UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Polyketide Synthase Engineering via Catalytic Domain Exchanges

Permalink
https://escholarship.org/uc/item/2ns1p4x6

Author
Eng, Clara Hoifung

Publication Date
2017

Peer reviewed|Thesis/dissertation

Polyketide Synthase Engineering via Catalytic Domain Exchanges

by

Clara Hoifung Eng

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D. Keasling, Chair
Professor John E. Dueber
Professor Wenjun Zhang

Fall 2017

Polyketide Synthase Engineering via Catalytic Domain Exchanges

Abstract

Polyketide Synthase Engineering via Catalytic Domain Exchanges

by

Clara Hoifung Eng

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Jay D. Keasling, Chair

Polyketides comprise a structurally diverse class of natural products that includes many highly effective pharmaceuticals. As such, the development of techniques to produce polyketide analogs with potentially improved pharmacokinetic or toxicological properties is a research area of great interest. The modular nature of PKSs has also motivated researchers to explore methods of employing engineered PKSs for the biological synthesis of industrially relevant small molecules, in addition to natural product drugs.

We utilize *in vitro* enzyme studies to elucidate the biosynthetic potential of polyketide synthases (PKSs), the large, multisubunit enzymes responsible for polyketide production, and further leverage the insights gained from benchtop experiments to inform alterations to culture conditions that enable the production of non-natural natural product analogs.

In addition, we report progress developing methods to harness our understanding of the modular structure of PKSs to reliably exchange ketoreductase (KR) domains between PKS systems. The importance of stereochemistry in determining biological activity makes an ability to understand and engineer KR domains, which dictate the stereochemistry at many of the stereocenters in polyketide products, essential to the objective of effectively exploiting PKSs for combinatorial biosynthesis. Our work to identify heuristics that facilitate reliable KR domain exchanges furnished a novel library of PKS variants harboring different KR domains, further enabling studies that provide insight into how KR domains function in the context of a PKS module.

Finally, we build on the knowledge gleaned from KR domain exchange experiments to develop more general strategies for PKS engineering. A limited theoretical understanding of the determinants of PKS fold and function poses a substantial barrier to the design of active variants, and identifying strategies to reliably construct functional PKS chimeras remains an active area of research. We develop and formalize a paradigm for the design of PKS chimeras, and implement a computational platform to streamline and simplify the process of designing experiments to test strategies for PKS engineering. Finally, we discuss the implications of this system for enabling the effective use of PKSs in synthetic biology applications.

To my parents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Polyketide Synthases as a Potential Platform for Combinatorial Biosynthesis

Polyketides and their derivatives are used broadly in medicine as antibiotics, antifungals, and immunosuppressants, among other applications [72], motivating research to engineer PKSs capable of producing novel drug analogs. More recently, researchers have also harnessed the biosynthetic potential of PKSs to produce commodity chemicals in an environmentally friendly manner [29, 81, 79].

The 6-deoxyerythronolide B (6-deB) synthase (DEBS) from *Saccharopolyspora erythraea*, shown in Figure 1.1, which produces the aglycon precursor to the antibiotic erythromycin, is the archetypal model system for understanding the structure and function of type I modular PKSs. PKSs are composed of a series of catalytic domains whose identity and order determine the structure of the final polyketide product in a predictable manner. A single PKS can include multiple protein subunits, each containing one or more modules, where each module is responsible for the extension of the nascent acyl chain by two carbon units. The polyketide product released from a PKS may undergo further modifications, such as hydroxylation, glycosylation, or methylation to form the final active compound [45].

Initiation of polyketide biosynthesis typically occurs through one of two mechanisms: in the first, the loading acyltransferase (AT_L) selects an acyl coenzyme A (CoA) starter substrate and transfers the acyl group from the CoA to the loading acyl carrier protein (ACP). The starter unit is then transferred directly to the ketosynthase (KS) domain of the first extension module. In the second of these mechanisms, the AT_L selects a 3-carboxylacyl-CoA substrate, such as malonyl-CoA, which is decarboxylated by an accompanying KS^Q prior to transfer to the KS domain of the first extension module. The KS^Q domain is thus optionally present, depending on the mechanism of polyketide biosynthesis initiation.

Each module, which is minimally made up of a set of ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) domains, carries out a KS-catalyzed decarboxylative condensation between an acyl starter or intermediate, and a dicarboxylic acid extender
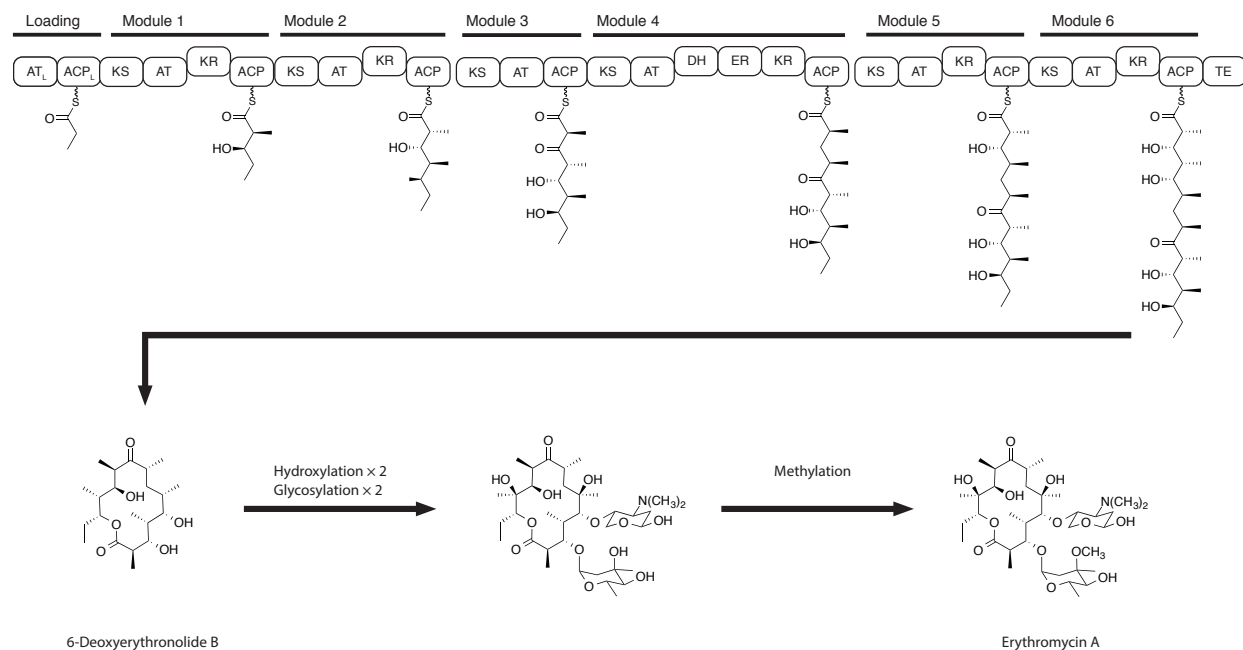
Figure 1.1: **Schematic of DEBS.** DEBS is comprised of three polypeptide subunits, each containing two modules. 6-Deoxyerythronolide B, the polyketide produced by DEBS undergoes hydroxylation, glycosylation, and methylation following release of the polyketide from the PKS to form the active compound, erythromycin.

unit, which is selected by the AT [10, 51]. The resulting β-ketone may be reduced by a ketoreductase (KR) domain, which determines the stereochemistry of the resulting β-hydroxyl group, as well as that of the α-substituent if present [59, 7, 24, 25, 2]. Additional dehydratase (DH) and enoyl reductase (ER) domains may further reduce the hydroxyl group to yield a methenyl or methylene group, respectively. The KR, DH, and ER make up a set of optional reducing domains that together are termed the reductive cassette. A schematic of a general PKS module is shown in Figure 1.2. In addition to these reductive domains, C- or O- methyltransferases may also be present. Release of a polyketide from a PKS is carried out by a thioesterase (TE) domain that catalyzes either hydrolysis or cyclization [45].

The potential interchangeability of PKS catalytic domains make PKSs an appealing prospective toolkit for combinatorial biosynthesis. The structure of the two-carbon addition to the nascent acyl chain that is effected by a single PKS module is determined by the substrate specificity of the AT domain, the stereochemical outcome modulated by the KR domain, and the number of active reducing domains in the reductive cassette. Over 20 distinct malonyl-CoA analogs have been observed as extender units in naturally occurring PKSs [81, 75, 12], and all possible combinations of stereochemical configurations at the α and β positions are made accessible by different KR domains. Developing methods to effectively
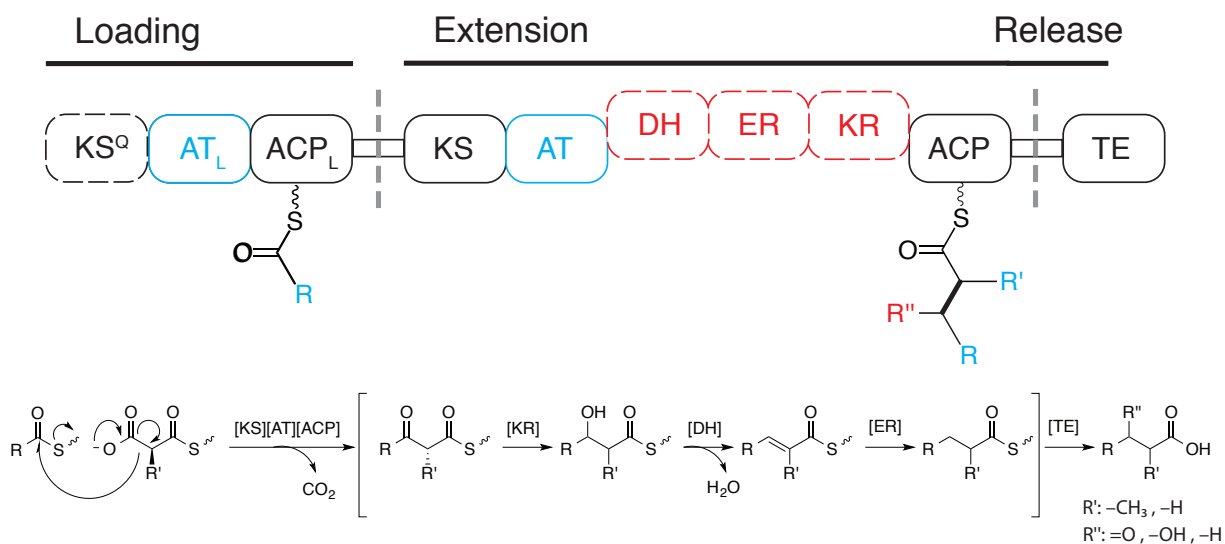
Figure 1.2: **Schematic of a PKS module.** Determination of polyketide chemical structure by a sequence of PKS catalytic domains. A loading AT_L (blue) selects a starter acyl-CoA and catalyzes the chain initiation reaction, thereby specifying the identity of the R group in the extension reaction product. This acyl-intermediate is transferred to the KS. The AT (blue) selects a malonyl-CoA derivative as an extender unit, which is subsequently transferred to the ACP, determining the R′ group in the extension reaction product. Next, the KS catalyzes a decarboxylative condensation between the KS bound acyl-intermediate and the ACP-bound extender unit to yield an ACP-tethered product, with the newly formed C-C bond shown in bold. The ketone located at the β position of the product may be optionally reduced by a KR to form a hydroxyl group, a DH to form an methenyl group, and a ER to form a methylene group. The presence or absence of these reductive domains (red) therefore determines the identity of the R″ group. This poly-β-ketone may undergo multiple extension steps before undergoing hydrolysis or cyclization by the TE to form either an acid or a lactone.

recombine existing PKSs thus offers a tantalizing promise of access to an expansive region of chemical space.

1.2 Using Enzymology to Inform Polyketide Synthase Engineering

To engineer PKSs, we first need to understand their intrinsic biosynthetic potential. While *in vivo* experiments can be used to make a rough assessment of a PKS's biosynthetic capabilities, the true enzymology of PKS enzymes may be obfuscated by variations in the cell environment. Thus, *in vitro* experiments are required to fully elucidate PKS enzymology. In

Chapter 2, we use *in vitro* enzyme experiments coupled with analytic chemistry techniques including high-performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC-MS), and gas chromatography-mass spectrometry (GC-MS) to reveal the unexpected broad substrate specificity of a truncated PKS subunit, and employ the resulting insights to produce a novel non-natural natural product analog *in vivo* by altering extracellular acyl-CoA concentrations [80, 82].

1.3 Engineering Ketoreductase Domains in Polyketide Synthases

Polyketides are produced with stringent stereochemical fidelity, and the importance of stereochemistry in determining bioactivity has made the determinants of stereochemistry in these pharmaceutically relevant molecules a research area of great interest. Previous work has demonstrated that AT domains specific for extender units with an α-substituent can only utilize substrates with (S) stereochemistry at the α position, such as (*2S*)-methylmalonyl CoA [10, 51]. This α-substituent undergoes inversion of configuration upon KS catalyzed condensation with the nascent poly-β-ketone [10, 73]. If present, the KR reduces the β-ketone on the resulting acyl-intermediate to form a hydroxyl group. The stereochemistry of this hydroxyl group has been demonstrated to be an intrinsic property of KR domains via domain exchange experiments [36]. In some cases, the α-substituent undergoes epimerization, a reaction that was also demonstrated to be catalyzed by the KR domain in recent isotope exchange experiments [24, 25]. Thus, KR domains determine the stereochemistry of both β-hydroxyl and α-alkyl groups. Naming conventions, shown in Figure 1.3, have been established to describe KRs that generate products exhibiting each of the four possible stereochemical outcomes [7, 59, 39]. We will use these naming conventions in our subsequent discussion of β-hydroxy and α-alkyl stereochemistry in PKS products.
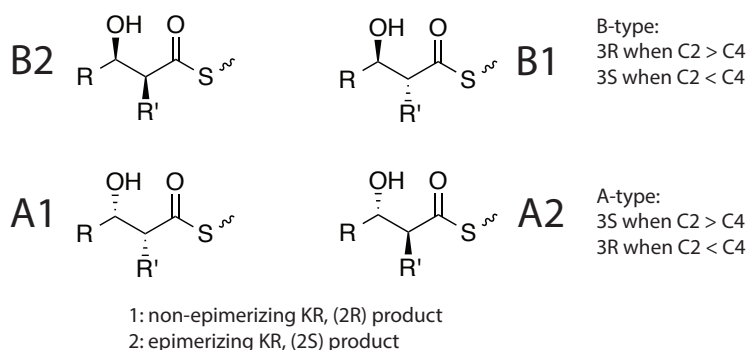
Figure 1.3: **KR naming conventions.** This naming convention is based on the four possible stereochemical outcomes from KR catalyzed reduction of the β-ketone in a growing poly-β-ketone that is produced by a PKS.

In Chapter 3, we investigate methods of engineering PKSs via both site-directed mutagenesis and domain exchanges, and report insights gained from domain exchange experiments in a truncated PKS model system. By designing chimeras that left a previously identified structural motif intact, we were able to develop a novel strategy for domain exchanges. Through this work, we reveal the importance of previously under-appreciated structural motifs for PKS engineering.

1.4 Developing General Polyketide Synthase Engineering Strategies

In spite of recent structural and mechanistic insights gained from structural biology techniques, including X-ray crystallography [41, 87, 71] and cryo-electron microscopy (cryo-EM) [16, 74], a comprehensive model of the protein-protein interactions and catalytic mechanisms that govern the fold and function of PKS modules remains elusive. Nevertheless, encouraging recent work has successfully combined intuition and expert domain knowledge with inspection of existing sequence alignments and X-ray crystal structures to design chimeric PKSs harboring a heterologous AT domain [81], a heterologous KR domain [18, 52, 36, 43], or a heterologous reductive cassette [29].

Our work [18] focuses on domain exchanges involving the KR domain. In our experiments, we employed the heuristic of seeking to minimize perturbations to the parent PKS module by conserving interaction elements believed to be important to PKS module fold and function, and by choosing donor KR domains that are evolutionarily closer to the KR present in the wild-type module. We also sought to minimize perturbations to the donor KR by favoring KR domains with native substrates that are more similar to the new substrate that would be presented by the parent PKS module. This allowed us to successfully carry out domain exchanges in our chosen model system. While it is not clear if these heuristics are generally applicable, our work provides an encouraging example of hypothesis-driven PKS design, demonstrating that taking advantage of established structural motifs can be an effective means of increasing the likelihood of success when engineering PKS chimeras.

In Chapter 4, we formalize the intuition gained from our KR domain exchange experiments into a paradigm for designing chimeric PKSs capable of producing small molecule compounds of interest, and discuss a computational platform developed to aid more generally in testing PKS design principles. In our experimental work, we found the existing computational tools lacking. In particular, existing databases were designed with a focus on interrogating genome sequences to predict metabolite structure [89]. We thus sought to develop software tools with the explicit objective of streamlining the design of experiments to test hypotheses regarding which donor catalytic domain features may be important in catalytic domain exchange experiments.

Includes material from published work:

Eng, C. H., Backman, T. W. H., Bailey, C. B., Magnan, C., Martin, H. G., Katz, L., Baldi, P., Keasling, J.D. ClusterCAD: a computational platform for type I modular polyketide synthase design. *Nucleic Acids Res.* (2017) gkx893, doi:10.1093/nar/gkx893.

Chapter 2

In vitro Enzyme Assays Inform Polyketide Synthase Engineering

2.1 The Lipomycin Synthase

The lipomycin synthase from *Streptomyces aureofaciens* Tü117 is a polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) hybrid, and is responsible for producing the acyclic, polyene antibiotic α-lipomycin [4]. Previous characterization of the cognate *lip* gene cluster resulted in the identification of genes encoding four PKS subunits, a NRPS, and a methyltransferase, that together produce β-lipomycin, the aglycon precursor to α-lipomycin.
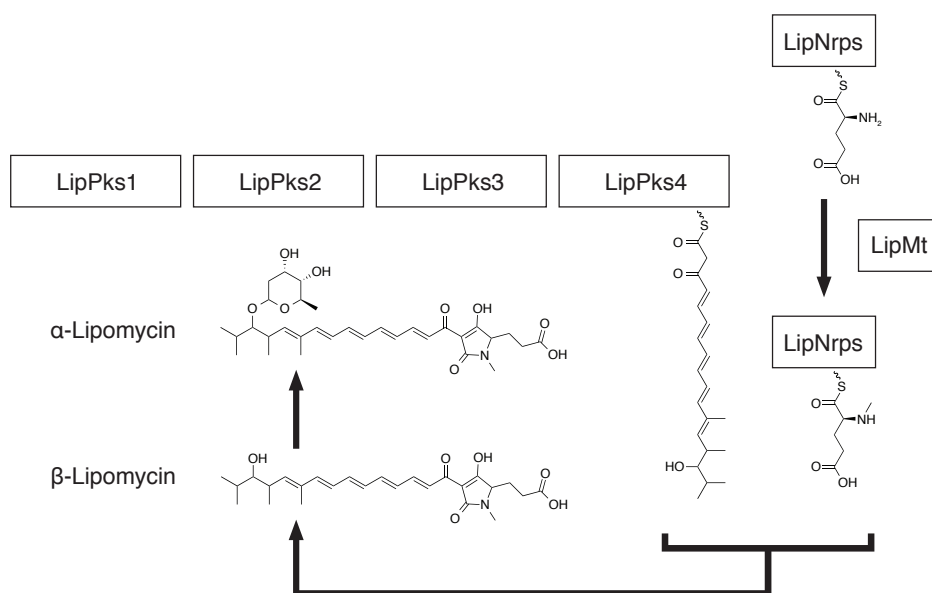
Figure 2.1: **Schematic of the lipomycin synthase.** The lipomycin synthase from *S. aureofaciens* Tü117 contains four PKS subunits, a NRPS, and a methyltransferase.

2.2 *In vitro* Characterization of LipPks1+TE

We focus on LipPks1, the first PKS subunit in the lipomycin synthase. LipPks1 contains a loading didomain (AT_L, ACP_L), a set of extension domains (KS, AT, ACP), and a single reducing domain (KR) which together make up the first module of the PKS. The AT_L domain of LipPks1 shows 50% amino acid sequence identity to the AT_L domain of the avermectin PKS from *Streptomyces avermitilis*, which is known to accept a variety of starter substrates [4, 27, 17]. The sequence similarity of the two AT_L domains suggests that the AT_L of LipPks1 may also exhibit a similarly relaxed substrate specificity.

In order to test this hypothesis, LipPks1 was expressed with the TE domain from DEBS appended to the C-terminus to enable the release of products from the terminal ACP [80]. Further, a N-terminal hexahistidine tag was used to facilitate purification of the resulting LipPks1+TE for *in vitro* analysis. Variants lacking either the arginine-rich N-terminal linker or the TE, (-NL)LipPks1+TE and LipPks1 respectively, were also constructed. The *Escherichia coli* strain K207-3 [55], whose genome contains the *sfp* gene from *Bacillus subtilis*, was used as an expression host. The *sfp* gene encodes a substrate promiscuous surfactin phosphopantetheinyl transferase that converts expressed *apo* PKS proteins to their corresponding *holo* forms.

Protein production was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). As shown in Figure 2.2, LipPks1+TE was almost exclusively produced in the soluble fraction, while a substantial amount of (-NL)LipPks1+TE was insoluble. Additionally, the (-NL)LipPks1+TE lysate contained a large amount of soluble protein that ran between 160 kDa and 260 kDa. These observations suggest that the proline and arginine-rich N-terminal linker contributes to stable protein folding, at least in *E. coli*. Surprisingly, a relatively low level of protein production was observed for LipPks1, which corresponds to a natural LipPks1 but lacks the C-terminal linker that is believed to bind the N-terminal linker of the LipPks2 subunit. This result suggests that the linker may itself stabilize the LipPks1 structure, or that adding a TE domain stabilized the homodimeric structure of LipPks1. Structural studies of DEBS, which is the most extensively studied multimodular PKS, demonstrated that each module forms a homodimer and that homodimerization is driven by the KS and TE domains [63, 66, 65].

To evaluate the activity of LipPks1+TE, the enzyme was first purified by nickel affinity chromatography followed by anion exchange chromatography to yield 2 mg of protein from 1 L of *E. coli* culture. Isobutyryl-CoA (**1a**), methylmalonyl-CoA, and NADPH were incubated in the absence or presence of LipPks1+TE, and the production of 3-hydroxy-2,4-dimethylpentanoate (**2a**) monitored by LC-MS. To verify the identity of the enzymatically produced product, we chemically synthesized **2a**. The ^{1}H-NMR data indicated that the synthesized compound was a mixture of diastereomers (*syn*:*anti* = 26:1) [68].

MS measurements were carried out in the selected ion-monitoring mode ($m/z = 145$). The enzyme experiments yielded a strong signal the presence of LipPks1+TE at 9.0 min, while the chromatogram of authentic standard contained two peaks at 8.2 min and 9.1 min. This led us to conclude that the product generated by LipPks1+TE has same stereochemistry

(a) Soluble.

(b) Insoluble.

(c) Purified LipPks1+TE.
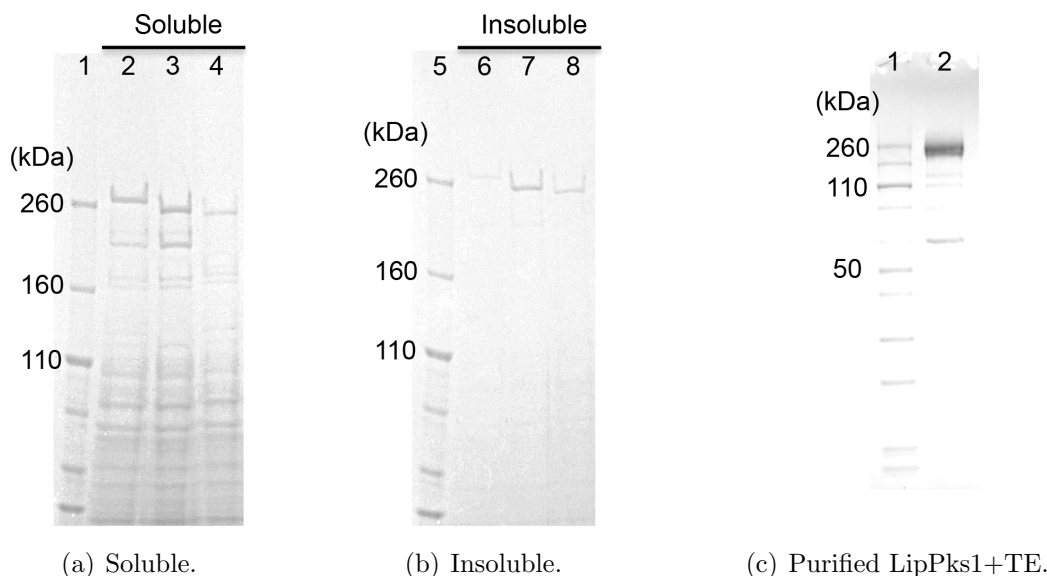
Figure 2.2: **SDS-PAGE analysis of recombinant LipPks1.** LipPks1+TE (259 kDa), (-NL)LipPks1+TE (239 kDa), and LipPks1 (229 kDa) are shown. Proteins were resolved using a 3-8% Tris-Acetate gel and stained with Coomassie blue. Lanes 1 and 5, molecular weight markers; Lanes 2 and 6, LipPks1+TE; Lanes 3 and 7, (-NL)LipPks1+TE; Lanes 4 and 8, LipPks1.

Table 2.1: Steady-state kinetics of LipPks1-TE

Substrate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})
isobutyryl-CoA	0.053	2.9	304.1
propionyl-CoA	0.056	13.4	70.3
n-butyryl-CoA	0.036	26.4	22.7
2-methylbutyryl-CoA	0.126	8.8	237.0
isovaleryl-CoA	0.290	128.1	38.0
pivaloyl-CoA	0.002	8.8	4.1

as that of the second peak, which corresponds to (*2S, 3S*)-1a and/or (*2R, 3R*)-**1a**. This was consistent with the amino acid sequence of the KR domain, which suggested the (*2S, 3S*) product. We also monitored the β-keto product of **1a** by LC-MS. No product was observed in the presence of NADPH. The steady-state kinetic parameters for the reaction were assessed at saturating conditions of methylmalonyl-CoA using a Michaelis-Menten model, and are shown in Table 2.1.

To analyze the substrate tolerance of the AT$_L$ domain of LipPks1, LipPks1+TE was

also incubated with one of acetyl-CoA, propionyl-CoA, n-butyryl-CoA, 2-methylbutyryl-CoA, isovaleryl-CoA, or pivaloyl-CoA at the saturation concentration of methylmalonyl-CoA (200 μM) in the presence of NADPH. While no product was observed when acetyl-CoA was used as a starter substrate, the remaining acyl-CoAs gave the corresponding products, 3-hydroxy-2-methylpentanoate (**2b**), 3-hydroxy-2-methylhexanoate (**2c**), 3-hydroxy-2,4-dimethylhexanoate (**2d**), 3-hydroxy-2,5-dimethylhexanoate (**2e**), and 3-hydroxy-2,4,4-trimethylpentanoate (**2f**), respectively. Production of each compound was confirmed by LC-MS analysis with a corresponding chemically synthesized standard. The broad substrate specificity of LipPks+TE is illustrated in Figure 2.3.
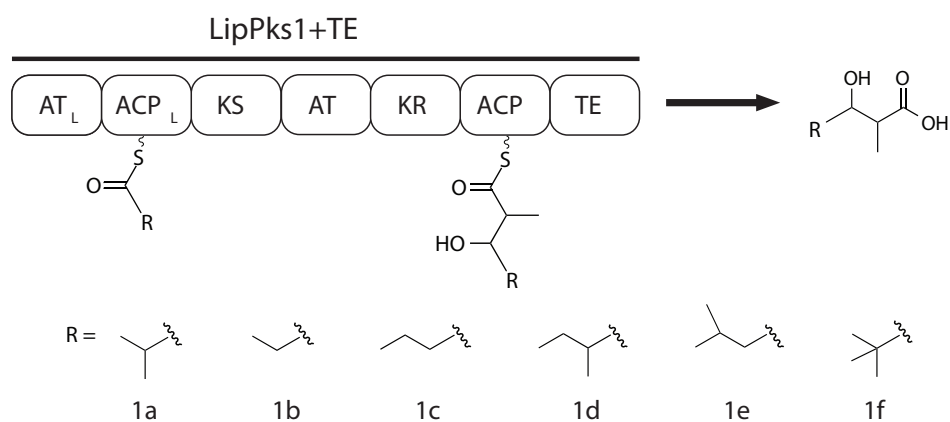
Figure 2.3: **Broad substrate specificity of LipPks1+TE.** LipPks1+TE was empirically demonstrated to accept propionyl-CoA, n-butyryl-CoA, 2-methylbutyryl-CoA, isovaleryl-CoA, and pivaloyl-CoA as substrates, in addition to the natural substrate, isobutyryl-CoA.

The steady-state kinetic parameters for hydroxyacid production by LipPks+TE using different starter substrates are shown in Table 2.1. The k_{cat} values for all substrates investigated, with the exception pivaloyl-CoA, were similar; the k_{cat} value for pivaloyl-CoA was substantially lower than the others. The basis for the relatively low k_{cat} with pivaloyl-CoA is not well understood, but may be related to the different rates of nucleophilic attack of the carbanion on the KS-linked substrates in the chain elongation reactions. Furthermore, the $k_{\text{cat}}/K_{\text{M}}$ values for isobutyryl-CoA and 2-methylbutyryl-CoA were comparable, which is consistent with the observation that both isobutyryl- and 2-methylbutyryl-started products have been isolated from *S. avermitilis* [61, 9]. Interestingly, the α-lipomycin analog that would be produced using 2-methylbutyryl-CoA as a starter was not observed in *S. aureofaciens* Tü117 [4, 46]. These data suggest that the production of α-lipomycin is tightly controlled by regulation of intracellular acyl-CoA concentrations, and that a lipomycin analog using 2-methylbutyryl-CoA as a starter unit has not been previously observed simply because LipPks1 has thus far only been characterized under conditions in which the intracellular concentrations of 2-methylbutyryl-CoA were far below the K_{M} for the enzyme.

The substrate specificity of the loading didomain in multimodular PKSs is usually assumed from the structure of the corresponding segment of polyketide products. However, this methodology only reveals the compounds produced under particular culture conditions, and might not reflect the actual substrate preference . The results in Table 2.1 emphasize that intrinsic substrate specificity can only be determined by *in vitro* kinetic analysis [49]. Such analysis of substrate specificity is crucial to our understanding of polyketide biosynthesis by multimodular PKSs and to the design of hybrid PKSs. We have uncovered broad substrate specificity of the loading didomain of the lipomycin synthase, which was not anticipated from the study of the PKS in its native host, and led us to speculate that novel lipomycin analogs can be produced by increasing intracellular acyl-CoA concentrations.

2.3   *In vivo* Production of the Novel Natural Product Analog 21-Methyl-α-Lipomycin

In vitro analysis of LipPks1+TE revealed that the enzyme is capable of efficiently utilizing 2-methylbutyryl-CoA as a starter unit, with k_{cat} and K_{M} values for 2-methylbutyryl-CoA that are comparable to those for isobutyryl-CoA, as shown in Table 2.1. These *in vitro* kinetic parameters suggest that an analog of α-lipomycin in which 2-methylbutyryl-CoA replaces isobutyryl-CoA in the chain initiation reaction, can be produced by increasing the intracellular concentration of 2-methylbutyryl-CoA in *S. aureofaciens* Tü117, although the expected α-lipomycin analog, 21-methyl-α-lipomycin, has not been previously reported in *S. aureofaciens* Tü117.

To test whether this observation was indeed the result of insufficient intracellular 2-methylbutyryl-CoA levels, *S. aureofaciens* Tü117 was cultured in 1 L of HA medium (1% malt extract, 0.4% yeast extract, 0.4% glucose and 1 mM $CaCl_2$) containing 20 mM isoleucine. As a control, the strain was also grown in 1 L of HA medium in the absence of isoleucine. After 7 days, mycelia of each culture were extracted with acetone. After the extracts were concentrated, they were combined with their corresponding supernatants and acidified with HCl (pH $\approx$ 4). The resulting aqueous solutions were extracted with ethyl acetate and concentrated. LC-MS analysis revealed that in media supplemented with 20 mM isoleucine, production of a methylated α-lipomycin analog, confirmed by structural analysis, is favored over that of α-lipomycin in a 23:8 ratio, as compared to a 1:16 ratio in unaltered media.

