**Title**
Finding Diamonds in the Rough: Resistance-Gene-Directed Discovery of Fungal Natural Products



**Author**
Yan, Yan

**Publication Date**
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Finding Diamonds in the Rough: Resistance-Gene-Directed Discovery of Fungal Natural Products

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemical Engineering

by

Yan Yan

2019

ABSTRACT OF THE DISSERTATION

Finding Diamonds in the Rough: Resistance-Gene-Directed Discovery of Fungal Natural Products

by

Yan Yan

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2019

Professor Yi Tang, Chair

Bioactive natural products have evolved to inhibit specific cellular targets and have served as lead molecules for health and agricultural applications for the last century. The post-genomics era has brought a renaissance in natural product discovery using synthetic biology tools. However, compared to traditional bioactivity-guided approaches, genome mining of natural products with specific and potent biological activities remains challenging. Here we proposed a resistance gene directed strategy to discover new natural products of desired bioactivity from genomic data. Our approach is based on the co-clustering of a self-resistance gene in the natural product biosynthetic gene cluster, which serves as a window to potential biological activity of the encoded compound. To demonstrate the feasibility, we first successfully located and verified the biosynthetic gene

cluster of a glyceraldehyde-3-phosphate dehydrogenase inhibitor heptelidic acid by using a resistance gene. We showed the glyceraldehyde-3-phosphate dehydrogenase homolog *hepG*, co-localized with biosynthetic genes, is able to confer self-resistance to the producing organism using in vitro biochemical analysis. Next we present the discovery and validation of a highly potent herbicide lead that targets a critical metabolic enzyme that is required for plant survival. We queried dihydroxyacid dehydratase in the branched-chain amino acid biosynthetic pathway in plants, which is often targeted for herbicide development. We also showed that the fungal sesquiterpenoid aspterric acid discovered using this method is a submicromolar inhibitor of DHAD and is effective as an herbicide in spray applications. The self-resistance gene *astD* was validated to be insensitive to aspterric acid and was deployed as a transgene in the establishment of plants that are resistant to aspterric acid. This herbicide-resistance gene combination complements urgent efforts in overcoming weed resistance. Our discovery demonstrates the potential of using a resistance-gene directed approach in the mining of bioactive natural products.

The dissertation of Yan Yan is approved.

Kendall N. Houk

Steven E. Jacobsen

Yvonne Y. Chen

Yi Tang, Committee Chair

University of California, Los Angeles

2019

## Table of Contents

## List of figures

**Acknowledgements**

Section 2.1.1 contains materials from Yan Yan, Qikun Liu, Steven E. Jacobsen, Yi Tang. "The impact of natural product discovery in agriculture." *EMBO Rep.* **2018**, 19: e46824.

Section 2.2 contains materials from Yan Yan, Qikun Liu, Xin Zang, Shuguang Yuan, Undramaa Bat-Erdene, Calvin Nguyen, Jianhua Gan, Jiahai Zhou, Steven E. Jacobsen, Yi Tang. "Resistance-gene directed discovery of a natural product herbicide with a new mode of action." *Nature* **2018**, 559, 415-418.

This work was supported by the NIH (1DP1GM106413 and 1R35GM118056) to Y.T. and CAS (XDB20000000) to J.Z.. S.E.J is an Investigator of the Howard Hughes Medical Institute. Q.L. is supported by the NIH (F32) Postdoctoral Fellowship.

First and foremost, I would like to thank my advisor Professor Yi Tang for his advice, inspiration and guidance on my projects at UCLA. He is always patient and confident in my projects when they are not going smoothly. He never doubted my abilities when some of my projects got scooped. He continuously remained tolerant on my behaviors, even when sometimes I made careless mistakes. He is also very supportive in helping me start my future scientific research in China. I really appreciate all of these favors! He also set up an excellent example for me on how to guide scientific research, organize a lab, collaborate with others and inspire students and postdocs. It is my great honor to work and learn in his lab!

I would like to thank my collaborator Professor Steven Jacobsen from Department of Molecular, Cell, and Developmental Biology at UCLA. I appreciate his very generous help on plant experiments and for providing inspiration on the herbicide project! He also helped us to patent our discovery. I am very grateful that he agreed to join my doctoral thesis defense committee at the last minute to help me graduate on time!

I would like to thank my collaborator Professor Jiahai Zhou from State Key Laboratory of Bio-organic and Natural Products Chemistry in Shanghai Institute of Organic Chemistry. He is always supportive in helping me to determine protein structures, and has also given us invaluable suggestions. I still remember how he stayed up for the whole night to collect X-ray diffraction data of DHAD with Professor Tang at the National Synchrotron Radiation Research Center in Taiwan.

I would like thank my family, especially my wife Hong Zhang. She came to Los Angeles to accompany and take care of me for three years in aspects of both life and research. She tried many preliminary plant growth assays and mammalian cell associated experiment for my projects. She also came up with many brilliant ideas to troubleshoot problems within my projects. I am always grateful that she is very patient and waits for me to do experiment for hours in the school parking lot from time to time!

I would like to thank Dr. Qikun Liu, Professor Hsiao-Ching Lin, Xin Zang, Professor Shuguang Yuan, Undramaa Bat-Erdene and Jan Schlemmer as the major coworkers on my projects. Qikun Liu from Professor Jacobsen's lab helped me a lot to do plant associated experiments, and also helped me to write and revise the manuscript. Professor Hsiao-Ching Lin initiated the heptelidic acid project and mentored me to get familiar with the lab during my first year. Xin Zang from Professor Zhou's lab helped us to crystallize both the DHAD-aspterric acid complex and the GAPDH-heptelidic acid complex. Professor Shuguang Yuan from the Laboratory of Physical Chemistry of Polymers and Membranes at Ecole Polytechnique Fédérale de Lausanne in Switzerland helped us on homology modeling of AstD, and docking of the substrate or aspterric acid to plant DHAD. Undramaa Bat-Erdene helped me to do yeast associated experiments to prove the function of AstD as a resistance gene in vivo, and she is trying to do genome editing to confer herbicide tolerance to crops. Jan Schlemmer helped me to improve the production of aspterric acid in engineered yeast strains.

All my collaborators and coworkers are all very professional, responsible, reliable and respectable colleagues, who made a lot of very intellectual contributions to my project. It is my great honor to work with and learn from every one of you!

I would like to thank Stanford Genome Technology Center for the *Saccharomyces cerevisiae* DHY *ΔURA3* strain. The diffraction data of *holo*-pDHAD was collected at beamline BL19U1 in Shanghai Synchrotron Radiation Facility (SSRF). The molecular modeling was performed at the Interdisciplinary Centre for Mathematical and Computational Modeling in Warsaw (GB70-3 & GB71-3).

I would also like to thank Professor Tang, Professor Kendall Houk, Professor Steven E. Jacobsen, and Professor Yvonne Chen for joining my doctoral thesis defense Committee, as well as Professor Neil Garg who joined my prospectus oral exam committee. Thank you for all the comments and advice to help me improve the quality of my projects and the skills to present my work.

I would like to thank John Billingsley, Nicholas Liu and Leibniz Hang who are in the same class as me. They helped me a lot with the Chemical Engineering and Chemistry classes in my first year because I had almost no Chemical Engineering background when I came to UCLA. I can still remember working on homework and preparing for all kinds of exams together at the Weyburn study room. It is hard to imagine how I could pass those required exams without them!

At last I would like to thank all Tang lab members and alumni who helped and worked together with me in the lab:

Graduate students: Dr. Muxun Zhao, Dr. Ralph Cacho, Dr. Anthony DeNicola, Dr. Carly Bond, Sunny Hung, Danielle Yee, Eun Bin Go, Joshua Misa, Cooper Jamieson.

Postdoctoral fellows: Dr. Wei Xu, Dr. Mancheng Tang, Dr. Zou Yi, Dr. Shushan Gao, Dr. Thomas Kakule, Dr. Xia Yu, Dr. Li Li, Dr. Yang Hai, Dr. Mengbin Chen, Dr. Masao Ohashi, Dr. Zhuan Zhang, Dr. Bora Shin, Dr. Dewei Gao,

Visiting scientists: Dr. Xuming Mao, Dr. Dehai Li, Dr. Michio Sato, Jan Schlemmer, Dr. Youming Ying, Dr. Yongxiang Song, Dr. Quansheng Du, Dr. Tan Dan, Dr. Liu Ling and Dr. Wei Cheng.

**Vita**

| | | |
|---|---|---|
| 2014.9 – present | graduate student | • University of California, Los Angeles (UCLA)<br>Major: Chemical Engineering |
| 2013.8 – 2014.7 | research assistant | • Shanghai Institute of Organic Chemistry (SIOC), Chinese Academy of Sciences, China |
| 2009.9 – 2013.7 | M.S | • Shanghai Institute of Organic Chemistry (SIOC), Chinese Academy of Sciences, China<br>Major: Organic chemistry |
| 2005.9 – 2008.7 | B.S | • China Agricultural University (CAU), China<br>Major: Honors' Program (Life Science) |

Publications

Yan Yan, Qikun Liu, Steven E. Jacobsen, Yi Tang. "The impact of natural product discovery in agriculture." *EMBO Rep.* **2018**, 19: e46824.

Yan Yan, Qikun Liu, Xin Zang, Shuguang Yuan, Undramaa Bat-Erdene, Calvin Nguyen, Jianhua Gan, Jiahai Zhou, Steven E. Jacobsen, Yi Tang. "Resistance-gene directed discovery of a natural product herbicide with a new mode of action." *Nature* **2018**, 559, 415-418. (highlighted by C&EN news, *Nat. Plants*, *Mol. Plant*, and *Pest Manag. Sci.*)

Zhenhua Tian, Peng Sun, Yan Yan (Co-first author), Zhuhua Wu, Qinfei Zheng, Shuaixiang Zhou, Hua Zhang, Futao Yu, Xinying Jia, Dandan Chen, Attila Mandi, Tibor Kurtan, Wen Liu. "An enzymatic [4+2] cyclization cascade creates the pentacyclic core of pyrroindomycins." *Nat. Chem. Biol.* **2015**, 11, 259-265.

Chenchen Chang, Rong Huang, Yan Yan (Co-first author), Hongmin Ma, Zheng Dai, Benying Zhang, Zixin Deng, Wen Liu, Xudong Qu. "Uncovering the formation and selection of benzylmalonyl-CoA from the biosynthesis of splenocin and enterocin reveals a versatile way to introduce amino acids into polyketide carbon scaffolds." *J. Am. Chem. Soc.* **2015**, 137, 4183–4190.

Lihan Zhang, Takahiro Mori, Qingfei Zheng, Takayoshi Awakawa, Yan Yan, Wen Liu, Ikuro Abe. "Rational control of polyketide extender units by structure-based engineering of a crotonyl-CoA carboxylase/reductase in antimycin biosynthesis." *Angew. Chem. Int. Ed.* **2015**, 54, 13462-13465.

Lihan Zhang, Jing Chen, Takahiro Mori, Yan Yan, Wen Liu, Ikuro Abe. "Crystallization and preliminary X-ray diffraction analysis of AntE, a crotonyl-CoA carboxylase/reductase from *Streptomyces sp.* NRRL 2288." *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2014**, 70, 734-737.

Yan Yan, Jing Chen, Lihan Zhang, Qingfei Zheng, Ying Han, Hua Zhang, Daozhong Zhang, Takayoshi Awakawa, Ikuro Abe, and Wen Liu. "Multiplexing of combinatorial chemistry in antimycin biosynthesis expands the molecular diversity and utility." *Angew. Chem. Int. Ed.* **2013**, 52, 12308-12312.

Yan Yan, Lihan Zhang, Takuya Ito, Xudong Qu, Yoshinori Asakawa, Takayoshi Awakawa, Ikuro Abe, Wen Liu. "Biosynthetic pathway for high structural diversity of a common dilactone core in antimycin production." *Org. Lett.* **2012**, 14, 4142–4145.

Patents

Yan Yan, Yi Tang, Qikun Liu, Steven E. Jacobsen. "Herbicidal compositions and methods involving aspterric acid" **2017**, priority number: US201762474528.

Wen Liu, Yan Yan. "The preparation and application of antimycin analogues" **2013**, priority number: CN201310380322.

## 1. Introduction

### 1.1 Natural products

Natural products (NPs) are small molecule chemicals produced by living organisms with diverse properties such as toxins, siderophores, pigments, antibiotics, cytostatics, immunosuppressants and anticancer agents[1-3]. As NPs coevolved together with biomacromolecules in nature, they possess potent bioactivities allowing the producer to kill competitors by inhibiting certain functional enzymes that are essential for survival[4-6]. NPs are not only beneficial to their producers – they can also benefit human beings. The most famous NP, penicillin, has saved millions of lives throughout the world since it was discovered in the 1940s[7]. By inhibiting bacterial cell wall formation, this molecule produced by *Penicillium* is able to effectively suppress the growth of bacteria[7,8]. Since the discovery of this first antibiotic, scientists found out that microorganisms are prolific producers of bioactive NPs. Among these potent bioactive NPs, some may be developed into important human therapeutics to treat diseases, while others can be applied as pesticides to increase crop production by controlling weeds, plant pathogens and insects[9-11].

Between 1981 and 2010, 34% of small molecule agents approved by the US Food and Drug Administration (FDA) for treating infectious diseases, cancer, lipid disorders and immunomodulation were NPs or their direct derivatives[3] (Fig. 1). For example, vancomycin is used to treat Gram-positive bacterial infections by inhibiting cell wall biosynthesis[12]. Taxol, which blocks microtubule assembly, is a chemotherapy agent used to treat different types of cancers[13]. Through inhibition of a specific phosphatidylinositol 3-kinase-related kinase known as mTOR, rapamycin is used as an immunosuppressant in organ transplantation[14]. From 1997 to 2010, NPs and their derivatives comprise more than 35% of all new registered pesticide ingredients[9]. For

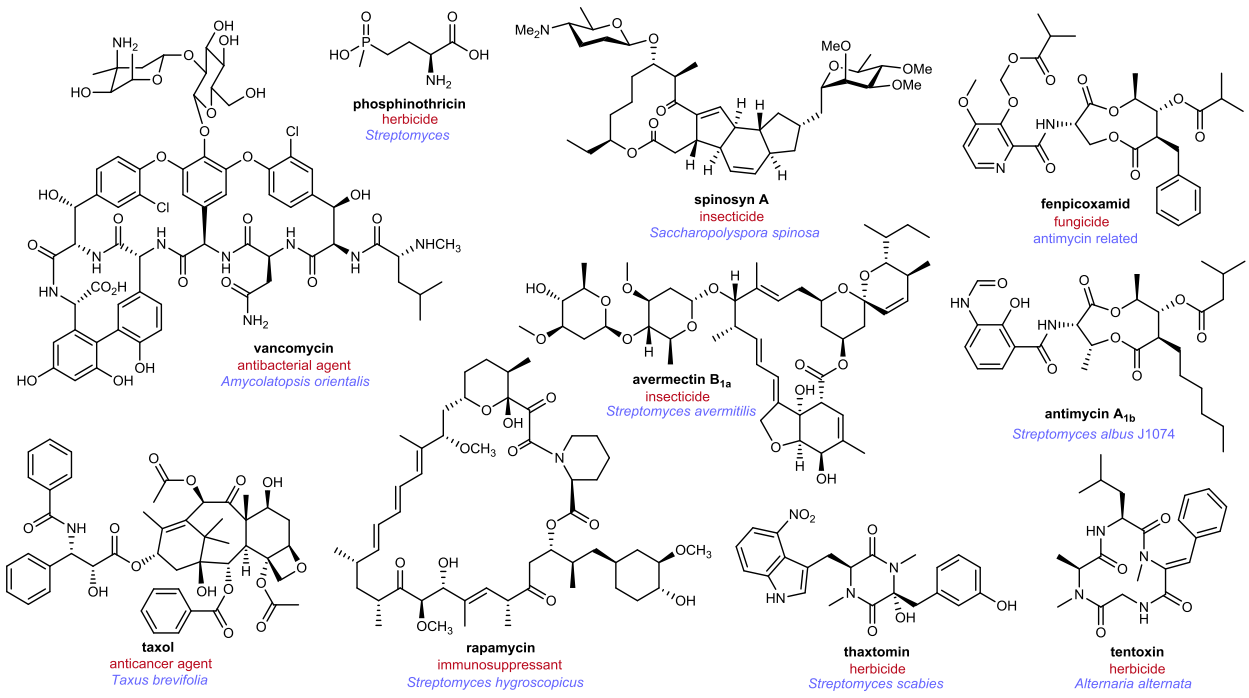



*Figure 1. The structures of representative pharmaceuticals and pesticides of natural product origin.*

*The bioactivities (red) and sources (blue) of natural molecules are listed under their names.*

example, the natural insecticide spinosyns and avermectins produced by *Saccharopolyspora spinose* and *Streptomyces avermitilis* respectively can effectively paralyze insects by hyperexcitation of their nervous system[15,16]. By targeting glutamine synthetase, glufosinate is able to kill plants due to a buildup of ammonia in the thylakoid lumen, which leads to decoupling of photophosphorylation[17] (Fig. 1). The antimycin related commercialized fungicide fenpicoxamid is an inhibitor of cellular respiration[18] (Fig. 1). Phosphinothricin, also known as glufosinate, produced by *Streptomyces* has been commercialized as an herbicide under the tradename of Finale® by Bayer[19] (Fig. 1). There are also NPs that possess novel modes of action compare to existing pesticides, such as thaxtomin and tentoxin, which are able to disrupt cellulose biosynthesis and energy transfer respectively[20] (Fig. 1). For some NPs, however, even though their potent

bioactivity and critical mode of action have been validated, their unaffordable cost of production has limited large scale application in the field.

### 1.2 Machinery of natural product biosynthesis

Although the structures and bioactivities of secondary metabolites varies dramatically, all of them are assembled from a common reservoir of primary metabolite building blocks using conserved biosynthetic machineries. These conserved machineries that have been well characterized to date include polyketides assembled by polyketide synthases (PKSs), non-ribosomally synthesized peptides assembled by non-ribosomal peptide synthetases (NRPSs), ribosomally synthesized and post-translationally modified peptides (RiPPs) and terpenoids assembled by terpene synthases (TSs)[21-25].

#### 1.2.1 Polyketide synthase

Polyketides comprise a large family of natural products in bacteria and fungi, which include many important clinical therapeutics such as erythromycin, rapamycin, lovastatin and avermectin[21,26] (Fig. 1). They are secondary metabolites derived from precursors containing multiple carbonyl methylene groups, which are assembled on a PKS using simple $C_2$ units as substrates[26] (Fig. 2). Very similar to fatty acid synthases, PKSs are a family of multi-domain enzymes that need several catalytic domains to work in coordination to condense simple building blocks of acyl-CoAs to form a polyketide backbone[24,27,28]. For example in the biosynthetic pathway of erythromycin A, 6-deoxyerythronolide B synthase DEBS1, DEBS2 and DEBS3 work in coordination to form the polyketide backbone precursor 6-deoxyerythronolide B using one propionyl-CoA as starter unit and 6 methylmalonyl-CoAs as extender units[29]. PKSs are usually above 100 kD, making them very large in size. They can be further divided into modules that together are responsible for one complete round of C2 unit extension of the growing polyketide

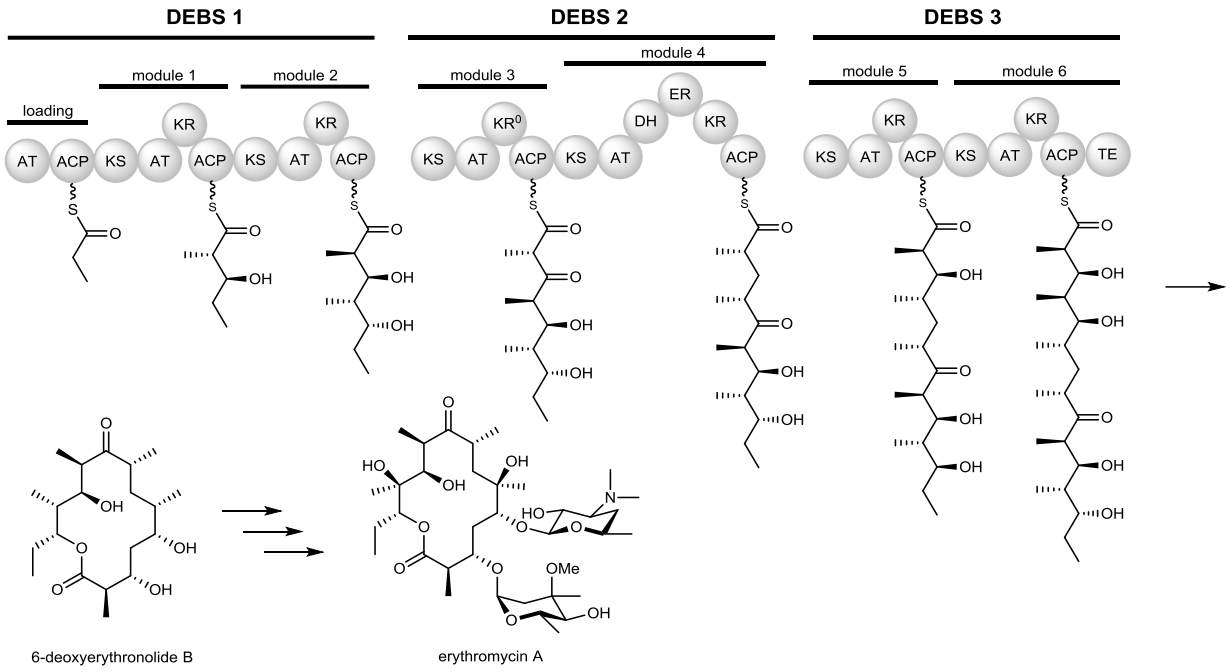



*Figure 2. Example of polyketide biosynthesis machinery.*

*The polyketide scaffold of 6-deoxyerythronolide was assembled on 6-deoxyerythronolide B synthase. Erythromycin A was synthesized through further post PKS assembly line modifications.*

chain. There are multiple catalytic domain within each module, such as an acyl carrier protein (ACP), acyltransferase (AT), β-ketoacyl-S-ACP synthase (KS), ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), and thioesterase (TE).

DEBS has seven ACP domains, one at the N terminus of a loading didomain, and six others at the C terminus of each chain-elongation module. The elongating chain is covalently tethered to the ACP via a phosphopantetheinyl prosthetic group, which facilitates the accession of the growing chain to other catalytic domains within the PKS[21,30]. Thus the ACP domains must be primed by a posttranslational modification of the hydroxyl group of a serine residue with phosphopantetheine using a PKS specific 4'-phosphopantetheine transferase (PPTase)[31]. Besides ACP, there is an AT domain on each module, and the loading didomain of the DEBS. The AT domain is responsible

for loading the acyl from the starter unit or extender units to the phosphopantetheinyl-SH group of the ACP. The reaction follows a ping-pong mechanism involving covalent catalysis. Besides the loading function, AT domains are also the key gatekeepers to selectively load the starter or extender units due to the intrinsic structural specificity at the active site[32,33].

Each module of DEBS also contains a KS domain, which catalyzes a polyketide chain elongation reaction[21]. The cysteine at the active of the KS first attacks the carbonyl of the thioester bond between the growing polyketide chain and ACP of the previous module to form a covalent acyl-*S*-cysteine bond between the elongating chain and the KS domain. Then a decarboxylative condensation of this thioester with an alkylmalonyl-S-ACP is carried out to produce a 2-alkyl-3-ketoacyl-*S*-ACP product. The configuration of C-2 in this 2-alkyl-3-ketoacyl-*S*-ACP product is also controlled by KS domain during the condensation[34].

In comparison with the ACP, AT and KS domains, the KR domain is not a necessary catalytic domain to complete a round of $C_2$ unit elongation. Although each module of DEBS contains a KR domain, the KR domain in module 3 is not active. KR domains are NADPH dependent, and catalyze a nucleophilic hydride attack of the keto at C-3 of the 2-alkyl-3-ketoacyl-*S*-ACP product formed by the KS. This reduction reaction generates 2-alkyl-3-hydroyl-*S*-ACP as a resulting product, meanwhile the configuration of C-3 is also determined by the specific structural properties of the active site[21,35].

DH domain is a tailoring domain on the polyketide assembly line that functions immediately following the KR domain. In DEBS, only module 4 contains this domain, which catalyzes the dehydration of the 2-alkyl-3-hydroyl-*S*-ACP to give the α,β-unsaturated product using a general base. The ER domain usually further tailors the α,β-unsaturated product by catalyzing a

nucleophilic hydride attack of the α,β-unsaturated double bond to give a saturated product[21]. Similar to the KS and KR domain, the ER domain also controls the configuration of C-2 of the saturated product. There are also other tailoring domains involved in PKS machinery, such as the methyl transferase (MT) domain, which functions as a *C*-methyl transferase to catalyze a methylation reaction at C-2 position of the growing polyketide chain[21].

After the assembly of the polyketide backbone is completed, the product needs to be released from the enzyme. One of the typical polyketide chain releasing mechanisms is through catalysis via the TE domain[21]. This domain belongs to the α/β-hydrolases family, which contains an active site catalytic triad of serine, histidine and aspartate residues[36]. For example, in erythromycin A biosynthesis, the His259 promotes nucleophilic attack by Ser142 on the acyl-ACP thioester to form an enzyme bound intermediate, which is followed by another nucleophilic attack by its C-13 hydroxyl to release the lactonized product 6-deoxyerythronolide B. Besides macrolactonization to terminate polyketide assembly, the TE domain may also catalyze product releasing through other mechanisms, such as hydrolysis, macrothiolactonization, hydrolysis followed by decarboxylation (dehydration), and Claisen condensation[29].

Based on organization of the catalytic domains and polyketide chain elongation mechanisms, PKSs are generally classified into three types[37,38]. PKSs like DEBS, which are multifunctional and organized into distinct linear acting modules are classified as modular type I PKSs. In addition, there are also type I PKSs which have iteratively acting catalytic domains: the polyketide chain is elongated on one module, and each catalytic domain on this module is used iteratively for chain extension[38]. Type II PKSs are multienzyme complexes that are comprised of a set of individual enzymes[37]. Similar to the catalytic domains in type I PKSs, these enzymes work in coordination to assemble the polyketide chain. Typical type II PKSs only contain minimal enzyme units to

elongate the polyketide chain, which are two ketosynthase units ($KS_\alpha$ and $KS_\beta$) and an ACP, and the polyketide products are usually aromatic. Instead of using the AT domain to load starter units or extender units, the KS is able to directly accept the acyl from CoA for chain elongation. Different from type I and type II PKSs, type III PKSs are independent of the ACP; they only have individual KSs as catalytic domains[38]. Instead of using ACP, they directly use CoA as the polyketide chain carrier to facilitate chain extension.

### 1.2.2 Nonribosomal peptide synthetase

Non-ribosomally synthesized peptides (NRPs) are another remarkable family of secondary metabolites that are derived from peptide precursors[21]. NRPs such as penicillin, vancomycin, and tentoxin, are produced predominantly by bacteria and fungi; NRPs are assembled by nonribosomal

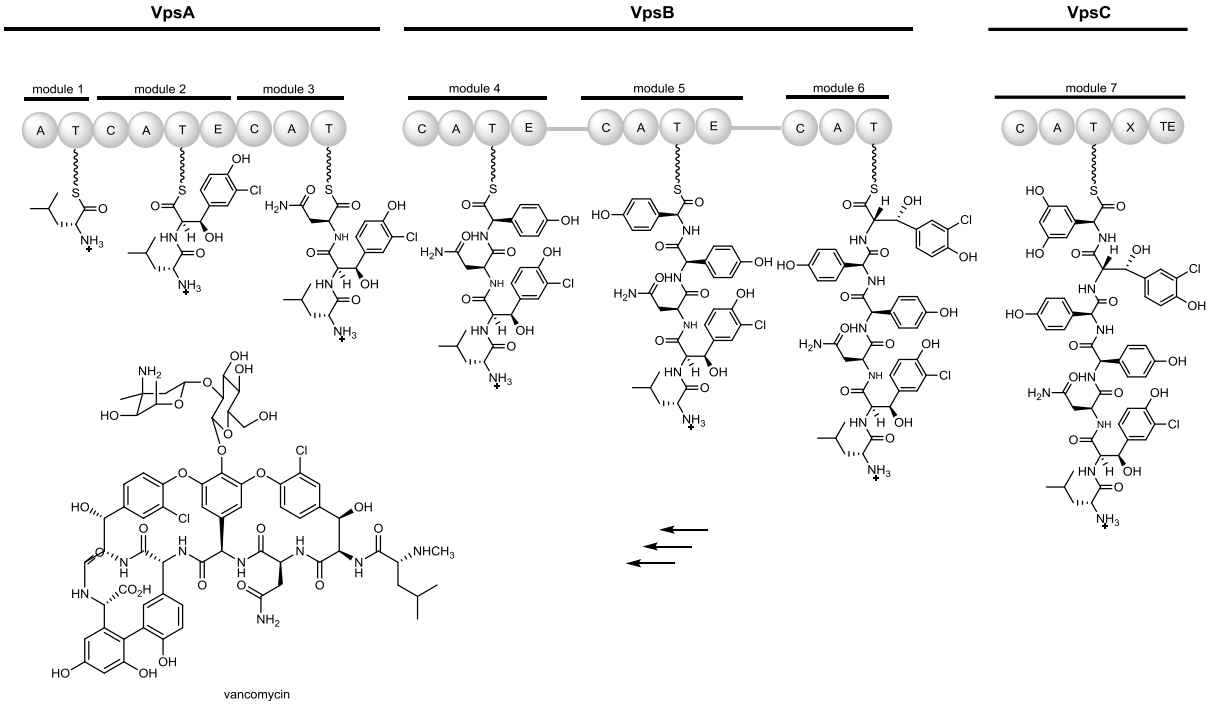



*Figure 3. Example of nonribosomal peptide biosynthesis machinery.*

*The vancomycin backbone was assembled on the NRPS. X = domain of unknown function.*

peptide synthetases (NRPSs), instead of ribosomes, and use amino acids as substrates. Similar to type I PKSs, NRPSs are megasynthetases comprised of multiple catalytic domains to work in coordination for polypeptide assembly[39,40]. Typical catalytic domains within an NRPS include the adenylation domain (A), thiolation domain (T), condensation domain (C), epimerization domain (E) and TE.

Similar to the ACP in PKSs, the T domain contains a covalently tethered phosphopantetheinyl prosthetic group, which is responsible for carrying the elongating polypeptide chain to facilitate the accession of the growing chain to other catalytic domains within the NRPS. Due to its essential function, there is also a T domain on every module of the NPRS[21] (Fig. 3).

In analogy to AT domain in PKSs, the NRPS adenylation domain is responsible for activating the amino acid substrates and loading them onto an NRPS assembly line. For example, in vancomycin biosynthesis, leucine is first adenylated, and then tethered to the T domain through a thioester bond[41]. Besides recruiting free amino acids, the A domain also acts as the gatekeeper to select amino acids due to its conserved structural properties at the active sites. As a result, the substrate specificity of A domains can be predicted based on the specificity-conferring residues at the active site of a given NRPS[42,43]. For example in vancomycin biosynthesis, the NRPSs including VpsA, VpsB and VpsC contain seven modules in total, all of which have T domains and A domains (Fig. 3).

Like KS domains in PKSs, C domains are able to catalyze chain-elongating of the polypeptide[21,44,45]. Although the C domain is not necessary on the loading module, it is essential within the elongating modules. The peptide bond is formed by attacking of the upstream carbonyl of the peptidyl thioester tethered to an upstream T domain using the free amine of the downstream

aminoacyl thioester tethered to another downstream T domain[21,45]. Meanwhile, chain elongation by one aminoacyl residue is completed together with the translocation of the polypeptide chain from the upstream T domain, to the next downstream T domain. In vancomycin biosynthesis, a heptapeptide chain is assembled on the NRPS by six consecutive condensation events catalyzed by six C domains on the assembly line[41] (Fig. 3).

Compared to ribosome dependent peptides, NRPs possess non-proteinogenic amino acids, which include D-amino acids, modified proteinogenic amino acids as well as other acid substrates that do not have amines on the α-carbon[21]. One way to generate D-amino acids is by epimerizing the α-carbon of aminoacyl group bound to the T domain in the presence of E domain on NRPSs. The E domain promoted epimerization is initiated by the generation of a mixture of racemic D and L T-tethered aminoacyl, wherein the downstream C domain will only accept the D-aminoacyl at the enantioselective donor sites. Like the tailoring domains in PKSs, they are not essential for polypeptide chain elongation. For example, only three out of seven modules in the vancomycin NRPSs have E domains (Fig. 3).

To generate peptides with a greater variety of non-proteinogenic amino acid residues, there are also other tailoring domains such as methyltransferase domains (MT), oxidase domains (Ox), and cyclization domains (Cy)[21,46-48]. These tailoring domains work in coordination with the other catalytic domains on the polypeptide assembly line to generate diverse structures of NRPs. After the chain extension of the polypeptide is completed, NRPSs also use TE domains to release the product from the assembly line. Like the machinery of PKS, the active site serine of TE domain carries out a nucleophilic attack on the carbonyl of the peptidyl thioester to give a covalent TE-substrate intermediate[49]. The peptide chain can be further released by either hydrolysis or cyclization through intramolecular peptide bond formation[50]. Compared to the diversified types of

PKSs machinery, all NRPSs are modular and resemble the type I PKS assembly machinery. Like the modules on iterative type I PKS, some catalytic domains on NRPSs are also able to be used more than once, resulting in the possibility for iterative assembly of peptides on the NRPS[51].

### 1.2.3 Polyketide synthase and nonribosomal peptide synthetase hybrid

Besides PKSs and NRPSs, there is also another type of multi domain enzyme that recruits both type I PKS and NRPS assembly machinery, termed PKS-NRPS hybrid assembly lines[21]. Due to high similarity of the assembly machinery of type I PKS and NRPS, it is reasonable to think that the growing intermediate chain can be transferred and elongated between a type I PKS and an NRPS to make a polyketide-NRP hybrid product. Some hybrid assembly lines recruit additional PKS machinery, while others recruit more NRPS machinery. For example, in antimycin biosynthesis: AntC is an NRPS, which has two modules in charge of a two-step polypeptide chain extension; AntD is PKS, which has only one domain that is responsible for the last step chain extension and release of the product[52,53] (Fig. 4).

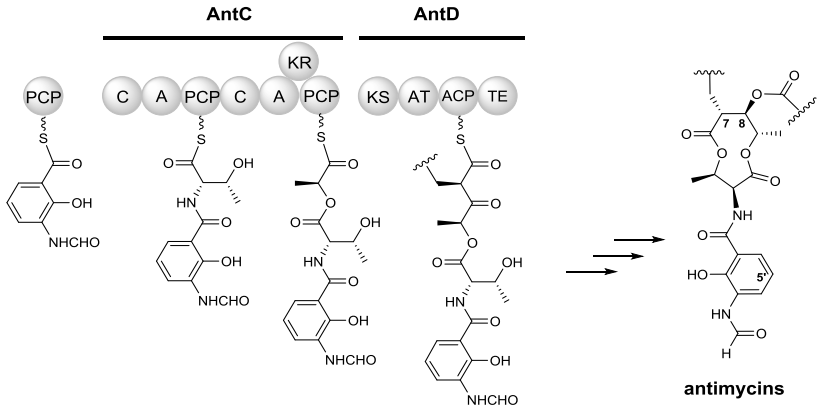



*Figure 4. Example of PKS-NRPS hybrid assembly line machinery.*

*Antimycin biosynthesis using a PKS-NRPS hybrid machinery.*

### 1.2.4 Terpenoid synthase

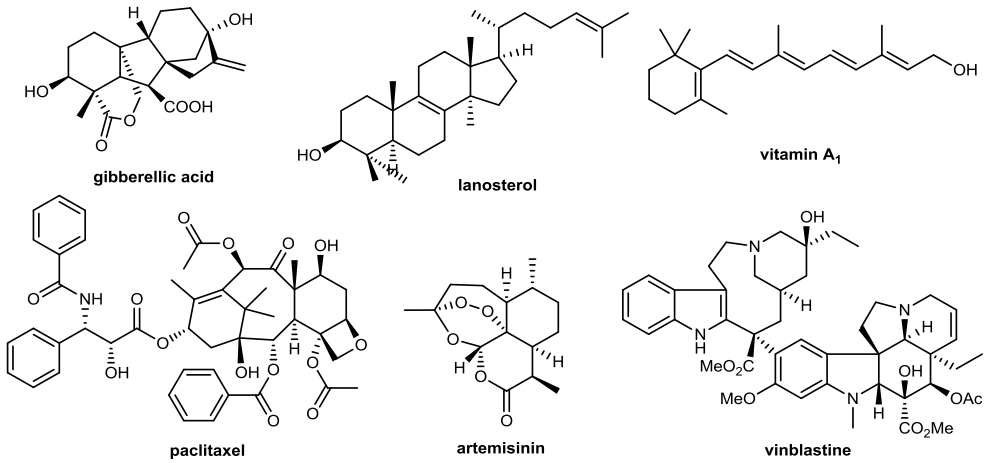



*Figure 5. Important terpenoid related natural products as pheromones, vitamins and pharmaceuticals of natural product origin.*

Terpenoids are critical for nearly every single living organism on this planet due to their essential role in primary metabolism as vitamins, hormones, and steroids[23,54] (Fig. 5). Besides their essential role in cell survival, they also belong to a versatile family of natural products produced by bacteria, fungi, plants, and insects[55,56]. Some terpenoids have remarkable therapeutic properties, such as the antimalarial agent artemisinin, and the anticancer agents paclitaxel and vinblastine (Fig. 5)[57,58]. The diverse structures of these terpenoids provide specific bioactivities, but make them challenging to obtain using organic synthesis. They all share a distinct, common structural feature, which are the $C_5$ isoprene units in the molecule derived from dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) precursors. There are two different primary metabolic pathways in different organisms which generate these $C_5$ units for terpenoids biosynthesis, which include mevalonate pathway and methylerythritol (MEP) phosphate pathway[59,60].

The assembly machinery of terpenoids is different from the biosynthesis of polyketides and NRPs, which are assembled using modular multi-domain enzyme complexes. Instead, the

biosynthesis of structurally diversified terpenoids share another common assembly machinery which involves prenyltransferases and terpene synthases[23]. Firstly, the $C_5$ building blocks are usually coupled using prenyltransferases to form longer linear building blocks such as the $C_{10}$ unit geranyl diphosphate (GPP), $C_{15}$ unit farnesyl diphosphate (FPP), $C_{20}$ unit geranylgeranyl

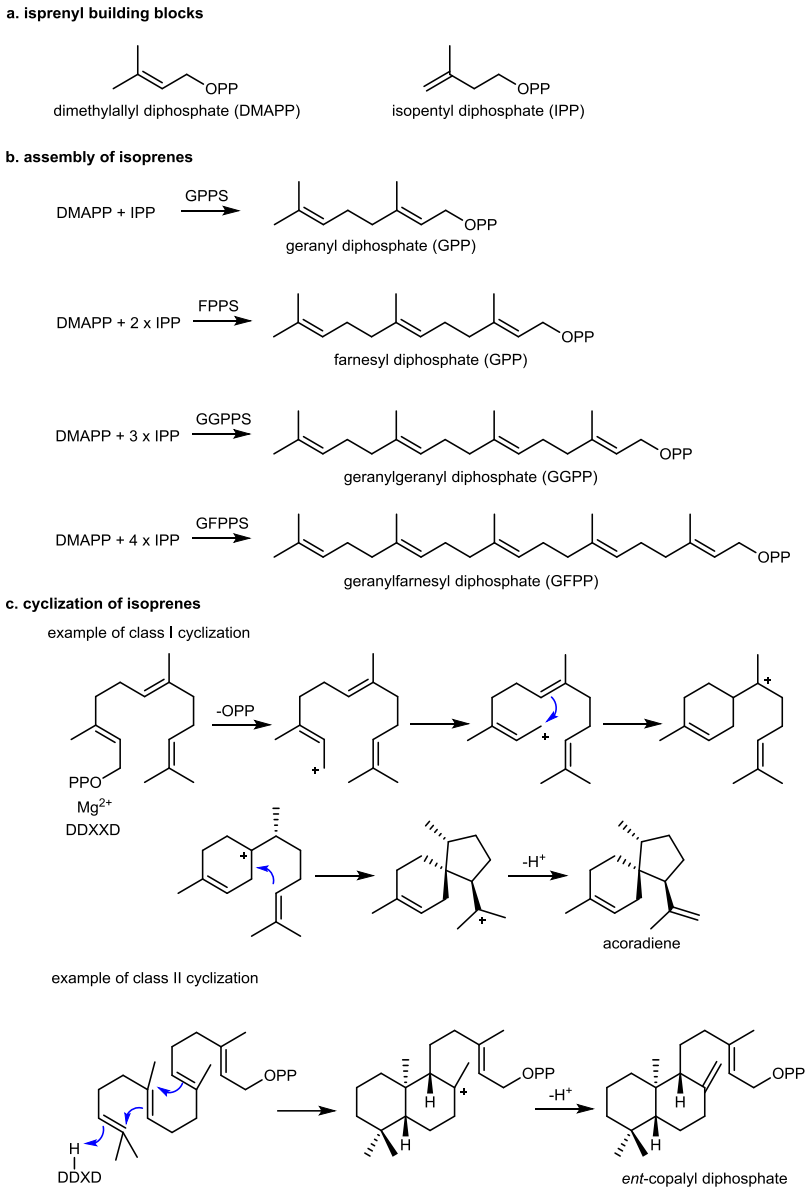



*Figure 6. Terpenoid assembly line machinery.*

***a.*** *isoprene building blocks.* ***b.*** *examples of isoprene assembly using prenyltransferases.* ***c.*** *examples of isoprene cyclization using terpenoid cyclases.*

diphosphate (GGPP), and $C_{25}$ unit geranylfarnesyl diphosphate (GFPP). These linear substrates are further processed through carbocation rearrangement or carbon-carbon bond formation using a terpene synthase (Fig 6).

Prenyltransferases are responsible for a head-to-tail connection of the isoprenes[61]. For example, FPP synthase catalyzes the formation of FPP using one DMAPP and two IPPs. The distinct feature of all prenyltransferases is a conserved DDXXD amino acid residue motif that is able to bind metal cations (mostly $Mg^{2+}$) at the active site[61]. Other modes of connection between the isoprenes are also reported, however the head-to-tail connection is the most common observed in terpenoid biosynthesis[62]. The reaction is initiated by coordination of the diphosphate group of DMAPP to $Mg^{2+}$, which is followed by cleavage of the diphosphate group to generate an allylic carbocation as an intermediate. The electrons on the C-C double bond then attack the carbocation to form a new C-C bond between DMAPP and IPP[23]. Finally, proton extraction is carried out, and a new C-C double bond is formed (Fig 6). Besides coupling reactions between linear diphosphate substrates, there are also prenyltransferases that can catalyze the connection between linear diphosphate substrates to aromatic rings or heteroatoms through a nucleophilic attack of the carbocation by electrons on the aromatic ring or lone paired electron of the heteroatoms[23,63].

To establish more complex isoprenoid scaffold, the linear isoprene diphosphate needs to be further processed by terpenoid cyclases. The enzymes of this family catalyze some of the most complex naturally occurring reactions and result in changes of bonding, stereochemistry, and hybridization of the carbon atoms on the substrates[23]. Based on the reaction mechanisms, the terpenoid cyclases can be grouped into two class. Class I terpenoid cyclases, including the prenyltransferases, have a metal cation at the active site to coordinate with the diphosphate group of the substrate. This coordination facilitates the leaving of the pyrophosphate anion to generate a

carbocation on the substrate, which undergoes a series of carbocation rearrangements to form the product[64]. On the other hand, Class II terpenoid cyclases, such as squalene synthase and ent-Copalyl diphosphate synthase, initiate the reaction by protonation of a carbon-carbon double bond or epoxide to generate a carbocation, which undergoes a series of rearrangement to give the product[64] (Fig 6). The catalytic general acid of Class II terpenoid cyclases is the central aspartate in the characteristic sequence motif DXDD.

### 1.3 Natural products resistance

As reviewed in Section 1.1, NPs have a plethora of very potent bioactivity targeting the physiology of natural competitors. However, the toxic natural of small-molecules is often a double-edged sword, as NPs may be potentially harmful to the organisms that produce these compounds, especially if the target pathway is also essential to the producers. As a result, producers must develop strategies to avoid being affected by the NPs, which are produced and accumulated to kill the biological enemies. Organisms have evolved several mechanisms to overcome this problem

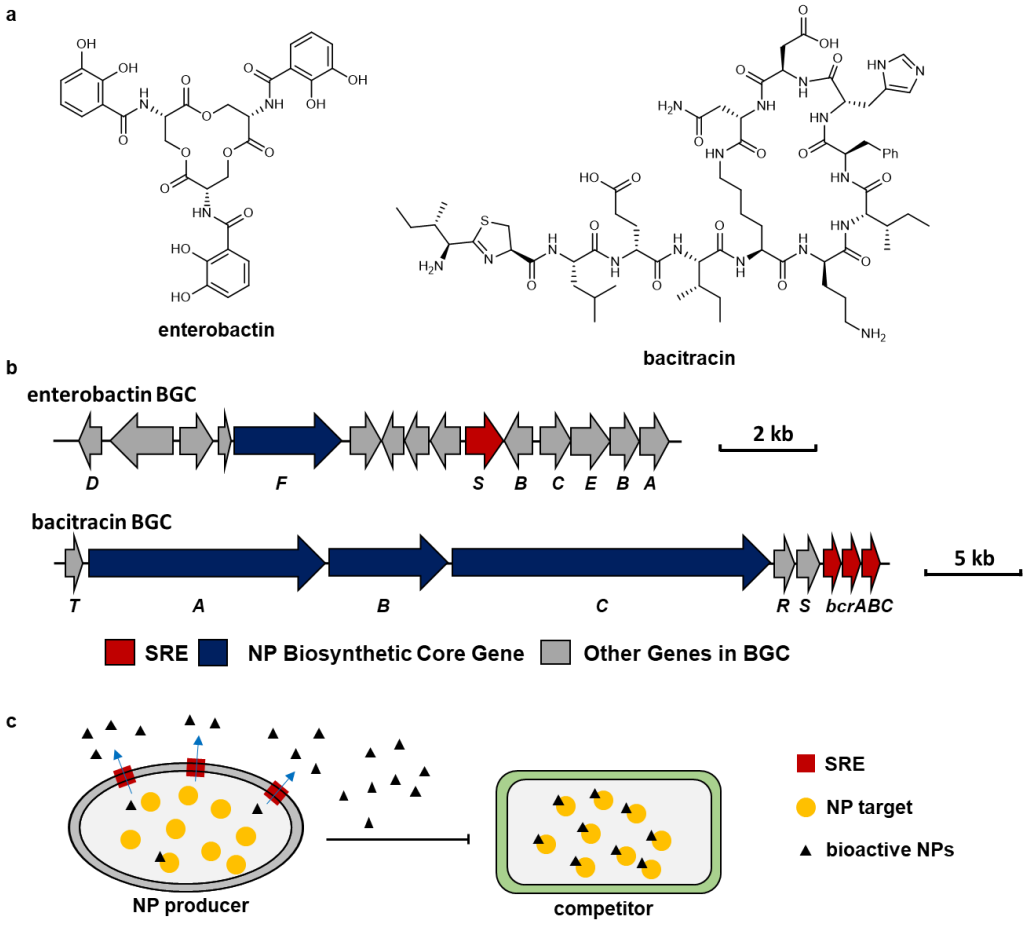



*Figure 7. Self-resistance mechanism using multidrug resistance transporters.*

***a.*** *the structure of enterobactin and bacitracin.* ***b.*** *BGC of enterobactin and bacitracin.* ***c.*** *the scheme of self-resistance using multidrug resistance transporters.*

including employing efflux pumps to pump the toxic compounds out of cells, developing enzymes that are able to perform chemical modification to toxic NPs, biosynthesis of NPs as prodrugs which are further converted to be active after uptake by the enemies, sequestering the active NPs to avoid binding to the biological target, and modification of the biological target to decrease binding affinity of the toxic NPs[65].

