Title
Exploring the programming rules of highly reducing polyketide synthases by analyzing their kinetic parameters

Author
Wang, Bo

Publication Date
2014

UNIVERSITY OF CALIFORNIA

Los Angeles

Exploring the Programming Rules of Highly Reducing Polyketide Synthases by Analyzing Their Kinetic Parameters

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Chemical and Biomolecular Engineering

By

Bo Wang

2014

ABSTRACT OF THE THESIS

Exploring the programming rules of highly reducing polyketide synthases

by analyzing their kinetic parameters

By

Bo Wang

Master of Science in Chemical Engineering

University of California, Los Angeles, 2014

Professor Yi Tang, Chair

Fungal highly reducing polyketide synthases (HR-PKS), which contains three β-keto processing domains: ketoreductase (KR), dehydratase (DH) and enoylreductase (ER), have higher degree of complexity in their biosynthetic programming rules than that of bacterial PKSs. In order to decipher the relationship between protein sequences and the structures of their products, model fungal HR-PKS LovB and its product lovastatin were studied. In each elongation step, LovB chooses specific combinations of modification domains. This thesis aimed at exploring the programming rules in HR-PKS based on kinetic studies. Specifically, methods for quantifying kinetic parameters of the KR and

MT (methyltransferase) domains of HR-PKS were established. Ketoreduction reaction was measured based on NADPH consumption using Plate Reader and methylation reaction was measured according to the desired ions intensity in the positive ionization mode of LC-MS. The fusarielins family was also studied and served as a parallel and complementary system to lovastatin.

The thesis of Bo Wang is approved.

James Liao

Harold Monbouquette

Yi Tang, Committee Chair

University of California, Los Angeles

2014

Table of Contents

List of Tables and Figures

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Professor Yi Tang, for accepting me into his research group and for all the guidance and support during these two years of my graduate studies. I really appreciate his patience and kindness.

Next, I would like to acknowledge Wei Xu who introduced me the experimental techniques and fundamental knowledge involved in my research. He was very helpful to every details of my experiments. I wish him all the best in the future.

Thirdly, I am very grateful to Professor James C. Liao and Professor Harold Monbouquette for their advice and encouragement for my dissertation.

Also, I would like to thank all of my labmates for their help and support these two years. I have a great time with them and the knowledge I've learnt was already beyond research itself.

Lastly, I would like to thank my parents and girlfriend for their love and encouragement. There are many individuals I am greatly appreciating and want to express my thanks to. I would not have completed the dissertation without their great help and guidance.

I Introduction

1 Fungal polyketide synthases

The enzymes that synthesize polyketides are collectively referred to as polyketide synthases (PKSs) and their enzymology is parallel to that of the fatty acid synthases (FASs) [1]. Fungal polyketides (PKs) are the products of fungal polyketide synthases (PKSs). Lots of them, such as lovastatin and penicillin, play an important role in agricultural or pharmaceutical fields [2, 3].

Like bacterial PKSs, the minimal enzyme components involved in chain extension of fungal PKs are also the ketosynthase (KS), malonyl-CoA:ACP transacylase (MAT) and the acyl carrier protein (ACP). Malonyl coenzyme A (malonyl-CoA) often serves as the monomers for PKs biosynthesis. The KS domain is responsible for decarboxylative thio-Claisen condensations and C–C bonds are formed during chain elongation, the MAT domain is used to introduce the extender unit during each cycle of elongation, and the ACP domain serves as the platform for the acyl chain assembling and chain elongation. The carbon at the α-position of the monomers, which is often nucleophilic and is attached to ACP domain, can attack keto group of the extension units which is directly attached to KS domain. This is where decarboxylative thio-Claisen condensations happens and C–C bonds are formed.[4] (See **Figure 1**)

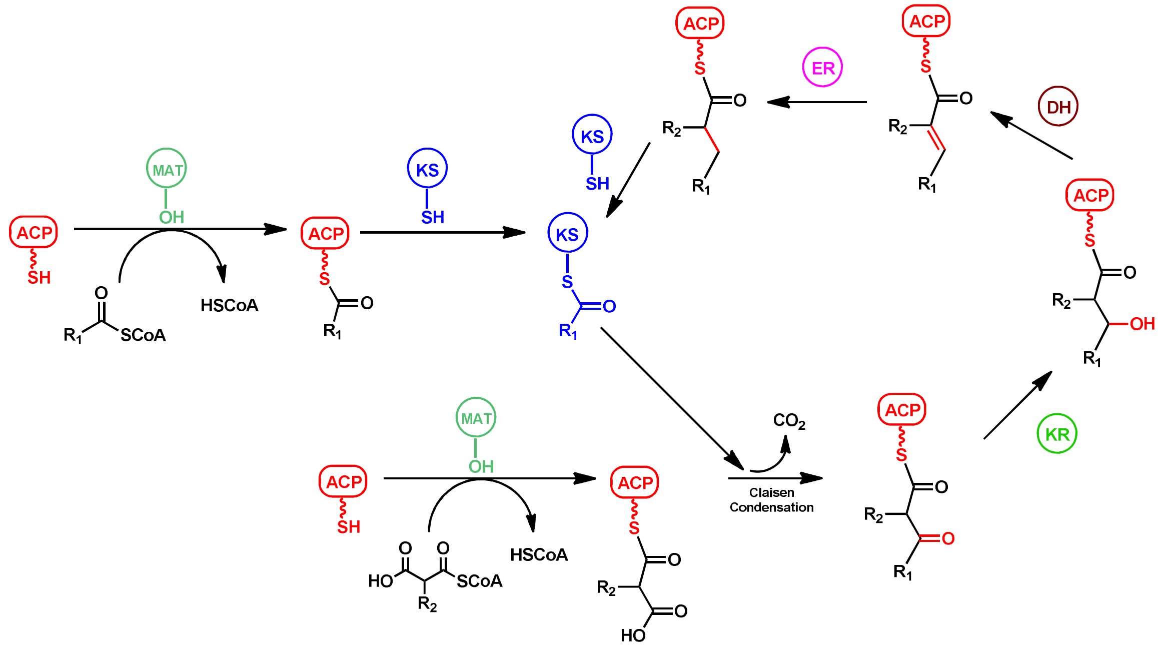


Figure 1. The fundamental reactions of PKS

The PKSs are currently classified into types I, II, and III subgroups. Type I PKSs are megasynthases and their catalytic domains are typically found in a single polypeptide. They can operate in an assembly line fashion. Type II PKSs are composed of mostly dissociated, monofunctional enzymes that function repeatedly in the synthesis of aromatic polyketide compounds. Type III PKSs, which differ from type I and II PKSs, primarily use coenzyme A thioesters directly as substrates without the need for an acyl carrier protein. Fungal PKs are typically biosynthesized by fungal iterative type I PKSs (IPKSs)[5].

Based on the extent of β-keto reduction, IPKSs can be further divided into four groups, which are non-reducing PKSs (NR-PKS), partially reducing PKSs (PR-PKS), highly reducing PKSs (HR-PKS) and highly reducing PKSs-nonribosomal peptide synthetase hybrids (HRPKS-NRPS hybrids, one or two modules of NRPSs are fused to the Cterminal end of a HR-PKS). In contrast to the NR-PKSs, HR-PKSs contains three β-keto processing domains: ketoreductase (KR), dehydratase (DH) and enoylreductase (ER), which introduce a much higher degree of complexity in their biosynthetic programming.

After Claisen condensation where the chain is attached to ACP domain, the next elongation step will happen if the chain is transferred to KS domain freeing the ACP domain for use. The three β-keto processing domains will have functions on the polyketide chain attached to ACP domain before the chain is transferred to KS domain. The HR-PKSs will be responsible for choosing which β-keto processing domain will be used first and which will be used second etc. When the desired chain length is reached, the product can be offloaded from the PKSs via different release mechanisms, such as hydrolysis or macrocyclization catalyzed by a thioesterase (TE, see **Figure 2**) [6].

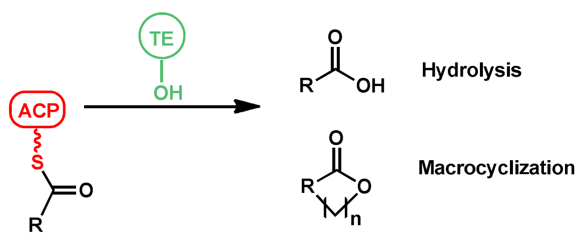


Figure 2. Hydrolysis and macrocyclization catalyzed by TE

2 The programming rules of Type I modular PKSs

Although the function of each domain is well-understood, this is not enough if we want to engineer the protein and get whatever (such as drug-oriented) structure and chain length we want. To get to this point, we need to know the reason why and when each fundamental step will happen. Also, one of the most important things is when the elongation will stop. In fact, we are still quite far from engineering the protein sequence and get whatever structure we want. Sometimes, we even have difficulty in predicting the products of PKSs and even the backbone of polyketides with the protein sequence in hand.

Before we discuss HR-PKSs (iterative PKSs, type I) which have higher degree of complexity in their biosynthetic programming, let's first try to predict the backbone of the products of one modular type I PKS (DEBS) [6, 7]. DEBS has served as the model system to study and engineer modular type I PKSs. From the protein sequence of DEBS, it is clear that DEBS contains one loading module and six extension modules. So, we can predict that the chain elongation happens six times. Including the start unit, the backbone of the products should contain 14 carbons (Monomers often have two carbons for the backbone). What's more, we even can predict the modifications in each module. We clearly know that KR functions 5 times, DH functions once and ER functions once.

In fact, the total biosynthesis of 6-deoxyerythronolide B (6-dEB) and erythromycin, which are the products from DEBS, is the most successful example of using *E. coli* as a heterologous host. 6-dEB is the 14-membered macrocyclic core (just as predicted) of the antibiotic erythromycin synthesized by *Saccharopolyspora erythraea*. The feature of type I modular PKSs has greatly enabled the rational manipulation of the domains and modules towards the generation of structurally altered products. Through the engineering of AT and β-processing domains of DEBS, over 50 different macrolides have been synthesized[8].

