTCTTCTCAAATACTCCAAGATCACGCACTGTGTTCCGCGTTTCGAG
AACTTCGAAGGGTATCTTGATTTCGAAGATGTCCACAAGGTCGCCGGGGCAATTGCAACGGATGCTTTGACCCATGGTGACTCG
GAGAAGGGCAAGTCTGATGCGCGATTTGTACATCACTCCAGCGGCCATCAAGTCTCCATGCAAGACTTCAAAAGCCGAATGGAG
ACTTTGTTCACTTGCCCGTTTGAGGAGGTCTCCATGACCGAGTGGATTGACCGCGCCTTGCAGGCTGGAATCGATCCACTAATC
ACCGGATACTTGGAGGCGATGACATCGAAGGGCGAGACCATTCGCTTCCCATATATGGGAGCCACAGGGTCGCTTTGA

>ApnB

ATGGCGGCTCTTAGTCTTGCAGATCTGGCCTCCCAGATCGAGTCTGTTGCTGCTTCTACACCAGACACACACGCGAATGATGTT
GACCGAAAGCGCCTGCTAAGCGCGTGTGATCAGCTCCGCAAAAAATTGGAAACGCCGTTCGAATTCACCCTGCGGACGGTTTTT
GCAGTATGTTTTTCACCCTTGTCCCTGGCGGCGGTAGCGGCCGCGTGAAAGTATCAAGTACGGTGGACTAATGTGGCTACGTGG
TGCAGGGCTATCCCGCCATAGCCCTTCGGCTTGGTATTGATATGAAAATCTTTGATGCGGTAGCGCAGCATATCGAGTCTTCGG
GCGAGAAGACCATCACTCTGTCGCAGCTGGCGGAGATGACCAAAGCTGATCCTCTGCTTGTGTGTAAGTCACTCAGTCCTGCCT
TTCTGTCGGAAGCAGAATCCTTAGAGCAGGCTGTAAAAGACCCCCGAAACAGAAATTGTTGATTACATGAAGACTGACGACTCA
ATTCTTCTTTGCTTTTTTTTCTCCAGCACGTGTCATGCGCTTCCTGTCGGCATTGGGAATCTTCAAAGAAGTCGAACAAGATGT
GTACGCTTCAACTCCTGTGGCAGCAGGCTACGTCTCGAGCTCGCCTCTCTCTGCGGCGGTGATTCACGTGTAAGTTCTTGACTC
TTACTCCCCAGAGTCCCTGTTCGATTCACTGACGGCCTCTGGAAGTACGCACTTCTCAATGGTCCTCACTCAGCTCCCGGCCTA
CTTCCATGAGAAAGGCTGGCAGAATCCCAACAATGTCCTCGATGGCCCCTTCCAATTCGCCACCAAGACAAAGTCGCACTACTT
TGACTTCCTCTCCGGAGAACCTTATTACCAACAAGCCTTCAACACGGTCATGACCATCTCTCACCGTCGCCAAGGCAAGAACTG
GTTCGATTTCTTCCCAGTGGAGGAGAAACTGGGCGGCGCGACAAGCGACTCCGATGTTGTCCTCGTCGACGTGGGCGGTAGCCA
GGGAGGTGATATTATCGCCTTCCAAAAGCAATTCCCCAGCGTCAAGGGTCGTCTCGTTCTGCAAGACCTACCAATTGTTATCGA
GGCAATTAAAGATGGCGAGCTCCCCGCGGGTATTGAGGCGCAGGGATACGACTTTTTCGACCCGCAGCCCGTCAAGGGTGCCAA
AGCCTACTATCTCCGCACCGTGCTGCATGATTGGCCCGATGAGCAGGCGAGACAGATTCTGGCTCGTGTTCGTGAGGCGATGGC
GCCTGATTCATTACTGCTCATTAACGAGACTTTGCTCCCAGAGTCAGGCATTGCGCTGTCCTCGGCGCAGGCGGATCTGACCAT
GATGGTCTCATTTGCTTCGTTGGAACGAACAAAGGCCCAGTTTGAACAGCTGCTGAATGATTCTGGGTTTGAGCTGGTGAAGGT
TTGGATGCCGGAGGGATTCACGGCGAGCTCGGCGGAGCTCGCCTCGCAGGCGACTCTGCTCGAAGCGAGACTGAAGTCTTCTTA
G

>ApnC

ATGATCGACCCCAAGACAAGCCAGAAACACATCCTCATCGTGGGCGCCGGACTCACGGGCCTCATCCTGGCACAGGCGTTGAGG
CATCTGAAGGCAACCCAAAATCCGGCTCACAATCAAGTGCGATACACGTACTCGCTCTTCGAGCGTGATCCGTATGCTTTCGCT
CGCGGAGCAGGGTGGAGTCTCACCATTCACTGGGCCCTCACTGATCTGCGGAGCATTCTGCCCTCAGACATCCTGGCTGGTTTC
AAGGACTGTCTGGTCAATCCCGGTGCCGCTGAGGAGGGTAATCCCGGCAAATTTCAGTTTCTCGATCTTCGAACAGGAGCGGCG
AAGGAGTCCTGGCCGATCCCTGCCAACGCCGCCTCGCGAGTGTCGCGTGAGAAACTGCTGGCGCTGCTGATGAAGGATTTAGAT
ATTCAGGTACGTTGAATCAGCCCACCTTGATCCTCATGCCCTTACTATGATGCAGTGAGCTCATTTCCAGTGGATAGTGGTCTA
AGCAAATCGCCGACATCACACACCCCAGTGACTCCTCCGTGACGGCGCACTTTTCTGATGGAAGCAGCGCCACGGGCGATCTAC
TCGTGGGCTGCGATGGGGCGCGTTCAATGGTTCGCCGCATTCTCTGCCCCGATATTGCCATGCCATATCGGCTTCCTGTGCGCC
TAATTGGGCTCCGCGTGCTCTATCCCGTGGAAAAAGTGCGCCAACACTGCGAGCCCATCGATGTGCACTTCTTCCAGGGTGGTG
ATCCCCAGACCAACGTGTACTTCTGGTTCTCGTTCATCCATCTACCTCGGCCTACAGACCCAGACACCCATGCCACTTGCCAGA
TCATGATGAGCTGGCCTTACCAGCCTGGATTCCTGGGCAAAGAAACCCCGGTAGAGATGCCCGAAAGCAATGCCGATCGTTTGG
CCTGGATGCGCAAGCTGGCGGCCGATTGGGCTGAGCCTTTTCGGCAGCTTGTCTATGATCTGCCCCCGGAGACGGAGCTGAGAG
AGATCGTCCTCGAGGATTGGCCGCCAAAGAAGGGTGCGTGGGATCATCACAACGGGACAGTGACGCTGGTGGGTGACGCGGCTC
ATGGAATGACAATGTGTAAGTCCCCAAGCTTCTATCACATCCTTCGTGGATGGAGAATGATGCCGGAGTCTAGGCTAACGGGGA
GAAATTGCAGTCCGTGGAGAGGCGGCCAATCATGGGGTGATTGATGTCTCGAGACTCATTAAACTATTCACTGCGAGCACCAAA
GAGGAGGTCCCGATGAAGTTGGAAGATATTGTGCGAGAGTATGAGTCAGAGATGATTGAGCGTACACAGCCGGCTGTGCTCAAG
TCACGACAAGCGTGCATTGATGCTCATCATTACGAGAGTGTGAATGGGTCTAGTCCGCTGATTTCGAAACGGGCCATGAAAGAT
TGA

Appendix A. (continued) Gene sequences used in this study

>ApnD
ATGGTCAAGTGGTTTGTTCTGATTCTATGCATCTTGCTTGTGCTCAAGATCGCCGCTAGACGCCGAAGCAGCAGAATCATTAAC
CAGCATGGCAAGACCATCCCCGAAGTGCCTGGCGATGCTCGCGTGTCCAAATTTGCTTTCAGGTACAAGAGAATCATATTTTTG
TGGGAAATCATCGATGGAGAGAGCCAAGCTAACAAGAGTCGAATAGCCAGCAGCTGTCTGATCAAGGCAAGGCTCTCGCCAGAG
AGGAGCCGTTCATCATCCGTAATGGCCGGGCAAGAGAGCTGGTGGTGACAAAACCAGAGCACATTTATGACTTTTACAAGGGTG
ATACCAAGCGTACGTCTTGCACAATCTCAGTGAGAGAAAGGGGAGGGGAAAAAACTCATGACCCACGCACAGATCATCCTAAAC
CACCGTATCTCAATATGGGAAAGTACTTCAGCGGGTGAGTTCTTTGGTTTTGCATTTTTGTAGCGCTCCTTACGTCTGCCAGTA
TGATTCGAAATATTAAAATGGGATTTTCAGTATTCTGGGACATGCAGTTGGAGCACTTGCAGGGGACCGTTGGTCGGTCATCCG
TCGCTACTTTGACCCCGAGTTCTCCTTTCAAGTCGCGCGACAAGCGATTCCGCAACTCAGCGCGAGTATTGATCGATGGCTCGA
TGATCTACCCCGGCAAACAACGACGCAAATCATAAATTCAAGGGGAGGGTTTGCACTAGAGCTCAAGAAGCCTTGTCGATTTCT
GCCGCTCCGACTCGCGGCTGAGTTTGTGTATGGGGAGGTCTTTAATGATGAGGTGGTGTGCCCTATCTATTTGTTTTGTTTTGT
TTTTGAAGCCTTCCATGATATTTAAAAACCACTGGCTAACGTGAATCATGACTCTCACTAGCTTTTTGCGGCTCTACTGCAACT
TAATGTCCTTCACGAAGTGATTTTGCACGATGTGATTGCTAACAAGCGACTGGCGACAAATTTAGGCTGTTGGTTGGACCGCAG
TGCGACTAAACGCATGAATGAGTTCCGTTTACAGTGGAAGAAGTTCAATCTGGGAATCATTCACTCGGCCCGACGAGGCTCGCA
TCTCTGTCCTGCGGAGAAGATCTACCGTGGGGTTGAAACCGGTGACTTGAAACTCGAGGAGGTCAGTATGCCTCTACTATCCTT
TCAATTCCTTGATTGTAAACCCGCATTGATCAATACACAGTTTCTACACACTCTGGACGAAATTCTCTTCGCCAACGTCGATGT
CAGCTCCGCCGTCCTCAACACTCTCTTCGAGCATCTTGCTTCCAACACAACCTTCCAACAAAGGCTGTGCGAAGAGATTACAGC
ACAAACTCAGACTCATACCCAGACCCCGAGCCCTACCAGCATTGACACAGATACTACCACACATACCGGAAAATATCTCTCCAA
ACAAGATACCCTTCTCCACTATGCTGTGATGGAGGCTATGCGATTTTCACCTGCTTTCGGTCAGTCCACATCCCCCTATAGCAC
CTCTTTCGCTCTTTAAATCTTCAATTCTGCTTCGTGCATATAATCTGCTCCACGCTTATACGTCCCAGCATTCTCTCTGCCCGA
GTGCACAGCCGTCCCCAAAGAAATTGGAGGGTATCGTGTTCCCGCACGGTGTCCCGTCGTGATAGACGCTAAACGTCTGAATGC
CGACCCAGCGACCTGGGGCAAAGACGCCGATGTATATCGCCCAGAACGGTTTCGTGAGATTGCATCCTCCAAATTACGGTACGG
GTTCATGCGATTCGGCGTGGGTGCTGCGAGCGGCCGGTGTCTCGGAAAGCATTTGGCGGATGCGATTTTCAAGTTGAGCTTGAT
GGCTGTGCTGCAGCGATATAGTCTTCATTTGGGGCAGAATGGGTCGGAGATTGAACTAAGATATGTTCAGCGGAAAGAATGA

>ApnE
ATGATCCCTCCAAGGCAACAGACTGCTTTGAAAATCACCGCGGAAGGTCGCATTGCCGCAATCTCATGTGCTCTCCCATCCATC
CACGATGACGAGTTGCTCGTTCGCGTGCGAAGCATCGCTTTGAACCCGTTCGACGCCAAATCGGCCGAGATGTCACCTACAGTT
GGCGCGACTCTGGGCTGCGATTTTGCAGGTGATGTAGTCGCCATGGGATCCCAGACCAAGAGTCGAAATTTCAACATTGGCGAC
CGAGTCTGCGGCTGGGTTTTTGGCAACAATCCCAACCGCTTGGATAATGGGGCTTTTGCAGAGTATACAGCTGTGCCCGCAGAT
TTGGTATTTCGAATTCCCGGCAAGATGAACTACAACGAAGCGGCTACGTTGGGTGTAGGGTTGGCTACCGTGGGCATGTCGCTA
TATCACTGTTTACAATTGCCAATGGAGCCCGAGAGGACGCAGAAGTCGCCGTATGTTCTGGTGTACGGTGGTTCGACAGCGACC
GGGACGCTTGCAATTCAGATTCTTGCACGGTAAGCGCGGTATTCTGTAGTTTGATTTTTCAATTCAAAATCAATGGTGTGAAAG
AGTCGTCTTCTGACTGAGGTTTCTCGCGCTCCTCTACAGCTCGGGATATACTCCCATCACCACCTGCTCACCACACAATTTCCA
CCTCGTCCAATCTCTTGGCGCGGTCGCAGCGTTCGATTACCACTCTCCCACATGCGGGCGCCAACTTCGGGATTTGTCGTCTGG
AACGATCCACTACGCACTCGACTGCATCACCGATACTCGCTCCATGGCCATTTGCTACGAAGCCATCGGACCATCGGGCGGTCA
CTATCTCAGTCTGGACCCGTTCCCCATCCGAGGCCACACACGACGAAGCGTCAAACCAAATTGGGTGCTGAGTCTGACCATGTA
CAATGAGCCAATTCCGTGGAAGAGGCCGTTCAAGCGAGATGCACGCCCTCAGGATCGAGAGTTTGCAAAGGAATGGTACCAGAT
TGCTCAGAAGATCATCGATGCGGGCGAGATTAAGCCGCTGCAGTCGGAGGTCAGGCCGGACAGTTGGACGGGTATACCGAAGGG
ATTGGATCTCCTCCAGAGAGGAGAGGTTTCGGGGAAGAAACTTGTGTACGAAGTAGCGAGTCATTGA

>ApnF
ATGCACACTGAGAACCTCGAGCCAGGAACCGTCCTCATCGTGGGTGGCGGTCCTGTCGGCCTCATCACGGCCACAACGCTAGCC
AAATACGGCGTGCGCAGTGTGATCCTCGAGCGTAATTTGACGACCACAAAATGGCCTAAAATGGACCTGACCAATGCTCGCTCG
GTGGAAATCTACCAGGCTCTAGGCATTGCAGAAGAGCTGCGAAAAGTTGCTGTGCCGGGCCACTACCCCTTCACATGTCTCTTT
TCCAGCGGGCTGCATGCCGAGAAAGCGATCACTGCGTGGAATCTGCCAAGTCCGGATGAGTATCAGAGAAGGAGTAGGGAGAAA
AACGATGGAAGCATGCCTTCGGAACCGTGGCTACGTGTTTCCCAGGAGATCTTTGAGGCGTGGCTCAAGGGTCTGGGGATGGAA
AACCCGCTCATCGATTTCCGTGCTGGATGGGAGGTGACCAGCGCGCATGAATTCGACTCGGGCGCGCAGGTAAATGCTATTCAT
CCCGAGACGAAAGAAGAGTGGTGCATCAACGCGGACTTTGCAGTGGGATGTGACGGTGCTCACAGTGCTATACGCAAGAGCCTA
GATATTCCCTTGGAAGGAGGCCCGATGTGAGTACTGGCCAACTCGTTTGTGGAGTCAACATGCAGATGAGAGAAAAGCGGGAAA
AAAAAGAACTTGCTGACAGATCTCTTCAGTCACGGATATGCAGTTCTTATTCATTTCAAGTCCACAGATCTCACGCGTATCCAA
AAGCAAGGTCAATTTTGGCACTTGTTCTTCCCCAATGCTGCCAGTGATGGTGGGTCCATCAAGGGAGCCGTCATTGCGCAGGAC
GAAGTCGACACTTGGACCATTCACCGTTTCATGAAACCCGACGTGGACCACACGCAACTTTCCTCGGAGGATATCGTCTACGAC
CTTCTGGGTGGAATGAGTGGCCGACCTTTCCCAATCCAAATCGATGAGGTCCTTGTTCGCTCAACATGGACCCCAAGTGTTGCA
CTTGCGCGTTCTTACGCAGGGTCCAAGGCGAAAATCTTCATCGCGGGAGATGCCTGCCATCAAACCGTGCCCACGGGTGGGTAT
GGAATGAATACCGGTATCGCGGATGGCTACGATATCGGCTGGAAGCTGGCGGCCGTCATACAAGGGTGGGCGGGGCCCGCGGCA
CTTCTATCTTACGAGCAAGAGCGCCGACCAGTGGGTGAATTGGCGCTCCAGTGGTCGAAAGTCCACATGGGCAACCTGATGAAG
ATGTCTGCCGAGCTGGGCCTCGACGCCAATGTGATCGACTCTGATGTTGAGGCCGGCGTTCAAATGCGAACAGCCATGCACGAG
TACCTGCAAACACACGATGGCCATAACCAAAGTGTTGGTGTGGAGATGGGATATCGCTACTCTTCCAATCTGTGTGTCGCTGGT
CCACTCGATGCCGTGCTTTCGCCGCCAGAGTTCCATCCCAGAAAGTACACACCAAGCACGATGCCGGGTTACAGAGCTCCTCAT
GTGTATTTGACGACGGGAAAGGCTATCTCGCACTTATTTGGTGACGGATTCACACTTGTTGCGTTTCCAGAAAGTGAGGAACTA
GCCACATCGATCGAGTATTTCAATATCTCAGCTTCGACGAGAAAGGTACCCCTCGAAGTTGTTCAGCTACCAGGGGAGGTACAC
GCCCACAAGGTCTGGGGCGCCCCGCTCACTCTTGTCAGACCCGACGGGTTTGTCTCCTGGCATGGTGACAAGGTGAGTAACCAG
CAAGAAGCCGATCGTATCGTGGCTCAGGCGAGCGGTTCTCTTTCTGACTCCCCGAATGACTGCGAAGAATCGCAAGAAAGCAGT
GTGTTGTAG