### 1.3.1 Efflux pumps

Multidrug resistance transporters such as major facilitator superfamily (MFS) and ATP-Binding-Cassette (ABC) transporters are commonly observed in NP biosynthetic gene clusters (BGCs)[66]. For example, a transporter gene was located in the siderophore enterobactin BGC, which is named EntS. It was demonstrated that EntS is responsible for exporting the molecules to the periplasmic space, which are then further pumped out of the cell outer membrane using other transporters (Fig 7). The bioactivity of enterobactin is to sequester iron from the environment due to its high affinity to a ferric cation[67]. Using this mechanism, the producers are even capable of diverting iron from other competitive organisms, especially in environments in which the concentration of ferric ion is extremely low. Similarly, *bcrABC* genes in the bacitracin BGC encode an ABC transporter, which was also demonstrated to protect the producer by transferring bacitracin outside the cell[68] (Fig 7).

### 1.3.2 Chemical modification

Another widespread strategy to detoxify the bioactive NPs is chemical modification using specific enzymes. The most famous examples of this self-resistance mechanism have been developed by scientists and applied as reporter genes to select for genetic transformants, such as β-lactamase, chloramphenicol hydrolase and hygromycin B phosphotransferase[69-71]. β-lactams such as penicillins were the first class of antibiotics known to humans. They can effectively target

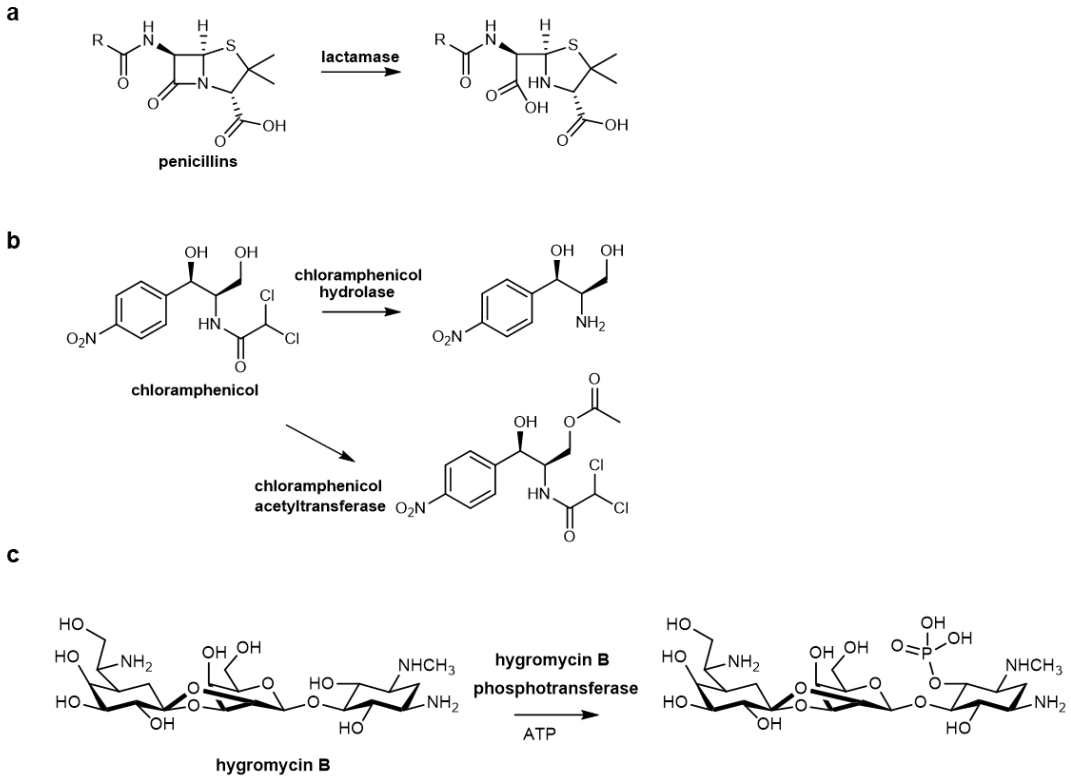



*Figure 8. Self-resistance mechanism through chemical modifications of NPs.*

***a.*** *The β-lactam is hydrolyzed using lactamase.* ***b.*** *chloramphenicol can be deactivated by both chloramphenicol hydrolase and chloramphenicol acetyltransferase.* ***c.*** *hygromycin B is deactivated using hygromycin B phosphotransferase.*

the DD-transpeptidase of the cell wall biosynthesis pathway in gram-negative bacteria through opening of the four membered lactam ring which has high ring strain. To avoid being damaged by β-lactam antibiotics, some bacteria employ a β-lactamase to survive by hydrolyzing the four membered β-lactam ring[69] (Fig 8). Chloramphenicol is another antibiotic used to treat bacterial infections by blocking protein synthesis. It can effectively bind to the 23S rRNA of the 50S ribosomal subunit to prevent the peptidyl transfer in the bacterial ribosome during protein synthesis. To survive when high concentration of chloramphenicol is produced, the producing organism *Streptomyces venezuelae* evolved a hydrolase that can catalyze the removal of the dichloroacetyl moiety to detoxify this compound[70] (Fig 8). Another more common resistance mechanism

employed by other bacteria to survive in the presence of high concentration of chloramphenicol is using a chloramphenicol acetyltransferase (CAT) to covalently attach an acetyl group to detoxify chloramphenicol[72] (Fig 8). Hygromycin B, produced by *Streptomyces hygroscopicus*, belongs to the aminoglycoside antibiotics family; it can effectively kill bacteria, fungi and higher eukaryotic organisms through interfering with protein synthesis. Therefore, the gene conferring hygromycin B resistance is an ideal selectable reporter gene utilized for gene transfer experiments in both prokaryotic and eukaryotic organisms. A hygromycin B phosphotransferase of bacterial origin was discovered to confer resistance to hygromycin B, which is capable of performing a phosphorylation reaction to detoxify hygromycin B[71] (Fig 8).

### 1.3.3 Prodrug of active natural products

Utilization of a prodrug, as an inactive precursor to a bioactive drug, is a widely adopted strategy in modern drug delivery[73]. This approach can be used to improve the delivery accuracy, as well as physiological and physical properties of the drugs. After being delivered to the target, the prodrug is further metabolized to become a bioactive pharmaceutical, which can prevent

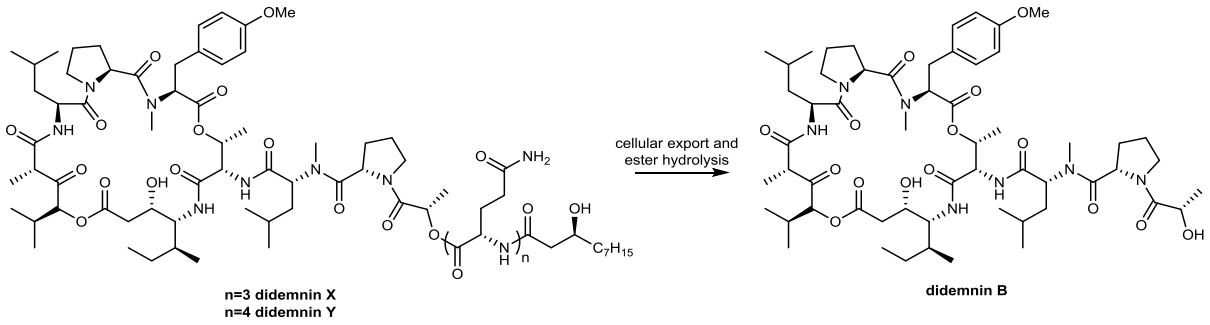



*Figure 9. Example of self-resistance mechanism using a prodrug strategy.*

*Didemnin B was first synthesized as its deactivated form didemnin X and Y, which are further activated by hydrolysis during exportation out of the producer.*

undesirable side effects. To avoid toxicity of the NPs, this prodrug strategy is also prevalently adopted by NP producers as a self-resistance mechanism: the NPs will not become active until uptake by competing organisms[74]. For instance, the anticancer agent didemnin B isolated from a marine tunicate was produced by its symbiotic bacteria *Tistrella mobilis*[75]. It was not directly assembled by the NRPS and stored in the cell as didemnin B. Instead, other structurally similar compounds didemnin X and Y are synthesized as a prodrug first, which are modified with one or two glutamine and a β-hydroxyl fatty acid by the NRPS. The cytotoxicity of didemnin X and Y are very low compared to the mature compound didemnin B. The prodrug didemnin X and Y was demonstrated to be converted to didemnin B after cellular export and ester hydrolysis, revealing the potent didemnin B (Fig 9).

### 1.3.4 Sequestering the active natural products

Sequestering a bioactive compound from accessing its cellular target is another common self-resistance mechanism. This is usually accomplished by noncovalently but tightly binding to a protein, which stabilizes the warhead of the bioactive NPs. One of the most notable examples is

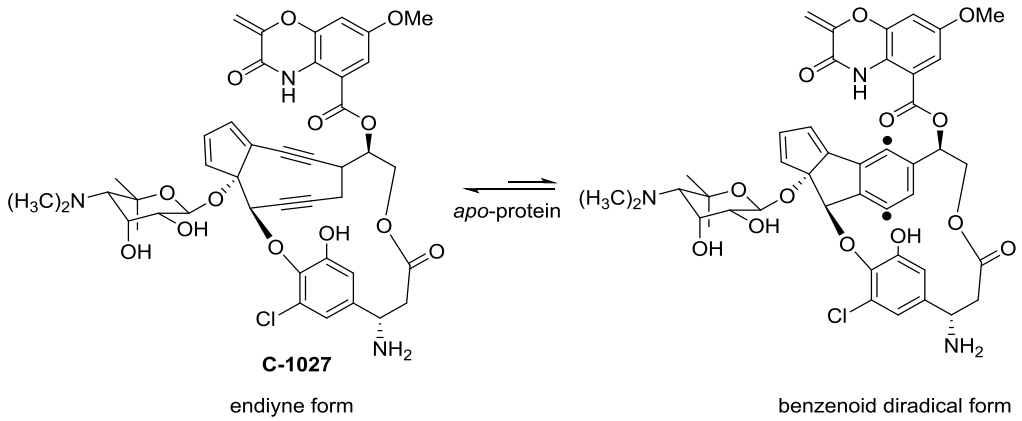



*Figure 10. Example of self-resistance mechanism using a sequestering strategy.*

*The nine-membered ring enediyne C-1027 is stabilized using apo-protein to prevent the producer.*

the self-resistance mechanism of NPs from nine-membered ring enediyne family[76]. For example, the enediyne moiety with two acetylenic groups conjugated to a double bond of C-1027 is in equilibrium with its benzenoid diradical form at physiological conditions[77]. However, the diradical is very unstable, and readily abstracts hydrogen from deoxyribose of DNA when binding to the minor grooves of DNA. The newly formed carbon radicals of deoxyribose further react with $O_2$ and lead to DNA double strand breaks, which will eventually cause cell death. Scientists were curious about how the producing strain is able to survive in the presence of this highly reactive NPs since its mode of action was figured out. It was then demonstrated that a constitutively expressed *apo*-protein is required for production of C-1027, which is capable of noncovalently binding to C-1027 to stabilize its enediyne form[76]. As a result, the *apo*-protein prevents the transformation of C-1027 into its active form to protect the producing strain (Fig 10).

### 1.3.5 Modification of targets

Alternatively, in order to acquire self-resistance of NPs, the targets within the producing organisms could be mutated to prevent NP binding. The self-resistance enzymes (SREs) are capable of performing the same function as the NP targeted housekeeping enzyme, however, it is sufficiently mutated to be not affected by the NP. Because the self-resistance gene is required for survival at the time of NP production, the gene encoding its SRE is frequently co-localized in the NP BGC. A well-known example is the lovastatin self-resistance mechanism: a 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR), which is also the biomolecular target, was found to be encoded by *ORF8* in the BGC of lovastation[78] (Fig 11).

Recently, in the post genomic era, various self-resistance mechanisms have caught the attention of natural product researchers. For instance, the BGC of the fatty acid synthase inhibitor thiolactomycin from *Salinispora pacifica* was successfully located within the genome of the producing organism using a second copy of fatty acid synthase, which was later proven to be a SRE[79] (Fig 11). The antituberculosis agent griselimycin was found to target DNA polymerase,

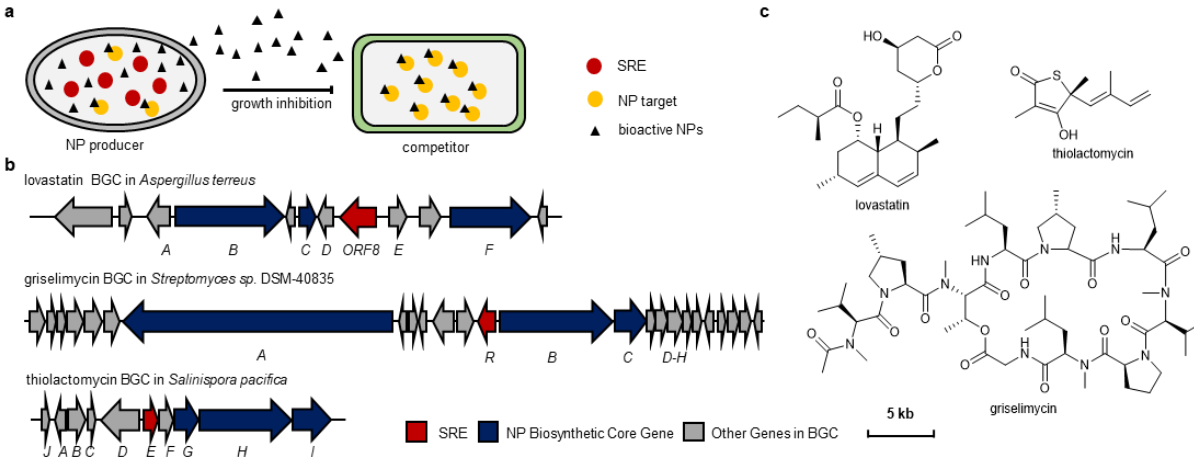



*Figure 11. Example of self-resistance mechanism via target alternation.*

***a,** the function of self-resistance enzymes to confer resistance of the host in the presence of bioactive NPs. **b,** co-localization of the resistance genes (red) and the NP biosynthetic genes (blue). **c,** structure of NPs using target alternation strategy as self-resistance mechanisms.*

hinted by the presence of a self-resistance gene that encodes DNA polymerase in its BGC[80] (Fig 11). Deep insights of the self-resistance mechanism not only provides the potential for application (such as developing self-resistance genes as markers for selection of recombinant DNA transformation) but can also provide insight in treating multi-drug resistance pathogens[65].

## 1.4 Approaches to discover novel natural products

### 1.4.1 Activity and structure guided natural products discovery

The application of NPs precedes recorded human history by thousands years, and our earliest ancestors were able to use medicinal plants to treat wounds or diseases[81]. Traditional Chinese medicine is one of the well-known example of the extensive use of natural products as medicines since ancient times for more than 2000~3000 years[82]. However, it was not until the 19$^{th}$ century, during which the development of organic chemistry, analytical chemistry, biology, and biochemistry paved the way to the golden era of NPs. Using modern technologies, scientists are able to isolate NPs, and determine their chemical structures, biological activities and medicinal properties[83]. In 1805, pure morphine was isolated from opium and became the first pure NP medicine commercialized by Merck in 1826[81]. Subsequently, many famous NPs drugs were isolated from medicinal plant extracts such as quinine from *Cinchona ledgeriana*, caffeine from *Coffea arabica*, and nicotine from *Nicotiana tobacum*[81,84,85].

The discovery of the first microbe-derived antibiotic penicillin is considered to be a significant milestone in modern NP research[86]. In 1928, Alexander Fleming observed a zone of inhibition near *Penillium* colonies on a petri dish growing *Staphylococcus.* He found this mold can secrete something that was capable of killing many harmful bacteria species. Ten years later, scientists confirmed that penicillin was the bioactive component, and started clinical trials. This bioactivity guided approach to discover bioactive NPs led scientists to effectively find many new and valuable NPs, which include some of the most famous pesticides or pesticide leads such as spinosyns, avermectins, glufosinate and antimycins[9] (Fig 1).

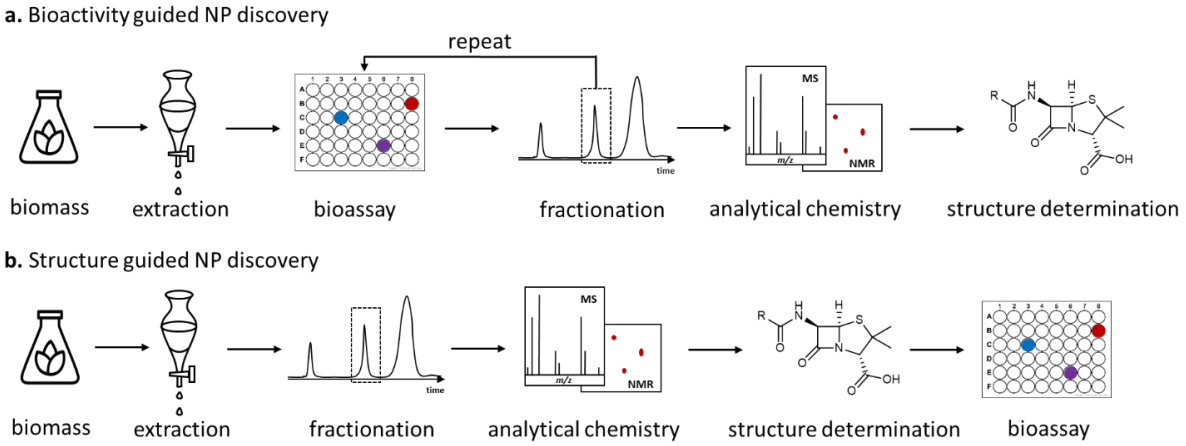



*Figure 12. Traditional strategies of natural product discovery.*

***a.*** *the scheme of bioactivity guided NP discovery, in which, the NPs are screened using bioactivity assay.* ***b.*** *the scheme of structure guided NP discovery, in which, the structure of all the pure NPs was determined using model analytical techniques.*

In a brief summary of NP discovery history, bioactivity-guided compound purification has been the used as a widespread strategy for the development of new drugs. The process can be simplified as follows: first, a biomass with excellent medicinal properties is subjected to fractionation based on the physical properties; then, the bioactivity of each fraction is determined and those without the desired bioactivity are discarded; and last, the first two steps are performed iteratively until the pure bioactive compound is obtained (Fig 12)[83]. Since an increasing number of bioactive NPs have developed into new drugs, scientists and pharmaceutical companies have focused on NP research[87].

Later on, with the development of purification technologies, more and more structurally diversified NPs were isolated prior to any knowledge of their bioactivity[88]. These compounds may then be subjected to high throughput screening of bioactivities and then developed into new therapeutics or agricultural agents[3,9] (Fig 12). The explanation of why so many NPs have very promising medicinal properties is: NPs have co-evolved together with large biomolecules

including proteins and DNA, which results in high proximity in influencing the physiology of relating organisms by interaction with specific biomolecules. These strategies can be summarized as the structure guided NP discovery approach, which has proven fruitful in leading NP research to its golden age[5].

### 1.4.2 Genome guided natural products discovery

Due to increasing drug resistance to existing molecules and a dwindling pipeline of new drug leads, the need to discover novel NPs and generate chemical diversity is increasingly critical[89,90]. Besides plants, microbes such as bacteria and filamentous fungi are considered to be rich sources of NPs. Therefore, as the amount of microbial genome information becomes increasingly abundant through next generation sequencing, genome-guided mining of new NPs for therapeutic application is emerging as a potentially powerful approach to discover new NPs[91].

Genomics has revolutionized every aspect of biology, and has also revitalized NP research. Rapid development of DNA sequencing technologies has ushered in a new golden age of NP research, with an ever-increasing number of whole genome sequencing data available[91-93]. Genes encoding NP biosynthetic enzymes are frequently colocalized to facilitate expression and regulation in microorganisms. These NP biosynthetic genes are physically colocalized and typically referred to as BGCs[94]. Bioinformatic analysis has revealed that many microbes in fact encode far more BGCs than the NPs that are produced under growth conditions used for pesticide screening[95]. A majority of BGCs remain silent due to the complex regulation of BGCs in nature and our inability to reproduce the natural environmental cues that are needed to turn them on. As a result, more than 90% of BGCs have therefore remained as genomic "dark matter" and encode secondary metabolites that have been elusive as opportunities for NPs discovery[95]. It is tantalizing

to imagine the quantity and impact of bioactive compounds that could be accessible to humans if all of the microbial biosynthetic capabilities were realized[96].

The first biosynthetic gene was characterized to be responsible of actinorhodin biosynthesis in 1980 with the establishment of *Streptomyces* genetics[97]. After this, researchers started to understand the biosynthetic logic and genetic basis for the production of these compounds. In the post-genomic era, detecting and analyzing secondary metabolite BGCs within a gigantic collection of microbial genomic databases are well established using existing bioinformatics tools[90]. To date, thousands of BGCs have been characterized and deposited in the database known as "minimum information for biosynthetic gene clusters" (MIBiG). Although the structures and bioactivities of these secondary metabolites vary dramatically, all of them are assembled from a common reservoir

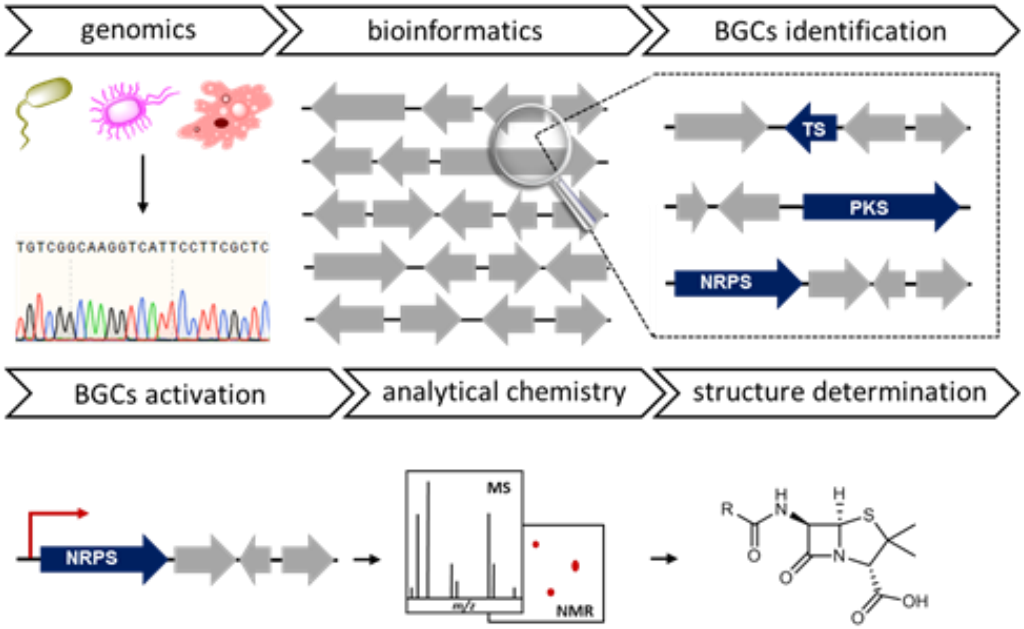



*Figure 13. Workflow of genome guided NP discovery.*

*This approach starts from the genomic sequence of organisms of interests, which are used to identify BGCs of putative NPs. The BGCs of interests are then activated using synthetic biology tools, which is followed by metabolic analysis, compound purification and structure elucidation.*

of primary metabolite building blocks using conserved biosynthetic machineries[98]. Such conserved machineries that have been well characterized to date include polyketides assembled by polyketide synthetases (PKSs), non-ribosomally synthesized peptides assembled by non-ribosomal peptide synthetases (NRPSs), ribosomally synthesized and post-translationally modified peptides (RiPPs) and terpenoids assembled by terpene synthase (TS). Besides shedding light on the sophisticated biosynthetic mechanisms of NPs, these conserved biosynthetic enzymes also provide us a window to discover new NPs[21-23].

The knowledge of NP biosynthesis enables a genome guided strategy to facilitate NP discovery, which is to predict and isolate natural products based on genomic sequences without knowing the structure and bioactivity beforehand[91,95]. Genome mining of highly conserved NP biosynthetic core enzymes such as PKSs, NRPSs and TSs that have not been characterized to be related to any known NPs is a typical approach to find new BGCs (Fig 13). Meanwhile, focusing on the tailoring enzymes that are able to modify the precursors also provides scientists with another way to identify new BGCs[96].

A more sophisticated approach to carry out genome mining is based on the phylogeny tree of biosynthetic genes[96]. The idea is to compare a homologous biosynthetic gene that is diversified within different organisms. Like the evolution of organisms, biosynthetic genes also gradually diversified in function as genomic sequences altered among different hosts. Therefore, a phylogenetic study of how the function is diversified as the DNA sequence alters provides hints for the discovery of new biosynthetic enzymes, which may further relate to the biosynthesis of novel NPs.

Based on Darwin's theory of evolution, the NPs are also subjected to gradual natural selection and evolution. Natural selection preserves and accumulates minor advantageous genetic mutations: "Natural selection acts only by taking advantage of slight successive variations; she can never take a great and sudden leap, but must advance by short and sure, though slow steps." Thus it is not difficult to imagine the huge number of BGCs that exist with slight variations. One random BGC may not produce NPs with desired bioactivity and medicinal properties. As a result, characterizing all of the orphan BGCs one by one is labor intensive and time consuming; thus, prioritizing these orphan BGCs according to a proposed mode of action provides a strategy to address this challenge[96].

## 2. Results and discussions

### 2.1 Genome mining of heptelidic acid biosynthetic gene cluster

#### 2.1.1 Introduction

Bioinformatic analysis has revealed that many microbes in fact encode far more BGCs than the NPs that are produced under growth conditions used for drug screening[99]. More than 90% of BGCs have therefore remained as genomic "dark matter" and encode NPs that have been elusive as opportunities for drug discovery. It is tantalizing to imagine the quantity and impact of therapeutic compounds that could be accessible to humans if all of the microbial biosynthetic capabilities were realized. Consequently, developing new strategies and tools to explore this

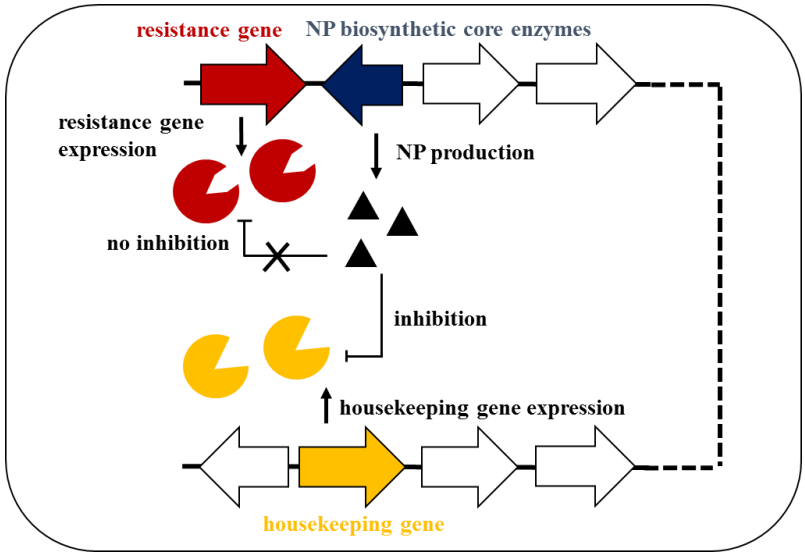



*Figure 14. The relationship between natural product biosynthetic core gene, resistance gene and housekeeping gene in the natural product producing host.*

*The housekeeping gene shown in yellow encodes a critical metabolic enzyme (yellow sphere) for the survival of the host, and this enzyme can bind to the produced natural product tightly to abolish its function. The resistance gene shown red encodes a self-resistance enzyme, which has the same function of housekeeping gene, however, the natural product is not able to bind to this enzyme due to some mutations at the active site. Thus the cell is still able to survival when producing the toxic natural product.*

hidden biosynthetic potential in nature, and to reinvigorate NP drug discovery are highly significant[100-102].

In my research project, we developed an innovative resistance gene directed NP discovery approach to effectively and efficiently translate microbial genome information into bioactive compounds[103]. The rationale of this approach is that a host organism producing NPs must have a method of self-protection; this self-protection is frequently achieved through the coexpression of an alternative homolog of the target enzyme that is insensitive to the NP (Fig 14). Therefore, a resistance gene directed approach has proven to be an effective strategy to bridge the gap between orphan BGCs and bioactivities in the search for new NPs[65].

The co-localization of a self-resistance enzyme in the BGCs of NPs offers a unique and exciting way to bridge bioactivity guided NP discovery and genome guided NP discovery, and to predict biological activity of the NPs produced by a sequenced organisms[103]. For example, a second copy of HMGR, the target of lovastatin, is also encoded in the lovastatin BGC as *ORF8* in

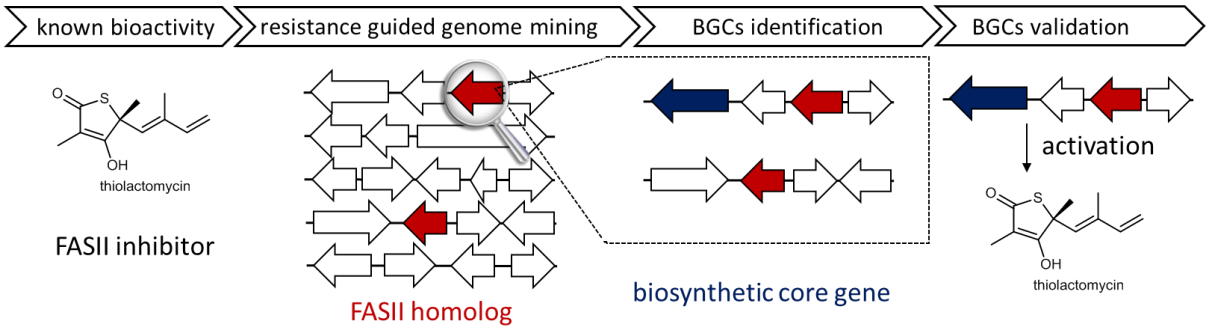



*Figure 15. Workflow of resistance gene guided genome mining of the biosynthetic gene cluster of thiolactomycin.*

*The BGC of a fatty acid synthase (FASII) inhibitor thiolactomycin is successfully located in the genome of producing host Salinispora tropica using FASII. This FASII is a second copy in the genome, which is proved to be not sensitive to thiolactomycin. Heterologous expression of this cluster further also proved that it is responsible of producing thiolactomycin.*

*Aspergillus terreus*[78] (Fig 11). From a reverse-engineering perspective, if the lovastatin BGC was silent, as 90% of the BGCs in *Aspergillus terreus* are, then by locating the second copy of HMGR as an SRE, we would have been able to predict HMGR as the target of the NP produced by the BGC.

This self-resistance mechanism has also inspired scientists in NP research recently. Tang and coworkers were able to successfully locate the BGC of thiotetronic acid in *Salinispora tropica* in 2015 (Fig 15)[79]. Although thiotetronic acid was isolated about 30 years ago and is known to be a fatty acid synthase inhibitor when the project was established, scientists had no clue about its biosynthesis[104,105]. They proposed a fatty acid synthase that is insensitive to the NP inhibitor could be encoded in the BGC to confer resistance when thiolactomycin is produced by *Salinispora tropica.* Thus, they first searched through the genome to look for fatty acid synthase (FASII) as putative antibiotic resistance gene. Then they found one FASII homolog was close to a hybrid PKS-NRPS, which had not been characterized in the biosynthesis of any known NPs. To

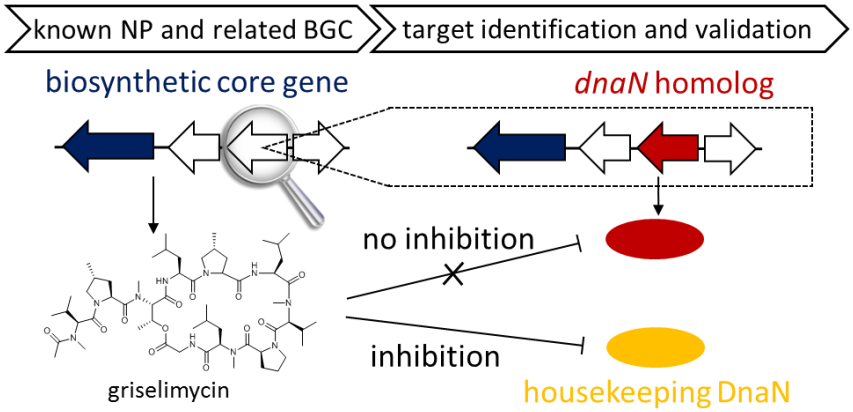



*Figure 16. Workflow of resistance-gene guided molecular target discovery of griselimycin. Molecular target of griselimycin is determined by the hint of a dnaN gene within its BGC, which was further proved to be the self-resistance gene.*

demonstrate this BGC is actually able to produce a FASII inhibitor, they expressed the gene cluster in a heterologous host *Streptomyces coelicolor* M1152 and identified a new product as thiolactomycin. This thiotetronic acid was isolated from other organisms and demonstrated to be a FASII inhibitor[105]. This research indicates it is practical to find a BGC related to NP of desired bioactivity using resistance gene as a guide.

Besides locating the BGCs of a NP of known bioactivity, the resistance gene in the BGCs may also shed lights to the bioactivity of its related NP. Griselimycin is a cyclic peptide isolated from *Strepotmyces* strains in 1960s[106]. Although it has excellent broad spectrum antibacterial activity including *Mycobacterium tuberculosis* (TB), further developing of this NP to human therapeutics as anti-TB drugs was impeded because of poor pharmacokinetics properties[80]. The analogs of griselimycin have been made and tested, however, efforts were abandoned when rifampin became available to treat TB[107]. Understanding the mode of action of bioactive NPs is critical to drug development and improvement. Recently, Kling *et al.* discovered the molecular

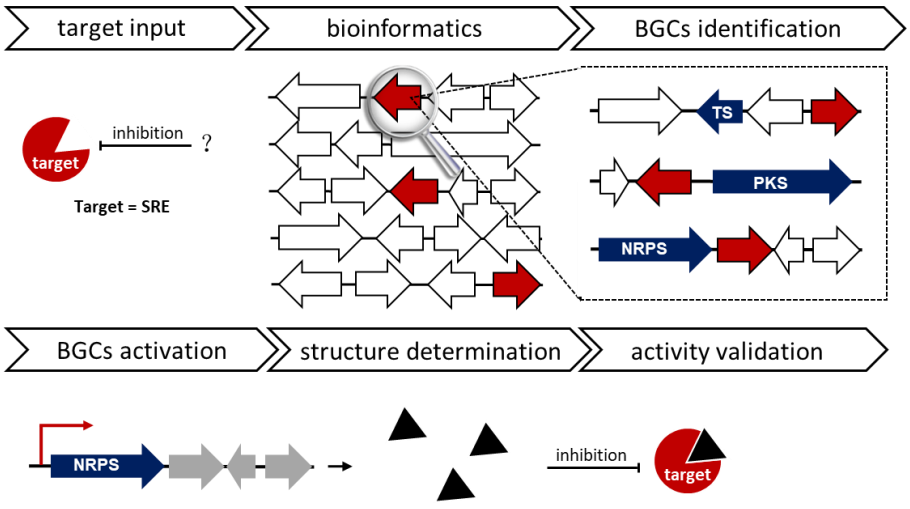



*Figure 17. Workflow of resistance-gene guided genome mining strategy to discover NPs of desired bioactivities.*

target of antibiotic griselimycins is DNA polymerase sliding clamp DnaN, which was hinted by a DnaN homolog encoded in the BGC that was demonstrated to be the SRE[80] (Fig 16). This discovery unveiled the novel mode of action of griselimycin, and its promising potential to be developed into new anti-TB drugs. This research outcome again indicated that the BGCs of NPs contain much more information, which is not just limited to how a NP is made, but also how that NP functions.

All the above examples illustrate the feasibility of using the resistance gene as a window to predict the function of NPs "a priori," without any prior knowledge of structure. The only information we need is the genomic sequence and the bioactivity to query. In other words, our goal is to find gene clusters that contain both biosynthetic core genes and a resistance gene, and then we activate the cluster to check whether a bioactive NP can be produced or not (Fig 17). Here we propose a resistance gene guided genome mining approach to discover NPs of desired bioactivities. Although it is still alignment based approach to predict BGCs using known biosynthetic core enzymes, which means it is not able to discover new biosynthetic enzyme, it is capable of helping us to prioritize the increasing number of cryptic and silent biosynthetic gene clusters to direct us to tap into bioactive NPs with desired mode of action.

### 2.1.2 Results and discussions

To test our hypothesis, we first used a NP for which the bioactivity was known but the BGC remained unknown. We wanted to check if we could identify the right BGC when using the known bioactivity as an input for resistance gene guided genome mining. After that, we will move onto discovery of NPs of a desired function using our proposed approach.

Heptelidic acid (**1**), also known as avocettin or koningic acid, was initially isolated from *Trichoderma virens* in the 1980s[108,109] (Fig 18). It was reported to have multiple bioactivities such as antibacterial, antimalarial and apoptosis regulator activity[110-112]. Heptelidic acid was further identified to be an inhibitor of a potent anticancer therapeutic target glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[113]. Similar to its structural analog pentalenolactone, the electrophilic epoxide moiety covalently blocks the active-site cysteine of GAPDH, which results in irreversible inactivation[114,115]. Isotopic acetate incorporation experiments indicated that heptelidic acid is derived from a cadalene type sequiterpene precursor, followed by an enzymatic Baeyer-Villiger type oxygen insertion occurring afterwards to form the lactone[116] (Fig 18). However, the gene

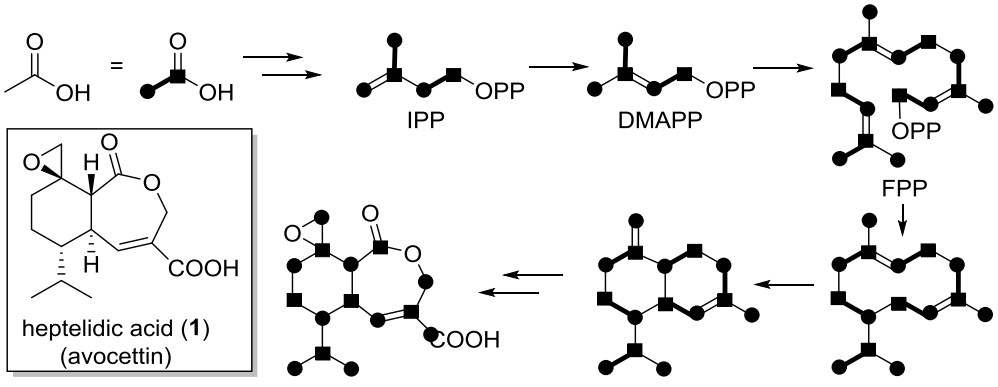



*Figure 18. Biosynthesis of heptelidic acid study using isotope labeling.*

cluster and biosynthetic pathway of heptelidic acid remained unknown at the onset of this project in 2014, despite a recent report of the BGC[117].

GAPDH, the biomolecular target of heptelidic acid, is in an essential pathway in the breakdown of glucose to obtain energy[118]. It catalyzes the sixth step of glycolysis to convert glyceraldehyde 3-phosphate to D-glycerate 1,3-bisphosphate using the cofactor nicotinamide adenine dinucleotide ($NAD^+$) as an oxidizing agent. The reaction is accomplished through covalent catalysis followed by general base catalysis. First, the free thiol of the cysteine residue of GAPDH attacks the carbonyl of glyceraldehyde 3-phosphate to form a hemithioacetal intermediate that covalently binds to GAPDH. Then the oxidation is carried out by $NAD^+$ in the active site, which abstracts a hydride to form the thioester intermediate that is covalently bound to GAPDH. Finally, a phosphate anion attacks the carbonyl of the thioester to form the final phosphorylated product D-glycerate 1,3-bisphosphate (Fig 19). Heptelidic acid is able to covalently bind to the cysteine in active sites via epoxide ring opening reaction, blocking the function of GAPDH[119]. Due to the increased aerobic glycolysis metabolism is uniquely observed in cancer cells, instead of normal human cells, heptelidic acid is considered a potential anticancer agent[120].

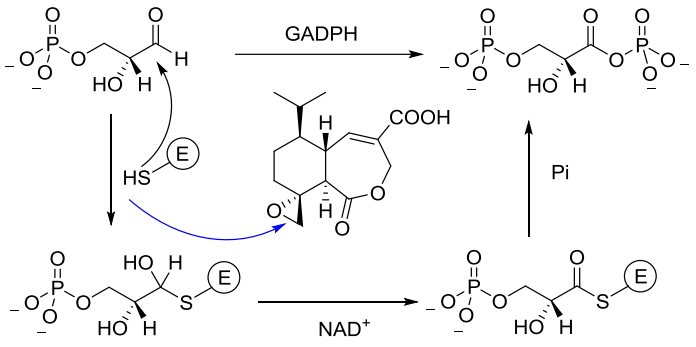



*Figure 19. Reaction and inhibition mechanism of GAPDH.*

To validate our proposal, we utilized our proposed resistance-gene guided genome mining approach to identify the BGC of heptelidic acid in the producing strain *T. virens*. In fungi, there is usually only one copy of the *gapdh* gene in the genome. In comparison, as we expected, bioinformatic analysis of the *T. virens* genomic sequence reveals there are two copies of *gapdh* in the genome (accession number: XP_013949968.1 and XP_013958680.1), which have 73% identity to one another. Bioinformatic analysis showed that one of the GAPDH homologs XP_013949968.1 has less identity (~70%) to the housekeeping fungal GAPDH across species, while the other homolog XP_013958680.1 has higher identity (~80%). So from this result, we predicted XP_013949968.1 is the resistance gene, and XP_013958680.1 is the housekeeping gene. And as we expected, XP_013949968.1 is co-localized with the natural product biosynthetic core gene, which is responsible of terpenoid biosynthesis. This putative terpenoid BGC, with a length of 18 kb, encodes a resistance gene (*hepG*), terpene synthase (*hepA*), four P450s (*hepC*, *hepD*, *hepE* and *hepH*), antibiotic biosynthesis monooxygenase (*hepB*) and an MFS transporter (*hepF*)

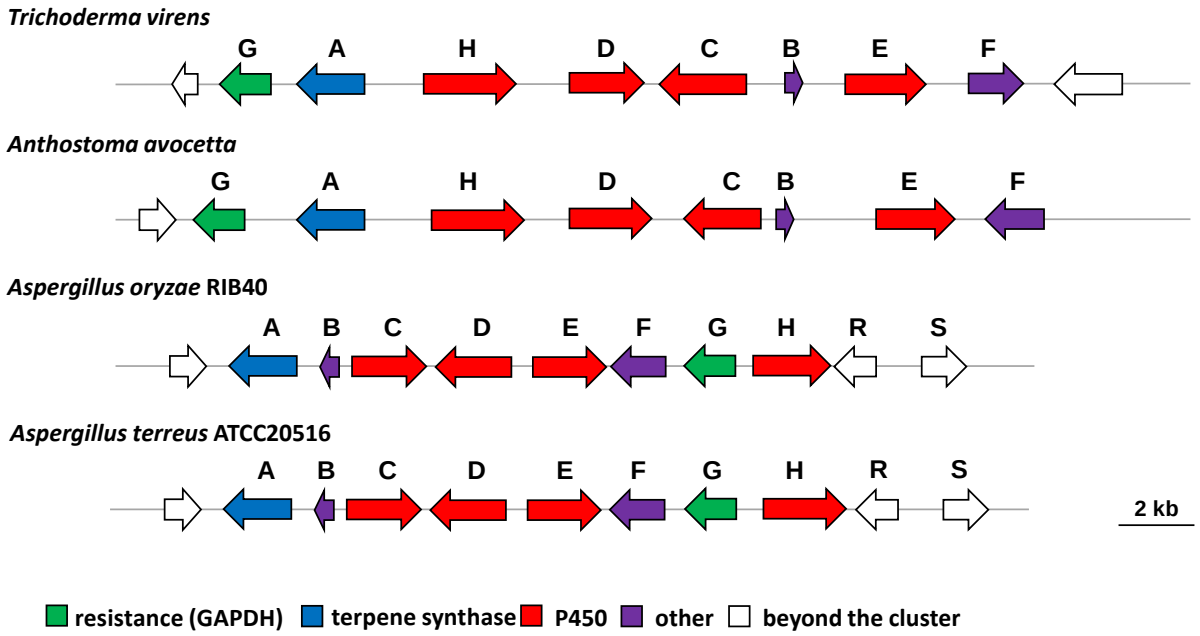



*Figure 20. Conserved BGC of heptelidic acid among several fungal species.*

(Table S1). These genes are predicted to be in charge of heptelidic acid biosynthesis, self-resistance and transport. In addition, the BGC, including the resistance gene, is also found to be conserved within several other fungal species including the reported producing strain *Anthostoma avocetta* and *Aspergillus terreus* ATCC20516 (Fig 20).

Further bioinformatics analysis showed that in addition to *T. virens*, the *hepG* presents the other two producers *A. avocetta* and *A. terreus* is also a second copy of the housekeeping *gapdh* gene which further indicated that the resistance gene is essential to help the host to survive when

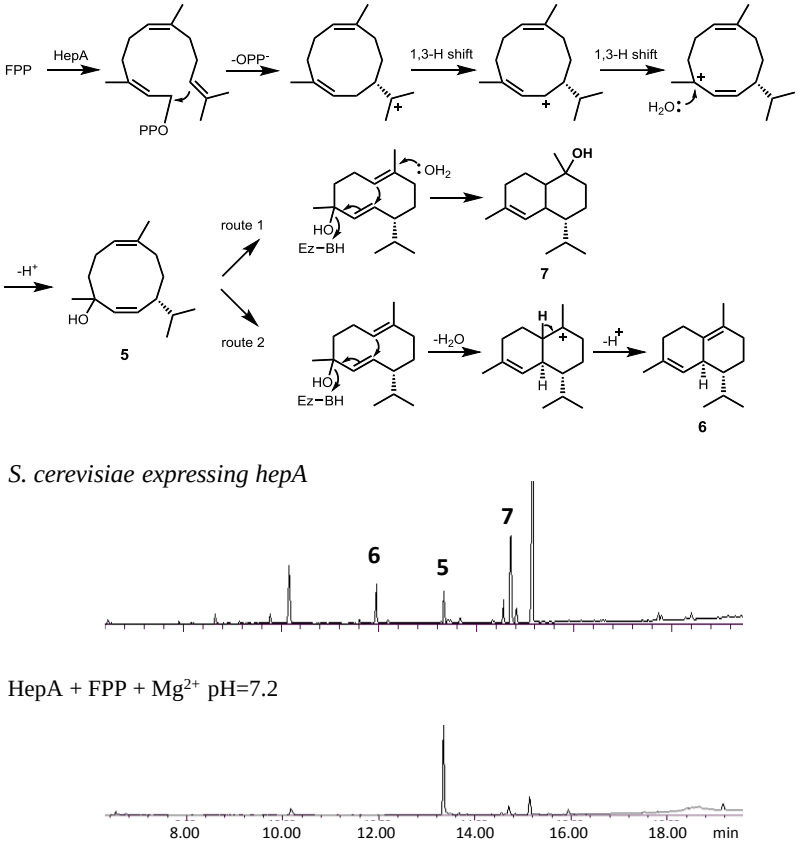



*Figure 21. HepA is a sesquiterpene cyclase.*

*The scheme proposed reaction to generate the sesquiterpene product **5** using HepA, and spontaneous reaction to form **7** and **8** (up). S. cerevisiae expressing HepA is able to produce compound **5**, **6** and **7** (middle). HepA is only able to catalyze the formation of **5** in the in vitro assay.*

heptelidic acid is produced. To validate our hypothesis, we first saught to prove this cluster is indeed responsible for heptelidic acid production. The most ideal approach is to knockout the *hepA* gene, however, gene knockout in *T. virens* failed due to a high rate of non-homologous end joining (NHEJ)[121]. We therefore tried to do heterologous expression of the whole cluster in *Aspergillus nidulans*, however, no product was detected. Thus, we decided to study the function of each gene one by one and try to piece together the obtained information to prove this gene cluster is in charge of heptelidic acid biosynthesis.