In order to confirm the structure of the analog, we purified the analog by HPLC, obtaining 10 mg from 1 L of the HA+isoleucine culture, a yield comparable to that of α-lipomycin purified from 1 L of HA culture (9 mg). LC-MS analysis of α-lipomycin showed the presence of two sets of intense signals at $m/z = 588.4$, 589.4, 590.4 and 591.4 and at $m/z = 458.3$, 459.3, 460.3 and 461.3, which correspond to [M+H]$^+$ α-lipomycin (theoretical exact mass = 588.31) and β-lipomycin (theoretical exact mass = 458.25), respectively. Because we observed a single peak in the LC, we concluded that the MS signals for β-lipomycin were the fragmented ions of α-lipomycin. Analysis of the analog also showed the presence of two sets

(a) α-Lipomycin

(b) 21-Methyl-α-lipomycin
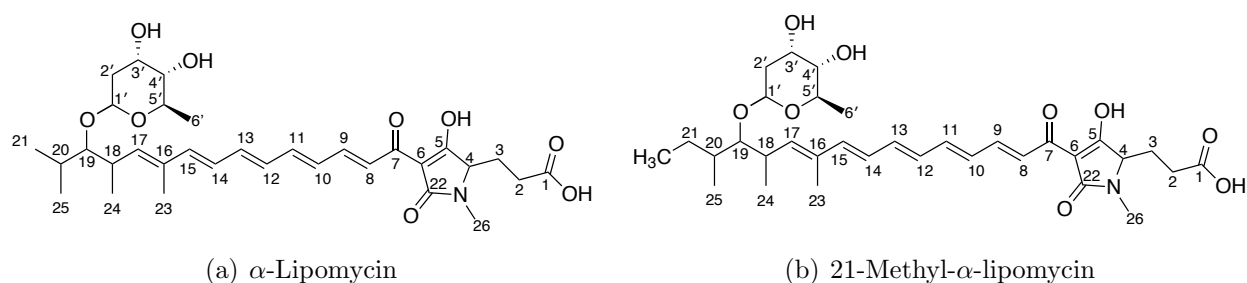
Figure 2.4: **Chemical structures of α-lipomycin and 21-methyl-α-lipomycin.**

of intense signals at $m/z = 602.4$, 603.4, 604.4 and 605.4 and at $m/z = 472.3$, 473.3, 474.3 and 475.3. The theoretical exact mass of $[\text{M} + \text{H}]^+$ of 21-methyl-α-lipomycin is 602.33. The fragmented ions indicate that both α-lipomycin and the analog have the same sugar moiety, and that the $+14$ mass originates from a change in the backbone structure.

The orange-red color of α-lipomycin stems from the conjugated double-bond structure. The UV spectrum of α-lipomycin exhibited an absorbance at 448 nm with $\epsilon_{448} = 29{,}800$. As expected, the spectrum of the analog was nearly identical to that of α-lipomycin, $\lambda_{\max} = 442$ nm with $\epsilon_{442} = 27{,}600$. These data indicate that the analog has a very similar p-electron conjugation system to α-lipomycin.

The structure of the methylated α-lipomycin analog was fully elucidated by NMR. The ^{1}H-NMR spectra for the natively produced antibiotic and the analog were very similar, with the exception of the chemical shifts corresponding to the protons H20, H21, and H25 in both compounds, in addition to the H$_3$C group of the analog, as shown in Figure 2.5 and in Table 2.2. The reported proton and carbon signals were assigned by analysis of 2D NMR spectra (DQF-COSY, TOCSY, and HSQC). The proton signal for H21 in the analog is a multiplet at δ_H 1.68 p.p.m., whereas the corresponding proton in α-lipomycin is a doublet at δ_H 0.99 p.p.m. Furthermore, a new proton signal, corresponding to the additional H$_3$C group, appeared at 0.92 p.p.m. in the ^{1}H-NMR spectra for the analog. On the basis of these NMR spectra, in conjunction with the MS and UV data, we determined the analog structure to be 21-methyl-α-lipomycin.

We also report the minimum inhibitory concentration (MIC) of β-lipomycin and the analog against *B. subtilis*, a representative Gram-positive bacterium and *E. coli*, a representative Gram-negative bacterium. The potencies were the same for the two compounds: MIC for *B. subtilis* = 0.78 μg/ml and MIC for *E. coli* >200 μg/ml, indicating that the activity is not affected by the structure of the starter substrate.

Table 2.2: ^{1}H and ^{13}C NMR chemical shifts of compounds in methanol-d_4

| Position | α-Lipomycin | | 21-Methyl-α-lipomycin | |
	δ_H	δ_C	δ_H	δ_C
2	2.14 (m, 1H)	25.05	2.13 (m, 1H)	24.98
	2.26 (m, 1H)		2.25 (m, 1H)	
3	2.26 (m, 3H)	29.02	2.27 (m, 3H)	29.04
4	3.90 (t, 1H)	66.24	3.87 (t, 1H)	66.39
8-15	6.26-8.11 (8H)	121.73-145.97	6.25-8.09 (8H)	122.25-145.67
17	5.71 (d, 1H)	139.88	5.72 (d, 1H)	139.70
18	2.87 (m, 1H)	37.47	2.88 (m, 1H)	37.09
19	3.33(dd, 1H)	88.63	3.42 (dd, 1H)	88.69
20	1.79 (m, 1H)	32.21	1.55 (m, 1H)	39.33
21	0.99 (d, 3H)	20.41	1.68 (m, 1H)	25.91
23	1.86 (s, 3H)	12.79	1.85 (s, 3H)	12.75
24	1.05 (d, 3H)	18.62	1.07 (d,3H)	18.90
25	0.96 (d, 3H)	18.80	0.91 (d, 3H)	16.26
H$_3$C	—	—	0.92 (d, 3H)	11.97
1'	4.87 (t, 1H)	100.42	4.87 (t,1H)	100.62
2'	1.65 (m, 1H)	39.57	1.65 (m, 1H)	39.56
	2.05 (m, 1H)		2.05 (m, 1H)	
3'	4.03 (dt, 1H)	69.39	4.03 (dt, 1H)	69.39
4'	3.18 (dd, 1H)	74.50	3.18 (dd, 1H)	74.48
5'	3.71 (m, 1H)	70.59	3.73 (m, 1H)	70.61
6'	1.27 (d, 3H)	18.62	1.26 (d, 3H)	18.62
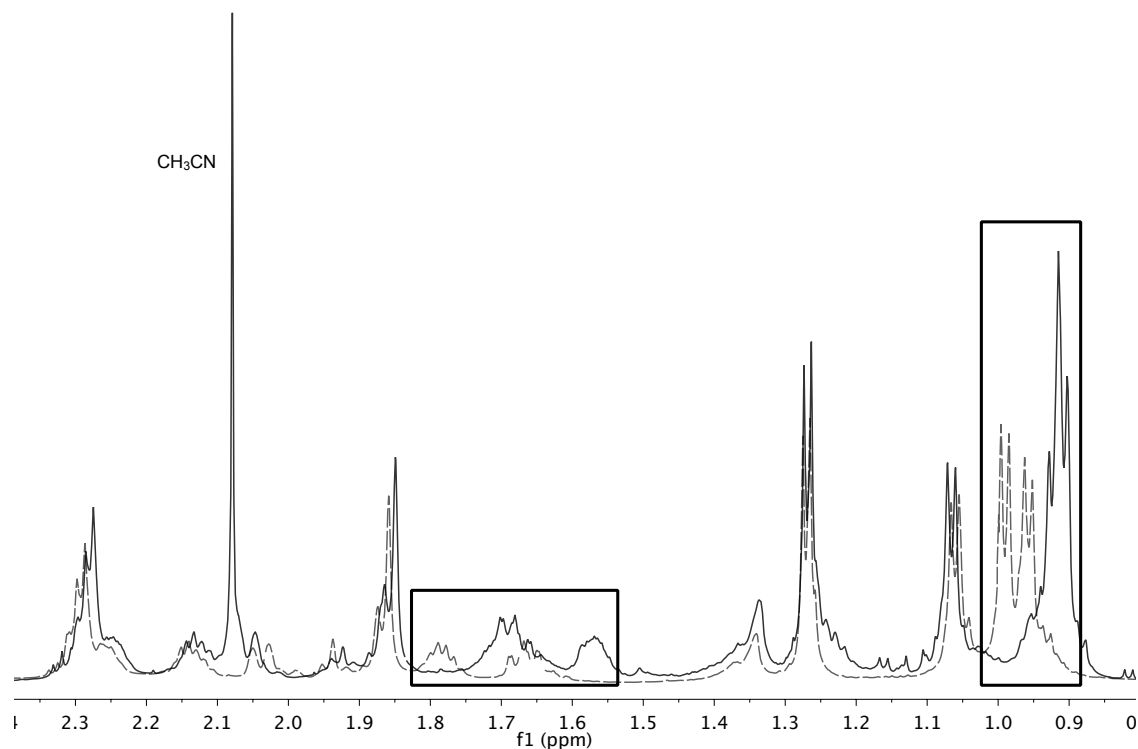
Figure 2.5: **Overlay of ^{1}H-NMR spectra of α-lipomycin and 21-methyl-α-lipomycin.** α-Lipomycin is shown as a dotted line and 21-methyl-α-lipomycin as a solid line. The regions containing the primary differences between the two spectra are boxed.

2.4   Conclusion

The avermectin family of polyketides is produced in *S. avermitilis* by a single PKS system whose loading didomain is capable of using isobutyryl-CoA or 2-methylbutyryl-CoA to start the synthesis [5, 34]. Because only a single lipomycin congener was isolated from *S. aureofaciens* Tü117 [46, 4, 60, 83], it was assumed that the LipPks1 loading didomain was specific for isobutyryl-CoA. However, our kinetic studies of LipPks1 revealed that the $k_{\mathrm{cat}}/K_{\mathrm{M}}$ values were comparable between isobutyryl-CoA and 2-methylbutyryl-CoA [80], suggesting that the lipomycin synthase was capable of producing a methylated analog of α-lipomycin. In this work, we demonstrated that increasing the intracellular concentration of 2-methylbutyryl-CoA enabled the production of 21-methyl-α-lipomycin in the native producer.

The distribution of products generated by an enzyme *in vivo* is directly dependent on relative intracellular substrate concentrations, hence the intrinsic substrate specificity of an enzyme can only be determined *in vitro*. By using *in vitro* enzyme analysis of LipPks1+TE

to inform modifications to the culture medium that modulated intracellular acyl-CoA concentrations, we achieved production of a novel antibiotic analog in the native producer. This illustrates how *in vitro* analysis may be applied to reveal methods of producing natural product analogs *in vivo*, thereby expanding the space of drug candidates and accelerating drug discovery.

Includes material from published work:

Yuzawa, S., **Eng, C. H.**, Katz, L., Keasling, J. D. Enzyme analysis of the polyketide synthase leads to the discovery of a novel analog of the antibiotic α-lipomycin. *J. Antibiot.* (2014) **67**, 199–201, doi:10.1038/ja.2013.110.
Yuzawa, S., **Eng, C. H.**, Katz, L., Keasling, J. D. Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase. *Biochemistry* (2013) **52**, 3791–3793, doi:10.1021/bi400520t.

Chapter 3

Altering Stereochemistry via Ketoreductase Domain Engineering

3.1 Introduction to Engineering Ketoreductase Domains in Polyketide Synthases

Type I PKSs are responsible for the production of structurally diverse polyketide natural products, including those that function as antibacterial, antiparasitic, and antitumor agents. The medicinal value of polyketides has fueled an interest in exploring the chemical space of natural product analogs. While the chemical modification of existing polyketides has facilitated the production of more active derivatives in some cases [72], the structural and stereochemical complexity of these natural products remains a substantial barrier to employing semi-synthetic methods. The stringent chemo-, regio- and stereospecificity of enzyme catalysts makes protein engineering a promising alternative.

Given the importance of stereochemistry in determining biological activity, we sought to discover reliable strategies of altering polyketide stereochemistry. AT domains exclusively utilize α-substituted substrates with (S) stereochemistry [10, 51]. This α-substituent undergoes inversion of configuration upon KS catalyzed condensation of an extender unit with the nascent acyl donor chain [10, 73], as shown in Figure 3.1. When present, the KR carries out NADPH-dependent reduction of the β-ketone, exercising stereoselective control over β-hydroxyl stereochemistry mediated by different binding modes that control the face of keto group attack by the 4-pro-S hydride of the NADPH cofactor [59, 7]. The KR domain may also perform additional epimerization of the α-substituent [24, 25, 2], with stereospecific reduction carried out only on one epimer. A schematic of the proposed mechanisms of KR-catalyzed reduction and epimerization are shown in Figure 3.2. Naming conventions, shown in Figure 3.1, have been established to describe KRs that generate products exhibiting each of the four possible stereochemical outcomes [7, 59, 39].

While previous studies have reported the application of both site-directed mutagenesis and domain exchange strategies in attempts to engineer KR stereochemistry, the success of
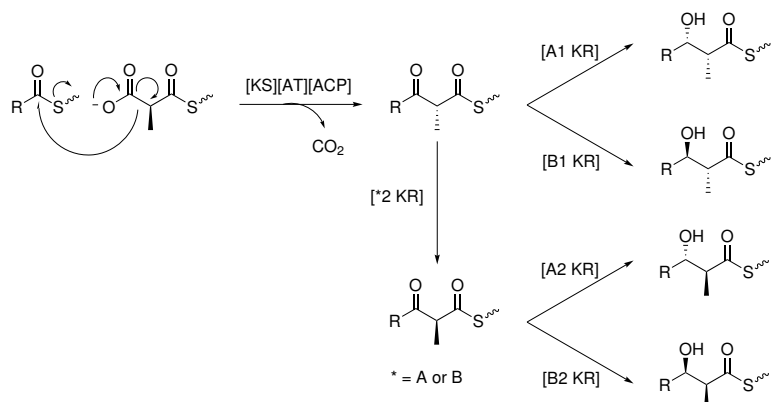
Figure 3.1: **Possible stereochemical outcomes of KR-catalyzed reduction.** Naming conventions based on the four possible stereochemical outcomes from KR-catalyzed reduction of the β-ketone in a growing poly-β-ketone.

(a) Reduction
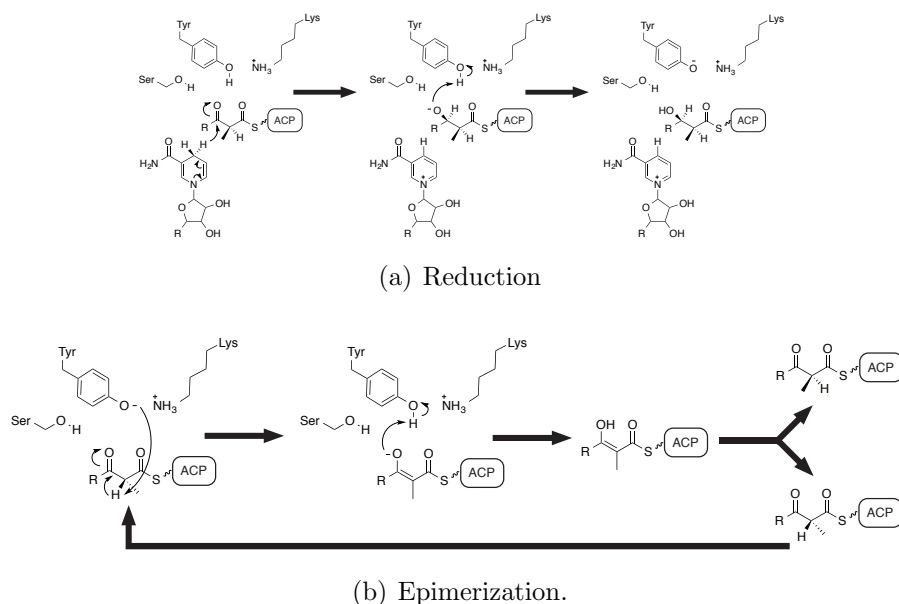
(b) Epimerization.

Figure 3.2: **Proposed mechanism of KR-catalyzed reduction and epimerization.**

these efforts have been mixed, motivating our desire to develop more generalizable methods applicable to a variety of parent PKS modules and donor KR domains.

Site-directed mutagenesis

Although site-directed mutagenesis has yielded some successes in efforts to alter the stereochemistry of isolated KR domains, this strategy has proven less fruitful in the context of

full PKS modules. Researchers have previously identified "fingerprint" residues conserved in KR domains that produce products with each possible stereochemical outcome at the α and β positions [7]. These KR "fingerprints" have been used to inform mutagenesis studies aimed at altering KR stereochemistry [3, 88, 86].

However, the few successes achieved in isolated KR domains acting on N-acetylcysteamine substrate analogs (SNACs)[78] have proven non-transferable to the context of a full PKS module acting on ACP-bound substrates. The point mutations used to alter EryKR1 stereochemistry from B2 type to A2 type in an isolated KR domain acting on a SNAC substrate analog [3] either had no effect or abolished activity in the context of a bimodular PKS [47]. Similarly, while point mutations were successful in altering AmpKR2 stereochemistry from A1 type to A2 type in a isolated KR domain presented with a SNAC substrate analog, after incorporation into a bimodular PKS, the mutated AmpKR2 failed to yield the expected product when presented with an ACP-bound substrate [86]. Interestingly, the mutated AmpKR2 was capable of reducing a SNAC substrate analog with altered stereochemistry even within the context of a bimodular PKS. One possibility is that these engineered KR domains are simply unable to compete with the downstream TE within a PKS module, leading to hydrolysis of the β-ketoacyl-ACP substrate [43, 47, 86].

Domain exchanges

There are several examples of successful domain exchange experiments achieving altered KR stereochemistry within the context of a full PKS module. In one early study, researchers carried out KR domain replacements in a three module, truncated DEBS, and found that replacement of EryKR2 with EryKR5 resulted in production of the expected tetraketide, while replacement of EryKR2 with RapKR2 from the rapamycin synthase yielded a triketide lactone, suggesting an aborted product due to failed recognition of the substrate by the third module in the truncated PKS [36]. This triketide lactone was the enantiomer of the product produced by a bimodular, truncated DEBS harboring the native EryKR2 [52]. While additional results have also been reported in the literature [37], success has generally been limited: previously reported engineered PKSs harboring non-native reductive loops typically either fail to produce the expected products or produce the expected product with greatly attenuated yield and/or stereospecificity [23, 37, 52].

Kellenberger and co-workers [43] suggest that failed domain exchange efforts may be partially explained by the use of sub-optimal fusion points between parent PKS modules and donor reductive loops, and advocate testing multiple fusion points for each parent-donor pair. The researchers replaced the reductive loop in the second module of a bimodular truncated DEBS with one of two synthetic oligonucleotide linkers, incorporating a number of restriction sites to facilitate the introduction of non-native reductive loops comprised of a subset of KR, DH, and ER domains. The researchers found that different reductive loops demonstrated higher activity and greater stereoselectivity when different restriction sites were used. However, there remains no reliable method for identifying optimal fusion points for KR domain exchange experiments.

Selecting a KR engineering strategy

Given that our primary interest was developing a strategy to engineer PKSs capable of producing novel natural product analogs, we chose a domain exchange strategy in order to increase our chances of maintaining activity in a PKS module context. We surmised that the previous failures in domain exchange experiments may be attributed not only to the use of suboptimal fusion points, but also to the use of suboptimal donor KR domains.

An additional layer of complexity stemmed from the fact that the stereochemical outcome of KR-catalyzed reduction appears to also be dependent on substrate context, and not solely on the KR domain itself. Studies investigating the activity of isolated KR domains on diketide-SNACs showed that KR domains that typically act on longer polyketide chains exhibit unexpected low stereospecificity on these shorter substrates [62]. Another study reported that the length of the hydrocarbon tail adjacent to the β-ketone, as well as the type of ester in β-keto ester substrate analogs are also important in the maintenance of stereochemical fidelity [26]. Experiments using an isolated A1-type KR domain suggested that for KR domains with native substrates containing an α-methyl group, the presence of that methyl group is also important for determining β-hydroxy stereochemistry [88].

We combined sequence alignments of KR domains with information on the predicted native substrates of each KR domain to rank the most promising donor KR domains for domain exchange experiments. By choosing donor KR domains that exhibited high sequence similarity with the KR domain in the parent PKS module, we sought to minimize perturbations to the remainder of the parent PKS module following a domain exchange. Similarly, by choosing donor KR domains with native substrates similar to the poly-β-ketone produced by the parent PKS module, we hoped to minimize perturbations to the function of the donor KR domain in its new PKS module context.

3.2 Ketoreductase Domain Exchanges in LipPks1+TE

In this study, we focused on altering α-methyl stereochemistry using KR domain exchanges. Previous work has demonstrated successful substitution of both A1 and B1 type heterologous KR domains for an A1 type KR domain (Ery KR$_2$) in bimodular and trimodular derivatives of DEBS with the TE from module 6 appended [52, 36, 43]. In select cases, domain exchanges successfully inverted β-hydroxyl stereochemistry with no loss of activity. More recently, researchers also used a bimodular derivative of DEBS to alter α-methyl stereochemistry at the expense of greatly decreased titers of products [2]. We sought to alter α-methyl stereochemistry with no loss of activity.

Our work also provided the first example of KR domain exchanges in an alternative model system, LipPks1+TE [80], a N-terminally hexahistidine tagged recombinant polyketide synthase subunit that consists of the first module of the lipomycin synthase from *S. aureofaciens* Tü117, with the TE from DEBS appended to the C-terminus. Similarly to DEBS module

2, LipPks1 harbors only a KR in its reductive cassette. However, unlike DEBS module 2, LipPks1 harbors an A2, rather than A1 type KR, and also natively contains a recently identified $\sim$55 amino acid dimerization motif that lies between the AT and KR in $\sim$50% of the PKS modules [88] that contain a single reducing catalytic domain. Previous KR domain exchange experiments have treated the dimerization motif as part of the KR domain when designing KR domain exchanges [43, 2]. We suspected that the dimerization motif may stabilize protein structure through interactions with the remainder of the PKS module, and sought to probe the importance of this dimerization motif biochemically.

As an initial test of the importance of the dimerization motif, we constructed expression vectors corresponding to variants in which the native Lip KR_1 was substituted for one of two donor KR domains: A1 type Amph KR_2 or B1 type Con KR_2. For each of these donor KR domains, two variants were constructed differing only in the dimerization motif, which was taken either from the parent PKS module, LipPks1, resulting in (Amph KR_2)LipPks1+TE and (Con KR_2)LipPks1+TE, or from the PKS module that furnished the donor KR domain, yielding (Amph DE,KR_2)LipPks1+TE and (Con DE,KR_2)LipPks1+TE. An expression vector for a variant lacking a dimerization motif, (-DE)LipPks1+TE was also constructed.

The *E. coli* strain K207-3 [55], which has a substrate promiscuous surfactin phosphopantetheinyl transferase from *B. subtilis* integrated into the genome, was used as the expression host for all LipPks1+TE variants. With the exception of (-DE)LipPks1+TE, all variant PKS proteins could be well purified using Ni affinity chromatography followed by anion exchange chromatography. Each preparation resulted in 0.5-2 mg/L of purified protein. While (-DE)LipPks1+TE could not be purified, SDS-PAGE revealed that LipPks1+TE and (-DE)LipPks1+TE were expressed similarly, with half of the total protein recovered in the insoluble fraction, as shown in Figure 3.3. This suggests that the dimerization motif may be important for protein stability, rendering variants lacking the dimerization motif more susceptible to denaturation outside of the native cellular environment.

To evaluate reductive ability, each LipPks1+TE variant was incubated overnight with isobutyryl-CoA (**1a**), methylmalonyl-CoA, and NADPH. The subsequent production of 3-hydroxy-2,4-dimethylpentanoic acid (**3a**) was quantified using LC-MS with an authentic standard. In order to distinguish between variants that were not competent in condensation, and those that were competent in condensation but incapable of reduction, LipPks1+TE variants were also incubated overnight against **1a** and methylmalonyl-CoA in the absence of NADPH to form a diketide product independent of the reductive ability of the KR. The diketide was then heated overnight to form a more stable ketone, 2-methylpentan-3-one (**2a**), with titers again determined by LC-MS with an authentic standard. The ketone and hydroxyacid products monitored in this study are shown in Figure 3.4

(Amph KR_2)LipPks1+TE produced **2a** and **3a** in titers comparable those of wild-type LipPks1+TE, as shown in Table 3.1. By contrast, (Amph DE,KR_2)LipPks1+TE exhibited greatly attenuated activity in both condensation and reduction assays. While neither of the two Con KR_2 mutants yielded observable **3a**, only (Con KR_2)LipPks1+TE was condensation competent.

To determine whether a strategy of retaining the native DE is broadly applicable in
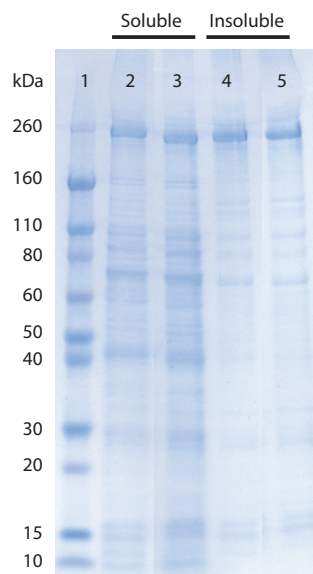
Figure 3.3: **SDS-PAGE analysis of recombinant LipPks1+TE.** LipPks1+TE (259 kDa) and (-DE)LipPks1+TE (253 kDa) are shown. Proteins were resolved using a 4-15% Tris-Glycine gel and stained with Coomassie blue. Lane 1, molecular weight marker; Lanes 2 and 4, LipPks1+TE; Lanes 3 and 5, (-DE)LipPks1+TE.
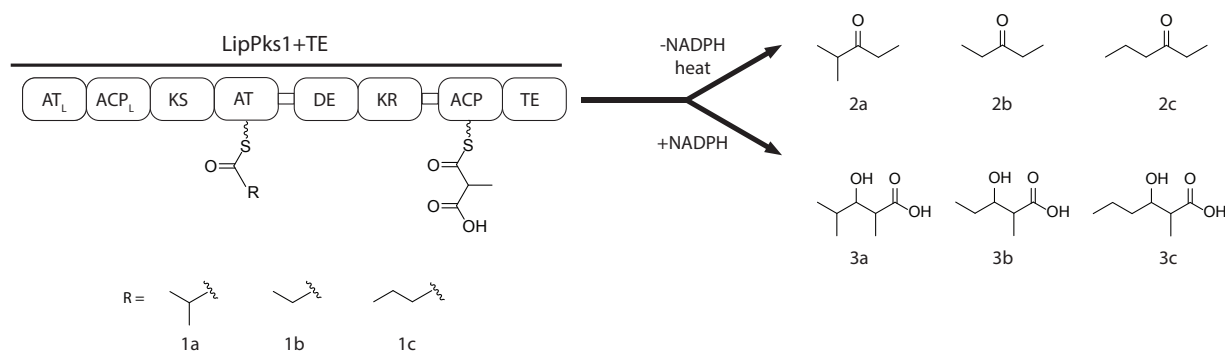
Figure 3.4: **Products generated by LipPks1+TE variants in this study.**

KR domain exchanges, we substituted additional B type (Bor KR$_1$), B2 type (Ery KR$_1$), A1 type (Spn KR$_3$, Ery KR$_6$) and A2 type (Amph KR$_1$) KR domains into LipPks1+TE, in each case keeping the native DE intact. Interestingly, with the exception of Con KR$_2$, condensation activity was almost completely retained no matter which donor KR was used. This suggests that the dimerization motif may confer the increased overall stability required for condensation, or may have important interactions with other regions of the PKS module that are mechanistically important for condensation activity. One possibility is that Con KR$_2$ has an unusual structure that disrupts the dimerization motif. It was further noted that all

of the A-type domain exchange variants were capable of achieving at least moderate levels
of reduction, though none of the B-type domain exchanges were successful. One explanation
is that dimerization motif plays a role in A-type activity, and hinders B-type activity.

Table 3.1: Production of corresponding ketones (**2a-c**) or acids (**3a-c**) LipPks1+TE variants
after overnight incubation with substrates **1a-c** in TTNs. Mean and standard error of
experiments performed in triplicate are reported.
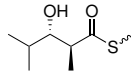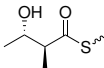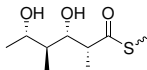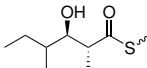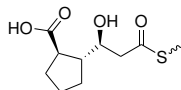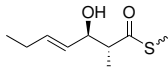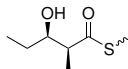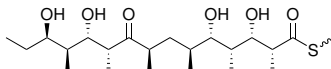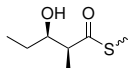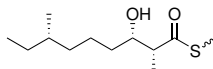
Substrate		1a		1b		1c	
Product		2a	3a	2b	3b	2c	3c
Lip KR$_1$	A2	84.4 ± 13.2	65.2 ± 2.9	99.8 ± 14.3	27.2 ± 1.3	93.0 ± 3.0	49.0 ± 6.6
Amph KR$_2$	A1	70.0 ± 5.0	60.4 ± 7.0	43.6 ± 3.0	19.6 ± 2.4	66.8 ± 7.7	48.2 ± 3.1
Amph KR$_2$[a]	A1	15.9 ± 2.4	7.3 ± 3.8	0.0 ± 0.0	3.6 ± 1.3	13.4 ± 2.2	7.9 ± 3.7
Con KR$_2$	B1	17.4 ± 4.4	0.0 ± 0.0	9.1 ± 0.7	0.0 ± 0.0	23.1 ± 0.4	1.1 ± 0.9
Con KR$_2$[a]	B1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.8 ± 0.6
Ery KR$_1$	B2	55.1 ± 12.4	0.0 ± 0.0	51.0 ± 7.5	0.6 ± 0.1	61.8 ± 2.6	0.0 ± 0.0
Bor KR$_1$	B	73.0 ± 7.6	0.0 ± 0.0	33.2 ± 4.6	0.0 ± 0.0	55.8 ± 5.9	0.9 ± 0.7
Spn KR$_3$	A1	66.9 ± 5.5	22.7 ± 0.7	94.6 ± 13.8	27.2 ± 1.1	104.9 ± 7.8	57.9 ± 2.4
Amph KR$_1$	A2	70.4 ± 2.5	25.6 ± 0.4	91.7 ± 7.7	34.2 ± 5.4	114.3 ± 7.7	20.1 ± 1.3
Ery KR$_6$	A1	50.4 ± 6.8	18.4 ± 0.2	17.3 ± 4.0	4.4 ± 2.0	30.0 ± 4.0	14.4 ± 3.9

[a] dimerization motif from donor KR domain substituted for that of LipPks1

LipPks1+TE was previously demonstrated to accept a set of diverse starter substrates
including propionyl-CoA and butyryl-CoA [80] in addition to the predicted starter substrate
isobutyryl-CoA [4]. The substrate promiscuity of our chosen PKS model system provided us
with a unique opportunity to probe the effect of substrate structure on the activity of KR
domain exchange variants. To the best of our knowledge this is the first report of the effect
of substrate structure on ketoreductase activity within the context of a PKS module.

Comparison of titers, shown in Table 3.1 resulting from incubation with different starter
substrates revealed that KR activity is substrate dependent. While Amph KR$_2$ was the
most active variant when **1a** was used as the starter substrate, when the starter substrate
was changed to **1b** or **1c**, the most active variants utilized Amph KR$_1$ or Spn KR$_3$ and
Amph KR$_2$ or Spn KR$_3$ respectively. Though it was not possible to identify clear, broadly-
applicable trends, we were able to draw loose correlations between native substrate structure,
shown in Table 3.2, and the outcome of reduction. For example, Amph KR$_2$ which natively
has a very small substrate, was most active when the smallest starter unit, **1b** was used,
while Spn KR$_3$, which has a longer, less substituted native substrate, favored the starter
unit **1c**. It was also notable that condensation ability of each variant relative to wild-type
is also substrate dependent. This suggests that KR substitution may alter the structure of
the PKS module such that condensation is more favorable with specific substrates.

Table 3.2: Properties of donor ketoreductase domains.