Appendix A. (continued) Gene sequences used in this study

>ApnG
ATGAGTCTGCTTCACATGTTTTCGAAGGAGAGTTTGCCCACTCTCCTTCAACACACCACTTTTCTCGATGGCATCAAATATGCC
CTGTCTGCATGGGTCATTTATTCCTGCTGCATGATTGCTGTTGACTGGGTTGTGTACGAACACAAACGCAAGCAGCATGGGTGT
GGCAAGATCCCCCGATACCCTCATCGCGACCCATTTTTCGGCTTCGACATTGTTCTCGGTATGGCTAAAGCCTTGAAGAATGAC
TACTTTCTTGTTTGGCTCAATAAGGTGCACCGAGACCTACCAAAGACCTTTTTGGTCAATTTTGTCGGCACTCGGTTCATTTAC
ACCATCGAACCTGAGAACATGAAAAGTATGTCCGCCATCAACTGGCAGGACTTTGCAGTTGGTCCGATGCGACGCAATAACAAG
GCGACTGCTCCATTCGCGGACAAGGGCGTTAATACCGTCGATGGGCATGAGTGGGAATTCAGTAGATTTCTGATCAAGCCCTTT
TTCAAGCCGGAAACCTTCAGGGATACCAGTCGTCTGTCCATTCACGTCGACCGAATTCTGGATCTGCTGCCAGCGGACGGCGAA
ACGGTCAATATTCAGCCACTCATTCAACGTTGGGTATGTTCGCTTTTCTTCTCATGGGCTGGAATTGAAATTCGGAAAAGCTTG
AACTCAAACACAGCTAACCGGTGTGGTTCATAGTTCCTTGATGTGACTACTGAGTCCCTCTTTGGAGATTCAATTGAGTCTCTG
GTCTACCCAGAGAGAGCGCCCATTTGCTGGGCCATGGTGGACGTTCTCCGAGGTCTTCGGCTTCGACTCCAGTGGTACAAGTAT
ATCTGGCTCTTCCGTCACCAGGCATGGCTCGATGCGGTTGACGTGGTCCACAAATATTTGAACAGCCATATTGATCGAACATAT
AAAGAGCTCGACGAATACAAGAGCCAAGGGAAAGATCCCGAACAAGCCGATCGCAAGGATCTTCTGTGGTACATGGCGAGCAAT
CTTCCTCAGGACAAGGAGGCTCTCCGGTCTCAGATCTGCTTGATCTTTGTGCCCAATAACGATACCACCTCCATCTTCATCAGC
CATATCCTCTGGAATCTTGCCCGTCATCCAGAGATTTACGAGAAATGCCGGCAGGAGGTACTCGCGCTGGGCGACGTGGAGTTG
ACATTCTCGGTACTGCGAAACATGAAATATCTCATTGGCATTTTGAACGAGAGTGAGTATCTCTGCCCGACCCTCCTACTCTAT
CAAAATGAGTATCTCATAAACTTACACAACTTAAAATAGCACATCGCCTTTATCCCAATGGTGTGACCCAGGTGCGAAAGTGCA
TCCATGACACCACATTACCCATGGGTGGAGGTCCAGATGGCAAGCAACCAATCTTTGTCCGTAAGGGGGATGTCGTCCAGGTTA
ACAAGAATGTCATCCATCGAGATCCAGATATCTGGGGTCCTGACGTTGAGGAATTCCGGCCTGAGCGATGGGAGAATCTTCGGC
CCTACTGGAGCTTCGTGCCCTTCGGTGGAGGTCCACGTCGTTGTCCCGCTCAGATGCTGGTCACAGCGGAAGCTAGCTACTTCC
TGGCTCGGCTGATGCGCGTCTACAAGAGGATTGAAGCCAGAGATCCGAACCCTTATGTTGGAGTGATGCGGGTGGGACCTTCAA
ATAAAACTGGCGTTCAAATTGCGCTGTTTAAGGAGTGA

>ApnS
ATGTCGGGCTCCGTTCAGAAAAACGTGATGCTGGCGTCCAAGGGTAAGTCCGCCGCTGTATCATCTCTTCCCCCCCCCTATTGC
TTGCTCAATGTCAGCTGACACGTAGGGTCCCCTCCAGCATCGACCATGCTCGGTCTACGCCGTGCCGTAGTAGCGTCCCCGTCC
CTAATCAGGCCGACCATGGGCCTGGTGTGGCAACAACAAGTTCGGTACAACTCCCCAAAGCATCGCGACAGATATATACACAAA
CACACATCGGCTAACCGTTCTTTCACACTTGTAGTCTGCAGACCACGAAACCCGACACAATACAATTGAAAGTGAACAAGCCGG
AAGATGGTCTGAGGGAACTCGAACGCCAACGTCTTCAGCGACCTGTAGCGCCCCATTTGTCCATCTATAAATGGCAGATTCACT
CAGTCAGCTCCGCTATGGAGCGAAACACTGGACTCCTGCTTTCTGGCGGCCTGTATCTCTTTGCGACCTCTTACTTGGCCGCTC
CCTGGCTGGGCTGGGACCTATCCTCGGCGACCCTCGCTGCCTCGTTTGGTGCCTTGCCGGTGGCTGCCAAGGCTGGTATCAAGT
TCTTCTTGGCCTGGCCCTTTACTTTCCATCTTTTCAACGGTATCCGGTATATCGCCTCCTCTGCCGGGTACACATTGACGAGCA
AGCCCCAGATTGTGAAGATTGCATGGGGCGTGGTCGGCTCATCGGCGCTGGCTGCCATTGGCTTGGTCGCTTGGGTTTGA

>ApnS$_2$
ATGTCACACGATCGGACATTGCGCACGTTAGCAGGGGGTCAAACGCTAGACGGCCAGCTCCTGCAATGCCCACCCTCTTTCTTT
CTGACTCTTTCATTCTCTCCCCTCTCAGCTCTCTCAGTGTGTCTCCATTAGGTCCTTGCGAAGCACCCGCACCCCATCCATCAT
GTCGGGCTCCGTTCAGAAAAACGTGATGCTGGCGTCCAAGGGTAAGTCCGCCGCTGTATCATCTCTTCCCCCCCCCTATTGCTT
GCTCAATGTCAGCTGACACGTAGGGTCCCCTCCAGCATCGACCATGCTCGGTCTACGCCGTGCCGTAGTAGCGTCCCCGTCCCT
AATCAGGCCGACCATGGGCCTGGTGTGGCAACAACAAGTTCGGTACAACTCCCCAAAGCATCGCGACAGATATATACACAAACA
CACATCGGCTAACCGTTCTTTCACACTTGTAGTCTGCAGACCACGAAACCCGACACAATACAATTGAAAGTGAACAAGCCGGAA
GATGGTCTGAGGGAACTCGAACGCCAACGTCTCCAGCGACCTGTAGCGCCCCATTTGTCCATCTATAAATGGCAGATTCACTCA
GTCAGCTCCGCTATGGAGCGAAACACTGGACTCCTGCTTTCTGGCGGCCTGTATCTCTTTGCGACCTCATACTTGGCCGCTCCC
TGGCTGGGCTGGGACCTATCCTCGGCGACCCTCGCTGCCTCGTTTGGTGCCTTGCCGGTGGCTGCCAAGGCTGGTATCAAGTTC
TTCTTGGCCTGGCCCTTTACTTTCCATCTTTTCAACGGTATCCGGTATATCGCCTCCTCTGCCGGGTACACATTGACGAGCAAG
CCCCAGATTGTGAAGATTGCATGGGGCGTGGTCGGCTCATCGGCGCTGGCTGCCATTGGCTTGGTCGCTTGGGTTTGAGGAAGA
CGACCATTGACGAGGAAGACTACCGCTTAGAATCGGGGCTTCGAGAAAGAGAGCAGTCATTCTGTATATGTATCCCAGAAGAGA
ATGTACAATTGACCGACAGATTTTGGAAGTAGCATGGCAGTGATAACAAGCAATCACCATCGGTCTCGCTCGGAGCCTCGGTGG
AGTGTATGAGATCGTTTCATCTCCACGTCTGTACTCCTGTGCGCTCAACACCGCAGAGACTACACGCACACACCCCCTCTTCTC
TCTCTCTCTCTCTCTCTCTCTAGTGTCTGAGATGCATAATAAGATATGACGGATTATTGCGCCATCAAAAGATCGATCGGTACT
CCATGCGCTCTCTGGAACAATTATCGAAACCTCTACATAATGATCTTTCTTCTCACAGGGGATCTAACATCAAACGGAGTCTCA
GGGCTTGTGCAGCACATCTCAACACGACGACTGATCTTAATTTGCTGTTGTCTTCGAGATGTGCTTCTGTAGCCCCGTCTCAGT
AAAGCAGTATACGTCAGGTCAGATTATTCTTGGGCCTATTATGTCTTGTCTTACCGATTTTGAACAAGGTCAGTGTAAGCAATA
CGCACTATGAAGGATTCTATTTCGTAGTGTCATTCCTCTACACTATCGTATGTGTGGTAA

>ApnT
ATGATGAGCTTCGTCATACTTCTGAGCTGGCTCTGGATACTGGTCGCTCGTATCCCCGGTGTGCCACTGTGCAGAAAGAATCTG
TGGATGCTGATGACGATTCGAAGTGTCACTGGCGTTGTTGGGAGTAAGTTATAAACAGATCTCATCAAACGAGAACCAAGATCA
AGAGGAACGACTTTGTTAACGAAGTGATTACCACTCCCCCCTCCAGTCTGGGGCTTCTACTATTCCCTACGTGCACTACCTCTT
TCCGAGGCGACGGTGATTAACTTTCTGTCGCCAATGGTAGCCGCATACGCTTCTAGTCTTCTGGGCAAAACACCGCCATTCACT
CGGACGCAGCAACTCGCCATCGCTGTCTCCTTTCTGGGAGTGACTCTCGTCTCCCAGCCATGGCTGAACATCCATGCTGCTGCC
CCCAGCACCACTCAAAAAGCATCCCATAGTAATTTTTTCTGCATGGGTGGAATGTGGGGTCTCTTACCGACGACGGTGATTTCC
AAGACGGTGGAACGTATTACCGCCATCGGGGCCGGCCTGGTGGGTGTCGTCGGCGGAGCCGCAGCCTACGTGGTGATGCCGCTG
ATCGGTCATGATGCGGATCCAGCAGTGACAGTGAACCATTTTGCGACGTGGACTGTTGTGCTGACCACATTAAGCTTGCCCATC
ACCGGGGGTTGACGCGTGGAGAATGCCTTATCCAGTGGAGTGGGCGCAACTGGTGTTTCTGGGGGTGTCTGGGTTTTCTGTCCAC
ATCTTGCTTGCCATGAGTATGCAGCAAGGAGGGAGCCCCCGGACTTTGAGTATGGTATACATCCAGATCGTTTTTCTGTTGATC
ATGGACGCGGTGGTGTGGGGTGAGAGTCCGAGCTGGGTGAGTGTGGGCGGGGGGATCTTGATTCTAGGAAGTGTGGTGACTACG
GTGATGCTCAAAGGGAGTCAAGCAGATGGGGAAGATGGGATAG

Appendix A. (continued) Gene sequences used in this study

>ApnU
ATGGCGCCCACTCTGGAGAAGTATCACTTGCTCAAGCATGATGATGACTCGTCTTCGAGTCGTCGCGATTCCGGTGGAGAGGAA
CTCACCGGTCATGAATATCTTCAGCGACGAGAGGAGGAGGGATTCAGGTCATGGTTCTCGAGAAAAGTGACGCTCATCACTTTG
GCCATTCTGTTTGGTGCTTCTATTTTCTCGAACATTCTCATGGCACTGGCTTTGACGAGGAATTATGAGACTCATGACCACGAG
TCTCACGAGCATGTGCCTGCTCAGCGAATTACGAAACATGGTGAGTCAATCAATTTTGGTCCCCTTTGATTTACTTATCTCGCT
AATCAAAACAACGTTCATCTTTTAGCGGGATTGGCGGATGATCATCCCGTTGCATGGCACGCTTCCACCCCCTACAGCTCCGAA
AGTGTTGAGGAGAGTGATCATGCCTGGGTCAACATGGAAATCGACTATATCAACGTGGCACTTACCAAGGAAGATGCCAAGGCC
CAGGGTCTTCCTCCCTCAGCACCTTTCCCATGGGATCAAGACTACAGCGTCTACATGATTACCTCCATGCACAGCATGCACTGT
CTGGTATGTCTCCCTCACCACCAGTGATATTAAATTGCACATACATTTCTGACTCTTCACCCACAGAAATCTCTCCACCGCTCC
AACCTTGAATACCGCAAGGGAGTCAAACAAAGCTATCCAACCGAGCATCTCATCCACTGCCTCGACAACGTCCGTCAAGATATC
ATGTGCGCTGCCGACGACACCCCACGCTATATCCCCGTGGACGCTTCAGGCACGGCCACCACCGCCGTCGGCCAGTACCGCCAG
TGCAAAGACTGGGAGAAAATGCGCCAATGGAGCAAGGCAAATGACGCCTGCTTCAGCTATAACGAGCTGATTCGCCACGAACTA
CTAGAGGATCAGCCATTCCCACACGCCTTGCGTTTCTGCTCCAAGGATAGCAAGTTCTTGCCCGCTGTGCAGAAGTATTACAAC
ATGTCGAGTGACTGGGTGCCCACCAGGCCTGATCCACCTTACCCGGAAGTCGGTGTGAAGTTTGATGGGTATGATCAGTACACT
TGA

>ApnV
ATGTCCTCGGTCGCACAGCATTCTCGAGAGCAGGAGCCTCTGCTCAGCTCCACCTCGGAGACAGGTCGGGAGCCAGAGGACTAT
AACATCGCATGGAATCTCCTCAGTGGTCAGTGCTTGTCGTTTGGAAGGGACACTGAAATTTTTGCGAACGCGATGGTTGCTTGA
GCGCCAAGGGGGAAAAATAAATAAATCAAAAGAACAGTTGCTGACTCGGGTGGACTGGCACCTTAGACACCGCAAGTGTCGCGC
AAATTGGCATATGGGTGGTGAGTCCTTTACGCGCGCCACCTGCGAGTATACCCTTTTGGCAATGACCAGGGTCGCCATTGGGCA
CAAGATGGCAGATCCAGTCCAGGACGAGAATACTGACGCTCAAGTGAAACCCAGCTGGCCATCCTTGTTTGGTACAATGTCTTG
TCGATGCCGTTGATGCTCTTCACAGGCCACCCAGTAAGGAAGCCCAATCTTGGTCCATTCCCCGATAGCATCATTTCACTGACC
CGGGGACTTTCTCTCTTCCTCACCTGCCGTAGCTTTTAGCCAGTACCGCTCTTCTTCTACAGGTCCAGGCAGCACTCATCCTTC
AACCGACCGTGACACCCAACCAGAAACAGCGGGGTACTCGCATCCACTTTACCCTGCAACTCCTCAGCATCCTCTGTTTCGTGG
CTGCGTTCATCGTGATTGAGCTGAACAAGGGCACCCATCCTCACTTTACCTCCCCGCACGGCGTTCTCGGTCTGATCACGATCA
TTCTTGTGGTGCTCCAAGCCCTGGTCGGAGTCGTGCAGTACTTCTTCCCCATCACCGTCCTGGGCAGTGTGGACGCGGGCAAAC
GGATCTACAAGTACCATCGCTGGACCGGCTATGTTCTGCTACTGCTGGAGGTTGCGACTGTCCTGGCGGCGACCCAGACCGGCT
TCAACCTGAACGCCATTCACATTCCCACTTGGGGCGTGGTCGTGGCGGTCGCCCTCGTCGTGGTCGGAGTGGGCGCCCGCGTCA
AGGCGCGCAAGTTGGGACTGGGTGATGCATAA

>gpdAp
ACTCCGGTGAATTGATTTGGGTGACGGGAGAGACCCAAGAGGGGCCAGAATAATAAGAATGGGGAAGGCGAAGGTACCGCCTTT
GGGGTCCAGCCACGCGACTCCAACATGGAGGGGCACTGGACTAACATTATTCCAGCACCGGGATCACGGGCCGAAAGCGGCAAG
GCCGCGCACTGCCCCTCTTTTTGGGTGAAAGAGCTGGCAGTAACTTAACTGTACTTTCTGGAGTGAATAATACTACTACTATGA
AAGACCGCGATGGGCCGATAGTAGTAGTTACTTCCATTACATCATCTCATCCGCCCGGTTCCTCGCCTCCGCGGCAGTCTACGG
GTAGGATCGTAGCAAAAACCCGGGGGATAGACCCGTCGTCCCGAGCTGGAGTTCCGTATAACCTAGGTAGAAGGTATCAATTGA
ACCCGAACAACTGGCAAAACATTCTCGAGATCGTAGGAGTGAGTACCCGGCGTGATGGAGGGGGAGCACGCTCATTGGTCCGTA
CGGCAGCTGCCGAGGGGGAGCAGGAGATCCAAATATCGTGAGTCTCCTGCTTTGCCCGGTGTATGAAACCGGAAAGGACTGCTG
GGGAACTGGGGAGCGGCGCAAGCCGGGAATCCCAGCTGACAATTGACCCATCCTCATGCCGTGGCAGAGCTTGAGGTAGCTTTT
GCCCCGTCTGTCTCCCCGGTGTGCGCATTCGACTGGGCGCGGCATCTGTGCCTCCTCCAGGAGCGGAGGACCCAGTAGTAAGTA
GGCCTGACCTGGTCGTTGCGTCAGTCCAGAGGTTCCCTCCCCTACCCTTTTTCTACTTCCCCTCCCCCGCCGCTCAACTTTTCT
TTCCCTTTTACTTTCTCTCTCTTTCCTCTTCATCCATCCTCTCTTCATCACTTCCCTCTTCCCTTCATCCAATTCATCTTCCA
AGTGAGTCTTCCTCCCCATCTGTCCCTCCATCTTTCCCATCATCATCTCCCCTCCCAGCTCCTCCCCTCCTCTCGTCTCCTCAC
GAAGCTTGACTAACCATTACCCCGCCACATAGACACATCTAAACA