First, we cloned the cDNA from *T. virens* and expressed the terpene synthase HepA in *Saccharomyces cerevisiae*. When HepA was expressed for 3 days, we identified a series of sesquiterpenes produced including compounds **5**, **6,** and **7** (Fig 21). However, if HepA was purified from the expressing *S. cerevisiae* strain, and subjected to in vitro biochemical assay using FPP,

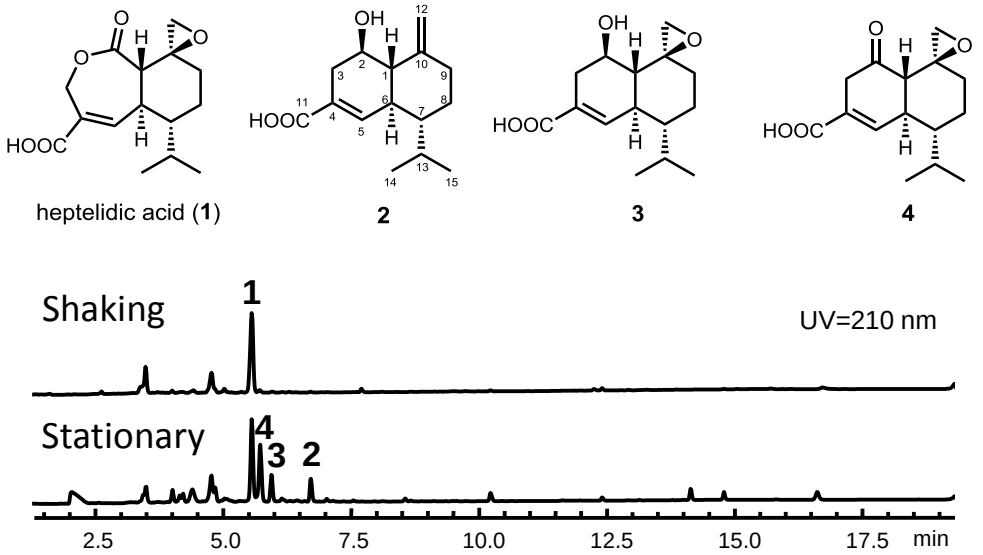



*Figure 22. Fermentation of heptelidic acid biosynthetic intermediates.*

*Only heptelidic acid can be obtained from fermentation of T. virens at shaking condition. However, compound **2-4**, which is the biosynthetic intermediates of heptelidic acid can be produced when growing T. virens at stationary condition, possibly due to the oxidation steps were not completed due to the lack of oxygen in the stationary growing conditions.*

only compound **5** can be detected as a product. Therefore, compound **5** is the product of HepA, compound **6** and compound **7** are metabolites of the yeast endogenous enzymes using compound **5** as substrates, which are shown in Figure 21. According to the reported isotope labeling results, **5** is a reasonable terpene precursor to heptelidic acid, which indicated this cluster is in charge of heptelidic acid biosynthesis[116] (Fig 18).

Next we focused on characterization of the cytochrome P450s including HepC, HepD, HepE, and HepH in the gene cluster. Fermentation of the wild type *T. virens* strain at different conditions was carried out to optimize the production of heptelidic acid. Compared to incubating the strain in the shaker, there were three more structurally related new metabolites generated when the strain was incubated at stationary condition. These new compounds **2**, **3** and **4** were isolated, and their structure were determined by NMR (Fig 22 and S3). Comparing these three new compounds to heptelidic acid, all of them are structurally related to each other and possibly biosynthetic intermediates of heptelidic acid. Compound **2** can be converted to **3** by epoxidation of the C10-C12 double bond; then dehydrogenation of hydroxyl group of compound **3** will give compound **4**. At last, a Baeyer-Villiger oxidation converts compound **4** to heptelidic acid. Although Baeyer-Villiger oxidation is known to be catalyzed by FAD dependent Baeyer-Villiger monooxygenases (BVMO)s, there is no flavoenzyme in the proposed biosynthetic gene cluster of heptelidic acid[122]. On the other hand, cytochrome P450s have also been reported to catalyze the Baeyer-Villiger oxidation in the biosynthetic pathway of brassinolide and brassinosteroid in *Arabidopsis*, which indicated that cytochrome P450s also have the potential to function as a BVMO. Therefore, we proposed one of the cytochrome P450s instead, is able to perform as a BVMO, because cytochrome P450s can activate the molecular oxygen, forming the peroxide anion Fe-$OO^-$, which will attack ketone to form Criegee intermediate. This unstable intermediate readily undergoes rearrangement

via expulsion of a 4α-hydroxyflavin anion Fe-O$^{-}$ and migration of a carbon-carbon bond to form the ester.

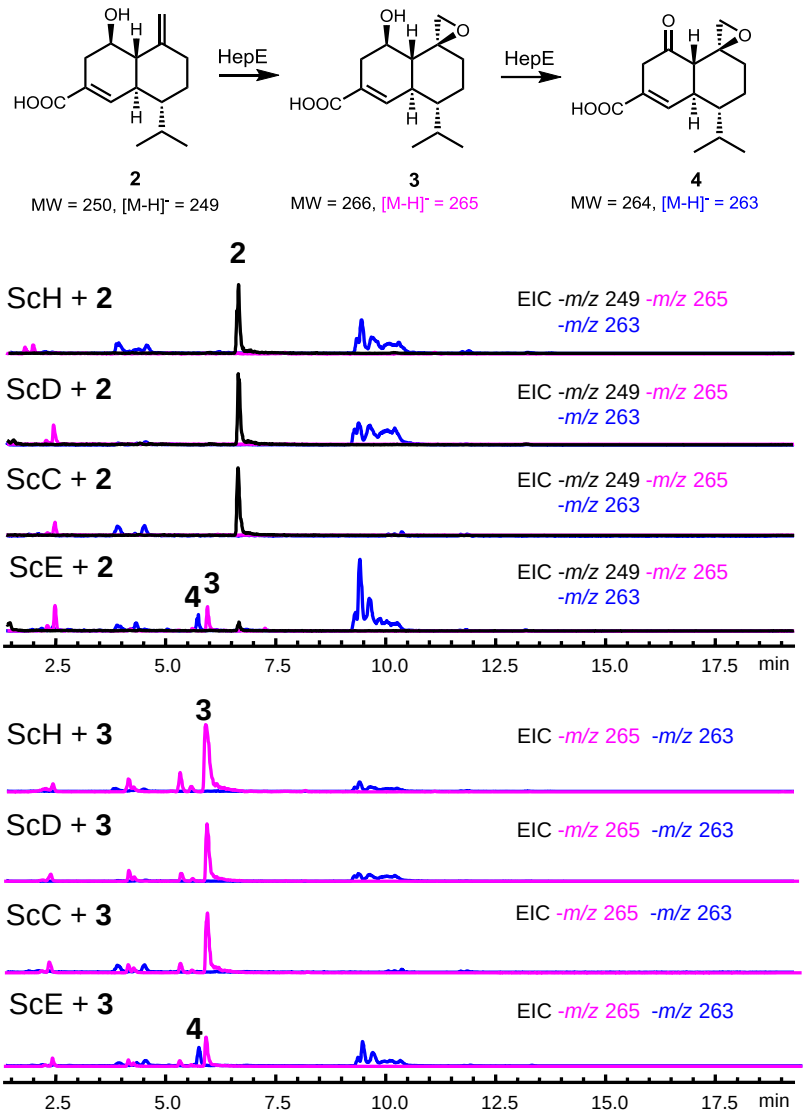



*Figure 23. Saccharomyces cerevisiae conversion of **2** to **3** and **3** to **4** using hepE.*

*Top: compound **2** was fed to S. cerevisiae expressing the cytochrome P450s within the BGC, only S. cerevisiae expressing HepE is able to convert **2** to **3**, and further conversion of **3** to **4** was also observed. Bottom: compound **3** was fed to S. cerevisiae expressing the cytochrome P450s within the BGC. Consistent with the results observed on top, only S. cerevisiae expressing HepE is able to convert compound **3** to compound **4**. These results indicated HepE is able to convert compound **2** to compound **4** via compound **3**.*

To validate our prediction, the cDNA of the cytochrome P450 oxidoreductase and all cytochrome P450s including *hepC*, *hepD*, *hepE*, *hepH*, and a cytochrome P450 reductase were obtained from the producing strain *T. virens*. These cytochrome P450s, together with their electron transfer partner, were heterologously expressed in *S. cerevisiae* under the *adh*2 promoter individually by strain ScC, ScD, ScE and ScH respectively. To investigate which cytochrome P450 could convert the biosynthetic intermediates, compound **2** was fed to cytochrome P450 expressing strains. After incubating these strains together with compound **2** for 12 h to carry out the bioconversion, each culture was extracted with acetone and subjected to HPLC-MS analysis (Fig 23). The results showed that only strain ScE is able to compound **2** into compound **3**, meanwhile ScE can further convert the biosynthetic intermediate compound **3** to compound **4**. In consistence, when compound **3** was fed to all cytochrome P450 expressing strains, only ScE can convert

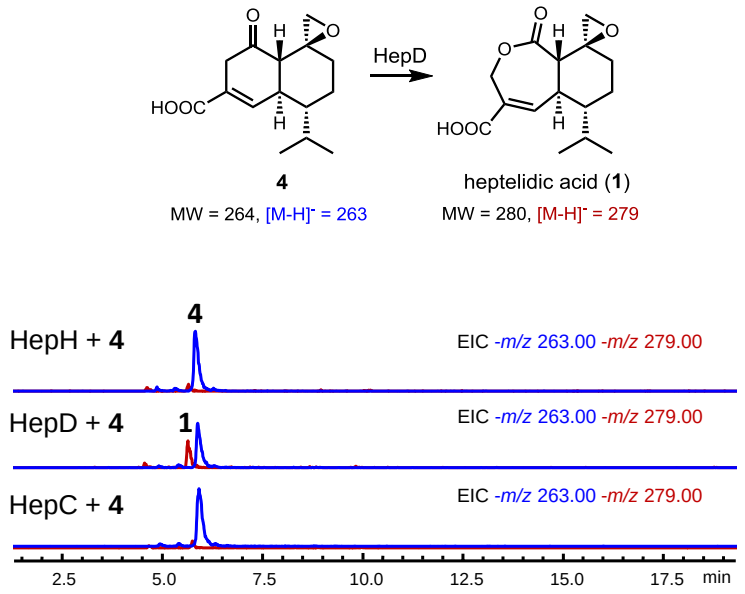



*Figure 24. Conversion of **4** to heptelidic acid using HepD.*

*Microsomes containing HepH, HepD, and HepC respectively together with their reducing partner cytochrome P450 reductase were incubated with **4** in the presence of NADPH, NADH, $FADH_2$ and $FMNH_2$. Only HepD is able to catalyze the reaction.*

compound **3** to compound **4**. Together, these indicated the function of HepE is to catalyze epoxidation of compound **2** to compound **3**, and dehydrogenation of compound **3** to compound **4**. However there is no final product heptelidic acid to be detected within the metabolites of ScE, which means the Baeyer-Villiger oxidation is catalyzed by another enzyme (Fig 23).

A similar feeding experiment was carried out using compound **4** as the substrate, however, no strains showed the ability to accomplish the transformation of compound **4** into heptelidic acid. Heptelidic acid is an irreversible inhibitor of GAPDH, which will covalently bind to the active site of GAPDH; therefore we proposed that even if there is a small amount of heptelidic acid produced by bioconversion, heptelidic acid may be still elusive from LC-MS detection when it covalently binds to GAPDH[115]. In addition, heptelidic acid may also affect the growth and protein expression of the yeast cells which may further decrease the conversion efficiency. Thus, we extracted the microsomes of the strains that express cytochrome P450s including HepC, HepD and HepH and their reducing partner cytochrome P450 reductase, and performed in vitro biochemical enzymatic assays (Fig 24). The results indicates that only HepD is able to catalyze a Baeyer-Villiger oxidation reaction to convert compound **4** into the final product heptelidic acid. This discovery added another example of a cytochrome P450 which functions as a BVMO in secondary metabolism[123].

To further validate that HepG is the resistance gene, we next investigated the catalytic efficiency of the GAPDHs in the presence of heptelidic acid. In comparison, catalytic performance of GAPDHs from representative animal and fungi were also determined in the presence of heptelidic acid. All GAPDHs were first incubated with different concentrations of heptelidic acid for 10 min, then the reaction was initiated by adding glyceraldehyde-3-phosphate as substrate, and the catalytic efficiency was determined by measuring the initial reaction rate. The result showed that GAPDHs from human and yeast are very sensitive to heptelidic acid, with $IC_{50}$s of around

0.15 μM and 0.4 μM respectively (Fig 25). On the other hand, the $IC_{50}$s of the *T. virens* housekeeping GAPDH and HepG are around 100 μM and 1600 μM respectively, which showed that both of the GAPDH in the producing strain are more tolerant to heptelidic acid than the GAPDH in non-producing organisms. We proposed that HepG is the resistance gene, which has around 16-fold greater tolerance to heptelidic acid, however, the catalytic efficiency of HepG is around 10-fold lower than the housekeeping GAPDH in fungi. This indicates that HepG originally evolved from a normal GAPDH, and sacrificed catalytic efficiency to obtain resistance to

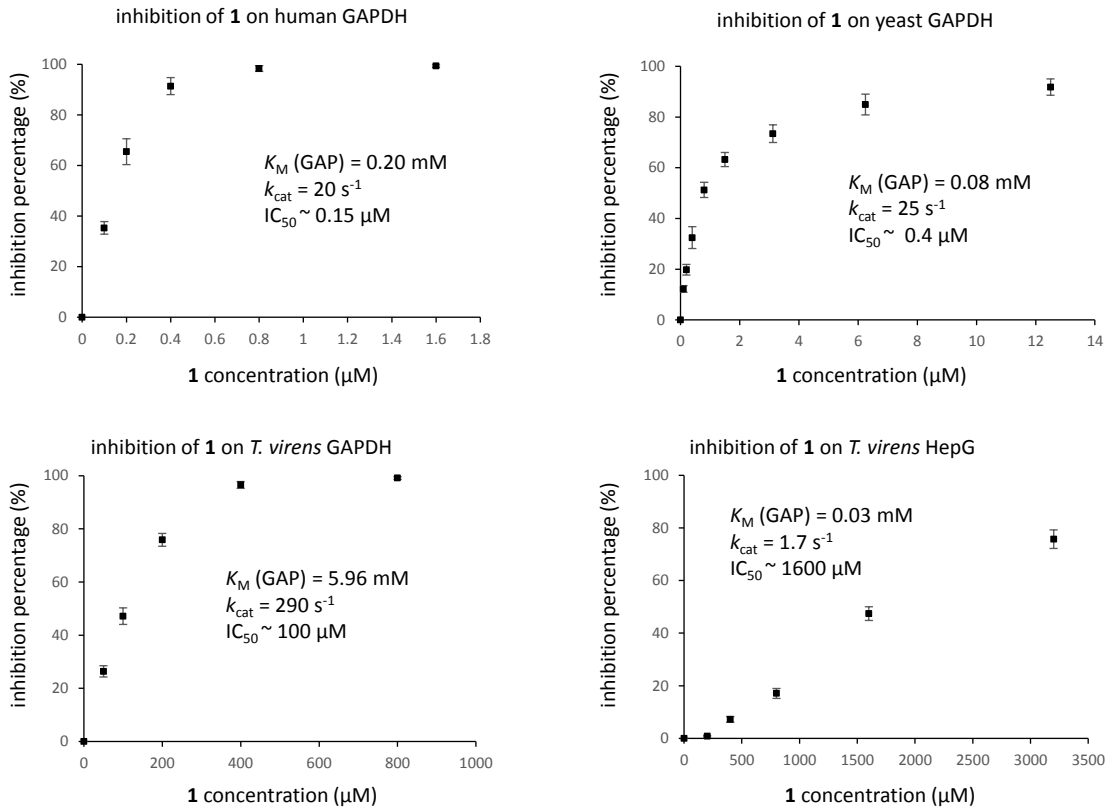



*Figure 25. Inhibition assay GAPDHs using heptelidic acid.*

*Biochemical analysis of GAPDHs were carried out in 100 μL reaction mixture containing 20 mM Tris-HCl pH 8.5, 0.2 mM EDTA, 1 mM DTT, 10 mM sodium arsenate, 1 mM $NAD^+$, and.* The reaction mixtures were first incubate with 0.01 ~ 3200 *μL heptelidic acid (**1**) for 10 min at 37 ˚C, then 0.01 ~ 1 mM glyceraldehyde-3-phosphate was added to initiate the reaction. The reaction rate was monitored by plate reader as the production of NADH, which has UV absorption at 340 nm.*

heptelidic acid. Meanwhile, the housekeeping GAPDH of the producer also evolved to be tolerant to heptelidic acid during coexistence with the toxic heptelidic acid. Compared to GAPDHs of yeast and human, the housekeeping GAPDH has a higher $K_M$ value, which indicates it sacrificed its substrate specificity to obtain tolerance to heptelidic acid, which is much higher than GAPDHs' in other species but still lower than the SRE HepG.

### 2.1.3 Conclusions

We proposed a resistance gene guided genome mining strategy to discover NPs of desired activity. To test our hypothesis, we first tried to locate the BGC of a known GAPDH inhibitor heptelidic acid using the resistance gene. We found that there are two copies of *gapdh* within the genome of the producing strain *T. virens* all together (XP_013949968.1 and XP_013958680.1). One of the *gapdh* (XP_013949968.1) is co-localized with terpenoid biosynthetic genes, which were predicted to be involved in heptelidic acid biosynthesis.

We demonstrated that this BGC is responsible for the biosynthesis of heptelidic acid using both an in vitro biochemical assay and in vivo biotransformation. HepG was also verified to be tolerant to heptelidic acid compare to the housekeeping tGAPDH. The validation of the heptelidic acid BGC proved our proposed resistance gene guided genome mining approach is feasible to identify the BGC of a NP with known bioactivity.

At this point we are still curious about the biomolecular mechanism that confers HepG tolerance to heptelidic acid, considering the identity of HepG and housekeeping GAPDH is around 70%. We are planning to elucidate the mechanism of self-resistance at the molecular level using structural biology approaches.

### 2.1.4 Materials and Methods

**General materials and methods**

Biological reagents, chemicals, media and enzymes were purchased from standard commercial sources unless stated. Plant, fungal, yeast and bacterial strains, plasmids and primers used in this study are summarized in Tables S2 and S3. DNA and RNA manipulations were carried out using Zymo ZR Fungal/Bacterial DNA Microprep™ kit and Invitrogen Ribopure™ kit respectively. DNA sequencing was performed at Laragen, Inc. The primers and codon optimized gblocks were synthesized by IDT, Inc.

**Construction of *Saccharomyces cerevisiae* strains.**

Plasmid pXW55 (*URA3* marker) digested with *Nde*I and *Pme*I was used to introduce the *hepA* gene (XP_013949969.1)[124]. A 1.2 kb fragment containing *hepA* obtained from PCR using primers HepA-xw55-recomb-F and HepA-xw55-recomb-R was cloned into pXW55 using yeast homologous recombination to afford pHepA-xw55. The plasmid pHepA-xw55 was then transformed into *Saccharomyces cerevisiae* RC01 to generate strain TY11[124].

Plasmid pXW06 (*TRP1* marker) digested with *Nde*I and *Pme*I was used to introduce the *hepH* gene (XP_013949970.1)[124]. A 1.6 kb fragment containing *hepH* obtained from PCR using primers HepH-xw06-recomb-F and HepH-xw06-recomb-R were cloned into pXW06 using yeast homologous recombination to afford pHepH-xw06. Plasmid pXW55 (*URA3* marker) digested with *Nde*I and *Pme*I was used to introduce the *T. virens* NADPH-cytochrome P450 reductase gene (XP_013956415.1). A 2.1 kb fragment containing *T. virens* cytochrome P450 reductase obtained from PCR using primers TvCPR-xw55-recomb-F and TvCPR-xw55-recomb-R was cloned into pXW55 using yeast homologous recombination to afford pTvCPR-xw55. The plasmid pHepH-

xw06 and pTvCPR-XW55 was then transformed into *Saccharomyces cerevisiae* RC01 to generate strain TY12.

Plasmid pXW02 (*LEU2* marker) digested with *Nde*I and *Pme*I was used to introduce the *hepD* gene (XP_013949971.1)[124]. A 1.6 kb fragment containing *hepD* obtained from PCR using primers HepD-xw02-recomb-F and HepD-xw02-recomb-R were cloned into pXW02 using yeast homologous recombination to afford pHepD-xw02. The plasmid pHepD-xw02 and pTvCPR-XW55 was then transformed into *Saccharomyces cerevisiae* RC01 to generate strain TY13.

Plasmid pXW06 (*TRP1* marker) digested with *Nde*I and *Pme*I was used to introduce the *hepC* gene (XP_013949972.1)[124]. A 1.6 kb fragment containing *hepC* obtained from PCR using primers HepC-xw06-recomb-F and HepC-xw06-recomb-R were cloned into pXW06 using yeast homologous recombination to afford pHepC-xw06. The plasmid pHepC-xw06 and pTvCPR-XW55 was then transformed into *Saccharomyces cerevisiae* RC01 to generate strain TY14.

Plasmid pXW02 (*LEU2* marker) digested with *Nde*I and *Pme*I was used to introduce the *hepE* gene (XP_013949974.1)[124]. A 1.6 kb fragment containing *hepE* obtained from PCR using primers HepE-xw02-recomb-F and HepE-xw02-recomb-R were cloned into pXW02 using yeast homologous recombination to afford pHepE-xw02. The plasmid pHepE-xw02 and pTvCPR-XW55 was then transformed into *Saccharomyces cerevisiae* RC01 to generate strain TY15.

**Fermentation and compound analyses and isolation.**

*Trichoderma virens* was maintained on PDA (potato dextrose agar, BD) 7 d for sporulation or in liquid PDB medium (potato dextrose broth, BD) for RNA extraction. *T. virens* was maintained in PDB medium for production of heptelidic acid and its biosynthetic intermediates at 28°C. To isolate heptelidic acid and its biosynthetic intermediates, fermentation broth of *T. virens*

was centrifuged (5000 rpm, 10 min), and supernatant was extracted three times with ethyl acetate. The organic phase was dried over sodium sulfate, concentrated to oil form, and subjected to HPLC purification.

HPLC-MS analyses were performed using a Shimadzu 2020 EVLC-MS (Phenomenex® Luna, 5μ, 2.0 × 100 mm, C-18 column) using positive and negative mode electrospray ionization. The elution method was a linear gradient of 5-95% (v/v) acetonitrile/water in 15 min, followed by 95% (v/v) acetonitrile/water for 3 min with a flow rate of 0.3 mL/min. The HPLC buffers were supplemented with 0.05% formic acid (v/v). HPLC purifications were performed using a Shimadzu Prominence HPLC (Phenomenex® Kinetex, 5μ, 10.0 × 250 mm, C-18 column). The elution method was a linear gradient of 65-100% (v/v) acetonitrile/water in 25 min, with a flow rate of 2.5 mL/min. GC-MS analyses were performed using Agilent Technologies GC-MS 6890/5973 equipped with a DB-FFAP column. An inlet temperature of 240°C and constant pressure of 4.2 psi were used. The oven temperature was initially at 60°C and then ramped at 10°C/min for 20 min, followed by a hold at 240°C for 5 min.

**Biotransformation using *Saccharomyces cerevisiae* strains.**

To perform biotransformation of heptelidic acid biosynthetic intermediates, strain TY12 and TY14 were first grown in 2 mL uracil and L-tryptophan double drop-out medium at 28 °C for 1 d, strain TY13 and TY15 were first grown in 2 mL uracil and L-leucine double drop-out medium at 28 °C for 1 d. 0.3 mL of the culture of each strain was then transferred into 3 mL fresh YPD medium (yeast extract 10 g/L, peptone 20 g/L supplement with 2% dextrose), and the cultures were shaken at 28 °C at 250 r.p.m. for 2 d. 0.1 mg of heptelidic acid biosynthetic intermediate was added to each culture, and then the cultures were shaken at 28 °C at 250 r.p.m., with 1 d to perform

biotransformation. For product detection, cell cultures were extracted with 1 mL ethyl acetate. After evaporation of the organic phase, the crude extracts were dissolved in 100 μl methanol for LC–MS analysis.

**Microsome purification and biochemical assay.**

Strain TY12 and TY14 were first grown in 2 ml uracil and L-tryptophan double drop-out medium at 28 °C for 1 d, strain TY13 and TY15 were first grown in 2 ml uracil and L-leucine double drop-out medium at 28 °C for 1 d. The 2 ml culture of each strain was then transferred into 50 mL fresh YPD medium (yeast extract 10 g/L, peptone 20 g/L supplement with 2% dextrose), and the cultures were shaken at 28 °C at 250 r.p.m. for 2 d.

The microsomes were prepared according to the protocol. Briefly, the cells were harvested by centrifugation (3,750 r.p.m. at 4 °C for 10 min) and the cell pellet was washed with 10 ml of TES buffer (50 mM Tris–HCl, pH, 7.5, 1 mM EDTA, 0.6 M sorbitol). The cells were centrifuged as above, resuspended in 5 ml of TEG-M (50 mM Tris–HCl, pH, 7.5, 1 mM EDTA, 20% glycerol supplemented with 1.5 mM 2-mercaptoethanol), and 25 μL protease inhibitor cocktail (Sigma) was added. Zirconia/silica beads (0.5 mm in diameter, BioSpec Products) were added to surface of the cell suspension. Cell walls were disrupted manually by hand shaking in a cold room for 10 min at 30-s intervals separated by 30-s intervals on ice. Cell suspension were aliquoted and transferred to a 1.5 mL centrifuge tube. Finally, microsomes were obtained by differential centrifugation at 10,000g for 5 min at 4 °C to remove cellular debris (precipitant) followed by centrifugation at 100,000g for 30 min at 4 °C and remove the supernatant. The microsomal pellets were resuspended in 1.5 mL of TEG-M buffer (50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 20% glycerol, and 1.5 mM 2-mercaptoethanol) and stored frozen at −80 °C.

Assays for microsome bioactivity with **4** in TEG-M buffer were performed at 100 μL scale with microsomal cytochrome P450 and the reducing partner, 0.4 mM NADPH, 100 μM FAD, and 100 μM FMN 28 °C for 10 h.

**Protein expression, purification and biochemical assay.**

To express and purify HepG (XP_013949968.1), primers HepG-pET-F and HepG-pET-R were used to amplify a 1.0 kb DNA fragment containing *hepG*. The PCR product was cloned into pET28a using *Nde*I and *Not*I restriction sites. The resulted plasmid HepG-pET was transformed into *E. coli* BL21 (DE3) to obtain TY16. To express and purify tGAPDH (XP_013958680.1), primers tGAPDH-pET-F and tGAPDH-pET-R were used to amplify a 1.0 kb DNA fragment containing *T. virens* GAPDH. The PCR product was cloned into pET28a using *Nde*I and *Not*I restriction sites. The resulted plasmid tGAPDH-pET was transformed into *E. coli* BL21 (DE3) to obtain TY17. To express and purify hGAPDH (AAP36549.1), a 1.0 kb fragment containing codon optimized *Homo sapiens gapdh* was synthesized. The fragment was cloned into pSJ2 ($Amp^R$ marker) using *Eco*RI and *Xho*I restriction sites. The resulting plasmid hGAPDH-pSJ2 was transformed into *E.coli* BL21 (DE3) to give TY18.

All GAPDHs fused a 6×His-tag with a molecular weight ~38 kD were expressed at 16°C 220 rpm for 20 h after 100 μM IPTG induction (IPTG was added when $OD_{600}$ = 0.8). Cells of 1 L culture were then harvested by centrifugation at 5000 rpm at 4°C. Cell pellet was resuspended in 15 mL Buffer A10 (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 8% glycerol, 10 mM imidazole). The cells were lysed by sonication, and the insoluble material was sedimented by centrifugation at 16000 rpm at 4°C. The protein supernatant was then incubated with 3 mL Ni-NTA for 4 h with slow, constant rotation at 4°C. Subsequently the Ni-NTA resin was washed with 10 column

volumes of Buffer A50 (Buffer A + 50 mM imidazole). For elution of the target protein, the Ni-NTA resin was incubated for 10 min with 6 mL Buffer A250 (Buffer A + 250 mM imidazole). The supernatant from the elution step was then analyzed by SDS-PAGE together with the supernatants from the other purification steps. The elution fraction containing the recombinant protein was buffer exchanged into storage buffer (50 mM Tris-HCl pH 8.5, 50 mM NaCl, 10% glycerol, 5 mM DTT, 5 mM GSH).

*In vitro* activity assays were carried out in 50 µL reaction mixture containing 20 mM Tris pH 8.75, 0.2 mM EDTA, 10 mM sodium arsenate, 1 mM DTT, 1 mM $NAD^+$, and 0.01 µM of purified GAPDH enzyme with or without 3200 ~ 0.1 µM heptelidic acid. After 10 min incubation at 30˚C, the reaction was initiated by adding 1 ~ 0.008 mM D-glyceraldehyde-3-phosphate hydrate. The reaction was monitored by UV absorption at 340 nm to detect the production of NADH.

The inhibition percentage of heptelidic acid on GAPDHs determined using *in vitro* biochemical assays are calculated by following equation:

$$\text{inhibition percentage} = 1 - \frac{\text{initial reaction rate with heptelidic acid}}{\text{initial reaction rate without heptelidic acid}}$$

## 2.2 Discovery of an herbicide with new mode of action

### 2.2.1 Introduction

After validating our proposal by successfully locating the BGC of heptelidic acid using a resistance gene guided method, we wanted to further use this approach to solve a more difficult problem. Our ultimate goal is to discovery new NPs with desired bioactivity, rather than find BGCs of existing NPs with known bioactivities. Thus we tried to use both pharmaceutical and pesticide targets as a query to search for NPs with new modes of action; herein we will show one successful example to explain how to perform resistance gene guided genome mining to discover new NPs of desired bioactivity.

Weeds, one of the major causes of tremendous worldwide crop loss, are effectively controlled by herbicides[125-127]. As herbicides are increasingly applied in crop production worldwide, however,

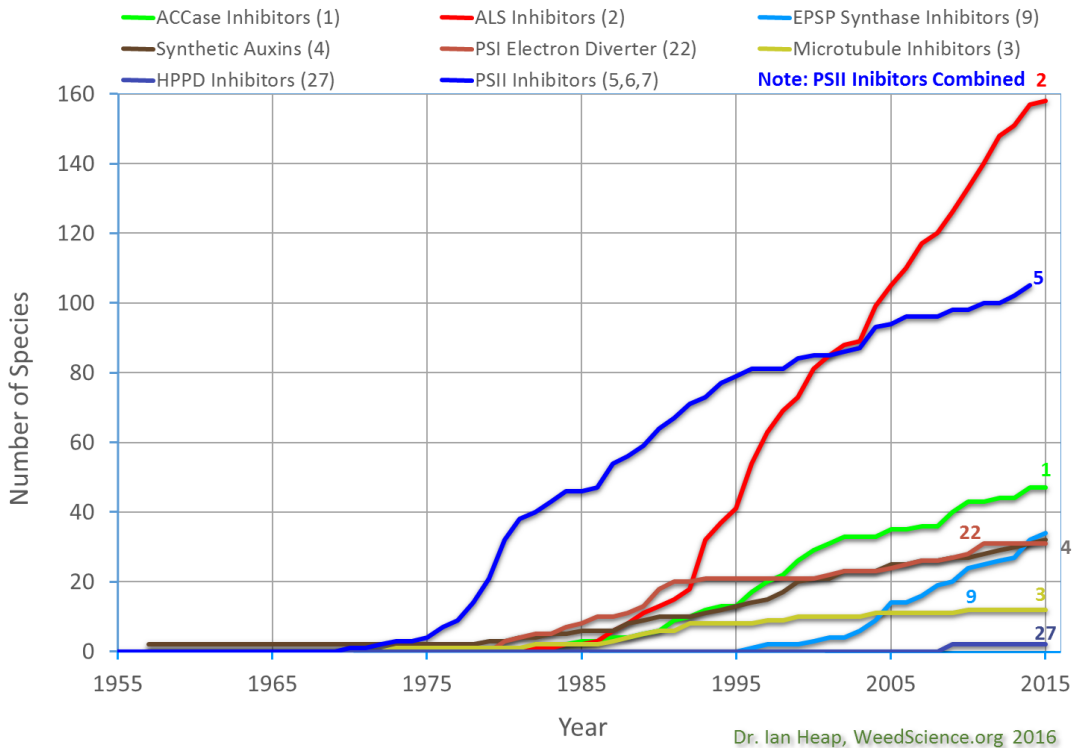



*Figure 26. Number of resistant species for several herbicide sites of action (WSSA codes).*

the demand of an herbicide with a novel mode of action becomes ever more urgent mainly due to continuously emerging weed resistance[128-130] (Fig 26). According to an international survey of herbicide resistant weeds, there are currently 498 unique cases of herbicide resistant weeds globally within 255 species including 148 dicots and 107 monocots[131]. Weeds have evolved resistance to 23 of the 26 known herbicide sites of action and to 166 different herbicides[131]. It is estimated that crop production in the United States would decline by 20% without the application of herbicides[127]. The constant and often heavy usage of herbicides results in many weeds evolving herbicide resistance[128,129].

The evolution of herbicide resistance in weeds has led to an urgent need for new herbicides with novel modes of action[126,128,132,133]. However, there is no herbicide with a new mode of action that has been commercialized for more than 30 years[129]. The BCAAs biosynthetic pathway is

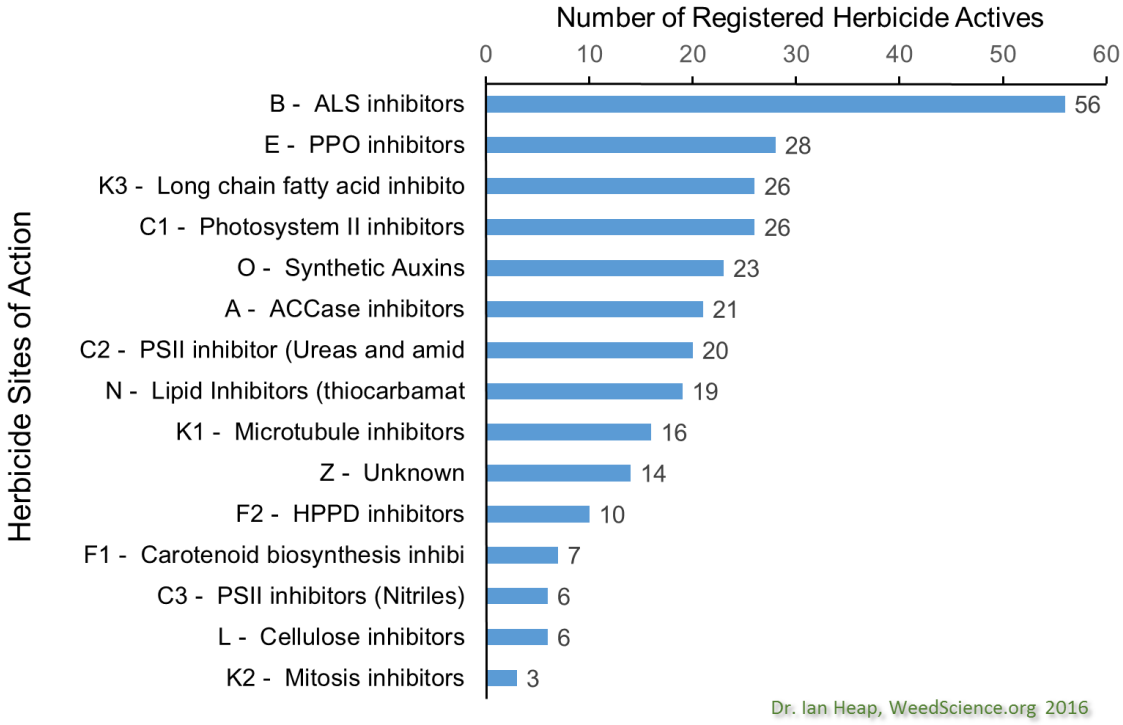



*Figure 27. Number of registered herbicides for the major herbicide sites of action (HRAC codes).*

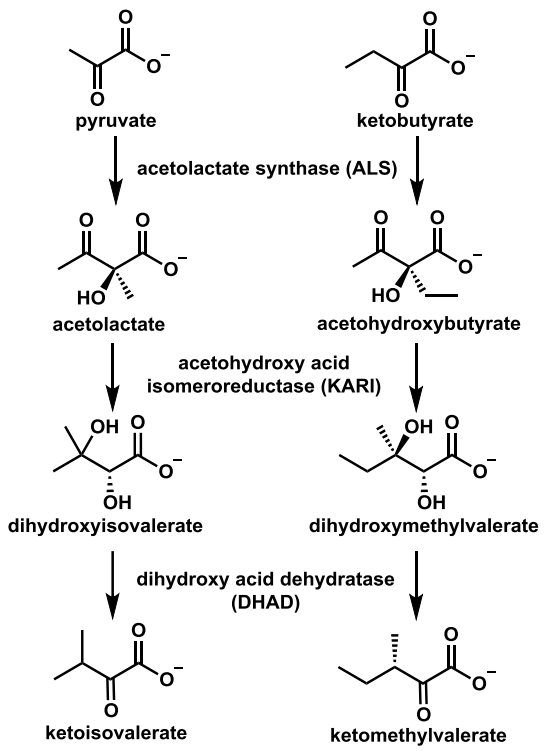



*Figure 28. Partial branched chain amino acid biosynthetic pathway.*

essential for plant growth[134,135]. It is not present in animals and is therefore a validated target for highly specific weed control agents[134,135] (Fig 27). The BCAA biosynthetic pathway in plants is carried out by three enzymes: acetolactate synthase (ALS), acetohydroxy acid isomeroreductase (KARI), and dihydroxy acid dehydratase (DHAD) (Fig 28). Among them, ALS has been the target for commercially successful herbicides since 1980, and currently is the second largest class of active herbicidal products in weed control for many non-transgenic crops, such as sulfonylureas, imiazolinones and triazolopyrimidines[135]. Although potent and selective inhibitors of KARI and DHAD have also been identified, these inhibitors show weak herbicidal activity[136-138]. Acorrding to the sequence comparison studys across species, KARI can be divided into two classes. Class I KARI (a short form ~400 amino acid residues) was found in fungi and most bacteria, and Class II KARI (a long form ~600 amino acid residues) is in all plants. As a result, KARI is not well conserve among plant and fungal species, the identity is only about 30%. Compare to KARI, DHAD is the more conserved in both plant and fungi, thus it is resonable to predict fungi may produce NPs to target this enzyme in plants (Fig 29). As a result, DHAD was selected as the SRE to carry out resistance gene guided genome mining to discover an herbicide with a new mode of action.

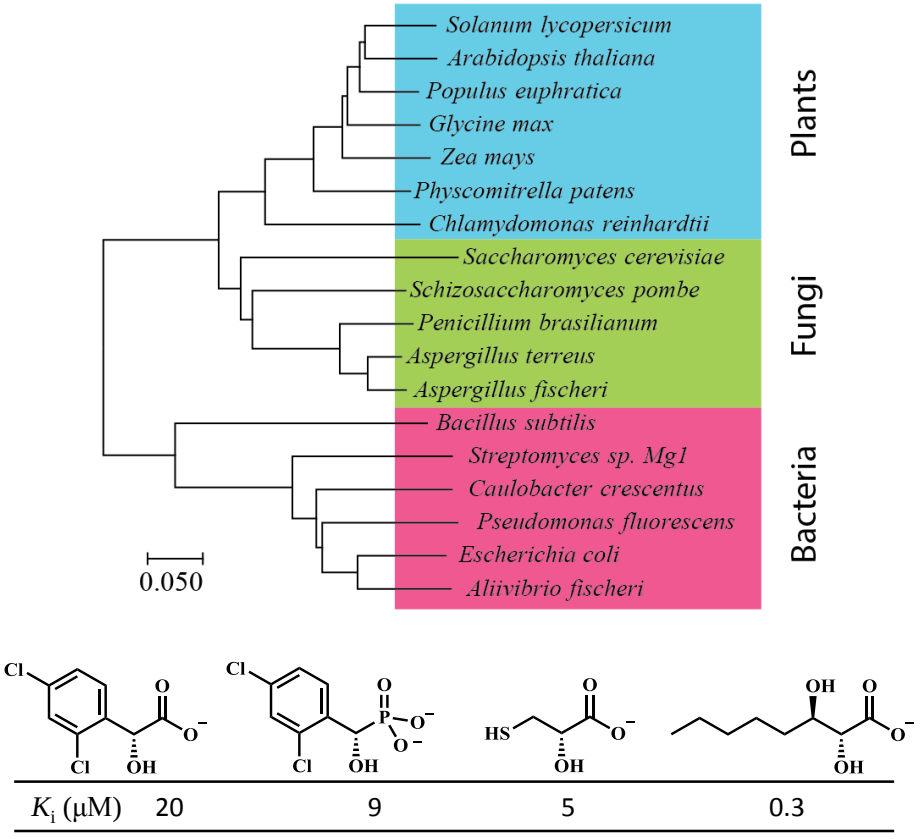



*Figure 29. Phylogenetic tree of DHAD among bacteria, fungi and plants (top). Representatives of small molecules that inhibit DHAD in vitro, but fail to inhibit plant growth (bottom).*

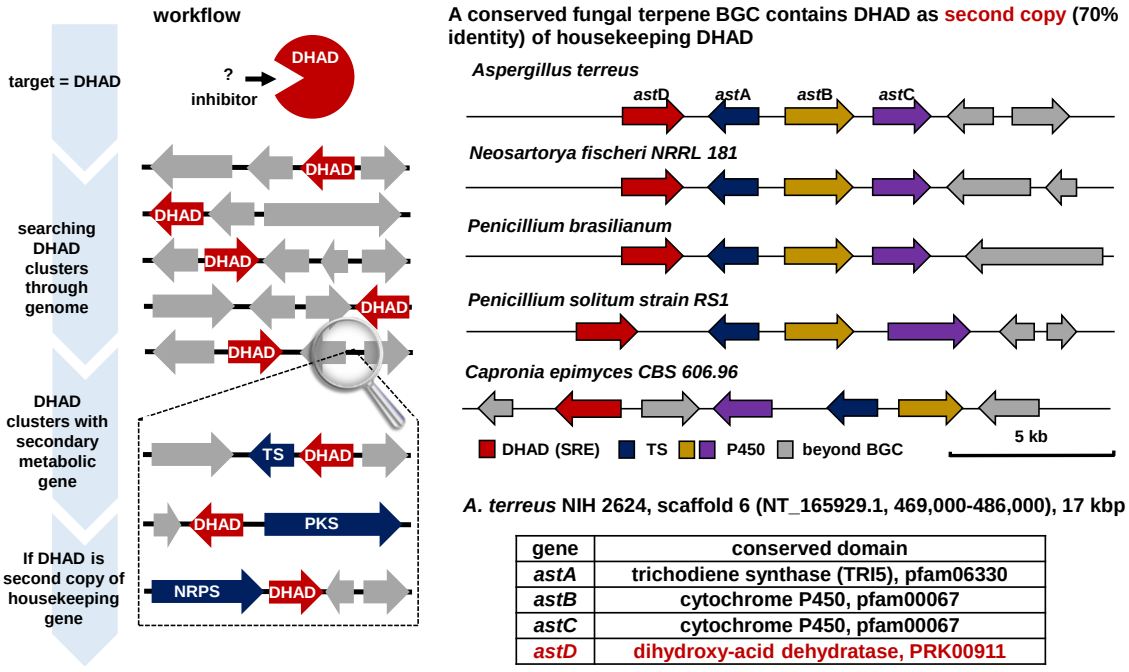
*Phylogenetic tree of DHAD among bacteria, fungi and plants. The evolutionary history was inferred by using the Neighbor-Joining method (MEGA7). Units represent the number of amino acid substitutions per site.*

Given the success of targeting ALS for herbicide development[128], it is surprising that no herbicide that targets either of the other two enzymes has been developed. DHAD is an essential and highly conserved enzyme among plant species which catalyzes β-dehydration reactions to yield α-keto acid precursors to isoleucine, valine and leucine (Fig 28 & 29)[135-139]. Efforts toward synthetic DHAD inhibitors resulted in compounds with submicromolar $K_i$; however, the compounds have no *in planta* activity (Fig 29)[136].

### 2.2.2 Results and discussions

Filamentous fungi are prolific producers of natural products (NPs), many of which have biological activities that aid the fungi in colonizing and killing plants[20,140]. Therefore, fungal NPs represent a promising source of potential leads for herbicide development. The abundance of sequenced fungal genomes enables genome mining of new NPs with novel biological activities[141-143]. Although no NP inhibitors of DHAD are known to date, we reason that a fungal NP with this property might exist, given the indispensable role of BCAA biosynthesis in plants[134,135].

To identify NP biosynthetic gene clusters that may encode a DHAD inhibitor, we hypothesized that such a cluster must contain an additional copy of DHAD that is insensitive to




| gene | conserved domain |
|---|---|
| *astA* | trichodiene synthase (TRI5), pfam06330 |
| *astB* | cytochrome P450, pfam00067 |
| *astC* | cytochrome P450, pfam00067 |
| *astD* | dihydroxy-acid dehydratase, PRK00911 |

*Figure 30. Resistance gene guided genome mining to discover a natural product inhibitor of dihydroxyacid dehydratase.*

*A 17 kb gene cluster from A. terreus containing four ORFs, which are also conserved among several other fungal species. AstA is homologous to sesquiterpene cyclase; AstB and AstC are predicted to be P450 monooxygenases; astD is predicted to encode a DHAD, and is proposed to confer self-resistance in the presence of the NP produced in the cluster.*

the inhibitor, thereby providing the required self-resistance for the producing organism to survive[79]. The presence of a gene encoding an SRE is frequently found in microbial NP gene clusters, as highlighted by the presence of an insensitive copy of HMGR or IMPDH in the gene clusters for lovastatin (that targets HMGR) or mycophenolic acid (that targets IMPDH), respectively (Fig 11)[78,144]. This phenomenon has been used to predict molecular targets of NPs, as well as to identify gene clusters for NPs of known activities[79,80,145].