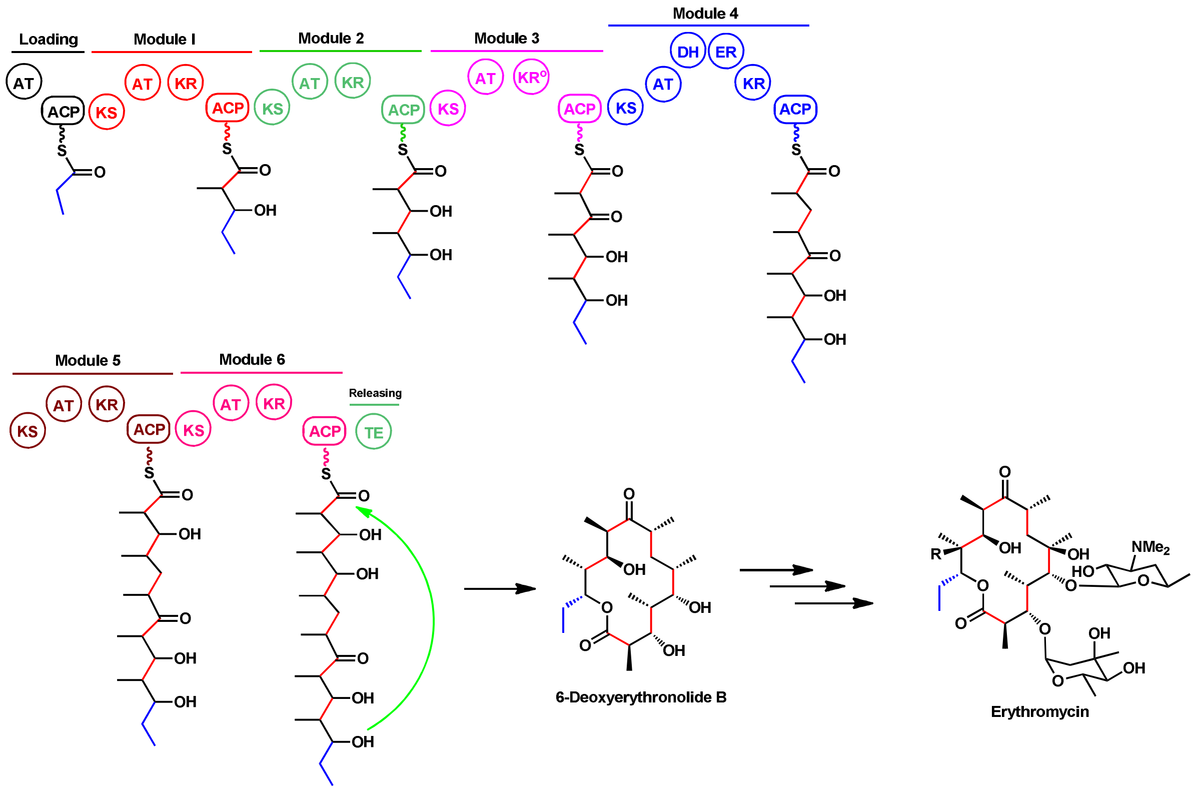


Figure 3. Biosynthetic pathway of 6-deoxyerythronolide B

3 The programming rules of highly reducing iterative polyketide synthase

Then, let's try to predict the backbone of the HR-PKSs' products. Our group reported efficient expression of the lovastatin nonaketide synthase (LovB: HR-PKS), which is a highly reducing iterative polyketide synthase, from an engineered strain of *Saccharomyces cerevisiae*[9]. Its catalytic function was completely reconstituted in the presence and absence of cofactors (NADPH: the reduced form of nicotinamide adenine dinucleotide phosphate and SAM: *S*-adenosylmethionine) and the enoyl reductase LovC.

Based on the protein structure of LovB (**Figure 4**), we would like to determine what the backbone's chain length of the products should be and how many times the KR domain, MT domain or DH domain will be used during the biosynthetic pathway. It seems we cannot answer any of these questions since LovB is not a modular type I PKS but a monomodule megasynthase. Then, will LovB produce many products with random chain length or modifications?

The answer is no. LovB is the key enzyme for the biosynthesis of lovastatin which is a cholesterol-lowering drug from *Aspergillus terreus* [9]. Biosynthesis of lovastatin proceeds via dihydromonacolin L. The chain length and modification steps are controlled accurately and efficiently. From the protein sequence, we cannot even predict the backbone of the products since we know nothing about the chain length and their modifications.

From the proposed mechanism of dihydromonacolin L biosynthesis, we summarize that KR domain was used 8 times, MT domain was used once, DH domain was used 6 times and ER domain was used 3 times (**Figure 4**). The necessity of such complex programming rule is evident in the product of LovB (Dihydromonacolin L) in which the decalin core is formed through the intramolecular Diels–Alder cyclization of a precisely prepared triene at the hexaketide stage[10, 11]. The domains chosen by LovB were so complex and precise. There must be a well-established programming rule involved. Before trying to understand how the chain length or intramolecular cyclization were controlled, it is very helpful to carefully analyze the biosynthetic pathway step by step from the aspects of basic organic chemistry.

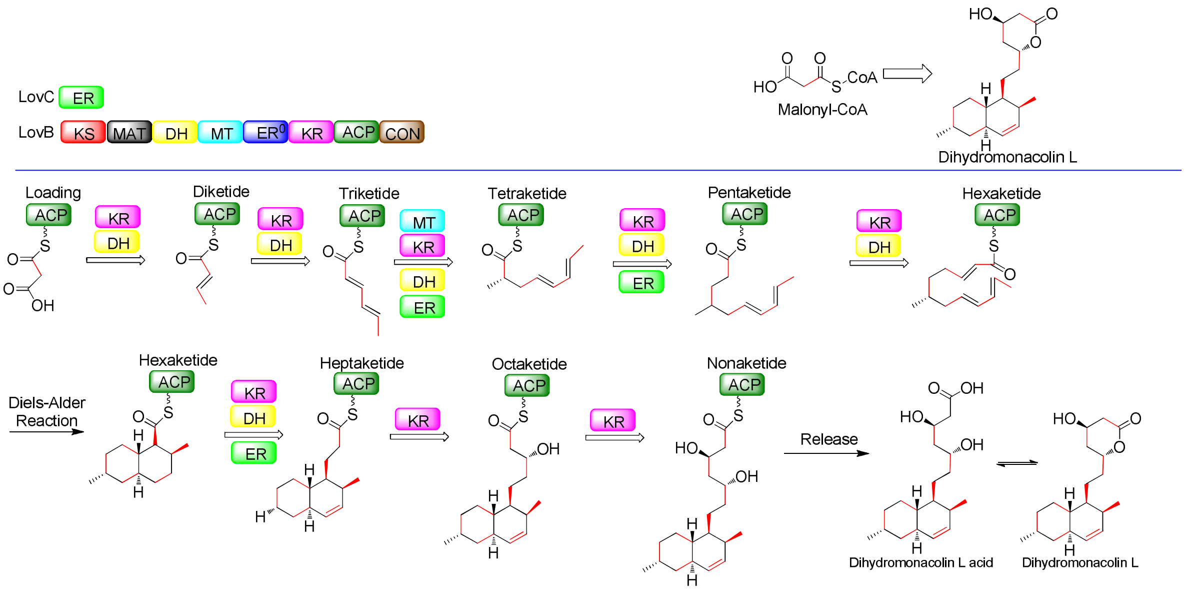
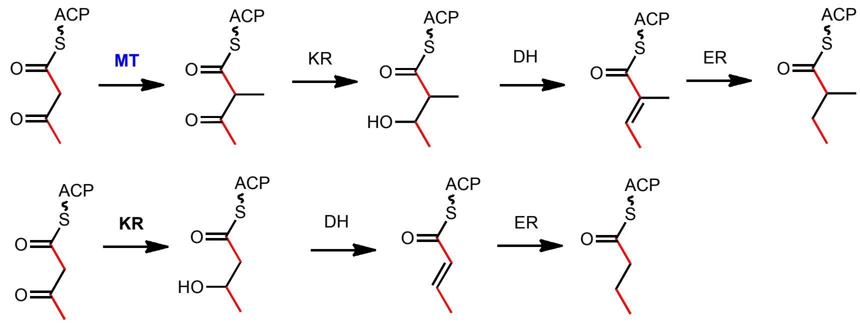


Figure 4. Proposed mechanism of dihydromonacolin L

For each elongation step, there may be four enzymes involved for modifications: MT, KR, DH and ER (**Figure 4**). As shown below, the α-carbon of the starting material is nucleophilic due to the two keto groups nearby. It can attack SAM (cofactor required by MT domain) to get methylated or keto group attached to KS domain to finish elongation. But if one of the keto groups is reduced by KR domain, the nucleophilic activity of the α-carbon was significantly reduced and MT domain may not be active enough to modify the α-position any more. So, it is very important to know if KR domain functions before or after MT domain. If the KR domain functions first, the MT will not show activity in this elongation step.



Therefore, there will not be a methyl group at α-position. From the biosynthetic pathway of dihydromonacolin L, all the modifications for each step begin with KR domain except the third step of elongation (**Figure 4**). So, only during the modification of tetraketide, MT domain functions before KR domain which introduce a methyl group to its α-position. Then, why and how LovB can only choose MT domain to function before KR domain during tetraketide modifications since both KR and MT domain are present. Can LovB use the MT domain like a switch (Turn it on only during the modification of tetraketide) or does the MT domain still have function on other intermediates like diketide and triketide? The mimic of intermediates of lovastatin could be used for further study[12]. Lots of questions remain to be answered.