KR	Type	DE	Length (a.a.)	Product in native PKS
Lip KR$_1$	A2	+	499	
Amp KR$_1$	A2	+	483	
Amp KR$_2$	A1	+	488	
Ave KR$_1$	B1	+	488	
Bor KR$_1$	B	-	496	
Con KR$_2$	B1	+	481	
Ery KR$_1$	B2	-	474	
Ery KR$_6$	A1	-	437	
Pik KR$_1$	B2	-	481	
Spn KR$_3$	A1	+	480	

We also sought to consider the impact of KR domain exchanges on stereochemical fidelity. GC-MS was used to evaluate the stereochemical outcome of reduction catalyzed by the A-type LipPks1+TE variants when **1a** was used as the starter substrate. Stereochemical assignments were made by comparison to authentic standards, and the results summarized in Table 3.3. Elution order was consistent with previous studies [57]. GC-MS analysis demonstrated that all of the LipPks1+TE variants with the exception of (Ery KR$_6$)LipPks1+TE produced enantiomerically pure product consistent with expectations, given the known KR type. The A1 type Ery KR$_6$ has a larger native substrate, and consequently a larger active site, which may make it more difficult for Ery KR$_6$ to distinguish between epimers, should the substrate undergo water catalyzed epimerization. Levels of the A2-type product were too low to be detected by LC-MS.

While we were unable to assign exact stereochemistry for products resulting from incubation with starter substrates **1b** and **1c**, in both cases different retention times were observed for *syn* (A2/B1) and *anti* (A1/B2) products. This was verified by comparing the retention

Table 3.3: Stereochemistry of KR-catalyzed reduction by LipPks1+TE variants. Observed products for each combination of enzyme and starter substrate are marked with a bullet (●).

Substrate		LC-MS						GC-MS	
		1a		1b		1c			
Stereochemistry		A1/B2	A2/B1	A1/B2	A2/B1	A1/B2	A2/B1	A1	A2
Lip KR$_1$	A2		●		●		●		●
Amph KR$_2$	A1	●		●		●		●	
Amph KR$_2{}^a$	A1	●		●		●		●	
Spn KR$_3$	A1	●		●		●		●	
Amph KR$_1$	A2		●		●		●		●
Ery KR$_6$	A1	●		●		●		●	●

time of the wild-type Lip KR$_1$ product, which was expected to be *anti* (A1) [31, 33] to that of the authentic synthetic standard, which was determined by NMR to contain a mixture of diasteromers (26:1 *syn*:*anti*). The *syn*/*anti* stereochemistry of the products corresponding to the **1b** and **1c** starter substrates are shown in Table 3.3.

It is notable that only peaks corresponding to the expected *syn*/*anti* outcome were observed for all products and LipPks1+TE variants, suggesting that stereochemistry is relatively robust to alterations in substrate structure, contrary to studies that investigated the activity of isolated KR domains on diketide-SNACs [62] and β-keto esters [26]. However, when an additional starter substrate, isovaleryl-CoA (**1d**), was used, as shown in Supplementary Table B.5, we observed that incubation with Amph KR$_1$ results in the formation of two peaks, with relative peak areas that varied between experiment repetitions. This suggests that in select cases, substrate structure may impact stereochemical outcome even in the context of a full PKS module.

3.3 Conclusion

By retaining a dimerization motif, we were able to reliably use KR domain exchanges to alter α-methyl stereochemistry in a PKS model system, in some cases with no loss of activity. Furthermore, the broad substrate specificity of LipPks1+TE enabled us to probe the substrate dependence of KR activity in the context of a full PKS module. We found that the choice of KR domain impacts both condensation and reductive ability, and may also affect stereochemical fidelity. These findings provide the first insight into the importance of considering both dimerization motif retention and substrate structure in future KR domain exchange experiments.

Includes material from published work:

Eng, C. H., Yuzawa, S., Wang, G., Baidoo, E. E. K., Katz, L., Keasling, J. D. Alteration
of Polyketide Stereochemistry from *anti* to *syn* by a Ketoreductase Domain Exchange in a
Type I Modular Polyketide Synthase Subunit.

Chapter 4

Computer-Aided Polyketide Synthase Design

4.1 Motivation for Developing ClusterCAD

ClusterCAD is a web-based toolkit designed to leverage the colinear structure and deterministic logic of type I modular PKSs for synthetic biology applications. The unique organization of these megasynthases, combined with the diversity of their catalytic domain building blocks, has fueled an interest in harnessing the biosynthetic potential of PKSs for the microbial production of both novel natural product analogs and industrially relevant small molecules. However, a limited theoretical understanding of the determinants of PKS fold and function poses a substantial barrier to the design of active variants, and identifying strategies to reliably construct functional PKS chimeras remains an active area of research. In this work, we formalize a paradigm for the design of PKS chimeras and introduce ClusterCAD as a computational platform to streamline and simplify the process of designing experiments to test strategies for engineering PKS variants. ClusterCAD provides chemical structures, including stereochemistry for the intermediates generated by each PKS module, as well as sequence- and structure-based search tools that allow users to identify modules based either on amino acid sequence or on the chemical structure of the cognate polyketide intermediate. ClusterCAD can be accessed at `https://clustercad.jbei.org` and at `http://clustercad.igb.uci.edu`.

Previous successes engineering chimeric PKSs harboring a heterologous AT domain [15, 53, 30, 81], a heterologous KR domain [36, 43, 18], or a heterologous reductive cassette [37, 52, 53, 23, 43, 29] suggest that it is possible to identify combinations of junctions and catalytic domains that facilitate the construction of active PKS chimeras. While codifying the intuition that has previously enabled successful domain exchanges is impossible, one commonly used heuristic is to seek to design a chimeric PKS that is as similar to a naturally occurring PKS as possible. Using this guiding principle, we propose the following paradigm for designing a chimeric PKS capable of producing a particular small molecule compound of

interest:

1. Identify a truncated PKS parent to use as a starting point for engineering by searching for a module known to produce a polyketide intermediate with maximal structural similarity to the compound of interest.

2. Decompose the difference between the natural product of the truncated PKS starting point and the compound of interest into a series of catalytic domain exchanges.

3. Select donor modules for the required catalytic domain exchanges on the basis of sequence similarity to the truncated PKS parent.

ClusterCAD was developed to enable the design of chimeric PKSs by following this paradigm. It's structural search tool facilitates the identification of a truncated PKS parent to use as an engineering starting point given a small molecule target, and it's sequence search tool enables the selection of donor modules for catalytic domain exchanges based on sequence similarity to the truncated PKS parent. However, it is import to emphasize that ClusterCAD is intended to augment, rather than supersede, the biochemical intuition of PKS synthetic biologists.

Sequence similarity between PKS modules, and structural similarity between their cognate polyketide intermediates, provide only a rudimentary method of predicting compatibility between PKS polypeptide sequences. It is thus also important to consider additional factors when designing a chimeric PKS. For example, modules that originate from well-characterized clusters, or that have been previously determined to be well-expressed in the host organism of choice, are particularly attractive choices for engineering.

4.2 Design and Implementation of ClusterCAD

Data curation

The gene clusters in ClusterCAD are sourced from the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database [54]. To construct each ClusterCAD entry, gene clusters in version 1.3 (September 3, 2016) of the MIBiG database labeled as type I modular PKSs were first identified, and subsequently annotated using antiSMASH (the antibiotics and Secondary Metabolite Analysis SHell) [70], and the resulting output parsed using a list of explicitly recognized catalytic domains. Cluster analysis was terminated if a subunit containing an unrecognized catalytic domain was encountered, potentially resulting in premature truncation of that subunit. Domain annotations, which include predictions for AT substrate specificity and KR stereospecificity, informed *in silico* simulation of biocatalysis using reaction SMARTS operators [35], which were implemented using the RDKit cheminformatics software library [48], yielding a predicted intermediate chemical structure for each PKS module. Known final chemical structures were taken from the MIBiG database if

available, or were otherwise generated from the text description of the final structure in the MIBiG database using the ChemAxon JChem Base naming tool (version 17.2.27.0, 2017) [69]. Chemical structures for well-characterized polyketide products that could not be obtained by either of these methods were manually incorporated. Comparisons of the predicted and known final product chemical structures, close inspection of established amino acid sequence motifs [39, 15], and an extensive literature search were used to manually curate each ClusterCAD entry by correcting subunit order, removing invalid subunits, accounting for iterating modules, and correcting misannotated catalytic domain activity. The SCRATCH software suite [13, 50] was used to generate secondary structure and relative solvent accessibility predictions for the PKS subunits in ClusterCAD, and the resulting predictions incorporated into the database.

Organization

The primary objective of ClusterCAD is to provide computational assistance to synthetic biologists seeking to engineer PKSs. ClusterCAD provides an intuitive interface for users to browse through the clusters in the database, as well as search tools developed to assist in the design of chimeric PKSs. An example entry of a PKS cluster as displayed in the ClusterCAD web interface is shown in Figure 4.1.

1. *Structure search tool*: The structure search tool enables the identification of a truncated PKS parent to use as a starting point for PKS engineering, and takes as input either a small molecule chemical structure in the form of a SMILES string, or a structure that is drawn in an interactive graphical user interface. This tool searches a database of the predicted polyketide intermediates produced by the PKS modules in ClusterCAD. Matches to the query structure are ranked using atom pair (AP) Tanimoto similarity scores [11], and returned matches are displayed with the maximum common substructure [58] between the query and target structures highlighted.

2. *Sequence search tool*: The sequence search tool enables researchers to select candidate catalytic domains for domain exchange experiments. Support for flexible queries enables researchers to test hypotheses regarding which domain-domain interactions may be important in facilitating successful domain exchanges. The sequence search tool takes as input an amino acid sequence, and performs a protein-protein BLAST+ search against a database containing all of the subunits in ClusterCAD [8]. The results page enables deeper interrogation of the returned results by highlighting catalytic domains on the basis of annotations such as AT substrate specificity, KR stereochemical outcome, and reductive domain activity.

Figure 4.1: **ClusterCAD entry for the borrelidin PKS.** The ClusterCAD page for each cluster contains links to the corresponding entries in the MIBiG and NCBI nucleotide databases. Catalytic domain annotations can be optionally displayed based on domain type. Additional interactive web page elements provide access to further information.

4.3 Implications of ClusterCAD for Polyketide Synthase Design

ClusterCAD provides the first PKS database that specifically targets engineering applications. Existing PKS databases have focused on identifying new metabolites through genome mining or on predicting metabolite structure from nucleotide or amino acid sequence [89]. However, many of the most promising applications of PKSs are premised on harnessing their ability to generate customized molecular architectures.

From an engineering perspective, ClusterCAD offers substantial improvements over existing databases. By taking advantage of resources such as the MIBiG database and the antiSMASH command-line tool for cluster identification and annotation, we were able to focus on designing tools tailored to synthetic biologist's needs without attempting to optimize our software for genome mining or metabolite structure prediction. Furthermore, given the strong commitment from the synthetic biology and natural products communities to update MIBiG and antiSMASH with continuous improvements, populating ClusterCAD using

an end-to-end pipeline that utilizes these tools means that ClusterCAD can easily be kept up-to-date as novel biosynthetic clusters are discovered and as our understanding of PKSs develops.

While building on MIBiG and antiSMASH aids in automating the maintenance of ClusterCAD, it also means that inaccuracies in MIBiG or in the antiSMASH software may be propagated into ClusterCAD. In order to address this issue, we also employ extensive manual curation of each ClusterCAD entry to provide predicted polyketide intermediates that are as accurate as possible. ClusterCAD has a persistent corrections record that overwrites the output generated by antiSMASH. This enables ClusterCAD to incorporate new MIBiG entries and systematic improvements to antiSMASH, while also guaranteeing that clusters that have previously been manually corrected will remain correct. Because these records are persistent and override the automatically generated output, the clusters will only need to be further corrected if new empirical findings are generated with regard to the cluster. ClusterCAD thus combines computational annotations of AT substrate specificity and KR stereospecificity with domain expert application of heuristics obtained from the literature [15, 39, 85, 40] to provide well-curated predictions of the polyketide intermediates naturally provided by each of the PKS modules in ClusterCAD. However, the reliance on manual curation means that some errors may be missed; we include links on each cluster page that can be used to submit corrections to the ClusterCAD maintainers for review.

ClusterCAD is further limited in that it does not support inclusion of chemistry effected by "on-line" tailoring enzymes that act on ACP-tethered polyketide intermediates, but that are independent of the PKS [64]. However, we believe that showing the direct product of the PKS provides the information that is most relevant for engineering, since the on-line tailoring enzyme is not expected to be naturally present in a heterologous host.

A substantial benefit of ClusterCAD is that it allows researchers to compare PKS modules based both on amino acid sequence similarity and on cognate polyketide intermediate similarity. The substructure search in ClusterCAD is of particular utility when determining which modules would be good candidates for PKS engineering. Previous experimental evidence supports the assertion that the structure of the intermediate can have a significant impact on the catalytic function of individual domains. KS domains, for example, have been reported to have "gatekeeping" functionality, in which they are unable to condense intermediates that are chemically dissimilar; structural differences that have been empirically determined to abolish activity range from variations in steric bulkiness to having an opposite stereochemical configuration [76, 56]. Other domains, such as the KR domain [18] and the DH domain [22] have also been demonstrated to have poor catalytic activity on substrates that are much smaller or bulkier than ones they encounter in nature.

The substructure search in ClusterCAD is one of the first that matches against polyketide intermediates associated with non-terminal PKS modules, making it more robust to false positives than similar search tools which only support substructure queries in the context of PKS final products [14]. Matching directly against the polyketide intermediates produced by each module also avoids undesirable matches to structural motifs produced by tailoring enzymes, which are of limited utility for PKS engineering. To the best of our knowledge,

SBSPKSv2 [44] is the only other PKS database that currently supports structure-based queries that match against polyketide intermediates. However, ClusterCAD boasts a larger number of type I modular PKS intermediates and utilizes AP descriptors, which exhibit greater sensitivity to uncommon chemical motifs than the substructure fingerprint approach used by SBSPKSv2.

Anecdotal evidence supports the assertion that seeking catalytic domains with greater sequence similarity improves domain exchange success rates [18]. While this heuristic is challenging to verify experimentally given the low throughput of domain exchange experiments, it is our hope that ClusterCAD will assist PKS researchers in elucidating the importance of sequence similarity in domain exchange experiments by streamlining the design of experiments to test their hypotheses. For instance, by facilitating chimeric PKS design, ClusterCAD will enable researchers to probe potential domain-domain interactions which may be important for engineering.

The sequence search tool in ClusterCAD highlights the domains in each returned module that overlap with the query, and also allows the user to highlight results based on catalytic domain activity. Identifying junctions in PKS chimeras that reliably leave functionally important structural elements intact [43, 18] remains an open area of research. ClusterCAD provides secondary structure and relative solvent accessibility predictions that will aid in the identification and comparison of homologous structural elements across PKSs, assisting researchers in junction design. As far as we are aware, ClusterCAD is the only PKS database that provides precomputed values for the predicted secondary structure and relative solvent accessibility of each subunit. By providing support to compare PKS modules on the basis of metrics that may contribute to compatibility between PKS polypeptide sequences, ClusterCAD aims to facilitate experimental research to identify robust strategies for PKS engineering.

Using ClusterCAD to design a chimeric PKS that produces adipic acid

While engineered PKSs have long been an attractive strategy to produce pharmaceutically relevant natural products and their analogs, more recent work has also sought to harness the chemical diversity accessible by PKS to provide an environmentally friendly alternative to chemical synthesis for commodity chemicals. Previous work by Hagen *et al.* has demonstrated that it is possible to engineer a novel PKS capable of producing adipic acid [29], a widely used commodity chemical. As a proof-of-concept, we demonstrate how ClusterCAD could be applied to engineer this novel PKS.

1. *Identify a truncated PKS starting point that naturally produces a polyketide intermediate that is structurally similar to adipic acid.* Hagen *et al.* selected the first module from the borrelidin PKS as a starting point based on previous biochemical studies establishing that BorMod1 tolerates succinyl-CoA as a starter, possibly because of structural similarity to the native priming unit, trans-1,2-cyclopentanedicarboxylic acid [28].

Adipic acid was used to query the ClusterCAD chemical structure search tool, which returned BorMod1 as the highest-scoring result, with a naturally produced intermediate that had an AP Tanimoto similarity score of 0.32 when compared against adipic acid.

2. *Identify a donor reductive loop with active ER, DH, and KR from a module which naturally incorporates a malonyl-CoA extender unit.* In order to produce adipic acid using BorMod1, incorporation of an extender unit lacking an α-substituent, and reduction by a full complement of active reducing domains was required. Hagen *et al.* thus chose four reductive loop donors that naturally accepted malonyl-CoA, harbored a DH, ER, and KR in the reductive cassette, and that came from a "standalone" module, in which the corresponding open reading frame or subunit contained only a single module. Donor loops from standalone modules were favored based on the intuition that domain exchanges using reductive cassettes derived from a standalone module would retain more architectural similarity to BorMod1, itself a standalone module, and thereby increase the likelihood of a functional chimera.

Using the ClusterCAD sequence search tool, we sought to identify donor modules that not only met these criteria, but also exhibited maximal sequence similarity to BorMod1. We queried the ClusterCAD sequence search tool with the amino acid sequence of BorA2, the subunit containing BorMod1, and ranked the potential donor modules by bit score. We then filtered for standalone modules containing a malonyl-CoA specific AT and a complete reductive cassette. IdmO, from the indanomycin PKS, and SpnB, from the spinosyn PKS, were the only two returned results that matched all three criteria. Chimeras utilizing these two donor modules were experimentally tested by Hagen *et al.*, along with chimeras harboring reductive cassettes from the donor modules AurB, from the aureothin PKS, and NanA2, from the nanchangmycin PKS. AurB, which is not contained in ClusterCAD, and SpnB were empirically determined to yield the most active chimeras. NanA2, which was not returned by ClusterCAD, but is contained in the database, yielded the least active chimera.

3. *Identify a DH domain that exhibits higher levels of activity on 3-hydroxyadipic acid.* The chimeras constructed by Hagen *et al.* yielded a relatively large proportion of the incompletely reduced product 3-hydroxyadipic acid, potentially due to the DH from the donor reductive cassettes having a poor substrate compatibility with the terminal carboxyl moiety, an unusual feature in PKS biosynthetic intermediates. To overcome this substrate incompatibility, Hagen *et al.* substituted a heterologous DH into the engineered BorMod1 variant. The DH from BorMod2 was selected based on the observation that it is known to demonstrate catalytic activity on a substrate containing a terminal carboxy moiety. The intuition employed by Hagen *et al.* was consistent with the sequence-similarity based output of the ClusterCAD search tool, which returned BorMod2 as the module containing an active DH that had the highest bit score with BorMod1.

ClusterCAD in Context

We sought not only to provide first PKS database that specifically targets synthetic biologists interested in PKS engineering, but also more generally to provide tools for computer-aided PKS design. We have therefore released ClusterCAD under a BSD style license at `https://github.com/JBEI/clusterCAD`. This open source release provides Python object representations of catalytic domains, modules, and clusters, enabling researchers to construct and simulate the activity of chimeric PKSs *in silico*. Furthermore, we sought to establish a method to describe PKSs in a human and computer readable manner using the JSON data-interchange format, as shown in Figure 4.2. The availability of such a PKS representation will aid in the development of additional software for computational PKS design by allowing researchers to easily describe novel PKSs in a computer-readable fashion.

4.4   Conclusion

To the best of our knowledge, ClusterCAD is the first web-based toolkit developed specifically to aid in designing chimeric PKSs. By providing chemical structures with stereochemistry for the polyketide intermediates produced by each module, as well as relative solvent accessibility and secondary structure predictions for each subunit, we provide rich information about each cluster to inform chimeric PKS design. Furthermore, our intuitive user interface, combined with the ability to query for modules based on either sequence- or structure- similarity provide powerful tools for PKS engineering that bring us closer to the promise of harnessing PKSs for combinatorial biosynthesis.

```json
{
  "description": "Borrelidin",
  "genbankAccession": "AJ580915.1",
  "mibigAccession": "BGC0000031.1",
  "architecture": {
    "borA1": {
      "0": {
        "iterations": 1,
        "domains": {
          "AT": {
            "substrate": "trans-1,2-CPDA"
          }
        }
      }
    },
    .
    .
    .
    "borA6": {
      "0": {
        "iterations": 1,
        "domains": {
          "AT": {
            "substrate": "mal"
          },
          "KR": {
            "type": "B1",
            "active": true
          },
          "TE": {
            "ring": 1,
            "cyclic": true
          }
        }
      }
    }
  }
}
```

Figure 4.2: **Partial JSON representation of the borrelidin PKS.** This representation captures the hierarchical organization of PKSs, as well as the important catalytic properties of their components.

Appendix A

Chapter 2 Supplementary Information

A.1   Experimental Procedures

Authentic Synthetic Standards

All chemicals were purchased from Sigma-Aldrich unless otherwise described. All NMR spectra were obtained on either a 400 MHz or a 600 MHz Bruker spectrometer. Chemical shifts (in ppm) were referenced to $CDCl_3$ or D_2O at room temperature (δ ^1H = 7.26 and δ ^{13}C = 77.16 for $CDCl_3$, δ ^1H = 4.79 for D_2O). A synthetic procedure analogous to that described for the synthesis of **1a** was also used to obtain each of the final products **1b-f**.

3-Propionylbenzo[d]oxazol-2(3H)-one (S1)

Following a reported procedure [6], a stirred solution of 2-benzoxazolonine (1 equiv, 102 mmol) in $(CH_3)_2CO$ (90 mL) was treated with K_2CO_3 (1.15 equiv, 117 mmol) before addition of propionic anhydride (1 equiv, 102 mmol). The resulting mixture was stirred at room temperature (r.t.) for 3 h. The precipitated product was collected under filtration and dried in vacuo. Recrystallization in Et_2O afforded a light tan solid, S1 (81.8 mmol, 80% yield). ^1H NMR ($CDCl_3$, 400 MHz) δ 8.09 (m, 1H), 7.24 (m, 3H), 3.15 (q, 2H), 1.29 (t, 3H).
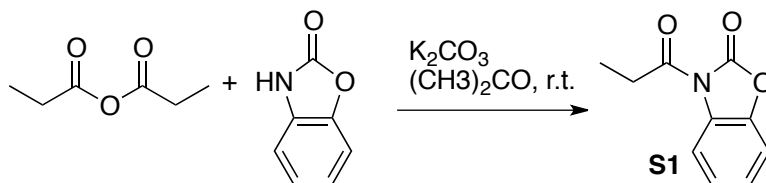
3-(3-Hydroxy-2,4-dimethylpentanoyl)benzo[d]oxazol-2(3H)-one (S2)

Following a modified reported procedure [6], to a solution of **S1** (1 equiv, 5.23 mmol) in anhydrous CH_2Cl_2 (8 mL) at 0°C under N_2 was added $TiCl_4$ (1.1 equiv, 5.75 mmol) over

5 min, and the resulting bright yellow slurry stirred vigorously for 15 min. Triethylamine (1.1 equiv, 5.75 mmol) was then added to the solution over 5 min, and the consequent blood red solution stirred for an additional 50 min. Isobutyraldehyde (2 equiv, 10.46 mmol) was added slowly over 1 h, and the mixture was stirred for a further 1.5 h before quenching with one volume 1M HCl. The aqueous phase was extracted with EtOAc. The combined organic layers were then washed with 1 M HCl followed by saturated NaHCO$_3$, and finally concentrated by rotary evaporation. Column chromatography followed by recrystallization in Et$_2$O/hexanes (1:1) afforded a white solid, **S2** (2.54 mmol, 49% yield), which was used directly in the next step of the reaction procedure.

3-Hydroxy-2,4-dimethylpentanoic acid (1a)

Following a modified previously reported procedure [21], to a solution of **S2** (1 equiv, 1.07 mmol) in THF (3 mL) at 0°C was added a solution of LiOH (2 equiv, 2.14 mmol) and H$_2$O$_2$ (7 equiv, 7.49 mmol) in water (1 mL). The reaction was stirred at 0°C for 4 h, and then quenched with one volume Na$_2$SO$_4$. The organic layer was removed via rotary evaporation. The aqueous layer was then washed with CH$_2$Cl$_2$, acidified with one volume 1 M HCl, and extracted with EtOAc. The combined organic layers were washed with 1 M HCl and then concentrated using a rotary evaporator. Column chromatography afforded a mixture of diastereomers as a tan/white solid, **1a** (0.919 mmol, 86% yield). ^{1}H NMR (400 MHz, CDCl$_3$) δ 3.64 (dd, 1H), 2.72 (qd, 1H), 1.73 (m, 1H), [1.24, 1.21; 1:25] (d, 3H), [1.03, 0.99; 23:1] (d, 3H), [0.94, 0.89; 1:30] (d, 3H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 182.03, 76.84, 41.92, 30.80, 19.20, 18.83, 9.93. ESI-MS (**1a** – H) C$_7$H$_{13}$O$_3^-$ Mw = 145.18, Observed = 145.1.
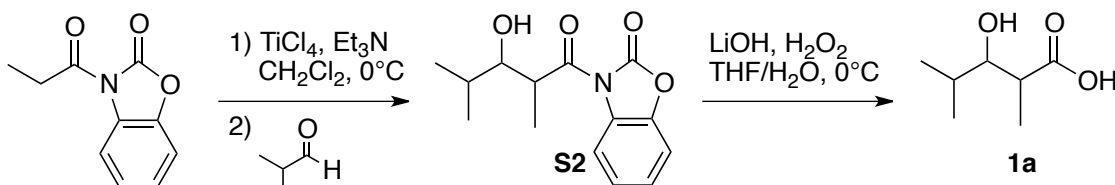
3-Hydroxy-2-methylpentanoic acid (1b)

Yellow oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.93 (s br, 2H), 3.88 (ddd, 1H), 2.60 (qd, 1H), 1.50 (m, 2H), 1.18 (d, 3H), 0.96 (t, 3H). ^{13}C NMR (150 MHz, CDCl$_3$) δ 181.13, 73.53, 43.93, 26.75, 10.46, 10.39. ESI-MS (**1b** H) C$_6$H$_{11}$O$_3^-$ Mw = 131.16, Observed = 131.2.
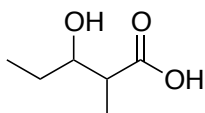
3-Hydroxy-2-methylhexanoic acid (1c)

Brown oil; ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.01 (ddd, 1H), 2.64 (qd, 1H), 1.55 (m, 2H), 1.47 (m, 2H), 1.25 (d, 3H), 0.99 (t, 3H). ^{13}C NMR (150 MHz, CDCl$_3$) δ 180.93, 71.67, 44.30, 35.96, 19.32, 14.06, 10.56. ESI-MS (**1c** − H) C$_7$H$_{13}$O$_3^-$ Mw = 145.18, Observed = 145.1.
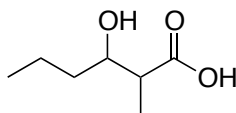
3-Hydroxy-2,4-dimethylhexanoic acid (1d)

Brown oil, mixture of diastereomers; ^{1}H NMR (400 MHz, CDCl$_3$) δ 3.73 (dd, 1H), 2.71 (dq, 1H), 1.76 (m, 1H), 1.48 (m, 1H), [1.26 (d), 1.22 (d), 1.18 (d) ; 1:6:24 (3H)], 1.18 (m, 1H), [0.96 (d), 0.84 (d); 1:4 (3H)], 0.90 (t, 3H). ^{13}C NMR (100 MHz, CDCl$_3$) δ [182.09, 191.73; 4:1], 75.32, [42.30, 41.66; 1:4], [37.12, 37.00; 1:4], [25.96, 25.02; 1:4], [15.04, 14.20; 4:1], [11.22, 11.00; 1:4], 9.35. ESI-MS (**1d** − H) C$_8$H$_{15}$O$_3^-$ Mw = 159.21, Observed = 159.2.
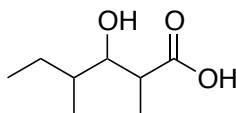
3-Hydroxy-2,5-dimethylhexanoic acid (1e)

Tan/white solid, mixture of diastereomers; ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.05 (ddd, 1H), 2.59 (qd, 1H), 1.78 (m, 1H), 1.48 (m, 1H), 1.21 (d, 3H), 1.19 (m, 1H), [0.95 (d), 0.93 (d); 1:1, 6H]. ^{13}C NMR (150 MHz, CDCl$_3$) δ 179.43, 70.03, 44.46, 42.76, 24.83, 23.57, 22.00, 10.72. ESI-MS (**1e** − H) C$_8$H$_{15}$O$_3^-$ Mw = 159.21, Observed = 159.2.
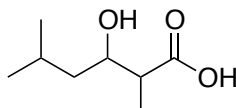
3-Hydroxy-2,4,4-trimethylpentanoic acid (1f)

White solid, mixture of diastereomers; 1 H NMR (400 MHz, CDCl$_3$) δ 3.68 (d, 1H), 2.75 (qd, 1H), [1.38 (d), 1.25 (d); 5:3 (3H)], [0.96 (s), 0.94 (s); 3:2 (9H)]. ^{13}C NMR (150 MHz, CDCl$_3$) δ 180.35, [82.76, 78.27; 18:5], 39.28, [36.14; 35.86; 18:5], [26.76, 26.26; 18:5], [18.52, 12.41; 18:5]. ESI-MS (**1f** − H) C$_8$H$_{15}$O$_3^-$ Mw = 159.21, Observed = 159.2.
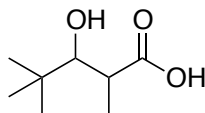
2-Methylbutyryl-CoA

Following a previously reported procedure [77], 2-methylbutyryc acid (1 equiv, 26 μmol), CoA (lithium salt, 1.1 equiv, 29 μmol), and PyBOP (1.5 equiv, 39 μmol) were dissolved in a solution of THF (0.39 mL) and 4% K_2CO_3 (0.39 mL) and stirred under N_2 for 40 min. The reaction mixture was diluted by addition of 2 mL water and injected onto ZORBAX SB-C18 column (50 mm length, 9.4 mm internal diameter, 5 μM particle size). The mobile phase was composed of 50 mM NaH_2PO_4 (pH = 4.2) in water (solvent A) and methanol (solvent B). The acyl-CoA product was separated using the following gradient: 2.5% to 31% B for 12 min, 31% to 95% B for 1 min, held at 95% B for 2 min, 95% to 2.5% B for 2 min, held at 2.5% B for 2 min. A flow rate of 5 mL/min was used throughout. After removal of methanol, the eluted product was lyophilized and desalted by reinjection onto the same column. For desalting, the mobile phase was composed of water (solvent A) and methanol (solvent B). The product was separated using the following gradient: 2.5% to 31% B for 12 min, 31% to 95% B for 1 min, held at 95% B for 2 min, 95% to 2.5% B for 2 min, held at 2.5% B for 2 min. Lyophilization of the eluted product facilitated isolation of a white solid, 2a (6.5 μmol, 25% yield). ^{1}H NMR (600 MHz, D_2O) δ 8.66 (s, 1H), 8.39 (s, 1H), 6.27 (d, 1H), 4.94 (m, 2H), 4.68 (t, 1H), 4.2 (t, 2H), 4.11 (s, 1H), 3.92 (dd, 1H), 3.65 (dd, 1H), 3.53 (t, 2H), 3.41 (t, 2H), 3.07 (t, 2H), 2.71 (h, 1H), 2.51 (t, 2H), 1.68 (m, 1H), 1.53 (m, 1H), 1.18 (d, 3H), 0.98 (s, 3H), 0.92 (t, 3H), 0.86 (s, 3H). ESI-MS (2a + H) $C_{26}H_{45}N_7O_{17}P_3S^+$ Mw = 852.17, Observed = 852.2.

Pivaloyl-CoA

Pivalic acid (1 equiv, 26 μmol) was used instead of 2-methylbutyric acid in the experimental procedure described above to obtain a white solid, 2b (10.0 mol, 38% yield). ^{1}H NMR (600MHz, D_2O) δ 8.75 (s, 1H), 8.50 (s, 1H), 6.30 (d, 1H), 4.96 (m, 2H), 4.69 (t, 1H), 4.34 (t, 1H), 4,13 (s, 1H), 3.95 (dd, 1H), 3.68 (dd, 1H), 3.55 (t, 2H), 3.42 (t, 2H), 3.06 (t, 2H), 2.52 (t, 2H), 1.28 (s, 9H), 1.02 (s, 3H), 0.89 (s, 3H). ESI-MS (2b + H) $C_{26}H_{45}N_7O_{17}P_3S^+$ Mw = 852.17, Observed = 852.2.

Cloning

Gene Synthesis

The gene encoding an N-terminally hexahistidine tagged *S. aureofaciens* Tü117 LipPks1+TE with codons optimized for expression in *E. coli* was ordered from DNA2.0.