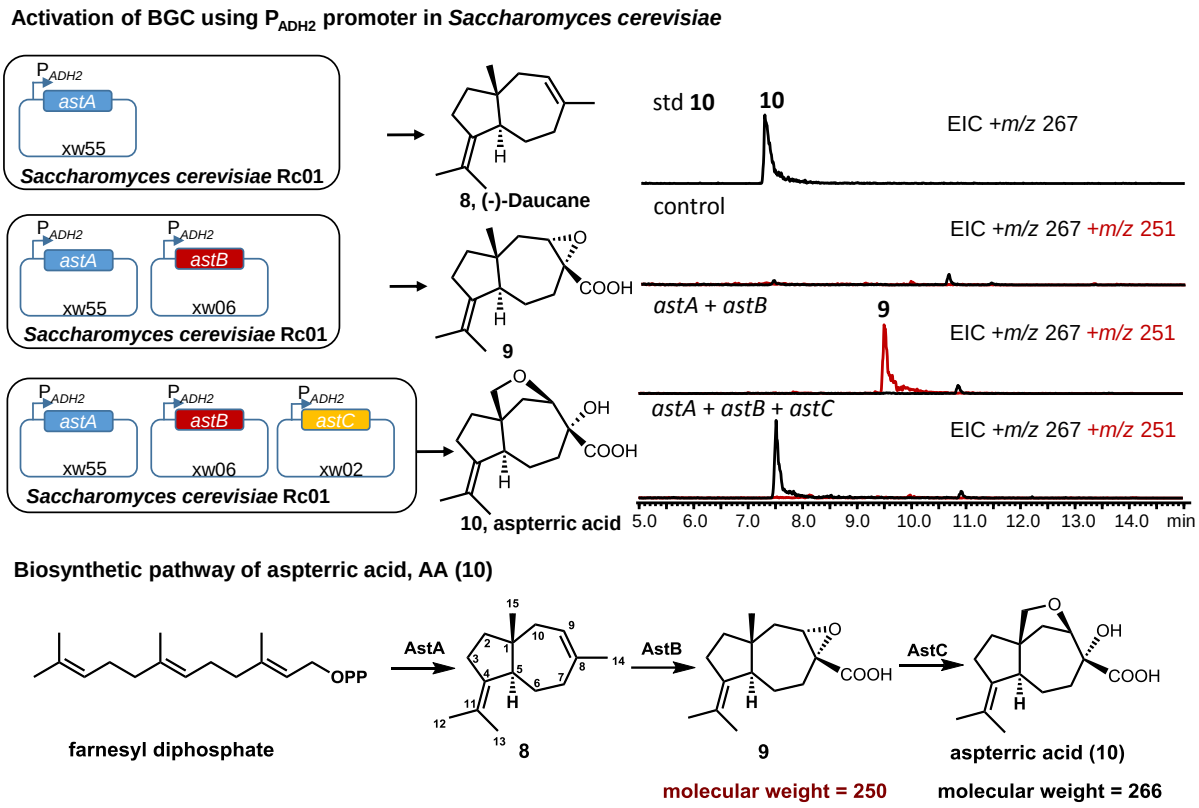



*Figure 31. Reconstitution of ast gene cluster in Saccharomyces cerevisiae.*

*HPLC-MS traces of metabolites produced from S. cerevisiae RC01 expressing different ast genes under $P_{ADH2}$ promoter control. i: S. cerevisiae without expression plasmids. ii: S. cerevisiae transformed with plasmids expressing astA and astB produces **9**. iii: S. cerevisiae transformed with plasmids expressing astA-C produces **AA** at a titer of 20 mg/L. The experiments were repeated independently with similar results for 3 times. Bottom, Proposed biosynthetic pathway of **AA**. AstA cyclizes farnesyl diphosphate (FPP) into (-)-daucane **1**, while the P450 enzymes AstB and AstC sequentially transform **8** into **9** and **10 (AA)**, respectively.*

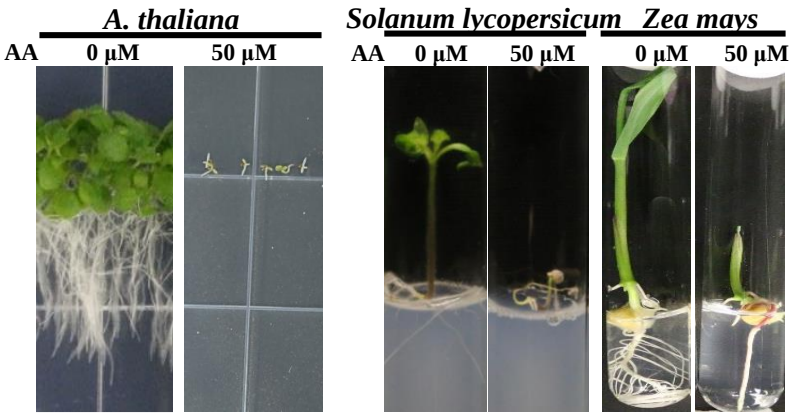



*Figure 32. Growth inhibition of plants on agar using aspterric acid.*

*Left, 2-week old Arabidopsis thaliana growing on MS media containing no* ***AA*** *(left) or 50 µM* ***AA*** *(right). The picture shown is representative of 3 replicates. Right, Same as in left, except for 2-week old dicot Solanum lycopersicum and monocot Zea mays. The picture shown is representative of 2 replicates.*

To identify possible self-resistance enzymes, we scanned sequenced fungal genomes to search for co-localizations of genes encoding DHAD with biosynthetic core enzymes, such as terpene cyclases, polyketide synthases, etc[21,23]. We identified a well-conserved set of four genes across multiple fungal genomes (Fig 30), including the common soil fungus *Aspergillus terreus* that is best known to produce lovastatin[146]. The conserved gene cluster include genes that encode a sesquiterpene cyclase homolog (*astA*), two cytochrome P450s (*astB* and *astC*), and a homolog of DHAD (*astD*) (Fig S1). Genes outside of this cluster are not conserved across the identified genomes and are hence unlikely to be involved in NP biosynthesis. AstD is the second copy of DHAD encoded in the genome, and is ~70% similar to the housekeeping copy that is well-conserved across fungi (Fig S2). Therefore, AstD is potentially a self-resistance enzyme that confers resistance to the encoded NP. Like a majority of biosynthetic gene clusters in sequenced fungal genomes, the *ast* cluster has not been associated with the production of a known NP[147,148].

To identify the NP encoded by the *ast* cluster, we heterologously expressed *astA*, *astB*, and *astC* genes in the host *Saccharomyces cerevisiae* RC01[124]. New compounds that emerged were purified and their structures were elucidated with NMR spectroscopy (Fig S3 and Table S5). RC01 expressing only *astA* produced a new sesquiterpene (**8**), which was confirmed to be (-)-daucane[149] (Fig S4). RC01 expressing both *astA* and *astB* led to the biosynthesis of a new product that was

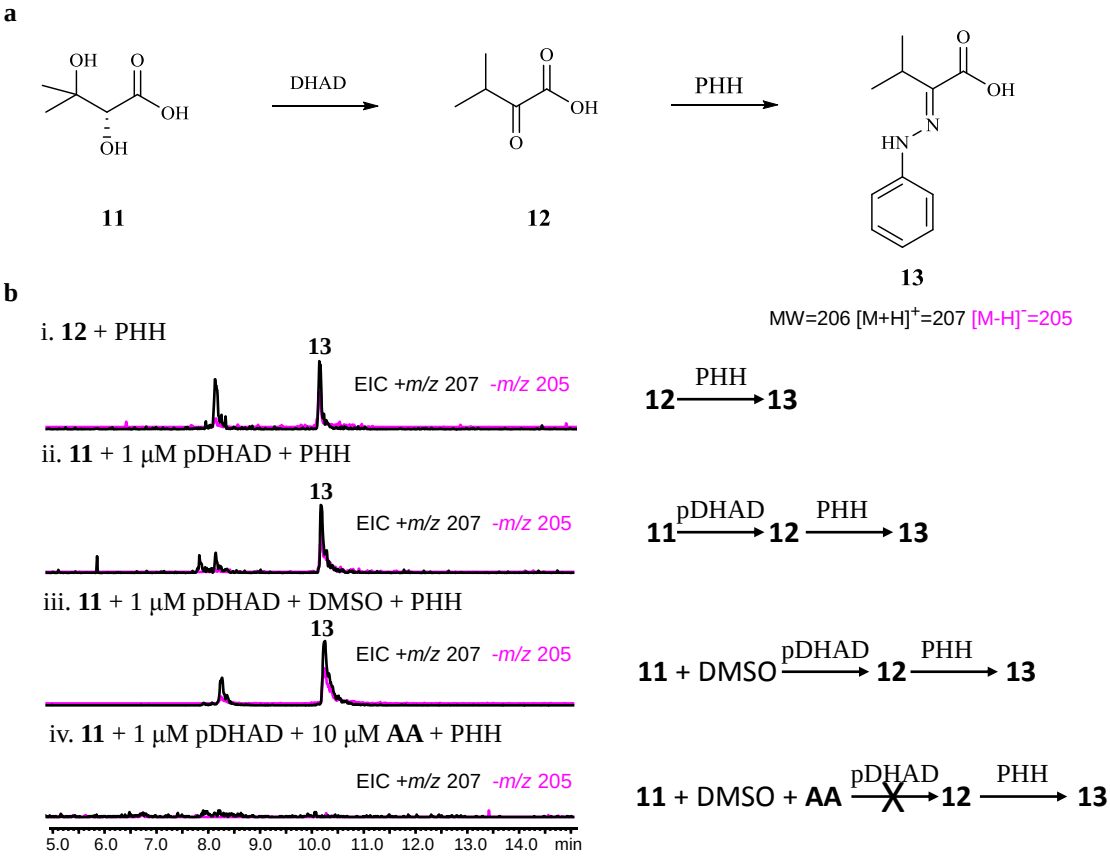



*Figure 33. Inhibition assay of aspterric acid on pDHAD.*

***a,*** *Assaying DHAD activities in converting the dihydroxyacid* ***11*** *into the α-ketoacid* ***12.*** *Formation of* ***12*** *can be detected on HPLC by chemical derivatization using phenylhydrazine (PHH) to yield* ***13****.* ***b,*** *LC-MS traces of the biochemical assays of A. thaliana DHAD (pDHAD). Extracted ion chromatogram (EIC) of positive ion mass of* $[M+H]^+=207$ *is shown in black, and the negative anion mass of* $[M+H]^-=205$ *is shown in pink. i. The derivatization reaction was validated by using the authentic* ***12****. ii. The bioactivity of pDHAD in converting* ***11*** *into* ***12*** *was validated. iii. Addition of DMSO to pDHAD enzymatic reaction mixture has no effect. iv. Addition of 10 μM* ***AA*** *to the reaction mixture abolished pDHAD activity. The experiments were repeated independently for 3 times with similar results.*

structurally determined to be the α-epoxy carboxylate (**9**) (Fig 31). When *astA*, *astB* and *astC* were expressed together, a new compound (**10**) became the dominant product (~ 20 mg/L). Full structural determination revealed the compound to be the tricyclic aspterric acid (**AA**), which is a previously isolated compound[150] (Fig 31). The biosynthetic pathway for **AA** is therefore concise: following cyclization of farnesyl diphosphate by AstA to create the carbon skeleton in **1**, AstB

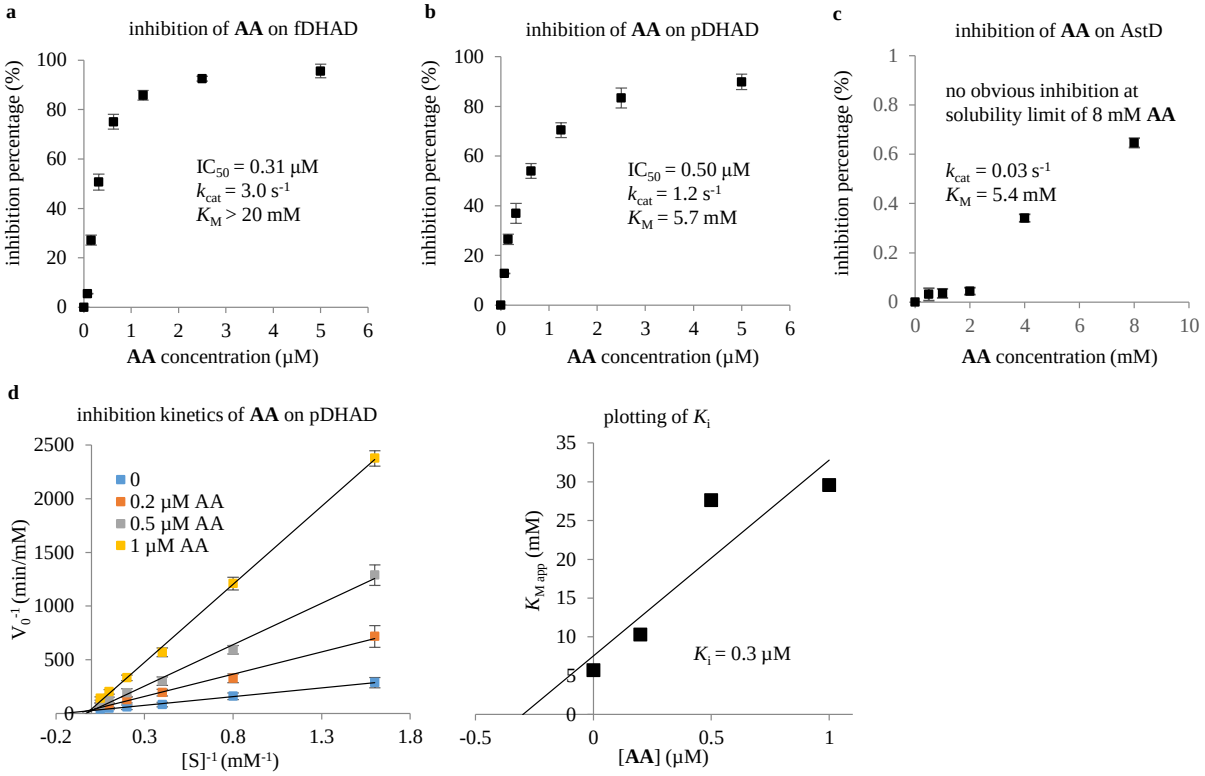



*Figure 34. Inhibition kinetics of aspterric acid on DHADs.*

*Three DHAD enzymes were assayed, including pDHAD (**p**lant DHAD from A. thaliana), fDHAD (**f**ungal housekeeping DHAD from A. terreus) and AstD (DHAD homolog within ast cluster). $IC_{50}$ and $K_i$ values of **AA** were measured based on inhibition percentage at different **AA** concentrations. Center values are averages, errors bars are s.d.; n = 3 biologically independent experiments. **a,** Plot of the inhibition percentage of 0.5 μM fDHAD as a function of **AA** concentration. **b,** Plot of the inhibition percentage of 0.5 μM pDHAD as a function of **AA** concentration. **c,** Plot of the inhibition percentage of 0.5 μM AstD as a function of **AA** concentration. **d,** Analysis of inhibitory kinetics of **AA** on pDHAD using the Lineweaver-Burk method at different concentrations of **AA** (left). Linear fitting of apparent Michaelis constant ($K_{M,app}$) as a function of **AA** concentration yields the inhibition constant ($K_i$) of **AA** on pDHAD (right).*

catalyzes oxidation of **8** to yield the epoxide **9.** Further oxidation by AstC at carbon 15 yields an alcohol, which can undergo intramolecular epoxide opening to create **AA** (Fig 31).

Upon its initial discovery, **AA** was shown to have inhibitory activity towards *Arabidopsis thaliana*, however, the mode of action was not known[151]. Our resistance-gene directed approach led to rediscovery of this compound with DHAD as a potential target. We first confirmed that **AA** is able to potently inhibit *A. thaliana* growth in an agar-based assay (Fig 32 & S6). **AA** was also an effective inhibitor of root development and plant growth when applied to a representative monocot (*Zea mays*) and dicot (*Solanum lycopersicum*) (Fig 31& S6). To test if **AA** indeed targets DHAD, we expressed and purified housekeeping DHAD from both *A. terreus* (XP_001208445.1, fDHAD) and *A. thaliana* (AT3G23940, pDHAD), as well as the putative self-resistance enzyme AstD (Fig S7). Both housekeeping DHAD enzymes converted dihydroxyisovalerate to ketoisovalerate (pDHAD: $k_{cat}$ = 1.2 $sec^{-1}$, $K_M$ = 5.7 mM) as expected[152,153]. The enzyme activities, however, were inhibited in the presence of **AA** (Fig 33 & 34). The $IC_{50}$ values of **AA** towards fDHAD and pDHAD were 0.31 μM and 0.50 μM at an enzyme concentration of 0.50 μM,

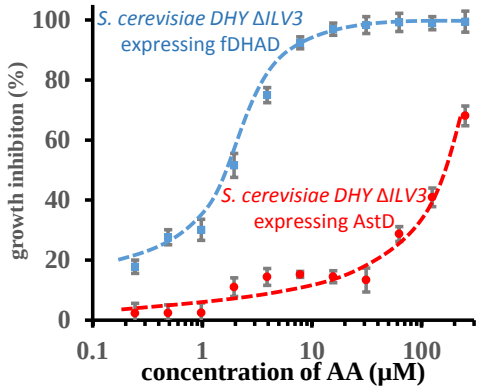



*Figure 35. Verification of the self-resistance function of AstD.*

*Growth inhibition curve of **AA** on S. cerevisiae ΔILV3 strains expressing fungal housekeeping fDHAD (blue) or AstD (red) in isoleucine, leucine and valine (ILV) dropout media. The plot shows mean values ± s.d. (error bars); n = 3 biologically independent experiments.*

respectively (Fig 34). **AA** was further determined to be a competitive inhibitor of pDHAD with a $K_i$ = 0.30 μM (Fig 34). **AA** displayed no significant cytotoxicity towards human cell lines up to 500 μM concentration, which is consistent with the lack of DHAD in mammalian cells (Fig S8).

AstD catalyzes the identical β-dehydration reaction as DHAD, albeit with a significantly more sluggish turnover rate ($k_{cat}$ = 0.03 $sec^{-1}$, $K_M$ = 5.4 mM). However, the enzyme was not inhibited by **AA**, even at the solubility limit of 8 mM (Fig 34). To determine if AstD can confer resistance to **AA**-sensitive strains, we developed a yeast based assay. The DHAD encoded by *ILV3* was first deleted from *Saccharomyces cerevisiae* strain DHY *ΔURA3*, which resulted in an auxotroph that requires exogenous addition of Ile, Leu and Val to grow[154]. We then introduced either *fDHAD* or *astD* episomally, both of which allowed the strain to grow in the absence of the three BCAAs (Fig S5). However, yeast expressing fDHAD was approximately 100 times more

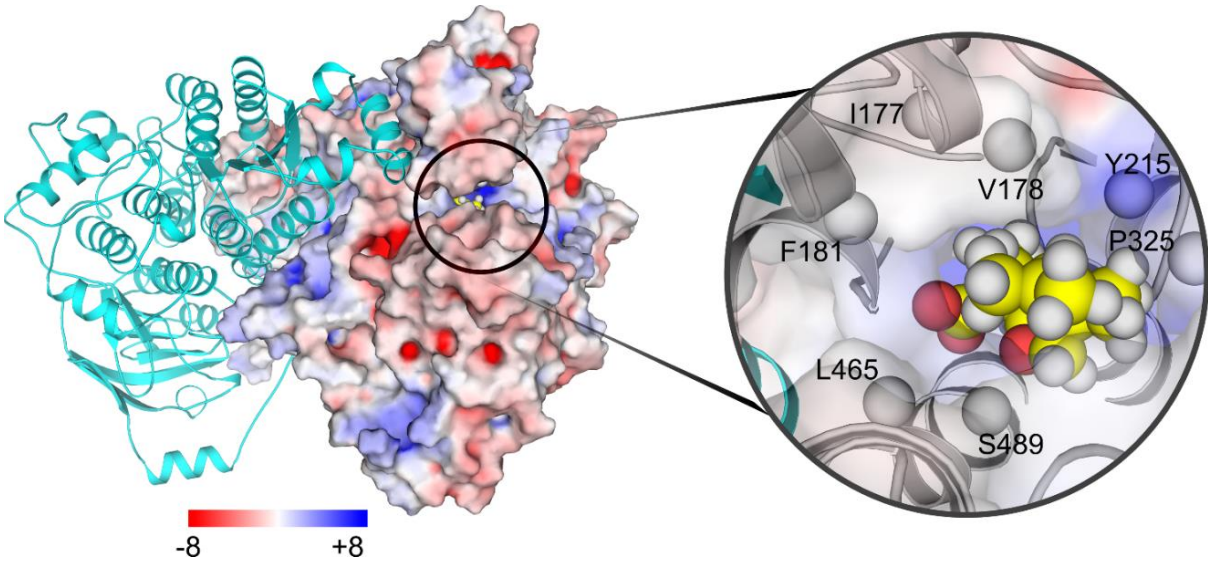



*Figure 36. Binding of aspterric acid to DHAD active site.*

*Crystal structure of the dimeric holo A. thaliana DHAD (pDHAD) containing the cofactor 2Fe-2S cluster and a $Mg^{2+}$ ion with the docked **AA** in the active site. One of the pDHAD monomers is show in cyan, whereas the other one is shown in electrostatic surface representation. The docked **AA** is shown in the inset in spaced-filled model. The hydrophobic portions of **AA** are surrounded by several hydrophobic residues (white spheres) from both monomers.*

sensitive to **AA** ($IC_{50}$ of 2 μM) compared to yeast expressing AstD ($IC_{50}$ of 200 μM) (Fig 35). Growth inhibition curve of **AA** on *S. cerevisiae ΔILV3* strains expressing fungal housekeeping fDHAD (blue) or AstD (red) in isoleucine, leucine and valine (ILV) dropout media showed that *astD* is the resistance gene that can confer **AA** tolerance to *S. cerevisiae*. Collectively, the biochemical and genetic assays validated **AA** as the first natural product inhibitor of fungal and plant DHAD; and AstD serves as the self-resistance enzyme in the *ast* biosynthetic gene cluster.

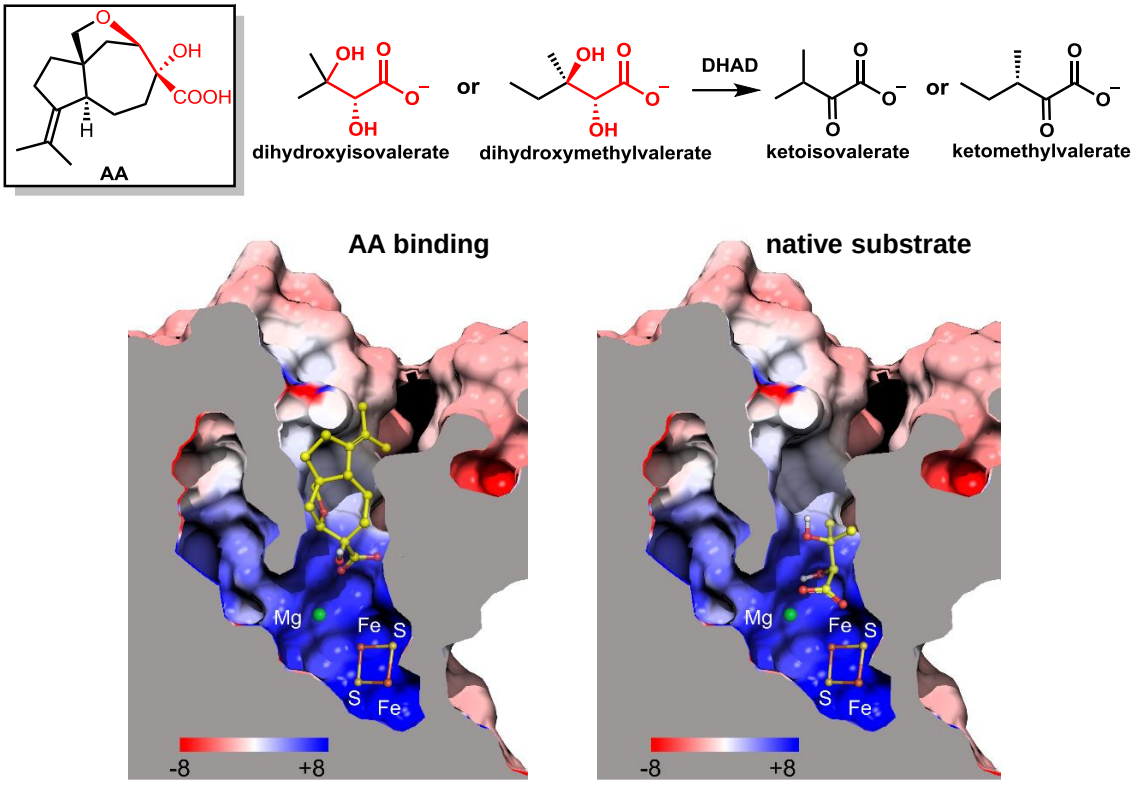



*Figure 37. Comparison of the binding mode of **AA** and native substrate to the active site of dihydroxyacid dehydratase.*

*The structures comparison of both **AA** and native substrates, in which the similar structures are highlighted in red (top). Shown at the bottom is cross-section electrostatic map of modeled holo-pDHAD in the binding site. Red map: the normalized negatively charged regions; blue map: the normalized positively charged regions; white map: the hydrophobic regions. The docked **AA** in the active site of pDHAD is shown on the left, while the docked native substrate dihydroxyisovalerate is shown on the right. The docking studies suggest the hydrophobic entrance to the reaction chamber preferentially binds the bulkier, tricyclic **AA**.*

The (*R*)-α-hydroxyacid and (*R*)-configured β-ether oxygen moieties in **AA** mimic the (*2R, 3R*)-dihydroxyl groups present in natural substrates such as dihydroxyisovalerate. The β-ether oxygen in **AA** is in position to coordinate to the 2Fe-2S cluster that is a required cofactor in both fungal and plant DHAD[152,153]. To understand potential **AA** mechanism of action, we determined the crystal structure (2.11 Å) of the pDHAD complexed with 2Fe-2S cluster (*holo*-pDHAD) (Fig 36). We identified a binding chamber at the homodimer interface, similar to that found in the *holo* bacterial L-arabinonate dehydratase[155] (Fig 36). The interior of the chamber is positively charged (2Fe-2S and $Mg^{2+}$) while the entrance is lined with hydrophobic residues. The modeled binding mode of α,β-dihydroxyisovalerate and **AA** predicted by computational docking are shown in Fig 37. The pocket is sufficiently spacious to accommodate the bulkier **AA**, and provide stronger hydrophobic interactions than the native substrate with a 5.3 ± 0.3 kcal/mol gain in binding energy (Fig 37). Based on the *holo*-pDHAD structure, we constructed a homology model of AstD to determine potential mechanism of resistance (Fig 38). Comparison of pDHAD and the modeled AstD structures shows that while most of the residues in the catalytic chamber are conserved, the hydrophobic region at the entrance to the reactive chamber in AstD is more constricted as a result of two amino acid substitutions (V496L and I177L). Narrowing of the entrance could therefore sterically exclude the bulkier **AA** from binding in the active site, while the smaller, natural substrates are still able to enter the chamber (Fig 38).

Superimpositions of the active sites of *holo*-pDHAD and homology modeled AstD indicate the 3D structures are similar. The structure of AstD was constructed by homology modeling based on the structure of *holo*-pDHAD. The structure of *holo*-pDHAD is in white; the crystal structure of AstD is in green (Fig 38). Comparison of the active sites in the crystal structure of pDHAD and the modeled structure of AstD showed the main different residues. The cartoon represents

superimposed binding sites of pDHAD (white) and AstD (green) (Fig 38). The shift of a loop in AstD, where L518 (correspond to V496 in pDHAD) is located, coupled with a larger L198 residue (correspond to I177 in pDHAD) lead to a smaller hydrophobic pocket of AstD than that in pDHAD

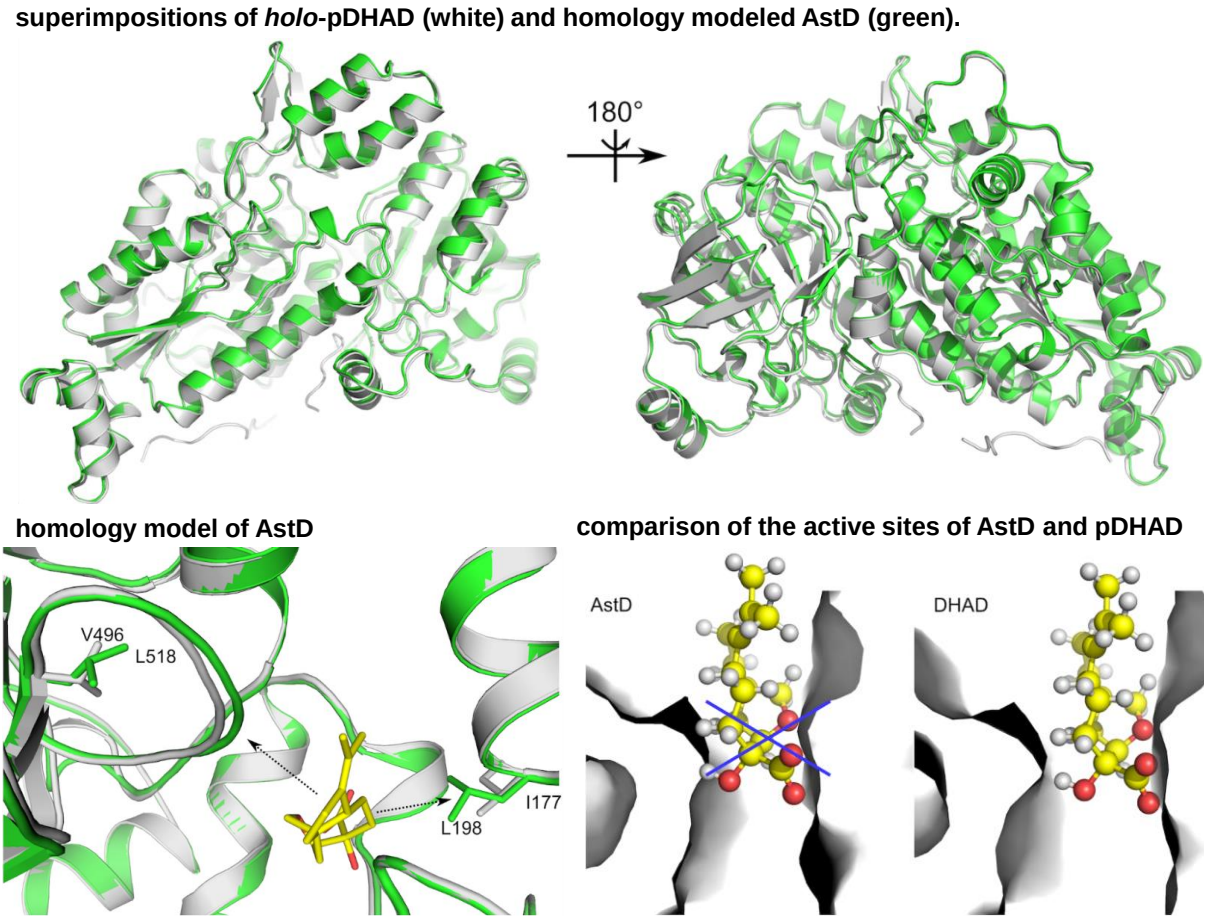



*Figure 38. X-ray Structure of holo-pDHAD and homology model of AstD.*

*Top, superimpositions of holo-pDHAD and homology modeled AstD. The structure of AstD was constructed by homology modeling based on the structure of holo-pDHAD. The structure of holo-pDHAD is in white; the crystal structure of AstD is in green. Bottom left, Comparison of the active sites in the crystal structure of pDHAD and the modeled structure of AstD. The cartoon represents superimposed binding sites of pDHAD (white) and AstD (green). The shift of a loop in AstD, where L518 (correspond to V496 in pDHAD) is located, coupled with a larger L198 residue (correspond to I177 in pDHAD) lead to a smaller hydrophobic pocket of AstD than that in pDHAD. Bottom right, the surface of binding sites of AstD (left) and pDHAD (right). The smaller hydrophobic channel in modeled AstD cannot accommodate the* ***AA*** *molecule (yellow balls-and-sticks).*

(Fig 38). The surface of binding sites of AstD (left) and pDHAD (right) showed the resistance mechanism at biomolecular level. The smaller hydrophobic channel in modeled AstD cannot accommodate the **AA** molecule (yellow balls-and-sticks).

As we showed in the introduction section, there are many NP resistance mechanisms to evolve which were utilized by the producers or (or targets) to survive in the presence of lethal concentration of toxic compounds[89,156,157]. Some of these resistance mechanisms have been applied by scientists to develop "marker" genes for cloning selection, others were used to develop herbicide tolerant crops to facilitate agricultural production[158,159].

To explore the potential of **AA** as an herbicide in the field, **AA** first needed to be developed into a spray formulation. Afterwards, we performed spray treatment of *A. thaliana* with **AA** to check if the bioactivity was still retained. There are many commercialized formulations available which enhance the effectiveness of herbicides, such as by increasing adherence of the droplet to the surface of weeds, improving the adsorption of active ingredients into weeds, and facilitating the cuticle penetration of the herbicide to weeds[160-163].

Therefore we added **AA** into a commercial glufosinate formulation known as Finale® at a final **AA** concentration of 250 μM[17]. We then sprayed **AA** solution onto glufosinate resistant *A. thaliana*. Finale® alone had no observable inhibitory effects on plant growth, but adding **AA** severely inhibited plant growth (Fig 39). This result indicated that the potential of **AA** on weeds could be further improved, considering the formulation of Finale® is optimized for glufosinate rather than **AA**. Further optimization of the adjuvant will help **AA** to overcome barriers impeding the movement of herbicide from leaf surface into interior of cells.

In addition, *A. thaliana* plants treated with **AA** before flowering failed to form normal pollen, which was also observed previously[151]. We found that the pistil of treated plants could still be

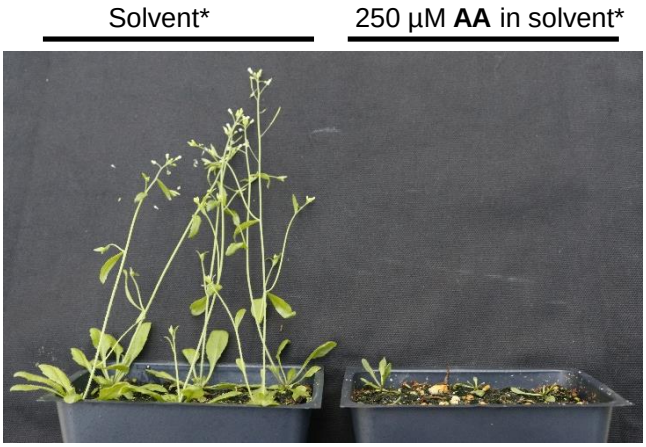



*Figure 39. Spray assay of **AA** on A. thaliana.*

*Glufosinate resistant A. thaliana was treated with (right) or without (left) **AA** in the solvent, which is a commercial glufosinate based herbicide marketed as Finale®. To improve the wetting and penetration, **AA** was firstly dissolved in ethanol and then added to solvent (0.06 g/L Finale® Bayer Inc. + 20 g/L ethanol) to make 250 µM **AA** spraying solution. The control plants were treated with solvent containing ethanol only. Spraying treatments began upon the seeds germination, and were repeated once every two days with approximately 0.4 mL **AA** solution per time per pot for 4 weeks. The picture shown below is taken after one month of treatment. The application rate of **AA** is approximately 1.6 lb/acre, which is comparable to the commonly used herbicide glyphosate (0.75~1.5 lb/acre). The experiments were repeated independently for 3 times with similar results.*

successfully pollinated using healthy pollen from the untreated *A. thaliana*, indicating that **AA** preferentially affects pollen but not egg formation (Fig S10 and S11). This effect was also observed with a lower concentration of **AA** (100 µM). Thus, in addition to its herbicidal properties, **AA** could potentially be used as a chemical hybridization agent for hybrid seed production[164-166].

We next investigated whether plants expressing *astD* are resistant to **AA** (Fig 40). This was motivated by the successful combination of glyphosate and genetically modified crops that are selectively resistant to glyphosate (Roundup Ready®)[167,168]. The *A. terreus astD* gene was codon optimized and the *N*-terminus was fused to a chloroplast localization signal derived from pDHAD[135,169,170]. Wild type or *astD* transgene-expressing *A. thaliana* was then grown on media

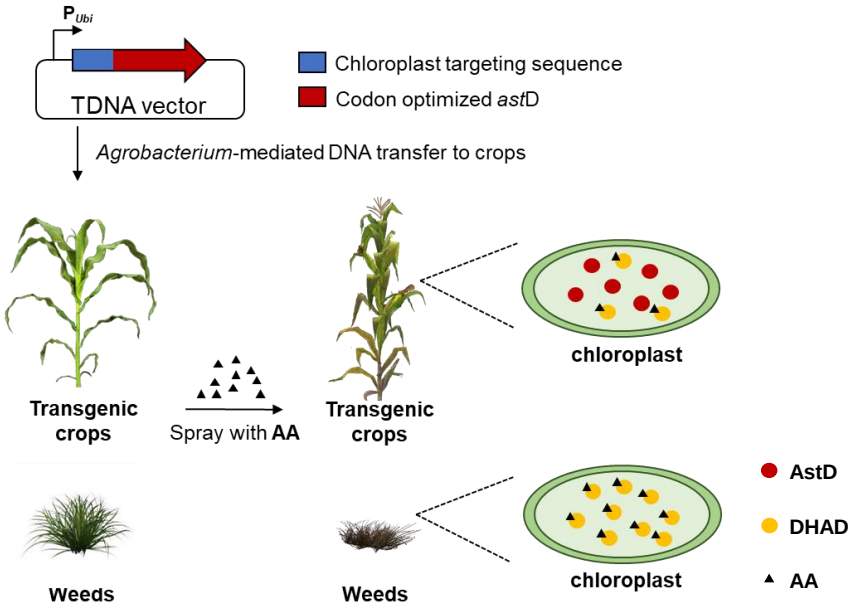
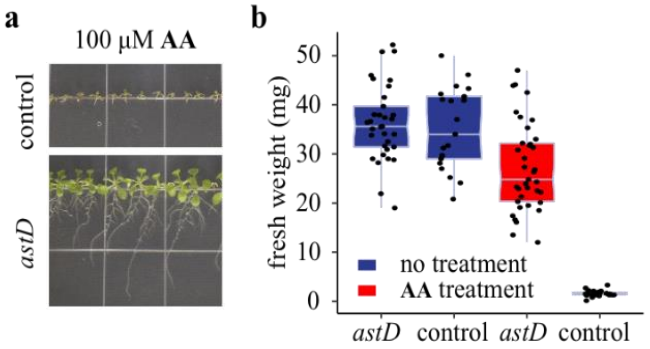



*Figure 40. The strategy to construct **AA** tolerant crops.*

*The resistance gene astD fused with a chloroplast targeting sequence at the N-terminus is integrated onto the genome of crops to confer the crops **AA** tolerance.*

that contained 100 μM **AA**. In the presence of **AA**, the growth of wild-type plants was strongly inhibited, and arrested at the cotyledon stage (Fig. 41a). In contrast, the growth of *astD* transgenic plants was relatively unaffected by **AA**, as indicated by the normally expanded rosette leaves, elongated roots, and whole plant fresh weight (Fig. 41a and b). The expression of AstD was verified by western blot (Fig. S12). A spray assay was also performed using T2 *astD* transgenic *A. thaliana* plants, which showed no observable growth defects under such treatment (Fig. 41c). In contrast, the control plants carrying the empty vector showed a strong growth inhibitory phenotype when treated with **AA** (Fig. 41c). Quantitative measurements of plant height showed AstD effectively confers **AA** resistance to *A. thaliana* (Fig. 41d).

**AA** tolerant *Arabidopsis* (*astD* transgenic) growing on agar




**AA** tolerant *Arabidopsis* (*astD* transgenic) growing in soil

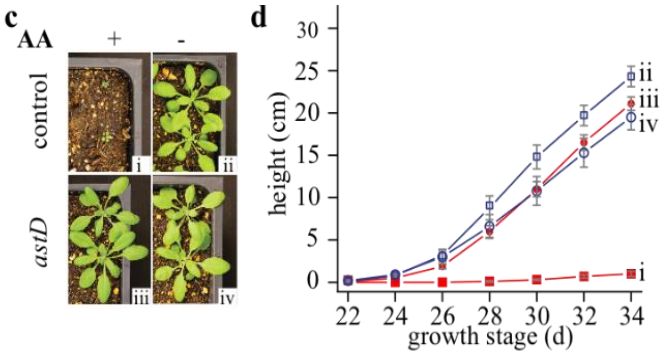



*Figure 41.* ***AA****-resistance of Arabidopsis plants expressing astD transgenes.*

***a,*** *Phenotype of 10-day old A. thaliana with (lower) and without (upper) astD transgene growing on media containing 100 µM* ***AA****. Control plants were transformed with a vector that carries the glufosinate ammonium selection marker but no astD transgene. The picture shown is representative of 3 replicates.* ***b,*** *Fresh weight of 3-week old Arabidopsis seedlings growing on media with (red box) and without (blue box) 100 µM* ***AA****; Box plots show the median and extend of the 1st to 3rd quartile range, with individual data points overlaid; n = 21 biologically independent experiments.* ***c,*** *glufosinate-resistant Arabidopsis with (lower) and without (upper) astD transgene growing in soil were sprayed with glufosinate ammonium with (left) and without (right) 250 µM* ***AA.*** *i. control sprayed with 250 µM* ***AA*** *+ glufosinate ammonium. ii. control sprayed with glufosinate ammonium only. iii. astD transgenic Arabidopsis sprayed with 250 µM* ***AA*** *+ glufosinate ammonium. iv. astD transgenic Arabidopsis sprayed with glufosinate ammonium only. The picture shown is representative of 3 replicates.* ***d,*** *Quantification of the height of Arabidopsis treated the same as in* ***c****; The plot shows mean values ± s.d. (error bars); n = 12 biologically independent experiments.*

### 2.2.3 Conclusions

In this section, an effective herbicide target, DHAD, was selected to perform resistance guided genome mining in order to discover an herbicide with a new mode of action. We first found a conserved four-gene-cassette containing DHAD and terpenoid biosynthetic genes. By activating this cluster in *S. cerevisiae*, we identified the product of this cluster as **AA**. Further biochemical analysis indicated **AA** is indeed a potent inhibitor of DHAD. Excitingly, this result not only demonstrated the feasibility of our proposal, but also enabled the identification of NPs with a new mode of action.

To further prove **AA** can be applied as an herbicide to kill weeds in the field, we demonstrated that **AA** can be developed into a spray formulation to effectively inhibit the growth of *Arabidopsis*. We also validated that introducing the resistance gene *astD* as a transgene or editing the sequence of the plant DHAD endogenous gene could be used to create **AA**-resistant crops. Thus, we suggest that **AA** is a promising lead for development as a broad-spectrum commercial herbicide.

### 2.2.4 Materials and Methods

#### General materials and methods

Biological reagents, chemicals, media and enzymes were purchased from standard commercial sources unless stated. Plant, fungal, yeast and bacterial strains, plasmids and primers used in this study are summarized in Tables S2 and S3. DNA and RNA manipulations were carried out using Zymo ZR Fungal/Bacterial DNA Microprep™ kit and Invitrogen Ribopure™ kit respectively. DNA sequencing was performed at Laragen, Inc. The primers and codon optimized gblocks were synthesized by IDT, Inc.

#### Plasmid and strain construction

#### Expression of *ast* genes in *Aspergillus nidulans* for cDNA isolation.

Plasmids pYTU, pYTP, pYTR digested with *Pac*I and *Swa*I were used as vectors to insert genes[171]. A *gpda* promoter was generated by PCR amplification using primers Gpda-pYTU-F and Gpda-R with pYTR serving as template. Genes to be expressed were amplified through PCR using the genomic DNA of *Aspergillus terreus* NIH2624 as a template. A 4.5 kb fragment obtained using primers AstD-pYTU-recomb-F and AstA-pYTU-recomb-R was cloned into pYTU together with a *gpda* promoter by yeast homologous recombination to obtain pAstD+AstA-pYTU. Yeast transformation was performed using Frozen-EZ Yeast Transformation II Kit™ (Zymo Research). A 2.4 kb fragment obtained using primers AstB-pYTR-recomb-F and AstB-pYTR-recomb-R was cloned into pYTR by yeast homologous recombination to obtain pAstB-pYTR. Similarly, a 2.3 kb fragment obtained using primers AstC-pYTP-recomb-F and AstC-pYTP-recomb-R was cloned into pYTP by yeast homologous recombination to obtain pAstC-pYTP.

All three plasmids (pAstD+AstA-pYTU, pAstB-pYTR and pAstC-pYTP) were transformed into *A. nidulans* following standard protocols to result in the *A. nidulans* strain TY01[171]. TY01 was cultured in liquid CD-ST medium (20 g/L starch, 20 g/L peptone, 50 mL/L nitrate salts and 1 mL/L trace elements) at 28°C for 3 days. Total RNA of TY01 was extracted with the Invitrogen Ribopure™ kit, and total cDNA of TY01 was obtained using the SuperScript III reverse transcriptase kit (Thermo Fisher Scientific). The cDNA fragment of *astA* was PCR amplified using primers AstA-xw55-recomb-F and AstA-xw55-recomb-R. The cDNA fragment of *astB* was PCR amplified using primers AstB-xw06-recomb-F and AstB-xw06-recomb-R. The cDNA fragment of *astC* was PCR amplified using primers AstC-xw02-recomb-F and AstC-xw02-recomb-R. The cDNA fragment of *astD* was PCR amplified using primers AstD-pXP318-F and AstD-pXP318-R. All the introns were confirmed to be correctly removed by sequencing.

**Construction of *Saccharomyces cerevisiae* strains.**

Plasmid pXW55 (*URA3* marker) digested with *Nde*I and *Pme*I was used to introduce the *astA* gene[124]. A 1.3 kb fragment containing *astA* obtained from PCR using primers AstA-xw55-recomb-F and AstA-xw55-recomb-R was cloned into pXW55 using yeast homologous recombination to afford pAstA-xw55. The plasmid pAstA-xw55 was then transformed into *Saccharomyces cerevisiae* RC01 to generate strain TY02[124].

Plasmid pXW06 (*TRP1* marker) digested with *Nde*I and *Pme*I was used to introduce the *astB* gene[124]. A 1.6 kb fragment containing *astB* obtained from PCR using primers AstB-xw06-recomb-F and AstB-xw06-recomb-R were cloned into pXW06 using yeast homologous recombination to afford pAstB-xw06. The plasmid pAstB-xw06 was then transformed into TY02 to generate strain TY03.

Plasmid pXW02 (*LEU2* marker) digested with *Nde*I and *Pme*I was used to introduce the *astC* gene[124]. A 1.6 kb fragment containing *astC* obtained from PCR using primers AstC-xw02-recomb-F and AstC-xw02-recomb-R were cloned into pXW02 using yeast homologous recombination to afford pAstC-xw02. The plasmid pAstC-xw02 was then transformed into TY03 to generate strain TY04.

*URA3* gene was inserted into *ilv3* locus of *Saccharomyces cerevisiae* DHY Δ*URA3* strain to generate UB01. A 879 bp homologous recombination donor fragment with 35-40 bp homologous regions flanking *ilv3* ORF was amplified using primers ILV3p-URA3-F and ILV3t-URA3-R using yeast gDNA as template. The PCR product was gel purified and transformed into *Saccharomyces cerevisiae* DHY Δ*URA3,* and selected on uracil dropout media to give UB01. The resulting strain was subjected to verification by colony PCR with primers ILV3KO-ck-F and ILV3KO-ck-R and the amplified fragment was sequence confirmed.

The *URA3* gene inserted into *ilv3* locus of *Saccharomyces 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 *ilv3* ORF was amplified using primers ILV3KO-F and ILV3KO-R, gel purified and transformed into UB01, and counterselected on 5-fluoroorotic acid (5-FoA) containing media to give UB02. The resulting strain was subjected to verification by colony PCR with primers ILV3KO-ck-F and ILV3KO-ck-R and the amplified fragment was sequenced confirmed.

The empty plasmid pXP318 (*URA3* marker) was transformed into UB02 to generate TY05[172].

Plasmid pXP318 digested with *Spe*I and *Xho*I was used as vector to introduce gene encoding fDHAD[172]. The cDNA of *Aspergillus terreus* NIH 2624 served as template for PCR amplification.

A 1.7 kb fragment obtained using primers fDHAD-pXP318-F and fDHAD-pXP318-R were cloned into pXP318 using yeast homologous recombination to afford fDHAD-pXP318. Then, fDHAD-pXP318 was transformed into UB02 to generate TY06. fDHAD was driven by a constitutive promoter *TEF1*.

Plasmid pXP318 digested with *Spe*I and *Xho*I was used as vector to introduce *astD* gene[172]. The cDNA isolated from TY01 served as the template for PCR amplification. A 1.8 kb fragment obtained using primers AstD-pXP318-F and AstD-pXP318-R were cloned into pXP318 using yeast homologous recombination to give AstD-pXP318. A FLAG-tag was also add to the *N*-terminal of AstD. Then, AstD-pXP318 was transformed into UB02 to generate TY07. AstD was driven by a constitutive promoter *TEF1*.

**Fermentation and compound analyses and isolation**

A seed culture of *S. cerevisiae* strain was grown in 40 mL of synthetic dropout medium for 2 d at 28°C, 250 rpm. Fermentation of the yeast was carried out using YPD (yeast extract 10 g/L, peptone 20 g/L) supplement with 2% dextrose for 3 d at 28°C, 250 rpm.