>glaAp
CCTGATCTTCCGAACTGGTCGTACCTGGCGACCTATGACTATGGCACCCCAGTTCTGGGGACCTTCCACGGAAGTGACCTGCTG
CAGGTGTTCTATGGGATCAAGCCAAACTATGCAGCTAGTTCTAGCCACACGTACTATCTGAGCTTTGTGTATACGCTGGATCCG
AACTCCAACCGGGGGGAGTACATTGAGTGGCCGCAGTGGAAGGAATCGCGGCAGTTGATGAATTTCGGAGCGAACGACGCCAGT
CTCCTTACGGATGATTTCCGCAACGGGACATATGAGTTCATCCTGCAGAATACCGCGGCGTTCCACATCTGATGCCATTGGCGG
AGGGGTCCGGACGGTCAGGAACTTAGCCTTATGAGATGAATGATGGACGTGTCTGGCCTCGGAAAAGGATATATGGGGATCATG
ATAGTACTAGCCATATTAATGAAGGGCATATACCACGCGTTGGACCTGCGTTATAGCTTCCCGTTAGTTATAGTACCATCGTTA
TACCAGCCAATCAAGTCACCACGCACGACCGGGGACGGCGAATCCCCGGGAATTGAAAGAAATTGCATCCCAGGCCAGTGAGGC
CAGCGATTGGCCACCTCTCCAAGGCACAGGGCCATTCTGCAGCGCTGGTGGATTCATCGCAATTTCCCCCGGCCCGGCCCGACA
CCGCTATAGGCTGGTTCTCCCACACCATCGGAGATTCGTCGCCTAATGTCTCGTCCGTTCACAAGCTGAAGAGCTTGAAGTGGC
GAGATGTCTCTGCAGGAATTCAAGCTAGATGCTAAGCGATATTGCATGGCAATATGTGTTGATGCATGTGCTTCTTCCTTCAGC
TTCCCCTCGTGCAGATGAGGTTTGGCTATAAATTGAAGTGGTTGGTCGGGGTTCCGTGAGGGGCTGAAGTGCTTCCTCCCTTTT
AGACGCAACTGAGAGCCTGAGCTTCATCCCCAGCATCATTACACCTCAGCA

>AmyBp
GATTAAAGGTGCCGAACGAGCTATAAATGATATAACAATATTAAAGCATTAATTAGAGCAATATCAGGCCGCGCACGAAAGGCA
ACTTAAAAAGCGAAAGCGCTCTACTAAACAGATTACTTTTGAAAAAGGCACATCAGTATTTAAAGCCCGAATCCTTATTAAGCG
CCGAAATCAGGCAGATAAAGCCATACAGGCAGATAGACCTCTACCTATTAAATCGGCTTCTAGGCGCGCTCCATCTAAATGTTC
TGGCTGTGGTGTACAGGGGCATAAAATTACGCACTACCCGAATCGATAGAACTACTCATTTTTATATAGAAGTCAGAATTCATG
GTGTTTTGATCATTTTAAATTTTTATATGGCGGGTGGTGGGCAACTCGCTTGCGCGGGCAACTCGCTTACCGATTACGTTAGGG
CTGATATTTACGTAAAAATCGTCAAGGGATGCAAGACCAAAGTAGTAAAACCCCGGAGTCAACAGCATCCAAGCCCAAGTCCTT
CACGGAGAAACCCCAGCGTCCACATCACGAGCGAAGGACCACCTCTAGGCATCGGACGCACCATCCAATTAGAAGCAGCAAAGC

GAAACAGCCCAAGAAAAAGGTCGGCCCGTCGGCCTTTTCTGCAACGCTGATCACGGGCAGCGATCCAACCAACACCCTCCAGAG
TGACTAGGGGCGGAAATTTAAAGGGATTAATTTCCACTCAACCACAAATCACAGTCGTCCCCGGTATTGTCCTGCAGAATGCAA
TTTAAACTCTTCTGCGAATCGCTTGGATTCCCCGCCCCTGGCCGTAGAGCTTAAAGTATGTCCCTTGTCGATGCGATGTATCAC
AACATATAAATACTAGCAAGGGATGCCATGCTTGGAGGATAGCAACCGACAACATCACATCAAGCTCTCCCTTCTCTGAACAAT
AAACCCCACAGAAGGCATTT

>PogpdAp
TTTGCTCCAGGAATACATGTGAGCTTACTGATTCTACTTTCGACATCTACCATTTATGTACGTACTGCAGTTGTACGTATTAGC
TGTTTCTAAGTGTATATGCGCCCATTGCCCTCTTTAGCCCAAATACATACGTCAATGTACATACTACGCAGAAATGCAGATAGG
ACCACTGCTATGCATCATCACTATCTTGTACAAGATGAACCTGGCAAACGAAGACGTTGAGTACCAAGCGACACCCACTATTAT
TACTTTCCAGTGAAACGACCACAACGTCCGGCGTTGTTGCATCGACTCTCAGGGTCGGATTTCAGAGCTTGCAGCCACATGAAG
TTTCCTGCACACAATTTTACTTGAACTCAAATGAGTCATTTGGACTTAGGTTGTTCATTCGGTGCTTGGTAGCCTCACTGCTTG
AACCGCAAATCGATGAACTTGCAAATCTATCAAGAAGTGCATGGCCATGGAGCTGTAGACTCATTAGTACAGGATCGATTATCC
CGTTCTGTGTGATAACTGTAATATGGAGTATATGTGTATGTGTATTAACACGGTAAAGCTGTACAGTACTGTCAAACGGACGAT
TGCTGTGAGCGTGGTACTGAATGTATGTATCATGGTATCGTGCCCACTTGATAGTTACAATGGATGTACCATGCGTTGATTTCA
TACCCCTCGAAAAAAGATTCAGGGTGAGTCGATGTTGCGTCAGCCTTGCCCAAAAACTACTTCTGCCCTCCATCTCCCCCTCCT
CCTGATCTTTTTTTTCTCTCCTTCCCCATCCTCTCCATTCATCCATCCGCATCACTTTCAAGGTAATTTTTCACCAGCTCCATT
CTGATTCTTAATCGCAGAGCTCTCTTTACTCCCGCCGCCTGCCCCTCTTCCCCTCTCCACGACCCCACCTTTCACTTTTCGCCT
CCCAGCTCCCAGCTCGCAGCTCGTCTGGCGCAGACAGCAAGCTTACAAGTGCATACAGAACACTTCAAACAATCGCAAAA

>PEgpdAp
GATTCGTCCAGGGCTTCCCAAGGTTGGAATATTGGGGACCCCAGTAGTTTAAACCCCTCAAGGCTCAAATAATTCAAATAATTC
AAAACAAAACAATCCAAAGAAAAATCTGCCAGATCCCGTGGCCCCATATGTTGAGATACCGGTGACTCCTTTCAGGGCGGAGGG
AGCCACCAAATTATATCACCATACTCGATCCGTCCACCGCAGGGGTTTGAACAATACGGTTGGCCCGCACAAGTGGAGGATCAC
AGGGATTTATGTAGTTAGAAGGGTTACTGGGATAATTATATTAATTCTATGAGTTTAACTGGTTGTATGTATATTTACTCCGTA
CAACGTAGTTAGTTTACATATACAACATAGGTACCTTACCTAAGGTATGTTTTTGCCCCTCAATGCCATTGCCCCTCCAACCCT
CTCAGCGTCATAGGTCTCCCCCTTAGTCAACCTCGGGAATCTCTTCAATTTAGTACCTAAGATTATCGTGATCATTTCCGTTGTA
CCCGGATAGACACCGAGTAGGCAAAATATTCAAGAGAAAGCTCCGAAAAATAGTAACAACCCGAAGAGCAGGTTGAGTTTGTCG
CATTCCTATTGGTAGACGGGAGCTACCGAGGGGCAGAGTAGGGCCCCAATATCGTGGCTCTTCCTGCTTTCCCCGGTGTATGAA
ACCGGAAATGGGTTGCTGGGCAGATGTGCAAGCGGCAACAGTCGGGAAATCTAGTTCGGTCTGATCTTTGTAGAAGGGGGGATG
TTGGCTGTGCTGGCTGTGAGTTGACAGTCCCCGTCCGTCTCTTCGGTGAGGGGTGATCTGTCCATTGCGTCAGTCATCCATTCA
TTAATCCTCCCCCTCCCCCAGATCTTTTTTTTCTTCTTCTCTTTCTTCTCCATCCTAGTTCCTCCCTTTCTTCAATTACTTCAC
TATTACAGTAAGCTCATCATTCTCTTATCATCTATCCCTCTTATACATGATCTAACAACTTCTAGTAAACCGCAATC

>ADH2p
GCAAAACGTAGGGGCAAACAAACGGAAAAATCGTTTCTCAAATTTTCTGATGCCAAGAACTCTAACCAGTCTTATCTAAAAATT
GCCTTATGATCCGTCTCTCCGGTTACAGCCTGTGTAACTGATTAATCCTGCCTTTCTAATCACCATTCTAATGTTTTAATTAAG
GGATTTTGTCTTCATTAACGGCTTTCGCTCATAAAAATGTTATGACGTTTTGCCCGCAGGCGGGAAACCATCCACTTCACGAGA
CTGATCTCCTCTGCCGGAACACCGGGCATCTCCAACTTATAAGTTGGAGAAATAAGAGAATTTCAGATTGAGAGAATGAAAAAA
AAAAAAAAAAAAAAAGGCAGAGGAGAGCATAGAAATGGGGTTCACTTTTTGGTAAAGCTATAGCATGCCTATCACATATAAATAG
AGTGCCAGTAGCGACTTTTTTCACACTCGAAATACTCTTACTACTGCTCTCTTGTTGTTTTTATCACTTCTTGTTTCTTCTTGG
TAAATAGAATATCAAGCTACAAAAAGCATACAATCAACTATCAACTATTAACTATATCGTAATAC

>MLS1p
CATTGGCCGATGAAGTTAGTCGACGGATAGAAGCGGTTGTCCCCTTTCCCGGCGAGCCGGCAGTCGGGCCGAGGTTCGGATAAA
TTTTGTATTGTGTTTTGATTCTGTCATGAGTATTACTTATGTTCTCTTTAGGTAACCCCAGGTTAATCAATCACAGTTTCATAC
CGGCTAGTATTCAAATTATGACTTTTCTTCTGCAGTGTCAGCCTTACGACGATTATCTATGAGCTTTGAATATAGTTTGCCGTG
ATTCGTATCTTTAATTGGATAATAAAATGCGAAGGATCGATGACCCTTATTATTATTTTTCTACACTGGCTACCGATTTAACTC
ATCTTCTTGAAAGTATATAAGTAACAGTAAAATATACCGTACTTCTGCTAATGTTATTTGTCCCTTATTTTTCTTTTCTTGTCT
TATGCTATAGTACCTAAGAATAACGACTATTGTTTTGAACTAAACAAAGTAGTAAAAGCACATAAAAGAATTAAGAAA

>PCK1p
ATAGGAAAAAACCGAGCTTCCTTTCATCCGGCGCGGCTGTGTTCTACATATCACTGAAGCTCCGGGTATTTTAAGTTATACAAG
GGAAAGATGCCGGCTAGACTAGCAAGTTTTAGGCTGCTTAACATTATGGATAGGCGGATAAAGGGCCCAAACAGGATTGTAAAG
CTTAGACGCTTCTGGTTGGACAATGGTACGTTTGTGTATTAAGTAAGGCTTGGCTGGGGATAGCAACATTGGGCAGAGTATAGA
AGACCACAAAAAAAAAGGTATATAAGGGCAGAGAAGTCTTTGTAATGTGTGTAACTTCTCTTCCATGTGTAATCAGTATTTCTAC
TTACTTCTTAAATATACAGAAGTAAGACAGATAACCAACAGCCTTTCCCAGATATACATATATATCTTTATTTCAGCTTAAACA
ATAATTATATTTGTTTAACTCAAAAATAAAAAAAAAAAACCAAACTCACGCAACTAATTATTCCATAATAAAATAACAAC

>PRM9t
GACAGAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATG
ATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTC
GTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAAT

Appendix A. (continued) Gene sequences used in this study

>CPS1t
GCGCAATGATTGAATAGTCAAAGATTTTTTTTTTTTAATTTTTTTTTTTTTAATTTTTTTTTTTTTTTCATAGAACTTTTTATTTA
AATAAATCACGTCTATATATGTATCAGTATAACGTAAAAAAAAAAACACCGTCAGTTAAACAAAACATAAATAAAAAAAAAAAG
AAGTGTCAAATCAAGTGTCAAAT

>SPG5t
CAAAGACGTTGTTTCATCGCGCTATTACCAAGAAGGTTACTTTACTTGTTCTTGCACATGGACGCACGTTGTGTGTTCATATAT
ATATATATATATATATATATATATTTGTGCTTGTTTTCATTGTCTCTATAGTTAATACATTCTATTTTTATCGTTATATTTGCATT
CTCTTCGCATAAAAACTTCATGAAAATTCGGCAGAAAATAAGC

>TfG8H
ATGGGACAGTCCCGCCGACCCCACACCGTCTACCTCGATCCGGCCAAAGGAGTCGACATCCCCGCGCAGCGACGCGAACTCCTC
GACAAAGGCCCGGTGGTACGCGTGGCTTTCCCCGGCAATTTGGAGGTCTGGGCACTGACTCATGATGCGCCTTTACGTAATGCC
TTAGCAGACGAGTCTGTTTTCGTTCGCGGGTGGCGCAACTGGCGCGCGCTTATGGCAGGGGAAGTTGACCCTACTCATCCGGTC
GCCAACATGTTGCGCGTCGAGAGCATGCTGGCCCGCTCTGGAGCGGACCATAAGCGCATGCGTGGGTTGGTGCAAGCTGCGTTT
ACCCGCCGCCGTGTGGAAGCGCTTCGCCCGCGCATCGAAGAAATTACTAATGAACTGCTTGATCGTATGGCAGAGTCGGACGGT
GTAGTTGACCTGAAAGCTGCTTATAGCTTCCCGTTACCTATTCGCGTGATCTCTGAGCTTTTAGGGTTAAACGAGGAAGACCAT
CTTACTCTTCAGACTCTTGTGACACGCACACTGAGTGGCACCGATCCGGAAGCAAATGCAGACGCCTTCACATTTGTAGCTTCC
CTGATTGAAGCGAAGCGTAAAAATCTTGATGACGGTCTGATTTCTGCTATGATTGAGGCCCGCGCGGAAGATGGAGATCGCTTA
TCTGAGACTGAGCTTATTCACAATACTCTTCTTTTAATCATCGGTGGTTTCGAAACGACGATGGGGATGATTAGCAATTCAGTA
CAGCTGTTGTTGACGCACCCCGACCAACTTCATCTTTTACGTACCGGACAGGCCAGTTGGGGAGAATGCCATCGAGGAATGTCTT
CGTTTTGAGTCAGCTGTAGTCATGTTACCTTTCCTTTATACGACACGCGACGTTGAAATCGACGGGATTACTATCCCCGCGGGA
GACGCTGTATTAATTGGCTTTGGTCCCGCTAACCGCGATCCACAGGCGTACGACGACCCTGACCGCTTCGATATCACCCGCCCA
CGCCCGCGCCATCTTGCGTTCGGACACGGAGCACACCTGTGCCTGGGAGCCGCATTAGCGCGTCTGGAACTGCTGATCGCACTG
CCCGCCTTATTTGAACGTTTTCCAGATATCACGTTAGTCGGTGAGGCTCCGCCAACCCCTACGGTTTTCATGAATCACCCACTT
AGCCGCCCCGTTTTACTTCGTCCGAAACCATAA

>FpR
ATGGCTGATTGGGTAACAGGCAAAGTCACTAAAGTGCAGAACTGGACCGACGCCCTGTTTAGTCTCACCGTTCACGCCCCCGTG
CTTCCGTTTACCGCCGGGCAATTTACCAAGCTTGGCCTTGAAATCGACGGCGAACGCGTCCAGCGCGCCTACTCCTATGTAAAC
TCGCCCGATAATCCCGATCTGGAGTTTTACCTGGTCACCGTCCCCGATGGCAAATTAAGCCCACGACTGGCGGCACTGAAACCA
GGCGATGAAGTGCAGGTGGTTAGCGAAGCGGCAGGATTCTTTGTGCTCGATGAAGTGCCGCACTGCGAAACGCTATGGATGCTG
GCAACCGGTACAGCGATTGGCCCTTATTTATCGATTCTGCAACTAGGTAAAGATTTAGATCGCTTCAAAAATCTGGTCCTGGTG
CACGCCGCACGTTATGCCGCCGACTTAAGCTATTTGCCACTGATGCAGGAACTGGAAAAACGCTACGAAGGAAAACTGCGCATT
CAGACGGTGGTCAGTCGGGAAACGGCAGCGGGGTCGCTCACCGGACGGATACCGGCATTAATTGAAAGTGGGGAACTGGAAAGC
ACGATTGGCCTGCCGATGAATAAAGAAACCAGCCATGTGATGCTGTGCGGCAATCCACAGATGGTGCGCGATACACAACAGTTG
CTGAAAGAGACCCGGCAGATGACGAAACATTTACGTCGCCGACCGGGCCATATGACAGCGGAGCATTACTGGTAA

>YkuN
ATGGCGAAGGCATTGATCACATACGCTTCAATGTCTGGCAATACCGAGGACATTGCTTTCATCATCAAGGACACCCTTCAGGAA
TACGAGCTTGACATCGACTGTGTTGAAATTAACGATATGGATGCATCCTGCCTTACATCTTATGATTATGTGTTAATCGGGACA
TATACATGGGGAGACGGTGACTTACCATACGAAGCCGAGGACTTCTTTGAAGAGGTAAAACAGATCCAACTGAATGGACTGAAG
ACCGCTTGCTTTGGGTCGGGCGATTACTCATACCCTAAATTTTGCGAAGCTGTTAACTTGTTTAACGTAATGTTGCAGGAAGCT
GGTGCGGCAGTATATCAGGAAACTTTGAAAATCGAGTTGGCACCTGAGACCGACGAAGATGTTGAATCATGTCGTGCCTTTGCT
CGTGGATTTTTAGCATGGGCAGACTACATGAACAAAGAGAAGATCCACGTCAGTTAA