4 The relationship between lovastatin and fusarielins family

Besides lovastatin, the structure of another interesting family (fusarielins, See **Figure 6**) was analyzed [13, 14]. Its gene cluster (FSL gene cluster) was characterized recently [15]. This family is produced by *Fusarium graminearum* and is predicted to be synthesized by a HR-PKS (FSL_PKS) which is similar to LovB (**Figure 5**).

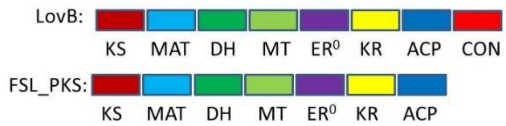


Figure 5. Linear organization of LovB and FSL_PKS

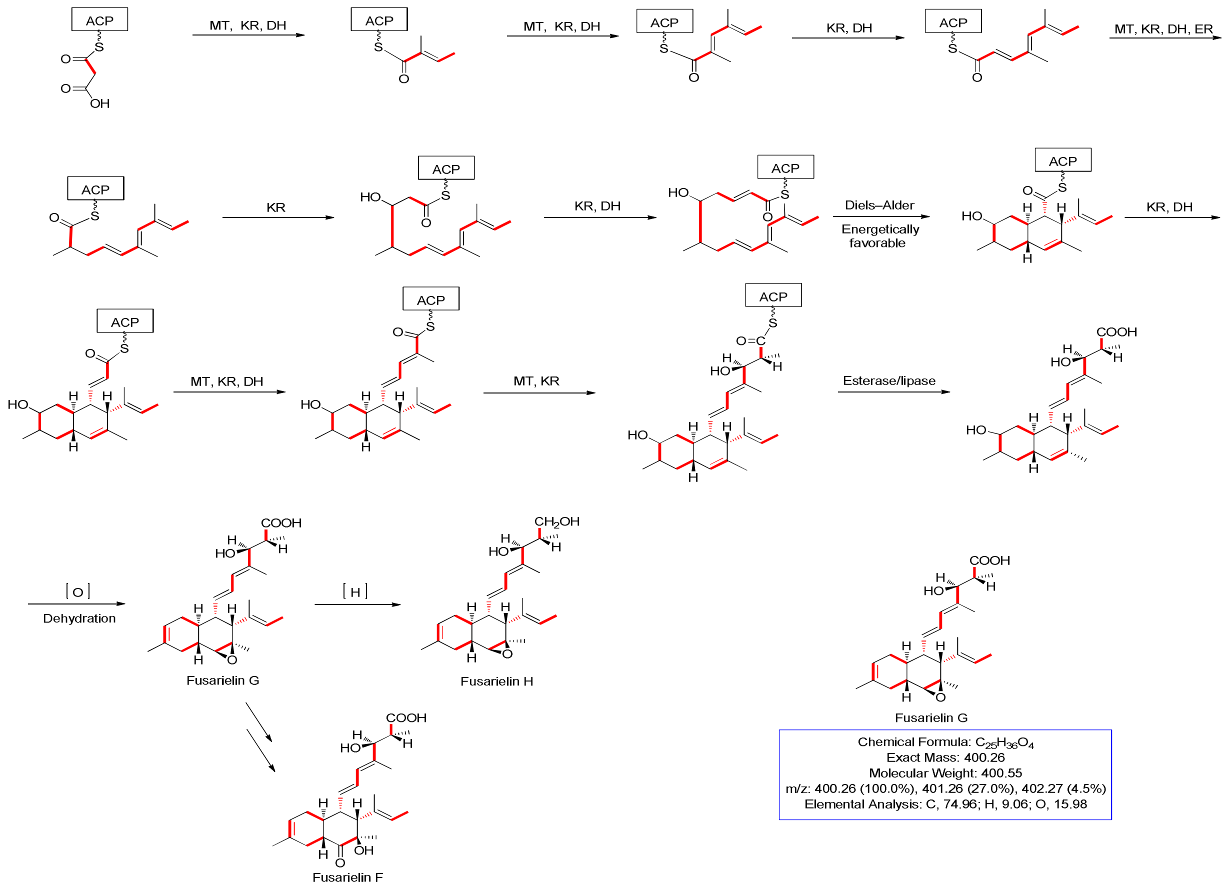


Figure 6. Proposed biosynthetic pathway of fusarielins

In comparing the first few proposed biosynthetic steps of the lovastatin and fusarielins (**Figure 7**), the MT domain of FSL_PKS functions before KR domain on diketide, triketide and pentaketide. But the KR domain functions before the MT domain on tetraketide which is just parallel and complementary to LovB (e.g. MT domain functions before KR domain on tetraketide). It will be a very good parallel model if we can compare LovB's and FSL_PKS's functions. So, the reconstitution of FSL_PKS in vitro is needed for study. Also, reconstitution of the biosynthetic pathway of fusarielins is very meaningful which has not been studied.

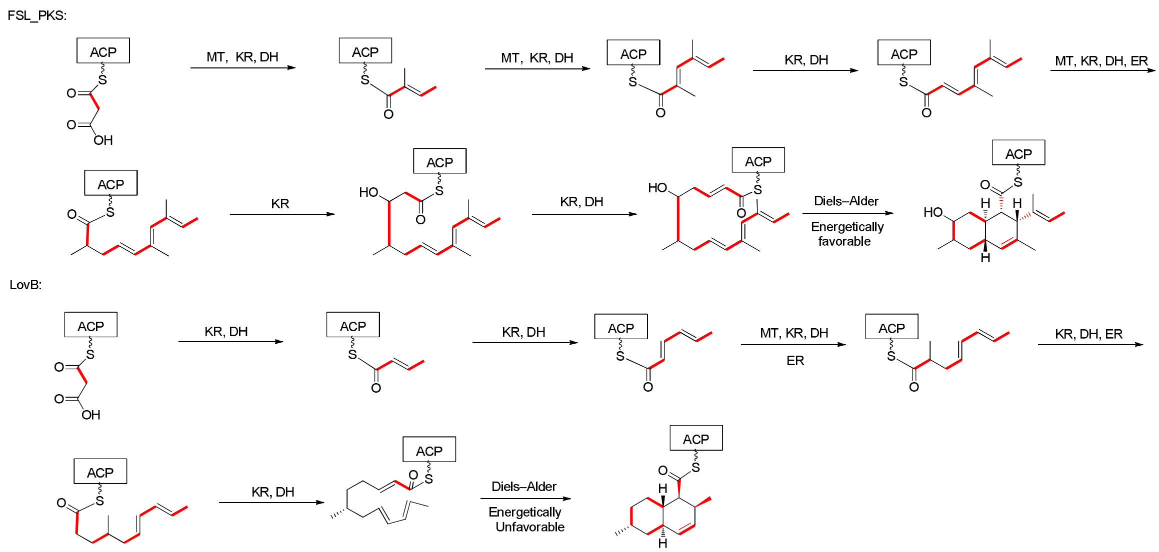


Figure 7. The comparison of biosynthetic pathways of lovastatin and fusarielins

A lot of polyketide biosynthetic gene clusters were identified using bioinformatics analysis. Although some protein sequences of the PKSs can be obtained accurately, predicting structures of their products is still one of the biggest challenges. First, it is very difficult to reconstitute large enzymes (e.g. HR-PKS) and generate soluble protein. Second, the programming rules of HR-PKS are still not well-understood. Third, a number of polyketide compounds remain unknown because they are encoded by cryptic biosynthetic clusters. Therefore, a better understanding of these biosynthetic pathways and the programming rules behind them are of significant importance in order to improve the drug discovery process.

II Materials

All chemicals were purchased from Sigma-Aldrich unless otherwise specified, and were used as received. *Saccharomyces cerevisiae* strain BJ5464-NpgA was used as the yeast expression host. Liquid Chromatography Mass Spectrometer (LC-MS) was conducted with a Shimadzu 2010 EV using positive electrospray ionization and a Phenomenex Luna 5u 2.0 × 100 mm C18 reverse-phase column. Bio-TEK powerwave XS Plate Reader was used to measure the NADPH consumption. 96-well UV plates with UV transparent flat bottoms from Corning Incorporated were used for the kinetic studies.

All other chemicals were from standard sources.

III Methods and Results

1. Exploring the programming rules of HR-PKS

1.1 Protein purification for LovB and LovB-DH^o

The expression plasmids, YEplovB-6His harboring the LovB or pXW11 harboring the LovB-DH^o (Dehydratase domain was site-mutated and ketoreduction products will accumulate), were transformed into *Saccharomyces cerevisiae* strain BJ5464-NpgA by using S. c. EasyCompTM Transformation Kit (Invitrogen). For 1 L yeast culture, the cells were grown at 28 °C in YPD media with 1% dextrose for 72 hours. The cells were harvested by centrifugation (3750 rpm, 15 minutes, 4 °C), resuspended in 30 mL lysis buffer (50 mM NaH_2PO_4, pH = 8.0, 0.15 M NaCl, 10 mM imidazole) and lysed with sonication on ice (Sonicate for 1 minute, then cool down for 1 minute. Repeat for 9 times). Cellular debris was removed by centrifugation (35000 g, 1 hour, 4 °C). Ni-NTA agarose resin was added to the supernatant (2~3 mL/L of culture) and the solution was stirred at 4 °C overnight. The protein/resin mixture was loaded into a gravity flow column and proteins were purified with increasing concentration of imidazole in Buffer A (50 mM Tris-HCl, pH = 7.9, 2 mM EDTA, 2 mM DTT). Purified proteins were concentrated and buffer exchanged into buffer E (50 mM Tris-HCl, 100 mM NaCl, pH=7.9) containing 10% glycerol. The concentrated enzyme solutions were aliquoted and flash frozen. Protein concentrations were determined with the Bradford (Biorad) assay using BSA as a standard. Purified proteins are run on the SDS gel.