HPLC-MS analyses were performed using a Shimadzu 2020 EVLC-MS (Phenomenex® Luna, 5μ, 2.0 × 100 mm, C-18 column) using positive and negative mode electrospray ionization. The elution method was a linear gradient of 5-95% (v/v) acetonitrile/water in 15 min, followed by 95% (v/v) acetonitrile/water for 3 min with a flow rate of 0.3 mL/min. The HPLC buffers were supplemented with 0.05% formic acid (v/v). HPLC purifications were performed using a Shimadzu Prominence HPLC (Phenomenex® Kinetex, 5μ, 10.0 × 250 mm, C-18 column). The elution method was a linear gradient of 65-100% (v/v) acetonitrile/water in 25 min, with a flow rate of 2.5 mL/min. GC-MS analyses were performed using Agilent Technologies GC-MS

6890/5973 equipped with a DB-FFAP column. An inlet temperature of 240°C and constant pressure of 4.2 psi were used. The oven temperature was initially at 60°C and then ramped at 10°C/min for 20 min, followed by a hold at 240°C for 5 min.

To isolate compound **8**, the fermentation broth of TY02 was centrifuged (5000 rpm, 10 min), and cell pellet was harvested and soaked in acetone. The organic phase was dried over sodium sulfate, concentrated to oil form, and subjected to silica column purification with hexane. To isolate compound **9**, the fermentation broth of TY03 was centrifuged (5000 rpm, 10 min), and supernatant was extracted three times with ethyl acetate. The organic phase was dried over sodium sulfate, concentrated to oil form, and then and subjected to HPLC purification. To isolate compound **AA**, the fermentation broth of TY04 was centrifuged (5000 rpm, 10 min), and supernatant was extracted three times with ethyl acetate. The organic phase was dried over sodium sulfate, concentrated to oil form, and subjected to HPLC purification.

**Protein expression, purification and biochemical assay**

To express and purify pDHAD, primers pDHAD-pET-F and pDHAD-pET-R were used to amplify a 1.7 kb DNA fragment containing *pdhad* (AT3G23940). The PCR product was cloned into pET28a using *Nhe*I and *Not*I restriction sites. The resulting plasmid pDHAD-pET was transformed into *E.coli* BL21 (DE3) to give TY08. To express and purify fDHAD (XP_001208445.1), primers fDHAD-pET-F and fDHAD-pET-R were used to amplify a 1.6 kb DNA fragment containing *fdhad*. The PCR product was cloned into pET28a using *Nde*I and *Not*I restriction sites. The resulted plasmid fDHAD-pET was transformed into *E. coli* BL21 (DE3) to obtain TY09. To express and purify AstD (XP_001213593.1), primers AstD-pET-F and AstD-pET-R were used to amplify a 1.6 kb DNA fragment containing *astD*. The PCR product was cloned into pET28a using *Nde*I and *Not*I restriction sites. The resulted plasmid AstD-pET was

transformed into *E. coli* BL21 (DE3) to obtain TY10. All DHADs fused a 6×His-tag with a molecular weight ~62 kD were expressed at 16°C 220 rpm for 20 h after 100 μM IPTG induction (IPTG was added when $OD_{600}$ = 0.8). Cells of 1 L culture were then harvested by centrifugation at 5000 rpm at 4°C. Cell pellet was resuspended in 15 mL Buffer A10 (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 8% glycerol, 10 mM imidazole). The cells were lysed by sonication, and the insoluble material was sedimented by centrifugation at 16000 rpm at 4°C. The protein supernatant was then incubated with 3 mL Ni-NTA for 4 h with slow, constant rotation at 4°C. Subsequently the Ni-NTA resin was washed with 10 column volumes of Buffer A50 (Buffer A + 50 mM imidazole). For elution of the target protein, the Ni-NTA resin was incubated for 10 min with 6 mL Buffer A300 (Buffer A + 300 mM imidazole). The supernatant from the elution step was then analyzed by SDS-PAGE together with the supernatants from the other purification steps. The elution fraction containing the recombinant protein was buffer exchanged into storage buffer (50 mM Tris-HCl pH 7.2, 50 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, 5 mM DTT, 5 mM GSH).

*In vitro* activity assays were carried out in 50 μL reaction mixture containing storage buffer, 10 mM (±)-sodium α,β-dihydroxyisovalerate hydrate (**4**) and 0.5 μM of purified DHAD enzyme. The reaction was initiated by adding the enzyme. After 0.5 h incubation at 30˚C, the reactions were stopped by adding equal volume of ethanol. Approximately 0.1 volume of 100 mM phenylhydrazine (PHH) was added to derivatize the product 3-methyl-2-oxo-butanoic acid (**5**) into **6** at room temperature for 30 min. 20 μL of the reaction mixture was subject to LC-MS analysis. The area of the HPLC peak with UV absorption at 350 nm were used to quantify the amount of **6**. (Fig. 33).

The inhibition percentage of **AA** on DHADs determined using *in vitro* biochemical assays are calculated by following equation:

$$\text{inhibition percentage} = 1 - \frac{\text{initial reaction rate with } \mathbf{AA}}{\text{initial reaction rate without } \mathbf{AA}}$$

**Growth inhibition assay of *S. cerevisiae* on plates or in the tubes.**

*S. cerevisiae* was grown in isoleucine, leucine and valine (ILV) dropout media (20 g/L glucose, 0.67 g/L Difco$^{TM}$ Yeast Nitrogen Base w/o amino acids, 18 mg/L adenine, 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 76 mg/L, tyrosine 76 mg/L) to test growth inhibition of **AA** on *S. cerevisiae*. *S. cerevisiae* was incubated at 28˚C until $OD_{600}$ of the control strain without **AA** treatment reached about 0.8. The ratio of yeast $OD_{600}$ in media with **AA** treatment to yeast $OD_{600}$ in media without **AA** was calculated as the percentage of growth inhibition. The inhibition curve was plotted as percentage of inhibition versus **AA** concentrations. To further prove **AA** affects BCAA biosynthesis, isoleucine, leucine and valine was also complemented to the media with or without treatment of **AA**. The growth curves of TY05, TY06 and TY07 were also plotted in Fig. S5. The $OD_{600}$ was recorded for every 20 min over a total of 50 h. Percent inhibition. The growth inhibition percentage of **AA** on *S. cerevisiae* strain is calculated by dividing the cell density ($OD_{600}$) of the **AA**-treated strain to the corresponding untreated strains when $OD_{600}$ reaches ~ 0.8 using following equation:

$$\text{growth inhibition percentage} = 1 - \frac{\text{OD}_{600} \text{ of } \mathbf{AA} \text{ treated strain}}{0.8}$$

in which 0.8 is the $OD_{600}$ of untreated strain.

**Growth inhibition assay of plants on plates or in the tubes.**

MS (2.16 g/L Murashige and Skoog basal medium, 8 g/L sucrose, 8 g/L agar) media was used to test the growth inhibition of **AA** on *A. thaliana*, *Solanum lycopersicum*, and *Zea mays*. *A. thaliana, S. lycopersicum*, and *Z. mays* were grown under long day condition (16/8 h light/dark) using cool-white fluorescence bulbs as the light resource at 23˚C. **AA** was dissolved in ethanol and added to the media before inoculating strains or growing plants. The media of control treatment contains the same amount of ethanol, but without **AA**.

**Plant growth inhibition assay by spraying**

**AA** was firstly dissolved in ethanol and then added to solvent (0.06 g/L Finale® Bayer Inc. + 20 g/L EtOH). The control plants were treated with solvent containing ethanol only. *A. thaliana* that are resistant to glufosinate (containing the *bar* gene) were grown under long day condition (16/8 h light/dark) using cool-white fluorescence bulbs as the light resource at 23˚C. Spraying treatments began upon the seed germination, and was repeated once every two days with approximately 0.4 mL **AA** solution per time per pot.

**Structure determination of *holo*-pDHAD**

The gene encoding pDHAD (residues 35–608) was cloned into pET21a derivative vector pSJ2 with an eight histidine (8×His) tag and a TEV protease cleavage site at the *N*-terminus. The following primers were used for cloning: the forward primer DHAD-F and the reverse primer DHAD-R. The double mutant K559A/K560A for efficient crystallization was designed using the surface entropy reduction prediction (SERp) server[173]. Mutations were generated by PCR using the forward primer K559AK560A-F and reverse primer K559AK560A-R. All constructed plasmids were verified by DNA sequencing.

pDHAD purified under aerobic conditions was found to contain no iron-sulfur cluster (*apo* form). Hence we performed [2Fe-2S] Cluster reconstitution under the atmosphere of nitrogen in an anaerobic box. The protein was incubated with $FeCl_3$ at the ratio of 1:10 for 1 h on ice and then

10 equivalents of $Na_2S$ per protein was added drop-wise every 30 min for 3 h. The reaction mixture was then incubated overnight. Excess $FeCl_3$ and $Na_2S$ were removed using a SephadexTM G-25 Fine column (GE Healthcare)[155].

The reconstituted *holo*-pDHAD was crystallized in an anaerobic box. The proteins (at 10 mg/mL) were mixed in a 1:1 ratio with the reservoir solution in a 50 μL volume of 2 μL and equilibrated against the reservoir solution, using the sitting-drop vapor diffusion method at 16°C. Crystals for diffraction were observed in 0.1 M sodium acetate pH 5.0, 1.5 M ammonium sulfate after 5 d.

All crystals were flash-cooled in liquid nitrogen after cryo-protected with solution containing 25% glycerol, 1.5 M ammonium sulfate, 0.1 M sodium acetate pH 5.0. The data were collected at 100K and at the Beam Line 19U1 in Shanghai Synchrotron Radiation Facility (SSRF). Diffraction data of *holo*-pDHAD was collected at the wavelength of 0.97774 Å. The best crystals diffracted to a resolution of 2.11 Å. The Ramachandran plot favored (%), allowed (%) and outlier (%) are 98.05, 1.60, and 0.36 respectively. All data sets were indexed, integrated, and scaled using the HKL3000 package[174]. The crystals belonged to space group $P4_22_12$. The statistics of the data collection are summarized in Table S6.

The *holo*-pDHAD structure was solved by the molecular replacement method Phaser embedded in the CCP4i suite and the L-arabinonate dehydratase crystal structure (PDB_ID: 5J83) as the search model. All the side chains were removed during the molecular replacement process[175,176]. The resulting model were refined against the diffraction data using the REFMAC5 program of CCP4i[177]. Based on the improved electron density, the side chains of *holo*-pDHAD protein, iron sulfur cluster, water molecule, acetate ion, sulfate ions, and magnesium ion were manually built using the program WinCoot[178]. The $R_{work}$ and $R_{free}$ values of the structure are 17.67%

and 22.15%, respectively. The detailed refinement statistics are summarized in Table S6. The geometry of the model was validated by WinCoot. Structural factor and coordinate of *holo*-pDHAD have been deposited in the Protein Bank (PDB code: 5ZE4).

**Homology modelling of AstD and docking of substrate or AA into active site of *holo*-pDHAD**

The structure of *holo*-pDHAD was prepared in Schrodinger suite software under OPLS3 force field[179]. Hydrogen atoms were added to reconstituted crystal structures according to the physiological pH (7.0) with the PROPKA tool in Protein Preparation tool in Maestro to optimize the hydrogen bond network[155,180]. Constrained energy minimizations were conducted on the full-atomic models, with heavy atom coverage to 0.5 Å. The homology model was performed in Modeller 9.18[181], using the crystal structure of *holo*-pDHAD solved in this work as a template. Sequence alignment in Modeller indicated that AstD and pDHAD shared 56.8% sequence identity and 75.0% sequence similarity (Fig. S9). All the highly conserved residues and motifs were properly aligned. A total of 2000 models were generated for each target in Modeller with the fully annealed protocol. The optimal models were chosen for docking studies according to DOPE (Discrete Optimized Protein Energy) score.

All ligand structures were built in Schrodinger Maestro software[155]. The LigPrep module in Schrodinger software was introduced for geometric optimization by using OPLS3 force field[179]. The ionization state of ligands were calculated with Epik tool employing Hammett and Taft methods in conjunction with ionization and tautomerization tools[182]. The docking of a ligand to the receptor was performed using Glide[183]. We included cofactors observed in crystal structure during the docking. Since both water and $SO_4^{2-}$ occupied the catalytic site, they were excluded prior to docking. Cubic boxes centered on the ligand mass center with a radius 8 Å for all ligands defined the docking binding regions. Flexible ligand docking was executed for all structures. Ten

poses per ligand out of 20,000 were included in the post-docking energy minimization. The best scored pose for the ligand was chosen as the initial structure for further study. The MM/GBSA method was introduced to evaluate the ligand binding affinity based on the best scored docking pose in Schrodinger software. Figures are prepared in PyMOL and Inkscape[184,185]. Both of native substrate α,β-dihydroxyisovalerate and **AA** were docked into the catalytic site of pDHAD. The cross-section electrostatic surface map shows this unique catalytic pocket has a positively charged internal and a hydrophobic entrance, which binds to negatively charged "head" and hydrophobic "tail" of substrate or **AA** respectively. Thus the negatively charged "head" can lead both of the substrate and **AA** into the catalytic chamber. The bulky hydrophobic tricyclic moiety of **AA**, however, provides stronger hydrophobic interactions to the entrance and blocks the entrance of active site due to the hydrophobic residues at the entrance, including G68, A71, I72, I134, A133, M141, V212, F215, M498 and P501. In contrast, the smaller "tail" of native substrate provides less interactions to entrance because the smaller size limits efficient hydrophobic contact to nearby residues. This implies that once **AA** binds to pDHAD, it can prevent substrate approaching the active site. We also introduced molecular mechanics generalized Born and surface area (MM/GBSA) continuum solvation method, an widely used approach for relative binding energy calculation, to evaluate the relative binding affinity for both ligands[186]. The MM/GBSA calculations had been done in Prime[187] (Schrödinger 2015 suite). The MM/GBSA energy was calculated using following equation:

$$\Delta G_{bind} = E_{complex} - E_{protein} - E_{ligand}$$

*E* denotes energy and includes terms such as protein−ligand van der Waals contacts, electrostatic interactions, ligand desolvation, and internal strain (ligand and protein) energies, using VSGB2.0 implicit solvent model with the OPLS2005 force field. The solvent entropy is also included in the

VSGB2.0 energy model, as it is for other Generalized Born (GB) and Poison−Boltzmann (PB) continuum solvent models.

MM/GBSA calculation shows that the relative binding energy for **AA** and α,β-dihydroxyisovalerate is -18.6 ± 0.3 kcal/mol and -13.3 ± 0.2 kcal/mol respectively, which shows the binding constant of **AA** to active site is about 6000 times greater than α,β-dihydroxyisovalerate. This further confirms that **AA** is a competitive inhibitor of pDHAD.

**Cytotoxicity assay of AA**

Cell proliferation experiments were performed in a 96-well format (five replicates per sample) using melanoma cell line A375 and SK-MEL-1. **AA** treatments were initiated 24 h postseeding for 72 h, and cell survival was quantified using CellTiter-GLO assay (Promega).

**Cross experiment of *A. thaliana***

To make male sterile *A. thaliana*, **AA** was added to chemical hybridization agent (CHA) formulation (250 μM **AA**, 2% ethanol, 0.1% Tween-80, 1% corn oil in water), which has less inhibition effect on the growth of *A. thaliana*. Flowers of the **AA** treated col-0 were selected as the female parent. The non-treated *A. thaliana* containing a glufosinate resistant gene were used as male parent to donate pollen. 2-week old F1 progeny resulting from the cross were treated by Finale (11.3% glufosinate-ammonium) at 1:2000 dilution. The results are summarized in Fig. S10.

**Construction of the transgenic plants**

The coding sequence of AstD was codon optimized for *A. thaliana*. A chloroplast localization signal (CLS) of 35-amino acid residues derived from the *N*-terminal of *A. thaliana* DHAD (MQATIFSPRATLFPCKPLLPSHNVNSRRPSIISCS) was fused to N-terminus of the codon optimized AstD. A 3×FLAG-tag was inserted between the CLS and the codon optimized AstD

(Supplementary results). The gene block containing CLS, FLAG-tag and *astD* was synthesized and then cloned into pEG202 vector using Gateway LR Clonase II Enzyme Mix (ThermoFisher scientific). The original CaMV 35S promoter of pEG202 was substituted by Ubiquitin-10 promoter to drive the expression of AstD. The construct was electro-transformed into *Agrobacterium tumefaciens* strain Agl0 followed by *A. thaliana* transformation using the standard floral dip method[188]. The *A. thaliana* Col-0 ecotype was transformed. Positive transgenic plants were selected using the glufosinate resistance marker, and were tested for survival in presence of **AA**.

**Protein expression verification with western blot**

Approximately 0.5 gram of leaf tissue of transgenic *A. thaliana* was grounded in liquid nitrogen. Proteins were homogenized in 2× SDS buffer followed by 5-min centrifuge at 21,000 g to remove undissolved debris. The supernatant containing resolved proteins were loaded onto a 4-12% Bis-Tris gel, and separated using MOPS running buffer. Transfer was conducted using iBlot2 dry transfer device and PVDF membrane. The total proteins were stained with Ponceau to demonstrate equal loading. Western blotting was performed using Sigma monoclonal anti-FLAG M2-Peroxidase antibody, followed by detection using Amersham ECL Prime detection reagent.

## 3. Conclusions

NPs are naturally occurring small molecule organic compounds produced by living organisms. These compounds have coevolved together with biomacromolecular targets in nature, and possess potent bioactivities and selectivity to kill or limit the growth of competitors through the inhibition of vital metabolic enzymes. NPs are structurally complex, includng multiple stereocenters, complex ring systems and rich assemblies of functional groups that serve as warheads, hydrogen bond donors/acceptors and target recognition moieties. Because of their potent inhibitory activities, NPs have been used to benefit humans as both pharmaceuticals and agrochemicals.

The post-genomics era has brought a renaissance in natural product discovery through genomic sequence directed approaches. This allows us to tap into cryptic NP biosynthetic gene clusters, which are not turned on under standard laboratory growing conditions. However, the bioactivities of NPs discovered using this approach are usually elusive. In this thesis, we proposed a resistance gene directed approach to discover novel NPs. The rationale for this approach is that the host organism producing NPs targeting housekeeping enzymes must have a self-protection mechanism, which includes the use of a self-resistance enzyme, a mutated version of housekeeping enzyme which can carry out the same function. We first successfully located the BGC of heptelidic acid using a self-resistance GAPDH, meanwhile, we also verified that the SRE encoded in the BGC of heptelidic acid has more tolerance than the housekeeping GAPDH. These results indicated that resistance gene guided approach is feasible to predict the bioactivity of a NP based on genomic sequence.

To prove that our approach is not just limited to locating the gene cluster of a NP with a known bioactivity, we carried out the resistance-gene directed NP discovery approach using not only popular human therapeutic targets, but also potential pesticide targets. To our delight, when

an herbicide target dihydroxy acid dehydratase (DHAD) was used as a query, a conserved BGC was located among several fungal species. We targeted the third enzyme, DHAD, within the branched chain amino acid (BCAA) biosynthesis pathway, which is conserved in fungi, bacteria, and plants, but is absent in humans and animals. The first enzyme in this pathway, acetolactate synthase (ALS) is the most targeted enzyme for herbicide development with over 50 commercialized compounds. On the other hand, there is no inhibitor of DHAD that has been reported to work in planta.

Using the target-guided genome mining approach, we scanned fungal genomes in publicly available databases for a BGC that encodes a possible resisting copy of DHAD. The cluster from *Aspergillus terreus* was successfully introduced into *Saccharomyces cerevisiae*, which led to the production of aspterric acid at 20 mg/L. Although this compound was isolated previously, its molecular target was not known. We tested the inhibitory properties of this compound on three different DHADs. As we hypothesized, aspterric acid is a potent competitive inhibitor of the housekeeping DHAD enzyme from *A. terreus* and the plant DHAD from *Arabidopsis thaliana*. In contrast, the predicted self-resistance DHAD (AstD) was completely insensitive to aspterric acid up to the solubility limit. Kinetic analysis showed that AstD is about ~20 fold slower than the housekeeping DHADs, indicating that during evolution, nature has sacrificed activity for resistance to the NP. Structural analysis of both housekeeping DHAD and AstD revealed the molecular basis of aspterric acid inhibition, as well as possible mutations that led to AstD insensitivity. Aspterric acid represents the first known NP inhibitor of DHAD, and is discovered using the resistance-gene directed genome mining approach.

Although an increasing number of weeds have evolved resistance to current herbicides, there hasn't been an herbicide with new mode of action commercialized for more than 30 years. Our

discovery could thus lead to the first new class of commercial herbicides in more than 30 years. We next addressed the question of how to make this broad-spectrum herbicide useful in specifically killing weed but not crops. Inspired by the successful of combination of "Roundup" and "Roundup-Ready crops", which was introduced into the market by Monsanto, we have considered the possibility of developing **AA** tolerant crops. Fortunately, the naturally occurring **AA** resistance gene within its BGC make this a straightforward foal. By introducing the resistance gene into the model plant *Arabidopsis*, we successfully demonstrated that the resistance gene is able to confer **AA** tolerance to plants, which may pave the way to developing **AA** tolerant crops in the future.

As it was said by Charles Dickens in his great novel *A Tale of Two Cities*, "It is the best of times, it is the worst of times"[189]. This exactly described the situation of NP discovery. Although we are still unsure whether there are enough bioactive NPs to meet the demand of continuously emerging resistance, we indeed have an increasing amount of synthetic biology tools, genomic sequencing data and knowledge of secondary metabolism to discover NPs possessing novel bioactivities. NP discovery is like mining diamonds in the rough. Ironically, our novel fungal treasure had been ignored for 40 years since it was first isolated. Likewise, Mother Nature has already come up with a plethora of NPs to solve our current problems. The question is whether or not we can find them.

## 4. Appendices

### 4.1 Supplementary results

**Structure determination of compounds**

Compound **2**, colorless oil readily dissolved in ethyl acetate and chloroform, had a molecular formula $C_{15}H_{22}O_3$, as deduced from LC-MS $[M+H]^+$ *m/z* 251, $[M-H]^-$ *m/z* 249. $^1$H NMR (500 MHz, $CDCl_3$): $\delta$ 7.14 (1H, s), 4.88 (2H, d, 35.3), 4.09 (1H, q, 9.3), 2.92 (1H, dd, 17.4, 5.8), 2.41 (1H, d, 12.0), 2.25~2.15 (2H, m), 2.10~1.85 (3H, m), 1.87 (1H, d, 12.5), 1.44 (1H, m), 1.20 (1H, m), 0.94 (3H, d, 6.8), 0.77 (3H, d, 6.9). $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 171.9, 149.5, 141.1, 127.6, 104.8, 66.7, 51.2, 46.4, 45.9, 36.9, 32.3, 27.3, 26.8, 21.3, 15.1.

Compound **3**, colorless oil readily dissolved in ethyl acetate and chloroform, had a molecular formula $C_{15}H_{22}O_4$, as deduced from LC-MS $[M+H]^+$ *m/z* 267, $[M-H]^-$ *m/z* 265. $^1$H NMR (500 MHz, $CDCl_3$): $\delta$ 7.10 (1H, s), 3.75 (1H, m), 3.21 (1H, s), 2.79 (1H, dd, 15.9, 5.4), 2.70 (1H, d, 3.4), 2.23 (1H, td, 6.9, 3.0), 2.18~1.97 (3H, m), 1.95~1.85 (2H, m), 1.43-1.23 (3H, m), 0.97 (3H, d, 6.9), 0.81 (3H, d, 6.9). $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 171.6, 139.5, 128.2, 68.0, 64.6, 49.8, 45.9, 45.5, 42.9, 34.8, 32.6, 26.7, 23.9, 21.3, 15.1.

Compound **4**, colorless oil readily dissolved in ethyl acetate and chloroform, had a molecular formula $C_{15}H_{20}O_4$, as deduced from LC-MS $[M+H]^+$ *m/z* 265, $[M-H]^-$ *m/z* 263. $^1$H NMR (500 MHz, $CDCl_3$): $\delta$ 7.31 (1H, d, 2.0), 3.89 (1H, dd, 5.5, 1.5), 3.27 (1H, dt, 20.2, 2.6), 2.96 (1H, dt, 20.1, 2.2), 2.89 (1H, d, 12.0), 2.57~2.52 (1H, m), 2.51 (1H, d, 5.7), 2.34~2.22 (1H, m), 1.95~1.80 (2H, m), 1.51 (1H, tt, 11.7, 3.1), 1.36~1.23 (2H, m), 1.03 (3H, d, 6.9), 0.83 (3H, d, 6.9). $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 204.3, 170.1, 141.9, 127.9, 58.1, 51.4, 51.1, 46.1, 44.2, 41.1, 35.5, 26.9, 23.0, 21.5, 15.2.

Compound **8**, colorless oil readily dissolved in hexane and chloroform, had a molecular formula $C_{15}H_{24}$, as deduced from EI-MS $[M]^+$ *m/z* 204, and showed $[\alpha]_D^{22} = -30°$ (*n*-hexane; *c* = 0.1). GC-MS 70 eV, *m/z* (relative intensity): 204 $[M]^+$ (42), 189 (5), 161 (35), 136 (100), 133 (10), 121 (70), 119 (25), 107 (20), 105 (27), 93 (21), 91 (26), 79 (13), 77 (15), 69 (20), 55 (12), 43 (12), 41 (13), 38 (21); $^1$H NMR (500 MHz, $CDCl_3$): $\delta$ 5.37 (1H, m), 2.20-2.10 (5H, m), 2.10-2.00 (2H, m), 1.95 (1H, d, 15.3), 1.75 (3H, s), 1.71 (3H, q, 1.7), 1.61 (3H, brs), 1.44 (1H, dd, 11.4, 7.2), 1.36 (1H, m), 1.31 (1H, dd, 11.3, 2.6), 0.73 (3H, s); $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 138.4, 138.3, 122.4, 122.2, 57.4, 42.6, 41.4, 40.3, 34.5, 29.6, 27.3, 25.0, 23.3, 20.6, 19.2. Both of the NMR and MS spectrums are identical to a known compound (+)-daucane, however, the optical rotation is opposite which led to the assignment of **8** to be (-)-daucane.

Compound **9**, colorless oil readily dissolved in ethyl acetate and chloroform, had a molecular formula $C_{15}H_{22}O_3$, as deduced from LC-MS $[M+H]^+$ *m/z* 251, $[M-H]^-$ *m/z* 249. $^1$H NMR (500 MHz, $CDCl_3$): $\delta$ 8.09 (1H, brs), 3.25 (1H, t, 7.4), 2.71 (1H, dd, 14.6, 6.5), 2.48 (1H, dd, 14.8, 6.3), 2.36 (1H, dd, 14.0, 6.6), 2.26 (1H, m), 2.15 (1H, dd, 16.3, 8.9), 2.08 (1H, d, 12.0), 1.84 (1H, q, 13.1), 1.73 (3H, d, 2.3), 1.59 (3H, d, 2.2), 1.48~1.35 (3H, m), 1.31 (1H, td, 11.5, 9.0), 0.86 (3H, s). $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 176.0, 135.8, 123.2, 60.1, 59.8, 59.4, 44.1, 40.5, 38.8, 30.6, 29.3, 24.9, 23.8, 20.6, 17.8.

Compound **10** is a colorless oil readily dissolved in acetone and chloroform, had a molecular formula $C_{15}H_{22}O_4$, as deduced from LC-MS $[M+H]^+$ *m/z* 267, $[M-H]^-$ *m/z* 265. $^1$H NMR (500 MHz, $CDCl_3$): $\delta$ 4.29 (1H, d, 8.5), 3.92 (1H, d, 8.3), 3.48 (1H, d, 8.3), 2.42 (1H, dd, 14.9, 7.3), 2.37~2.28 (2H, m), 2.25 (1H, dd, 13.0, 4.4), 2.20~2.17 (1H, m), 2.12 (1H, d, 13.4), 2.01 (1H, m), 1.80~1.65 (2H, m), 1.71 (3H, s), 1.64~1.54 (1H,m), 1.60 (3H, s), 1.50 (1H, m); $^{13}$C NMR (125 MHz, $CDCl_3$):

$\delta$ 178.2, 134.5, 125.2, 82.9, 76.3, 75.6, 55.4, 53.0, 36.6, 36.2, 33.8, 32.2, 23.6, 23.4, 20.9. **10** is identical to aspterric acid (**AA**) as reported.

**The codon-optimized sequence of *Homo sapiens gapdh* for expression in *E. coli BL21 (DE3)*.** The restriction digestion site of *Eco*RI and *Xho*I was underlined at 5' end and 3' end respectively.

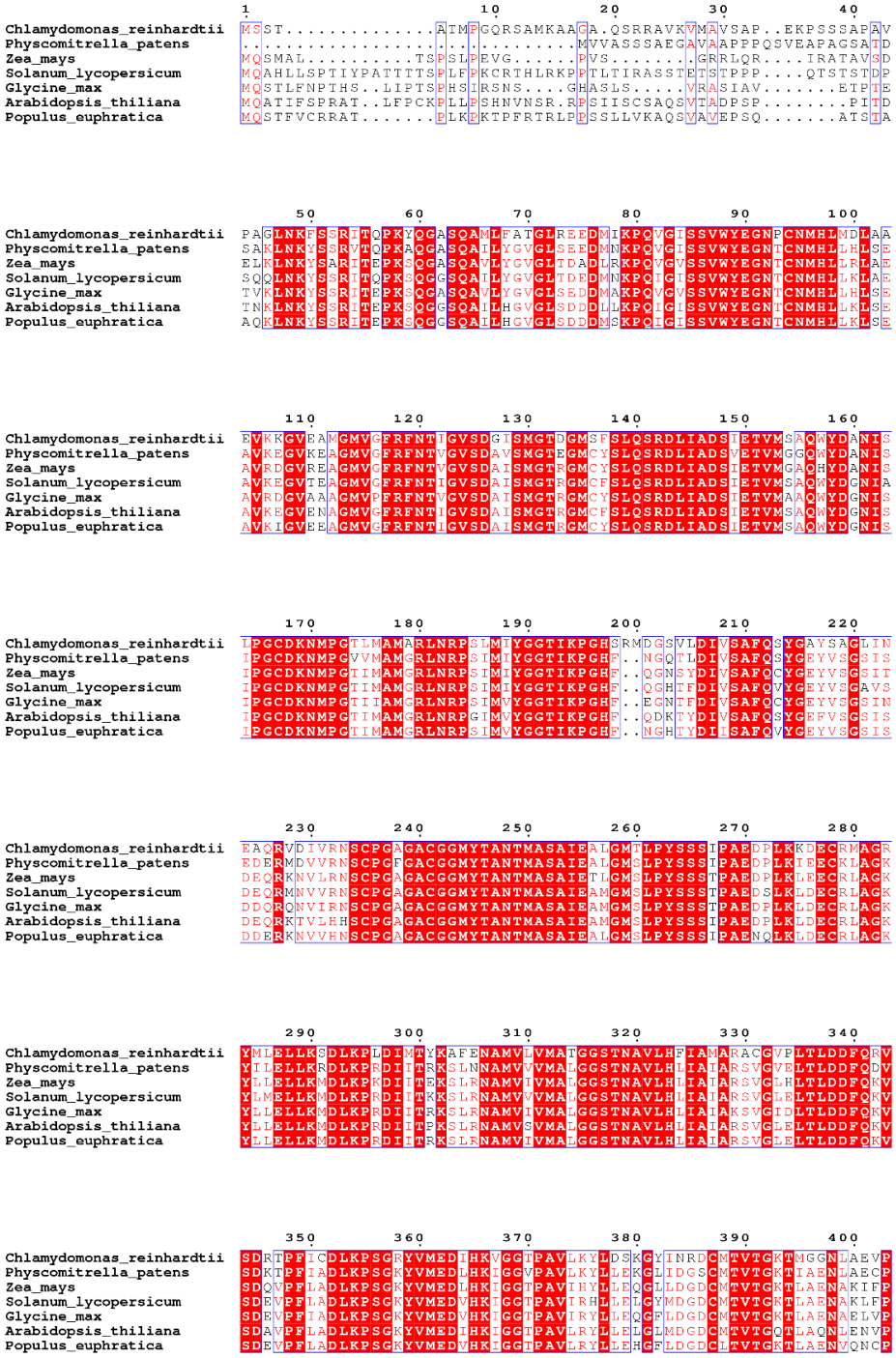
GAATTCATGGGTAAAGTTAAAGTGGGTGTTAATGGTTTTGGTCGCATCGGTCGTCTG
GTGACACGCGCCGCCTTTAATAGCGGCAAAGTGGATATAGTTGCCATCAATGATCCA
TTTATCGATCTGAATTATATGGTGTATATGTTTCAGTATGATAGTACACATGGCAAAT
TTCATGGTACCGTGAAAGCCGAAAATGGCAAATTAGTTATCAATGGCAATCCTATCA
CGATCTTTCAGGAACGCGATCCGAGTAAAATCAAATGGGGCGATGCGGGTGCGGAA
TATGTGGTTGAATCTACGGGCGTGTTTACGACTATGGAAAAAGCCGGCGCACATCTT
CAGGGCGGTGCCAAGCGGGTTATTATTAGCGCTCCGTCAGCAGATGCTCCTATGTTT
GTTATGGGCGTTAATCATGAAAAATATGATAATAGTCTGAAAATCATCTCTAATGCT
TCTTGCACTACAAATTGCCTCGCCCCGCTGGCTAAAGTGATTCATGATAATTTTGGC
ATCGTGGAAGGCTTAATGACGACAGTGCACGCTATTACAGCGACCCAGAAAACCGT
GGATGGTCCGAGCGGTAAACTGTGGCGCGATGGTCGCGGTGCCTTACAGAATATTAT
CCCAGCTTCTACCGGCGCAGCCAAAGCAGTGGGTAAAGTTATTCCAGAACTGAATG
GTAAACTGACGGGCATGGCCTTTCGCGTTCCTACCGCCAATGTGTCAGTTGTGGATT
TAACCTGTCGCTTAGAAAAACCTGCGAAATATGATGACATTAAAAAAGTTGTGAAA
CAGGCATCTGAAGGTCCACTGAAAGGTATCTTAGGCTATACGGAACATCAGGTTGTC
AGCTCAGATTTTAATTCAGATACACATTCTAGTACGTTTGATGCAGGTGCCGGCATT
GCCCTGAATGATCATTTTGTTAAACTGATTAGCTGGTATGATAATGAATTTGGCTAT
AGCAATCGTGTTGTGGATCTGATGGCACACATGGCGTCTAAAGAATAACTCGAG

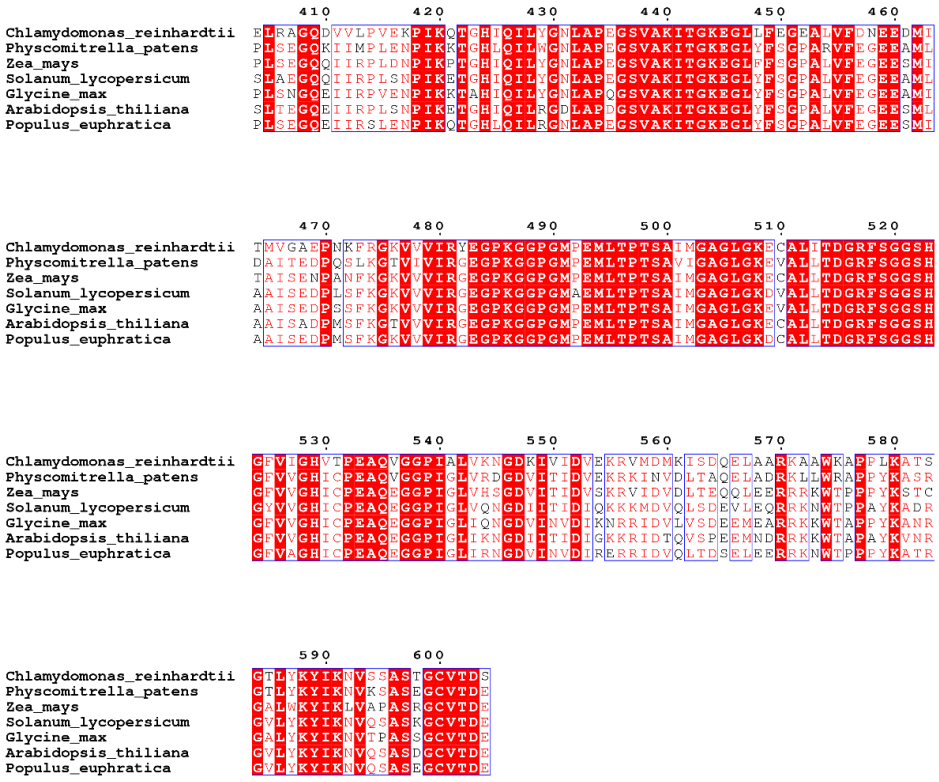
**The codon-optimized sequence of *astD* for expression in *A. thaliana*.** The pDHAD chloroplast localization signal is in blue, and the FLAG-tag is in red.

ATGCAAGCCACCATCTTCTCTCCACGCGCCACTCTCTTCCCCTGTAAACCTCTCCTCC
CTTCCCACAATGTCAACTCTCGCCGTCCCTCAATCATCTCCTGCTCCGACTACAAAG
ACGATGACGACAAACACATGGACTACAAAGACGATGACGACAAAGACTACAAAGA
CGATGACGACAAATTCGCATCTCGTATCCGATCAAGAGCCCTTGGTCTTCACCCCCG
AGCTCGATTTGAAAACACTCGACTTCCGGCCAGTACTACCGGAAGGAGGTATAAGA
GCGACGAGACCCTGAACAGAATTAGCTCAAAAATCACTCAGCCGAAGTCCCAGGGG
GCATCTCAGGCGATGCTCTATGCTACCGGCTTGACAGAGGAAGACATGTCAAAGCC
GCAGGTCGGAATCTCATCCGTTTGGTTTGAGGGGAATCCATGCAACATGCATCTTCA
CGATTTGTCCGCCATCGTTCGTGACTCTGTTCACCGAGCAGGTCTTGTCCCCATGAGG
TTTAATTCAGTTGGCGTTTCAGACGGGATCTCTATGGGTACGAAGGGGATGAGATAT
TCCCTCCAGTCCAGAGAACTGATAGCTGACGGTATTGAGACCGTAATGAACGCTCAA
TGGTATGATGCCAACGTGTCTCTGCCTGGCTGTGACAAAAATATGCCGGGCGTCCTC
ATGGCGATGGGTAGAACCAATCGTCCTTCCATCATGGTCTACGGAGGTAGCATCAAA
CCCGGTTGCTCTGCCAAGGGACAAAAATTAGATCTGGTATCTGCTTTTCAAAGCTAT
GGACAGTTCATAACGGGTCAGATAGATGAAAAGGAACGATTTGATATAATTAGAAA
CGCATGTCCTGGTCGTGGAGCTTGCGGTGGAATGTACACGGCAAACACACTAGCTAC
CGCTATAGAGACTATGGGTATGACAGTGCCCGGCTCTTCTAGCTGTCCGGCGGATGA
CCCTAAAAAACTAGTGGAGTGCGAGAATATAGGCGAGGTGGTAAAGACGATGCTCC
GAGAAGATATCAAACCGCGAGACGTTCTCACCCGTCAGGCTTTTGAAAATGCGATG
ATTGTGGTCAACATACTAGGGGGTTCTACAAATGCCGTACTCCATCTAATAGCCATA
GCTGATTCCGTAGGGATCAAACTGACGATAGATGATTTCCAAGCCGTATCTGATAAA
ACACCGTTCTTGGCCGATCTAAAACCCTCAGGGAAGTACTTAATGAACGATTTGTAC
AACATCGGCGGCACCCCGGCGCTTTTAAAGTATCTCCTTAAGGAAGGACTTATTGAC
GGAAGTGGCATAACTGTCACTGGTAAAACCATGAAAGAAAACGTGGCCTCTTGGCC
CGATTTTCCCTCAGATCAAGACATAATCCGACCGCTCTCAAACCCGATTAAGCCCTC
CGGCCATTTACAAATTCTCAGGGGGTCTCTAGCGCCGGGAGGGTCAGTAGGTAAGA
TTACGGGTAAGGAGGGGCTCAGATTTGAGGGCACCGCAAAGTGCTATGACTACGAG
GATGCCTTCATAGAGAGTCTTGAGAGAGGTGAGATTAAAAAGGGAGAGAAGACAGT
GGTGATAATAAGGTATGAAGGCCCAAAGGGGGGCCCGGGCATGCCTGAAATGCTCA
AACCAAGCGCCGCGATTATGGGTGCGGGCTTAGGTCAAGACGTTGCACTTCTGACA
GATGGGCGTTTCTCAGGGGGATCACACGGCTTTCTAATAGGTCACATTGTACCAGAG
GCGATGGAGGGGGGTCCCATAGCTTTAGCACGTGACGGGGATCGTATTGTCATCGA
CGCCGAGGAGCGAGTAGTCGATCTTGAGATACCGACCGAAGAGTTGGAGAAACGTC
GAAAGGAGTGGAAAGCCCCCCCACTTCGATACCAAAAGGGCACGCTCAAAAAATAT
TGCACGCTTGTTAGCGATGCATCTCACGGTTGCGTGACAGACGGGCCGATTTAG

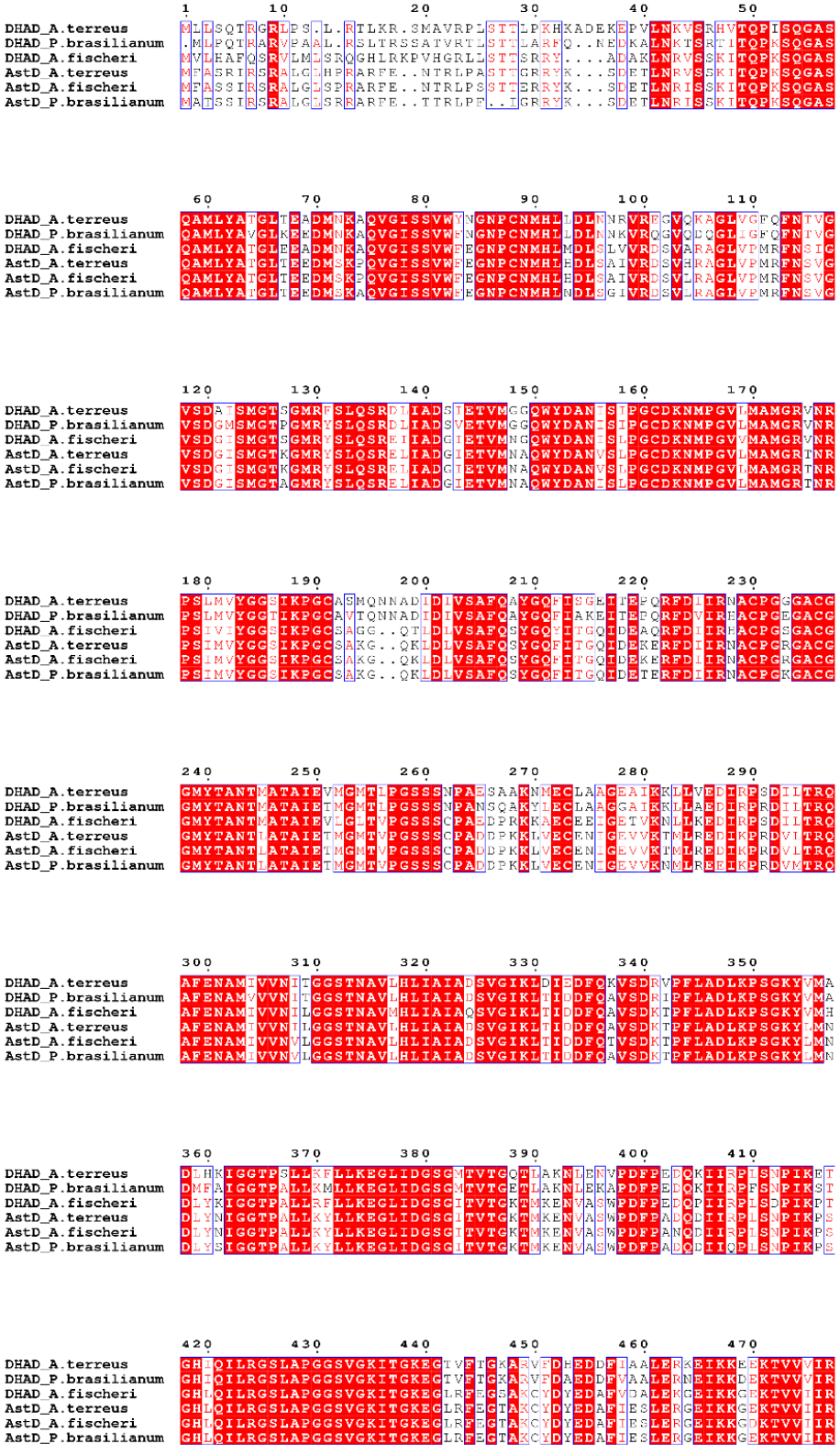
## 4.2 Supplementary figures

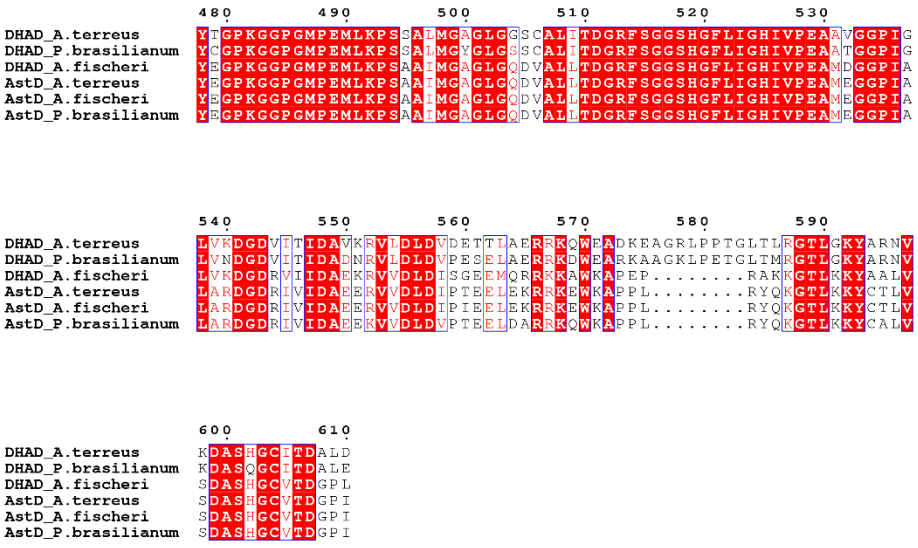
**Supplementary Figure 1. Alignment of amino acid sequences of DHADs from different plant species.** The identities of DHAD sequences among flowering plant are ~80%. The lack of identity at the *N*-terminal of these DHAD results from the differences in chloroplast localization signals from different species.