>GOR
ATGACTAAAACTAATTCTCCAGCCCCATCTGTCATTACTTGCAAGGCTGCTGTCGTTTGGAAATCCGGTGAACCACCAAAGGTC
GAAGAGATCCAAGTTGATCCACCCAAGGCTTCTGAAGTTCGCATTAAGATGTTGTGTGCTTCCTTGTGCCACACCGATTTCTTG
GCTTGTAATGGTCTGCCAGTTCCATTGTTTCCCAGAATTCCAGGTCACGAAGGTGTTGGTATGATCGAATCTGTCGGTGAAAAC
GTCACCAACTTGAAGGAAGGTGACATTGTCATGCCATTGTACTTGGGTGAGTGTGGCGAATGCTTGAATTGCAAGTCCGGCAGG
ACTAACTTGTGTCATAAGTATCCGTTGGGTTTTTCTGGCCTGTTGTTGGATGGCACTTCCAGGATGAGCATTGGCGAACAAAAA
GTCTACCACCACTTCTCTTGTTCCACCTGGTCTGAATACATTGTTATTGAGGCCGCCTACGCAGTTAAAGTTGACCCAAGGGTT
AGCTTGCCACATGCTTCTTTCCTGTGTTGCGGTTTTACTACTGGCTTTGGCGCCACTTGGAGAGATGTTAATGTTGTCAAAGGC
TCTACTGTCGCTGTTTTGGGTTTAGGTGCTGTCGGTTTGGGTGCTGTTCAAGGCGCTAAATCTCAAGGTGCCTCCAGGATCATT
GGTTTAGACATTAACGATAAGAAGAGGGAGAAAGGCGAAGCTTTCGGCATGACCGAATTCATCAACCCCAAGGGCTCCAATAAG
TCCATCTCCGAATTGATCAACGAAGCTACTGGTGGTCTAGGTTTGGACTACGTTTATGAATGCACTGGTGTCCCAGCTCTGTTG
AACGAAGCCATTGAGTCCTCTAAAGTTGGTCTGGGTACTGCCGTCTTGATTGGTGCTGGTCTAGAAACCTCTGGTGAAATCAAA
TTCATTCCCCTGTTGTGCGGCAGAACTGTTAAAGGTTCCATTTACGGTGGTGTTAGGCCAAAGTCCGACTTGCCAACTCTGATT
GAGAAGTGCATTAACAAGGAGATTCCAATGGACGAGCTGATGACCCATGAGGTGTCTCTGTCCGAGATCAACAAGGGTTTCGAG
TACTTGAAGCACCCAGACTGTGTCAAAGTTGTTATTAAGTTCTAA

Appendix A. (continued) Gene sequences used in this study

>ISY
ATGTCCTGGTGGTGGAAAAGGTCTATTGGTGCTGGCAAAAACTTGCCAAACCAAAACAAGGAAAACGGTGTCTGCAAGTCTTAC
AAATCTGTCGCCTTGGTCGTCGGTGTTACTGGTATTGTTGGTTCTTCTCTGGCTGAGGTTTTGAAGTTGCCAGATACTCCAGGT
GGTCCATGGAAAGTTTATGGTGTTGCTAGAAGACCATGTCCAGTCTGGTTGGCTAAGAAGCCAGTCGAGTACATCCAGTGTGAC
GTCTCCAATAACCAAGAAACCATTTCTAAGCTGTCTCCCCTGAAAGACATCACTCACATCTTCTATGTCTCCTGGATTGGCTCT
GAGGATTGCCAGACTAATGCCACCATGTTCAAGAACATCTTGAACTCCGTTATCCCAAATGCTTCCAACTTGCAGCACGTCTGC
CTACAAACCGGCATTAAGCATTACTTCGGCATTTTCGAAGAGGGTTCCAAAGTCGTTCCACATGATTCCCCCTTTACCGAAGAT
TTGCCACGCTTGAACGTCCCAAACTTTTATCACGACCTGGAAGACATTTTGTACGAGGAGACAGGCAAAAATAACCTAACCTGG
TCCGTTCACAGGCCAGCTTTGGTTTTCGGTTTTTCCCCATGCTCCATGATGAATATCGTCTCTACTCTGTGCGTCTACGCTACT
ATTTGCAAGCATGAGAACAAGGCTCTGGTTTACCCAGGTTCCAAGAATTCCTGGAATTGCTATGCTGATGCTGTCGATGCTGAC
TTGGTTGCTGAGCATGAAATTTGGGCTGCTGTTGATCCAAAGGCCAAAAACCAGGTTCTGAATTGCAACAACGGCGACGTCTTC
AAATGGAAACATATCTGGAAGAAGCTGGCTGAAGAGTTTGGTATCGAGATGGTCGGTTATGTTGAAGGCAAAGAACAGGTCAGC
CTGGCCGAATTGATGAAAGATAAGGATCAAGTCTGGGACGAAATCGTCAAGAAAAACAACCTGGTGCCAACTAAGTTGAAGGAG
ATTGCCGCCTTCTGGTTTGCCGATATCGCCTTTTGCTCTGAAAACTTGATCTCTTCCATGAACAAGTCCAAGGAGCTGGGTTTC
CTAGGCTTCAGGAACTCTATGAAGTCTTTCGTCTCCTGTATCGACAAGATGAGAGACTACAGATTCATTCCATAA

>NmMLPL
ATGGCGTCGAAACTGGAGATCGAAATCGAACTTAAATCGGACGTGGAAAAAATGTGGAAACATTTCAAGGAGTTTACAAAACTG
TTCCCGAAGGCTTGCCCGCATTTGTATGAAAAGATTGATGTGATTGAGGGGGATGGCATCTCGGTGGGGACTATTTTCGTGTCA
ACGTTGAAGCCTACAGAGTTAAATCCAGTAGTGATGGTCACGAAAGAAAAGATTGATTTTTTAGATGACGAAAATAAAATGTTA
CGCTATTCTTACATGGAGGGTGAGATTTTAAAGAACTACAAGAACCTTCGTGGGACGGTACACATGTCCAGCTCGAAGTCCGGT
GGAACTATCTTCAAGTACTCGGGTGAATTTGAAAAGGCGAACGAACAGGTCCCCGACCCCGTATTCTTTAAGGACTTCATGGTC
ATTGTCTTTCAAGGGCTTGACGACTACATCTTGAAGGGTATGAATCACACTTGTCAAAATTAG

>NEPS1
ATGGCATCCACAGCGAACCCGATGCAGGTAATGAAAAAAAAGTTAGAAGGCAAAGTAGTGATTGTAACTGGCGGTGCATCAGGT
ATAGGACAAACCGCTGCTCGTGTGTTCGCACAGCATGGTGCGAGGGCGGTAGTCATCGCAGACATACAGTCAGAGGTGGGGAAA
TCAGTTGCTAAAAGTATCGGTGATCCTTGCTGTTATGTTCAATGTGATGTATCCGACGAAGAAGAAGTTAAGAGTATGATTGAG
TGGACGGCGTCAGCTTACGGCGGGTTAGATATGATGTTCTCTAATGTGGGGATAATGTCCAAAAGTGCCCAAACTGTAATGGAT
TTGGATCTTCTGGAATTTGACAAAGTGATGAGAGTCAACGCAAGAGGAATGGCAGCATGCCTGAAACACGCAGCCCGTAAAATG
GTCGAGTTGGGCACGAGGGGGACTATAATCTGCACAACAACTCCGCTGTCAAGTCGTGGTGGCCAATCAATGACAGACTATGCG
ATGTCAAAACACGCGGTTATGGGTCTGGTCAGGTCAGCCAGCATCCAATTAGGTGCCCACGGAATTAGAGTGAATTGTGTTACT
CCTTCAGTCGTGTTAACGCCGTTAGCACAAAGAATGGGTCTGGCGACGCCAGATGATTTCCATACGCATTTTGGTAATTTTACT
TCACTAAAGGGCGTGTACCTTACCCCAGAGCAAGTAGCTGAGGCGGTTGTGTATCTTGCCTCAGATGACGCCGCTTTCATCACC
GGGCATGACCTGGTACTTGATGGGGGCCTTCTATGTTTACCTTTCTTCGCACCGTCATAG

>NoxE
ATGAAAATCGTAGTTATCGGTACAAACCACGCAGGCATTGCTACAGCGAATACATTACTTGAACAATATCCCGGGCATGAAATT
GTCATGATTGACCGTAATAGCAACATGAGTTATCTAGGTTGTGGCACAGCAATTTGGGTTGGAAGACAAATTGAAAAACCAGAT
GAATTATTTTATGCCAAAGCAGAGGATTTTGAGGCAAAAGGGGTAAAAATTTTGACTGAAACAGAAGTTTCAGAAATTGATTTT
GCTAATAAGAAAGTTTATGCAAAAACTAAATCTGATGATGAAATAATTGAAGCTTACGACAAGCTTGTTTTAGCAACAGGTTCA
CGTCCAATTATTCCTAATCTACCAGGCAAAGACCTTAAGGGAATTCATTTTCTGAAACTTTTTCAAGAAGGTCAAGCAATTGAC
GCAGAATTTGCCAAAGAAAAAGTCAAGCGTATCGCAGTCATTGGTGCAGGATATATCGGTACAGAGATTGCGGAAGCAGCTAAA
CGTCGGGGTAAAGAAGTTCTTCTCTTTGACGCTGAAAATACTTCACTTGCATCATATTATGATGAAGAATTTGCCAAAGGAATG
GATGAAAACCTTGCTCAACATGGAATTGAACTTCATTTTGGAGAACTGGCCAAAGAATTTAAAGCGAATGAGGAAGGTTATGTA
TCACAAATCGTAACCAACAAGGCGACTTATGATGTTGATCTTGTCATCAATTGTATTGGTTTTACTGCCAACAGTGCCTTGGCA
AGTGATAAGTTAGCTACCTTCAAAAATGGCGCAATCAAGGTGGATAAGCATCAACAAAGTAGTGATCCAGATGTTTACGCGGTA
GGTGATGTTGCGACAATTTATTCTAATGCCTTGCAAGATTTTACTTATATCGCTCTTGCCTCAAACGCTGTTCGGTCAGGAATT
GTCGCAGGACACAATATTGGTGGAAAAGAATTAGAATCTGTTGGTGTTCAAGGTTCTAATGGTATTTCGATTTTTGGTTACAAT
ATGACTTCTACAGGACTTTCTGTTAAAGCTGCTAAAAAATTAGGTTTAGAAGTTTCATTTAGTGATTTTGAAGATAAACAAAAA
GCTTGGTTTCTTCATGAAAACAACGATAGTGTGAAAATTCGTATCGTATATGAGACAAAAAGTCGCAGAATTATTGGAGCACAA
CTTGCTAGTAAAAGTGAGATAATTGCAGGAAATATAAATATGTTCAGTTTAGCGATTCAAGAGAAAAAAACAATTGATGAACTA
GCTTTGCTTGATTTATTCTTTCTCCCCCACTTCAACAGTCCATATAATTATATGACAGTTGCAGCTTTGAATGCCAAATAA

>FumC
ATGAATACAGTACGCAGCGAAAAAGATTCGATGGGGGCGATTGATGTCCCGGCAGATAAGCTGTGGGGCGCACAAACTCAACGC
TCGCTGGAGCATTTCCGCATTTCGACGGAGAAAATGCCCACCTCACTGATTCATGCGCTGGCGCTAACCAAGCGTGCAGCGGCA
AAAGTTAATGAAGATTTAGGCTTGTTGTCTGAAGAGAAAGCGAGCGCCATTCGTCAGGCGGCGGATGAAGTACTGGCAGGACAG
CATGACGACGAATTCCCGCTGGCTATCTGGCAGACCGGCTCCGGCACGCAAAGTAACATGAACATGAACGAAGTGCTGGCTAAC
CGGGCCAGTGAATTACTCGGCGGTGTGCGCGGGATGGAACGTAAAGTTCACCCTAACGACGACGTGAACAAAAGCCAAAGTTCC
AACGATGTCTTTCCGACGGCGATGCACGTTGCGGCGCTGCTGGCGCTGCGCAAGCAACTCATTCCTCAGCTTAAAACCCTGACA
CAGACACTGAATGAGAAATCCCGTGCTTTTGCCGATATCGTCAAAATTGGTCGTACTCACTTGCAGGATGCCACGCCGTTAACG
CTGGGGCAGGAGATTTCCGGCTGGGTAGCGATGCTCGAGCATAATCTCAAACATATCGAATACAGCCTGCCTCACGTAGCGGAA
CTGGCTCTTGGCGGTACAGCGGTGGGTACTGGACTAAATACCCATCCGGAGTATGCGCGTCGCGTAGCAGATGAACTGGCAGTC
ATTACCTGTGCACCGTTTGTTACCGCGCCGAACAAATTTGAAGCGCTGGCGACCTGTGATGCCCTGGTTCAGGCGCACGGCGCG
TTGAAAGGGTTGGCTGCGTCACTGATGAAAATCGCCAATGATGTCCGCTGGCTGGCCTCTGGCCCGCGCTGCGGAATTGGTGAA
ATCTCAATCCCGGAAAATGAGCCGGGCAGCTCAATCATGCCGGGGAAAGTGAACCCAACACAGTGTGAGGCATTAACCATGCTC
TGCTGTCAGGTGATGGGGAACGACGTGGCGATCAACATGGGGGGCGCTTCCGGTAACTTTGAACTGAACGTCTTCCGTCCAATG
GTGATCCACAATTTCCTGCAATCGGTGCGCTTGCTGGCAGATGGCATGGAAAGTTTTAACAAACACTGCGCAGTGGGTATTGAA
CCGAATCGTGAGCGAATCAATCAATTACTCAATGAATCGCTGATGCTGGTGACTGCGCTTAACACCCACATTGGTTATGACAAA
GCCGCCGAGATCGCCAAAAAAGCGCATAAAGAAGGGCTGACCTTAAAAGCTGCGGCCCTTGCGCTGGGGTATCTTAGCGAAGCC
GAGTTTGACAGCTGGGTACGGCCAGAACAGATGGTCGGCAGTATGAAAGCCGGGCGTTAA

>MaeB
ATGGATGACCAGTTAAAACAAAGTGCACTTGATTTCCATGAATTTCCAGTTCCAGGGAAAATCCAGGTTTCTCCAACCAAGCCT
CTGGCAACACAGCGCGATCTGGCGCTGGCCTACTCACCAGGCGTTGCCGCACCTTGTCTTGAAATCGAAAAAGACCCGTTAAAA
GCCTACAAATATACCGCCCGAGGTAACCTGGTGGCGGTGATCTCTAACGGTACGGCGGTGCTGGGGTTAGGCAACATTGGCGCG
CTGGCAGGCAAACCGGTGATGGAAGGCAAGGGCGTTCTGTTTAAGAAATTCGCCGGGATTGATGTATTTGACATTGAAGTTGAC
GAACTCGACCCGGACAAATTTATTGAAGTTGTCGCCGCGCTCGAACCAACCTTCGGCGGCATCAACCTCGAAGACATTAAAGCG
CCAGAATGTTTCTATATTGAACAGAAACTGCGCGAGCGGATGAATATTCCGGTATTCCACGACGATCAGCACGGCACGGCAATT
ATCAGCACTGCCGCCATCCTCAACGGCTTGCGCGTGGTGGAGAAAAACATCTCCGACGTGCGGATGGTGGTTTCCGGCGCGGGT
GCCGCAGCAATCGCCTGTATGAACCTGCTGGTAGCGCTGGGTCTGCAAAAACATAACATCGTGGTTTGCGATTCAAAAGGCGTT
ATCTATCAGGGCCGTGAGCCAAACATGGCGGAAACCAAAGCCGCATATGCGGTGGTGGATGACGGCAAACGTACCCTCGATGAT
GTGATTGAAGGCGCGGATATTTTCCTGGGCTGTTCCGGCCCGAAAGTGCTGACCCAGGAAATGGTGAAGAAAATGGCTCGTGCG
CCAATGATCCTGGCGCTGGCGAACCCGGAACCGGAAATTCTGCCGCCGCTGGCGAAAGAAGTGCGTCCGGATGCCATCATTTGC
ACCGGTCGTTCTGACTATCCGAACCAGGTGAACAACGTCCTGTGCTTCCCGTTCATCTTCCGTGGCGCGCTGGACGTTGGCGCA
ACCGCCATCAACGAAGAGATGAAACTGGCGGCGGTACGTGCGATTGCAGAACTCGCCCATGCGGAACAGAGCGAAGTGGTGGCT
TCAGCGTATGGCGATCAGGATCTGAGCTTTGGTCCGGAATACATCATTCCAAAACCGTTTGATCCGCGCTTGATCGTTAAGATC
GCTCCTGCGGTCGCTAAAGCCGCGATGGAGTCGGGCGTGGCGACTCGTCCGATTGCTGATTTCGACGTCTACATCGACAAGCTG
ACTGAGTTCGTTTACAAAACCAACCTGTTTATGAAGCCGATTTTCTCCCAGGCTCGCAAAGCGCCGAAGCGCGTTGTTCTGCCG
GAAGGGGAAGAGGCGCGCGTTCTGCATGCCACTCAGGAACTGGTAACGCTGGGACTGGCGAAACCGATCCTTATCGGTCGTCCG
AACGTGATCGAAATGCGCATTCAGAAACTGGGCTTGCAGATCAAAGCGGGCGTTGATTTTGAGATCGTCAATAACGAATCCGAT
CCGCGCTTTAAAGAGTACTGGACCGAATACTTCCAGATCATGAAGCGTCGCGGCGTCACTCAGGAACAGGCGCAGCGGGCGCTG
ATCAGTAACCCGACAGTGATCGGCGCGATCATGGTTCAGCGTGGGGAAGCCGATGCAATGATTTGCGGTACGGTGGGTGATTAT
CATGAACATTTTAGCGTGGTGAAAAATGTCTTTGGTTATCGCGATGGCGTTCACACCGCAGGTGCCATGAACGCGCTGCTGCTG
CCGAGTGGTAACACCTTTATTGCCGATACATATGTTAATGATGAACCGGATGCAGAAGAGCTGGCGGAGATCACCTTGATGGCG
GCAGAAACTGTCCGTCGTTTTGGTATTGAGCCGCGCGTTGCTTTGTTGTCGCACTCCAACTTTGGTTCTTCTGACTGCCCGTCG
TCGAGCAAAATGCGTCAGGCGCTGGAACTGGTCAGGGAACGTGCACCAGAACTGATGATTGATGGTGAAATGCACGGCGATGCA
GCGCTGGTGGAAGCGATTCGCAACGACCGTATGCCGGACAGCTCTTTGAAAGGTTCCGCCAATATTCTGGTGATGCCGAACATG
GAAGCTGCCCGCATTAGTTACAACTTACTGCGTGTTTCCAGCTCGGAAGGTGTGACTGTCGGCCCGGTGCTGATGGGTGTGGCG
AAACCGGTTCACGTGTTAACGCCGATCGCATCGGTGCGTCGTATCGTCAACATGGTGGCGCTGGCCGTGGTAGAAGCGCAAACC
CAACCGCTGTAA