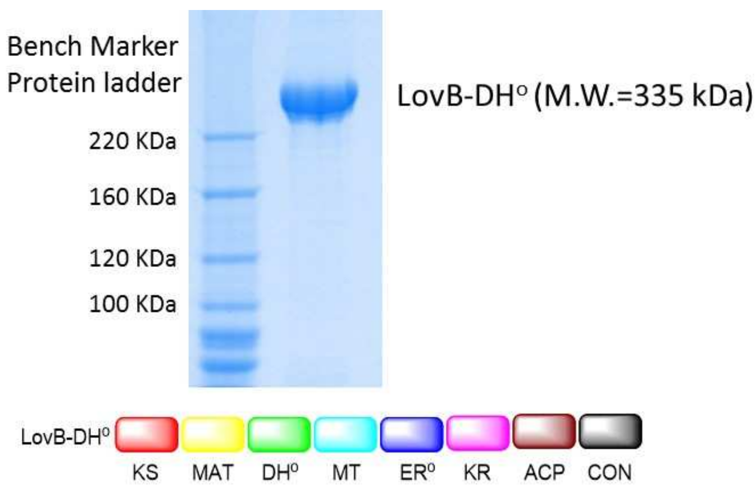


Figure 8. SDS-PAGE of the purified of LovB-DH^0 from S. cerevisiae strain

1.2 Essays

1.2.1 KR essays

For in vitro assays ketoreduction study (KR domain)[16], the final concentrations of LovB-DH^o were 20~32.85 μM, with cofactor concentrations as 0.05 or 0.7 mM NADPH, 100 mM NaH_2PO_4 (pH = 7.4) buffer at room temperature. The total volume of the reaction is 100 μL (for testing the reactions) or 200 μL (Measure with Plate Reader). The reactions were quenched after 1~24 hrs and extracted twice with 99% ethyl acetate (EA)/1% acetic acid (AcOH). The resultant organic extracts were evaporated to dryness, redissolved in methanol, and then analyzed by LC-MS.

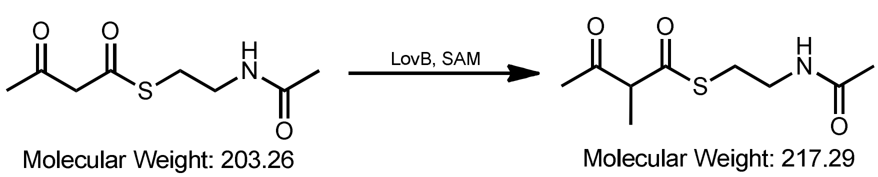
1.2.2 MT essays

For in vitro assays methylation study (MT domain), the final concentrations of LovB were 20~40 μM, with cofactor concentrations as 1 mM SAM, 100 mM NaH_2PO_4 (pH = 7.4) buffer at room temperature. The reactions were quenched after 1~24 hours and extracted twice with 99% ethyl acetate (EA)/1% acetic acid (AcOH). The resultant organic extracts were evaporated to dryness, re-dissolved in methanol, and then analyzed by LC-MS.

1.3 Kinetic studies

1.3.1 Pre-experiments on model diketide and tetraketide intermediates of lovastatin

We used a series of *β*-keto-S-N-acetyl cysteamine (SNAC) compounds and assayed them in the presence of LovB and its required cofactors (SAM or NADPH). We first try to test if the MT domain of LovB has activity on diketide **1a** (AASNAC) since the first step of modification on diketide was ketoreduction not methylation. 100 μL reaction was set up with 20 uM LovB, 1mM SAM and 1mM **1a**. After 16 hours, the reaction was quenched as stated in MT essays above and analyzed by LC-MS.



1a **1b**

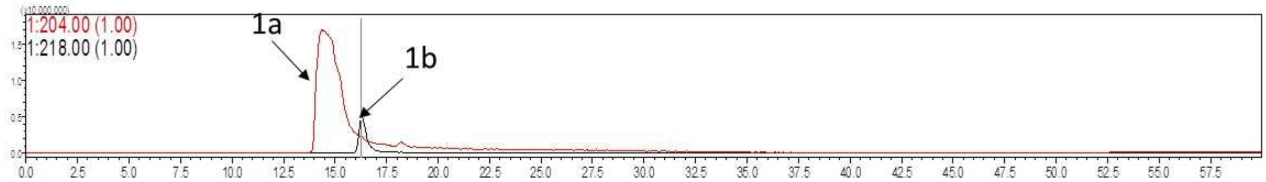
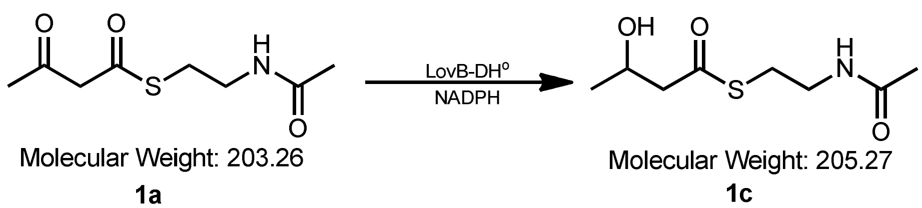


Figure 9. LC-MS analysis of AASNAC methylation essay

The LC-MS analysis of this in vitro assay was shown above in **Figure 9**. The traces shown are the selected ion monitoring of desired ions in the positive ionization mode. Trace 1a is $[M+H]^+$ at 204 for **1a** and trace 1b is $[M+H]^+$ at 218 for **1b**. At the same time, we set up another essay with 20 uM LovB-DH^o, 1mM NADPH and 1 mM **1a**. We expected to see the happening of ketoreduction and **1c** with molecular weight 205.27 could be detected. After 16 hours, we analyzed the reaction with LC-MS.



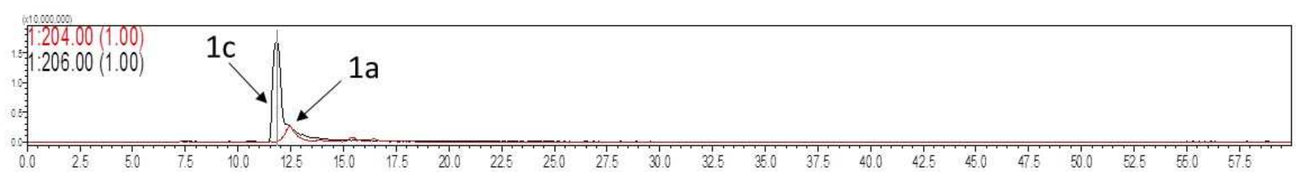


Figure 10. LC-MS analysis of AASNAC ketoreduction essay

At this time, trace 1a is $[M+H]^+$ at 204 for **1a** and trace 1c is $[M+H]^+$ at 206 for **1c** (**Figure 10**). It's really interesting to see that both ketoreduction and methylation reactions can happen on diketide **1a** since the natural product lovastatin was formed without methylation in its first elongation step.

If methylation can happen on diketide, then could ketoreduction happen first on tetraketide? If ketoreduction could happen on tetraketide first, the MT domain may not function on the reduction product since the nucleophilic activity of the tetraketide will be dramatically reduced. We tried the same reaction conditions: 40 uM LovB-DH^{o} , 1 mM model tetraketide SNAC **2a**, 1 mM NADPH or SAM, 16 hours.