410 420 430 440 450 460
Chlamydomonas_reinhardtii
Physcomitrella_patens
Zea_mays
Solanum_lycopersicum
Glycine_max
Arabidopsis_thiliana
Populus_euphratica
470 480 490 500 510 520
530 540 550 560 570 580
590 600

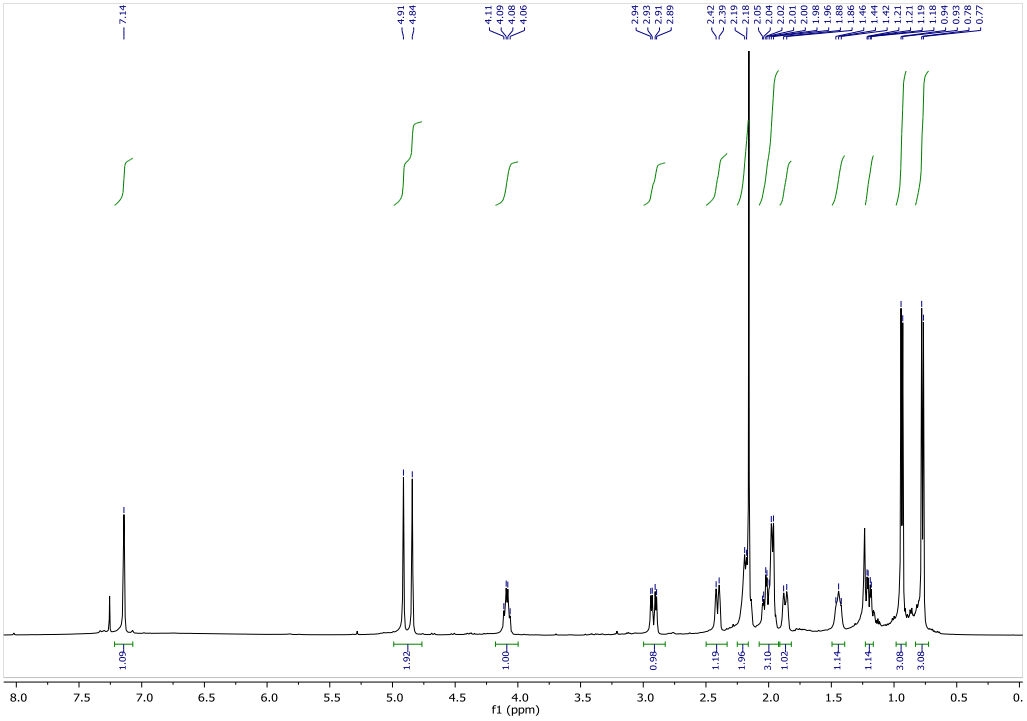
**Supplementary Figure 2. Alignment of amino acid sequences of AstD and housekeeping DHAD from different strains.**



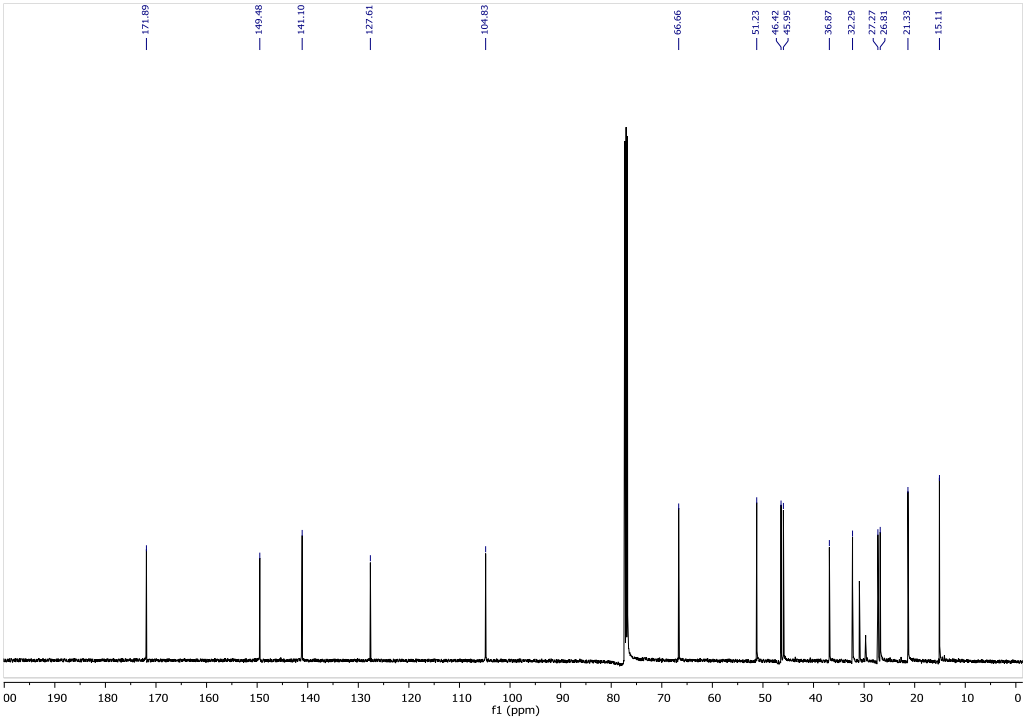

480 490 500 510 520 530
DHAD_A.terreus YTGPKGGPGMPEMLKPSSALMGAGLGSSCALFTDGRFSGGSHGFLIGHIVPEAAVGGPIG
DHAD_P.brasilianum YCGPKGGPGMPEMLKPSSALMGYGLGSSCALFTDGRFSGGSHGFLIGHIVPEAATGGPIG
DHAD_A.fischeri YEGPKGGPGMPEMLKPSAALMGAGLGQDVALFTDGRFSGGSHGFLIGHIVPEAMDGGPIA
AstD_A.terreus YEGPKGGPGMPEMLKPSAALMGAGLGQDVALFTDGRFSGGSHGFLIGHIVPEAMEGGPIA
AstD_A.fischeri YEGPKGGPGMPEMLKPSAALMGAGLGQDVALFTDGRFSGGSHGFLIGHIVPEAMEGGPIA
AstD_P.brasilianum YEGPKGGPGMPEMLKPSAALMGAGLGQDVALFTDGRFSGGSHGFLIGHIVPEAMEGGPIA
540 550 560 570 580 590
DHAD_A.terreus LVKDGDVITIDAVKRVLDLDVDETTIAERRKQWEADKEAGRLPPTGLTLRGTLGKYARNV
DHAD_P.brasilianum LVNDGDVITIDADNRVLDLDVPESELAERRKDWEARKAAGKLPETGLTMRGTLGKYARNV
DHAD_A.fischeri LVKDGDRVIIDAEKRVVDLDISGEEMQRRKKAWKAPEP.....RAKKGTLKKYAALV
AstD_A.terreus LARDGDRIVIDAEERVVDLDIPTEELEKRRKEWKAPPL.....RYQKGTLKKYCTLV
AstD_A.fischeri LARDGDRIVIDAEERVVDLDIPIEELEKRRKEWKAPPL.....RYQKGTLKKYCTLV
AstD_P.brasilianum LARDGDRIVIDAEEKVVDLDVPTEELDARRKQWKAPPL.....RYQKGTLKKYCALV
600 610
DHAD_A.terreus KDASHGCITDALD
DHAD_P.brasilianum KDASQGCITDALE
DHAD_A.fischeri SDASHGCVTDGPL
AstD_A.terreus SDASHGCVTDGPI
AstD_A.fischeri SDASHGCVTDGPI
AstD_P.brasilianum SDASHGCVTDGPI

**Supplementary Figure 3. NMR analyses of the compounds in this study.** The experiments were repeated independently for 3 times with similar results.

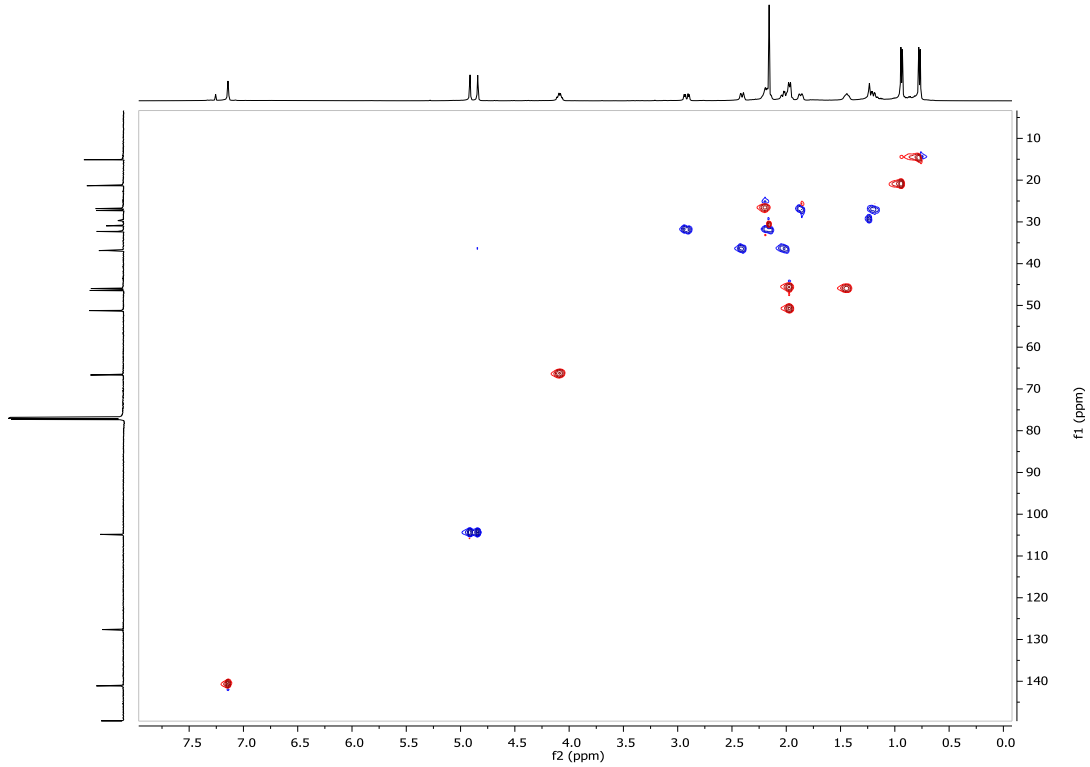
**$^{1}$H NMR of compound 2 (500 MHz, $CDCl_3$):**



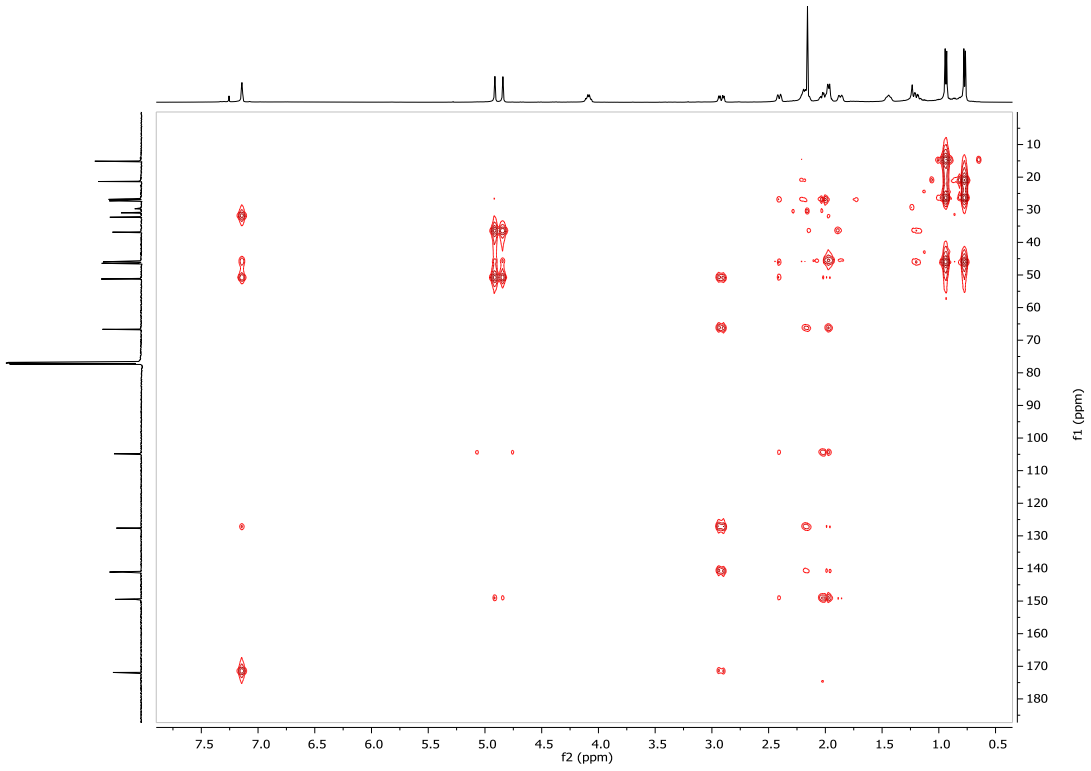

**$^{13}$C NMR of compound 2 (125 MHz, $CDCl_3$):**



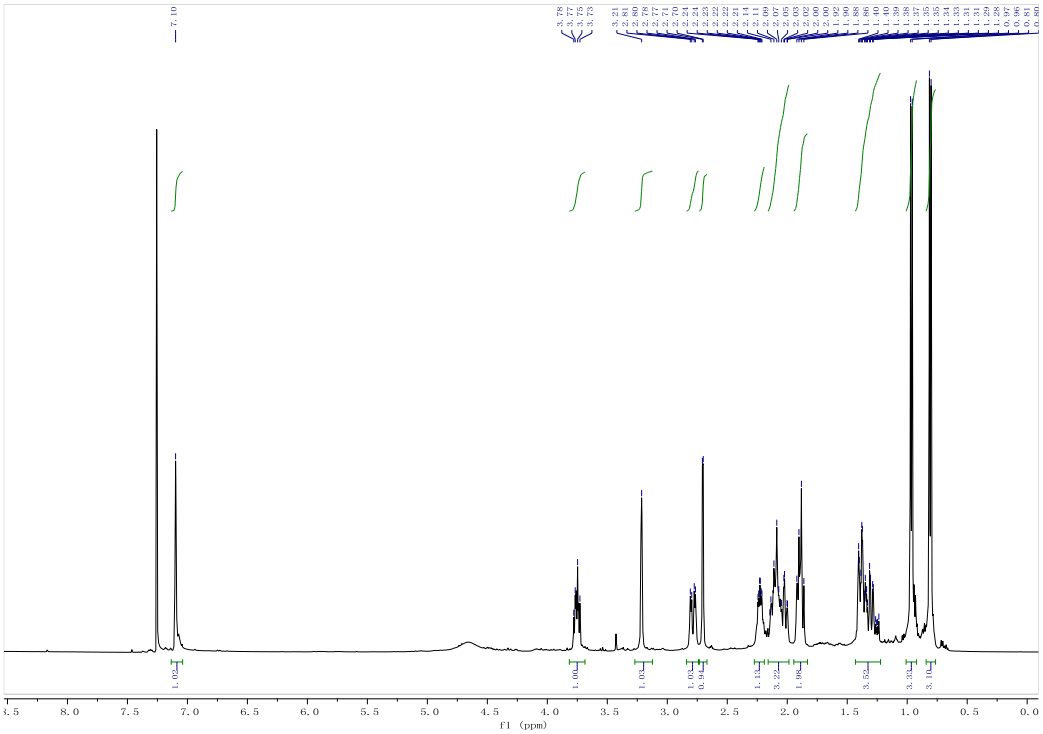
**HSQC of compound 2 (500 MHz, $CDCl_3$):**



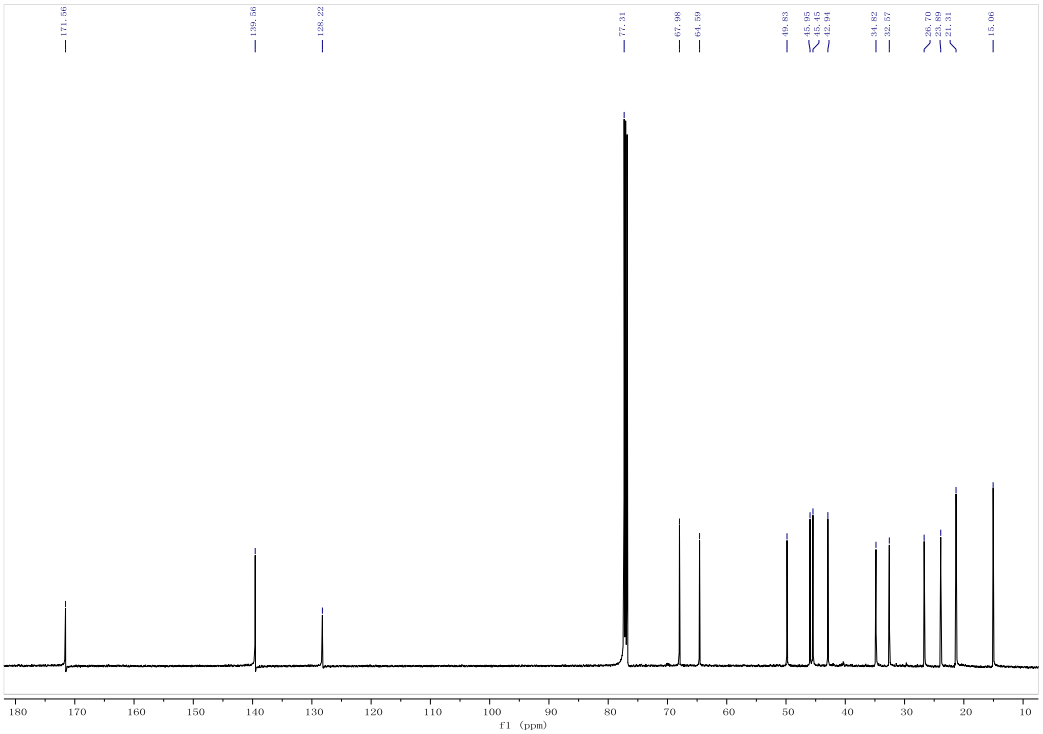

**HMBC of compound 2 (500 MHz, $CDCl_3$):**



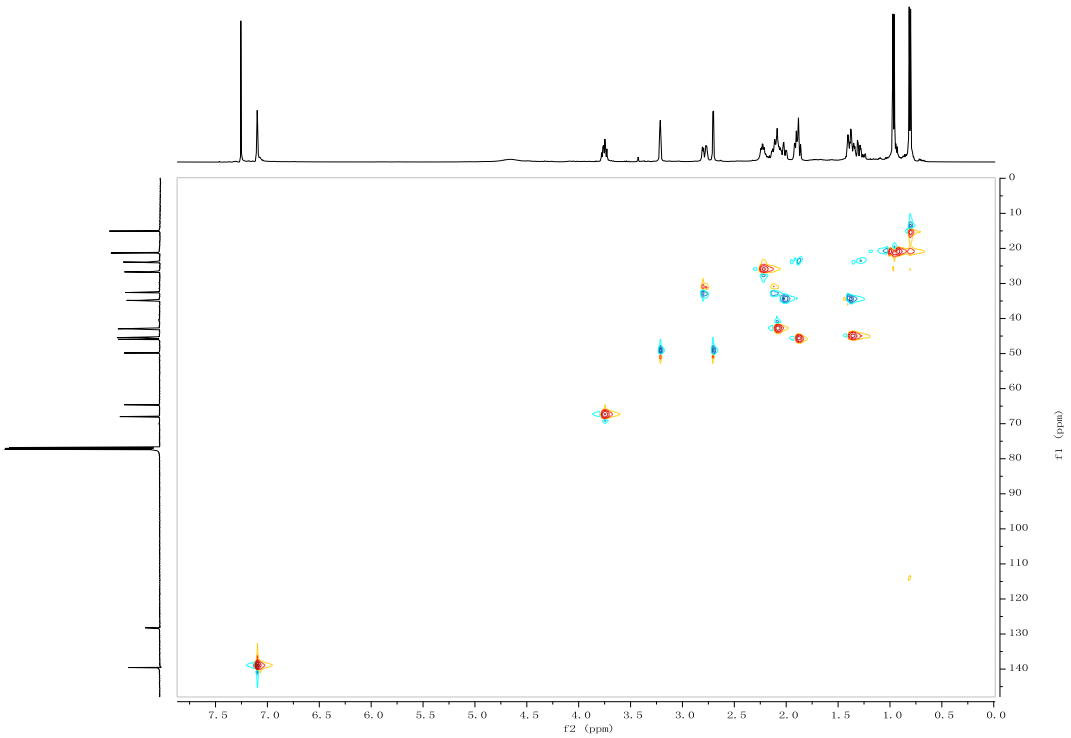
**$^{1}$H NMR of compound 3 (500 MHz, $CDCl_3$):**



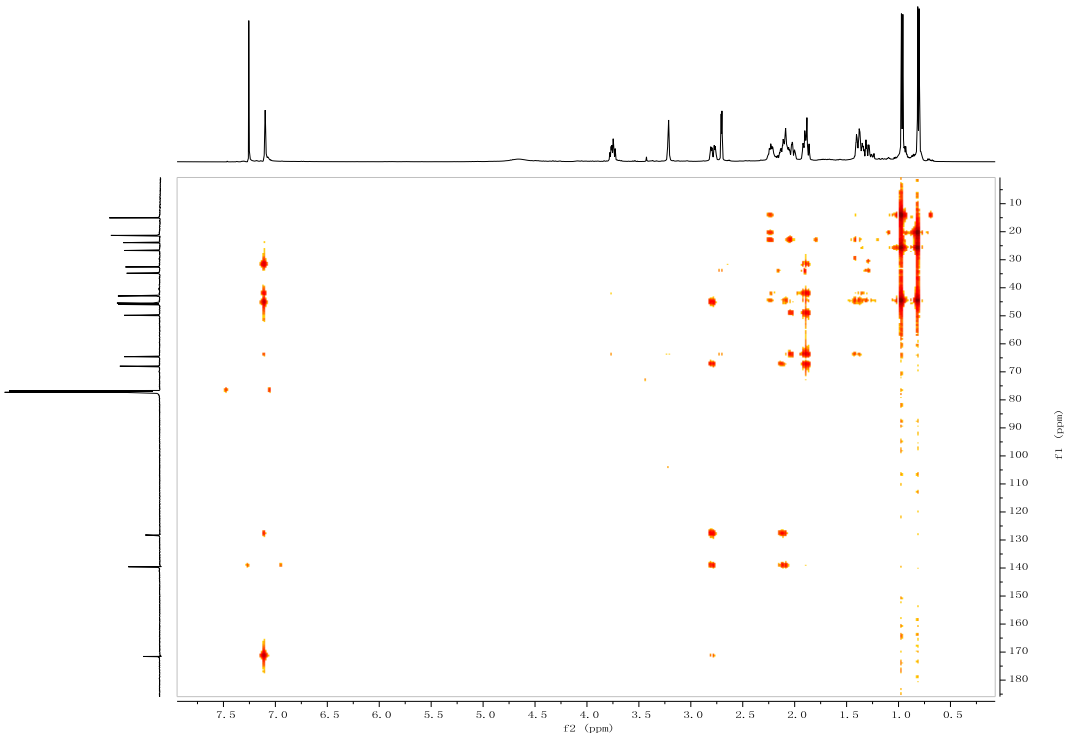
**$^{13}$C NMR of compound 3 (125 MHz, $CDCl_3$):**



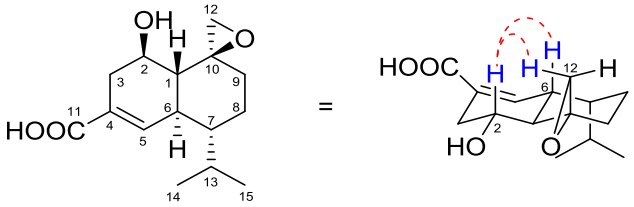
**HSQC of compound 3 (500 MHz, $CDCl_3$):**



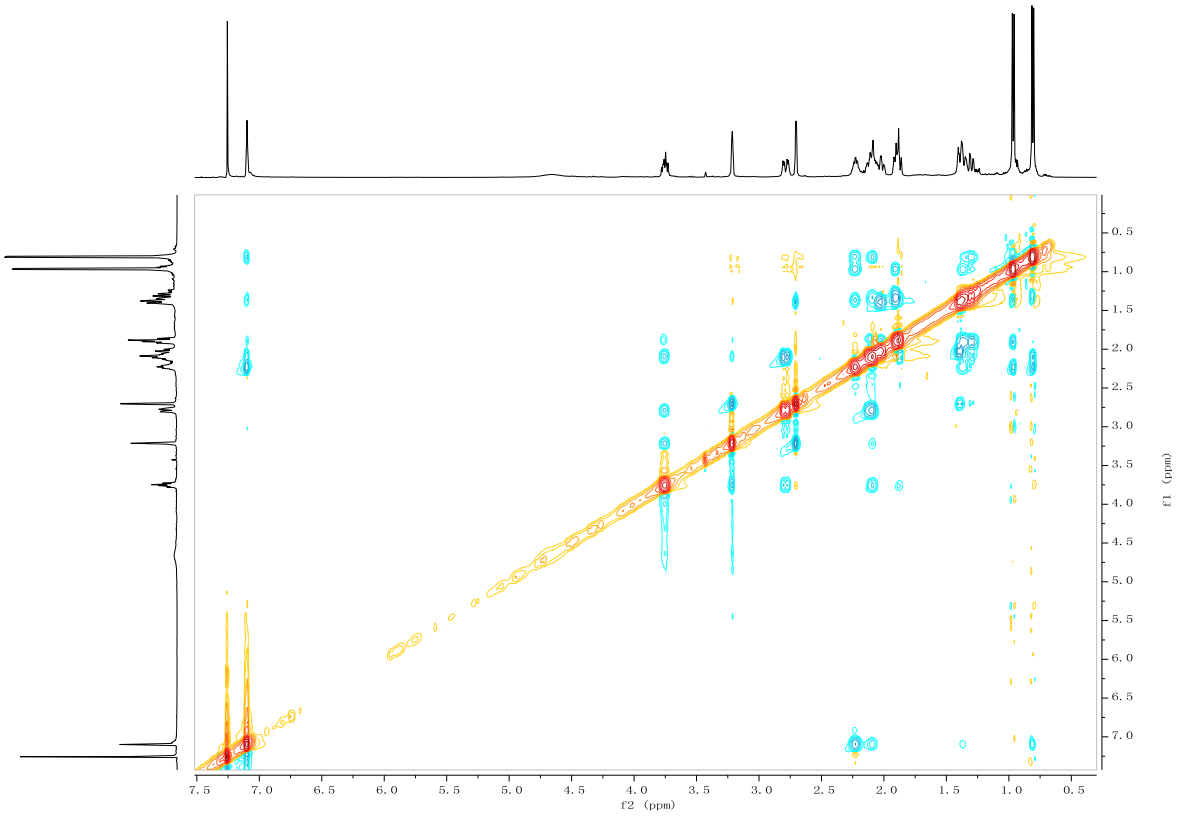

**HMBC of compound 3 (500 MHz, $CDCl_3$):**



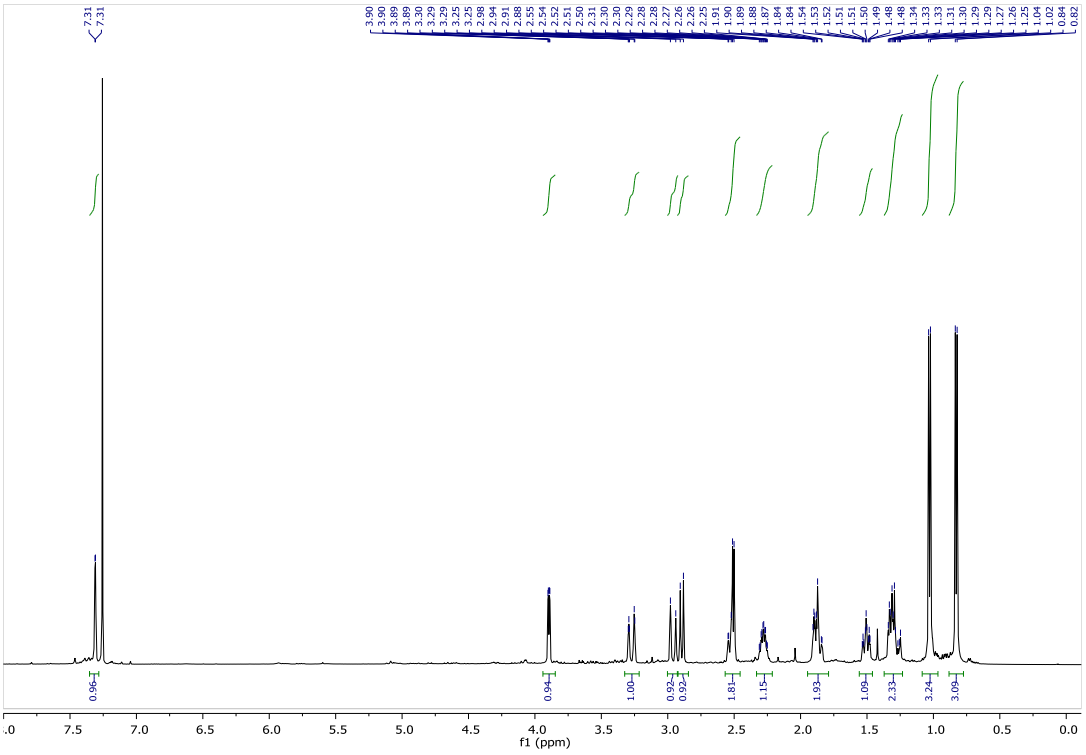
**NOESY of compound 3 (500 MHz, $CDCl_3$):**



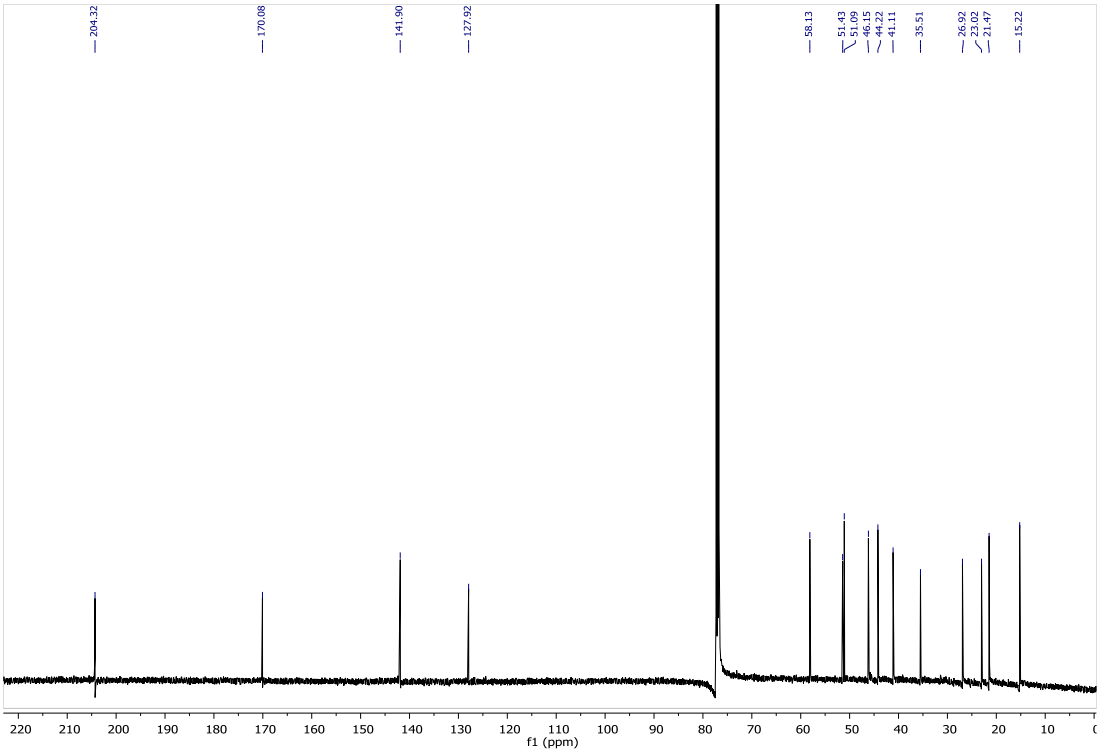

avo-266 (**3**)



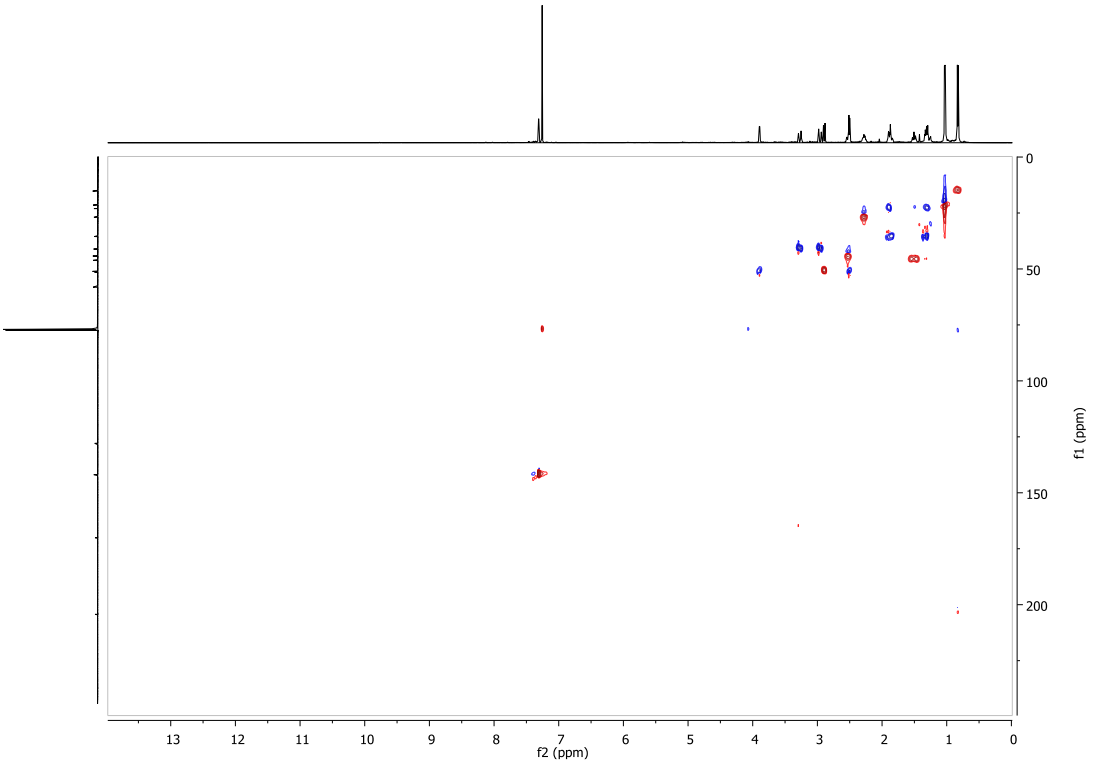
## $^{1}$H NMR of compound 4 (500 MHz, $CDCl_3$):



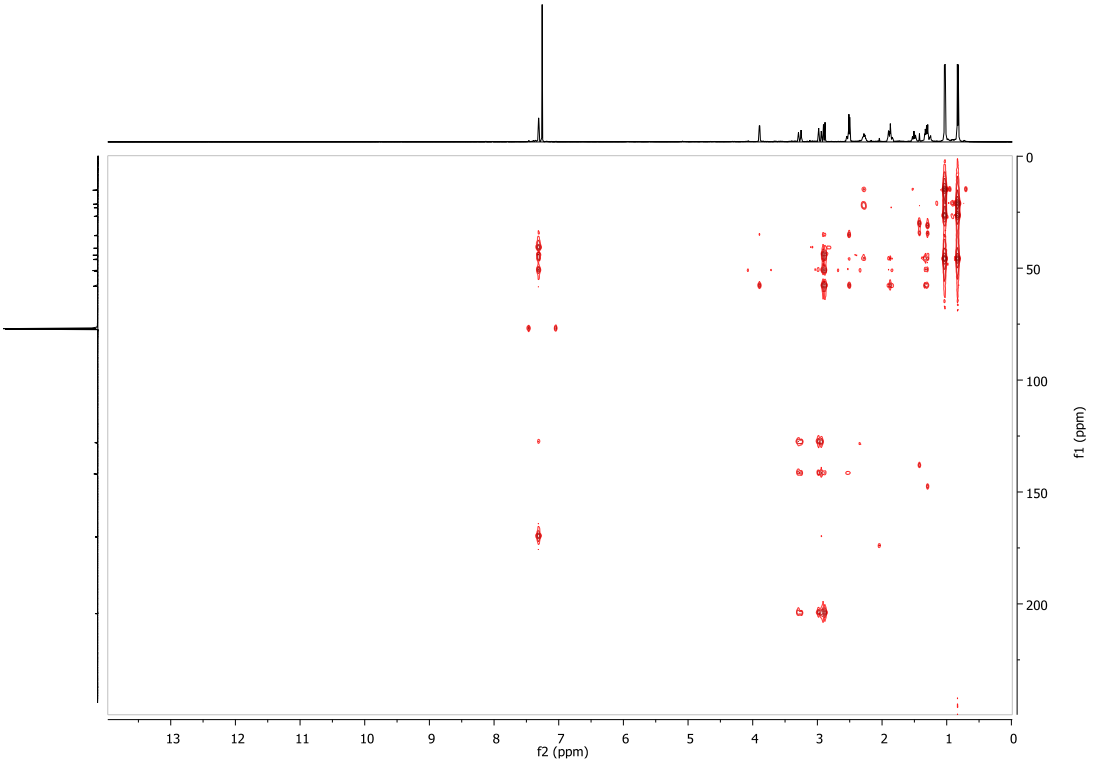

## $^{13}$C NMR of compound 4 (125 MHz, $CDCl_3$):



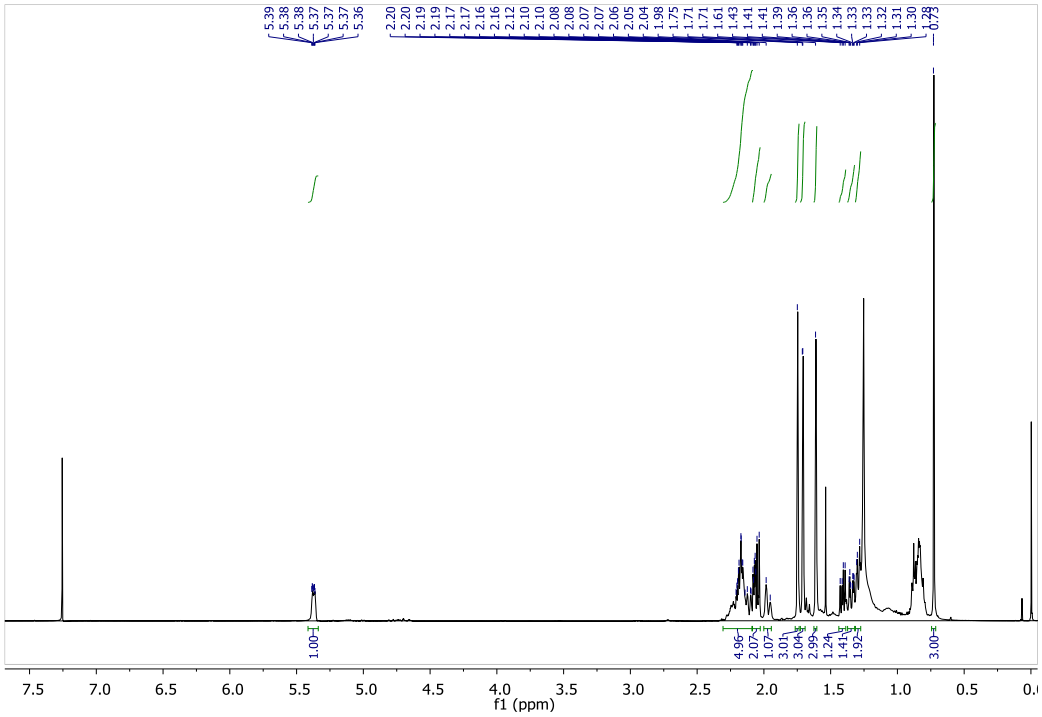
**HSQC of compound 4 (500 MHz, $CDCl_3$):**



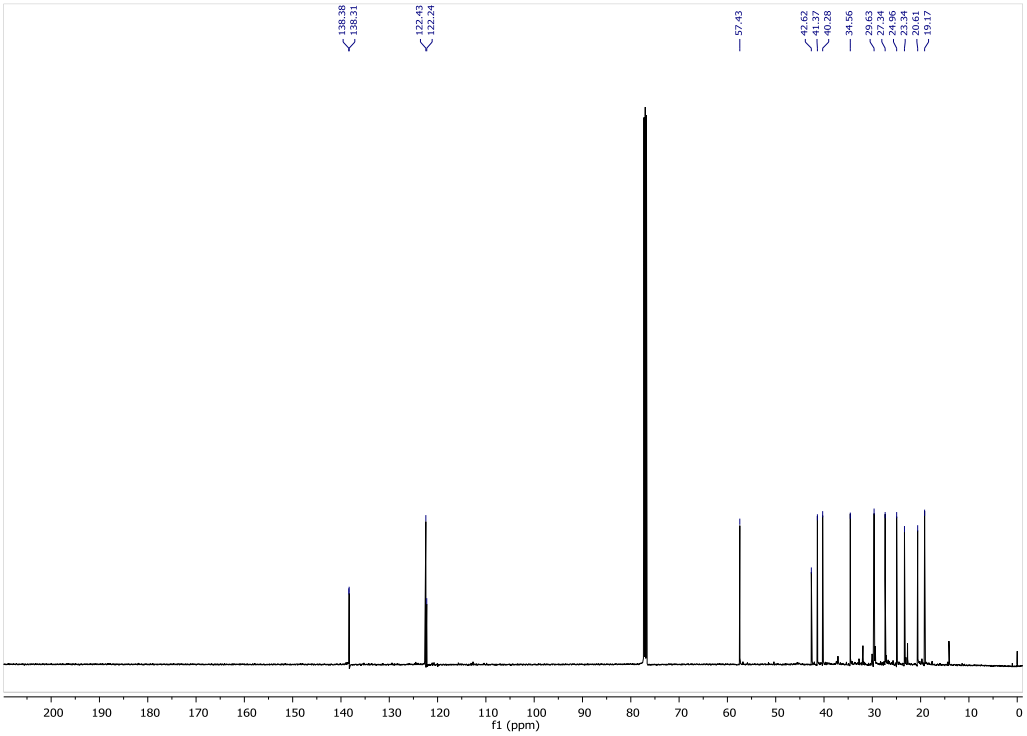

**HMBC of compound 4 (500 MHz, $CDCl_3$):**



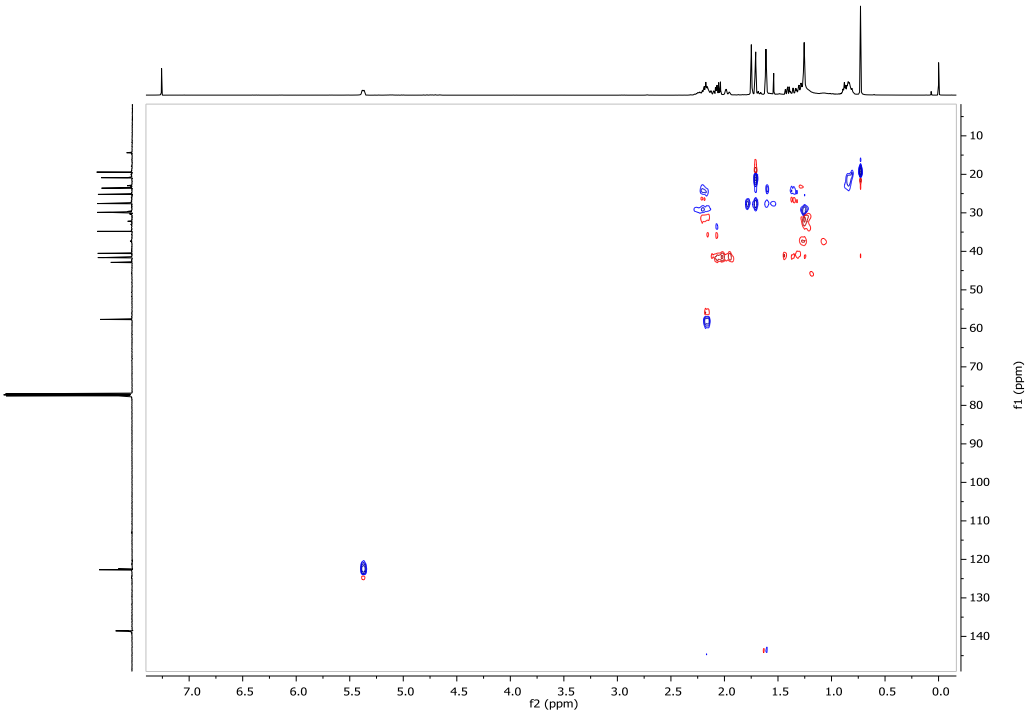
**$^{1}$H NMR of compound 8 (500 MHz, $CDCl_3$):**



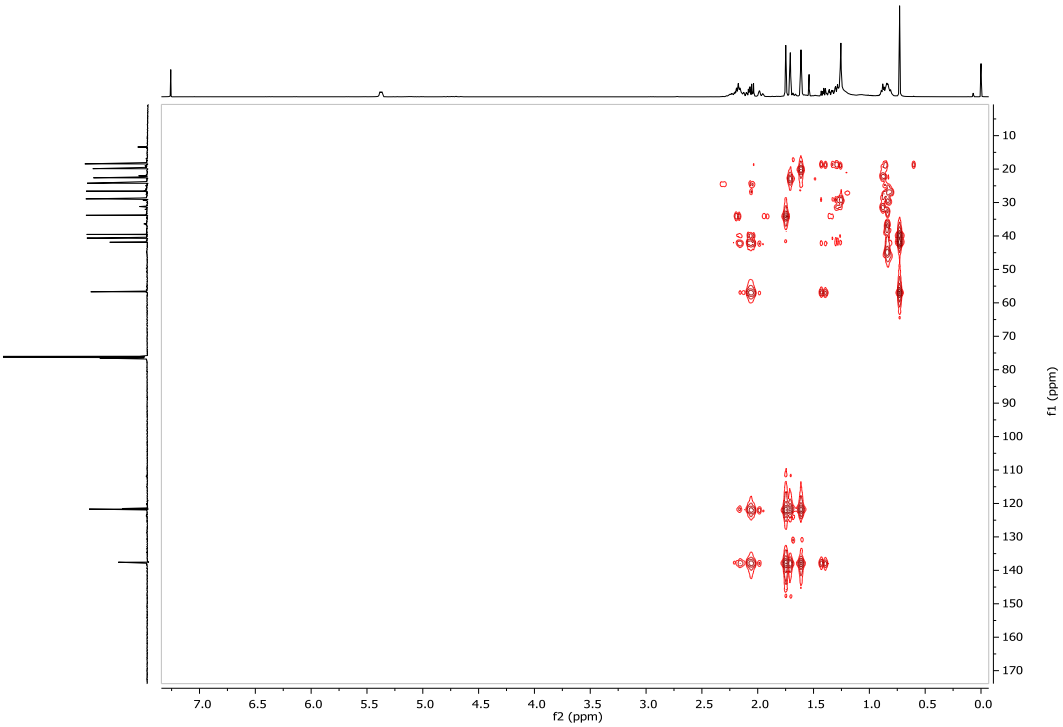
**$^{13}$C NMR of compound 8 (125 MHz, $CDCl_3$):**



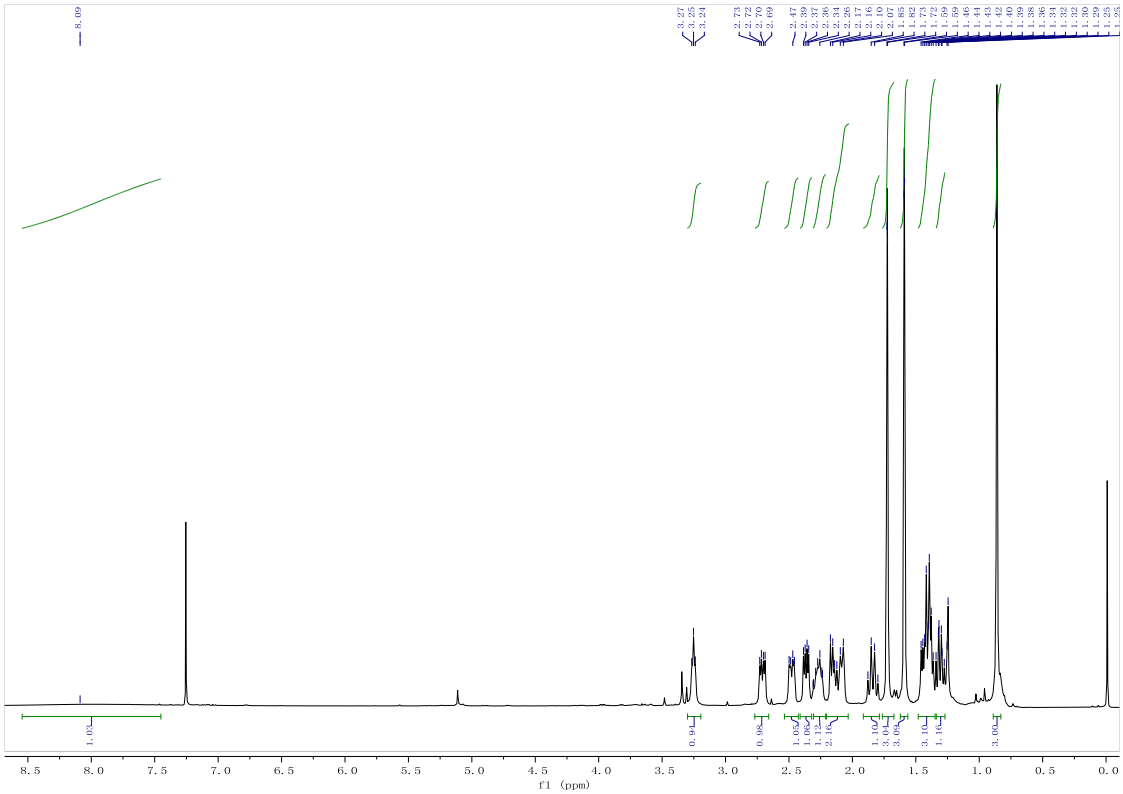
**HSQC of compound 8 (500 MHz, $CDCl_3$):**