Appendix B. Plasmids and primer sequences used in this study

Construct	PRIMERS USED FOR CLONING	Amplicon
A, nidulans expression constructs		
pAUB60001 - pYTU – glaAp.HarA	TTCATCCCCAGCATCATTACACCTCAGCATTAATATGCCTGAATCAATTGCCATCATTTC	HarA
	ATTCGAAGCTCAAGGACTCGCTT	
	ATGGTGATTCCGGAGGGTACG	
	ACTCTCAGGTGACGATTCAAGAAGG	
	ATTGCACAAAAAGGTGCCTAGTTACTTG	
	ATGGGCACCAGGAGATCCG	
	ATGACATTGCAGTCAAAGATGGATTCAG	
	AGTGGAGGACATACCCGTAATTTTCTGGGCATTTGTGGGATTATCGGCAGAAGTTTGAAG	
pAUB60002 - pYTR – gpdAp.HarE	CCATTACCCCGCCACATAGACACATCTAAACATTAATATGACCATTATCATTCCGGGCAG	HarE
	GCTAAAGGGTATCATCGAAAGGGAGTCATCCAATTTTAGCTTGCAGGTTCGTCAATAGGT	
pAUB60003 - pYTP – AmyBp.HarG	TTCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGGCTCTGCAATCATGGGGAG	HarG
	GAGACCCAACAACCATGATACCAGGGGATTTGAAACTGCCTATGAAGTTATTGCGAAAGC	
pAUB60004 - pYTR – gpdAp.HarE – glaAp.HarF	CCATTACCCCGCCACATAGACACATCTAAACATTAATATGACCATTATCATTCCGGGCAG	HarE
	GCTAAAGGGTATCATCGAAAGGGAGTCATCCAATTTTAGCTTGCAGGTTCGTCAATAGGT	
	CCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGACCTCACAAACTCTCGTTCC	HarF
	CTCTTGGGTCTCTCCCGTCACCCAAATCAATTCACCGGAGTCCATCAAGCAAGACGGGCG	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	GCTGAGGTGTAATGATGCTGGG	
pAUB60005 – pYTP - gpdAp.HarD - amyBp.HarG	AGTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGCACTCCGGTGAATTGATTTGGGTG	gpdAp
	TGTTTAGATGTGTCTATGTGGCGGG	
	TTGACTAACCATTACCCCGCCACATAGACACATCTAAACAATGGTAAACGCCACTTGGGC	HarD
	TATCATTTATAGCTCGTTCGGCACCTTTAATCGTCAGGTTACATTCAGGGGCTAAATATG	
	TTCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGGCTCTGCAATCATGGGGAG	HarG
	GAGACCCAACAACCATGATACCAGGGGATTTGAAACTGCCTATGAAGTTATTGCGAAAGC	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGT	
pAUB60006 – pYTR – glaAp.HarC – gpdAp.HarE	GTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGCCCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	AGCCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGCTGTTCCATCTGGAGATG	HarC
	TGGGTCTCTCCCGTCACCCAAATCAATTCACCGGAGTCCAGCTGCAGTTACTTTCAGTCC	
	ACTCCGGTGAATTGATTTGGGTG	gpdAp
	TGTTTAGATGTGTCTATGTGGCGGG	
	CCATTACCCCGCCACATAGACACATCTAAACATTAATATGACCATTATCATTCCGGGCAG	HarE
	GCTAAAGGGTATCATCGAAAGGGAGTCATCCAATTTTAGCTTGCAGGTTCGTCAATAGGT	
pAUB60007 – pYTP – glaAp.HarB – gpdAp.HarD – amyBp.HarG	GTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGCCCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	CCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGCGAACGTTAATGTGAAGGAG	HarB
	ATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCCATGCCATGCCAGAGGATGAC	
	AGTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGCACTCCGGTGAATTGATTTGGGTG	gpdAp

	TGTTTAGATGTGTCTATGTGGCGGG	
	TTGACTAACCATTACCCCGCCACATAGACACATCTAAACAATGGTAAACGCCACTTGGGC	HarD
	TATCATTTATAGCTCGTTCGGCACCTTTAATCGTCAGGTTACATTCAGGGGCTAAATATG	
	TTCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGGCTCTGCAATCATGGGGAG	HarG
	GAGACCCAACAACCATGATACCAGGGGATTTGAAACTGCCTATGAAGTTATTGCGAAAGC	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGT	
pAUB60008 – pYTR – amyBp.HarC – gpdAp.HarE - glaAp.HarF	AGTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGCACTCCGGTGAATTGATTTGGGTG	gpdAp - HarE
	TCGAATTCCTGCAGCCCGGGGGATCCTTAATGAGGAAAAGCCCATTGATAAAAGCATCAT	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	CCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGACCTCACAAACTCTCGTTCC	HarF
	CTCTTGGGTCTCTCCCGTCACCCAAATCAATTCACCGGAGTCCATCAAGCAAGACGGGCG	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGT	
	CTCGCGGGTGTTCTTGACGATGGCATCCTGCGGCCGCGATTAAAGGTGCCGAACGAGCTA	HarC
	TCGCCAGGTACGACCAGTTCGGAAGATCAGGGAAACTGCCTATGAAGTTATTGCGAAAGC	
pAUB60009 – pYTP – glaAp.HarB – amyBp.HarG	GTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGCCCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	CCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGCGAACGTTAATGTGAAGGAG	HarB
	ATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCCATGCCATGCCAGAGGATGAC	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGT	
	TTCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGGCTCTGCAATCATGGGGAG	HarG
	GAGACCCAACAACCATGATACCAGGGGATTTGAAACTGCCTATGAAGTTATTGCGAAAGC	
pAUB70001 - pYTU - glaAp.ApnA	TTCATCCCCAGCATCATTACACCTCAGCATTAATATGCAAGACTTGATTGCTATTGTTGG	ApnA
	TCACGGTGAGGATAGTACTCCAGTC	
	ATGTCTCTACAGGCGTCACTGAAG	
	TCGAGATTCCAAACACGACAACGC	
	ATCCGTCATCACCAGTCCCTTC	
	TTGGGTCTCTCCCGTCACCCAAATCAATTCACCGGAGTCTTGCTCTAAGCGCCCCACTAC	
pAUB70002 - pYTR - gpdAp.ApnE – glAp.ApnD – AmyBp.ApnC	CTCGCGGGTGTTCTTGACGATGGCATCCTGCGGCCGCactccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	tgactaaccattaccccgccacatagacacatctaaacaatgatccctccaaggcaacag	ApnE
	GTCATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGggttcctggctcgaggttctc	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	GCTTCATCCCCAGCATCATTACACCTCAGCAatgagttccatcaattttcaatgggagga	HarD
	ATATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCctgtggggactccctgtcg	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCCCTTCTCTGAACAATAAACCCCACAGAAGGCATTTatgatcgaccccaagacaagcc	ApnC
	TGCTAAAGGGTATCATCGAAAGGGAGTCATCCAATTTggagaatgccttatccagtggag	
	TGAACAATAAACCCCACAGAAGGCATTTTTAATatgagtctgcttcacatgttttcgaag	ApnG

pAUB70003 - pYTP – AmyBp.ApnG – gpdAp.ApnB	ctctcccgtcacccaaatcaattcaccggagtcacctgatagggcaagatctctcaaaaa	
	actccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	tgactaaccattaccccgccacatagacacatctaaacaatggcggctcttagtcttgca	ApnB
	GAGACCCAACAACCATGATACCAGGGGATTTAAATtgcccaagccaaggggatgaaggct	
pAUB70004 - pYTP – AmyBp.ApnG – gpdAp.ApnB – glaAp.ApnU	TGAACAATAAACCCCACAGAAGGCATTTTTAATatgagtctgcttcacatgttttcgaag	ApnG
	ctctcccgtcacccaaatcaattcaccggagtcacctgatagggcaagatctctcaaaaa	
	actccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	tgactaaccattaccccgccacatagacacatctaaacaatggcggctcttagtcttgca	ApnB
	AGTCATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGaagccaaggggatgaaggct	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGG	
	GAGAGCCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAatggcgcccactctggaga	ApnU
	ACCCAACAACCATGATACCAGGGGATTTAAATtgttgaatcggtggtaattccttttgat	
pAUB70005 - pYTU - glaAp.ApnA – gpdAp.HarE	TTCATCCCCAGCATCATTACACCTCAGCATTAATatgcaagacttgattgctattgttgg	ApnA
	tcacggtgaggatagtactccagtc	
	atgtctctacaggcgtcactgaag	
	tcgagattccaaacacgacaacgc	
	atccgtcatcaccagtcccttc	
	ttgggtctctcccgtcacccaaatcaattcaccggagtcttgctctaagcgccccactac	
	actccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	gactaaccattaccccgccacatagacacatctaaacaatgatccctccaaggcaacaga	ApnE
	AACACAGTGGAGGACATACCCGTAATTTTCTGGGCATTTggttcctggctcgaggttctc	
pAUB70006 - pYTR - glAp.ApnD – AmyBp.ApnC – gpdAp.ApnV	aaccattaccccgccacatagacacatctaaacaTTAATatgtcctcggtcgcacagc	ApnV
	CATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGagtccgactttcatgtggagtcg	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	GCTTCATCCCCAGCATCATTACACCTCAGCAatgagttccatcaattttcaatgggagga	ApnD
	ATATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCctgtggggactccctgtcg	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCCCTTCTCTGAACAATAAACCCCACAGAAGGCATTTatgatcgaccccaagacaagcc	ApnC
	TGCTAAAGGGTATCATCGAAAGGGAGTCATCCAATTTggagaatgccttatccagtggag	
pAUB70007 - pYTP – AmyBp.ApnG – gpdAp.ApnB – glaAp.ApnU – PogpdAp.ApnT	TGAACAATAAACCCCACAGAAGGCATTTTTAATatgagtctgcttcacatgttttcgaag	ApnG
	ctctcccgtcacccaaatcaattcaccggagtcacctgatagggcaagatctctcaaaaa	
	actccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	tgactaaccattaccccgccacatagacacatctaaacaatggcggctcttagtcttgca	ApnB
	GAGACCCAACAACCATGATACCAGGGGATTTAAATtgcccaagccaaggggatgaaggct	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGG	
	GAGAGCCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAatggcgcccactctggaga	ApnU

	atcagtaagctcacatgtattcctggagcaaatgttgaatcggtggtaattccttttgat	
	tttgctccaggaatacatgtgagcttac	PogpdAp
	ttttgcgattgtttgaagtgttctgtatgc	
	acaagtgcatacagaacacttcaaacaatcgcaaaaatgactccagcgaaattcttccga	ApnT
	GTAGGAGTGATGAGACCCAACAACCATGATACCAGGGGATTTacgggacagtgacgctgg	
pAUB70008 - pYTR - glAp.ApnD – AmyBp.ApnC – gpdAp.ApnV – PogpdAp.ApnS	aaccattaccccgccacatagacacatctaaacaTTAATatgtcctcggtcgcacagc	ApnV
	CATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGagtccgactttcatgtggagtcg	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	GCTTCATCCCCAGCATCATTACACCTCAGCAatgagttccatcaattttcaatgggagga	ApnD
	ATATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCctgtggggactccctgtcg	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCCCTTCTCTGAACAATAAACCCCACAGAAGGCATTTatgatcgaccccaagacaagcc	ApnC
	gtagaatcagtaagctcacatgtattcctggagcaaaggagaatgccttatccagtggag	
	tttgctccaggaatacatgtgagcttac	PogpdAp
	ttttgcgattgtttgaagtgttctgtatgc	
	tacaagtgcatacagaacacttcaaacaatcgcaaaaatgtcgggctccgttcagaaaaa	ApnS
	AAGGGTATCATCGAAAGGGAGTCATCCAATTTaccacacatacgatagtgtagaggaatg	
pAUB70008-2 - pYTR - glAp.ApnD – AmyBp.ApnC – gpdAp.ApnV – PogpdAp.ApnS$_2$	aaccattaccccgccacatagacacatctaaacaTTAATatgtcctcggtcgcacagc	ApnV
	CATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGagtccgactttcatgtggagtcg	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	GCTTCATCCCCAGCATCATTACACCTCAGCAatgagttccatcaattttcaatgggagga	ApnD
	ATATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCctgtggggactccctgtcg	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCCCTTCTCTGAACAATAAACCCCACAGAAGGCATTTatgatcgaccccaagacaagcc	ApnC
	gtagaatcagtaagctcacatgtattcctggagcaaaggagaatgccttatccagtggag	
	tttgctccaggaatacatgtgagcttac	PogpdAp
	ttttgcgattgtttgaagtgttctgtatgc	
	ttacaagtgcatacagaacacttcaaacaatcgcaaaaatgtcacacgatcggacattgc	ApnS$_2$
	AAGGGTATCATCGAAAGGGAGTCATCCAATTTaccacacatacgatagtgtagaggaatg	
pAUB70009 - pYTR - glAp.ApnD – AmyBp.ApnC – gpdAp.ApnF – PogpdAp.ApnS	ccattaccccgccacatagacacatctaaacaTTAATatgtacactgagaacctcgagcc	ApnF
	TCATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGagtaagagaggtctcgcctcca	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	GCTTCATCCCCAGCATCATTACACCTCAGCAatgagttccatcaattttcaatgggagga	ApnD
	ATATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCctgtggggactccctgtcg	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCCCTTCTCTGAACAATAAACCCCACAGAAGGCATTTatgatcgaccccaagacaagcc	ApnC
	gtagaatcagtaagctcacatgtattcctggagcaaaggagaatgccttatccagtggag	
	tttgctccaggaatacatgtgagcttac	PogpdAp

	ttttgcgattgtttgaagtgttctgtatgc	
	tacaagtgcatacagaacacttcaaacaatcgcaaaaatgtcgggctccgttcagaaaaa	ApnS
	AAGGGTATCATCGAAAGGGAGTCATCCAATTTaccacacatacgatagtgtagaggaatg	
pAUB70009-2 - pYTR - glAp.ApnD – AmyBp.ApnC – gpdAp.ApnF – PogpdAp.ApnS$_2$	ccattaccccgccacatagacacatctaaacaTTAATatgtacactgagaacctcgagcc	ApnF
	TCATAGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGagtaagagaggtctcgcctcca	
	CCTGATCTTCCGAACTGGTCGTAC	glaAp
	TGCTGAGGTGTAATGATGCTGGG	
	GCTTCATCCCCAGCATCATTACACCTCAGCAatgagttccatcaattttcaatgggagga	ApnD
	ATATTGTTATATCATTTATAGCTCGTTCGGCACCTTTAATCctgtggggactccctgtcg	
	GATTAAAGGTGCCGAACGAGCTAT	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCCCTTCTCTGAACAATAAACCCCACAGAAGGCATTTatgatcgaccccaagacaagcc	ApnC
	gtagaatcagtaagctcacatgtattcctggagcaaaggagaatgccttatccagtggag	
	tttgctccaggaatacatgtgagcttac	PogpdAp
	ttttgcgattgtttgaagtgttctgtatgc	
	ttacaagtgcatacagaacacttcaaacaatcgcaaaaatgtcacacgatcggacattgc	ApnS$_2$
	AAGGGTATCATCGAAAGGGAGTCATCCAATTTaccacacatacgatagtgtagaggaatg	
pAUB80001 – pYTU – gpdAp.AscA	ACTAACCATTACCCCGCCACATAGACACATCTAAACATTAATATGGGTGAACCTATTGCC	AscA
	CGCTCGAGAGCTTCAAGCAC	
	GCCTTCATAACCCATTCATCCTCGG	
	GCCTGCGTTTCCAATGACCG	
	AACACTGACATACCTGCTCGTGG	
	ACACAGTGGAGGACATACCCGTAATTTTCTGGCTAGCATTTCGCCCCCATGGGGTATACA	
pAUB80002 – pYTR – gpdAp.AscE	accattaccccgccacatagacacatctaaacaTTAATATGTCAATTCCCAACACCCAGA	AscE
	TCTGCTAAAGGGTATCATCGAAAGGGAGTCATCCAATTTCTAATAGACAGCCAATCCAGC	
pAUB80003 – pYTP – AmyBp.AscG	TCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGACTCTTGTTTGGTATTTGCC	AscG
	AGTAGGAGTGATGAGACCCAACAACCATGATACCAGGGGATTTTGCATCATTGCGGCGAG	
pAUB80004 – pYTP – gpdAp.AscD - AmyBp.AscG	accattaccccgccacatagacacatctaaacaATGACTGTCTTAAACGATGATCGACTC	AscD
	ATCATTTATAGCTCGTTCGGCACCTTTAATCGCGGCCGCCTTCTCTACCGGGAGTCCGGG	
	GCGGCCGCGATTAAAGGTG	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	TCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGACTCTTGTTTGGTATTTGCC	AscG
	AGTAGGAGTGATGAGACCCAACAACCATGATACCAGGGGATTTTGCATCATTGCGGCGAG	
pAUB80005 – pYTR – gpdAp.AscE – AmyBp.AscB	accattaccccgccacatagacacatctaaacaTTAATATGTCAATTCCCAACACCCAGA	AscE
	CATTTATAGCTCGTTCGGCACCTTTAATCGCGGCCGCCTAATAGACAGCCAATCCAGCGG	
	GCGGCCGCGATTAAAGGTG	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCTGAACAATAAACCCCACAGAAGGCATTTATGGTTTCTATCAATTTCACGGAGCTAGC	AscB
	CTTCTGCTAAAGGGTATCATCGAAAGGGAGTCATCCAAGCTGATCGCAGAGCTGGAAGAG	
pAUB80006 – pYTP – glaAp.AscC - gpdAp.AscD - AmyBp.AscG	CCTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGCCCCATTCCATCCCCCAAGAA	AscC
	cttgggtctctcccgtcacccaaatcaattcaccggagtCTGGGGAGTGTTGTTACGGCG	
	actccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	accattaccccgccacatagacacatctaaacaATGACTGTCTTAAACGATGATCGACTC	AscD