6xHis-LipPks1+TE:

```
ATGGGCAGCAGCCACCACCACCACCACCACAGCAGCGGCCTGGTCCCGCGTGGTTCCCACATGGCAGGTCCACCGC
CGTTCCCGCGTCGTCGTGGCCCGAGCGGCCGTCGTCGTTGCGGCGGTCGTGCTACCCCGGGTAGCGTCCGTGACCG
TACGGGTCGTCGTCCGGCAGCCGTTCCGAGCCGTGCAGTTTGCGCAGCGGATCTGTGCGAAGAAAATGACGACGGC
TCGAAGAATGTTTCTGAACATCGCGGTAGCGCAGGTGGTTCCGTGCTGTTTCCACGTACCGGCACCGTCCTGCCGT
GGGTACTGACCGGTCCTGGCGCAGCGGCGGTTCGCGCACGCTCCGAAGCACTGCGCACGCACCTGCGTGCGAGCAC
CGAGTGGTCCCCTGCGGGCGTCGGTCAGGCGCTGCTGGCCGGTACGGGTGCGGGTGCCGATACCCACCGTGCCGTT
GTTCTGGCAGGCGACCGTGCCCAGACCCTGAACGCATTGGCAGCGCTGAGCGCAGGCGCAGACCACCGGCAGTTT
TCACCAGCACTCGTGCGGATGCAAGCCCGGCTGGCCCGGTGTTTGTGTTCCCGGGTCAAGGCTCGCAGTGGACCGG
TATGGCTCGTGAACTGCTGGACTCCGCACCGGTTTTCGCGCGTAAGCTGCACGACTGTGCAGACGCGTTTGCCCCG
TACCTGGGCCACAGCCTGCTGGATAGCGTCACCGGTGCAGCAGGTGGTCCAGAGCCTGTTGGCGCGGACGTCGTCC
AACCGGCGCTGTTCGCCGTTATGGTTGCGCTGACTGATCTGTGGAACGCGGCTGGCGTTGCACCGGGTGCACTGCT
GGGTCACTCCCTGGGTGAACTGGCAGCCGCGCATGTCGCGGGTGTTCTGTCCCTGGACGATTCTGCTCGCGTCGTG
GCGCGTTGGAGCCAAGCGCAGGCTACGTTGGCGGGTCGTGGTGACATGGTCAGCGTTCTGTTGCCTGCGGATGAAT
TGGCGGACCTGCTGGACCGCCGTTGGCCGGGTCGCTTGGTTGTGGCGGTTGAAAACGGTCCAGGTAGCGCGGTCGC
GAGCGGTGACCTGGACGCTGCGGCGGAACTGGTCGCACACCTGACCGCCGAAGGTATCCACGCGCGTCGCGTTGAC
GTGGGCCTGGCGGCTCACAGCCCGCACATTGACGCGATCCTGCCACGTATTCGCGCGGACATCGCGCCGATTCGTG
CGCATACGCCGAGCATCCCGGTTTATTCGGCGCTGCATGGTGGTGCACTGGATGGCACGCCAATGGACGCGGCGTA
CTGGTGTCGTAATCTGCGCTCCACTGTACGTTTCGCGGACGCGACCCGTGCAGCCCTGGAGGCAGGCCATACCACG
TTTGTGGAGGTAAGCCCACATCCGGTCCTGACTACGGCGATGGAGGTGAGCGCAACCCGTGCCGCGCACGCAGCAA
CTGTCCTGGGTACGCTGCGCCGTGGTGAGGGTGGTCCGAGCCGCTTCCTGGCGAGCCTGGCCGAACTGCATGTCAG
CGGTGGTGATGCCGATCTGCGTACGGTTCTGCCGGCTAGCCAGCGGCTGGCTTGCCGGAAACCGTTCTGACGGCG
GGTCCGCGTGGCGAGAGCGCGGATGGCGACTCTCGTCATGAGGTTCTGTGCGCACGCCTGGCACCGCTGGACCCAG
CGGAGCGTCGTGCCCAGCTGCTGACTGTTGTTCGTGAAAGCGCAGCTGCCGCGCTGGACGGCGACGATCAAGGTAG
CATTGACGGTCGTCGCACGTTCCGTGACCTGGGTATCACGTCGCTGGCAGCGGTGGGCATCCGTGATCGCCTGCAT
TCCGCAACCGGTCTGCGTCTGTCTCCGACCGTTGTGTTTGATCATCCGACCCCGGACGCACTGGCGGCACACTTGG
ACACCGAACTGTTCGGCACGGGCGCAGATGCCGAGCCGGCACCAGCTGCGGGTGGTCGTGCGGTGCCGCATGACGA
ACCAATTGCGATCGTGGGTATGGCGTGCCGTTACCCTGGCGGCGTTGGTGCACCGGCCGACCTGTGGCGTACCGTT
CTGGCCGGTGTCGACGCAGTTGGTCCGCTGCCGGCTGATCGTGGCTGGAATATTGCGGACGGTTACGATCCGGAGC
TGGCGGGTCCTGGTCGTTTTAGCCAGCGTGAGGGCGGCTTTCTGCACGACGCAGCTGAATTTGATGCGGAGTTCTT
TGGTATTAGCCCGCGTGAGGCATTGGCGATGGACCCGCAGCAACGTTTGGCTCTGGAAAGCGCCTGGGAAGCGATT
GAGGATGCGGGTCTGGACGCCCATAGCCTGCGTGGCAGCCGTACTGGCGTTTTTCTGGGCTTGATTACCCAGGATT
ATGGTCCTCGTGCGGGTGAGCCGACCACGCGTGCAGGTGCGGTGGAGGGTCACCTGTTCCTGGGTAGCACTGGCAG
CGTCGCAAGCGGTCGTCTGAGCTATACCTTGGGTCTGGAAGGTCCGTCTTTGACGATTGATACGGCATGTTCGAGC
AGCCTGGTGGCACTGCACGAAGCATGTCAAGCGCTGCGTACCGGTGATTGCGACATGGCTCTGACTGGTGGTGTGA
CGGTCATGCCGAGCACCGGCATGTTGGTCGAGTTCAGCCGTCAGCGTGGTCTGTCGCCTGACGGCCGTTGTAAAGC
CTTTTCTGCATCTGCCGACGGTTTTGGTCTGGCGGAAGGTGTCGGTATGCTGGTGGTTGAGCGTCTGAGCGATGCG
CGTCGTCTGGGCCATCGTGTGCTGGCGGTGGTGCGCGGTTCTGCGGTTAACCAAGATGGCGCGAGCAATGGCCTGT
CGGCGCCTAGCGGCCCAGCACAACAGCGCGTTATTCGCCAGGCGCTGGTCAACGCTGGCGTCCAAGCATCCCAAGT
GGACGTTGTCGAAGCACATGGCACCGGTACGAAACTGGGCGATCCGATTGAGGCTCAAGCCCTGCAAGCGACCTAT
GGCCAGGGCCGTCCGGCTGAGCGCCCGTTGTGGTTGGGTTCTCTGAAGTCCAATATCGGCCACGCGCAAGCGGCAG
CGGGTGTGGGCGGTGTTATCAAAATGGTCATGGCGTTGCGTGAAGGCGTCCTGCCACCGACCCTGCACGCAGACGA
GCCGAGCCCGCATATTGACTGGTCGGCGGGTCAGGTTCGTCTGCTGACCGAGGAACGCGAGTGGCCGGAGGCAGGT
```

```
CACCCTCGCCGTGCGGCAGTTTCGAGCTTCGGTGTTAGCGGTACCAACGCACATGTGATTCTGGAAGCTGCACCGG
GTACGGGTGGTGCGCCAGAAGTTTCGGACGGTGTCCTGGGTAGCGCGCCTGAAACGGTCCCGTGGGTGCTGAGCGC
TGCAAGCCCTGACGCATTGCGTGCACAAGCAGAGCGTCTGCGCGGTCATGTGGCGGAGCGTCCGGGTCTGGCTTCC
GCCGATGTCGCGTTTGCGCTGGCGACCCGTCGTACCGCGCTGGAATATCGCGCGGTGGCGGTTGGTGCGGAGCGCG
ACGAGCTGCTGGATACCTTGGACGCGCTGAGCGCCGGTCGTCCGGCACCGCGTGCTGTACCGGGTGACGCGGCTGC
GCATAGCCGTCGTCCGGTTTTCGTCTTTCCGGGTCAGGGTAGCCAGTGGGCAGGTATGGCGGTTGAACTGCTGGAC
AGCAGCCCGGTTTTTGCGGACAGCATGCACGCATGTTCCGAGGCCCTGAATGAATTTGTTGACTGGAACCTGCTGG
AAGTTCTGCGTAGCGGTGACGAAGAGCTGTCTAACCGTGTTGATGTCGTCCAACCGGTGCTGTGGGCAGTTATGGT
GAGCCTGGCAGCTCTGTGGCAAGCGTGTGGCGTCCGTCCTGCGGCGGTTGTGGGTCACAGCCAAGGTGAGATTGCA
GCTGCCGTTGTCGCAGGTGCACTGAGCCTGCGTGATGGTGCCCGCGTTGTTGCATTGCGTAGCGCAGTGATCGCGC
GTCTGCTGGCAGGTAAGGGTGCGATGGCGAGCGTGGCTCTGGCGTCTGACACCGTTCGTGAGCGCCTGACCCCGTG
GGAAGGTCGTCTGTCTCTGGCAGCGGTCAATGGTCCGAGCAGCAGCGTTGTTTGCGGCCATCTGGATGCACTGGAC
GAGTTCGTTAGCGCGTTGGAGCACGATGGCGTGCGTGTGCGTCGCATCGCGGTTGACTACGCAAGCCATAGCGTGT
TCGTGGAGCAGGCAGAAGAAGAGCTGCGTAATGTCCTGACCGAGGTGAGCCCTTTGCCGGGTCAAGTCCCTTTCTA
CAGCACCGTGACCGGTGCGGTTCTGGATACCACGACTCTGGACGCCGGCTACTGGTATCGTAACTTGCGTCAGACG
GTTCGTTTCGAAGAAACCGTGCGTGAGCTGACGCGCCGTGGCCACGACGCGTTCATCGAGGTGTCGGCTCATCCGG
TTCTGACCGTCGGCATTCAGGATACCTTGGAAGCCACCGGCACCCGCCATGCAGTCTGCGGTACGCTGCGTCGTGG
TGAGGGCGGTGCGCAGCGTTTGCTGACCAGCCTGGGTGAAGCGTGGGTTGCCGGCATTGCGGTGGACTGGAGCCGC
TTGACGCCGACGACGACCGCTGTCCAACTGCCGACCTACGCATTTCAGCATCAGCGTTACTGGCTGGATAGCACCA
CTGCAAACACTGGTGACCGTCCGGCAGCGGACCGTGACACCGCATTTTGGGAAGCTGTGCAGCACACCGACCTGGA
CGCCTTCGCTGCAGAATTGGACATTGCCCCGGATGCGCCGTTGGGCACCGTCTTGCCGGCTCTGGCTGACTGGCGT
CAACGCCTGCGTACGGCAGCGGCTGTTGACGCATGGCGTTACCGCACCGCCTTTAAACGTCTGCCAGATGCGCCAG
GTGCACCAGTCCTGACGGGCAGCTGGCTGGCCGTAGTTCCGGTGCGTCACCTGGATGATCCGAGCGTTACCACTAG
CCTGGATGCAGTTGCTAAAGCGGGTGCGGAAGTCGTTCAGTTGGCAATCGAAGATGCGGACGCGGACGTTGATCGT
CTGACTGAGCGCTTGCGTGGCCTGGTTGCCGGTCTGGGTGCCGCGCCGGCGGGCATTATGAGCTTCCTGGGTCTGG
ATGAAGAGCGTCATCGTGACCACCCGGCGATGCCGAGCGGTCTGGCCACCAGCTTGGCGCTGGTCCGCGCCTTGGG
TCGTGCGGGCATCGGTGCACCGCTGTGGATGGTTACGCGTGAGGCAGTGGCAGCGGGTCAAGACACGCACCCGCAT
GCGCCTCTGGGTAGCCTGATCTGGGGTCTGGGCCAAGTGACGGCTCTGGAGCACGCAGATCGCTGGGGTGGTCTGA
TCGATCTGCCCGGGTGTGTGTGATGCGCGCGTTGCCCGCATGCTGTGCGCGGGTCTGAGCGGCCGTGGTGCCGAAGA
TCAGCTGGCCCTGCGTCCGAGCGGCACTTTCGTCCGCCGTCTGGCGCATATCCCTGGCGAGCAACGTGCAGCACGT
CGTAGCTGGCAACCACGTGGTACGGTGATTGTCACCGGTGGTACGGGTGCGCTGGGTGCAGTCCTGGCCCGCTGGT
TGGCTACCGAGGACGCGGAGCACCTGGTGCTGACCGGCCGTCGTGGCGCGGACGCCCCTGGCGCGGAGCGTTTGCG
TGACGAGCTGGTCGCTACGGGCGCTCGTGTCACGCTGGCGGCGTGCGACGTGGCAGATCGCAAAGCCGTCGCCGCA
TTGCTGGACGAACTGGCTGCGGACGGCGAGACTGTTCGCGCAGTTCTGCACGCTGCGGGTGTCGCCGATCTGACGT
CGCTGGAGAATACCGGTCCAGAAGCGTTCGCGGCAGGCGTGGCCGCGAAGGTCGATGGTGCACTGCACCTGACCGA
ACTGTTGGATCACGATTCGCTGGATGCGTTTGTGTTGTTCAGCAGCATTGCGGGTGTTTGGGGTTCCGGCGACCAC
GGCGCGTATGCGGCTGCGAACGCATTTCTGAATGCGTTGGCAGAGTACAATCGTGCACGCGGTATCCCGACCACGA
GCATCGCATGGGGCGTTTGGAACGCGTTTGGCGTCGAGGGTGCAGGCGGTATCAGCGAGGCGGTTGATTTGGACCA
GCTGCATCGTCGCGGCCTGCCGCTGATTGAACCAGAGCTGGGTCTGACTGCACTGCGTCGCGCTCTAGACCGTGAC
GAAACGGTGCTGACGGTTGCTCCGGTTGCCTGGGAGCGCTTCTTTCCGCTGTTCTCCGCTGCACGTCCGCGTCCGT
TGTTTGAGGACTTGCCGCAAGTGCGTGCCCTGAGCGCACCTGTCCCGACGACGGCGGGTCCGGCCGTGGAACCAGG
TCGCCGTGGTAGCGGCCTGGGCGATTTGCCTCTGACGGATCGCGATAGCGCGCTGCTGGCCTTGGTCCGCGGTGAG
```

```
AGCGCATCCGTGCTGGGGTTACGAGCGTCCAGATCGCCTGGACCCGGACCGTGCGCTGCGTGATGTGGGTTTCGATA
GCCTGACGGCGATGGAACTGCGTAACCGTCTGGCTACCGCGACCGGCCTGACGGCTGCCTGCGGCCCTGGTGTTTGA
TCACCCGACCCCACTGGCGATCGCGGCGTATCTGAAAGCCGAGCTGACGAGCCAGCTGGACTCCGGTACGCCTGCA
CGTGAAGCTAGCTCTGCACTGCGTGACGGCTACCGCCAAGCGGGTGTGAGCGGCCGTGTTCGTAGCTACCTGGATC
TGTTGGCAGGTCTGTCGGATTTCCGCGAACATTTTGATGGTAGCGATGGTTTTAGCCTGGATCTGGTTGATATGGC
AGATGGTCCGGGTGAGGTGACCGTTATTTGCTGCGCGGGTACGGCTGCGATCTCTGGTCCGCATGAGTTCACCCGT
CTGGCAGGTGCCCTGCGCGGCATTGCACCTGTTCGTGCGGTGCCGCAGCCGGGTTACGAAGAAGGCGAACCGCTGC
CTTCTAGCATGGCGGCTGTTGCAGCTGTGCAGGCTGATGCTGTCATTCGTACGCAAGGCGATAAGCCGTTCGTGGT
TGCGGGTCACTCGGCAGGCGCGCTGATGGCGTACGCGCTGGCGACCGAACTGCTGGATCGTGGTCACCCGCCTCGC
GGTGTTGTCCTGATTGATGTGTACCCGCCTGGTCATCAGGACGCGATGAACGCGTGGCTGGAGGAACTGACGGCAA
CGTTGTTCGACCGTGAAACTGTTCGTATGGACGACACCCGTCTGACCGCGTTGGGTGCTTACGACCGCCTGACGGG
TCAATGGCGTCCTCGCGAAACCGGTTTGCCGACCCTGTTGGTTAGCGCGGGTGAACCAATGGGCCCGTGGCCGGAC
GATAGCTGGAAACCGACCTGGCCTTTCGAGCACGACACCGTTGCGGTCCCGGGTGATCATTTCACGATGGTTCAAG
AACATGCTGATGCGATTGCCCGTCACATTGACGCCTGGCTGGGTGGCGGCAACAGCTGA
```

Plasmid Construction

pSY044: The synthetic gene was designed to have a NdeI site and a BamHI site at the 5' and 3' ends, respectively. These restriction sites were used to subclone the gene into the NdeI and BamHI sites in pET30b to construct pSY044.

pSY066: A PCR amplified NdeI-HindIII fragment corresponding to the (-NL)lipPks1+TE gene, which encodes a N-terminally truncated version of the protein that begins with the amino acid sequence VFVFPGQG, was ligated into NdeI-HindIII-digested pET28b to construct pSY066.

pSY065: A PCR amplified NdeI-EcoRI fragment corresponding to the lipPks1 gene, which encodes a C- terminally truncated version of the protein that ends with the amino acid sequence AYLKAEL, was ligated into NdeI-EcoRI-digested pET28b to construct pSY065.

In vitro Kinetic Analysis of LipPks+TE

Protein Production and Purification

An E. coli K207-3 strain harboring pSY044 was grown in LB medium supplemented with appropriate antibiotics at 37°C until the OD 600 reached 0.4-0.5. The cultures were then cooled to 18°C and induced with 250 μM isopropyl-β-D-galactopyranoside for 16 h. The cells were harvested by centrifugation (4000g, 5 min) and resuspended in lysis/wash buffer (50 mM phosphate, pH 7.6, 300 mM NaCl, 10 mM imidazole, 4°C). The cells were lysed by sonication (8 × 30 sec) and cellular debris was removed by two subsequent centrifugations (4000g, 30 min, 4°C). Nickel-NTA agarose resin (Qiagen) was added directly to the supernatant (1 mL of resin per L of culture) and mixed for 1 h at 4°C. The resulting mixture was poured into a fritted column, washed with 10 resin volumes of lysis/wash buffer (4°C), and eluted

with 2 resin volumes of elution buffer (150 mM phosphate, pH 7.6, 50 mM NaCl, 150 mM imidazole, 4°C). The eluted protein was then applied to a HiTRAP Q anion exchange column (GE Healthcare), washed with 10 resin volumes of wash buffer (50 mM phosphate, pH 7.6, 200 mM NaCl, 8% glycerol, 4°C), and eluted at approximately 375 mM NaCl. The buffer was exchanged into stock buffer (50 mM phosphate, pH 7.6, 8% glycerol, 4°C) by dialysis using an Amicon Ultra-15 Centrifugal Filter, 100K device (Millipore). The resulting purified protein was stored at -80°C.

Starter substrates

The Bradford protein assay was used to measure the concentration of LipPks1+TE. The concentrations of solutions containing various acyl-CoA compounds (acetyl-CoA, propionyl-CoA, n-butyryl-CoA, isobutyryl-CoA, 2-methylbutyryl-CoA, isovaleryl-CoA, or pivaloyl-CoA) were determined using the absorbance at 260 nm and calibration against known CoA concentration standards. Each time point was set up in a reaction volume of 75 L. Samples containing 20 μM, 40 μM, 80 μM, 160 μM, or 320 μM of acyl-CoA were individually incubated with 200 μM methymalonyl-CoA and 500 μM NADPH in the presence of 0.5 μM LipPks1+TE in 100 mM phosphate buffer, pH 7.2, containing 2.5 mM TCEP at 23°C. The reactions were quenched at different time points by adding 75 L of methanol and subsequently filtered with an Amicon Ultra-0.5 mL Centrifugal Filter, 3K device (Millipore) at 23°C. 10 μl of the resulting solutions were analyzed by liquid chromatography-mass spectrometry (LC-MS). Authentic standards (**1a**, **1b**, **1c**, **1d**, **1e**, and **1f**) were used to quantify the products in the LC-MS analysis. We assumed that ionization efficiencies of diastereomers were identical.

Extension substrate

Each time point was set up in a reaction volume of 75 μL. Samples containing 10 μM, 20 μM, or 40 μM methylmalonyl-CoA were individually incubated with 500 μM isovaleryl-CoA and 500 μM NADPH in the presence of 0.5 μM LipPks1+TE in 100 mM phosphate buffer, pH 7.2, containing 2.5 mM TCEP at 23°C. The reactions were quenched at different time points by adding 75 μL of methanol and subsequently filtered with an Amicon Ultra-0.5 mL Centrifugal Filter, 3K device (Millipore) at 23°C. 10 μl of the resulting solutions were analyzed by LC-MS. Authentic standard 1e was used to quantify the products in the LC-MS analysis. We assumed that ionization efficiencies of diastereomers were identical.

Liquid Chromatography

LC separation of the enzymatically generated products conducted at 50°C with an Inertsil ODS-3 reverse-phase column (250 mm length, 2.1 mm internal diameter, 3 M particle size; GL Sciences) using a 1100 series high-performance liquid chromatography system (Agilent Technologies). The mobile phase was composed of 0.2% formic acid in water (solvent A) and 0.2% formic acid in methanol (solvent B). Products were separated using the following

gradient: 60% to 100% B for 10 min, held at 100% B for 2 min, 100% to 60% B for 1 min, held at 60% B for 17 min. A flow rate of 0.14 mL/min was used throughout.

Mass Spectrometry

The LC system was coupled to an Agilent Technologies LC-MSD SL electrospray ionization mass (ESI MS) spectrometer. Nitrogen gas was used as both the nebulizing and drying gas to facilitate the production of gas-phase ions. The drying and nebulizing gases were set to 10 L/min and 20 lb/in^2 , respectively, and a drying gas temperature of 300°C was used throughout. ESI was conducted in the negative-ion mode with a capillary voltage of 4 kV. Mass measurements were carried out in the selected ion monitoring mode (**1a**, m/z 145; **1b**, m/z 131; **1c**, m/z 145; **1d**, m/z 159; **1e**, m/z 159; **1f**, m/z 159) at 1.01 s/cycle with a dwell time of 1 s for the detection of [M-H]$^-$ ions. The instrument was tuned for a range of m/z 50 to 3000 via the Agilent ES tuning mix. Data acquisition and processing were performed by an Agilent Chemstation (Agilent technologies).

In vivo Production of α-Lipomycin and 21-Methyl-α-Lipomycin

Fermentation and Extraction

S. aureofaciens Tü117 was cultured in 1 L of HA medium containing or not-containing 20 mM isoleucine, pH adjusted to 7.3, in 2 L baffled flask at 30 °C on a rotary shaker at 200 r.p.m. After 7 days, mycelia were collected by centrifugation and extracted with 500 mL of acetone. After the extract was concentrated in vacuo, it was combined with the supernatant and acidified with HCl (pH $\approx$ 4). The resulting aqueous solution was then extracted with 500 mL of ethyl acetate. After removal of the organic phase *in vacuo*, the residue was dissolved in 5 mL of acetonitrile.

Liquid Chromatography

The crude extracts were diluted by 100-fold with 50% methanol in water. LC separation was conducted at 50°C with a ZORBAX Eclipse XDB-C8 column (150 mm length, 2.1 mm internal diameter, 3.5 μM particle size; Agilent Technologies) was used. The mobile phase was composed of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) with the following gradient: 15% to 100% B for 10 min, held at 100% B for 2 min, 100% to 15% B for 1 min, held at 15% B for 12 min. A flow rate of 0.23 mL min^{-1} was used throughout.

Mass Spectrometry

The LC system was coupled to an Agilent Technologies LC-MSD SL electrospray ionization mass (ESI MS) spectrometer. Nitrogen gas was used as both the nebulizing and drying gas to facilitate the production of gas-phase ions. The drying and nebulizing gases were set to

10 L min^{-1} and 20 lb in^{-2}, respectively, and a drying gas temperature of 300°C was used throughout. ESI was conducted in the positive-ion mode with a capillary voltage of 4 kV. Mass measurements were carried out in the selected ion monitoring mode (α-lipomycin, m/z 588.7; 21-methyl-α-lipomycin, m/z 602.7) at 1.01 s/cycle with a dwell time of 1 s for the detection of $[M+H]^+$ ions. The instrument was tuned for a range of m/z 50 to 3000 via the Agilent ES tuning mix. Data acquisition and processing were performed by an Agilent Chemstation (Agilent technologies).

Isolation

The crude extracts were diluted by 2-fold with water. LC separation was conducted at 23°C with a ZORBAX SB-C18 column (50 mm length, 9.4 mm internal diameter, 5 μM particle size; Agilent Technologies) with a detection wavelength of 460 nm. The mobile phase was composed of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) with the following gradient: 50% to 55% B for 10 min, held at 55% B for 2 min, 55% to 50% B for 1 min, held at 50% B for 2 min. A flow rate of 5 mL min^{-1} was used throughout. After removal of the organic phase *in vacuo*, the aqueous solution was lyophilized and dissolved in methanol.

UV analysis

UV spectra were obtained on a SpectraMax M2 (Molecular devices). To determine the molar extinction coefficient, a serial dilution of each compound was prepared in methanol and ϵ was calculated by applying the Beer-Lambert law, $A = \epsilon c l$.

NMR analysis

All NMR spectra (^{1}H NMR, ^{13}C NMR, DQF-COSY, TOCSY, and HSQC) for all samples were recorded on a Bruker Avance 600 MHz spectrometer fitted with a TCI cryoprobe. The solvent was methanol-d$_4$ and chemical shifts were referenced against residual water for ^{1}H NMR and deuterated solvent for ^{13}C NMR.

MIC analysis

B. subtilis and *E. coli* were grown in LB medium at 37°C overnight. Each culture was diluted by 100-fold into fresh LB medium and 150 μL dispensed into a 96-well plate. α-Lipomycin or 21-methyl-α-lipomycin was then added into the cultures at 200, 100, 50, 25, 12.5, 6.3, 3.1, 1.6, 0.78, 0.39, 0.20, 0.10 μg m^{-1} . The plates were incubated at 37°C overnight and the growth of *B. subtilis* or *E. coli* was measured by absorbance at 600 nm. Each experiment was performed in duplicate.

A.2 Characterization of Authentic Synthetic Standards
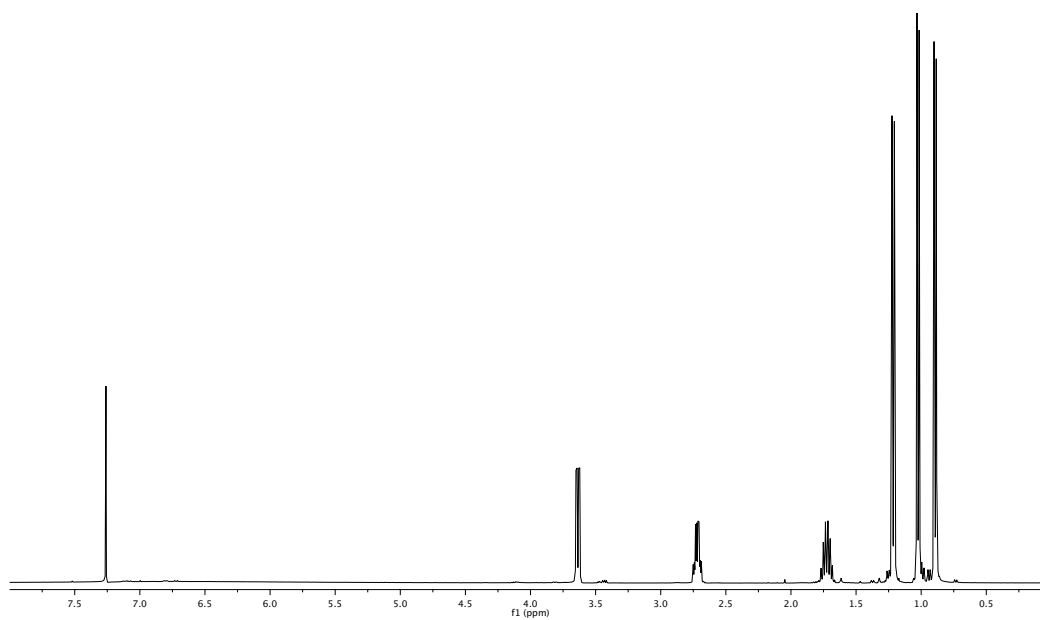
(a) ^{1}H NMR
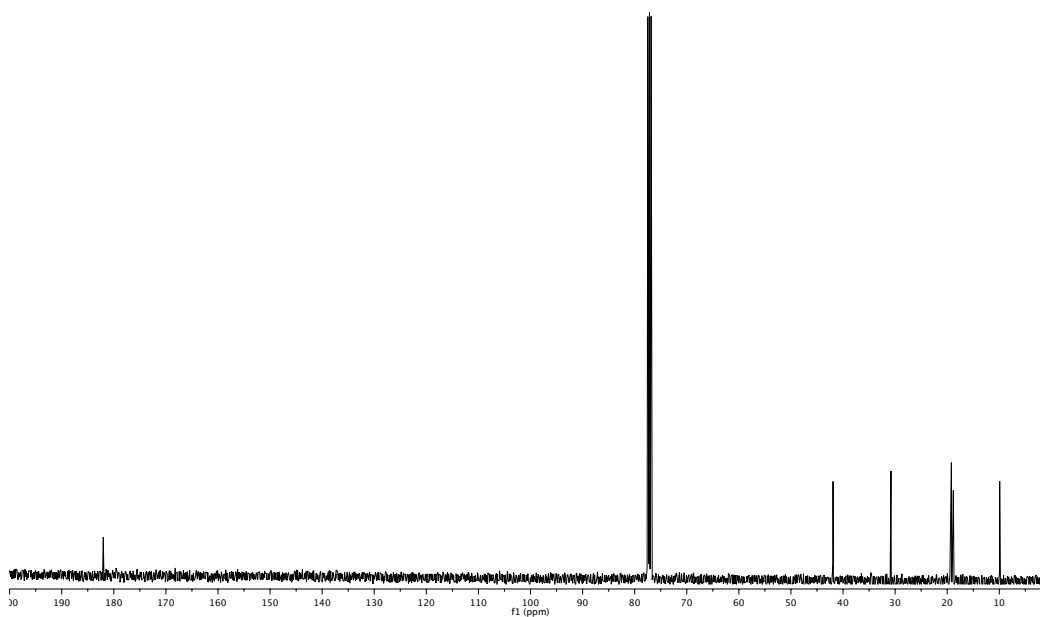
(b) ^{13}C NMR