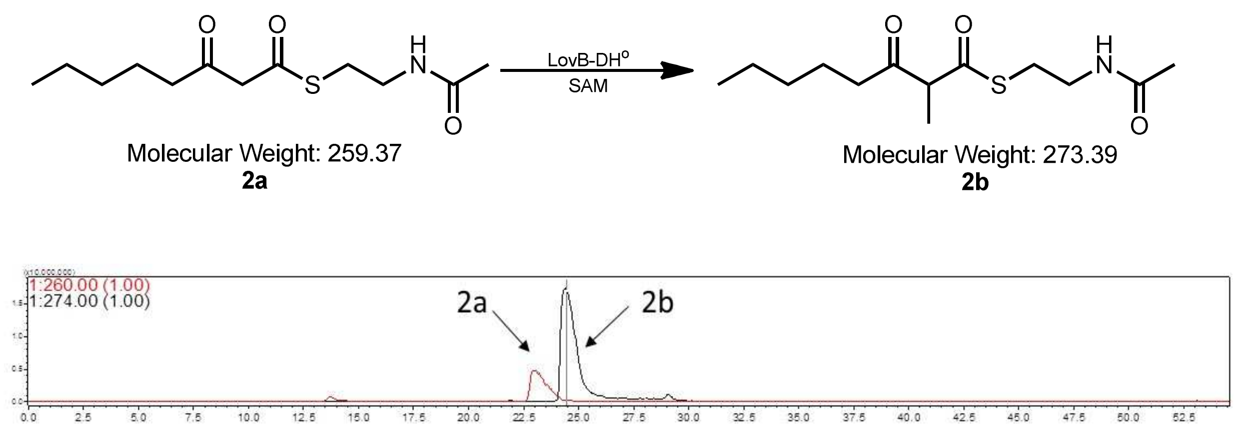


Figure 11. LC-MS analysis of model tetraketide SNAC methylation essay

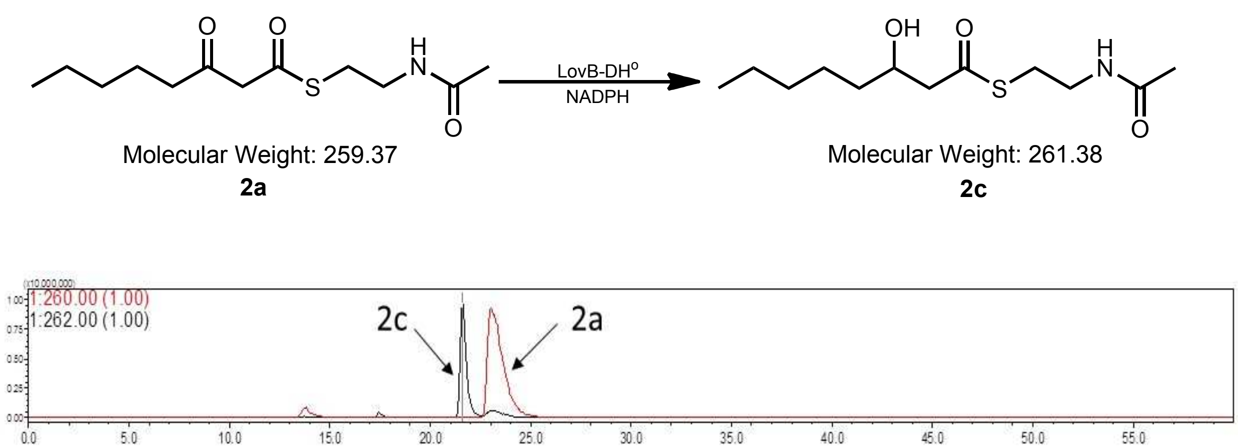


Figure 12. LC-MS analysis of model tetraketide SNAC ketoreduction essay

From the results above (**Figure 11** and **Figure 12**), we found that both methylation and reduction can also happen on tetraketide (same as diketide). MT domain of LovB will modify diketide if NADPH is not given. In comparing **Figure 9** and **Figure 10**, ketoreduction reaction seems much faster than the methylation reaction. KR domain of LovB will modify tetraketide if SAM is not given. After comparing **Figure 11** and **Figure 12**, methylation reaction seems faster than the ketoreduction reaction. Both ketoreduction reaction on diketide and methylation reaction on tetraketide happen during the biosynthesis of lovastatin. And these two reaction are both faster. In each step of elongation, LovB has the preference of choosing specific combination of modification domains based on the substrates.

Although all domains are present in the enzyme from beginning to the end of biosynthesis of lovastatin, LovB seems to choose which domain to use first based on kinetic selectivity. For example, if LovB chooses to use KR domain (Functions faster than MT domain) first in diketide, MT domain may not have activity on the reduced product any more. This may also be the case for triketide. Until tetraketide, LovB will choose MT domain to work faster than KR domain where methylation products start to appear in this step. For pentaketide, LovB will come back to choose KR domain first. From diketide to pentaketide, LovB may choose its modification domains (KR first or MT first) based on kinetic selectivity.

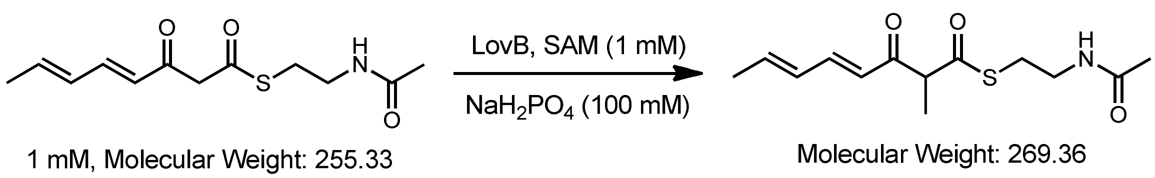
In order to prove our assumptions, we need to quantify the ketoreduction and methylation reactions. So, we need to find a way to get the kinetic parameters which describe the rates of these two reactions.

Michaelis–Menten equation we used here describes the rate of enzymatic reactions by relating reaction rate v to [S] which is the concentration of a substrate S. Its formula is given by $v = \frac{d[\mathrm{P}]}{dt} = \frac{V_{\max}[\mathrm{S}]}{K_m + [\mathrm{S}]}$. V_{max} represents the maximum reaction rate in the system, at maximum (saturating) substrate concentrations. The Michaelis constant K_m is equal to the substrate concentration at which the reaction rate is half its maximal value. Sometimes, a high K_m indicates weak binding; a low K_m indicates strong binding [17].

In proposed model of the enzymatic reaction ($E + S \rightleftharpoons ES \rightarrow E + P$), an enzyme E is binding to a substrate S to form a complex ES (k_f is the rate constant of ES formation and k_r is the rate constant of its reverse reaction), which is then converted into a product P and the enzyme E itself (k_{cat} is the rate constant where E and P are formed from ES). If the enzyme concentration is much less than the substrate concentration, the equation $v = \frac{d[\mathrm{P}]}{dt} = \frac{V_{\max}[\mathrm{S}]}{K_m + [\mathrm{S}]}$ can be rewritten to $v = \frac{d[\mathrm{P}]}{dt} = V_{\max}\frac{[\mathrm{S}]}{K_m + [\mathrm{S}]} = k_{cat}[\mathrm{E}]_0\frac{[\mathrm{S}]}{K_m + [\mathrm{S}]}$ which shows that $V_{\max} = k_{cat}[\mathrm{E}]_0$. $[E]_0$ is the enzyme concentration. k_{cat} (the turnover number) is the maximum number of substrate molecules converted to product per enzyme molecule per second. After getting K_m and k_{cat}, we can use the constant kcat/Km to measure of how efficiently an enzyme converts a substrate into product. After we get kinetic parameters from both ketoreduction and methylation of the same substrate, we can compare their K_m, k_{cat} and k_{cat}/K_m to discuss their kinetic selectivities [20].

In our reaction, LovB will be the enzyme. Model SNAC compounds and cofactors (SAM or NADPH) will be the substrates. For determining the constants V_{max} and K_m, we can run a series of enzyme assays at varying substrate concentrations [S] and measure the initial reaction rate v_0,which is measured after a relatively short time period. Since we have two substrates in our reactions, before we vary the substrates concentration of different model SNAC substrates, we need to make sure the concentration of other substrates, SAM or NADPH, can provide a saturated environment for the reactions.

The reason why we did not use natural tetraketide SNAC substrate **3a** is that the methylation reaction did not give us a single product as shown in **Figure 13**. **3a** was almost totally consumed after 2 hours. Compared to the results of methylation reaction above, **3b** might be the product we want. And there was another peak, which has the same molecular weight with **3b,** that is formed simultaneously. This will be very difficult for us to quantify the methylation reaction when the other peak is unknown. So, we chose to use the model substrates (**2a**) for further study which gave us a single peak besides the substrate. This may also indicate that the unknown peak in **Figure 13** could be another undesired methylation product.



3a **3b**

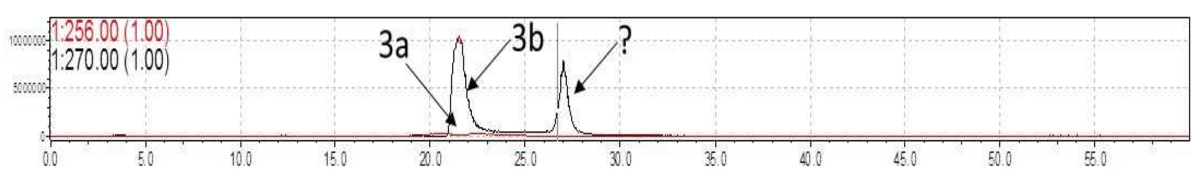


Figure 13. LC-MS analysis of natural tetraketide SNAC methylation essay

In order to quantify the kinetic activities of KR domain or MT domain, either the consumption of substrates, formation of products or consumption of the cofactors (NADPH or SAM) could be measured. NADPH consumption is widely used in quantification of NADPH involved enzymatic reactions. So, we first tried to measure the kinetic parameters of KR domain based on NADPH consumption using Plate Reader (340 kinetics).

1.3.2 Kinetic measurements

1.3.2.1 NADPH Standard Curve

96-well UV plates with UV transparent flat bottom were used here for measurements of NADPH 340nm absorption. Each well contains 200 μL solutions. The NADPH standard curve (**Figure 14**) is established to show the correlation between the 340nm absorptions from Plate Reader and NADPH concentration [18, 19]. This curve also shows the linear range of NADPH concentration we can use for our reaction. After increasing the NADPH concentration to 2.5 mM or 5 mM, the 340 nm absorption was no longer in the linear range.

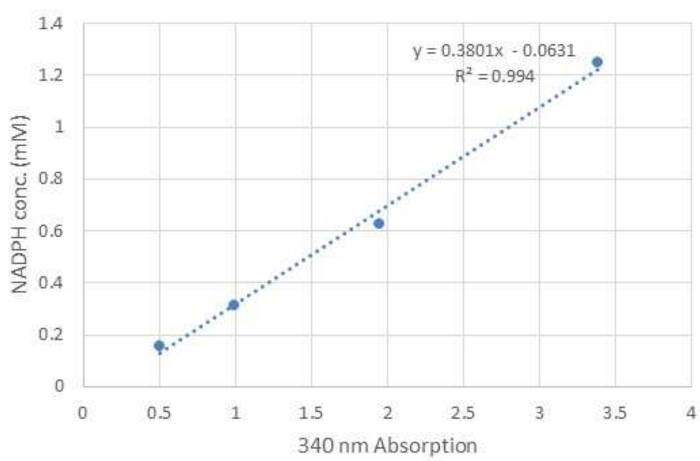


Figure 14. Correlation between NADPH concentration and 340 nm absorption

1.3.2.2 Measurements of kinetic parameters of ketoreduction reactions

Before varying the SNAC substrate concentration (**1a** was tested first) to get reaction rates, three reactions as a control (without LovB-DH° or NADPH or **1a**) were set up together with a ketoreduction reaction. We can use Plate Reader to collect data as the reaction goes on. Only the reaction with all the components can significantly reduce the 340nm absorption of the solution. See **Table 1** and **Figure 15** below. The enzyme concentration (32.85 μM) used here gave us a nearly straight line, which will be used in ketoreduction reactions later.