	ATCATTTATAGCTCGTTCGGCACCTTTAATCGCGGCCGCCTTCTCTACCGGGAGTCCGGG	
	GCGGCCGCGATTAAAGGTG	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	TCTCTGAACAATAAACCCCACAGAAGGCATTTTTAATATGACTCTTGTTTGGTATTTGCC	AscG
	AGTAGGAGTGATGAGACCCAACAACCATGATACCAGGGGATTTTGCATCATTGCGGCGAG	
pAUB80007 – pYTR – glaAp.AscU - gpdAp.AscE – AmyBp.AscB	GAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGAGTCGCTACCTTCGAAATATCAC	AscU
	gggtctctcccgtcacccaaatcaattcaccggagtCTCGATCCTTTTGTTCGAACCCAG	
	actccggtgaattgatttgggtg	gpdAp
	tgtttagatgtgtctatgtggcggg	
	accattaccccgccacatagacacatctaaacaTTAATATGTCAATTCCCAACACCCAGA	AscE
	CATTTATAGCTCGTTCGGCACCTTTAATCGCGGCCGCCTAATAGACAGCCAATCCAGCGG	
	GCGGCCGCGATTAAAGGTG	AmyBp
	AAATGCCTTCTGTGGGGTTTATTGTTC	
	CTCTGAACAATAAACCCCACAGAAGGCATTTATGGTTTCTATCAATTTCACGGAGCTAGC	AscB
	CTTCTGCTAAAGGGTATCATCGAAAGGGAGTCATCCAAGCTGATCGCAGAGCTGGAAGAG	
P. oxalicum constructs		
pPO70001 – pHyg - ApnUko	gggaacaaaagctggagctcggatccatttagcggcccgcctcaggtaatacgagagagc	ApnU upstream region
	aaaagtgctccttcaatatcatcttctgtcgagttcatggttttgattaaacagtgggga	
	tcgacagaagatgatattgaaggagcact	TrpCp – hpn - TrpCt
	aagaaggattacctctaaacaagtgtacctgt	
	gaatgcacaggtacacttgtttagaggtaatccttcttgatgcagtctcgtgggtctcaa	ApnU downstream region
	aaacgacggccagtgaattcgagctcggtaccctcgagtggaaatcactcgaaggctgca	
pPO70002 – pHyg - ApnSko	ggaacaaaagctggagctcggatccatttagcggccatgtcggcgatttgcttagaccac	ApnS upstream region
	ccaagcccaaaaagtgctccttcaatatcatcttctgtcgagatggatggggtgcgggtg	
	tcgacagaagatgatattgaaggagcact	TrpCp – hpn - TrpCt
	aagaaggattacctctaaacaagtgtacctgt	
	gcacaggtacacttgtttagaggtaatccttcttcatggcagtgataacaagcaatcacc	ApnS downstream region
	gacggccagtgaattcgagctcggtaccctcgagccacattagtccaccgtacttgatac	
pPO70003 – pHyg - ApnTko	acaaaagctggagctcggatccatttagcggcccgaaagacttccgagaaggagcaaagt	ApnT upstream region
	gctccttcaatatcatcttctgtcgacgaatctggaagatctcaaatatcaaagttgtc	
	tcgacagaagatgatattgaaggagcact	TrpCp – hpn - TrpCt
	aagaaggattacctctaaacaagtgtacctgt	
	acaggtacacttgtttagaggtaatccttcttgtgaaacaggtctttttctcgtgcatc	ApnT downstream region
	aaaacgacggccagtgaattcgagctcggtaccctcgagctctgtggatcagtctcgggg	
pPO70004 – pHyg - ApnVko	gaacaaaagctggagctcggatccatttagcggcctgagttcgaaatcgcctggtttgcc	ApnV upstream region
	gcccaaaaagtgctccttcaatatcatcttctgtcgaggtttggccaaaacgtgaactgg	
	tcgacagaagatgatattgaaggagcact	TrpCp – hpn - TrpCt
	aagaaggattacctctaaacaagtgtacctgt	
	agaatgcacaggtacacttgtttagaggtaatccttcttaggccgatacgacgtatgacc	ApnV downstream region
	ttgtaaaacgacggccagtgaattcgagctcggtaccctcgagacagtcgggtccaccgc	
pPO70005 – pHyg - ApnGko	aaagggaacaaaagctggagctcggatccatttagcggcctcgtccaccaaggaccactc	ApnG upstream region
	ccaaaaagtgctccttcaatatcatcttctgtcgagacccactattctccacggtaagac	
	tcgacagaagatgatattgaaggagcact	TrpCp – hpn - TrpCt
	aagaaggattacctctaaacaagtgtacctgt	

	cccagaatgcacaggtacacttgtttagaggtaatccttcttatgcggaggtctgcgcaa	ApnG downstream region
	aaacgacggccagtgaattcgagctcggtaccctcgagaggtccttgttcgctcaacatg	
pPO70006 – pBAR - ApnU	catgcggagagacggacg	AngpdAp
	gggaaaagaaagagaaaagaaaagagcagc	
	ctctccccaccagctgctcttttctttttctctttctttttcccatggcgcccactctggag	ApnU
	taaaggttcttggatgggaagatgaatatactgaagatggattccgattgccgcttcaga	
pPO70007 – pBAR - ApnT	catgcggagagacggacg	AngpdAp
	gggaaaagaaagagaaaagaaaagagcagc	
	ccaccagctgctcttttctttttctctttctttttcccatgactccagcgaaattcttccga	ApnT
	tgattaaaggttcttggatgggaagatgaatatactgaagatacgggacagtgacgctgg	
S. cerevisiae expression constructs		
pYUB10001 – pXP – TEF1p.pDHAD.CYC1t	aataACTAGTatggctagcgcccaatccg	pDHAD
	aataCTCGAGttactcgtcagtcacacatccatctgaag	
pYUB10002 – pXP – TEF1p.fDHAD.CYC1t	GCATAGCAATCTAATCTAAGTTTTAATTACAAAACTAGTatgcttctctctcagacccgg	fDHAD
	GAATGTAAGCGTGACATAACTAATTACATGACTCGAGttagtcaagagcatcggtgatgc	
pYUB10003 – pXP – TEF1p.AstD.CYC1t	TAGCAATCTAATCTAAGTTTTAATTACAAAACTAGTATGGACTACAAAGACGATGACGAC	AstD
	GCGTGAATGTAAGCGTGACATAACTAATTACATGACTCGAGCTAGATCGGTCCGTCCGTG	
pYUB60001 – pJB – ADH2p.HarD.PRM9t	TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCGCAAAACGTAGGGGCAAACAAACG	ADH2p
	GGTATTACGATATAGTTAATAGTTGATAGTTGATTG	
	tacaatcaactatcaactattaactatatcgtaataccatatggtaaacgccacttgggc	HarD
	TAAAGTTGTGTGCTAGTGTCTCCCGTCTTCTGTCtcagttcggactgaaagtaactgcag	
	GACAGAAGACGGGAGACACTAGCAC	PRM9t
	GTTCTACAAAATGAAGCACAGATGCTTCGTTGGCATTTTCAACATCGTATTTTCCGAAGC	
pYUB60002 – pJB – MLS1p.HarC.SPG5t	aTGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCCATTGGGCCGATGAAGTTAGTCG	MLS1p
	TTTCTTAATTCTTTTATGTGCTTTTACTACTTTGTTTAGTTC	
	TAAACAAAGTAGTAAAAGCACATAAAAGAATTAAGAAAatggctgttccatcgggagatg	HarC
	CCTTCTTGGTAATAGCGCGATGAAACAACGTCTTTGCttaatccctgacaactcgctttgca a	
	GCAAAGACGTTGTTTCATCGC	SPG5t
	TGTTCTACAAAATGAAGCACAGATGCTTCGTTGCTTATTTTCTGCCGAATTTTCATGAAG	
pYUB60003 – pJB – ADH2p.HarB.ADH2t	TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCGCAAAACGTAGGGGCAAACAAACG	ADH2p
	GGTATTACGATATAGTTAATAGTTGATAGTTGATTG	
	acaatcaactatcaactattaactatatcgtaataccataatggcgaacgttaatgtgaa	HarB
	ttgataatgaaaactataaatcgTGAAGGCATGTTTtcagtggtggtggtggtggtgctc	
	cgatttatagttttcatt	ADH2t
	taatattctacttttgctcccctcgagggtaccgagctcgaattcactggccgtcgttt	
pYUB60004 – pJB – ADH2p.HarD.PRM9t – MLS1p.HarC.SPG5t	TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCGCAAAACGTAGGGGCAAACAAACG	ADH2p
	GGTATTACGATATAGTTAATAGTTGATAGTTGATTG	
	tacaatcaactatcaactattaactatatcgtaataccatatggtaaacgccacttgggc	HarD
	TAAAGTTGTGTGCTAGTGTCTCCCGTCTTCTGTCtcagttcggactgaaagtaactgcag	
	GACAGAAGACGGGAGACACTAGCAC	PRM9t
	GGCATTTTCAACATCGTATTTTCCGAAGC	
	TGCAACGCTTCGGAAAATACGATGTTGAAAATGCCCACATTGGGCCGATGAAGTTAGTCG	MLS1p
	TTTCTTAATTCTTTTATGTGCTTTTACTACTTTGTTTAGTTC	
	TAAACAAAGTAGTAAAAGCACATAAAAGAATTAAGAAAatggctgttccatcgggagatg	HarC

	TCTTGGTAATAGCGCGATGAAACAACGTCTTTGCttaatccctgacaactcgctttgcaa	
	GCAAAGACGTTGTTTCATCGC	SPG5t
	TGTTCTACAAAATGAAGCACAGATGCTTCGTTGCTTATTTTCTGCCGAATTTTCATGAAG	
pYUB60005 – pJB – ADH2p.HarD.PRM9t – PCK1p.HarB.CPS1t	TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCGCAAAACGTAGGGGCAAACAAACG	ADH2p
	GGTATTACGATATAGTTAATAGTTGATAGTTGATTG	
	tacaatcaactatcaactattaactatatcgtaataccatatggtaaacgccacttgggc	HarD
	TAAAGTTGTGTGCTAGTGTCTCCCGTCTTCTGTCtcagttcggactgaaagtaactgcag	
	GACAGAAGACGGGAGACACTAGCAC	PRM9t
	GGATGAAAGGAAGCTCGGTTTTTTCCTATTGGGCATTTTCAACATCGTATTTTCCGAAGC	
	CAATAGGAAAAAACCGAGCTTCCTTTC	PCK1p
	GTTGTTATTTTATTATGGAATAATTAGTTGCGTG	
	AACTCACGCAACTAATTATTCCATAATAAAATAACAACatggcgaacgttaatgtgaagg	HarB
	AAAAAAAAAATCTTTGACTATTCAATCATTGCGCttatcttcgaagcctaacctccagca	
	GCGCAATGATTGAATAGTCAAAGATT	CPS1t
	TTGTTCTACAAAATGAAGCACAGATGCTTCGTTATTTGACACTTGATTTGACACTTCTTT	
pYUB60006 – pJB – PCK1p.HarB.CPS1t – MLS1p.HarC.SPG5t	TCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCCAATAGGAAAAAACCGAGCTTCCTTTC	PCK1p
	GTTGTTATTTTATTATGGAATAATTAGTTGCGTG	
	AACTCACGCAACTAATTATTCCATAATAAAATAACAACatggcgaacgttaatgtgaagg	HarB
	AAAAAAAAAATCTTTGACTATTCAATCATTGCGCttatcttcgaagcctaacctccagca	
	GCGCAATGATTGAATAGTCAAAGATT	CPS1t
	TTCTATCCGTCGACTAACTTCATCGGCCCAATGATTTGACACTTGATTTGACACTTCTTT	
	CATTGGGCCGATGAAGTTAGTCG	MLS1p
	TTTCTTAATTCTTTTATGTGCTTTTACTACTTTGTTTAGTTC	
	TAAACAAAGTAGTAAAAGCACATAAAAGAATTAAGAAAatggctgttccatcgggagatg	HarC
	TCTTGGTAATAGCGCGATGAAACAACGTCTTTGCttaatccctgacaactcgctttgcaa	
	GCAAAGACGTTGTTTCATCGC	SPG5t
	TGTTCTACAAAATGAAGCACAGATGCTTCGTTGCTTATTTTCTGCCGAATTTTCATGAAG	
pYUB60007 – pJB – ADH2p.HarD.PRM9t - PCK1p.HarB.CPS1t – MLS1p.HarC.SPG5t	TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCGCAAAACGTAGGGGCAAACAAACG	ADH2p
	GGTATTACGATATAGTTAATAGTTGATAGTTGATTG	
	tacaatcaactatcaactattaactatatcgtaataccatatggtaaacgccacttgggc	HarD
	TAAAGTTGTGTGCTAGTGTCTCCCGTCTTCTGTCtcagttcggactgaaagtaactgcag	
	GACAGAAGACGGGAGACACTAGCAC	PRM9t
	GGATGAAAGGAAGCTCGGTTTTTTCCTATTGGGCATTTTCAACATCGTATTTTCCGAAGC	
	CAATAGGAAAAAACCGAGCTTCCTTTC	PCK1p
	GTTGTTATTTTATTATGGAATAATTAGTTGCGTG	
	AACTCACGCAACTAATTATTCCATAATAAAATAACAACatggcgaacgttaatgtgaagg	HarB
	AAAAAAAAAATCTTTGACTATTCAATCATTGCGCttatcttcgaagcctaacctccagca	
	GCGCAATGATTGAATAGTCAAAGATT	CPS1t
	TTCTATCCGTCGACTAACTTCATCGGCCCAATGATTTGACACTTGATTTGACACTTCTTT	
	CATTGGGCCGATGAAGTTAGTCG	MLS1p
	TTTCTTAATTCTTTTATGTGCTTTTACTACTTTGTTTAGTTC	
	TAAACAAAGTAGTAAAAGCACATAAAAGAATTAAGAAAatggctgttccatcgggagatg	HarC
	TCTTGGTAATAGCGCGATGAAACAACGTCTTTGCttaatccctgacaactcgctttgcaa	
	GCAAAGACGTTGTTTCATCGC	SPG5t