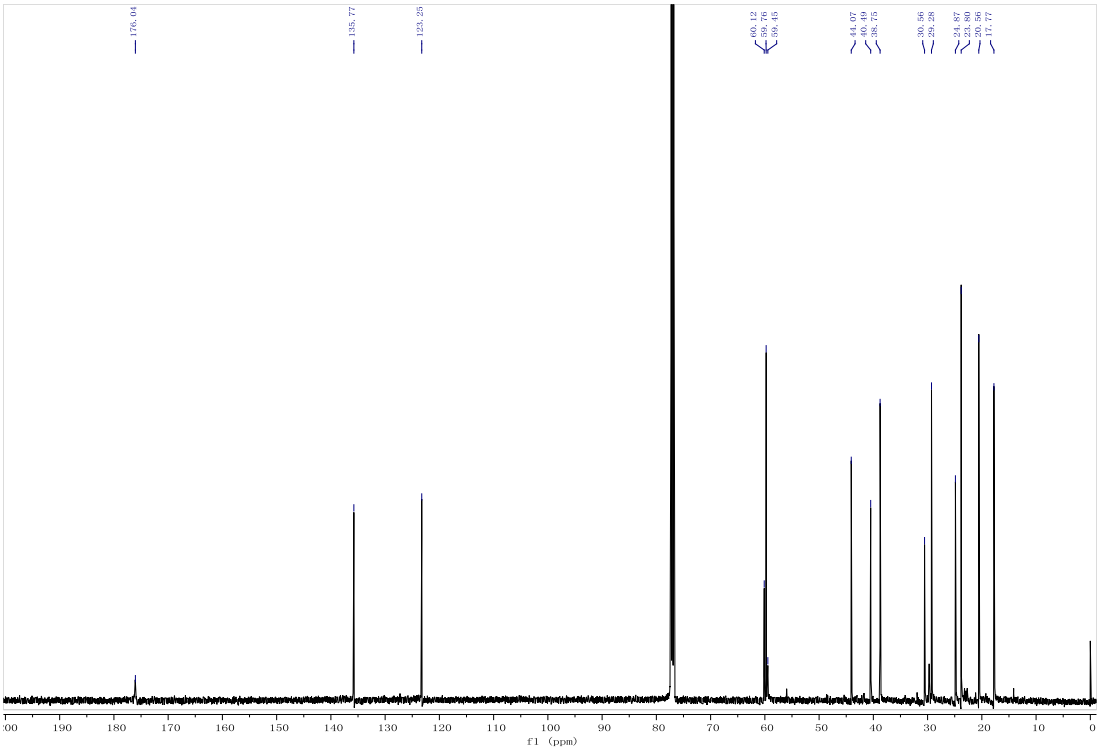

**HMBC of compound 8 (500 MHz, $CDCl_3$):**



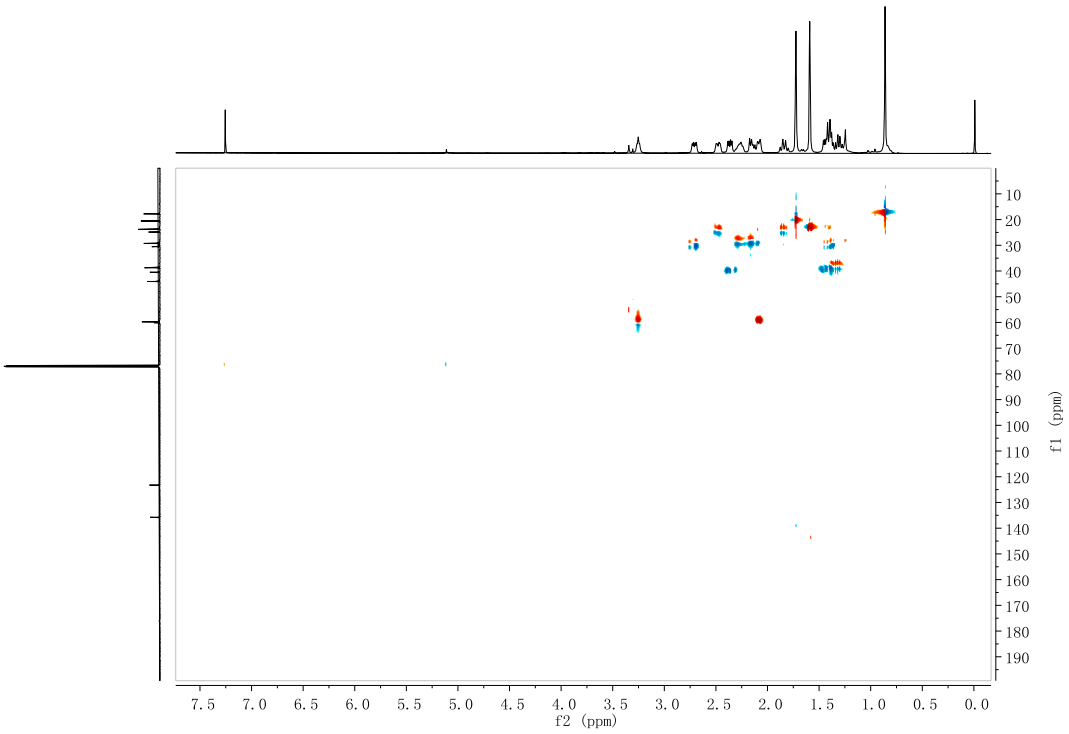
**$^{1}$H NMR of compound 9 (500 MHz, $CDCl_3$):**



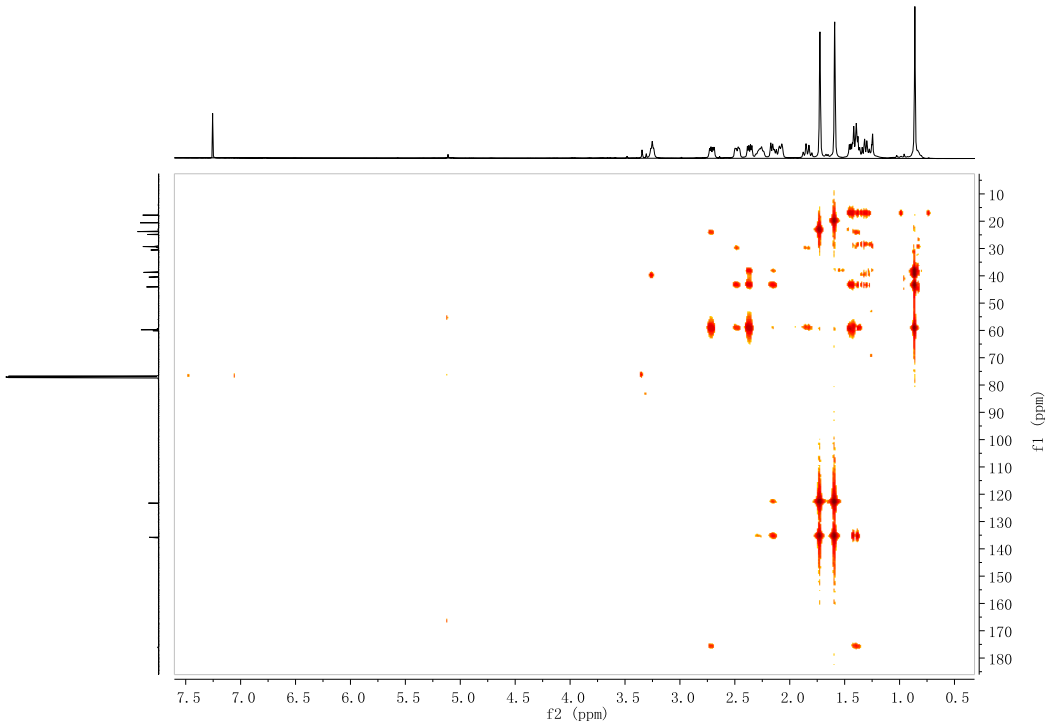
**$^{13}$C NMR of compound 9 (125 MHz, $CDCl_3$):**



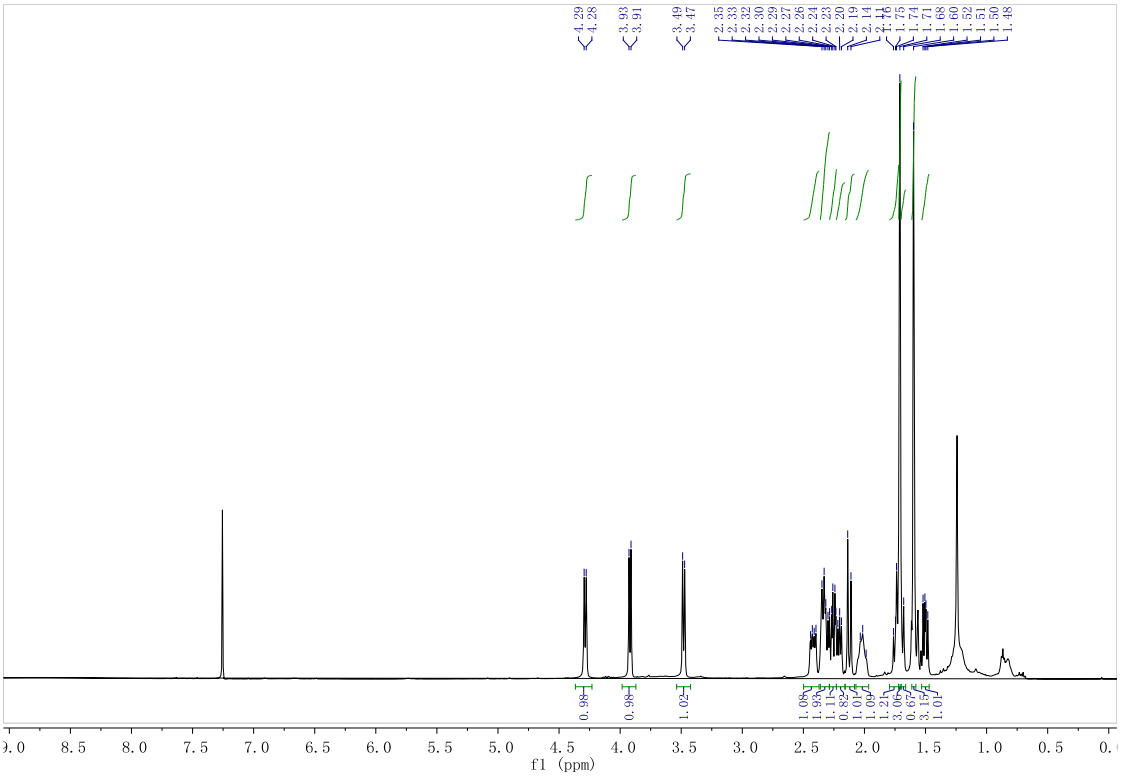
**HSQC of compound 9 (500 MHz, $CDCl_3$):**



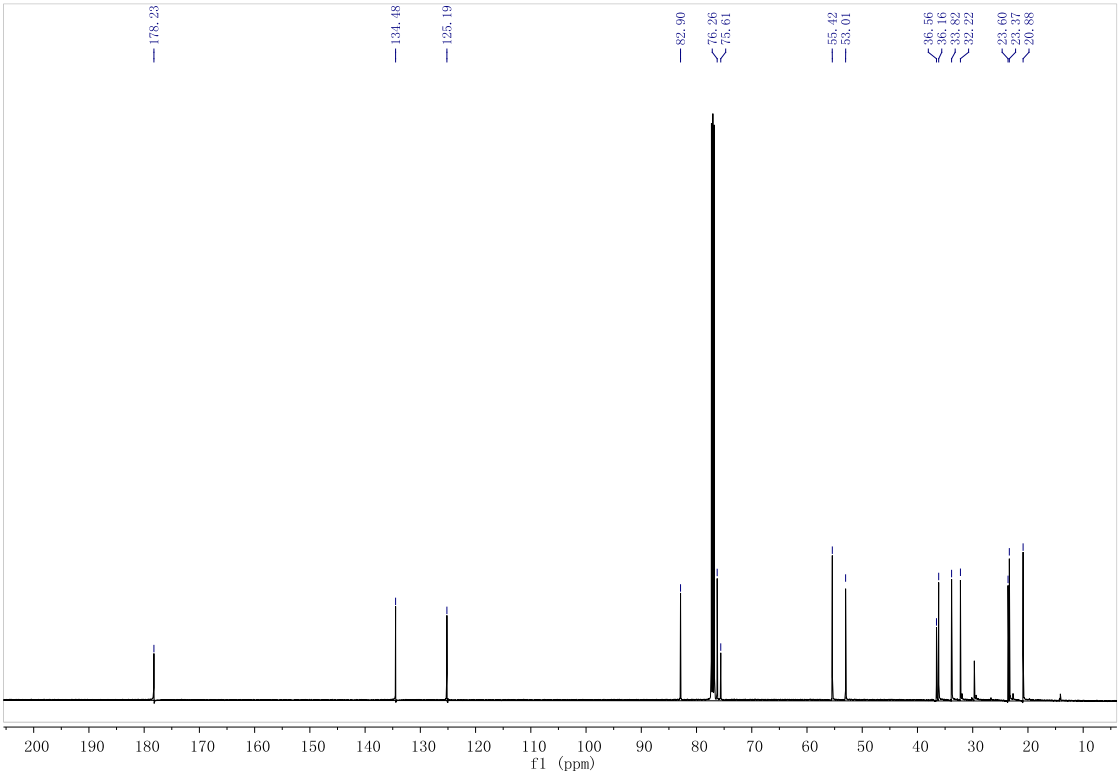

**HMBC of compound 9 (500 MHz, $CDCl_3$):**



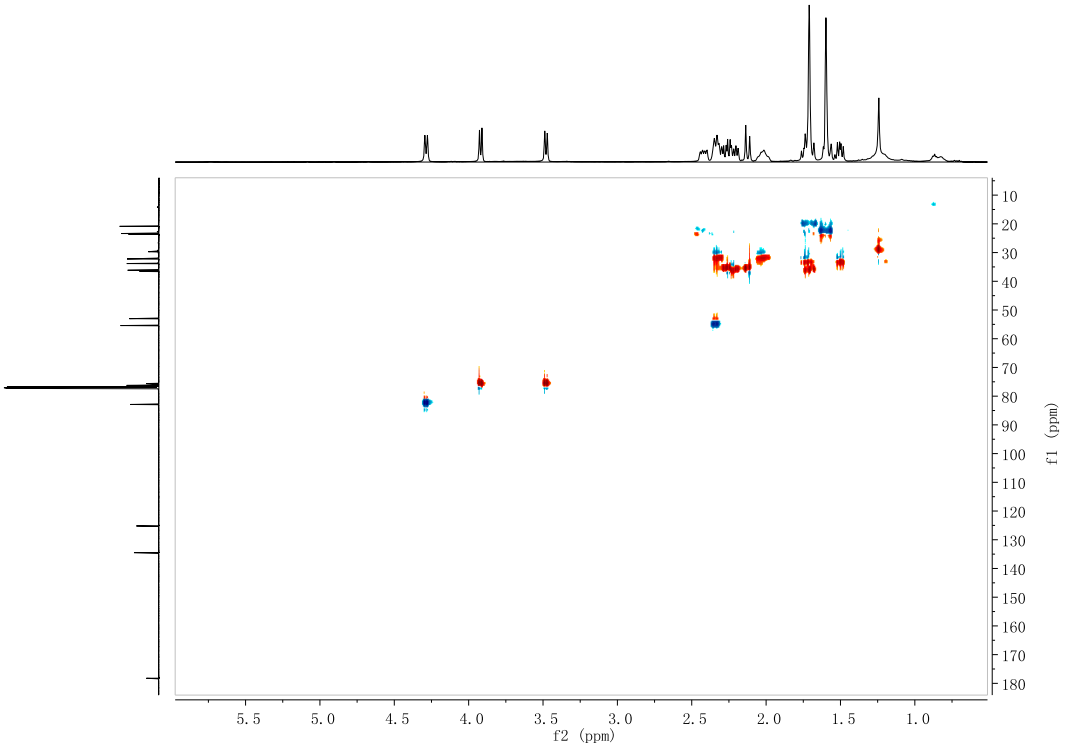
**$^{1}$H NMR of AA (500 MHz, $CDCl_3$):**



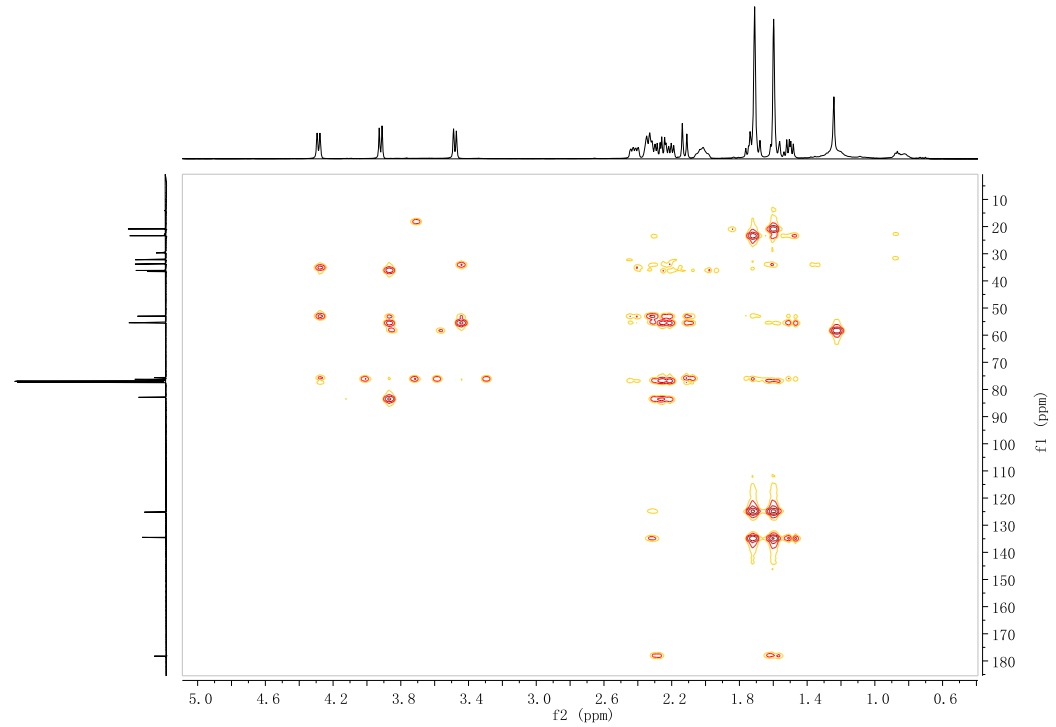

**$^{13}$C NMR of AA (125 MHz, $CDCl_3$):**



**HSQC of AA (500 MHz, $CDCl_3$):**

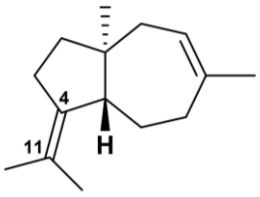



**HMBC of AA (500 MHz, $CDCl_3$):**



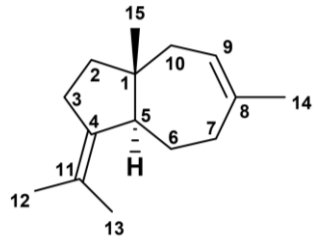
**Supplementary Figure 4. EI-MS of compound 8 by GC-MS analysis.**

The structure of compound **8** (top right) and its known enantiomer (+)-Dauca-4(11),8-diene (top left). The EI-MS of compound **8** (bottom). The EI-MS spectrum of (+)-Dauca-4(11),8-diene is reported as m/z (rel.int): 204 $[M]^+$ (22), 189 $[M\text{-}Me]^+$ (2), 161 (18), 148 (3), 136 (100), 133 (10), 121 (60), 119 (10), 107 (17) 105 (15), 93 (19), 91 (18), 79 (12), 77 (11), 55 (10), 41 (22). The EI-MS of both compound **8** and (+)-Dauca-4(11),8-diene are identical. The experiment was repeated independently with similar results for 7 times.



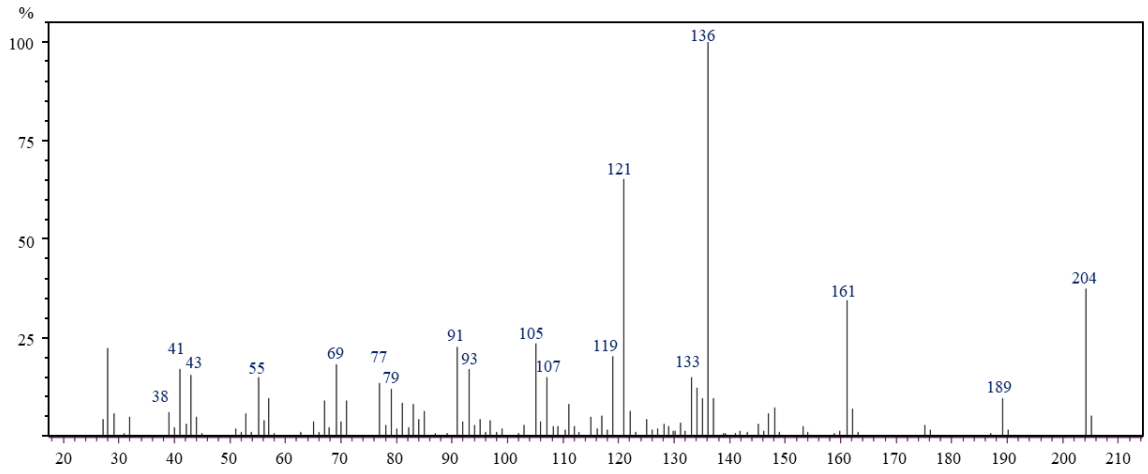

(+)-dauca-4(11),8-diene

$[\alpha]_D^{22} > +40°$ (*n*-hexane; $c < 0.1$)



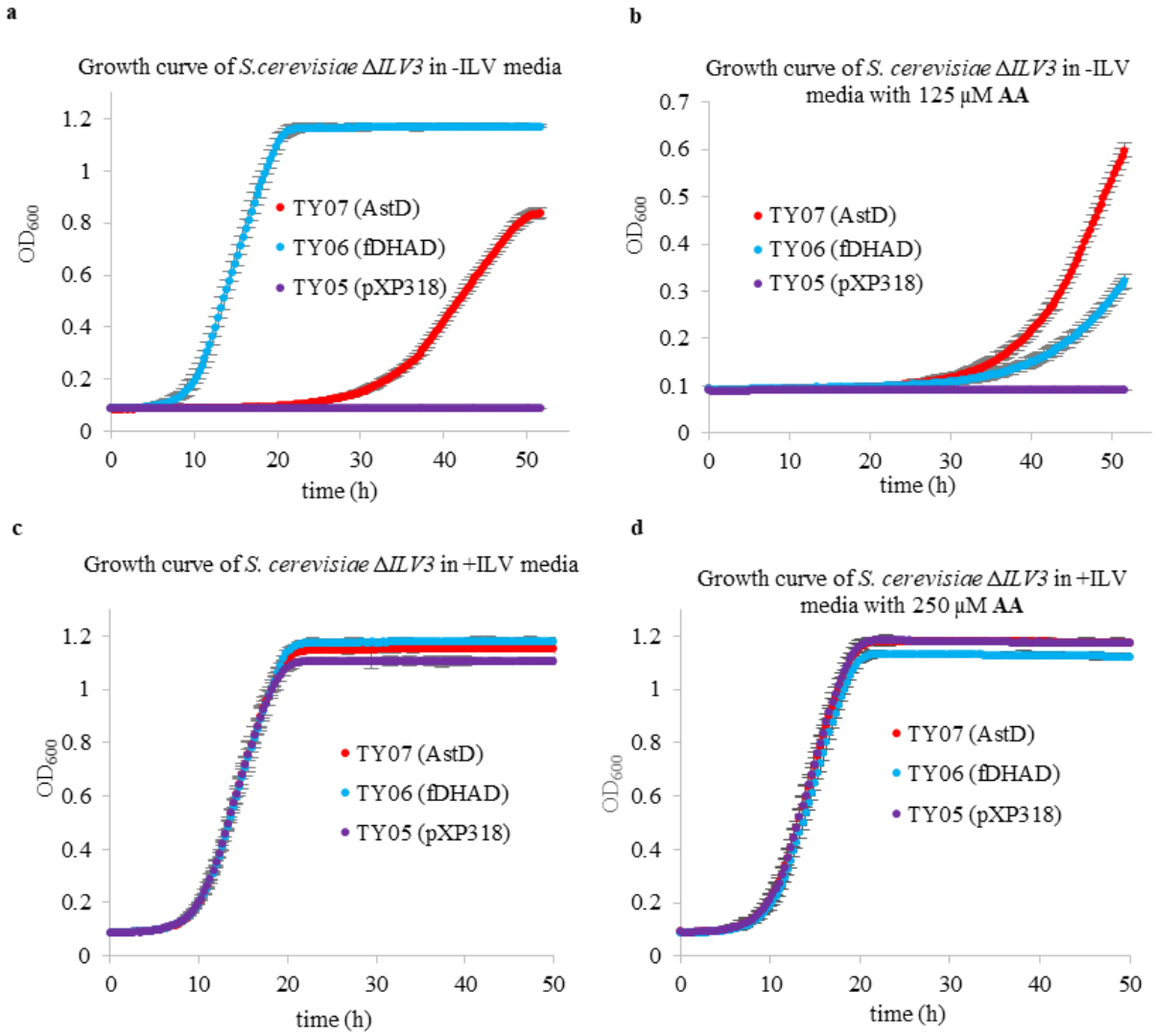

**1**, (-)-dauca-4(11),8-diene

$[\alpha]_D^{22} = -30°$ (*n*-hexane; $c = 0.1$)



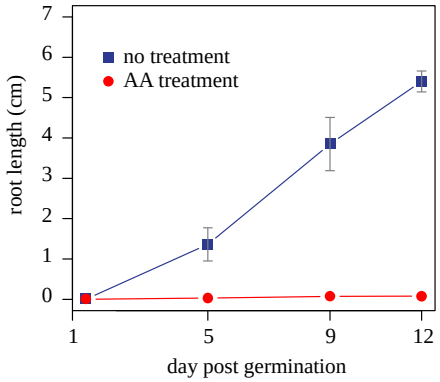
**Supplementary Figure 5. Growth curve of *S. cerevisiae* Δ*ILV3* expressing AstD and fDHAD.**

Three DHAD enzymes were assayed, including pDHAD (**p**lant DHAD from A. thaliana), fDHAD (**f**ungal housekeeping DHAD from A. terreus) and AstD (DHAD homolog within ast cluster). $IC_{50}$ and $K_i$ values of **AA** were measured based on inhibition percentage at different **AA** concentrations. Center values are averages, errors bars are s.d.; n = 3 biologically independent experiments. **a,** Plot of the inhibition percentage of 0.5 μM fDHAD as a function of **AA** concentration**. b,** Plot of the inhibition percentage of 0.5 μM pDHAD as a function of **AA** concentration**. c,** Plot of the inhibition percentage of 0.5 μM AstD as a function of **AA** concentration**. d,** Analysis of inhibitory kinetics of **AA** on pDHAD using the Lineweaver-Burk method at different concentrations of **AA** (left). Linear fitting of apparent Michaelis constant ($K_{M,app}$) as a function of **AA** concentration yields the inhibition constant ($K_i$) of **AA** on pDHAD (right).



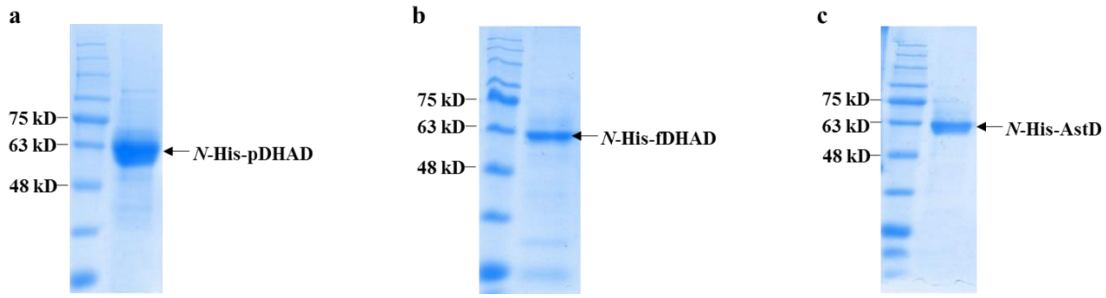
**Supplementary Figure 6. Root length of AA treated *Arabidopsis*.**

Wild type *A. thaliana* was grown on MS media with and without 250 µM **AA**. The lengths of roots were measured at four different time points after seed germination. The plot shows mean values ± s.d. (error bars); n=18 biologically independent samples.



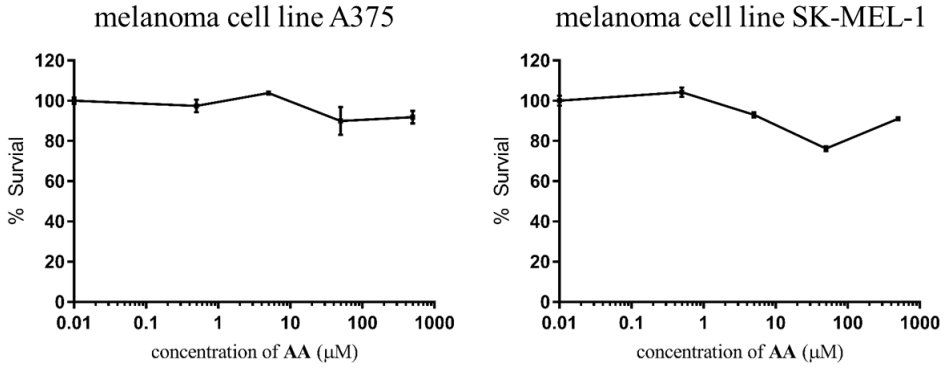

**Supplementary Figure 7. SDS-PAGE analysis of purified proteins.**

SDS-PAGE analysis of purified pDHAD (**a**), fDHAD (**b**) and AstD (**c**) from *E. coli* BL21 (DE3). The experiments were repeated independently with similar results for 5 times.



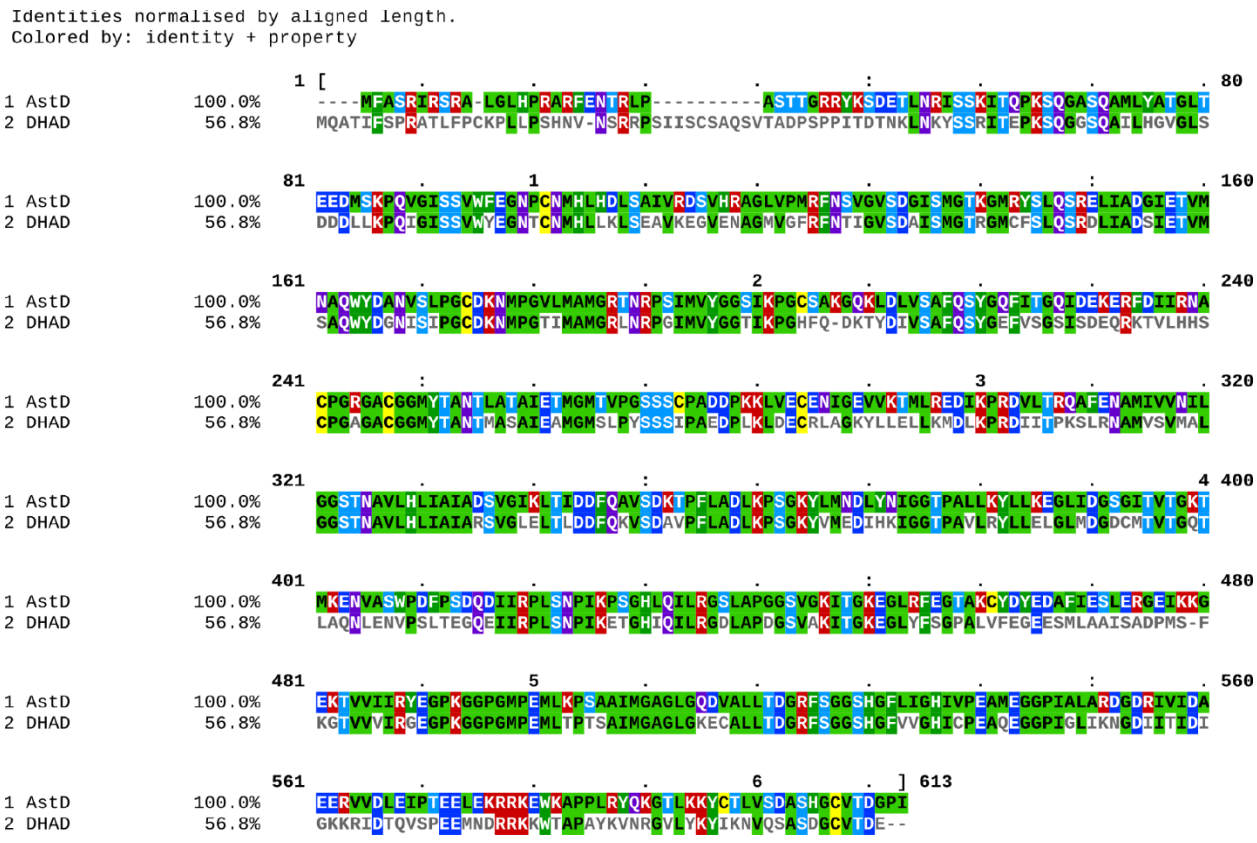
**Supplementary Figure 8. Cytotoxicity assay of AA.**

Percent growth inhibition of melanoma cell line A375 (left) and SK-MEL-1 (right) indicate **AA** has no significant cytotoxicity. Treatments of **AA** was initiated at 24 h postseeding for 72 h, cell viability was measured by CellTiter-GLO Luminescence (Promega) following the manufacturer's recommendations. Results are representative data in duplicate from three independent experiments (center values are averages, errors bars are s.d., n = 5 biologically independent experiments, error bars were determined from independent biological replicates).



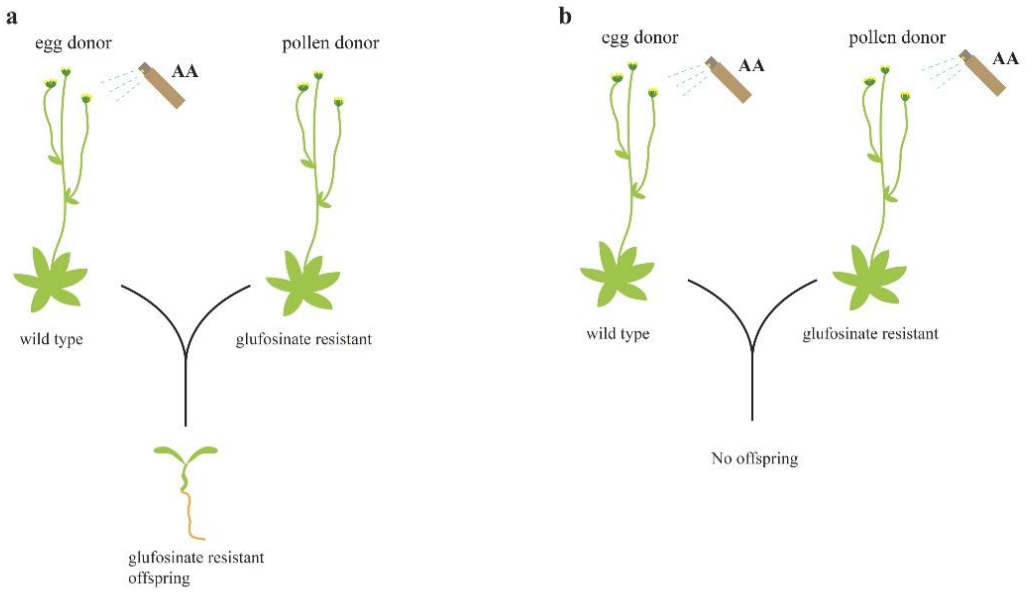
**Supplementary Figure 9. Sequence alignment between pDHAD and AstD.**

The sequence identity between pDHAD and AstD is 56.8%, whereas the similarity between them is 75.0%. Residues were colored according to their property and similarity.



**Supplementary Figure 10. Schematic illustration of results from the cross experiment.**

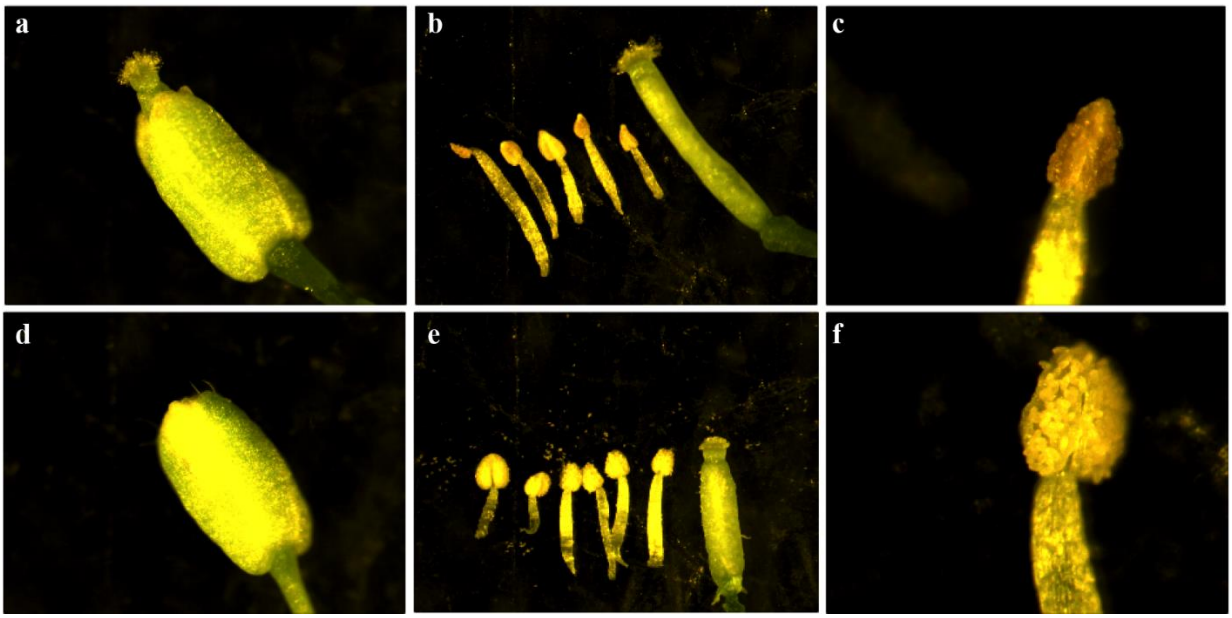
**a,** Wild type *A. thaliana* treated with 250 µM **AA** was pollinated with pollen from the un-treated plant that carries the glufosinate resistant gene. Offspring was obtained, and inherited the glufosinate resistance from the pollen donor. **b,** similar as in **a**, except that the pollen donor was also treated with 250 µM **AA**. No offspring was obtained from this cross. Similar results were obtained with the treatment of **AA** at 100 µM.




| female parent | male parent | offspring obtained | inherit resistance |
|---|---|---|---|
| **AA** treated wild type | un-treated Glufosinate resistant plant | Yes | Yes |
| **AA** treated wild type | **AA** treated glufosinate resistant plant | No | N/A |

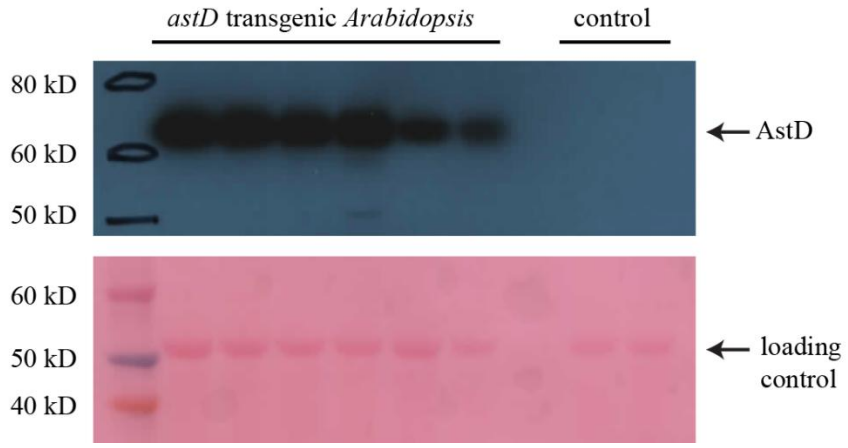
**Supplementary Figure 11. Specific inhibition of anther development in *A. thaliana*.**

Comparison of flower organs between the **AA** treated (**a-c**) and non-treated (**d-f**) *Arabidopsis*. **a** compare to **d**, the **AA** treated flower shows abnormal pistil elongation due to the lack of pollination. **b** compare to **e**, the **AA** treated flower is missing one stamen. **c** compare to **f**, the **AA** treated anther is depleted of healthy and mature pollen. The experiments were performed twice with similar results.



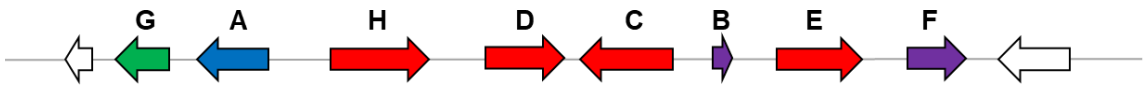
**Supplementary Figure 12. Verification of AstD expression in A. thaliana using western blot.**

Western blot verification of AstD expression in *A. thaliana* using anti-FLAG antibody (top); Ponceau staining shows equal loading (bottom). Six independent T1 transgenic plants (lane 1 to 6) and two wild-type plants (lane 8, and 9) were assayed. Lane 7 does not contain any sample. The experiment was not repeated.