Table 1. Substrate **1a** Ketoreduction

	Substrate	NADPH	Enzyme	NaH_2PO_4
Ketoreduction	30 mM	0.05 mM	32.85 uM	100 mM
Control 1	30 mM	0.05 mM	***	100 mM
Control 2	30 mM	***	32.85 uM	100 mM
Control 3	***	0.05 mM	32.85 uM	100 mM

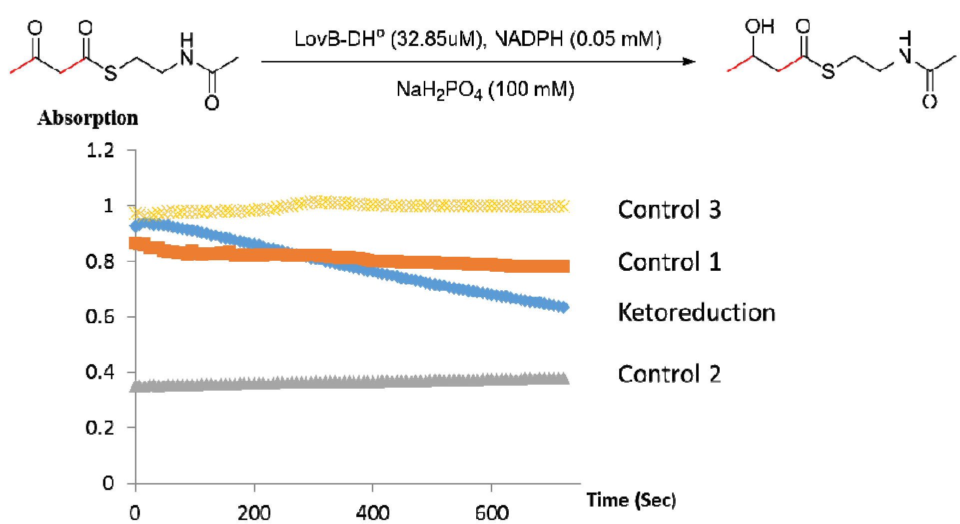
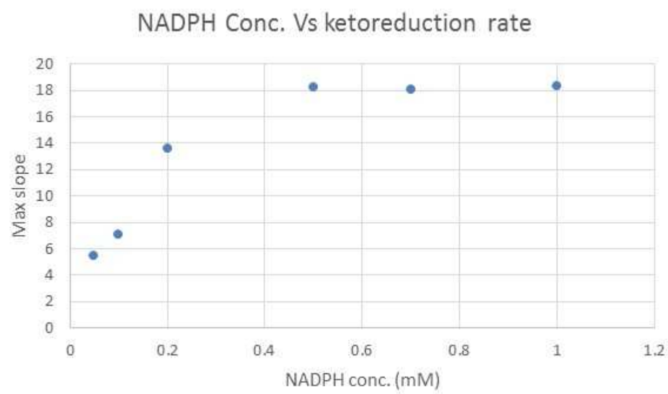


Figure 15. Pre-experiments of AASNAC ketoredution

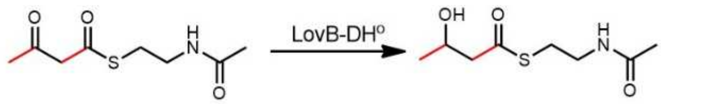
After we have an available enzyme concentration, we need to make sure the NADPH concentration is high enough during the ketoreduction reactions. Different NADPH concentrations were tested (5 mM **1a**, 32.85 uM LovB-DH^{o}). The max slope in **Table 2**, which corresponding to the initial reaction rate, was based on the reduction of 340nm absorption. From 0.5 mM of NADPH concentration, the reaction rate became stable (saturated) which can be used for further study. 0.7 mM NADPH was chosen here.

Table 2. Ketoreduction with different NADPH concentration

NADPH Conc.	Max Slope
0.05 mM	-5.4287114
0.1 mM	-7.1140233
0.2 mM	-13.594086
0.5 mM	-18.242236
0.7 mM	-18.054851
1 mM	-18.359286



The next step is to run a series of enzyme assays with different substrate concentrations (100 uM – 30000 uM) to get the correlation between initial rate and substrate concentration. The initial reaction rate is calculated based on the first few points read from Plate Reader (**Table 3**). The K_m and k_{cat} were calculated using Michaelis-Menten model (Software GraphPad Prism was used, **Figure 16**). From the data here, we could get kinetic parameters of AASNAC (**1a**) ketoredution that K_m is 6.007 mM, k_{cat} is 0.2242 min^{-1} and V_{max} is 7.365 uM/min. Also, k_{cat}/K_m is 0.03732 $mM^{-1}min^{-1}$.



Essay conditions: 32.85 uM LovB-DH° , AASNAC, 0.7 mM NADPH, 100 mM NaH_2PO_4.

Table 3. Substrate **1a** concentration and their initial rate

1a Conc. (uM)	Initial rate (uM/min)
30000	6
20000	6
10000	4.2
5000	3.6
1000	1.2
500	0.12
100	0.36

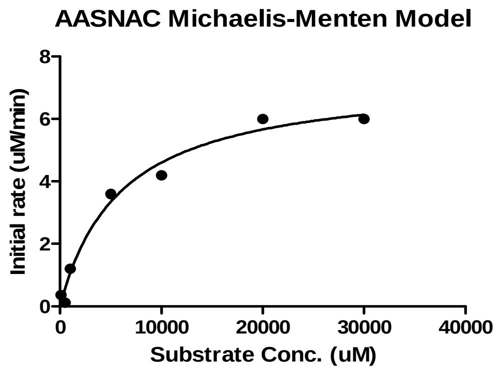
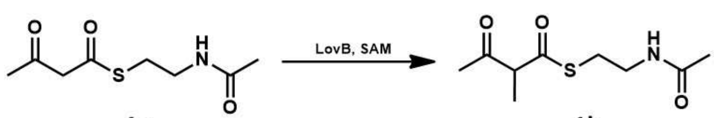


Figure 16. Results of AASNAC Michaelis-Menten Model

1.3.2.3 Methods and standard curve for measuring methylation reactions



1a **1b**

After setting up the methods for ketoreduction reactions, methylation reactions need to be quantified. AASNAC was also tested first here. There is almost no UV absorption at lower substrate concentration of both substrate and its desired product. Also, no enzymes were efficient and available to do some structure modification to facilitate the detection.

For our testing essays, the peak of the product cannot be distinguished clearly to know exact area of the peak. But we found out that selected ion monitoring of the desired ions intensity in the positive ionization mode of LC-MS is very easy to identify. $[M+H]^+$ (218 for AASNAC **1a**) was the peak we selected and used for determination of the area of the MS intensity.

One of the most important part here is to confirm if the MS intensity can be trusted and used for calculation. Eight standard **1b** concentration have been tested, each of the solutions was 100 uL and LC-MS injected 20 uL of the solutions.

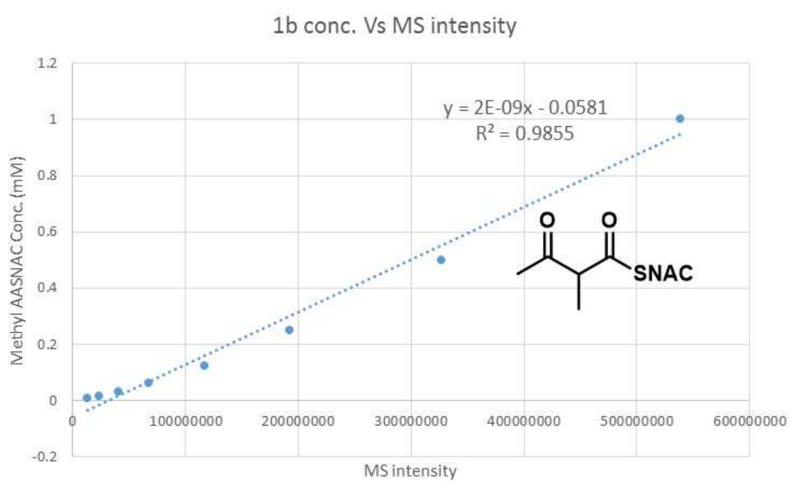


Figure 17. Correlation between **1b** concentrations and their corresponding MS intensities

Based on the results (see **Figure 17**), we got a very good linear correlation between **1b** concentrations and the corresponding MS intensities. Then, for each methylation reaction, we can measure the $[M+H]^+$ MS intensity of product **1b** using LC-MS and calculate the **1b** concentration based on the standard curve.

1.3.2.4 Measurements of kinetic parameters of methylation reactions

We tried to get reaction rates based on the method above (SAM concentration was not adjusted yet). Methylation reactions at each substrate concentration were set up as 400 uL or 500 uL reaction mix and aliquot the mix as 100 uL mix per 1.5 mL tube. The reactions from the same master mix were quenched after 0, 2, 4, 6, 8 hrs and extracted twice with 99% ethyl acetate (EA)/1% acetic acid (AcOH). The resultant organic extracts were evaporated to dryness, redissolved in 100 uL methanol, and 20 uL was injected to LC-MS by auto-sampling. The MS intensities were converted to the concentrations of the product. The concentration of LovB could be 20 uM or 40 uM, which was in the linear range. The next step will be testing the SAM concentration which can be used for further study.

1.4 Discussion

Methods for measuring the kinetic parameters of ketoreduction and methylation reactions were set up. NADPH consumption could be measured by Plate Reader (340 kinetics) as the reaction goes on. After enzyme concentration and NADPH concentration were selected and tested, different substrate (**1a**) concentrations and their corresponding initial reaction rate were used to calculate K_m, k_{cat} and k_{cat}/K_m. During quantifying the methylation reactions, $[M+H]^+$ MS intensities were recorded and converted into the product concentration based on the standard curve.