Figure A.1: Spectral data for product **1a**.
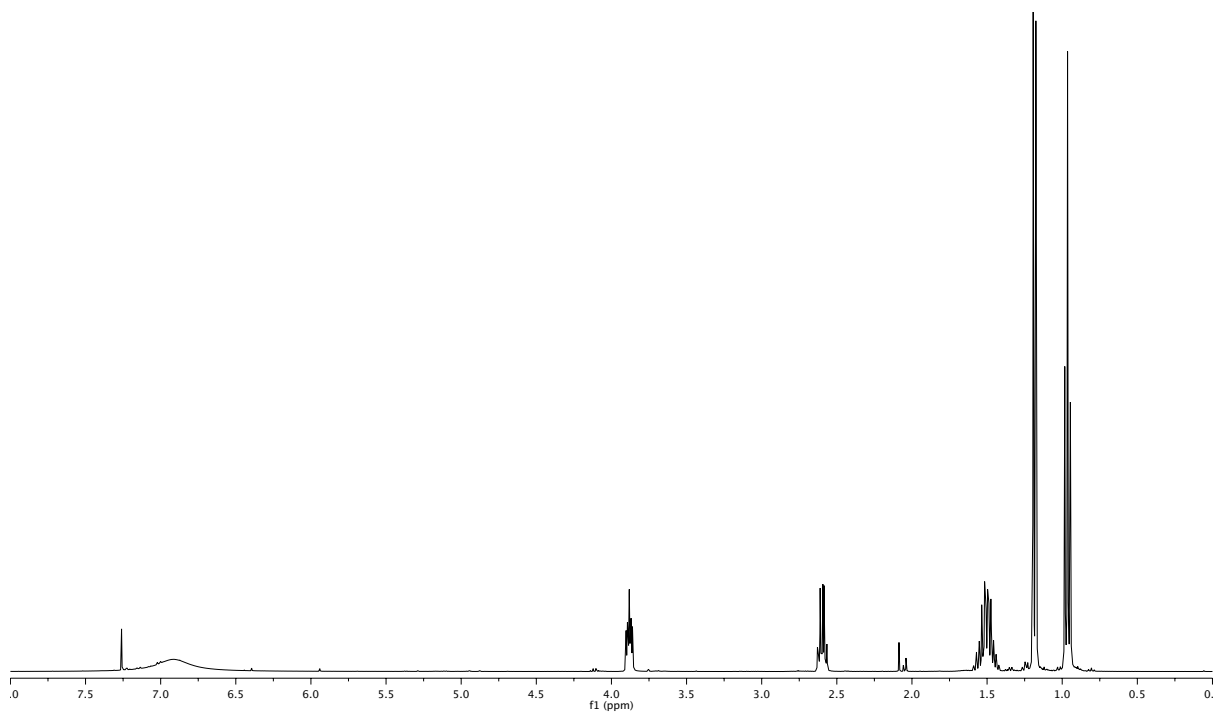
(a) ^{1}H NMR
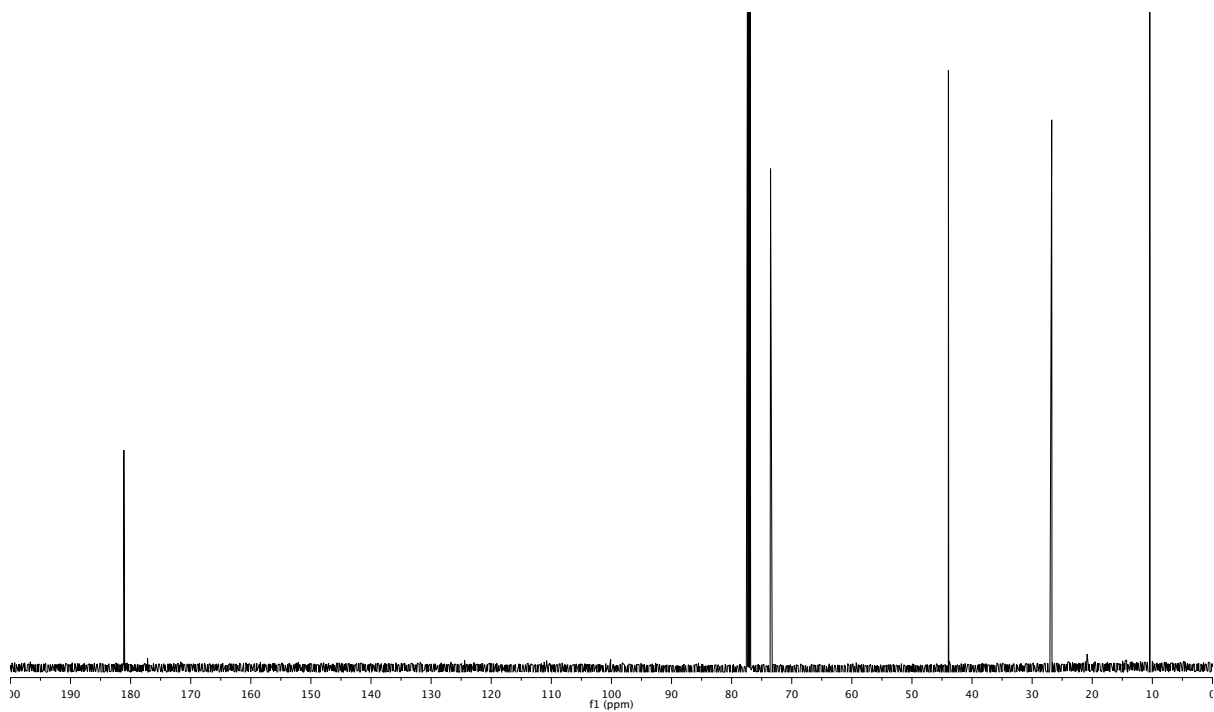
(b) ^{13}C NMR

Figure A.2: Spectral data for product **1b**.
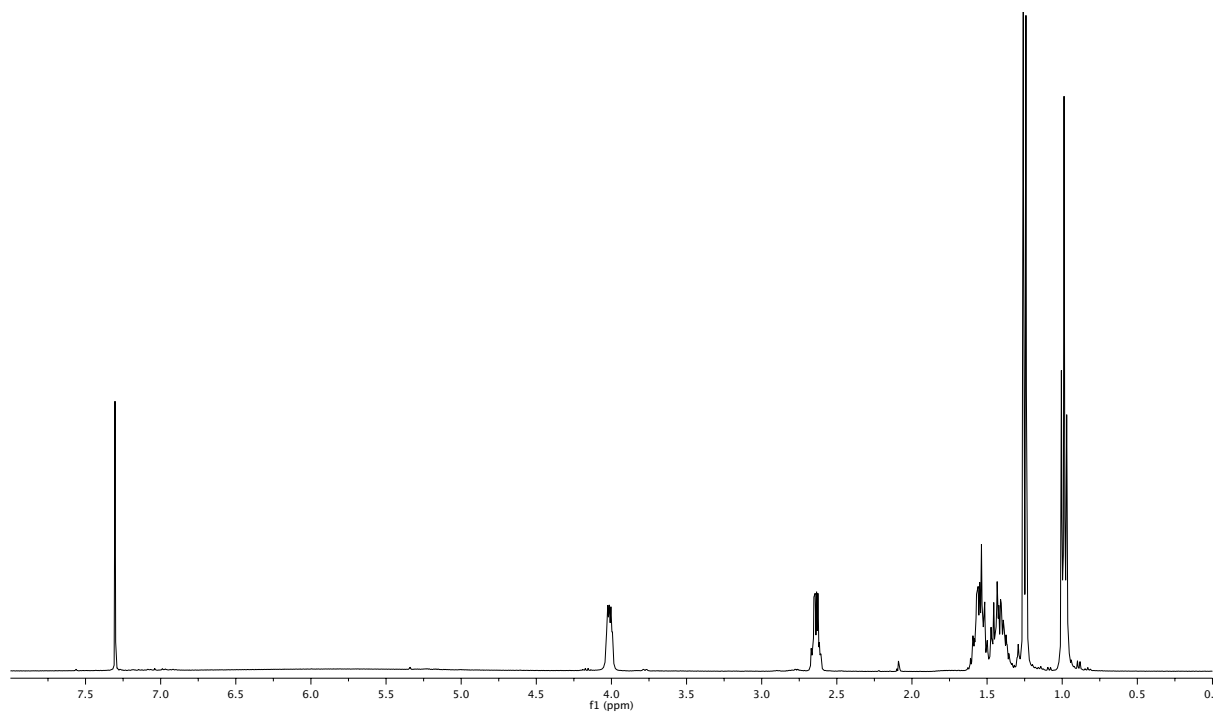
(a) ^{1}H NMR
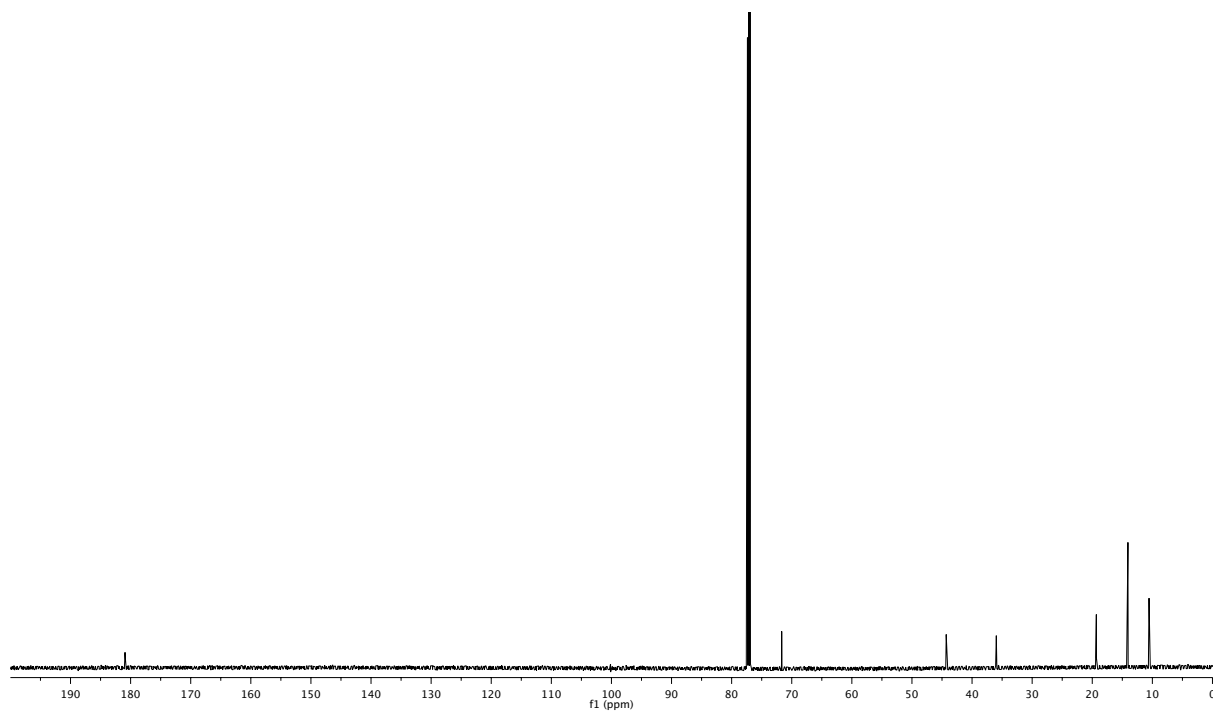
(b) ^{13}C NMR

Figure A.3: Spectral data for product **1c**.
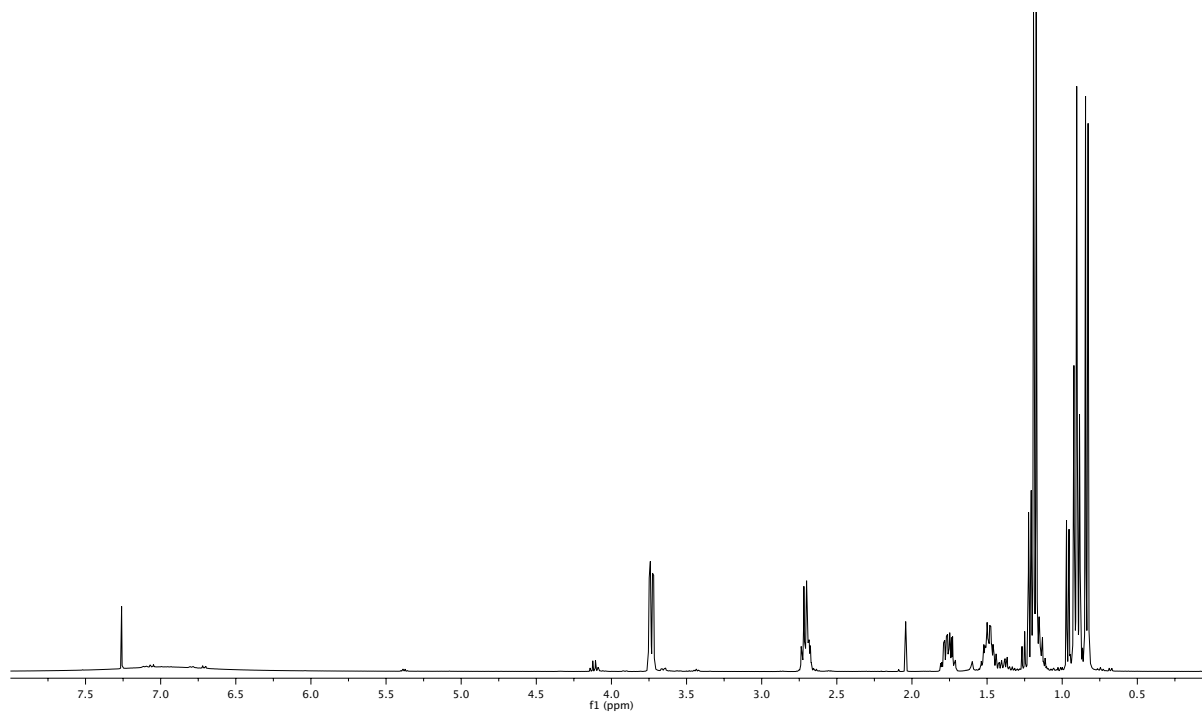
(a) ^{1}H NMR
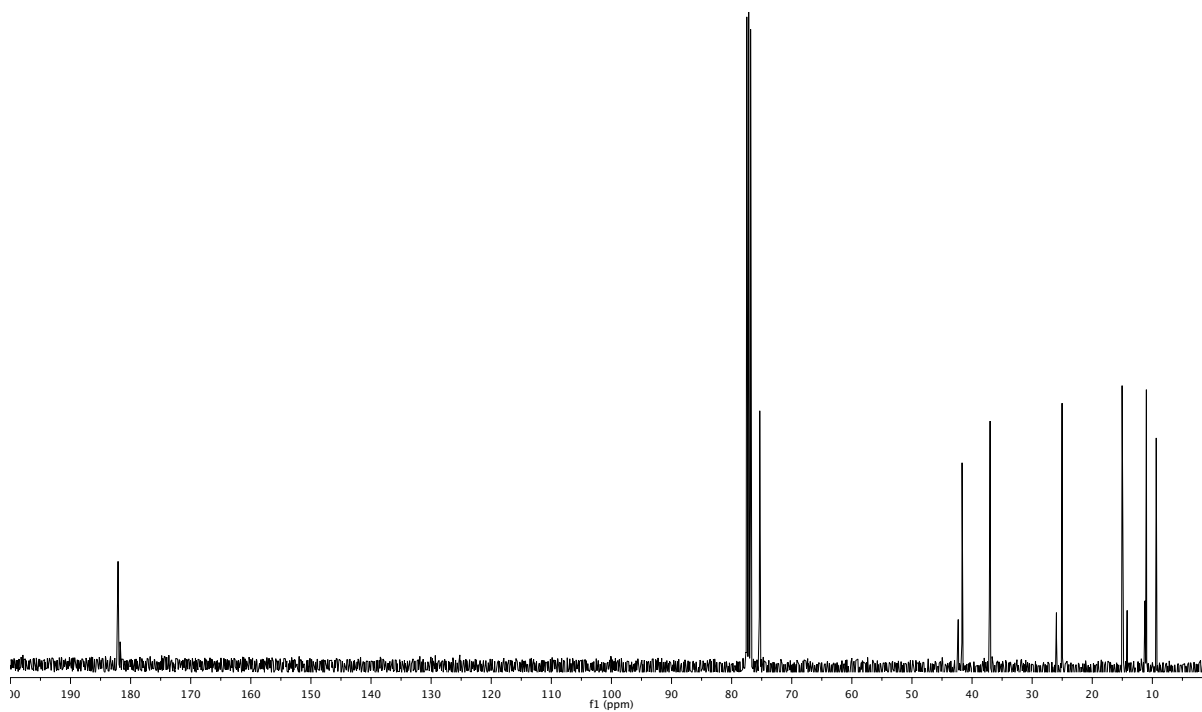
(b) ^{13}C NMR

Figure A.4: Spectral data for product **1d**.
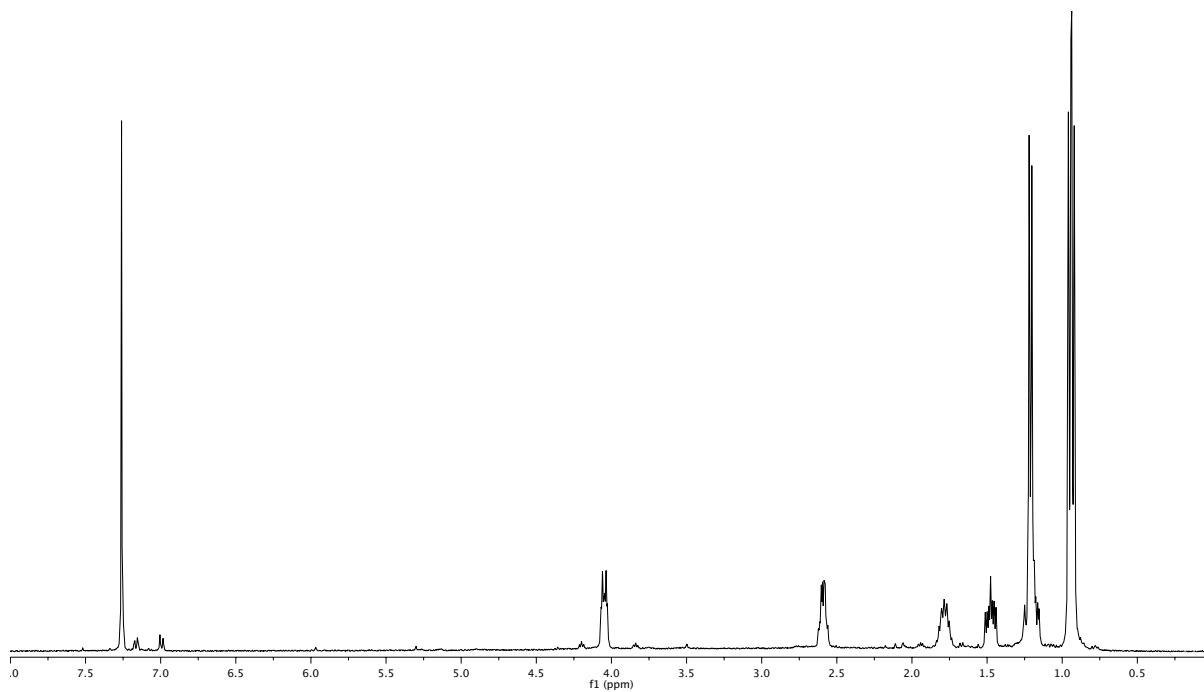
(a) ^{1}H NMR
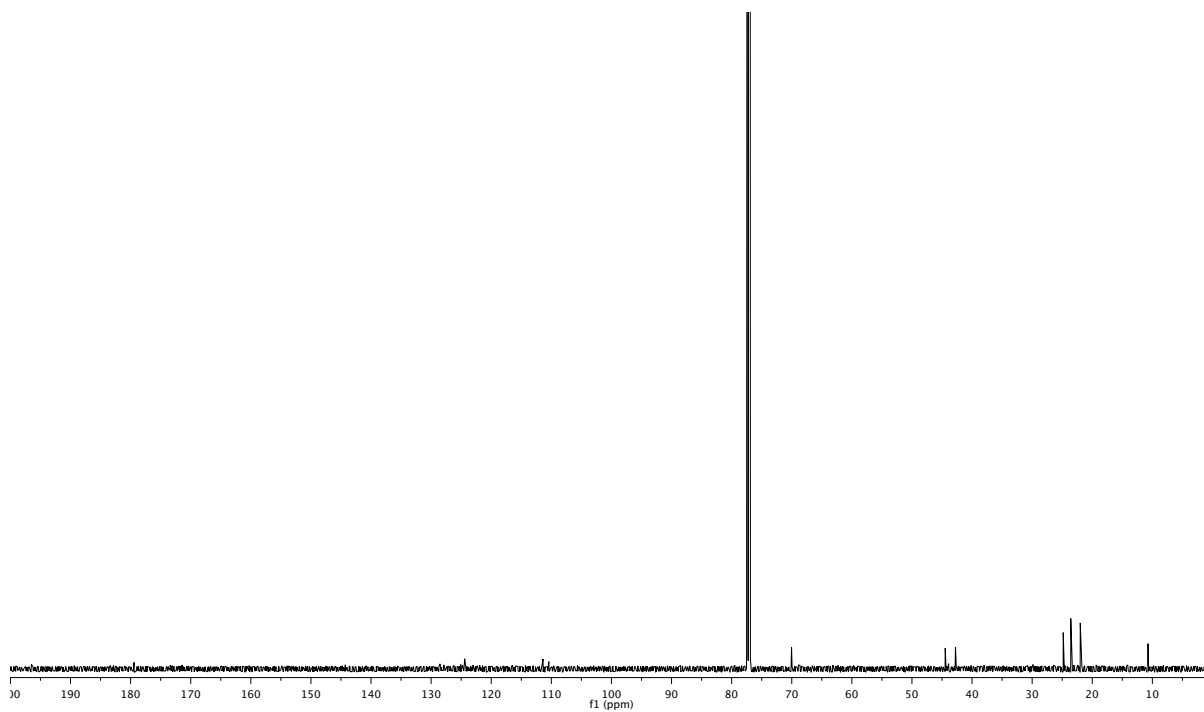
(b) ^{13}C NMR

Figure A.5: Spectral data for product **1e**.
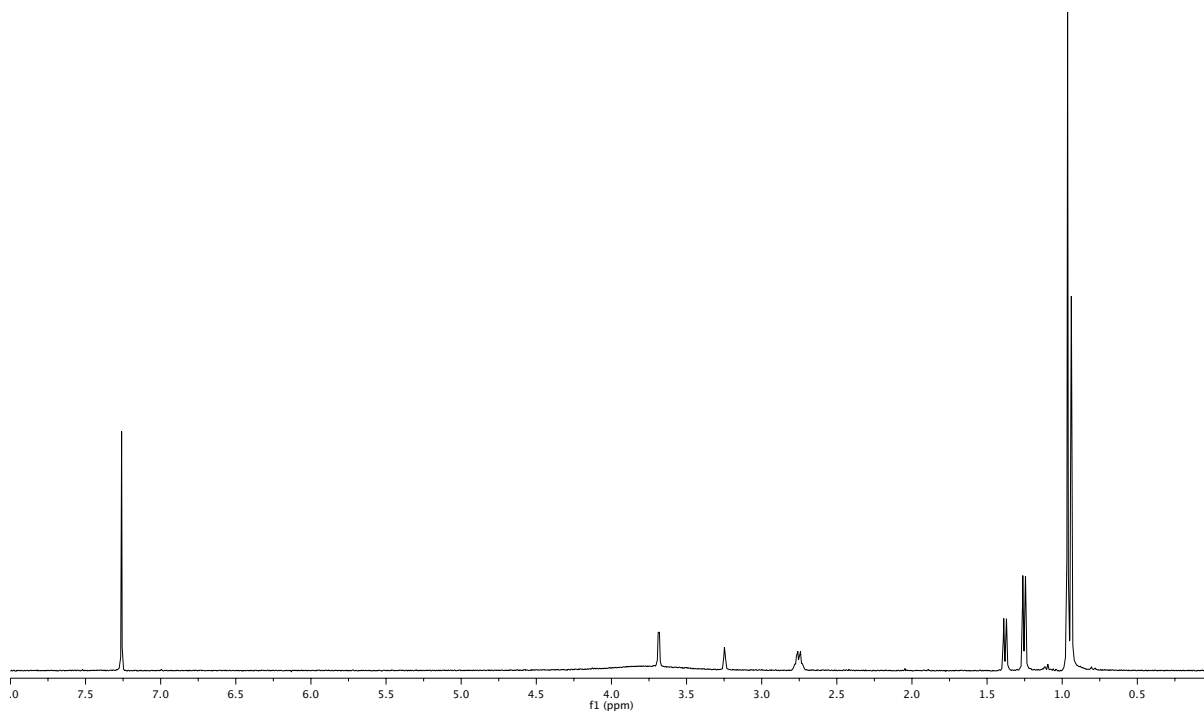
(a) ^{1}H NMR
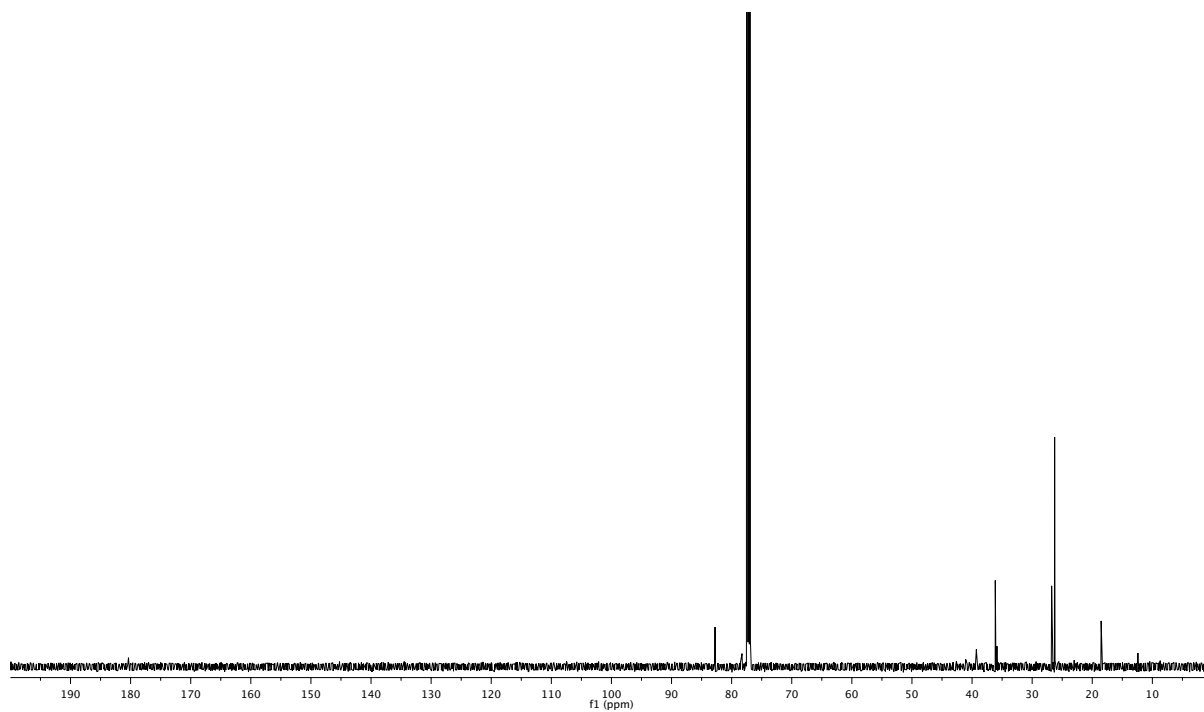
(b) ^{13}C NMR

Figure A.6: Spectral data for product **1f**.
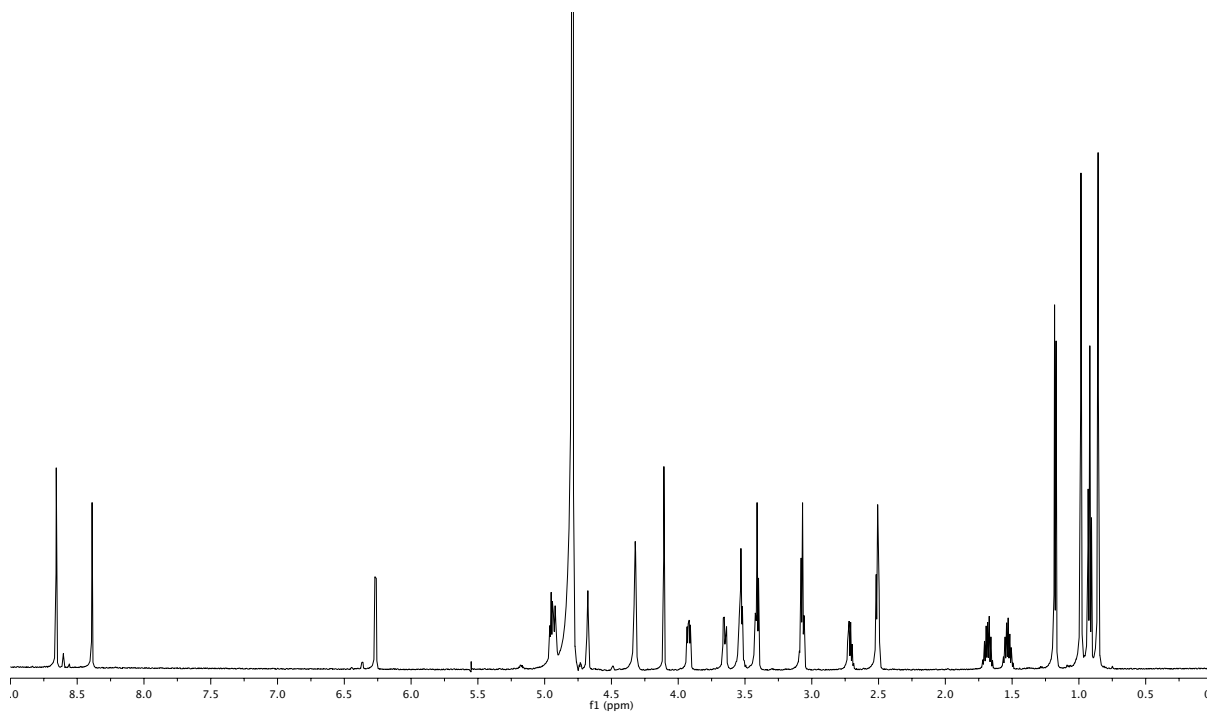
(a) ^{1}H NMR of 2-Methylbutyryl-CoA
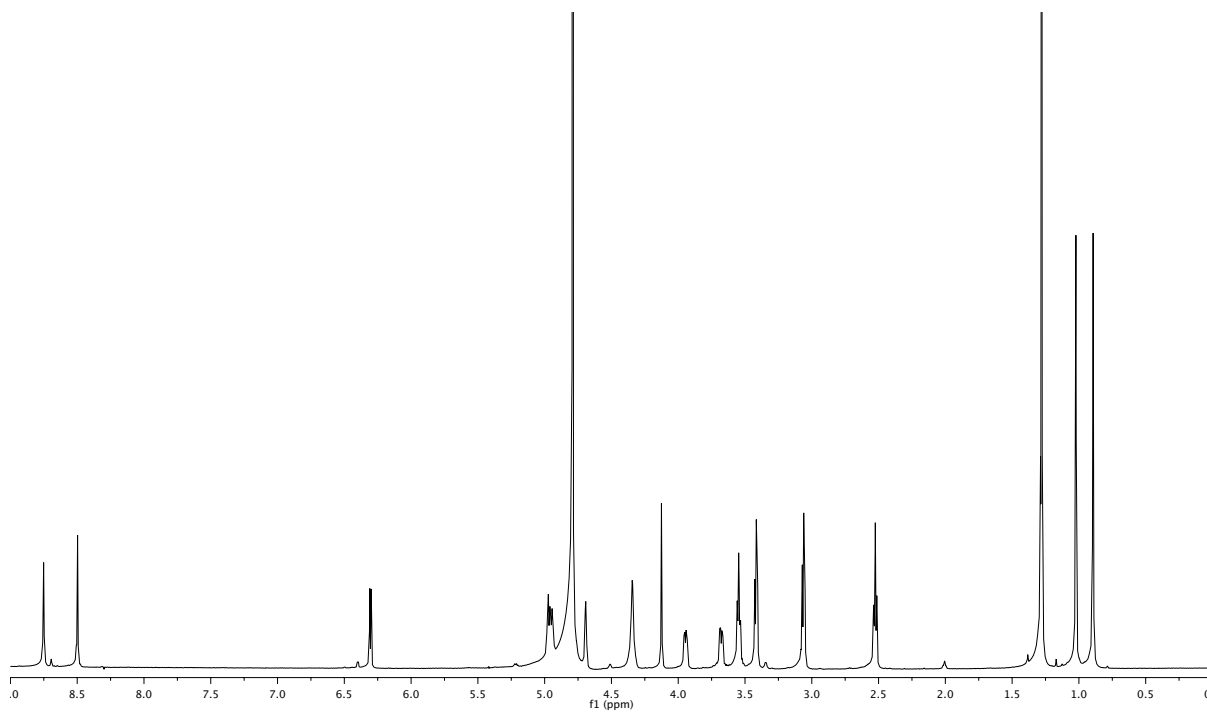
(b) ^{1}H NMR of Pivaloyl-CoA

Figure A.7: Spectral data for CoA products **2a-b**.