	TGTTCTACAAAATGAAGCACAGATGCTTCGTTGCTTATTTTCTGCCGAATTTTCATGAAG	
S. cerevisiae knock-out constructs		
pYUB20001 – pCRblunt – ILV3p.URA3.ILV3t	AGATATCCATCACACTGGCccttcgaggccctggct	ILV3p
	ctctacttttttcttggctgtttacaaaaaaatacaag	
	gtatttttttgtaaacagccaagaaaaaagtagagATGTCGAAAGCTACATATAAGGAAC	URA3
	atctctatatatatattcatcgattggggcctataatgcaTTAGTTTTGCTGGCCGCATC	
	tgcattataggccccaatcgatgaatatatatatagaga	ILV3t
	GGGCGAATTGGGCCCTCTAGATGCATGCacgttaggaccccagtatctcgg	
piCas9-ILV3	ATGATCGAACATGCATGCCCAGGTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	ILV3 targeting sgRNA
	TTTCTAGCTCTAAAACGGACCTGGGCATGCATGTTCGATCATTTATCTTTCACTGCGGAG	
piCas9-URA3	ATGATCTGCAAGGGCTCCCTAGCTACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	URA3 targeting sgRNA
	TTTCTAGCTCTAAAACGTAGCTAGGGAGCCCTTGCAGATCATTTATCTTTCACTGCGGAG	
E. coli expression constructs		
pEUB10001 – pET28a - pDHAD	atggctagcgcccaatccgtaaccgctg	pDHAD
	atagcggccgcttactcgtcagtcacacatccatctg	
pEUB10002 - pET28a – I177L	gacatatgaccttgtatctgctttccagag	I177L mutation
	cagatacaaggtcatatgtcttgtcttgaaaatg	
pEUB10003 - pET28a – I177F	caagacatatgactttgtatctgctttccagagttatggag	I177F mutation
	ggaaagcagatacccagtcatatgtcttgtcttgaaaatgg	
pEUB10004 - pET28a – I177W	gacaagacatatgactgggtatctgctttccagagttatggag	I177W mutation
	ggaaagcagatacccagtcatatgtcttgtcttgaaaatgg	
pEUB10005 - pET28a – V178L	gacatatgacattctgtctgctttccagagttatggagaatttg	V178L mutation
	ctggaaagcagacagaatgtcatatgtcttgtcttgaaaatg	
pEUB10006 - pET28a – V178I	gacatatgacattatttctgctttccagagttatggagaatttg	V178I mutation
	ctggaaagcagaaataatgtcatatgtcttgtcttgaaaatg	
pEUB10007 - pET28a – V178F	gacatatgacatttttttctgctttccagagttatggagaatttg	V178F mutation
	ctggaaagcagaaaaaatgtcatatgtcttgtcttgaaaatg	
pEUB10008 - pET28a – V178W	gacatatgacatttggtctgctttccagagttatggagaatttg	V178W mutation
	ctggaaagcagaccaaatgtcatatgtcttgtcttgaaaatg	
pEUB10009 - pET28a – V496L	cacggttttcttgttggccacatttgc	V496L mutation
	gccaacaagaaaaccgtgtgacccaccag	
pEUB10010 - pET28a – V496I	cacggttttattgttggccacatttgc	V496I mutation
	gccaacaataaaaccgtgtgacccaccag	
pEUB10011 - pET28a – V496F	cacggtttttttgttggccacatttgccc	V496F mutation
	gccaacaaaaaaaccgtgtgacccacc	
pEUB10012 - pET28a – V496W	cacacggtttttgggttggccacatttgccctgaag	V496W mutation
	gtggccaacccaaaaaccgtgtgacccaccag	
pEUB10013 - pET28a – V497L	cacggttttgttcttggccacatttgccc	V497L mutation
	gtggccaagaacaaaaccgtgtgacc	
pEUB10014 - pET28a – V497I	cacggttttgttattggccacatttgccc	V497I mutation
	gtggccaataacaaaaccgtgtgaccc	
pEUB10015 - pET28a – V497F	cacggttttgtttttggccacatttgccc	V497F mutation
	gtggccaaaaacaaaaccgtgtgaccc	
	cacacggttttgtttggggccacatttgccctgaag	

pEUB10016 - pET28a – V497W	caaatgtggccccaaacaaaaccgtgtgacccaccag	V497W mutation
pEUB10017 - pET28a – I177L – V496L	gacatatgaccttgtatctgctttccagag	I177L mutation
	cagatacaaggtcatatgtcttgtcttgaaaatg	
	cacggttttcttgttggccacatttgc	V496L mutation
	gccaacaagaaaaccgtgtgacccaccag	
pEUB10018 - pET28a – I177L – V4967L	gacatatgaccttgtatctgctttccagag	I177L mutation
	cagatacaaggtcatatgtcttgtcttgaaaatg	
	cacggttttgttcttggccacatttgccc	V497L mutation
	gtggccaagaacaaaaccgtgtgacc	
pEUB10018 - pET28a – V178L – V4967L	gacatatgacattctgtctgctttccagagttatggagaatttg	V178L mutation
	ctggaaagcagacagaatgtcatatgtcttgtcttgaaaatg	
	cacggttttgttcttggccacatttgccc	V497L mutation
	gtggccaagaacaaaaccgtgtgacc	
pEUB60001 – pET28a - HarB	ctggtgccgcgcggcagcatggcgaacgttaatgtgaaggagc	HarB
	tcagtggtggtggtggtggtgCTCGAGtcttcgaagcctaacctccagcaag	
pEUB60002 – pET28a - HarC	ttatCCatggctgttccatcgggagatg	HarC
	ttatCTCGAGatccctgacaactcgctttgcaatc	
pEUB20001 – pET28a – TfG8H	ttatCATATGGGACAGTCCCGCCGACCCCACACCGTCTACCTCGATCCGGCCAAAGGAGTCG ACA TCCCCGCGCAGCGACGCGAACTCCTCGACAAAGGCCCGGTGGTACGCGTGGCTTTCCCCGGC	TfG8H
	AtAActcgagTTATGGTTTCGGACGAAGTAAAACGGG	
pEUB20002 – pET28a – FpR	aataGCTAGCatggctgattgggtaacaggcaaag	FpR
	aataCTCGAGttaccagtaatgctccgctgtcatatg	
pEUB20003 – pET28a – YkuN	aataCATATGGCGAAGGCATTGATCACATAC	YkuN
	aataCTCGAGTTAACTGACGTGGATCTTCTCTTTGTTCATG	
pEUB20004 – pET28a – GOR	CAGCCATATGACTAAAACTAATTCTCCAGCCCCATC	GOR
	GGTGCTCGAGTGAACTTAATAACAACTTTGACACAGTCTGGG	
pEUB20005 – pET28a – ISY	CAGCCATATGTCCTGGTGGTGGAAAAGGTC	ISY
	AAGCTTGTCGTTATGGAATGAATCTGTAGTCTCTCATCTTGTCG	
pEUB20006 – pET28a – NmMLPL	ttatCATATGGCTAGCATGGCGTCGAAACTGG	NmMLPL
	ttatCTCGAGCTAATTTTGACAAGTGTGATTCATACCCTT	
pEUB20007 – pET28a – NEPS1	TAATGCTAGCATGGCATCCACAGCGAACC	NEPS1
	TAATGAGCTCCTATGACGGTGCGAAGAAAGGTAAAC	
pEUB20008 – pET22a – NoxE	CATATGAAAATCGTAGTTATCGG	NoxE
	CTCGAGTTATTTGGCATTCAAAGCT	
pEUB20009 – pET28a – FumC	tggacccatatgaatacagtacgcagcgaaaaagattcg	FumC
	TGGACCAAGCTTTTAACGCCCGGCTTTCATACTGC	
pEUB20010 – pET28a – MaeB	TGGCAAGCTAGCatggatgaccagttaaaacaaagtgcacttg	MaeB
	TGGCAACTCGAGTTACAGCGGTTGGGTTTGCG	

9 REFERENCES

1. Dias, D. A., Urban, S. & Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, 303–336 (2012).
2. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507–519 (2011).
3. Lobanovska, M. & Pilla, G. Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? *Yale J. Biol. Med.* **90**, 135–145 (2017).
4. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* (2020) doi:10.1021/acs.jnatprod.9b01285.
5. Nussbaum, F. von, Brands, M., Hinzen, B., Weigand, S. & Häbich, D. Antibacterial Natural Products in Medicinal Chemistry—Exodus or Revival? *Angew. Chem. Int. Ed.* **45**, 5072–5129 (2006).
6. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).
7. Corre, C. & Challis, G. L. New natural product biosynthetic chemistry discovered by genome mining. *Nat. Prod. Rep.* **26**, 977–986 (2009).
8. Bachmann, B. O., Van Lanen, S. G. & Baltz, R. H. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **41**, 175–184 (2014).
9. Challis, G. L. Genome Mining for Novel Natural Product Discovery. *J. Med. Chem.* **51**, 2618–2628 (2008).
10. Zerikly, M. & Challis, G. L. Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem* **10**, 625–633 (2009).
11. Amos, G. C. A. *et al*. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci.* **114**, E11121–E11130 (2017).
12. Han, R. *et al*. De Novo Deep Transcriptome Analysis of Medicinal Plants for Gene Discovery in Biosynthesis of Plant Natural Products. *Methods Enzymol.* **576**, 19–45 (2016).
13. Fisher, M. C., Hawkins, N. J., Sanglard, D. & Gurr, S. J. Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science* **360**, 739–742 (2018).
14. Ventola, C. L. The Antibiotic Resistance Crisis. *Pharm. Ther.* **40**, 277–283 (2015).
15. Swanton, C. J., Harker, K. N. & Anderson, R. L. Crop Losses Due to Weeds in Canada. *Weed Technol.* **7**, 537–542 (1993).
16. Gianessi, L. P. The increasing importance of herbicides in worldwide crop production. *Pest Manag. Sci.* **69**, 1099–1105 (2013).
17. Fisher, M. C., Hawkins, N. J., Sanglard, D. & Gurr, S. J. Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science* **360**, 739–742 (2018).
18. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* **5**, 175–186 (2007).
19. Handsfield, H. H., Clark, H., Wallace, J. F., Holmes, K. K. & Turck, M. Amoxicillin, a New Penicillin Antibiotic. *Antimicrob. Agents Chemother.* **3**, 262–265 (1973).
20. Farmer, J. A. *et al*. Comparative effects of simvastatin and lovastatin in patients with hypercholesterolemia. The Simvastatin and Lovastatin Multicenter Study Participants. *Clin. Ther.* **14**, 708–717 (1992).

21. Xie, X., Watanabe, K., Wojcicki, W. A., Wang, C. C. C. & Tang, Y. Biosynthesis of lovastatin analogs with a broadly specific acyltransferase. *Chem. Biol.* **13**, 1161–1169 (2006).
22. Xie, X. & Tang, Y. Efficient synthesis of simvastatin by use of whole-cell biocatalysis. *Appl. Environ. Microbiol.* **73**, 2054–2060 (2007).
23. Gao, X. *et al.* Directed Evolution and Structural Characterization of a Simvastatin Synthase. *Chem. Biol.* **16**, 1064–1074 (2009).
24. Neu, H. C. The Crisis in Antibiotic Resistance. *Science* **257**, 1064–1073 (1992).
25. Laxminarayan, R. *et al.* Antibiotic resistance—the need for global solutions. *Lancet Infect. Dis.* **13**, 1057–1098 (2013).
26. Cundliffe, E. SELF DEFENCE IN ANTIBIOTIC-PRODUCING ORGANISMS. *Br. Med. Bull.* **40**, 61–67 (1984).
27. Hopwood, D. A. How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol. Microbiol.* **63**, 937–940 (2007).
28. Almabruk, K. H., Dinh, L. K. & Philmus, B. Self-Resistance of Natural Product Producers: Past, Present, and Future Focusing on Self-Resistant Protein Variants. *ACS Chem. Biol.* **13**, 1426–1437 (2018).
29. Mak, S., Xu, Y. & Nodwell, J. R. The expression of antibiotic resistance genes in antibiotic-producing bacteria. *Mol. Microbiol.* **93**, 391–402 (2014).
30. Biggins, J. B., Onwueme, K. C. & Thorson, J. S. Resistance to Enediyne Antitumor Antibiotics by CalC Self-Sacrifice. *Science* **301**, 1537–1541 (2003).
31. Noda, M. *et al.* Self-protection Mechanism in d-Cycloserine-producing Streptomyces lavendulae GENE CLONING, CHARACTERIZATION, AND KINETICS OF ITS ALANINE RACEMASE AND d-ALANYL-d-ALANINE LIGASE, WHICH ARE TARGET ENZYMES OF d-CYCLOSERINE. *J. Biol. Chem.* **279**, 46143–46152 (2004).
32. Gandlur, S. M., Wei, L., Levine, J., Russell, J. & Kaur, P. Membrane Topology of the DrrB Protein of the Doxorubicin Transporter of Streptomyces peucetius. *J. Biol. Chem.* **279**, 27799–27806 (2004).
33. Danshiitsoodol, N., de Pinho, C. A., Matoba, Y., Kumagai, T. & Sugiyama, M. The Mitomycin C (MMC)-binding Protein from MMC-producing Microorganisms Protects from the Lethal Effect of Bleomycin: Crystallographic Analysis to Elucidate the Binding Mode of the Antibiotic to the Protein. *J. Mol. Biol.* **360**, 398–408 (2006).
34. Yan, Y., Liu, N. & Tang, Y. Recent developments in self-resistance gene directed natural product discovery. *Nat. Prod. Rep.* (2020) doi:10.1039/C9NP00050J.
35. Thiara, A. S. & Cundliffe, E. Cloning and characterization of a DNA gyrase B gene from Streptomyces sphaeroides that confers resistance to novobiocin. *EMBO J.* **7**, 2255–2259 (1988).
36. Steffensky, M., Mühlenweg, A., Wang, Z.-X., Li, S.-M. & Heide, L. Identification of the Novobiocin Biosynthetic Gene Cluster of Streptomyces spheroides NCIB 11891. *Antimicrob. Agents Chemother.* **44**, 1214–1222 (2000).
37. Olano, C. *et al.* Biosynthesis of the Angiogenesis Inhibitor Borrelidin by Streptomyces parvulus Tü4055: Cluster Analysis and Assignment of Functions. *Chem. Biol.* **11**, 87–97 (2004).
38. Kale, A. J., McGlinchey, R. P., Lechner, A. & Moore, B. S. Bacterial Self-Resistance to the Natural Proteasome Inhibitor Salinosporamide A. *ACS Chem. Biol.* **6**, 1257–1264 (2011).
39. Tang, X. *et al.* Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).

40. Peterson, R. M., Huang, T., Rudolf, J. D., Smanski, M. J. & Shen, B. Mechanisms of Self-Resistance in the Platensimycin- and Platencin-Producing Streptomyces platensis MA7327 and MA7339 Strains. *Chem. Biol.* **21**, 389–397 (2014).
41. Kling, A. *et al.* Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **348**, 1106–1112 (2015).
42. Kennedy, J. *et al.* Modulation of Polyketide Synthase Activity by Accessory Proteins During Lovastatin Biosynthesis. *Science* **284**, 1368–1372 (1999).
43. Regueira, T. B. *et al.* Molecular Basis for Mycophenolic Acid Biosynthesis in Penicillium brevicompactum. *Appl. Environ. Microbiol.* **77**, 3035–3043 (2011).
44. Yeh, H.-H. *et al.* Resistance Gene-Guided Genome Mining: Serial Promoter Exchanges in Aspergillus nidulans Reveal the Biosynthetic Pathway for Fellutamide B, a Proteasome Inhibitor. *ACS Chem. Biol.* **11**, 2275–2284 (2016).
45. Lin, H.-C. *et al.* The Fumagillin Biosynthetic Gene Cluster in Aspergillus fumigatus Encodes a Cryptic Terpene Cyclase Involved in the Formation of β-trans-Bergamotene. *J. Am. Chem. Soc.* **135**, 4616–4619 (2013).
46. Yue, Q. *et al.* Genomics-driven discovery of a novel self-resistance mechanism in the echinocandin-producing fungus Pezicula radicicola. *Environ. Microbiol.* **0**, (2018).
47. Yan, Y. *et al.* Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action. *Nature* **559**, 415–418 (2018).
48. Amorim Franco, T. M. & Blanchard, J. S. Bacterial Branched-Chain Amino Acid Biosynthesis: Structures, Mechanisms, and Drugability. *Biochemistry* **56**, 5849–5865 (2017).
49. Binder, S. Branched-Chain Amino Acid Metabolism in Arabidopsis thaliana. *Arab. Book Am. Soc. Plant Biol.* **8**, (2010).
50. Brown, H. M. Mode of action, crop selectivity, and soil relations of the sulfonylurea herbicides. *Pestic. Sci.* **29**, 263–281 (1990).
51. Singh, B. & Shaner, D. Biosynthesis of Branched Chain Amino Acids: From Test Tube to Field. *Plant Cell* **7**, 935–944 (1995).
52. Grandoni, J. A., Marta, P. T. & Schloss, J. V. Inhibitors of branched-chain amino acid biosynthesis as potential antituberculosis agents. *J. Antimicrob. Chemother.* **42**, 475–482 (1998).
53. Zhang, Y. *et al.* Synthesis and biological activities of novel 5-substituted-1,3,4-oxadiazole Mannich bases and bis-Mannich bases as ketol-acid reductoisomerase inhibitors. *Bioorg. Med. Chem. Lett.* **26**, 4661–4665 (2016).
54. Tsuda, Y. *et al.* Aspterric acid, a new sesquiterpenoid of the carotane group, a metabolite from Aspergillus terreus IFO-6123. X-Ray crystal and molecular structure of its p-bromobenzoate. *J. Chem. Soc. Chem. Commun.* **0**, 160–161 (1978).
55. Shimada, A. *et al.* Aspterric acid and 6-hydroxymellein, inhibitors of pollen development in Arabidopsis thaliana, produced by Aspergillus terreus. *Z. Naturforschung C J. Biosci.* **57**, 459–464 (2002).
56. Akasaka, T., Tanaka, M., Yamaguchi, A. & Sato, K. Type II Topoisomerase Mutations in Fluoroquinolone-Resistant Clinical Strains of Pseudomonas aeruginosa Isolated in 1998 and 1999: Role of Target Enzyme in Mechanism of Fluoroquinolone Resistance. *Antimicrob. Agents Chemother.* **45**, 2263–2268 (2001).
57. Nessar, R., Reyrat, J. M., Murray, A. & Gicquel, B. Genetic analysis of new 16S rRNA mutations conferring aminoglycoside resistance in Mycobacterium abscessus. *J. Antimicrob. Chemother.* **66**, 1719–1724 (2011).