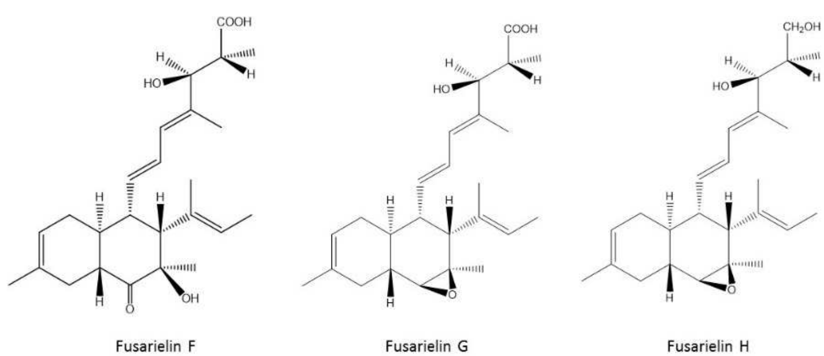
2 Biosynthesis of Fusarielins

2.1 Molecular Biology

2.1.1 General DNA Manipulation Techniques and putative fusarielin gene cluster

E. coli XL1-Blue and *E. coli* TOPO10 were used for cloning following standard recombinant DNA techniques. *E. coli* BL21 (DE3) was used for protein expression in *E. coli*. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs). *Saccharomyces cerevisiae* strain BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2- Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*) was used as the yeast expression host [9].

Putative fusarielin gene cluster in *Fusarium graminearum* was identified [15]. After analyzing the structure of fusarielins, we thought that three key enzymes (**Figure 18**) may be responsible for the biosynthesis of fusarielins: FSL1 (Polyketide synthase, FSL_PKS; Construct the backbone of fusarielins, 287.01 KDa), FSL2 (Esterase/lipase, FSL_TE; Release the products from ACP domain of polyketide synthase, 26.99 KDa) and FSL5 (Enoyl reductase, FSL_ER; Enoyl reduction during fusarielins biosynthesis, 39.32 KDa).



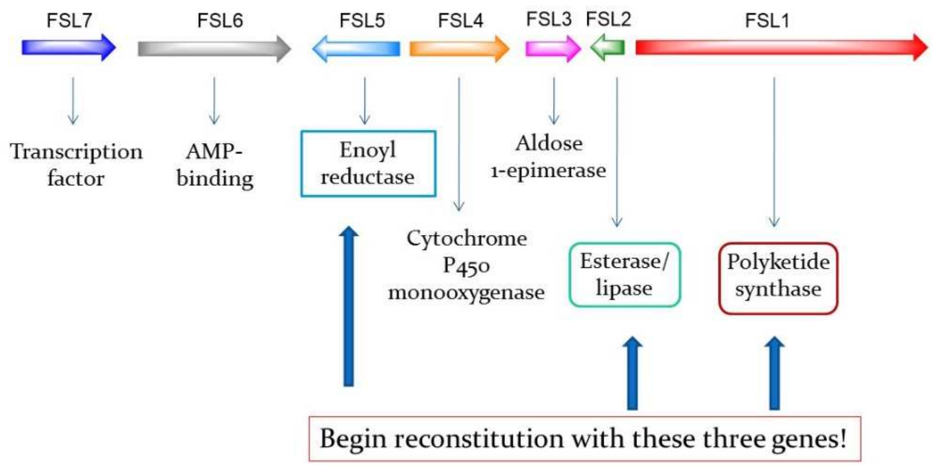


Figure 18. Structures of fusarielins and putative fusarielin gene clusters

The three enzymes need to be cloned and expressed. Based on prediction, FSL_TE has no introns, FSL_ER has one introns and FSL_PKS has five introns (three at the beginning and two in the end, **Figure 19**). All the introns have to be removed in order to achieve heterologous protein expression. Since the FSL_PKS is too large, it will be divided into three parts: PKS1 (front, ~2 K bp, three introns), PKS-mid (middle, ~4 K bp, no introns), PKS2 (end, ~ 2K bp, two introns). The FSL_PKS will be constructed using yeast recombination.

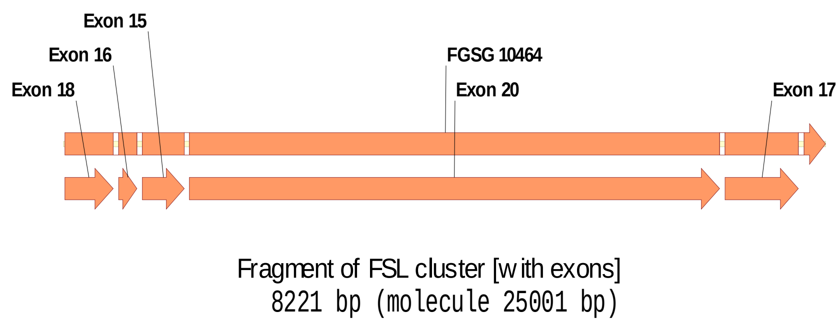


Figure 19. Putative FSL_PKS gene and the predicted position of introns

2.1.2 Plasmid construction

The genome DNA (gDNA) was extracted from *Fusarium graminearum* (*F. graminearum*) which was grown for about two weeks on solid Yeast Extract Sucrose (YES) agar[15]. FSL_PKS, FSL_ER and FSL_TE genes were cloned from the cDNA which was generated by RT-PCR from RNA of *F. graminearum* using poly-T or the primers listed in **Table 4**.

Table 4. Primer table for cloning

Primer Name	Sequence
FSL_N-His_ER-NdeI-F	AAC ATA TGC ATC ACC ATC ATC ATC ACA TGA AAC GTG TCT TTT TAC TC
FSL_ER-NotI-R	AAG CGG CCG CTT ATG CTA CTT GCG GAA TCA TG
FSL_ER-PmeI-R	AAG TTT AAA CTT ATG CTA CTT GCG GAA TCA TG
FSL_N-His_TE-NdeI-F	AAC ATA TGC ATC ACC ATC ATC ATC ACA TGA CTA TTC AAA GTA CTG CGA ATC
FSL_TE-NotI-R	AAG CGG CCG CTT ATG TTT GCG CAG AGT CCC AAC
FSL_TE-PmeI-R	AAG TTT AAA CTT ATG TTT GCG CAG AGT CCC AAC
FSL_PKS-SpeI-F	GCT AGC GAT TAT AAG GAT GAT GAT GAT AAG ACT AGT ATG CAA GGA CCG ACC AAC GAA C
FSL_PKS-RT_PCR1-R	GCT GGA TGA CTG AGT GTT GCG
FSL_PKS_PCR_Mid-f	CAC AGA GAG TGT CTA TCA TAG CAT CGA C
FSL_PKS_PCR_Mid-r	TAT TGA CGT CTG TGC CGA CCT G
FSL_PKS-RT_PCR2-f	CAA CTG GTT TGC CAA GCC AC
FSL_PKS-RT_PCR2-PmlI-R	TCA TTT AAA TTA GTG ATG GTG ATG GTG ATG CAC GTG CTT GGA CTT TTC CTT GAC CGG

2.1.2.1 Cloning of C-His6 FSL-PKS expression vector

DNA sequences of FSL-PKS were amplified in 3 pieces (PKS1, PKS-mid, PKS2), using the primers listed in **Table 4**. All of them were first ligated into pCR-Blunt vectors and hence created pBo3 (from gDNA, PKS-M), pBo23, pBo25 (from cDNA, PKS1 and PKS2) which carries an ADH2p-(FSL_PKS)-ADH2t cassette, was used as the vector template in the cloning. Yeast recombination (**Figure 19**) were performed using three corresponding PCR products(PKS1, PKS-mid, PKS2) and one expression vector pXW55 digested by SpeI/PmlI (a yeast expression vector with a uracil marker), to create FSL_PKS yeast expression vector pBo27.

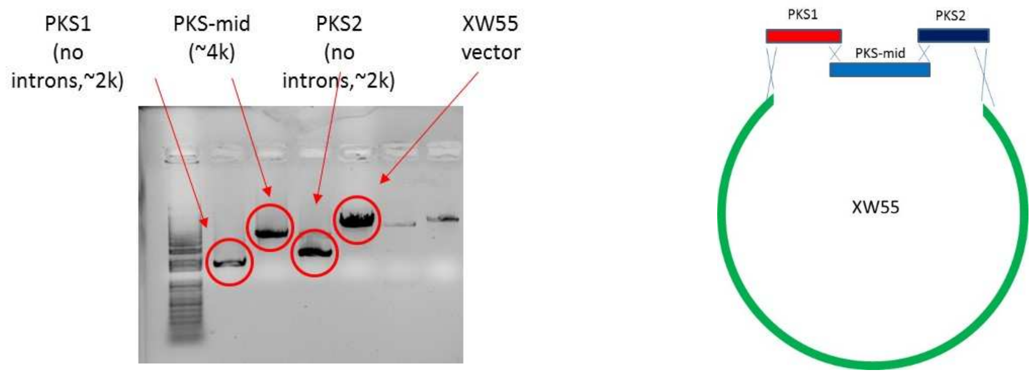


Figure 19. DNA gel showing the DNA pieces before yeast recombination and proposed yeast recombination method

2.1.2.2 Cloning of C-His6 FSL_ER *E. coli* and *Saccharomyces cerevisiae* expression vector

FSL_ER (from 2nd ATG) was cloned from cDNA using the primers listed in **Table 4**. The PCR product was inserted into pCR-Blunt and subsequently digested with NdeI and NotI and inserted into pET-23a (+) to yield *E. coli* expression vector pBo18. FSL_ER PCR products with PmeI/NdeI digestion ends were inserted into the digested pXW06 (a yeast expression vector with a tryptophan marker) to create pBo17.

2.1.2.3 Cloning of C-His6 FSL_TE *E. coli* and *Saccharomyces cerevisiae* expression vectors

FSL_TE was cloned from gDNA using the primers listed in **Table 4**. The PCR product was inserted into pCR-Blunt and subsequently digested with NdeI and NotI and inserted into pET-23a (+) to yield *E. coli* expression vector pBo9. FSL_TE PCR products with PmeI/NdeI digestion ends were inserted into the digested pXW02 (a yeast expression vector with leucine marker) or pXW06 (a yeast expression vector with a tryptophan marker) to create pBo26-4 or pBo11.