A.3   Characterization of α-Lipomycin and 21-Methyl-α-Lipomycin
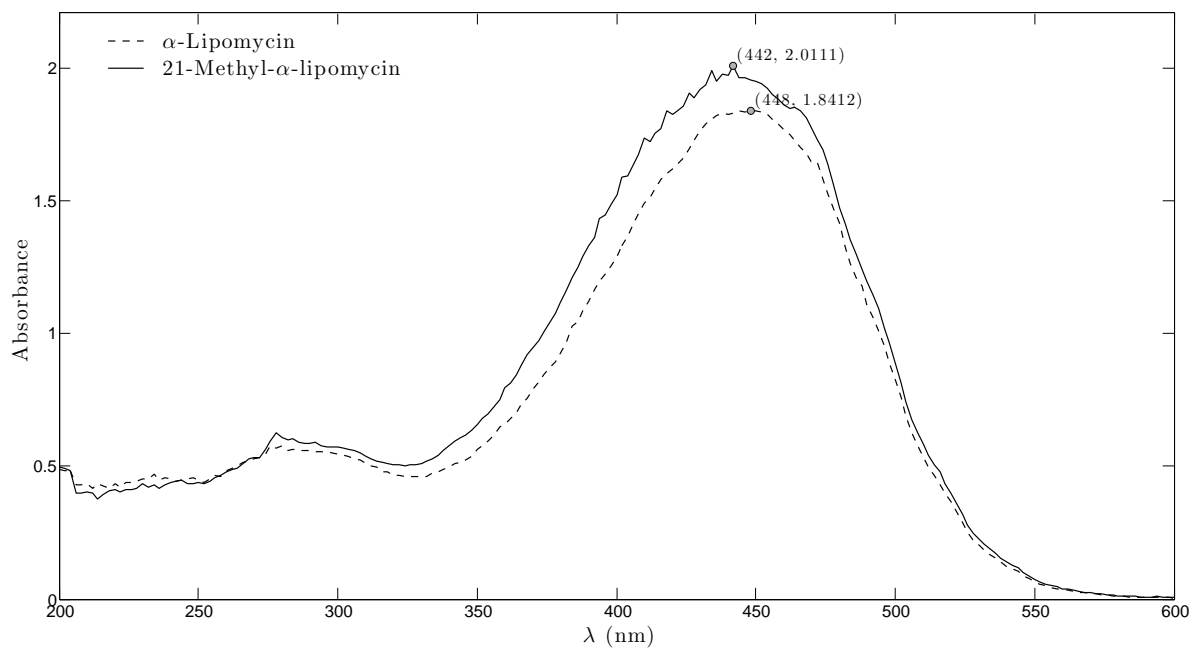
Figure A.8: UV-Vis absorption spectrum of α-lipomycin (dotted line) and 21-methyl-α-lipomycin (solid line).
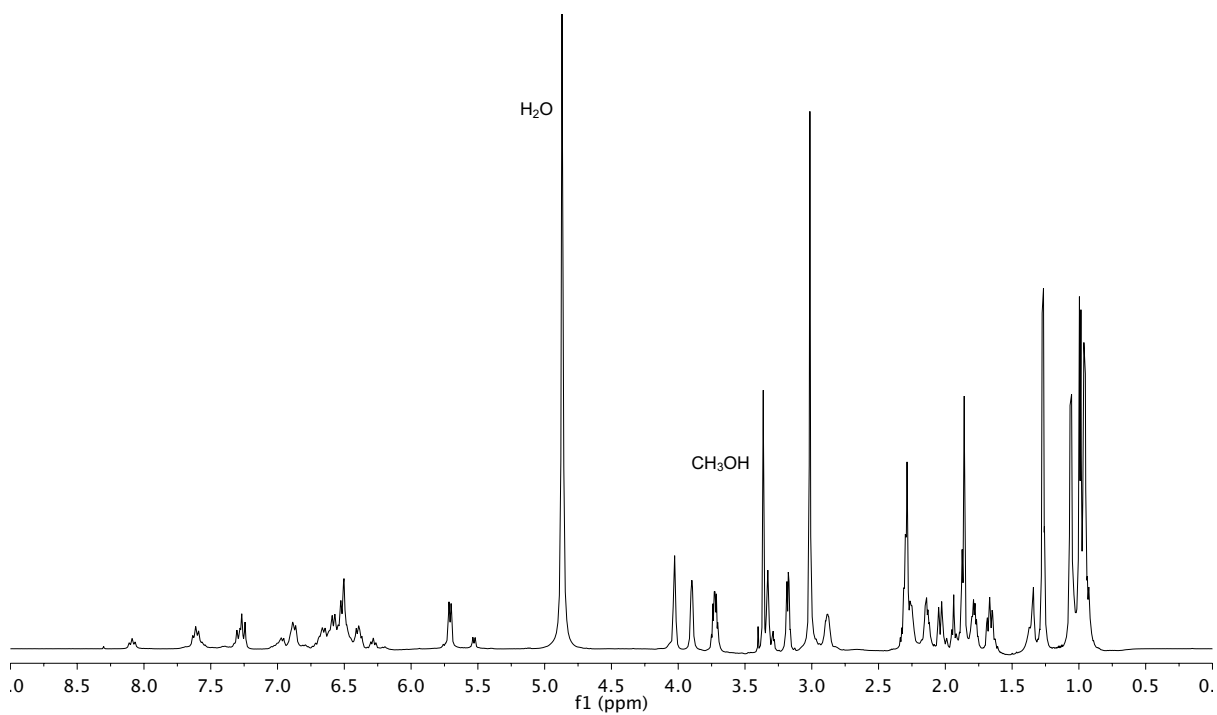
(a) ^{1}H NMR
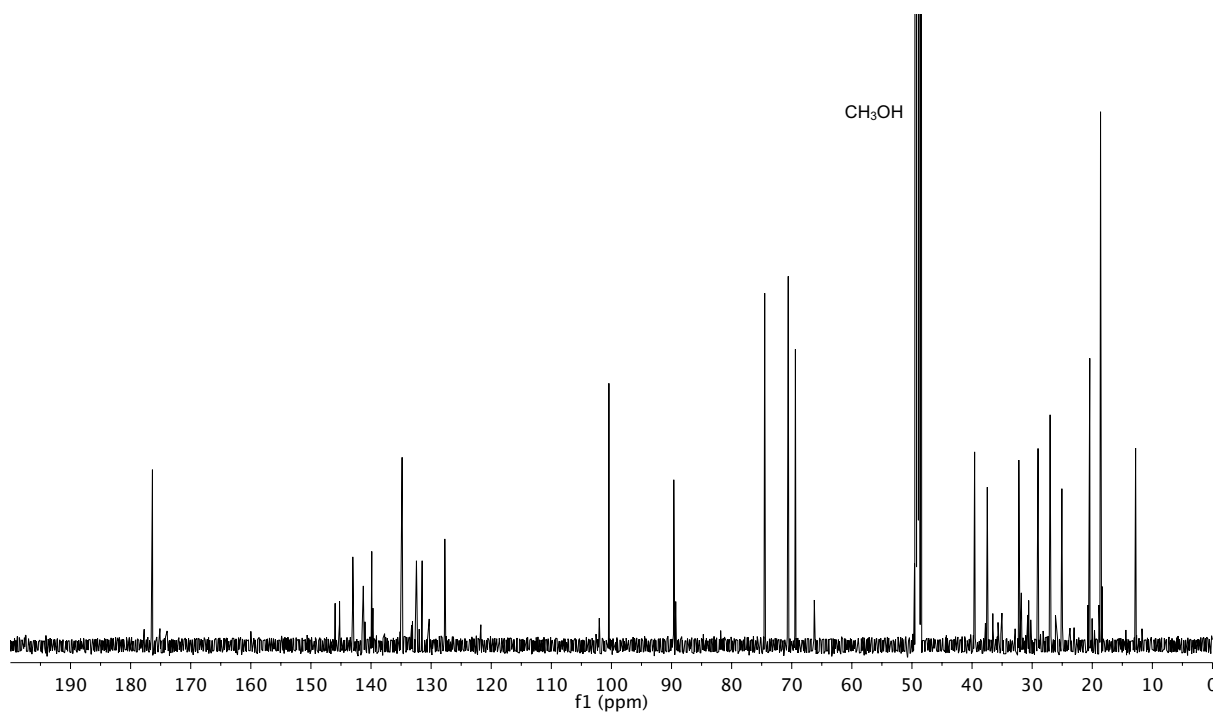
(b) ^{13}C NMR

Figure A.9: 1D NMR spectra of α-lipomycin.
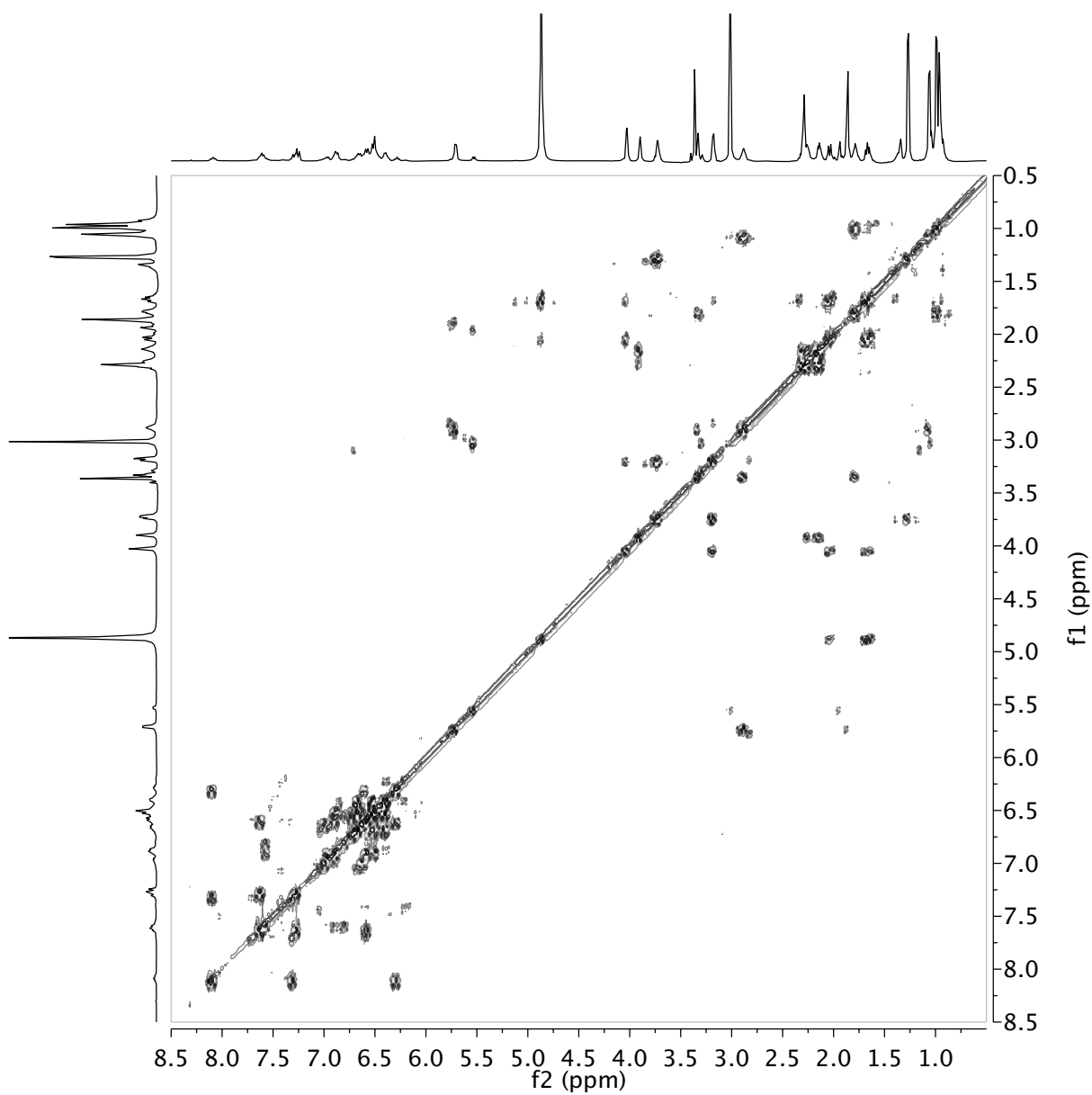
Figure A.10: DQF-COSY spectrum of α-lipomycin.
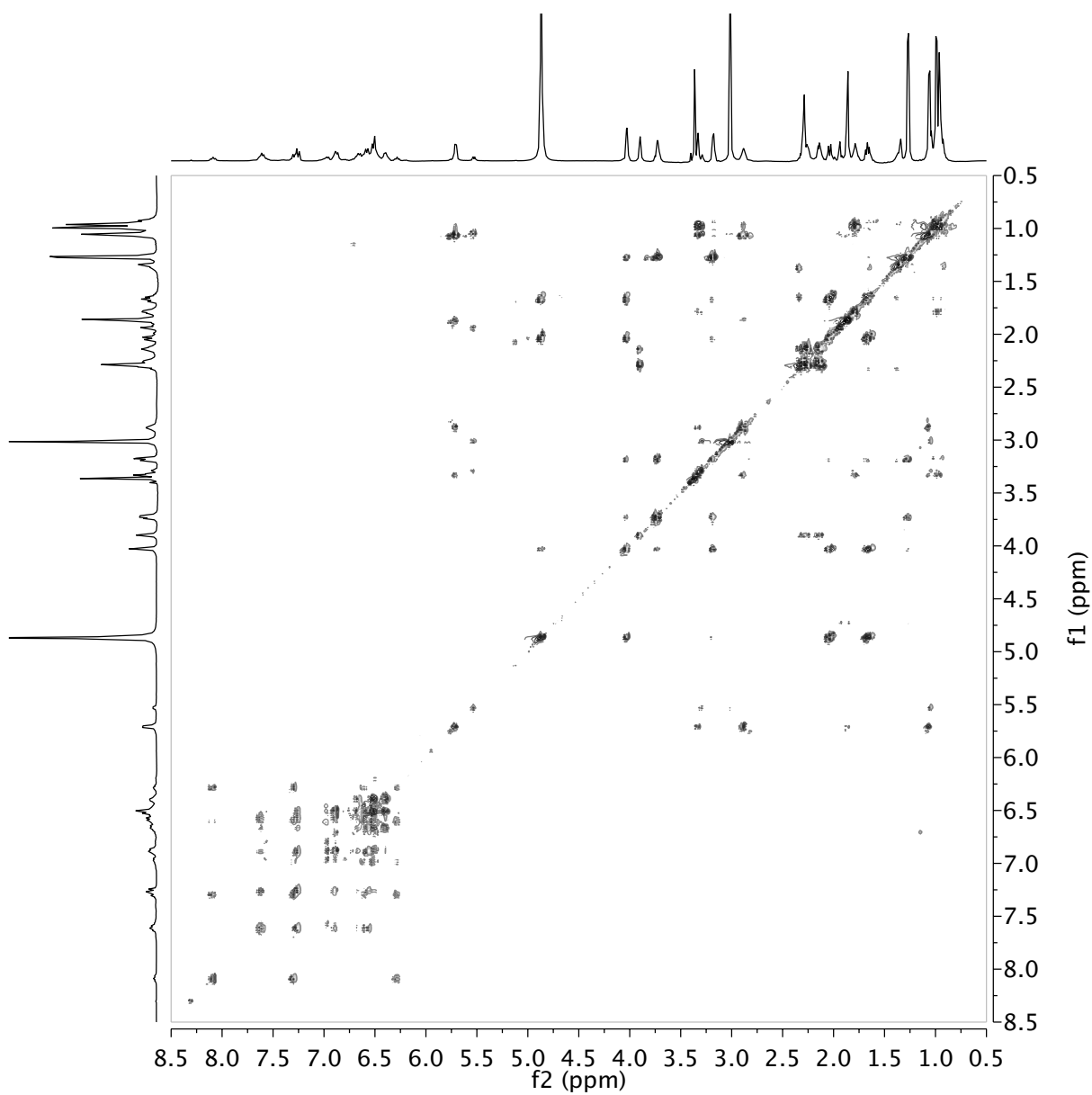
Figure A.11: TOCSY spectrum of α-lipomycin.
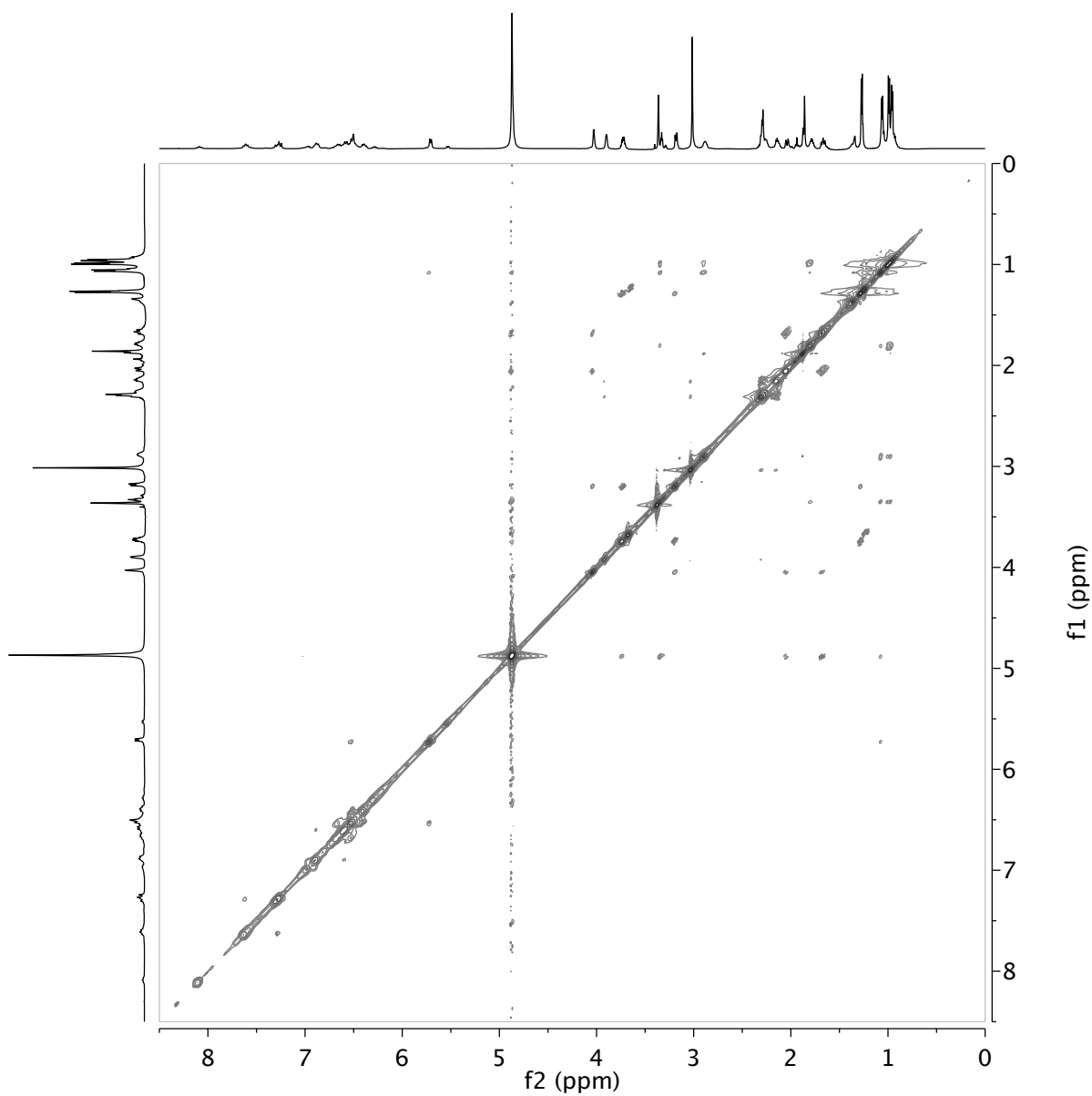
Figure A.12: NOESY spectrum of α-lipomycin.
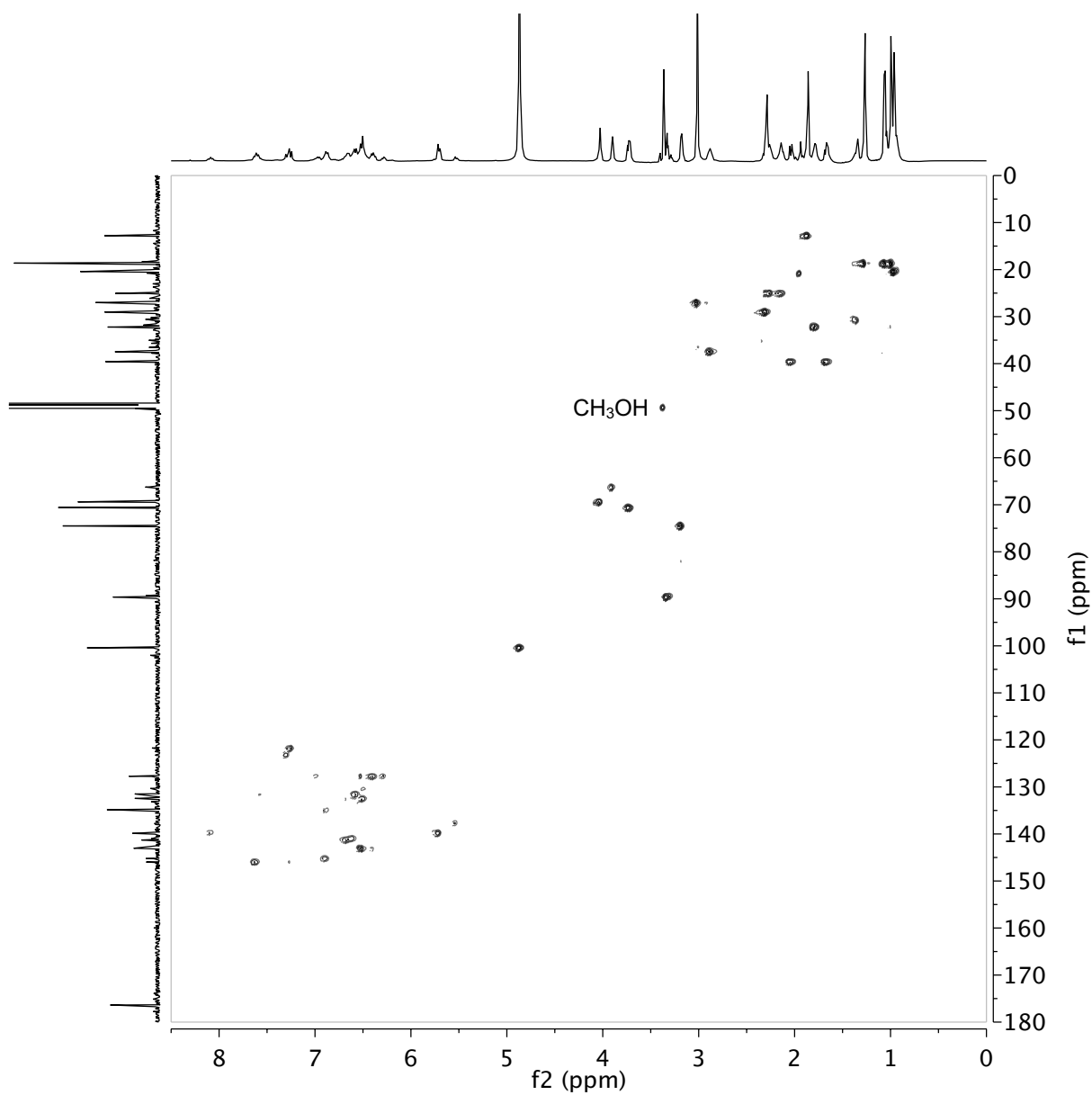
Figure A.13: HSQC spectrum of α-lipomycin.
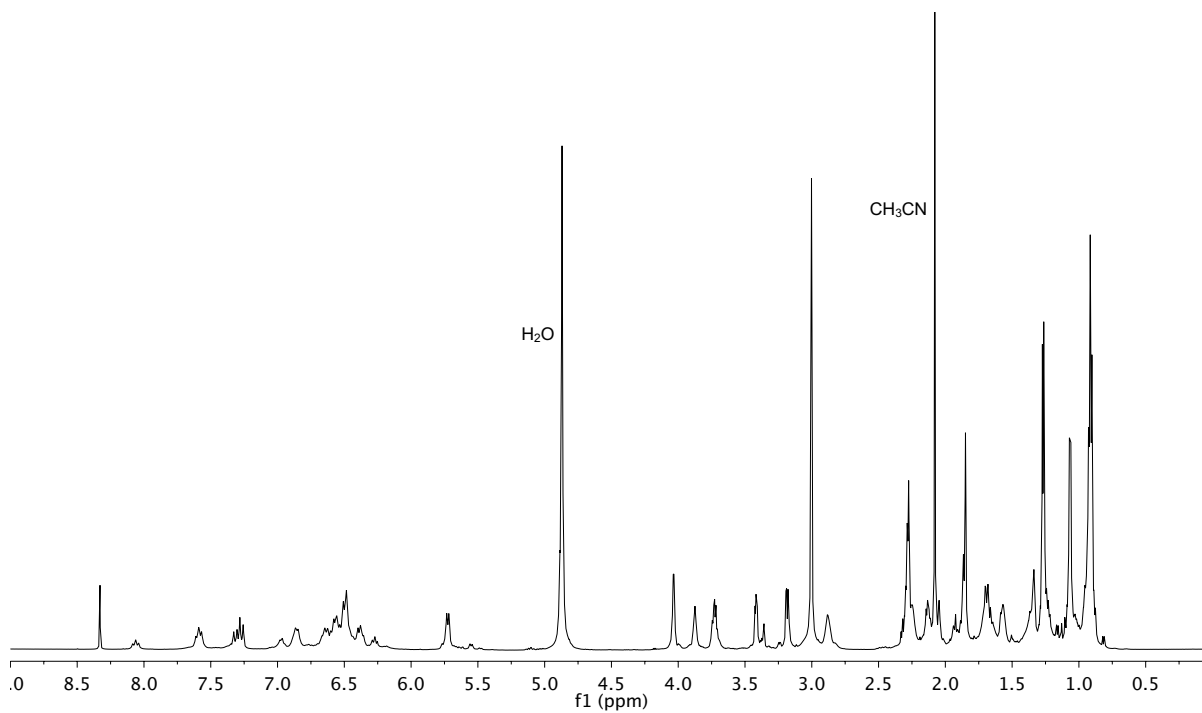
(a) ^{1}H NMR
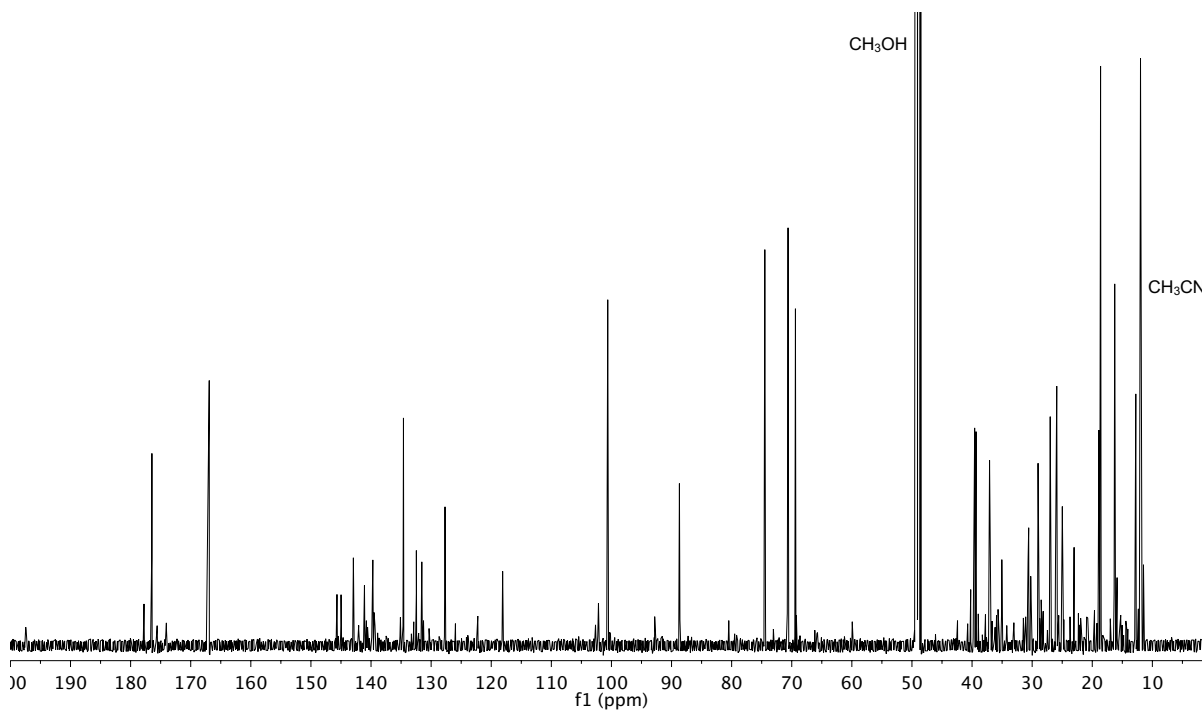
(b) ^{13}C NMR