## 4.3 Supplementary tables

**Supplementary Table 1. Heptelidic acid biosynthetic gene cluster of *T. virens*.**




*Trichoderma virens* Gv29-8, scaffold 92 (NW_014013657.1, 1,831,400-1,858,00), 26.6 kb

| gene | accession number | size (gene/protein) | BLASTP homologs | identity/similarity (%) | proposed function |
|---|---|---|---|---|---|
| *hepA* | XP_013949969.1 | 1424/377 | GAP85091.2 | 59/72 | terpene synthase |
| *hepH* | XP_013949970.1 | 1971/534 | EIT74228.1 | 66/80 | cytochrome P450 |
| *hepD* | XP_013949971.1 | 1730/512 | BAJ04474.1 | 59/75 | cytochrome P450 |
| *hepC* | XP_013949972.1 | 1759/438 | RAQ59865.1 | 58/74 | cytochrome P450 |
| *hepE* | XP_013949974.1 | 1847/528 | OOO06875.1 | 70/83 | cytochrome P450 |
| *hepB* | XP_013949973.1 | 456/115 | OOO06872.1 | 39/58 | antibiotic biosynthesis monooxygenase |
| *hepG* | XP_013949968.1 | 1147/336 | OAA55922.1 | 78/88 | GAPDH |
| *hepF* | XP_013949975.1 | 1572/503 | XP_008081852.1 | 59/75 | MFS multidrug transporters |

**Proposed functions of genes within the *ast* cluster in *A. terreus***


*A. terreus* NIH 2624, scaffold 6 (NT_165929.1, 469,00-486,00), 17 kb

| Gene | Accession number | Size (gene/protein) | BLASTP homologs | Identity/similarity (%) | Putative function |
|---|---|---|---|---|---|
| *astA* | XP_001213594.1 | 1230/409 | XP_001266526.1 | 94/97 | Terpene synthase |
| *astB* | XP_001213595.1 | 1760/512 | XP_001266527.1 | 94/96 | Cytochrome P450 |
| *astC* | XP_001213596.1 | 1716/538 | CEJ61176.1 | 84/89 | Cytochrome P450 |
| *astD* | XP_001213593.1 | 1874/598 | OJJ72940.1 | 98/98 | Dihydroxy-acid dehydratase |

**Supplementary Table 2. Microbial strains used in this study**

| strain | genotype | source |
|---|---|---|
| Fungi | | |
| *Trichoderma virens* Gv29-8 | wild type | NRRL |
| *Aspergillus terreus* NIH2624 | wild type | FGSC |
| *Aspergillus nidulans* A1145 | *ΔpyrG*, *ΔpyroA*, *ΔriboB* | [171] |
| TY01 | *Aspergillus nidulans* A1145 carrying AstD+AstA-pYTU, AstB-pYTR, AstC-pYTP | this study |
| *Saccharomyces cerevisiae* | | |
| RC01 | *MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 ura3-52::atCPR prb1 Δ1.6R can1 GAL* | [124] |
| TY11 | RC01 carrying pHepA-xw55 | this study |
| TY12 | RC01 carrying pHepH-xw06 & TvCPR-xw55 | this study |
| TY13 | RC01 carrying pHepD-xw02 & TvCPR-xw55 | this study |
| TY14 | RC01 carrying pHepC-xw06 & TvCPR-xw55 | this study |
| TY15 | RC01 carrying pHepE-xw02 & TvCPR-xw55 | this study |
| TY16 | BL21 (DE3) carrying HepG-pET28a | this study |
| TY17 | BL21 (DE3) carrying tGAPDH-pET28a | this study |
| TY18 | BL21 (DE3) carrying hGAPDH-pSJ2 | this study |
| TY02 | RC01 carrying pAstA-xw55 | this study |
| TY03 | TY02 carrying pAstB-xw06 | this study |
| TY04 | TY03 carrying pAstC-xw02 | this study |
| DHY *ΔURA3* | *MATα ura3Δ0* | this study |
| UB01 | DHY *ΔURA3 ilv3::URA3* | this study |
| UB02 | DHY *ΔURA3 ΔILV3* | this study |
| TY05 | UB02 carrying pXP318 | this study |
| TY06 | UB02 carrying fDHAD-pXP318 | this study |
| TY07 | UB02 carrying AstD-pXP318 | this study |
| *Escherichia coli* | | |
| DH10β | | NEB |
| BL21 (DE3) | | NEB |
| TY08 | BL21 (DE3) carrying AstD-pET28a | this study |
| TY09 | BL21 (DE3) carrying pDHAD-pET28a | this study |
| TY10 | BL21 (DE3) carrying fDHAD-pET28a | this study |

**Supplementary Table 3. Primers for PCR amplification in this study**

| primer | sequences of primer (5'→3') |
|---|---|
| HepA-xw55-recomb-F | tatggctagcgattataaggatgatgatgataagactagtatggctcaagtccagcgtgc |
| HepA-xw55-recomb-R | caaatttgtcatttaaattagtgatggtgatggtgatgcacagcagcgtgattcaccacg |
| HepH-xw06-recomb-F | tcaactatcaactattaactatatcgtaataccatatgatggagcaactcaaaactcggc |
| HepH-xw06-recomb-R | gtggtggtgactcgcgacctcatacacaagcttctaatccgaaacccacctcttattgtc |
| HepD-xw02-recomb-F | tatcaactattaactatatcgtaataccatatgatgtcttcgtttacatcccctgacatc |
| HepD-xw02-recomb-R | gataatgaaaactataaatcgtgaaggcatgtttaaacctaattctttcggggaatcatc |
| HepC-xw06-recomb-F | actatcaactattaactatatcgtaataccatATGatgctcgcatctgtccaatcccttg |
| HepC-xw06-recomb-R | tggtggtggtgactcgcgacctcatacacaagctttcagtgtgaaacattccaggtggag |
| HepE-xw02-recomb-F | caactatcaactattaactatatcgtaataccatatgatggacaccttcaatgccactcc |
| HepE-xw02-recomb-R | tgataatgaaaactataaatcgtgaaggcatgtttaaactcaccactcatccttccgtcg |
| TvCPR-xw55-recomb-F | tacaatcaactatcaactattaactatatcgtaataccatatggcggaactggacacg |
| TvCPR-xw55-recomb-R | aaatttgtcatttaaattagtgatggtgatggtgatgcacttatgaccagacatcctcctgg |
| HepG-pET-F | catcacagcagcggcctggtgccgcgcggcagccatatggttcccaaagttggcatcaac |
| HepG-pET-R | tctcagtggtggtggtggtggtctcgagtgcggccgcttactggccggcatccttttg |
| tGAPDH-pET-F | tggtgccgcgcggcagccatatgagtcacaaatcagctacaatggctcccatcaaggtcg |
| tGAPDH-pET-R | gtgctcgagtgcggccgcatctgatcagatcaaacacttatttggaggcatcgaccttgg |
| | |
| AstD-pYTU-recomb-F | gagagcctgagcttcatccccagcatcattacacctcagcaatgttcgcgtcgaggatcc |
| AstA-pYTU-recomb-R | gactaaccattaccccgccacatagacacatctaaacaatggacatgaataccttccccg |
| Gpda-pYTU-F | gtggaggacatacccgtaattttctgggcatttaaatactccggtgaattgatttgggtg |
| Gpda-R | tgtttagatgtgtctatgtggcggg |
| AstB-pYTR-recomb-F | aaccattaccccgccacatagacacatctaaacaatgctattccaagacctgtcttttcc |
| AstB-pYTR-recomb-R | gctaaagggtatcatcgaaagggagtcatccaggtactgcttgtattgaatcctagtttg |
| AstC-pYTP-recomb-F | cccttctctgaacaataaaccccacagaaggcatttatgggagcttctactttctcccag |
| AstC-pYTP-recomb-R | caacaaccatgataccaggggatttaaatttaattaaggttggggtttcatgcatatagc |
| AstA-xw55-recomb-F | tggctagcgattataaggatgatgatgataagactagtatggacatgaataccttccccg |
| AstA-xw55-recomb-R | atttgtcatttaaattagtgatggtgatggtgatgcacgtgttatgcgttgcctagcggg |
| AstB-xw06-recomb-F | caactatcaactattaactatatcgtaataccatatgctattccaagacctctcgtttcc |
| AstB-xw06-recomb-R | tacttgataatggaaactataaatcgtgaaggcatctacttgcagagacccataactcgc |
| AstC-xw02-recomb-F | atcaactatcaactattaactatatcgtaataccatatgggagcttctactttctccctg |
| AstC-xw02-recomb-R | ttgataatgaaaactataaatcgtgaaggcatgtttaaacctagcctcgtctctttattc |
| pDHAD-pET-F | atagctagcatgcaagccaccatcttctctcc |
| pDHAD-pET-R | atagcggccgcttactcgtcagtcacacatccatctg |
| fDHAD-pET-F | atacatatgcttctctctcagacccgg |
| fDHAD-pET-R | atagcggccgcttagtcaagagcatcggtgatgcag |
| AstD-pET-F | atacatatgttcgcgtcgaggatcc |
| AstD-pET-R | atagcggccgcctagatcggtccgtccgtgac |
| fDHAD-pXP318-F | gcatagcaatctaatctaagttttaattacaaaactagtatgcttctctctcagacccgg |

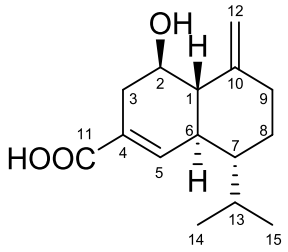
| | |
|---|---|
| fDHAD-pXP318-R | gaatgtaagcgtgacataactaattacatgactcgagttagtcaagagcatcggtgatgc |
| AstD-pXP318-F | tagcaatctaatctaagttttaattacaaaactagtatggactacaaagacgatgacgac |
| AstD-pXP318-R | gcgtgaatgtaagcgtgacataactaattacatgactcgagctagatcggtccgtccgtg |
| DHAD-F | acaggatccgcccaatccgtaaccgc |
| DHAD-R | cacgtcgacttactcgtcagtcacacatccat |
| K559AK560A-F | acataggagcagcaagaatagacacacaagtctcacccg |
| K559AK560A-R | gtctattcttgctgctcctatgtcaatggtgattatgtctc |

**Supplementary Table 4. Plasmids used in this study**

| plasmids | features | source |
|---|---|---|
| pHepA-xw55 | pXW55 expressing *hepA* | this study |
| pHepH-xw06 | pXW06 expressing *hepH* | this study |
| pHepD-xw02 | pXW02 expressing *hepD* | this study |
| pHepC-xw06 | pXW06 expressing *hepC* | this study |
| pHepE-xw02 | pXW02 expressing *hepE* | this study |
| pTvCPR-xw55 | pXW55 expressing *Trichoderma virens* cytochrome P450 reductase | this study |
| HepG-pET | pET28a expressing HepG | this study |
| tGAPDH-pET | pET28a expressing *Trichoderma virens* GAPDH | this study |
| hGAPDH-pSJ2 | pSJ2 expressing *homo sapiens* GAPDH | this study |
| | | |
| pYTU | protein expression vector in *A. nidulans* (*pyrG* marker) | [171] |
| pYTR | protein expression vector in *A. nidulans* (*riboB* marker) | [171] |
| pYTP | protein expression in *A. nidulans* (*pyroA* marker) | [171] |
| pAstD+AstA-pYTU | pYTU expressing *astA* and *astD* | this study |
| pAstB-pYTR | pYTR expressing *astB* | this study |
| pAstC-pYTP | pYTP expressing *astC* | this study |
| pXW55 | protein expression vector in *S. cerevisiae* (*URA3* marker) | [124] |
| pXW06 | protein expression vector in *S. cerevisiae* (*TRP2* marker) | [124] |
| pXW02 | protein expression vector in *S. cerevisiae* (*LEU2* marker) | [124] |
| pAstA-xw55 | pXW55 expressing *astA* | this study |
| pAstB-xw06 | pXW06 expressing *astB* | this study |
| pAstC-xw02 | pXW02 expressing *astC* | this study |
| pET28a | protein expression vector in *E. coli* BL21 (DE3) | Addgene |
| pDHAD-pET | pET28a expressing pDHAD | this study |
| fDHAD-pET | pET28a expressing fDHAD | this study |
| AstD-pET | pET28a expressing AstD | this study |
| pXP318 | protein expression vector in *S. cerevisiae* (*URA3* marker) | Addgene |
| fDHAD-pXP318 | pXP318 expressing fDHAD | this study |
| AstD-pXP318 | pXP318 expressing AstD | this study |
| pEG202 | protein expression vector in *A. thaliana* ($blp^R$ marker) | Addgene |
| pAstDo-pEG | pEG202 expressing codon optimized AstD | this study |

**Supplementary Table 5. NMR data and structure**

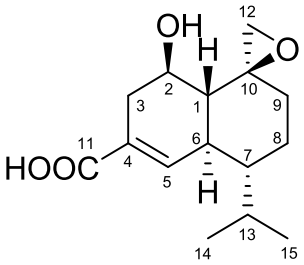
**$^{1}$H (500 MHz, $CDCl_3$) and $^{13}$C NMR (125 MHz, $CDCl_3$) of compound 2:**




**2**

| no. | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ | mult. | HMBC |
|---|---|---|---|---|
| 1 | 1.97 (1H, m) | 51.2 | CH | 149.5, 141.1, 66.7, 46.4, 45.9, 32.3 |
| 2 | 4.09 (1H, q, 9.3) | 66.7 | CH | |
| 3 | 2.92 (1H, dd, 17.4, 5.8) | 32.3 | $CH_2$ | 171.9, 141.1, 127.6, 66.7, 51.2 |
| 3’ | 2.15 (1H, m) | | | 141.1, 127.6, 66.7 |
| 4 | - | 127.6 | C | - |
| 5 | 7.14 (1H, s) | 141.1 | CH | 171.9, 127.6, 51.2, 46.4, 32.3 |
| 6 | 1.97 (1H, m) | 45.9 | CH | 149.5, 141.1, 127.6, 66.7, 51.2, 46.4, 27.3 |
| 7 | 1.44 (1H, m) | 46.4 | CH | |
| 8 | 1.87 (1H, d, 12.5) | 27.3 | $CH_2$ | 149.5, 45.9 ,36.9 |
| 8’ | 1.20 (1H, m) | | | 45.9 ,36.9 |
| 9 | 2.41 (1H, d, 12.0) | 36.9 | $CH_2$ | 149.5, 104.8, 51.2, 46.4, 27.3 |
| 9’ | 2.02 (1H, m) | | | 149.5, 104.8, 51.2, 46.4, 27.3 |
| 10 | - | 149.5 | C | - |
| 11 | - | 171.9 | C | - |
| 12 | 4.88 (2H, d, 35.3) | 104.8 | $CH_2$ | 149.5, 51.2, 45.9, 36.9, 27.3 |
| 13 | 2.20 (1H, m) | 26.8 | CH | 45.9, 27.3, 21.3, 15.1 |
| 14 | 0.94 (3H, d, 6.8) | 21.3 | $CH_3$ | 46.4, 26.8, 15.1 |
| 15 | 0.77 (3H, d, 6.9) | 15.1 | $CH_3$ | 46.4, 26.8, 21.3 |

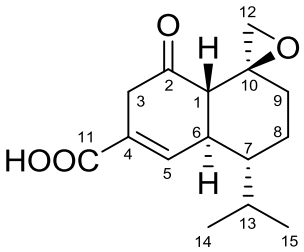
**$^{1}$H (500 MHz, $CDCl_3$) and $^{13}$C NMR (125 MHz, $CDCl_3$) of compound 3:**




**3**

| no. | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ | mult. | HMBC |
|---|---|---|---|---|
| 1 | 1.88 (1H, m) | 45.9 | CH | 68.0, 64.6, 49.8, 45.5, 42.9, 34.8, 32.6 |
| 2 | 3.75 (1H, m) | 68.0 | CH | 64.6, 42.9 |
| 3 | 2.79 (1H, dd, 15.9, 5.4) | 32.6 | $CH_2$ | 171.6, 139.5, 128.2, 68.0, 45.9 |
| 3’ | 2.12 (1H, m) | | | 139.5, 128.2, 68.0, 45.9 |
| 4 | - | 128.2 | C | - |
| 5 | 7.10 (1H, s) | 139.5 | CH | 171.6, 128.2, 64.6, 45.5, 42.9, 32.6 |
| 6 | 2.09 (1H, m) | 42.9 | CH | 45.9, 45.5, 23.9 |
| 7 | 1.36 (1H, m) | 45.5 | CH | 45.9, 34.8 |
| 8 | 1.91 (1H, m) | 23.9 | $CH_2$ | 64.6, 45.5, 42.9, 34.8 |
| 8’ | 1.29 (1H, m) | | | 64.6, 45.5, 42.9, 26.7 |
| 9 | 2.00 (1H, m) | 34.8 | $CH_2$ | 64.6, 49.8, 45.5 |
| 9’ | 1.38 (1H, m) | | | 64.6, 45.5, 23.9 |
| 10 | - | 64.6 | C | - |
| 11 | - | 171.6 | C | - |
| 12 | 3.21 (1H, s) | 49.8 | $CH_2$ | 64.6 |
| 12’ | 2.70 (1H, d, 3.4) | | | 64.6, 34.8 |
| 13 | 2.23 (1H, td, 6.9, 3.0) | 26.7 | CH | 45.5, 42.9, 23.9, 21.3, 15.1 |
| 14 | 0.97 (3H, d, 6.9) | 21.3 | $CH_3$ | 45.5, 26.7, 15.1 |
| 15 | 0.81 (3H, d, 6.9) | 15.1 | $CH_3$ | 45.5, 26.7, 21.3 |

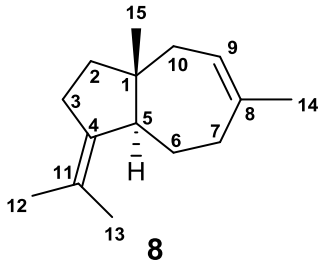
**$^{1}$H (500 MHz, $CDCl_3$) and $^{13}$C NMR (125 MHz, $CDCl_3$) of compound 4:**




**4**

| no. | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ | mult. | HMBC |
|---|---|---|---|---|
| 1 | 2.89 (1H, d, 12.0) | 51.1 | CH | 204.3, 141.9, 58.1, 51.4, 44.2 |
| 2 | - | 204.3 | C | - |
| 3 | 3.27 (1H, dt, 20.2, 2.6) | 41.1 | $CH_2$ | 204.3, 141.9, 127.9 |
| 3’ | 2.96 (1H, dt, 20.1, 2.2) | | | 204.3, 141.9, 127.9 |
| 4 | - | 127.9 | C | - |
| 5 | 7.31 (1H, d, 2.0) | 141.9 | CH | 170.1, 127.9, 51.1, 44.2, 41.1 |
| 6 | 2.57~2.52 (1H, m) | 44.2 | CH | 141.9, 127.9 |
| 7 | 1.51 (1H, tt, 11.7, 3.1) | 46.1 | CH | |
| 8 | 1.95~1.80 (1H, m) | 23.0 | $CH_2$ | 58.1, 46.1 |
| 8’ | 1.36~1.23 (1H, m) | | | 58.1, 46.1 |
| 9 | 1.95~1.80 (1H, m) | 35.5 | $CH_2$ | 58.1, 46.1 |
| 9’ | 1.36~1.23 (1H, m) | | | 58.1, 46.1 |
| 10 | - | 58.1 | C | - |
| 11 | - | 170.1 | C | - |
| 12 | 3.89 (1H, dd, 5.5, 1.5) | 51.4 | $CH_2$ | 58.1, 35.5 |
| 12’ | 2.51 (1H, d, 5.7) | | | 58.1, 35.5 |
| 13 | 2.34~2.22 (1H, m) | 26.9 | CH | 46.1, 21.5, 15.2 |
| 14 | 1.03 (3H, d, 6.9) | 21.5 | $CH_3$ | 46.1, 26.9, 15.2 |
| 15 | 0.83 (3H, d, 6.9) | 15.2 | $CH_3$ | 46.1, 26.9, 21.5 |

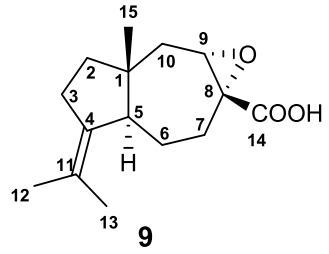
**$^{1}$H (500 MHz, $CDCl_3$) and $^{13}$C NMR (125 MHz, $CDCl_3$) of compound 8:**




**8**

| no. | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ | mult. | HMBC |
|---|---|---|---|---|
| 1 | - | 42.6 | C | - |
| 2 | 1.44 (1H, dd, 11.4, 7.2) | 40.3 | $CH_2$ | 138.4, 57.4, 42.6, 29.6, 19.2 |
| 2’ | 1.31 (1H, dd, 11.3, 2.6) | | | 42.6, 41.4, 29.6, 19.2 |
| 3 | 2.20 (1H, m) | 29.6 | $CH_2$ | 138.4, 42.6, 34.5 |
| 3’ | 2.15 (1H, m) | | | 138.4, 122.2, 57.4, 42.6, 40.3 |
| 4 | - | 138.4 | C | - |
| 5 | 2.16 (1H, m) | 57.4 | CH | 138.4, 42.6, 40.3, 34.5, 25.0 |
| 6 | 2.19 (1H, m) | 25.0 | $CH_2$ | 138.4,138.3, 57.4, 42.6, 34.5 |
| 6’ | 1.36 (1H, m) | | | 34.5 |
| 7 | 2.15 (1H, m) | 34.5 | $CH_2$ | 138.3, 122.4, 57.4 |
| 7’ | 2.07 (1H, m) | | | 138.3, 122.4, 57.4, 27.3, 25.0 |
| 8 | - | 138.3 | C | - |
| 9 | 5.37 (1H, m) | 122.4 | CH | |
| 10 | 2.00 (1H, m) | 41.4 | $CH_2$ | 138.3, 122.4, 57.4, 42.6, 40.3, 19.2 |
| 10’ | 1.95 (1H, d, 15.3) | | | 138.3, 122.4, 57.4, 42.6 |
| 11 | - | 122.2 | C | - |
| 12 | 1.61 (3H, brs) | 23.3 | $CH_3$ | 138.4, 122.2, 20.6 |
| 13 | 1.71 (3H, q, 1.7) | 20.6 | $CH_3$ | 138.4, 122.2, 23.3 |
| 14 | 1.75 (3H, s) | 27.3 | $CH_3$ | 138.3, 122.4, 34.5 |
| 15 | 0.73 (3H, s) | 19.2 | $CH_3$ | 57.4, 42.6, 41.4, 40.3 |

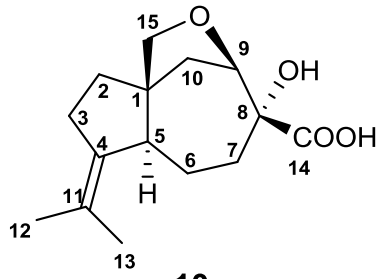
**$^{1}$H (500 MHz, $CDCl_3$) and $^{13}$C NMR (125 MHz, $CDCl_3$) of compound 9:**




**9**

| no. | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ | mult. | HMBC |
|---|---|---|---|---|
| 1 | - | 44.1 | C | - |
| 2 | 1.41 (1H, m) | 38.8 | $CH_2$ | 135.8, 60.1, 44.1, 29.3, 17.8 |
| 2’ | 1.31 (1H, td, 11.5, 9.0) | | | 44.1, 40.5, 29.3, 17.8 |
| 3 | 2.26 (1H, m) | 29.3 | $CH_2$ | 135.8 |
| 3’ | 2.15 (1H, dd, 16.3, 8.9) | | | 135.8, 123.2, 60.1, 44.1, 38.8 |
| 4 | - | 135.8 | C | - |
| 5 | 2.08 (1H, d, 12.0) | 60.1 | CH | |
| 6 | 2.48 (1H, dd, 14.8, 6.3) | 24.9 | $CH_2$ | 59.4, 44.1, 30.6 |
| 6’ | 1.84 (1H, q, 13.1) | | | 59.4, 30.6 |
| 7 | 2.71 (1H, dd, 14.6, 6.5) | 30.6 | $CH_2$ | 176.0, 60.1, 59.8, 59.4, 24.9 |
| 7’ | 1.39 (1H, m) | | | 176.0, 60.1, 59.8, 59.4, 24.9 |
| 8 | - | 59.4 | C | - |
| 9 | 3.25 (1H, t, 7.4) | 59.8 | CH | 176.0, 59.4, 40.5 |
| 10 | 2.36 (1H, dd, 14.0, 6.6) | 40.5 | $CH_2$ | 60.1, 59.8, 59.4, 44.1, 38.8 |
| 10’ | 1.44 (1H, m) | | | 60.1, 59.8, 59.4, 44.1, 17.8 |
| 11 | - | 123.2 | C | - |
| 12 | 1.59 (3H, d, 2.2) | 23.8 | $CH_3$ | 135.8, 123.2, 20.6 |
| 13 | 1.73 (3H, d, 2.3) | 20.6 | $CH_3$ | 135.8, 123.2, 23.8 |
| 14 | - | 176.0 | C | - |
| 15 | 0.86 (3H, s) | 17.8 | $CH_3$ | 59.8, 44.1, 40.5, 38.8 |
| 14-COOH | 8.09 (1H, brs) | - | COOH | |

**$^{1}$H (500 MHz, $CDCl_3$) and $^{13}$C NMR (125 MHz, $CDCl_3$) of compound 10 (AA):**




**10**

| no. | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ | mult. | HMBC |
|---|---|---|---|---|
| 1 | - | 53.0 | C | - |
| 2 | 1.73 (1H, m) | 33.8 | $CH_2$ | 134.5, 76.3, 53.0, 23.6 |
| 2’ | 1.50 (1H, m) | | | 134.5, 76.3, 55.4, 53.0, 23.6 |
| 3 | 2.42 (1H, dd, 14.9, 7.3) | 23.6 | $CH_2$ | 76.3, 55.4, 53.0, 33.8 |
| 3’ | 1.61 (1H, m) | | | 134.5, 55.4, 53.0, 33.8 |
| 4 | - | 134.5 | C | - |
| 5 | 2.34 (1H, m) | 55.4 | CH | 134.5, 125.2, 76.3, 53.0, 33.8, 23.6 |
| 6 | 2.20 (1H, m) | 36.6 | $CH_2$ | 75.6, 55.4, 53.0 |
| 6’ | 1.70 (1H, m) | | | 75.6, 53.0 |
| 7 | 2.32 (1H, m) | 32.2 | $CH_2$ | 178.2, 82.9, 75.6, 55.4 |
| 7’ | 2.01 (1H, m) | | | 75.6, 55.4 |
| 8 | - | 75.6 | C | - |
| 9 | 4.29 (1H, d, 8.5) | 82.9 | CH | 76.3, 75.6, 53.0, 36.2 |
| 10 | 2.26 (1H, m) | 36.2 | $CH_2$ | 82.9, 76.3, 75.6, 55.4 |
| 10’ | 2.12 (1H, d, 13.4) | | | 76.3, 75.6, 55.4, 53.0 |
| 11 | - | 125.2 | C | - |
| 12 | 1.71 (3H, s) | 20.9 | $CH_3$ | 134.5, 125.2, 23.4 |
| 13 | 1.60 (3H, s) | 23.4 | $CH_3$ | 134.5, 125.2, 20.9 |
| 14 | - | 178.2 | C | - |
| 15 | 3.92 (1H, d, 8.3) | 76.3 | $CH_2$ | 82.9, 55.4, 53.0, 36.2 |
| 15’ | 3.48 (1H, d, 8.3) | | | 55.4, 53.0, 33.8 |

**Supplementary Table 6. Data collection and refinement statistics (molecular replacement)**

| | *holo*-pDHAD |
|---|---|
| **Data collection** | |
| Space group | $P4_22_12$ |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 135.5, 135.5, 66.0 |
| α, β, γ (°) | 90, 90, 90 |
| Resolution (Å) | 47.89-2.11 (2.15-2.11) * |
| $R_{sym}$ or $R_{merge}$ | 0.189 (1.240) |
| $I$ / σ$I$ | 17.86 (2.33) |
| Completeness (%) | 100 (100) |
| Redundancy | 25.1 (23.1) |
| **Refinement** | |
| Resolution (Å) | 95.79-2.11 |
| No. reflections | 33235 (1714) |
| $R_{work}$ / $R_{free}$ | 0.1767 / 0.2216 |
| No. atoms | |
| Protein | 4208 |
| Ligand/ion | 24 |
| Water | 118 |
| *B*-factors | |
| Protein | 26.60 |
| Ligand/ion | 46.53 |
| Water | 26.22 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.007 |
| Bond angles (°) | 1.195 |

## 5. References


1 Dias, D. A., Urban, S. & Roessner, U. A historical over view of natural products in drug discovery. *Metabolites* **2**, 303-336 (2012).

2 Newman, D. J., Cragg, G. M. & Snader, K. M. The influence of natural products upon drug discovery. *Nat. Prod. Rep.* **17**, 215-234 (2000).

3 Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629-661 (2016).

4 Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829-837 (2004).

5 Shen, B. A new golden age of natural products drug discovery. *Cell* **163**, 1297-1300 (2015).

6 Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.* **8**, 423-435 (2010).

7 Miller, E. L. The penicillins: a review and update *J. Midwifery Women. Health* **47**, 426-434 (2002).

8 Waxman, D. J. & Strominger, J. L. Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. *Annu. Rev. Biochem.* **52**, 825-869 (1983).

9 Cantrell, C. L., Dayan, F. E. & Duke, S. O. Natural products as sources for new pesticides. *J. Nat. Prod.* **75**, 1231-1242 (2012).

10 Dayan, F. E., Owens, D. K. & Duke, S. O. Rationale for a natural products approach to herbicide discovery. *Pest Manag. Sci.* **68**, 519-528 (2012).

11 Duke, S. O. *et al.* Chemicals from nature for weed management. *Weed Sci.* **50**, 138-151 (2017).

12 Watanakunakorn, C. Mode of action and *in-vitro* activity of vancomycin. *J. Antimicrob. Chemother.* **14**, 7-18 (1984).

13 Weaver, B. A. How taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **25**, 2677-2681 (2014).

14 Li, J., Kim, Sang G. & Blenis, J. Rapamycin: one drug, many effects. *Cell Metab.* **19**, 373-379 (2014).

15 Kirst, H. A. The spinosyn family of insecticides: realizing the potential of natural products research. *J. Antibiot.* **63**, 101-111 (2010).

16 Wolstenholme, A. J. & Rogers, A. T. Glutamate-gated chloride channels and the mode of action of the avermectin/milbemycin anthelmintics. *Parasitology* **131**, S85-S95 (2005).

17 Hoerlein, G. Glufosinate (phosphinothricin), a natural amino acid with unexpected herbicidal properties. *Rev. Environ. Contam. Toxicol.* **138**, 73-145 (1994).

18 Huang, L.-S., Cobessi, D., Tung, E. Y. & Berry, E. A. Binding of the respiratory chain inhibitor antimycin to the mitochondrial $bc_1$ complex: a new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. *J. Mol. Biol.* **351**, 573-597 (2005).

19 Metcalf, W. W. & Donk, W. A. v. d. Biosynthesis of phosphonic and phosphinic acid natural products. *Annu. Rev. Biochem.* **78**, 65-94 (2009).

20 Dayan, F. E. & Duke, S. O. Natural compounds as next-generation herbicides. *Plant Physiol.* **166**, 1090-1105 (2014).

21 Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468-3496 (2006).

22 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160 (2013).

23 Christianson, D. W. Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* **117**, 11570-11648 (2017).

24 Khosla, C. Structures and mechanisms of polyketide synthases. *J. Org. Chem.* **74**, 6416-6420 (2009).

25 Chaitan Khosla, Rajesh S. Gokhale, John R. Jacobsen & Cane, D. E. Tolerance and specificity of polyketide synthases. *Annu. Rev. Biochem.* **68**, 219-253 (1999).

26 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688-4716 (2009).
27 Wakil, S. J. Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* **28**, 4523-4530 (1989).
28 Smith, S. & Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041-1072 (2007).
29 Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A. & Cane, D. E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* **76**, 195-221 (2007).
30 Byers, D. M. & Gong, H. Acyl carrier protein: structure–function relationships in a conserved multifunctional protein family. *Biochem. Cell Biol.* **85**, 649-662 (2007).
31 Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31**, 61-108 (2014).
32 Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **52**, 1839-1841 (2013).
33 Dunn, B. J., Watts, K. R., Robbins, T., Cane, D. E. & Khosla, C. Comparative analysis of the substrate specificity of *trans*- versus *cis*-acyltransferases of assembly line polyketide synthases. *Biochemistry* **53**, 3796-3806 (2014).
34 Keating, T. A. & Walsh, C. T. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Curr. Opin. Chem. Biol.* **3**, 598-606 (1999).
35 Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737-748 (2006).
36 Gokhale, R. S., Hunziker, D., Cane, D. E. & Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **6**, 117-125 (1999).
37 Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162-190 (2007).
38 Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285-295 (2003).
39 Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234-240 (2010).
40 Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**, 453-488 (2004).
41 Hubbard, B. K. & Walsh, C. T. Vancomycin assembly: nature's way. *Angew. Chem. Int. Ed. Engl.* **42**, 730-765 (2003).
42 Rausch, C., Huson, D. H., Kohlbacher, O., Weber, T. & Wohlleben, W. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799-5808 (2005).
43 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493-505 (1999).
44 Hur, G. H., Vickery, C. R. & Burkart, M. D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29**, 1074-1098 (2012).
45 Challis, G. L. & Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Struct. Biol.* **14**, 748-756 (2004).
46 Walsh, C. T., O'Brien, R. V. & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed. Engl.* **52**, 7098-7124 (2013).

47 Schneider, T. L., Shen, B. & Walsh, C. T. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**, 9722-9730 (2003).
48 Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci.* **113**, 12432-12437 (2016).
49 Kohli, R. M., Trauger, J. W., Schwarzer, D., Marahiel, M. A. & Walsh, C. T. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* **40**, 7099-7108 (2001).
50 Keating, T. A. *et al.* Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-*S*-enzyme breakdown in antibiotic and siderophore biosynthesis. *ChemBioChem* **2**, 99-107 (2001).
51 Juguet, M. *et al.* An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in Streptomyces ambofaciens. *Chem. Biol.* **16**, 421-431 (2009).
52 Yan, Y. *et al.* Biosynthetic pathway for high structural diversity of a common dilactone core in antimycin production. *Org. Lett.* **14**, 4142-4145 (2012).
53 Yan, Y. *et al.* Multiplexing of combinatorial chemistry in antimycin biosynthesis: expansion of molecular diversity and utility. *Angew. Chem. Int. Ed. Engl.* **52**, 12308-12312 (2013).
54 Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408-414 (2007).
55 Dewick, P. M. The biosynthesis of C5–C25 terpenoid compounds. *Nat. Prod. Rep.* **19**, 181-222 (2002).
56 Ajikumar, P. K. *et al.* Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* **5**, 167-190 (2008).
57 Krishna, S., Bustamante, L., Haynes, R. K. & Staines, H. M. Artemisinins: their growing importance in medicine. *Trends Pharmacol. Sci.* **29**, 520-527 (2008).
58 Caroline, M. S. & Diana, F. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential in the treatment of cancer. *Drugs* **48**, 794-847 (1994).
59 Buhaescu, I. & Izzedine, H. Mevalonate pathway: a review of clinical and therapeutical implications. *Clin. Biochem.* **40**, 575-584 (2007).
60 Rodríguez-Concepción, M. & Boronat, A. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. a metabolic milestone achieved through genomics. *Plant Physiol.* **130**, 1079-1089 (2002).
61 Liang, P.-H., Ko, T.-P. & Wang, A. H.-J. Structure, mechanism and function of prenyltransferases. *Eur. J. Biochem.* **269**, 3339-3354 (2002).
62 Poulter, C. D. Biosynthesis of non-head-to-tail terpenes. Formation of 1'-1 and 1'-3 linkages. *Acc. Chem. Res.* **23**, 70-77 (1990).
63 Tello, M., Kuzuyama, T., Heide, L., Noel, J. P. & Richard, S. B. The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cell. Mol. Life Sci.* **65**, 1459-1463 (2008).
64 Wendt, K. U. & Schulz, G. E. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* **6**, 127-133 (1998).
65 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).
66 Mousa, J. J. & Bruner, S. D. Structural and mechanistic diversity of multidrug transporters. *Nat. Prod. Rep.* **33**, 1255-1267 (2016).
67 Raymond, K. N., Dertz, E. A. & Kim, S. S. Enterobactin: An archetype for microbial iron transport. *Proc. Natl. Acad. Sci.* **100**, 3584-3588 (2003).

68 Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927-937 (1997).
69 Bush, K. Bench-to-bedside review: the role of β-lactamases in antibiotic-resistant Gram-negative infections. *Crit. Care* **14**, 224 (2010).
70 Tao, W. *et al.* Inactivation of chloramphenicol and florfenicol by a novel chloramphenicol hydrolase. *Appl. Environ. Microbiol.* **78**, 6295-6301 (2012).
71 Blochlinger, K. & Diggelmann, H. Hygromycin B phosphotransferase as a selectable marker for DNA transfer experiments with higher eucaryotic cells. *Mol. Cell. Biol.* **4**, 2929-2931 (1984).
72 W V Shaw, a. & Leslie, A. G. W. Chloramphenicol acetyltransferase. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 363-386 (1991).
73 Huttunen, K. M., Raunio, H. & Rautio, J. Prodrugs—from serendipity to rational design. *Pharmacol. Rev.* **63**, 750-771 (2011).
74 Reimer, D. & Bode, H. B. A natural prodrug activation mechanism in the biosynthesis of nonribosomal peptides. *Nat. Prod. Rep.* **31**, 154-159 (2014).
75 Xu, Y. *et al.* Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *J. Am. Chem. Soc.* **134**, 8625-8632 (2012).
76 Galm, U. *et al.* Antitumor antibiotics: bleomycin, enediynes, and mitomycin. *Chem. Rev.* **105**, 739-758 (2005).
77 Liu, W., Christenson, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**, 1170-1173 (2002).
78 Kennedy, J. *et al.* Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368-1372 (1999).
79 Tang, X. *et al.* Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10**, 2841-2849 (2015).
80 Kling, A. *et al.* Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **348**, 1106-1112 (2015).
81 Petrovska, B. Historical review of medicinal plants' usage. *Pharmacog. Rev.* **6**, 1-5 (2012).
82 Yu, F. *et al.* Traditional chinese medicine and kampo: a review from the distant past for the future. *J. Int. Med. Res.* **34**, 231-239 (2006).
83 Henke, M. T. & Kelleher, N. L. Modern mass spectrometry for synthetic biology and structure-based discovery of natural products. *Nat. Prod. Rep.* **33**, 942-950 (2016).
84 McLellan, T. M., Caldwell, J. A. & Lieberman, H. R. A review of caffeine's effects on cognitive, physical and occupational performance. *Neurosci. Biobehav. Rev.* **71**, 294-312 (2016).
85 Benowitz, N. L. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **49**, 57-71 (2009).
86 Siang Yong, T. & Yvonne, T. Alexander Fleming (1881-1955): discoverer of penicillin. *Singapore Med. J.* **56**, 366-367 (2015).
87 Li, J. W.-H. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-165 (2009).
88 Sticher, O. Natural product isolation. *Nat. Prod. Rep.* **25**, 517-554 (2008).
89 Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**, 277-283 (2015).
90 Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155-176 (2016).
91 Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111-129 (2015).
92 Koboldt, Daniel C., Steinberg, Karyn M., Larson, David E., Wilson, Richard K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27-38 (2013).

93 Shendure, J. *et al.* DNA sequencing at 40: past, present and future. *Nature* **550**, 345-353 (2017).
94 Jensen, P. R. Natural products and the gene cluster revolution. *Trends Microbiol.* **24**, 968-977 (2016).
95 Keller, N. P. Translating biosynthetic gene clusters into fungal armor and weaponry. *Nat. Chem. Biol.* **11**, 671-677 (2015).
96 Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988-1005 (2016).
97 Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature* **309**, 462-464 (1984).
98 Medema, M. H. *et al.* Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625-631 (2015).
99 Challis, G. L. Genome mining for novel natural product discovery. *J. Med. Chem.* **51**, 2618-2628 (2008).
100 Unkles, Shiela E., Valiante, V., Mattern, Derek J. & Brakhage, Axel A. Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem. Biol.* **21**, 502-508 (2014).
101 Young, E. & Alper, H. Synthetic biology: tools to design, build, and optimize cellular processes. *J. Biomed. Biotechnol.* **2010** (2010).
102 Jensen, M. K. & Keasling, J. D. Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Res.* **15**, 1-10 (2015).
103 Yan, Y., Liu, Q., Jacobsen, S. E. & Tang, Y. The impact and prospect of natural product discovery in agriculture. *EMBO Rep.* **19**, e46824 (2018).
104 Sasaki, H. *et al.* Thiolactomycin, a new antibiotic. II. structure elucidation. *J. Antibiot.* **35**, 396-400 (1982).
105 Hayashi, T., Yamamoto, O., Sasaki, H., Okazaki, H. & Kawaguchi, A. Inhibition of fatty acid synthesis by the antibiotic thiolactomycin. *J. Antibiot.* **37**, 1456-1461 (1984).
106 Abdalla, M. A. & McGaw, L. J. Natural cyclic peptides as an attractive modality for therapeutics: a mini review. *Molecules* **23**, 2080 (2018).
107 Menzies, D. *et al.* Effect of duration and intermittency of rifampin on *Tuberculosis* treatment outcomes: a systematic review and meta-analysis. *PLOS Med.* **6**, e1000146 (2009).
108 Yasuhiro, I. *et al.* A new sesquiterpene antibiotic, heptelidic acid producing organisms, fermentation, isolation and characterization. *J. Antibiot.* **33**, 468-473 (1980).
109 Yasuhiro, I., Shuji, T., Tatsuo, H. & Mamoru, A. Structure of heptelidic acid, a new sesquiterpene antibiotic from fungi. *J. Antibiot.* **33**, 525-526 (1980).
110 Jun, K. *et al.* Antitumor activity of heptelidic acid chlorohydrin. *J. Antibiot.* **47**, 1562-1563 (1994).
111 Yoshitake, T. *et al.* Antimalarial activity of radicicol, heptelidic acid and other fungal metabolites. *J. Antibiot.* **51**, 153-160 (1998).
112 Sirover, M. A. On the functional diversity of glyceraldehyde-3-phosphate dehydrogenase: biochemical mechanisms and regulatory control. *Biochim. Biophys. Acta* **1810**, 741-751 (2011).
113 Akira, E., Kenji, H., Kenji, H., Kaoru, S. & Tomimi, K. Specific inhibition of glyceraldehyde-3-phosphate dehydrogenase by koningic acid (heptelidic acid). *J. Antibiot.* **38**, 920-925 (1985).
114 Cane, D. E. & Sohng, J.-K. Inhibition of glyceraldehyde-3-phosphate dehydrogenase by pentalenolactone: kinetic and mechanistic studies. *Arch. Biochem. Biophys.* **270**, 50-61 (1989).
115 Cane, D. E. & Sohng, J.-K. Inhibition of glyceraldehyde-3-phosphate dehydrogenase by pentalenolactone. 2. identification of the site of alkylation by tetrahydropentalenolactone. *Biochemistry* **33**, 6524-6530 (1994).
116 Stipanovic, R. D. & Howell, C. R. The X-ray crystal structure determination, and biosynthetic studies of the antibiotic, heptelidic acid. *Tetrahedron* **39**, 1103-1107 (1983).

117 Nishimura, I. & Koyama, Y. Identification of a gene cluster for biosynthesis of the sesquiterpene antibiotic, heptelidic acid, in *Aspergillus oryzae*. *Biosci Biotechnol Biochem.* **22**, 1-8 (2018).
118 Chuang, D.-M., Hough, C. & Senatorov, V. V. Glyceradehyde-3-phosphate dehydrogenase, apoptosis, and neurodegenerative diseases. *Annu. Rev. Pharmacol. Toxicol.* **45**, 269-290 (2004).
119 Kato, M., Sakai, K. & Endo, A. Koningic acid (heptelidic acid) inhibition of glyceraldehyde-3-phosphate dehydrogenases from various sources. *Biochim. Biophys. Acta* **1120**, 113-116 (1992).
120 Gatenby, R. A. & Gillies, R. J. Why do cancers have high aerobic glycolysis? *Nat. Rev. Cancer* **4**, 891-899 (2004).
121 Catalano, V. *et al.* Use of a non-homologous end-joining-deficient strain (delta-*ku70*) of the biocontrol fungus *Trichoderma virens* to investigate the function of the laccase gene *lcc1* in sclerotia degradation. *Curr Genet* **57**, 13-23 (2011).
122 Walsh, C. T. & Wencewicz, T. A. Flavoenzymes: versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* **30**, 175-200 (2013).
123 Kim, T.-W. *et al.* Arabidopsis CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. *Plant Cell* **17**, 2397-2412 (2005).
124 Tang, M.-C. *et al.* Discovery of unclustered fungal indole diterpene biosynthetic pathways through combinatorial pathway reassembly in engineered yeast. *J. Am. Chem. Soc.* **137**, 13724-13727 (2015).
125 Oerke, E. C. Crop losses to pests. *J. Agric. Sci.* **144**, 31-43 (2005).
126 Gianessi, L. P. The increasing importance of herbicides in worldwide crop production. *Pest Manag. Sci.* **69**, 1099-1105 (2013).
127 Gianessi, L. P. & Reigner, N. P. The value of herbicides in U.S. crop production. *Weed Technol* **21**, 559-566 (2007).
128 Heap, I. Global perspective of herbicide-resistant weeds. *Pest Manag. Sci.* **70**, 1306-1315 (2014).
129 Gould, F., Brown, Z. S. & Kuzma, J. Wicked evolution: can we address the sociobiological dilemma of pesticide resistance? *Science* **360**, 728-732 (2018).
130 Burgos, N. R. *et al.* Review: confirmation of resistance to herbicides and evaluation of resistance levels. *Weed Sci.* **61**, 4-20 (2013).
131 Heap, I. *The international survey of herbicide resistant weeds*, <www.weedscience.com> (2019).
132 Soltani, N. *et al.* Potential corn yield losses from weeds in north America. *Weed Technol.* **30**, 979-984 (2017).
133 Swanton, C. J., Harker, K. N. & Anderson, R. L. Crop losses due to weeds in Canada. *Weed Technol.* **7**, 537-542 (2017).
134 Amorim Franco, T. M. & Blanchard, J. S. Bacterial branched-chain amino acid biosynthesis: structures, mechanisms, and drugability. *Biochemistry* **56**, 5849-5865 (2017).
135 Singh, B. K. & Shaner, D. L. Biosynthesis of branched chain amino acids: from test tube to field. *Plant Cell* **7**, 935-944 (1995).
136 Flint, D. H. & Nudelman, A. Studies on the active site of dihydroxy-acid dehydratase. *Bioorg. Chem.* **21**, 367-385 (1993).
137 Aulabaugh, A. & Schloss, J. V. Oxalyl hydroxamates as reaction-intermediate analogs for ketol-acid reductoisomerase. *Biochemistry* **29**, 2824-2830 (1990).
138 Schulz, A., Spönemann, P., Köcher, H. & Wengenmayer, F. The herbicidally active experimental compound Hoe 704 is a potent inhibitor of the enzyme acetolactate reductoisomerase. *FEBS Lett.* **238**, 375-378 (1988).
139 Binder, S., Knill, T. & Schuster, J. Branched-chain amino acid metabolism in higher plants. *Physiol. Plant.* **129**, 68-78 (2007).
140 Duke, Dayan, Romagni & Rimando. Natural products as sources of herbicides: current status and future trends. *Weed Res.* **40**, 99-111 (2000).

141 Bok, J. W. *et al.* Genomic mining for *Aspergillus* natural products. *Chem. Biol.* **13**, 31-37 (2006).
142 Palmer, J. M. & Keller, N. P. Secondary metabolism in fungi: does chromosomal location matter? *Curr. Opin. Microbiol.* **13**, 431-436 (2010).
143 Keller, N. P. Fungal secondary metabolism: regulation, function and drug discovery. *Nat. Rev. Microbiol.* **17**, 167-180 (2019).
144 Regueira, T. B. *et al.* Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. *Appl. Environ. Microbiol.* **77**, 3035-3043 (2011).
145 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).
146 Hendrickson, L. *et al.* Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chem. Biol.* **6**, 429-439 (1999).
147 Clevenger, K. D. *et al.* A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat. Chem. Biol.* **13**, 895-901 (2017).
148 Guo, C.-J. & Wang, C. C. C. Recent advances in genome mining of secondary metabolites in *Aspergillus terreus*. *Front. Microbiol.* **5**, 717 (2014).
149 Cool, L. G. *ent*-Daucane and acorane sesquiterpenes from *Cupressocyparis leylandii* foliage. *Phytochemistry* **58**, 969-972 (2001).
150 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).
151 Shimada, A. *et al.* Aspterric acid and 6-hydroxymellein, inhibitors of pollen development in *Arabidopsis thaliana*, produced by *Aspergillus terreus*. *Zeitschrift für Naturforschung. C, Journal of biosciences* **57**, 459-464 (2002).
152 Flint, D. H. & Emptage, M. H. Dihydroxy acid dehydratase from spinach contains a [2Fe-2S] cluster. *J. Biol. Chem.* **263**, 3558-3564 (1988).
153 Flint, D. H., Emptage, M. H., Finnegan, M. G., Fu, W. & Johnson, M. K. The role and properties of the iron-sulfur cluster in *Escherichia coli* dihydroxy-acid dehydratase. *J. Biol. Chem.* **268**, 14732-14742 (1993).
154 Velasco, J. A. *et al.* Cloning of the dihydroxyacid dehydratase-encoding gene (*ILV3*) from *Saccharomyces cerevisiae*. *Gene* **137**, 179-185 (1993).
155 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).
156 Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417-433 (2010).
157 Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, 42-51 (2014).
158 Selectable markers and reporter genes: a well furnished toolbox for plant science and genetic engineering. *Crit. Rev. Plant Sci.* **31**, 401-453 (2012).
159 Miki, B. & McHugh, S. Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J. Biotechnol.* **107**, 193-232 (2004).
160 Kirkwood, R. C. Use and mode of action of adjuvants for herbicides: A review of some current work. *Pest Manag. Sci.* **38**, 93-102 (1993).
161 Kerler, F., Riederer, M. & Schönherr, J. Non-electrolyte permeability of plant cuticles: A critical evaluation of experimental methods. *Physiol. Plant.* **62**, 599-602 (1984).

162 Xu, L., Zhu, H., Ozkan, H. E., Bagley, W. E. & Krause, C. R. Droplet evaporation and spread on waxy and hairy leaves associated with type and concentration of adjuvants. *Pest Manag. Sci.* **67**, 842-851 (2011).
163 Damak, M., Mahmoudi, S. R., Hyder, M. N. & Varanasi, K. K. Enhancing droplet deposition through in-situ precipitation. *Nat. Comm.* **7**, 12560 (2016).
164 Streiff, K., Blouet, a. & Guckert, A. Hybrid wheat seed production potential using the chemical hybridizing agent SC2053. *Plant Growth Regulation* **21**, 103-108 (1997).
165 Fleury, D. *et al.* Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *J. Exp. Bot.* **64**, 5411-5428 (2013).
166 D. Harold, M. in *Plant Breeding Reviews* (ed Janick Jules) Ch. 3, (1985).
167 Benbrook, C. M. Trends in glyphosate herbicide use in the United States and globally. *Environ. Sci. Eur.* **28**, 3 (2016).
168 Duke, S. O. & Powles, S. B. Glyphosate: a once-in-a-century herbicide. *Pest Manag. Sci.* **64**, 319-325 (2008).
169 Reyes-Prieto, A. & Moustafa, A. Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Sci. Rep.* **2**, 955 (2012).
170 Binder, S. Branched-chain amino acid metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* **8**, e0137 (2010).
171 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).
172 Fang, F. *et al.* A vector set for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast* **28**, 123-136 (2011).
173 Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: a web server for the design of crystallizable protein variants. *Protein Sci.* **16**, 1569-1576 (2007).
174 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).
175 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
176 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol Crystallogr* **67**, 235-242 (2011).
177 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol Crystallogr* **67**, 355-367 (2011).
178 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol Crystallogr* **66**, 486-501 (2010).
179 Harder, E. *et al.* OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput* **12**, 281-296 (2016).
180 Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7**, 2284-2295 (2011).
181 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1-5.6.37 (2016).
182 Greenwood, J. R., Calkins, D., Sullivan, A. P. & Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **24**, 591-604 (2010).
183 Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739-1749 (2004).

184 Yuan, S., Chan, H. C. S., Filipek, S. & Vogel, H. PyMOL and Inkscape bridge the data and the data visualization. *Structure* **24**, 2041-2042 (2016).
185 Yuan, S., Chan, H. C. S. & Hu, Z. Using PyMOL as a platform for computational drug design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **7**, e1298 (2017).
186 Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449-461 (2015).
187 Sirin, S. *et al.* A computational approach to enzyme design: predicting ω-aminotransferase catalytic activity using docking and MM-GBSA scoring. *J. Chem. Inf. Model.* **54**, 2334-2346 (2014).
188 Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743 (1998).
189 Dickens, C. & Schama, S. *A tale of two cities*. (New York: Vintage Books, 1990).