2.2 Protein Expression and Purification

2.2.1 Proteins Expression and Purification from *E. coli.*

The expression plasmids were transformed into *E. coli* BL21 (DE3) strain for protein expression through electroporation. The cells were grown at 37°C in 1L LB medium with 100 μg/mL ampicillin to an OD600 of 0.4-0.6. After that, the cells were incubated on ice for 10 minutes, and then induced with 0.1 mM isopropylthio-β-D-galactoside (IPTG) for 16 hours at 16°C. The cells were harvested by centrifugation (2500 g, 15 minutes, 4°C), re-suspended in 25 mL lysis buffer (50 mM Tris-HCl, 2 mM EDTA, 2 mM DTT, 500 mM NaCl, 5 mM imidazole, pH=7.9) and lysed through sonication on ice (Sonicate for 30 minute, then cool down for 1 minute. Repeat for 6 times). Cellular debris was removed by centrifugation (30,000 g, 30 min, 4°C). Ni-NTA agarose resin was added to the supernatant and the solution and was stirred at 4°C for at least 2 hours. The mixture was loaded into a gravity flow column and proteins were eluted with an increasing concentration of imidazole in buffer A (50 mM Tris-HCl, 500 mM NaCl, pH=7.9).

For the expression of FSL_TE, a chaperone plasmid pG-KJE8 was also co-transformed with pBo9 into *E. coli* BL21 (DE3) strain through electroporation. The transformants were incubated at 37°C in 1L LB medium containing 100 μg/mL ampicillin, 25 μg/mL, chloramphenicol and 5 ng/mL tetracycline to an OD600 of 0.4-0.6. The expression and purification steps were the same as above.

2.2.2 Expression and Purification of proteins in *Saccharomyces cerevisiae*

For FSL_TE and FSL_PKS an attempt was made to use the yeast strain BJ5464-NpgA as an expression host. The expression plasmids were transformed into *Saccharomyces* *cerevisiae* strain BJ5464-NpgA by using S. c. EasyCompTM Transformation Kit (Invitrogen). For 1L of yeast culture, the cells were grown at 25°C in YPD media with 1% dextrose for 72 hours. The cells were harvested by centrifugation (2500 g, 20 minutes, 4°C), re-suspended in 25 mL lysis buffer (50 mM NaH_2PO_4 pH 8.0, 0.15 M NaCl, 10 mM imidazole) and lysed through sonication on ice. Cellular debris was removed by centrifugation (35,000 g, 1 hour, 4°C). Ni-NTA agarose resin was added to the supernatant (2 mL/L of culture) and the solution was stirred at 4°C overnight. The protein/resin mixture was loaded into a gravity flow column and proteins were purified with increasing concentration of imidazole in Buffer A (50 mM Tris-HCl, pH=7.9, 2 mM EDTA, 2 mM DTT).

2.3 Results and Discussion

FSL_ER was expressed using *E. coli* BL21 (DE3) as expression host. We could not get FSL_TE expressed in vitro both in *E. coli* and in yeast. Three expression plasmids, (pBo27 harboring the FSL_PKS, pBo17 harboring the FSL_ER and pBo26-4 harboring the FSL_TE), were co-transformed into *Saccharomyces cerevisiae* strain BJ5464-NpgA by using S. c. EasyCompTM Transformation Kit (Invitrogen). The strain was cultivated and extracted. Unfortunately, LC-MS did not show any promising products.

FSL_PKS was expressed to get two truncated parts using *Saccharomyces cerevisiae* as expression host (**Figure 20**). After adding protease inhibiter, the FSL_PKS sometimes can be purified without truncated part. We guess the protein may be truncated between MAT domain and DH domain where the bond is relative week. But the concentration of FSL_PKS was very low and in vitro essays did not show any activities.

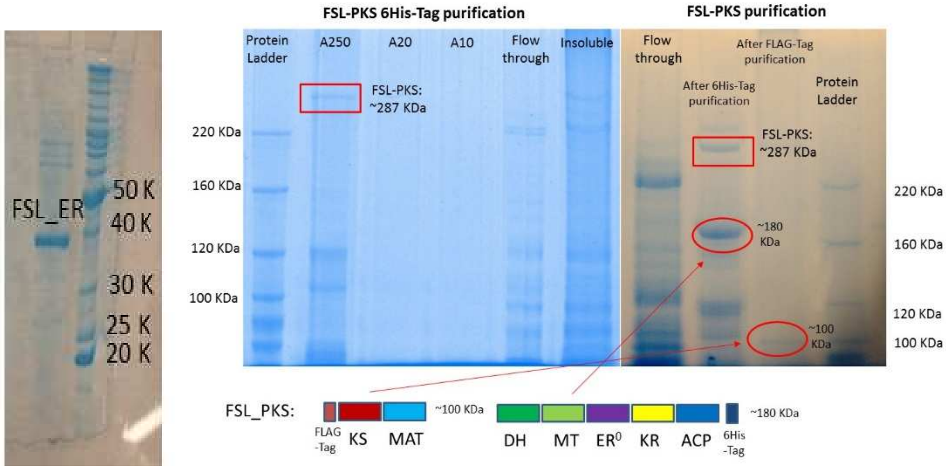


Figure 20. SDS-PAGE showing FSL_PKS and FSL_ER

IV Future studies

In conclusion, methods for measuring the kinetic parameters of KR domain and MT domain using NADPH consumption or MS intensity were established. The next step is to get the kinetic parameters of the methylation reaction using AASNAC as the substrate and to apply the methods to other model substrates. Kinetic parameters can be compared between KR and MT domains based on the same intermediate. The ratio of Kp (Kp=kcat/Km) between KR domain and MT domain is predicted to become much smaller during the biosynthesis of tetraketide where MT domain works faster than KR domain. This fundamental study will give us more knowledge from the aspects of kinetic selectivity to make it clearer about what the programming rule of HR-PKS is.

In addition, the fusarielins family's biosynthetic pathway is a good model to compare with lovastatin's. Its MT domain is proposed to work slower on tetraketide and work faster on diketide, triketide and pentaketide. The expression of the key enzymes should be looked at more carefully by focusing on their cloning and sequence analysis.

V References

1. Maier, T., M. Leibundgut, and N. Ban, *The crystal structure of a mammalian fatty acid synthase.* Science, 2008. **321**(5894): p. 1315-22.
2. Michael A. Fischbach, C.T. Walsh, and *Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms.* Chem. Rev, 2006. **106**: p. 3468-3496.
3. Winter, J.M. and Y. Tang, *Synthetic biological approaches to natural product biosynthesis.* Curr Opin Biotechnol, 2012. **23**(5): p. 736-43.
4. Sattely, E.S., M.A. Fischbach, and C.T. Walsh, *Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways.* Nat Prod Rep, 2008. **25**(4): p. 757-93.
5. Chooi, Y.H. and Y. Tang, *Navigating the fungal polyketide chemical space: from genes to molecules.* J Org Chem, 2012. **77**(22): p. 9933-53.
6. Xu, W., K. Qiao, and Y. Tang, *Structural analysis of protein-protein interactions in type I polyketide synthases.* Crit Rev Biochem Mol Biol, 2012.
7. Gao, X., P. Wang, and Y. Tang, *Engineered polyketide biosynthesis and biocatalysis in Escherichia coli.* Applied Microbiology and Biotechnology, 2010. **88**(6): p. 1233-1242.
8. Robert Mcdaniel, Claes Gustafsson, Hong Fu, Melanie Betlach, and Mary betlach, *Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products.* Proc. Natl. Acad. Sci. USA, 1999. **96**: p. 1846–1851.

9. Ma, S.M., et al., *Complete reconstitution of a highly reducing iterative polyketide synthase.* Science, 2009. **326**(5952): p. 589-92.

10. Kennedy, J., *Modulation of Polyketide Synthase Activity by Accessory Proteins During Lovastatin Biosynthesis.* Science, 1999. **284**(5418): p. 1368-1372.

11. Karine Auclair, Jonathan Kennedy, David J. Witter, and Richard Hutchinson, *Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels-Alder Reaction of a Substrate Analogue.* J. Am. Chem. Soc., 2000. **122**: p. 11519-11520.

12. Gao, Z., et al., *Investigation of fungal iterative polyketide synthase functions using partially assembled intermediates.* J Am Chem Soc, 2013. **135**(5): p. 1735-8.

13. Stajich, J.E., F.S. Dietrich, and S.W. Roy, *Comparative genomic analysis of fungal genomes reveals intron-rich ancestors.* Genome Biol, 2007. **8**(10): p. R223.

14. Gai, Y., et al., *Fusarielin E, a new antifungal antibiotic from Fusarium sp.* Chinese Chemical Letters, 2007. **18**(8): p. 954-956.

15. Sorensen, J.L., et al., *Production of novel fusarielins by ectopic activation of the polyketide synthase 9 cluster in Fusarium graminearum.* Environ Microbiol, 2012. **14**(5): p. 1159-70.

16. Zhou, H., et al., *A fungal ketoreductase domain that displays substrate-dependent stereospecificity.* Nat Chem Biol, 2012. **8**(4): p. 331-3.

17. Chen, W.W., M. Niepel, and P.K. Sorger, *Classic and contemporary approaches to modeling biochemical reactions.* Genes Dev, 2010. **24**(17): p. 1861-75.

18. Cunniff, B., et al., *A direct and continuous assay for the determination of thioredoxin reductase activity in cell lysates.* Anal Biochem, 2013. **443**(1): p. 34-40.

19. Obiero, J., et al., *Thioredoxin system from Deinococcus radiodurans.* J Bacteriol, 2010. **192**(2): p. 494-501.

20. Lehninger, A.L.; Nelson, D.L.; Cox, M.M. (2005). *Lehninger principles of biochemistry. New York: W.H. Freeman. ISBN 978-0-7167-4339-2.*