Figure A.14: 1D NMR spectra of 21-methyl-α-lipomycin.
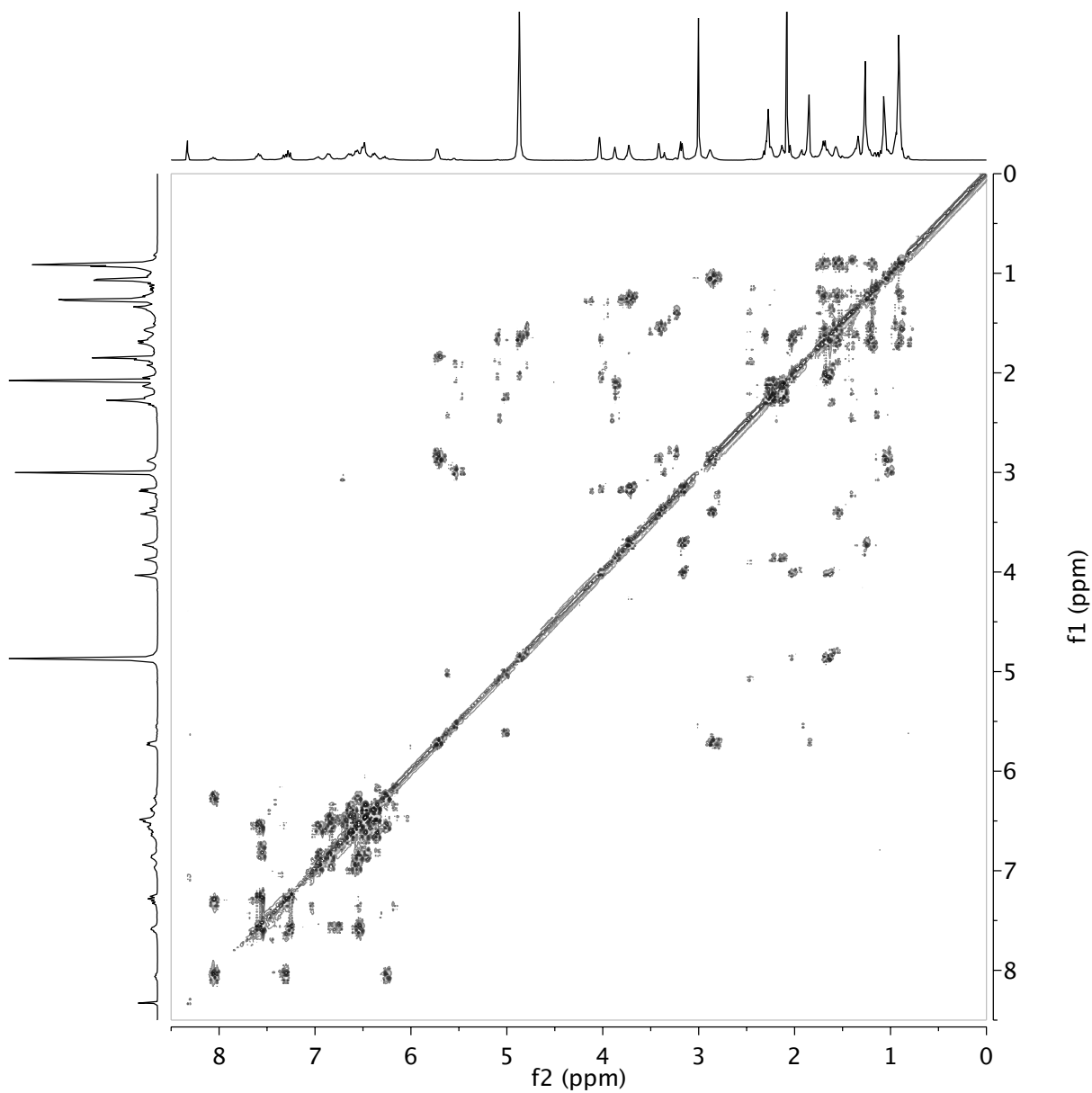
Figure A.15: DQF-COSY spectrum of 21-methyl-α-lipomycin.
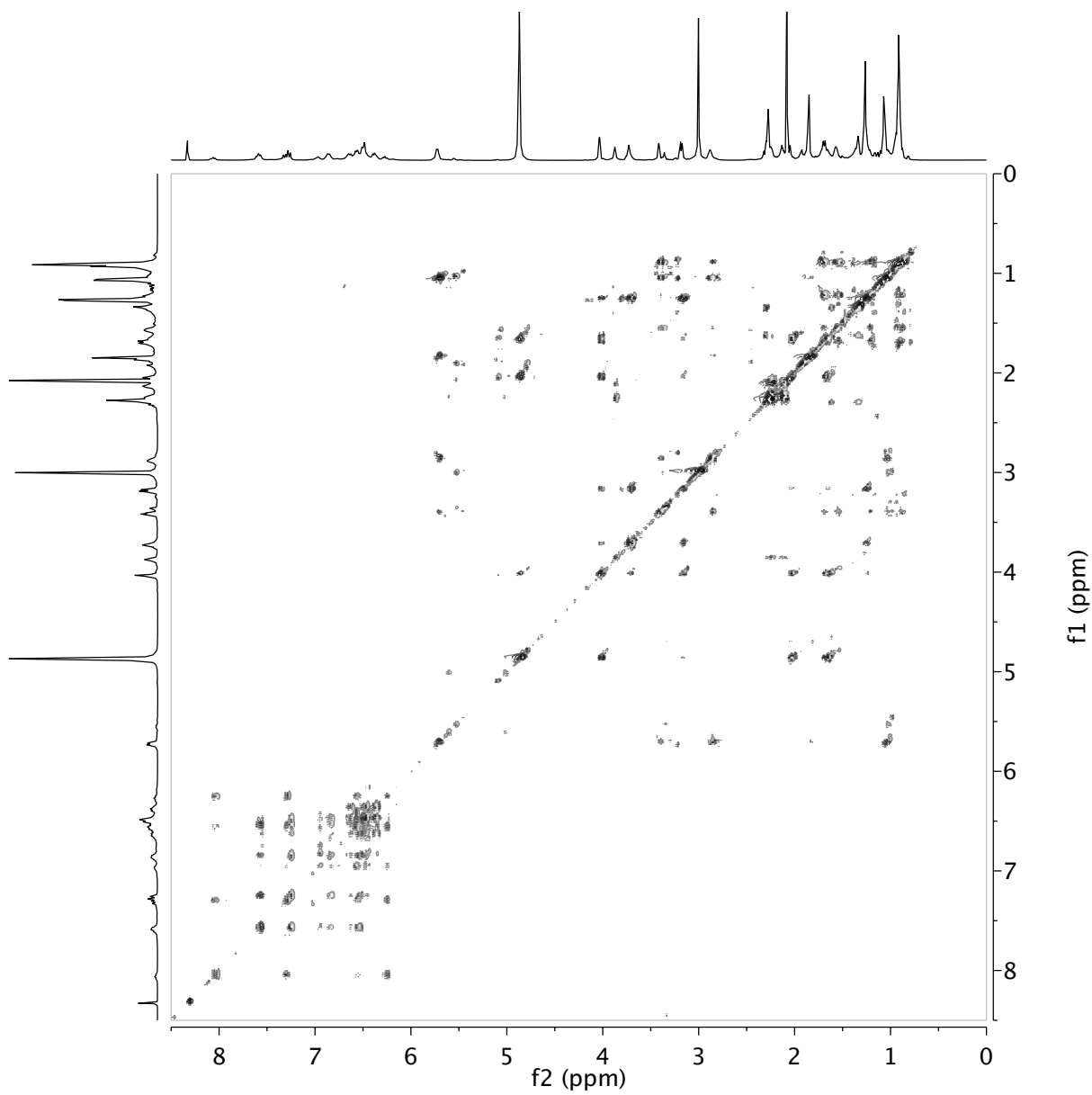
Figure A.16: TOCSY spectrum of 21-methyl-α-lipomycin.
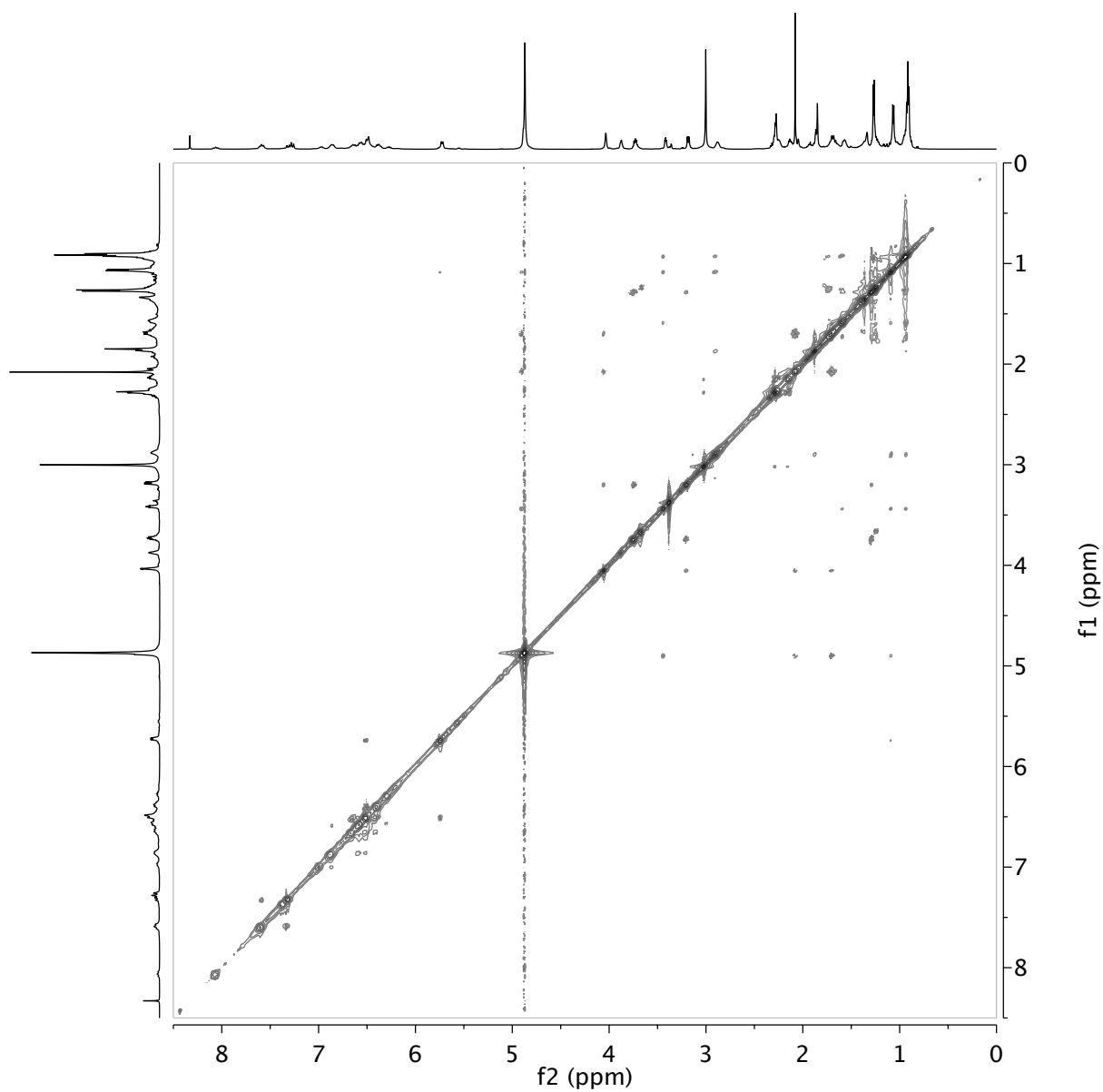
Figure A.17: NOESY spectrum of 21-methyl-α-lipomycin.
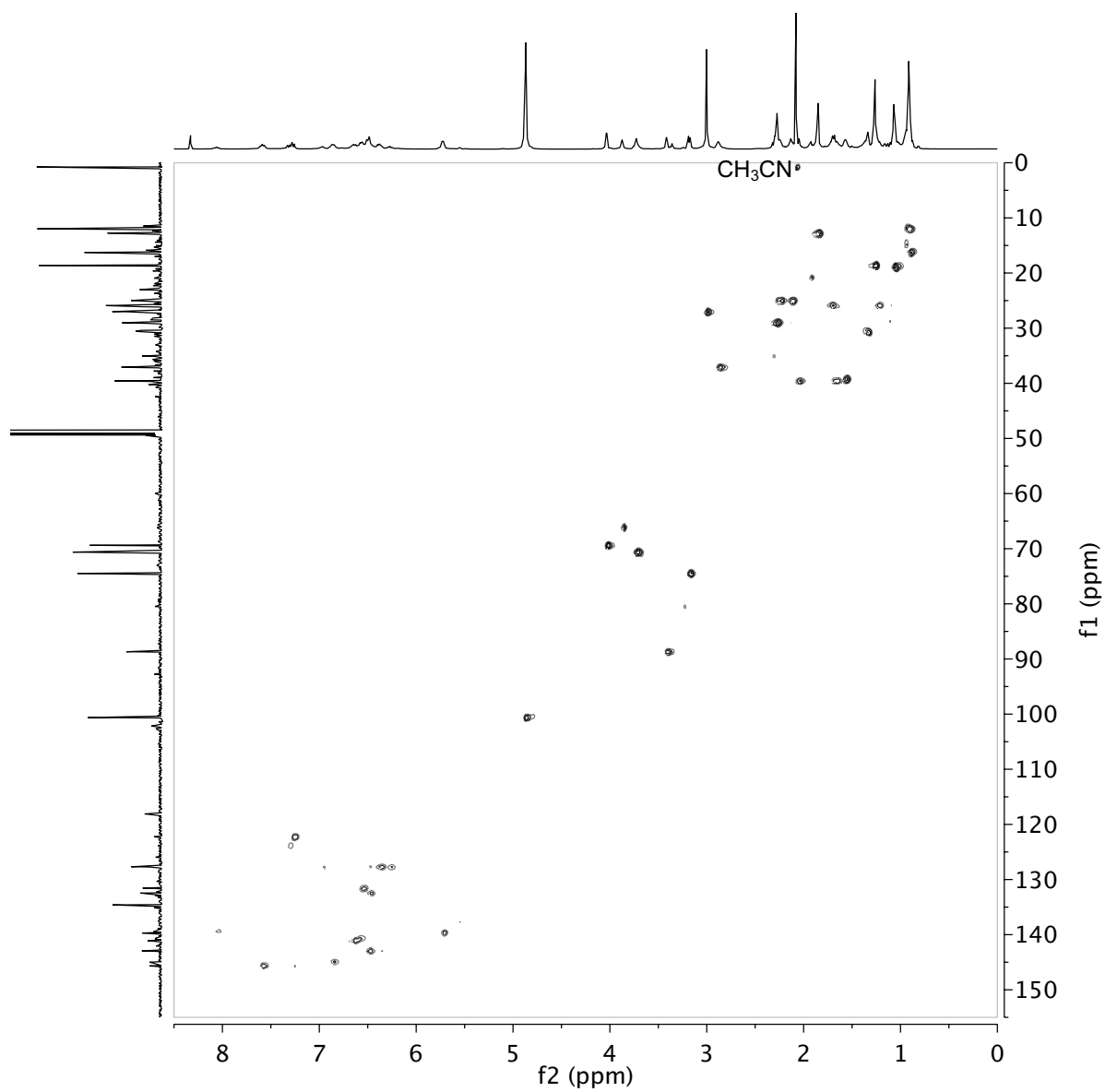
Figure A.18: HSQC spectrum of 21-methyl-α-lipomycin.

Appendix B

Chapter 3 Supplementary Information

B.1 Experimental Procedures

Authentic Synthetic Standards

Racemic mixtures of *syn* **3a-d** were synthesized as previously described [80]. (2*R*,3*S*)-**3a** (*syn* A1 type product) and (2*S*,3*S*)-**3a** (*anti* A2 type product) were purchased from Organic Consultants Incorporated. All other chemicals were purchased from Sigma-Aldrich unless otherwise described.

Junction Design for KR Domain Exchanges

All unique accession numbers associated with type I modular PKS KS domains present in the curated Clustermine360 database [14] were compiled into a local BLAST nucleotide database using BLAST+ [8]. The 45 homologs of LipPks1 present in this local database which had over 75% sequence coverage of the query were identified using `tblastn`. Next, the `mafft-linsi` algorithm from the Mafft software package [38] was used to perform a sequence alignment and provide an estimate of sequence conservation. The amino acid sequences corresponding to the KR and ACP structures present in the Protein Data Bank (PDB) [1, 42, 39, 88, 84, 86] were separately aligned with the amino acid sequence of LipPks1, again using the `mafft-linsi` algorithm. These two sequence alignments were used to select junctions for KR domain exchanges.

Cloning

A LipPks1+TE variant in which the DE was removed was generated by PCR amplication followed by phosphorylation and ligation. This DE deletion variant differed from wild-type LipPks1+TE only in that a 54 amino acid sequence preceding the KR starting with RPA and ending with AAV, shown in Supplementary Figure B.1, was removed. A nine amino acid spacer sequence between the AT and KR was left intact, consistent with what is typically

observed in modules lacking a native DE. Gene fragments encoding Amp DE_2, Amp KR_2, Con DE_2, and Con KR_2 were purchased as gBlocks from Integrated DNA Technologies. *E. coli* codon optimized synthetic genes encoding additional donor KR domains were ordered either from Genewiz (Amp KR_1, Pik KR_1, and Spn KR_3) or from Genscript (Ave KR_1, Ery KR_1, and Ery KR_6). Bor KR_1 was amplified from the plasmid pARH100 [28]. All cloning was performed in the *E. coli* strain DH10B. Plasmid backbones containing LipPks1+TE but lacking either the native DE and KR, or just the native KR were amplified from pSY044 [80]. All oligonucleotides used as primers in this study were designed using j5 DNA assembly design software [32], and are shown in Supplementary Tables B.1 and B.2. Plasmids were constructed using a Golden Gate protocol adapted from work by Engler *et. al.* [19, 20] for use with j5 [32].

Protein Expression and Purification

Expression vectors containing LipPks1+TE variants were transformed into the *E. coli* strain K207-3 [55]. Cultures were grown at 37°C in Luria-Bertani (LB) medium supplemented with the appropriate antibiotic until an OD_{600} of 0.35-0.45 was reached. The cultures were cooled to 18°C, induced with 250 μM isopropyl-β-D-galactopyranoside (IPTG), and further incubated at 18°C for 18 h. Cells were collected by centrifugation (5000 rpm, 5 min, room temperature), and resuspended in 50 mL suspension buffer (50 mM phosphate buffer, pH 7.6, 300 mM sodium chloride, 10 mM imidazole) per 1 L culture. The cells were lysed by sonication, and cellular debris removed by three consecutive centrifugation steps (8000 rpm, 15 min, 4°C). The supernatant was mixed with 10 mL Nickel-NTA agarose resin (Qiagen) per 1 L culture for 1 h at 4°C. The resulting mixture was added to a fritted column, washed with 18 resin volumes of suspension buffer, and the protein eluted with four resin volumes of elution buffer (150 mM phosphate buffer, pH 8.3, 50 mM sodium chloride, 150 mM imidazole). The eluted protein was further purified by anion exchange using a HiTRAP Q column (GE Healthcare). A mixture of 50 mM phosphate buffer, pH 7.6, 8% w/v glycerol (buffer A) and 50 mM phosphate buffer, pH 7.1, 500 mM sodium chloride, 8% w/v glycerol (buffer B) was used as the mobile phase, and the protein purified using the following gradient method: 0 to 40% B over 5 min at a flow rate of 2.5 ml/min, hold at 40% B for 10 min using a flow rate of 1 mL/min, 40 to 100% B over 40 min at a flow rate of 1 mL/min. The purified protein was concentrated using a Amicon Ultra-15 centrifugal filter unit with Ultracel-100 membrane (Millipore) and stored at -80°C.

In Vitro Production of 3a-d by LipPks1+TE Variants

Each LipPks1+TE variant (0.5 μM) was incubated in a 50 μL sample containing 100 mM sodium phosphate buffer (pH 7.2), 2.5 mM TCEP, 500 μM NADPH, 216 μM starter acyl-CoA (one of **1a-d**), and 236 μM methylmalonyl-CoA. After incubation at room temperature for 17.5 h, samples were quenched using 50 μL methanol [80].

Separation by LC was performed at 50°C on an Agilent 1100 series high-performance liquid chromatography system (Agilent Technologies) using an Inertsil ODS-3 reverse-phase column (250 mm length, 2.1 mm internal diameter, 3 μm particle size; General Life Sciences). A mixture of 0.1% formic acid in water (solvent A) and 0.1% formic acid in methanol (solvent B) was used as the mobile phase, and the products were separated using the following gradient method: 60 to 100% B over 10 min, hold at 100% B for 2 min, 100 to 60% B over 1 min, hold at 60% B for 17 min, all at a constant flow rate of 0.13 mL/min. A sample injection volume of 10 μL was used.

The LC system was coupled to an Agilent 1100 series mass spectrometer detector (Agilent Technologies). Atmospheric pressure ionization electrospray (API-ES) was used to produce gas-phase ions. Nitrogen gas at a flow rate of 10 L/min and pressure of 30 lb/in^2 was used for both nebulization and drying. A drying gas temperature of 330°C was used. Electrospray ionization (ESI) was performed in negative ion mode with a capillary voltage of 3.5 kV. The instrument was tuned for a m/z range of 50 to 3000 using the Agilent ESI tuning mix. [M-H]$^-$ ions were detected in selected ion monitoring (SIM) mode (**3a**, m/z 145; **3b**, m/z 131; **3c**, m/z 145; **3d**, m/z 159) with a fragmentor voltage of 70 V, gain of 5 EMV, and actual dwell of 746. The ionization efficiencies of stereoisomers were assumed to be identical. Data acquisition was performed using ChemStation software (Agilent Technologies), and data analysis was performed using MassHunter software (Agilent Technologies).

In Vitro Production of 2a-d by LipPks1+TE Variants

Each LipPks1+TE variant (0.5 μM) was incubated in a 50 μL sample containing 100 mM sodium phosphate buffer (pH 7.2), 2.5 mM TCEP, 216 μM starter acyl-CoA (one of **1a-d**), and 236 μM methylmalonyl-CoA in water. After incubation at room temperature for 17.5 h, samples were quenched using 50 μL methanol. The samples were then sealed, incubated at 50°C overnight, and subsequently chilled at 4°C for 1 hour to obtain ketone products [81].

Separation by LC was conducted at 55°C with a Kinetex XB-C18 reversed phase column (100 mm length, 2.1 mm internal diameter, 2.6 μm particle size; Phenomenex) using an Agilent 1200 Rapid Resolution LC system (Agilent Technologies). The mobile phase was composed of water (solvent A) and methanol (solvent B). **2a-b** and **2d** were each separated via the following gradient: 25 to 100% B over 1.5 min, hold at 100% B for 3 min, 100 to 25% B over 0.2 min, and hold at 25% B for an additional 5.3 min. **2c** was separated via the following gradient: 50 to 100% B over 1.5 min, hold at 100% B for 3 min, 100 to 50% B over 0.2 min, and hold at 50% B for an additional 5.3 min. For both separation methods, the sample injection volume was 10 μL, and the following flow rate program was used: hold constant at 0.24 mL/min from 0 to 4.5 min, increase from 0.24 to 0.4 mL/min over 0.2 min, and hold constant at 0.4 mL/min for an additional 5.3 min. This resulted in a total run time of 10 min for each method.

The Agilent 1200 Rapid Resolution LC system was coupled to an Agilent 6210 time-of-flight (TOF) mass spectrometer (Agilent Technologies). Nitrogen gas was used as both the nebulizing and drying gas to facilitate the production of gas-phase ions. The drying and

nebulizing gases were set to 10 L/min and 30 lb/in^2, respectively, and a drying gas temperature of 320°C was used throughout. Atmospheric pressure chemical ionization (APCI) was conducted in the positive-ion mode with capillary and fragmentor voltages of 3.5 kV and 100 V, respectively. The skimmer, OCT1 RF, and corona needle were set to 50 V, 170 V, and 4 μA, respectively. The vaporizer was set to 350°C. **2a-d** were detected via [M+H]$^+$ ions: **2a**, $m/z = 101.09609$; **2b**, $m/z = 87.08044$; **2c**, $m/z = 101.09609$; **2d**, $m/z = 115.11174$. The analysis was performed using an m/z range of 66 to 166. Data acquisition and processing were performed using MassHunter software (Agilent Technologies).

Stereochemical Purity of Enzymatically Produced 3a

GC-MS was performed on an Agilent 6890 gas chromatography system (Agilent Technologies) equipped with an Agilent 5973 mass selective detector (Agilent Technologies). 1 μL samples were injected onto a CycloSil-B chiral capillary column (30 m length, 0.25 mm internal diameter, 0.25 μm film thickness; Agilent Technologies) and GC separation was carried out with the following oven temperature program, following a previously reported [67] method: hold at 50°C for 2 min, ramp up to 90°C at a rate of 1°C/min, then ramp up to a final temperature of 200°C at a rate of 20°C/min. **3a** was detected by monitoring the fragment ions m/z 117, 88, and 45 in SIM mode. ChemStation software (Agilent Technologies) was used for data acquisition and analysis.

B.2 Schematic of Ketoreductase Domain Exchange Junctions in LipPks1+TE
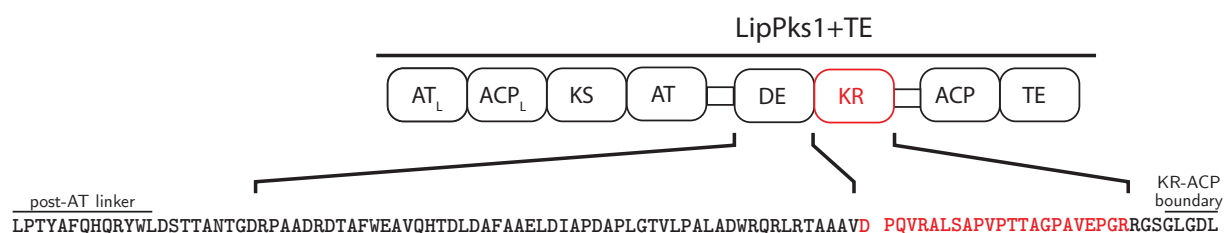
Figure B.1: KR domain exchange junctions selected in LipPks1+TE with the DE underlined, and the KR shown in red. The conserved post-AT linker [65] and the conserved motif that marks the beginning of the ACP [1] are also annotated.

B.3 Sequence Alignment of Donor Ketoreductase Domains

```
                        post-AT linker        start of DE
                        **.*.*:   .    *
LipKR1   LPTYAFQHQRYWLDSTTANTGDRPAAD------------------RDTAFWEAVQHTDLDAFAAELDIAPD---APLGTV   59
AmpKR1   LPTYAFQHERYWNDRAVVDRSASAPMD------------------AEFWAAVEQADVSALTEALGTDEE---S-VAAI    56
AmpKR2   LPTYPFQRQDFWPEPKTAAPAGTADAS------------------DAAFWELVQNQDLDALAGELGIPADDGAARLGDV   61
AveKR1   LPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPS---AGLDTV  77
BorKR1   LPTYPFQHQHYWLDVPPLFTASSAAQD-                                                     27
ConKR2   LPTYPFQRRRYWLEAPRASDSAAPADD------------------GLWREIEAQDSAALADTLGVDPG---A-VGEV    55
EryKR1   LPTYPFQRERVWLEPKPVARRSTEVDE-                                                     27
EryKR6   LPNYPFEPQRYWLAPEVSDQ-                                                            20
PikKR1   LPTYAFQAERYWLENTPAAL-                                                            20
SpnKR3   LPTYPFQRQRYWLDKGAASDEARAVSDP----------------AAGWFWQAVARQDLKSVSDALDLDAD---APLSAT   60
         1.......10.......20.......30.......40.......50.......60.......70.......80

                      end of DE   start of KR
                     ::   :            *              :   *      .
LipKR1   LPALADWRQRLRTAAAVDAWRYRTAFKRLPDAPGAPV--LTGSWLAVVPVRHLDDPSVTTSLDAVAKAGAEVVQLAIEDA  137
AmpKR1   LPGLASWRRTRGERSTVDSWRYRVVWEPLAQIPRAT---LDGTWLLVSAGGVDDD-----VAEVLEAGGAEVRRLVL-DE  127
AmpKR2   LPALSAWRTRSLTRSRADALRYHIEWNRVAEPGTAR---PAGRLLAVISPDHAGAPWVTAVLDAL---GPDTVRFEA--K  133
AveKR1   VPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQ--PHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTL-NH   153
BorKR1   ---------------GGWRYRIHWRRLGTRDSGDR--LSGRWLLLVPESDGTEPWVEGAEKMLAERGCEVVHVPI-AA    87
ConKR2   VPALSAWRSERIRGRRADSWRHRLGWRPLTLAPRAS---LSGRWLVVLPATDSG----ADVLAALTAAGAEPVPVVVSAP  128
EryKR1   ----------------VSALRYRIEWRPTGAGEPAR---LDGTWLVAKYAGTADFT-STAAREALESAGARVRELVV-DA   86
EryKR6   ---------------LADSRYRVWRPLATTPVD-----LEGGFLV-----HGSAP--ESLTSAVEKAGGRVVPVAS---   70
PikKR1   --------------ATGDDWRYRIDWKRLPAAEGSERTGLSGRWLAVTPEDHSAQA--AAVLTALVDAGAKVEVLTA-GA   83
SpnKR3   LPALSVWHRQERERVLADGWRYRVDWVRVAPQPVRR---TRETWLLVVPPGGIEEALVERLTDALNTRGISTLRLDV-PP  136
         ........90.......100.......110.......120.......130.......140.......150.......160

                       :       .  .:.*        .      :      .    ***  *
LipKR1   DADVDR-LTERLRGLVAG----LGAAPAGIMSFLGLDEERHRDHPAMPSGLATSLALVRALGRAGIGAPLWMVTREAVAA  212
AmpKR1   TCTDPAVLREHLTDT--------DGLTGIVSVLADAEEGCVRHPGLTLGLALSVSLAQALGEADVTAPLWFLTCGAFST   198
AmpKR2   GTDRAA-WAAQLAQLRED-----EGEFHAVVSLLAAAEALHTDHGSVPLGLAQTLLLAQALGDAGLTAPLWCLTRGGVAA  207
AveKR1   THTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHNPTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTT  233
BorKR1   TADRDA-MVGAVRESVED-----GRVDGVLSLLALDGRPHPDAAAVPTGLVATAQVVQVSDELGI-GPLWVATRQAVSV  159
ConKR2   DQDRES-LATSLRAASGA------EPVAGVFSLLAEETHPHPAQPDVPAGLSLTLALVQALGDTGTDAPLWCATRALAAV  201
EryKR1   RCGRDE-LAERLRSV---------GEVAGVLSLLAVDEAEPEEAPLALASLADTLSLVQAMVSAELGCPLWTVTESAVAT  156
EryKR6   -ADREA-LAAALREV--------PGEVAGVLS---------------VHTGAATHLALHQSLGEAGVRAPLWLVTSRAVAL  126
PikKR1   DDDREA-LAARLTALTTG------DGFTGVVSLL--------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSV  142
SpnKR3   AATSGE-LATELRAAADG------DPVKAILSLTALDERPHPECKDVPSGIALLLNLVKALGEADLRIPLWTITRGAVKA  209
         ........170.......180.......190.......200.......210.......220.......230.......240

                       *:. .  :* .   .* :*:*      .      :       : *:*:* *
LipKR1   G------------QDTHPHAPLGSLIWGLGQVTALEHADRWGGLIDLPGVCDARVARMLCAGLSGR-GAEDQLALRPSGT  279
AmpKR1   G------------PSDTVTRPLQSQIAGLGWTVAVEHPHRWGGGVDLPEALDARAAQRLVAALAGALGDDDQLAVRPAGV  266
AmpKR2   G------------RGDVLSSPVQGALWGLGRVIGLEHPDRWGGLIDLPETVDTRAAARLTGLLADA-GGEDQLAIRGSGV  274
AveKR1   H------------PNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQP-HHQTQLAIRTTGT  300
BorKR1   DGADEADGAGRTRKADDPADVAQAAVWGLGRVVAALEKPRLWGLVDLPARADERMRDLVAQALTAP-DAEDQLAVRADGI  238
ConKR2   G------------GGERPGAGAQAAVAGLGRTVALERPDTWGGLVDLPAVLDERAADRLGAVLATV-TDEDQVAVRSSGV  268
EryKR1   G------------PFERVRNAAHGALWGLVGRVIADLTWDCVPAGSVAELARHLAAVVSGG-AGEDQLALRADGV       223
EryKR6   G------------ESE-PVDPEQAMVWGLGRVMGLETPERWGGLVDLPAEPAPGDGEAFVACLGAD-GHEDQVAIRDHAR  192
PikKR1   G------------RLDTPADPDRAMLWGLGRVVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGA-TGEDQIAIRTTGL  209
SpnKR3   G------------PADRLLRPMQAQAWGLGRVAALEHPERWGGLIDLPDSLDGDVLTRLGEALTNG-LAEDQLAIRQSGV  276
         ........250.......260.......270.......280.......290.......300.......310.......320

                 **       *:.::***  *   .    ::  :.     ::*  .*  *  .:   *  .*   :*      *     .
LipKR1   FVRRLAHIPGEQRAARRSWQPRGTVIVTGGTGALGAVLARWL-ATEDAEHLVLTGRRGADAPGAERLRDELVATGARVTL  358
AmpKR1   LARRIVRASGDTRRKARSWKPRGTTLVTGGSGTLAPGLARHL-AAQGAEHLVLLSRRGADAPGAAELAAELQAAGTEVRF  345
AmpKR2   LARRLAHAAPAVPGSGKRPPVHGSVVLTGGTGGIGGRVARRL-AEQGAAHLVLTSRRGADAPGAAELRAELEQLGVRVTI  353
AveKR1   HTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTI  380
BorKR1   AVRRLVRSAAS--APADDWQPSGTVLVTGGTGGVGANVARWL-VTDQDIQHLLLVSRRGPDAPGAAELLAELSASGTSVTI  315
ConKR2   YGRRLLPVGPG-RPAEARHSPSGTVLITGGTGGTGAQVAREL-AERGAAHLLLLSRRGPDAPGADELRAELAALGARVTV  346
EryKR1   YGRRWVRAAAP-ATDDEWKPTGTVLVTGGTGGVGQIARWL-ARRGAPHLLLVSRSGPDADGAGELVAELAELAGARTTV   300
EryKR6   YGRRLVRAPLGTRES--SWEPGTALVTGGTGALGGHVARHL-ARCGVEDLVLVSRRGVDAPGAAELEAELVALGAKTTI  269
PikKR1   HARRLARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTI  288
SpnKR3   LARRLVPAPAN-QPAGRKWRPRGSALITGGLGAVGAQVARWL-AEIGAERIVLTSRRGNQAAGAAELEAELRALGAQVSI  354
         ........330.......340.......350.......360.......370.......380.......390.......400
```

```
                .** ::    :   ::         : :*.*:*.       :       .    * . * * :         *::**
LipKR1  AACDVADRKAVAALLDELAADGETVRAVLHAAGVA-DLTSLENTGPEAFAAGVAAKVDGALHLTELLDHDSLDAFVLFSS  437
AmpKR1  AACDITDPDAVAALLADLKAEGRTVRTVVHTAAVI-ELAALADTTVDAFADVVHAKVTGARILDELLDDEELDDFVLYSS  424
AmpKR2  AACDAADREALAALLAELPED-APLTAVFHSAGVAHDDAPVADLTLGQLDALMRAKLTAARHLHELTADLDLDAFVLFSS  432
AveKR1  TTCDTSNPDQLQQLLNTIPPQ-HPLTTVIHTAGIL-DDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSS  458
BorKR1  EPCDVTDADAVRRLIGAVPAE-RPLSTVVHAAGVL-DDCLIDALTPQRLAAALEVKAKGALNLHEAAGEAHL---VLFSS  390
ConKR2  AACDVADRAQLAAVVAAIPEF-EPLTTVVHAAGVL-DDGTLDALDPARLATLMRVKVTAARTLHEVTDGIGLADFVVFSS  424
EryKR1  AACDVTDRESVRELLGGIGDD-VPLSAVFHAAATL-DDGTVDTLTGERIERASRAKVLGARNLHELTRELDLTAFVLFSS  378
EryKR6  TACDVADREQLSKLLEELRGQGRPVRTVVHTAGVP-ESRPLHEIG--ELESVCAAKVTGARLLDELCP--DAETFVLFSS  344
PikKR1  AACDVADPHAMRTLLDAIPAE-TPLTAVVHTAGAL-DDGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSS  366
SpnKR3  VACDVTDRAEMSALLAE-----FDVTAVFHAAGVG-RLLPLAETDQNGLAEICAAKVRGAQVLDELCDSTDLDAFVLFSS  428
        .........410.......420.......430.......440.......450.......460.......470.......480


                .    *        *  .  ** *. *.     .::        ::  ** *                  .  *.  :  .    ::
LipKR1  IAGVWGSGDHGAYAAANAFLNALAEYNRARGIPTTSIAWGVWNAFGVEGAGGISEAVDLDQLHRRGLPLIEPELGLTALR  517
AmpKR1  TAGMWGSGVHAAVVAGNAYLSALAEQRRARGARATSIHWGKWPDDLE------RELADPHQIRRSGLEYLDPELAMTALT  498
AmpKR2  GAAVWGSGGQPGYAAANAYLDALAEHRRSLGLTASSVAWGTWGEVGMA-----TDPEVHDRLVRQGVLAMEPEHALGALD  507
AveKR1  AAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLA-----DSDKARAYLDRRGFRPMSPELATAAVT  533
BorKR1  LAGTTGTKGQGNYAAANAYLDALAERRRADGLPATSVAWGAWQGAGMV-----ADAAVAHRTRRYGLPLMSPDRAVATLR  465
ConKR2  FMGVLGSAGQGNYAAANAALDALVAERRAAGLPGTSVAWGAWAGGGM------VDDRIAERLRRLGITPMEPRAAVRAMT  498
EryKR1  FASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTWAGSGM------AEPAVADRFRRHGVIEMPPETACRALQ  452
EryKR6  GAGVWGSANLGAYSAANAYLDALAHRRRAEGRAATSVAWGAWAGEGM------ATGDL-EGLTRRGLRPMAPERAIRALH  417
PikKR1  VSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPWDGGGMA-----AGDGVAERLRNHGVPGMDPELALAALE  441
SpnKR3  GAGVWGGGGQGAYGAANAFLDTLAEQRRARGLPATSISWGSWAGGGM------ADGAAGEHLRRRGIRPMPAASAILALQ  502
        .......490.......500.......510.......520.......530.......540.......550.......560


                                                              end of KR  |  KR-ACP boundary
                :           :   :  *       .               :
LipKR1  RALDRDET-VLTVAPVAWERFFPLFSAARPRPLFEDLPQVR----ALSAPVPTTAGPAVEPGR-RGSGLGDL  583
AmpKR1  RVMEDDET-VIGLMDIDWGTYHDVFTAGRPSHLFDRIPEVAR----LLADRAAPAATAVATSG-LAARLQGV  564
AmpKR2  QMLENDDT-AAAITLMDWEMFAPAFTANRPSALLSTVPEAVS---ALSDEGPADGGADAAVPP-LRARLDGL  574
AveKR1  QAIADTERPYVVIADIDWSKI-------EHTSQTSDLVS------AAREREPAVQRPTPPAE--LHKTLAHQ  590
BorKR1  QVMAEPVA-TQVVADVDWQRFVADFTAVRPSRLLADLPEVR----SLGEQRKDGPGGQGEEDG-LASKLAAL  531
ConKR2  TALADHDE-LCLVADVDWQRFREA-AGLRSTALLSSLPG------AAGARTPAQTGPAEQDGQ-VRGQLAAL  561
EryKR1  NALDRAEV-CPIVIDVRWDRFLLAYTAQRPTRLFDEIDDARR--------AAPQAAAEP----RVGALASL  510
EryKR6  QALDNGDT-CVSIADVDWEFAVGFTAARPRPLLDELVT-------PAVG--------------AVPAVQAA  467
PikKR1  SALGRDET-AITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGANPLAERLAAA  512
SpnKR3  EVLDQDET-CVSIADVDWDRFVPTFAATRATRLFDEVPAARK-------AMPA-NGPAEPGGSPFARNLAEL  565
        .......570.......580.......590.......600.......610.......620.......630..
```

B.4   Gene Synthesis

The following genes were ordered from Integrated DNA Technologies as gBlocks.