58. Rahman, M. M. *et al*. The Crystal Structure of a Bacterial l-Arabinonate Dehydratase Contains a [2Fe-2S] Cluster. *ACS Chem. Biol.* **12**, 1919–1927 (2017).
59. Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. STRUCTURE, FUNCTION, AND FORMATION OF BIOLOGICAL IRON-SULFUR CLUSTERS. *Annu. Rev. Biochem.* **74**, 247–281 (2005).
60. Burton, M., Rigby, D., Young, T. & James, S. Consumer attitudes to genetically modified organisms in food in the UK. *Eur. Rev. Agric. Econ.* **28**, 479–498 (2001).
61. Costa-Font, M., Gil, J. M. & Traill, W. B. Consumer acceptance, valuation of and attitudes towards genetically modified food: Review and implications for food policy. *Food Policy* **33**, 99–111 (2008).
62. Bawa, A. S. & Anilakumar, K. R. Genetically modified foods: safety, risks and public concerns—a review. *J. Food Sci. Technol.* **50**, 1035–1046 (2013).
63. Paoletti, C. *et al*. GMO risk assessment around the world: Some examples. *Trends Food Sci. Technol.* **19**, S70–S78 (2008).
64. (7) Risk and Regulation: U.S. Regulatory Policy on Genetically Modified Food and Agriculture. *ResearchGate* https://www.researchgate.net/publication/238603026_Risk_and_Regulation_US_Regulatory_Policy_on_Genetically_Modified_Food_and_Agriculture.
65. Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J. & Nekrasov, V. Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **32**, 76–84 (2015).
66. Eid, A., Alshareef, S. & Mahfouz, M. M. CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* **475**, 1955–1964 (2018).
67. Hua, K., Tao, X. & Zhu, J.-K. Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnol. J.* **0**,.
68. Hua, K., Tao, X., Yuan, F., Wang, D. & Zhu, J.-K. Precise A·T to G·C Base Editing in the Rice Genome. *Mol. Plant* **11**, 627–630 (2018).
69. Li, C. *et al*. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **19**, 59 (2018).
70. Sun, Y. *et al*. Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase. *Mol. Plant* **9**, 628–631 (2016).
71. Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14**, 438–443 (2003).
72. Kluckova, K., Bezawork-Geleta, A., Rohlena, J., Dong, L. & Neuzil, J. Mitochondrial complex II, a novel target for anti-cancer agents. *Biochim. Biophys. Acta BBA - Bioenerg.* **1827**, 552–564 (2013).
73. Saraste, M. Oxidative Phosphorylation at the fin de siècle. *Science* **283**, 1488–1493 (1999).
74. Mori, M., Nonaka, K., Masuma, R., Ōmura, S. & Shiomi, K. Helminth Electron Transport Inhibitors Produced by Fungi. in *Physiology and Genetics: Selected Basic and Applied Aspects* (eds. Anke, T. & Schüffler, A.) 297–329 (Springer International Publishing, 2018). doi:10.1007/978-3-319-71740-1_10.
75. Omura, S. *et al*. ATPENINS, NEW ANTIFUNGAL ANTIBIOTICS PRODUCED BY PENICILLIUM SP. *J. Antibiot. (Tokyo)* **41**, 1769–1773 (1988).
76. Miyadera, H. *et al*. Atpenins, potent and specific inhibitors of mitochondrial complex II (succinate-ubiquinone oxidoreductase). *Proc. Natl. Acad. Sci.* **100**, 473–477 (2003).
77. Horsefield, R. *et al*. Structural and Computational Analysis of the Quinone-binding Site of Complex II (Succinate-Ubiquinone Oxidoreductase) A MECHANISM OF ELECTRON

TRANSFER AND PROTON CONDUCTION DURING UBIQUINONE REDUCTION. *J. Biol. Chem.* **281**, 7309–7316 (2006).
78. Jacob Jessen, H. & Gademann, K. 4-Hydroxy-2-pyridone alkaloids: Structures and synthetic approaches. *Nat. Prod. Rep.* **27**, 1168–1185 (2010).
79. Trecourt, F., Mallet, M., Mongin, O. & Queguiner, G. Total Synthesis of (.+-.)-Atpenin B. An Original 'Clockwise' Functionalization of 2-Chloropyridine. *J. Org. Chem.* **59**, 6173–6178 (1994).
80. Tréecourt, F., Mallet, M., Mongin, O. & Quéguiner, G. First synthesis of (±)-harzianopyridone by metalation of polysubstituted O-pyridylcarbamates. *J. Heterocycl. Chem.* **32**, 1117–1124 (1995).
81. Ohtawa, M. *et al.* Enantioselective total synthesis of atpenin A5. *J. Antibiot. (Tokyo)* **62**, 289–294 (2009).
82. Krautwald, S. *et al.* Bioisosteric Exchange of C -Chloro and Methyl Substituents: Synthesis and Initial Biological Studies of Atpenin A5 Analogues. *Angew. Chem. Int. Ed.* **55**, 4049–4053 (2016).
83. Selby, T. P., Hughes, K. A., Rauh, J. J. & Hanna, W. S. Synthetic atpenin analogs: Potent mitochondrial inhibitors of mammalian and fungal succinate-ubiquinone oxidoreductase. *Bioorg. Med. Chem. Lett.* **20**, 1665–1668 (2010).
84. Wang, H. *et al.* Synthesis and Antineoplastic Evaluation of Mitochondrial Complex II (Succinate Dehydrogenase) Inhibitors Derived from Atpenin A5. *ChemMedChem* **12**, 1033–1044 (2017).
85. Otsuka, T., Takase, S., Terano, H. & Okuhara, M. New angiogenesis inhibitors, WF-16775 A1 and A2. *J. Antibiot. (Tokyo)* **45**, 1970–1973 (1992).
86. Eley, K. L. *et al.* Biosynthesis of the 2-Pyridone Tenellin in the Insect Pathogenic Fungus Beauveria bassiana. *ChemBioChem* **8**, 289–297 (2007).
87. Dickinson, J. M., Hanson, J. R., Hitchcock, P. B. & Claydon, N. Structure and biosynthesis of harzianopyridone, an antifungal metabolite of Trichoderma harzianum. *J. Chem. Soc. Perkin 1* **0**, 1885–1887 (1989).
88. Halo, L. M. *et al.* Late Stage Oxidations during the Biosynthesis of the 2-Pyridone Tenellin in the Entomopathogenic Fungus Beauveria bassiana. *J. Am. Chem. Soc.* **130**, 17988–17996 (2008).
89. Ohashi, M. *et al.* SAM-dependent enzyme-catalysed pericyclic reactions in natural product biosynthesis. *Nature* **549**, 502–506 (2017).
90. Bergmann, S. *et al.* Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans. *Nat. Chem. Biol.* **3**, 213–217 (2007).
91. Liu, N. *et al.* Identification and Heterologous Production of a Benzoyl-Primed Tricarboxylic Acid Polyketide Intermediate from the Zaragozic Acid A Biosynthetic Pathway. *Org. Lett.* **19**, 3560–3563 (2017).
92. Wasil, Z. *et al.* One pathway, many compounds: heterologous expression of a fungal biosynthetic pathway reveals its intrinsic potential for diversity. *Chem. Sci.* **4**, 3845–3856 (2013).
93. Jencks, W. P. *Catalysis in Chemistry and Enzymology*. (Dover Publications, 1987).
94. Bamberger, E. Ueber die Reduction der Nitroverbindungen. *Berichte Dtsch. Chem. Ges.* **27**, 1347–1350 (1894).
95. Latham, J., Brandenburger, E., Shepherd, S. A., Menon, B. R. K. & Micklefield, J. Development of Halogenase Enzymes for Use in Synthesis. *Chem. Rev.* **118**, 232–269 (2018).

96. Harris, C. M., Kannan, R., Kopecka, H. & Harris, T. M. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *J. Am. Chem. Soc.* **107**, 6652–6658 (1985).
97. Xu, Z. *et al.* Halogen Bond: Its Role beyond Drug–Target Binding Affinity for Drug Discovery and Development. *J. Chem. Inf. Model.* **54**, 69–78 (2014).
98. Lu, Y. *et al.* Halogen bonding for rational drug design and new drug discovery. *Expert Opin. Drug Discov.* **7**, 375–383 (2012).
99. Sirimulla, S., Bailey, J. B., Vegesna, R. & Narayan, M. Halogen interactions in protein-ligand complexes: implications of halogen bonding for rational drug design. *J. Chem. Inf. Model.* **53**, 2781–2791 (2013).
100. Cavallo, G. *et al.* The Halogen Bond. *Chem. Rev.* **116**, 2478–2601 (2016).
101. Clark, T., Hennemann, M., Murray, J. S. & Politzer, P. Halogen bonding: the σ-hole. *J. Mol. Model.* **13**, 291–296 (2007).
102. Han, F.-S. Transition-metal-catalyzed Suzuki–Miyaura cross-coupling reactions: a remarkable advance from palladium to nickel catalysts. *Chem. Soc. Rev.* **42**, 5270–5298 (2013).
103. Nicolaou, K. C., Bulger, P. G. & Sarlah, D. Palladium-catalyzed cross-coupling reactions in total synthesis. *Angew. Chem. Int. Ed Engl.* **44**, 4442–4489 (2005).
104. Hartwig, J. F. Carbon-heteroatom bond formation catalysed by organometallic complexes. *Nature* **455**, 314–322 (2008).
105. Gkotsi, D. S., Dhaliwal, J., McLachlan, M. M., Mulholand, K. R. & Goss, R. J. Halogenases: powerful tools for biocatalysis (mechanisms applications and scope). *Curr. Opin. Chem. Biol.* **43**, 119–126 (2018).
106. Smith, D. R., Grüschow, S. & Goss, R. J. Scope and potential of halogenases in biosynthetic applications. *Curr. Opin. Chem. Biol.* **17**, 276–283 (2013).
107. Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. H. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **544**, 2–17 (2014).
108. Mudgal, R., Sandhya, S., Chandra, N. & Srinivasan, N. De-DUFing the DUFs: Deciphering distant evolutionary relationships of Domains of Unknown Function using sensitive homology detection methods. *Biol. Direct* **10**, (2015).
109. Goodacre, N. F., Gerloff, D. L. & Uetz, P. Protein Domains of Unknown Function Are Essential in Bacteria. *mBio* **5**, (2013).
110. Umemura, M. *et al.* Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in Aspergillus flavus. *Fungal Genet. Biol.* **68**, 23–30 (2014).
111. Ye, Y. *et al.* Unveiling the Biosynthetic Pathway of the Ribosomally Synthesized and Post-translationally Modified Peptide Ustiloxin B in Filamentous Fungi. *Angew. Chem. Int. Ed.* **55**, 8072–8075 (2016).
112. Umemura, M., Kuriiwa, K., Tamano, K. & Kawarabayasi, Y. Ustiloxin biosynthetic machinery is not compatible between Aspergillus flavus and Ustilaginoidea virens. *Fungal Genet. Biol.* **143**, 103434 (2020).
113. Ye, Y. *et al.* Heterologous production of asperipin-2a: proposal for sequential oxidative macrocyclization by a fungi-specific DUF3328 oxidase. *Org. Biomol. Chem.* **17**, 39–43 (2018).

114. Jiang, Y. *et al*. Biosynthesis of Cyclochlorotine: Identification of the Genes Involved in Oxidative Transformations and Intramolecular O,N-Transacylation. *Org. Lett.* **23**, 2616–2620 (2021).
115. O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23**, 532–547 (2006).
116. Br, L. *et al*. Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71–79 (2018).
117. Miettinen, K. *et al*. The seco-iridoid pathway from Catharanthus roseus. *Nat. Commun.* **5**, 1–12 (2014).
118. Lichman, B. R. *et al*. The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv.* **6**, eaba0721 (2020).
119. Billingsley, J. M., Anguiano, J. L. & Tang, Y. Production of semi-biosynthetic nepetalactone in yeast. *J. Ind. Microbiol. Biotechnol.* **46**, 1365–1370 (2019).
120. Billingsley, J. M. *et al*. Engineering the biocatalytic selectivity of iridoid production in Saccharomyces cerevisiae. *Metab. Eng.* **44**, 117–125 (2017).
121. Yee, D. A. *et al*. Engineered mitochondrial production of monoterpenes in Saccharomyces cerevisiae. *Metab. Eng.* **55**, 76–84 (2019).
122. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci.* **112**, 3205–3210 (2015).
123. Geu-Flores, F. *et al*. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138–142 (2012).
124. McElvain, S. M., Bright, R. D. & Johnson, P. R. The Constituents of the Volatile Oil of Catnip. I. Nepetalic Acid, Nepetalactone and Related Compounds. *J. Am. Chem. Soc.* **63**, 1558–1563 (1941).
125. Bowie, J. U. *et al*. Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production. *Trends Biotechnol.* **38**, 766–778 (2020).
126. Korman, T. P., Opgenorth, P. H. & Bowie, J. U. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun.* **8**, 15526 (2017).
127. Sherkhanov, S. *et al*. Isobutanol production freed from biological limits using synthetic biochemistry. *Nat. Commun.* **11**, 4292 (2020).
128. Black, W. B. *et al*. Engineering a nicotinamide mononucleotide redox cofactor system for biocatalysis. *Nat. Chem. Biol.* **16**, 87–94 (2020).
129. King, E., Maxel, S. & Li, H. Engineering natural and noncanonical nicotinamide cofactor-dependent enzymes: design principles and technology development. *Curr. Opin. Biotechnol.* **66**, 217–226 (2020).
130. Richardson, K. N., Black, W. B. & Li, H. Aldehyde Production in Crude Lysate- and Whole Cell-Based Biotransformation Using a Noncanonical Redox Cofactor System. *ACS Catal.* **10**, 8898–8903 (2020).
131. Opgenorth, P. H., Korman, T. P., Iancu, L. & Bowie, J. U. A molecular rheostat maintains ATP levels to drive a synthetic biochemistry system. *Nat. Chem. Biol.* **13**, 938–942 (2017).
132. Opgenorth, P. H., Korman, T. P. & Bowie, J. U. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nat. Commun.* **5**, 4113 (2014).
133. Collu, G. *et al*. Geraniol 10-hydroxylase11The nucleotide sequence newly reported in this paper has been deposited in the GenBank/EMBL data banks with the accession number AJ251269, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215–220 (2001).

134. Lichman, B. R. *et al*. Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71 (2019).
135. Campbell, A. *et al*. Engineering of a Nepetalactol-Producing Platform Strain of Saccharomyces cerevisiae for the Production of Plant Seco-Iridoids. *ACS Synth. Biol.* **5**, 405–414 (2016).
136. Billingsley, J. M. *et al*. Engineering the biocatalytic selectivity of iridoid production in Saccharomyces cerevisiae. *Metab. Eng.* **44**, 117–125 (2017).
137. Duan, Y., Liu, J., Du, Y., Pei, X. & Li, M. Aspergillus oryzae Biosynthetic Platform for de Novo Iridoid Production. *J. Agric. Food Chem.* **69**, 2501–2511 (2021).
138. Kouda, R. & Yakushiji, F. Recent Advances in Iridoid Chemistry: Biosynthesis and Chemical Synthesis. *Chem. – Asian J.* **15**, 3771–3783 (2020).
139. Lee, S., Paek, S.-M., Yun, H., Kim, N.-J. & Suh, Y.-G. Enantioselective Total Synthesis of a Natural Iridoid. *Org. Lett.* **13**, 3344–3347 (2011).
140. Sim, J., Yoon, I., Yun, H., An, H. & Suh, Y.-G. Divergent synthetic route to new cyclopenta[c]pyran iridoids: syntheses of jatamanin A, F, G and J, gastrolactone and nepetalactone. *Org. Biomol. Chem.* **14**, 1244–1251 (2016).
141. Harnying, W., Neudörfl, J.-M. & Berkessel, A. Enantiospecific Synthesis of Nepetalactones by One-Step Oxidative NHC Catalysis. *Org. Lett.* **22**, 386–390 (2020).
142. Sakai, K., Ishiguro, Y., Funakoshi, K., Ueno, K. & Suemune, H. A novel synthesis of cis-3,4-disubstituted cyclopentanones. *Tetrahedron Lett.* **25**, 961–964 (1984).
143. 1987 Sakai synthesis.pdf.
144. Bühler, C. von, Le-Huu, P. & Urlacher, V. B. Cluster Screening: An Effective Approach for Probing the Substrate Space of Uncharacterized Cytochrome P450s. *ChemBioChem* **14**, 2189–2198 (2013).
145. Bakkes, P. J. *et al*. Engineering of versatile redox partner fusions that support monooxygenase activity of functionally diverse cytochrome P450s. *Sci. Rep.* **7**, 9570 (2017).
146. Walsh, C. T. & Tang, Y. *The Chemical Biology of Human Vitamins*. (2018).
147. Walsh, C. T. & Tang, Y. *Natural Product Biosynthesis*. (2017).
148. Tang, M.-C., Zou, Y., Watanabe, K., Walsh, C. T. & Tang, Y. Oxidative Cyclization in Natural Product Biosynthesis. *Chem. Rev.* **117**, 5226–5333 (2017).
149. Morris, J. S., Caldo, K. M. P., Liang, S. & Facchini, P. J. PR10/Bet v1-like Proteins as Novel Contributors to Plant Biochemical Diversity. *ChemBioChem* **22**, 264–287 (2021).
150. R. Lichman, B. The scaffold-forming steps of plant alkaloid biosynthesis. *Nat. Prod. Rep.* **38**, 103–129 (2021).
151. Sellés Vidal, L., Kelly, C. L., Mordaka, P. M. & Heap, J. T. Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1866**, 327–347 (2018).
152. Lopez de Felipe, F. & Hugenholtz, J. Purification and characterisation of the water forming NADH-oxidase from Lactococcus lactis. *Int. Dairy J.* **11**, 37–44 (2001).
153. Spaans, S. K., Weusthuis, R. A., Van Der Oost, J. & Kengen, S. W. M. NADPH-generating systems in bacteria and archaea. *Front. Microbiol.* **6**, (2015).
154. Mordhorst, S. & Andexer, J. N. Round, round we go – strategies for enzymatic cofactor regeneration. *Nat. Prod. Rep.* (2020) doi:10.1039/D0NP00004C.
155. Bologna, F. P., Andreo, C. S. & Drincovich, M. F. Escherichia coli Malic Enzymes: Two Isoforms with Substantial Differences in Kinetic Properties, Metabolic Regulation, and Structure. *J. Bacteriol.* **189**, 5937–5946 (2007).

156. Morlock, L. K., Böttcher, D. & Bornscheuer, U. T. Simultaneous detection of NADPH consumption and H2O2 production using the Ampliflu™ Red assay for screening of P450 activities and uncoupling. *Appl. Microbiol. Biotechnol.* **102**, 985–994 (2018).
157. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* **11**, 449–455 (2014).
158. Chavez, A. *et al*. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
159. Zalatan, J. G. *et al*. Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160**, 339–350 (2015).
160. Lee, Y. J., Hoynes-O'Connor, A., Leong, M. C. & Moon, T. S. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Res.* **44**, 2462–2473 (2016).
161. Li, L. *et al*. Biochemical Characterization of a Eukaryotic Decalin-Forming Diels–Alderase. *J. Am. Chem. Soc.* **138**, 15837–15840 (2016).
162. Billingsley, J. M., Anguiano, J. L. & Tang, Y. Production of semi-biosynthetic nepetalactone in yeast. *J. Ind. Microbiol. Biotechnol.* **46**, 1365–1370 (2019).
163. Ma, S. M. *et al*. Complete Reconstitution of a Highly Reducing Iterative Polyketide Synthase. *Science* **326**, 589–592 (2009).
164. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J. Cell Mol. Biol.* **16**, 735–743 (1998).
165. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
166. Ohtawa, M. *et al*. Stereoselective Total Synthesis of Atpenins A4 and B, Harzianopyridone, and NBRI23477 B. *Chem. Pharm. Bull. (Tokyo)* **60**, 898–906 (2012).
167. Kumagai, H. *et al*. THE STRUCTURES OF ATPENINS A4, A5 AND B, NEW ANTIFUNGAL ANTIBIOTICS PRODUCED BY PENICILLIUM SP. *J. Antibiot. (Tokyo)* **43**, 1553–1558 (1990).