Amp DE_2 (229 b.p):
CACACCAGGTCTCATGGAACCTAAGACCGCCGCTCCGGCGGGCACCGCAGACGCCTCAGACGCGGCGTTTTGGGAA
TTAGTTCAAAATCAGGATCTTGATGCGCTGGCCGGTGAACTGGGTATTCCCGCCGATGATGGCGCTGCCCGGCTGG
GGGACGTACTTCCGGCGCTGTCAGCGTGGCGTACGCGTTCACTCACCCGTTCCCGTGCGGATGTGAGACCTGGTGT
G

Amp KR_2 (1492 b.p):
CACACCAGGTCTCAGATGCTCTAAGATATCATATTGAATGGAACCGCGTCGCGGAGCCGGGAACTGCTCGTCCGGC
GGGTCGTTTGCTGGCAGTGATTTCCCCCGACCATGCGGGCGCGCCTTGGGTCACGGCAGTCTTAGACGCCCTGGGA
CCGGACACGGTGCGTTTCGAAGCTAAGGGCACTGATCGTGCTGCATGGGCAGCCCAACTCGCCCAGCTGCGCGAGG
ATGAAGGGGAATTTCATGCAGTTGTGAGTCTGTTAGCGGCGGCGGAAGCCCTGCACACTGACCACGGCTCAGTCCC
ACTGGGCCTGGCCCAAACCTTACTGCTGGCCCAGGCGCTCGGCGATGCGGGCCTGACCGCGCCGTTATGGTGCCTC
ACCCGCGGGGGCGTTGCTGCAGGACGCGGTGATGTGCTCTCTTCACCGGTTCAGGGCGCTCTGTGGGGTCTGGGTC
GCGTCATTGGCCTGGAACACCCCGACCGCTGGGGTGGCCTGATTGATTTACCTGAAACAGTAGATACCCGGGCCGC
GGCCCGCTTGACGGGGCTGCTTGCGGACGCGGGCGGTGAAGATCAACTGGCGATCCGCGGTTCGGGCGTACTCGCC
CGTCGTCTGGCTCATGCAGCTCCTGCTGTGCCGGGGTCAGGCAAACGTCCGCCGGTGCATGGTAGTGTGCTTGTGA
CCGGAGGGACGGGCGGCATTGGTGGTCGCGTCGCTCGCCGCCTCGCGGAACAGGGAGCGGCACACCTGGTACTGAC
TTCACGTCGTGGTGCCGATGCCCCGGGCGCAGCCGAGCTGCGCGCTGAACTGGAACAACTCGGTGTACGCGTCACC
ATTGCCGCGTGTGACGCGGCGGATCGTGAAGCCCTCGCCGCACTGTTAGCGGAACTGCCGGAGGACGCTCCCCTGA
CGGCTGTATTCCACAGCGCGGGTGTTGCCCACGATGATGCCCCGGTAGCCGACCTGACCCTGGGTCAGCTGGATGC
GCTGATGCGTGCCAAGCTGACAGCAGCTCGTCATTTACATGAGCTCACCGCGGATCTTGACTTAGACGCCTTTGTA
CTTTTCTCATCAGGGGCGGCGGTATGGGGTTCCGGTGGTCAGCCGGGATATGCGGCCGCGAACGCGTATTTGGATG
CACTGGCCGAGCACCGCCGCAGTCTGGGTTTGACCGCCAGCAGCGTAGCATGGGGTACCTGGGGCGAAGTAGGCAT
GGCTACCGACCCTGAAGTTCATGACCGTCTCGTTCGTCAGGGCGTTCTGGCGATGGAGCCAGAACATGCACTGGGG
GCGCTTGACCAGATGCTTGAAAACGACGACACTGCCGCCGCGATTACGCTCATGGATTGGGAGATGTTTGCACCGG
CCTTTACAGCGAATCGGCCAAGCGCATTACTGAGCACGGTTCCCGAGGCGGTTAGCGCATTAAGCGATGAGGGCCC
AGCTGATGGAGGTGCAGATGCAGCTGTACCACCTTGAGACCTGGTGTG

Con DE_2 (211 b.p):
CACACCAGGTCTCATGGAAGCACCGCGGGCGTCGGATAGCGCGGCTCCGGCGGATGACGGGTTGTGGCGCGAAATT
GAAGCGCAGGATAGCGCTGCTTTGGCCGACACCCTTGGTGTTGACCCTGGCGCCGTAGGTGAGGTTGTGCCGGCTT
TGTCTGCTTGGCGGTCAGAACGTATCCGTGGCCGTCGCGCGGATATGAGACCTGGTGTG

Con KR_2 (1471 b.p):
CACACCAGGTCTCAGATAGTTGGAGACATAGACTAGGATGGCGTCCGCTGACCCTGGCACCGCGCGCTAGTTTATC
AGGGCGTTGGCTGGTTGTATTGCCGGCTACCGACTCTGGGGCCGATGTTCTGGCCGCGCTGACGGCAGCCGGTGCA
GAACCAGTACCAGTAGTTGTGAGCGCACCGGATCAGGATCGTGAATCTCTTGCAACATCTCTTAGAGCTGCATCTG
GTGCAGAACCTGTTGCAGGTGTATTCAGCTTGCTGGCTGAGGAGACGCACCCGCATCCTGCGCAGCCGGACGTACC
GGCGGGCTTATCGCTTACCTTGGCGCTGGTGCAGGCCCTGGGGGATACGGGTACGGATGCACCTCTGTGGTGCGCT

```
ACCCGCGCACTGGCAGCAGTTGGTGGCGGAGAACGTCCTGGTGCTGGCGCACAGGCAGCTGTCGCTGGCCTCGGTC
GCACTGTGGCACTTGAAAGACCGGACACGTGGGGAGGCTTGGTCGACTTGCCAGCGGTTCTGGATGAACGTGCTGC
AGATCGTCTCGGTGCGGTATTGGCAACCGTTACCGATGAAGACCAGGTGGCTGTACGGTCTTCCGGCGTATACGGT
CGTCGCCTTCTGCCTGTAGGTCCTGGAAGACCTGCAGAAGCTAGACATTCTCCATCAGGAACTGTTTTAATCACGG
GTGGGACCGGCGGCACGGGTGCGCAGGTGGCCCGCGAGTTAGCGGAGCGCGGTGCTGCACATCTTCTACTGTTAAG
TAGACGAGGTCCTGATGCTCCAGGAGCAGATGAACTGCGTGCTGAGTTAGCAGCTTTAGGTGCCCGCGTTACCGTA
GCAGCTTGTGATGTGGCAGATCGCGCACAGTTGGCGGCCGTCGTGGCTGCAATTCCAGAAGAAGAACCGCTGACAA
CGGTGGTGCACGCAGCTGGTGTACTGGACGACGGGACTCTGGATGCGCTTGACCCTGCCCGTCTGGCGACGCTGAT
GCGTGTTAAAGTGACCGCGGCCCGGACTTTGCACGAGGTTACCGATGGTATTGGTCTGGCAGATTTCGTAGTATTC
AGCAGCTTCATGGGCGTCTTGGGGAGCGCCGGACAGGGTAATTACGCCGCAGCAAATGCGGCGTTAGACGCTTTAG
TCGCAGAGCGTCGCGCGGCAGGCCTGCCGGGCACTTCCGTGGCGTGGGGTGCGTGGGCCGGTGGCGGTATGGTTGA
CGATCGCATCGCGGAGCGCTTACGGCGCTTGGGCATTACGCCTATGGAGCCGCGTGCGGCGGTGCGTGCAATGACT
ACCGCGCTGGCCGATCATGATGAACTCTGTTTAGTTGCCGACGTGGATTGGCAACGCTTTCGTGAAGCTGCGGGCC
TTCGCAGTACTGCGTTGCTGTCATCGTTGCCAGGTGCTGCAGGAGCTAGAACTCCTGCTCAAACAGGTCCAGCTGA
ACAAGATGGTCAGTGAGACCTGGTGTG
```

The following genes were ordered from Genewiz and delivered in pUC57-Kan flanked by a 5′-GCTAGCT sequence containing a BmtI cut site and a 3′-GGTACCGC sequence containing a Acc65I cut site.

Amp KR$_1$ (1449 b.p.):
```
GATAGCTGGCGTTATCGTGTGGTGTGGGAACCTCTGGCCCAGATTCCGCGCGCAACCCTGGATGGCACCTGGTTAC
TGGTTAGCGCAGGCGGTGTGGATGACGATGTTGCCGAAGTTCTGGAAGCAGGTGGCGCCGAAGTGCGTCGCCTGGT
TCTGGATGAAACATGCACCGATCCGGCCGTGCTGCGTGAACACCTGACCGATACAGATGGCTTAACCGGCATCGTT
AGCGTGCTGGCAGATGCAGAAGAAGGTTGCGTTCGCCATCCTGGTCTGACCCTGGGCCTGGCACTGAGTGTGAGCC
TGGCACAGGCCTTAGGTGAGGCAGATGTGACCGCCCCGCTGTGGTTTCTGACCTGTGGTGCCTTCAGCACCGGCCC
GAGCGATACAGTGACACGCCCGTTACAGAGCCAGATTGCAGGCCTGGGCTGGACCGTGGCCGTTGAACATCCGCAT
CGTTGGGGTGGTGGTGTTGACCTGCCTGAAGCATTAGATGCCCGCGCCGCACAGCGCTTAGTTGCAGCACTGGCCG
GTGCCCTGGGCGATGATGATCAGCTGGCAGTTCGTCCCGCTGGTGTTCTGGCCCGTCGTATTGTTCGCGCAAGTGG
CGATACCCGCCGCAAAGCACGTAGCTGGAAACCGCGCGGCACTACACTGGTTACCGGCGGCAGTGGTACACTGGCA
CCGGGTCTGGCCCGCCATTTAGCCGCACAGGGTGCAGAACATCTGGTGTTACTGAGCCGTCGTGGTGCCGATGCCC
CTGGCGCAGCAGAACTGGCAGCAGAGTTACAAGCAGCCGGCACTGAAGTTCGTTTTGCCGCCTGCGATATTACCGA
CCCGGATGCCGTGGCAGCCTTACTGGCCGATCTGAAGGCCGAAGGCCGTACCGTTCGTACCGTGGTGCATACCGCC
GCCGTTATCGAACTGGCCGCACTGGCCGATACAACCGTGGACGCATTCGCCGATGTGGTTCATGCCAAGGTTACCG
GTGCACGCATTCTGGACGAGCTGCTGGACGATGAAGAACTGGATGATTTTGTTCTGTATAGTAGCACCGCAGGTAT
GTGGGGCAGTGGCGTGCATGCAGCCTACGTTGCCGGCAACGCCTATCTGAGCGCACTGGCAGAACAACGCCGCGCA
CGTGGTGCACGCGCCACCAGCATTCACTGGGGTAAATGGCCGGATGACCTGGAACGTGAACTGGCAGATCCTCACC
AGATCCGTCGTAGCGGCCTGGAATATTTAGATCCTGAATTAGCTATGACCGCCCTGACACGTGTGATGGAGGATGA
TGAAACTGTGATCGGTCTGATGGACATTGACTGGGGCACATACCATGACGTTTTTACCGCAGGCCGTCCGAGCCAT
CTGTTCGATCGCATTCCGGAGGTGGCACGTCTGTTAGCCGACCGTGCCGCACCGGCAGCAACCGCAGTTGCAACCA
GTGGT
```

Pik KR$_1$ (1443 b.p.):

```
GATGACTGGCGCTATCGCATTGACTGGAAACGTCTGCCTGCCGCCGAAGGCAGTGAACGTACCGGTCTGAGCGGTC
GCTGGTTAGCCGTTACCCCTGAAGATCACAGTGCCCAGGCAGCCGCCGTGTTAACAGCCCTGGTGGATGCCGGTGC
AAAAGTTGAGGTGCTGACCGCAGGCGCCGATGATGATCGCGAAGCATTAGCCGCCCGTCTGACCGCACTGACAACC
GGTGACGGTTTTACCGGCGTTGTGAGCCTGCTGGATGGCCTGGTGCCGCAGGTTGCCTGGGTTCAGGCACTGGGCG
ATGCAGGTATTAAGGCACCGCTGTGGAGCGTTACCCAGGGTGCCGTGAGTGTTGGTCGTCTGGATACCCCGGCAGA
CCCGGATCGTGCAATGTTATGGGGCCTGGGCCGCGTTGTTGCACTGGAGCATCCGGAACGTTGGGCAGGTCTGGTT
GATCTGCCGGCCCAGCCTGATGCAGCAGCCCTGGCACATCTGGTGACCGCTTTATCTGGCGCAACCGGCGAAGATC
AGATTGCAATCCGCACCACCGGTCTGCATGCCCGTCGTCTGGCACGTGCACCGTTACACGGTCGCCGCCCGACACG
TGATTGGCAGCCTCATGGCACTGTTCTGATCACCGGCGGTACAGGCGCCTTAGGTAGCCATGCCGCACGTTGGATG
GCCCATCATGGTGCCGAACATCTGCTGCTGGTGAGCCGTAGTGGCGAGCAAGCACCTGGCGCCACACAGCTGACAG
CCGAACTGACAGCCAGTGGTGCCCGCGTTACCATCGCAGCCTGCGATGTGGCAGATCCGCATGCAATGCGCACCCT
GCTGGACGCCATTCCGGCAGAAACCCCGCTGACAGCCGTGGTGCATACCGCAGGTGCCCTGGATGATGGCATTGTG
GACACCCTGACCGCAGAACAGGTGCGCCGTGCCCATCGTGCAAAAGCCGTGGGTGCCAGTGTGTTAGATGAGCTGA
CCCGCGATCTGGACCTGGATGCCTTCGTGCTGTTTAGCAGCGTGAGTAGTACCCTGGGTATTCCGGGTCAGGGTAA
TTACGCCCCGCACAACGCCTATCTGGACGCACTGGCAGCACGTCGTCGCGCAACCGGTCGTAGCGCAGTGAGTGTG
GCATGGGGTCCGTGGGATGGTGGTGGTATGGCAGCCGGCGATGGTGTTGCCGAACGCCTGCGTAACCATGGCGTTC
CGGGTATGGACCCTGAACTGGCCCTGGCCGCACTGGAAAGCGCCTTAGGCCGCGACGAAACCGCCATCACCGTGGC
CGATATCGATTGGGACCGTTTCTACCTGGCATACAGCAGTGGCCGCCCTCAACCGCTGGTTGAAGAGCTGCCGGAA
GTTCGCCGCATTATTGATGCACGCGACAGTGCCACCAGCGGTCAAGGCGGTAGTAGCGCACAGGGCGCCAATCCG
```

Spn KR$_3$ (1440 b.p.):
```
GACGGTTGGCGCTATCGTGTTGATTGGGTGCGTGTTGCCCCTCAGCCGGTTCGTCGCACACGTGAAACCTGGCTGC
TGGTTGTTCCGCCGGGTGGTATTGAAGAAGCCCTGGTGGAACGCCTGACCGATGCCCTGAATACCCGTGGCATTAG
CACCCTGCGCCTGGATGTTCCGCCGGCAGCAACAAGTGGCGAATTAGCCACCGAACTGCGTGCCGCCGCAGATGGT
GATCCGGTTAAGGCCATTCTGAGCCTGACCGCACTGGATGAACGTCCGCATCCGGAATGTAAGGATGTGCCTAGCG
GCATTGCCCTGTTACTGAATCTGGTGAAGGCCCTGGGTGAGGCCGATCTGCGCATCCCGCTGTGGACAATTACTCG
TGGTGCAGTGAAAGCCGGTCCTGCCGATCGTCTGCTGCGTCCTATGCAGGCCCAGGCATGGGGTTTAGGCCGTGTT
GCAGCCCTGGAACATCCGGAACGTTGGGGTGGCCTGATTGATCTGCCGGACAGCCTGGATGGCGACGTGCTGACCC
GTCTGGGTGAAGCCCTGACCAATGGTCTGGCAGAAGATCAGCTGGCAATCCGTCAAAGCGGTGTGCTGGCACGCCG
TTTAGTTCCTGCCCCGGCAAATCAGCCGGCAGGCCGTAAATGGCGCCCTCGCGGTAGCGCACTGATTACCGGCGGT
CTGGGTGCAGTGGGTGCACAGGTTGCACGTTGGCTGGCCGAAATCGGCGCAGAACGCATTGTGCTGACCAGCCGTC
GCGGCAATCAAGCAGCCGGTGCAGCCGAACTGGAAGCCGAGCTGCGTGCACTGGGCGCACAGGTTAGCATTGTGGC
CTGCGATGTGACCGATCGTGCCGAAATGAGTGCCCTGCTGGCAGAGTTTGATGTGACAGCCGTGTTCCATGCAGCA
GGCGTGGGTCGCCTGCTGCCGTTAGCAGAAACCGACCAGAATGGCCTGGCAGAAATTTGTGCCGCCAAAGTGCGCG
GCGCACAAGTTCTGGACGAACTGTGCGACAGCACCGATCTGGACGCCTTTGTGCTGTTTAGTAGCGGTGCAGGCGT
TTGGGGTGGTGGCGGTCAGGGCGCTTATGGTGCAGCAAATGCATTCCTGGATACCCTGGCAGAGCAGCGTCGTGCA
CGCGGTTTACCGGCAACCAGCATTAGCTGGGGTAGCTGGGCAGGTGGTGGTATGGCAGATGGCGCCGCAGGTGAAC
ATCTGCGTCGTCGCGGTATCCGCCCTATGCCGGCAGCCAGTGCCATTCTGGCCTTACAAGAAGTGCTGGACCAGGA
TGAAACTTGTGTTAGCATTGCCGACGTTGATTGGGACCGCTTTGTTCCGACATTTGCAGCAACCCGTGCAACCCGT
CTGTTCGATGAGGTGCCGGCAGCACGCAAAGCAATGCCTGCAAATGGTCCGGCAGAACCGGGTGGTAGCCCG
```

The following genes were ordered from Genscript.

Ave KR$_1$ (1464 b.p.):
AATACCTGGACGTACCAGGAAACGTGGAAACCGCTGACGCTGCCGACCACCCATCAACCGCACCAGACCTGGTTGA
TCGCCATCCCGGAAACGCAAACCCACCATCCACATATTACCAACATCCTGACGAACTTGCACCATCACGGCATCAC
CCCAATCCCTCTGACCCTGAACCATACGCACACCAATCCTCAGCATCTGCATCATACCCTGCACCACACCCGTCAG
CAGGCACAAAACCATACCACGGGTGCAATTACCGGTCTGCTGAGCCTGTTGGCACTGGATGAAACGCCGCACCCGC
ACCACCCGCATACCCCTACGGGCACGCTGCTGAATTTGACGCTGACCCAGACCCACACGCAAACGCATCCGCCAAC
TCCGCTGTGGTATGCTACGACGAACGCAACGACGACGCATCCTAACGATCCATTGACCCACCCGACCCAGGCGCAG
ACCTGGGGCCTGGCGCGTACGACCTTGCTGGAGCACCCTACCCATACCGCGGGTATCATTGATCTGCCGACCACGC
CGACGCCTCATACCTTGCAGCATTTGACGCAGACGCTGACCCAGCCTCACCATCAGACGCAGCTGGCGATCCGTAC
TACCGGCACTCACACGCGCCGTCTGACCCCGACGACTCTGACGCCGACCCACCAGCCGCCGACCCCGACCCCGCAT
GGTACGACGCTGATCACGGGTGGCACGGGTGCGCTGGCGACGCACCTGACGCATCACCTGACGACCCACCAACCTA
CGCAGCACCTGCTGCTGACCAGCCGTACTGGTCCTCACACCCCGCACGCGCAGCATCTGACTACCCAGCTGCAACA
AAAGGGCATCCACCTGACCATTACCACGTGTGATACCTCTAATCCGGACCAGTTGCAACAGCTGCTGAACACGATT
CCGCCGCAGCACCCGCTGACTACCGTGATTCACACTGCTGGTATTCTGGACGACGCGACCCTGACCAACCTGACCC
CAACCCAGCTGAATAACGTCCTGCGTGCGAAAGCGCACTCTGCCCATCTGCTGCACCAGCTGACCCAACACACGCC
GCTGACTGCGTTTGTGCTGTACAGCTCTGCCGCTGCGACCTTCGGTGCTCCGGGCCAGGCCAATTACGCCGCAGCG
AACGCGTATCTGGATGCTCTGGCCCACCACCGTCACACGCATCACTTGCCGGCGACCTCTATTGCTTGGGGCACGT
GGCAGGGTAATGGCCTGGCGGACAGCGATAAAGCTCGTGCGTACCTGGACCGTCGCGGTTTCCGCCCGATGTCCCC
AGAACTGGCAACCGCTGCAGTTACTCAGGCGATCGCGGACACGGAGCGTCCGTACGTTGTTATTGCAGATATCGAT
TGGAGCAAGATTGAGCACACGAGCCAGACGAGCGATCTGGTTAGCGCAGCCCGTGAACGTGAGCCGGCAGTGCAAC
GTCCTACCCCGCCTGCCGAG

Ery KR$_1$ (1422 b.p.):
TCTGCACTGCGTTATCGCATTGAATGGCGTCCGACCGGTGCTGGTGAACCGGCACGTCTGGATGGTACGTGGCTGG
TTGCGAAATACGCAGGCACCGCTGACGAAACGTCAACCGCAGCTCGCGAAGCGCTGGAATCGGCAGGTGCTCGTGT
TCGCGAACTGGTTGTCGATGCACGTTGCGGTCGTGACGAACTGGCAGAACGTCTGCGCAGCGTCGGTGAAGTGGCG
GGCGTTCTGTCTCTGCTGGCCGTGGATGAAGCAGAACCGGAAGAAGCGCCGCTGGCGCTGGCCAGTCTGGCCGATA
CCCTGTCCCTGGTGCAGGCTATGGTTAGCGCGGAACTGGGTTGTCCGCTGTGGACGGTGACCGAATCTGCAGTTGC
TACCGGTCCGTTTGAACGTGTTCGTAACGCAGCACATGGCGCGCTGTGGGGTGTCGGCCGCGTGATTGCCCTGGAA
AATCCGGCAGTCTGGGGCGGTCTGGTCGATGTGCCGGCAGGTAGTGTTGCAGAACTGGCCCGTCACCTGGCAGCTG
TGGTTTCCGGCGGTGCAGGTGAAGATCAGCTGGCACTGCGTGCTGACGGTGTGTACGGTCGTCGCTGGGTTCGTGC
AGCAGCACCGGCAACCGATGACGAATGGAAACCGACGGGCACCGTTCTGGTCACCGGCGGCACCGGCGGCGTGGGC
GGCCAAATCGCACGTTGGCTGGCACGTCGCGGTGCTCCGCATCTGCTGCTGGTTAGCCGTTCTGGTCCGGATGCAG
ACGGTGCAGGCGAACTGGTTGCGGAACTGGAAGCACTGGGTGCTCGCACCACGGTCGCTGCATGCGATGTGACCGA
CCGTGAATCAGTGCGTGAACTGCTGGGCGGTATTGGCGATGACGTTCCGCTGTCGGCCGTCTTTCATGCCGCAGCT
ACGCTGGATGACGGCACCGTGGATACGCTGACCGGCGAACGTATCGAACGCGCGTCACGTGCCAAAGTTCTGGGTG
CGCGCAACCTGCACGAACTGACGCGTGAACTGGATCTGACCGCATTTGTGCTGTTCTCATCGTTTGCGAGTGCATT
CGGTGCGCCGGGTCTGGGCGGTTATGCGCCGGGTAATGCCTACCTGGATGGCCTGGCGCAGCAACGTCGCAGTGAC
GGTCTGCCGGCAACGGCTGTTGCATGGGGCACCTGGGCAGGTTCCGGTATGGCAGAAGGTCCGGTTGCAGATCGTT
TTCGTCGCCATGGCGTCATTGAAATGCCGCCGGAAACCGCATGCCGCGCTCTGCAGAACGCACTGGATCGTGCTGA
AGTGTGTCCGATTGTTATCGATGTCCGCTGGGACCGTTTCCTGCTGGCGTATACGGCACAGCGTCCGACCCGTCTG
TTTGATGAAATCGATGACGCACGTCGCGCAGCACCGCAAGCAGCTGCGGAACCG

Ery KR$_6$ (1311 b.p.):
GACTCCCGCTATCGTGTTGATTGGCGCCCGCTGGCCACCACGCCGGTGGATCTGGAAGGCGGTTTTCTGGTGCATG
GCTCAGCACCGGAAAGCCTGACCTCTGCGGTTGAAAAAGCCGGCGGTCGTGTGGTTCCGGTTGCTTCGGCGGATCG
TGAAGCACTGGCAGCTGCGCTGCGTGAAGTTCCGGGCGAAGTCGCGGGTGTGCTGAGTGTTCACACCGGTGCGGCA
ACCCACCTGGCACTGCACCAGTCCCTGGGTGAAGCAGGTGTTCGTGCTCCGCTGTGGCTGGTCACGAGCCGTGCGG
TGGCACTGGGCGAATCTGAACCGGTTGATCCGGAACAGGCCATGGTCTGGGGCCTGGGTCGTGTGATGGGTCTGGA
AACCCCGGAACGTTGGGGCGGTCTGGTGGATCTGCCGGCTGAACCGGCCCCGGGTGATGGTGAAGCATTTGTCGCA
TGCCTGGGCGCGGATGGTCATGAAGACCAGGTGGCAATCCGCGATCACGCTCGTTATGGCCGTCGCCTGGTTCGTG
CACCGCTGGGTACGCGTGAAAGCAGCTGGGAACCGGCTGGCACCGCCCTGGTGACCGGCGGTACGGGTGCCCTGGG
CGGTCATGTGGCGCGCCACCTGGCCCGTTGCGGCGTTGAAGATCTGGTGCTGGTTTCACGTCGCGGCGTGGACGCC
CCGGGTGCTGCGGAACTGGAAGCCGAACTGGTTGCCCTGGGTGCAAAAACCACGATTACCGCATGTGATGTGGCTG
ACCGCGAACAGCTGTCTAAACTGCTGGAAGAACTGCGCGGTCAGGGTCGCCCGGTTCGCACCGTTGTGCATACGGC
GGGCGTGCCGGAAAGCCGCCCGCTGCACGAAATCGGTGAACTGGAAAGCGTGTGCGCCGCAAAAGTCACGGGCGCA
CGTCTGCTGGATGAACTGTGTCCGGACGCTGAAACCTTTGTTCTGTTCAGTTCCGGTGCGGGTGTCTGGGGCTCAG
CCAACCTGGGTGCCTATTCGGCTGCGAATGCCTACCTGGATGCTCTGGCGCATCGTCGCCGTGCCGAAGGTCGTGC
CGCAACCAGTGTGGCATGGGGTGCGTGGGCAGGTGAGGGTATGGCCACCGGTGATCTGGAAGGTCTGACCCGCCGT
GGCCTGCGTCCGATGGCCCCGGAACGCGCTATTCGTGCGCTGCACCAGGCCCTGGATAACGGCGACACCTGTGTGA
GCATCGCAGATGTTGACTGGGAACGCTTTGCTGTTGGTTTTACGGCTGCGCGCCCGCGTCCGCTGCTGGATGAACT
GGTGACCCCGGCCGTCGGT

B.5 Primers

Table B.1: Primers used to probe the importance of the DE in KR domain exchange experiments.

Primer	Sequence	Target
debb_1f	CACACCAGGTCTCAACCTCGTGGTAGCGGCCTGGGC	pSY044
nodebb_2f	CACACCAGGTCTCATCAGCGTGGTAGCGGCCTGGGC	pSY044
backbone_r	CACACCAGGTCTCATCCAGCCAGTAACGCTGATGCTGAAATGCG	pSY044
lipde1_f	CACACCAGGTCTCATGGATAGCACCAGCAAACACTGG	pSY044
lipde1_r	CACACCAGGTCTCACATCAACAGCCGCTGCCGTACGC	pSY044
ampde2_f	CACACCAGGTCTCATGGAACC	Amp DE_2
ampde2_r	CACACCAGGTCTCACATCCGC	Amp DE_2
ampkr2_f	CACACCAGGTCTCAGATGC	Amp KR_2
ampkr2_r	CACACCAGGTCTCAAGGTGG	Amp KR_2
conde2_f	CACACCAGGTCTCATGGAAGCACCG	Con DE_2
conde2_r	CACACCAGGTCTCATATCCGCGCG	Con DE_2
conkr2_f	CACACCAGGTCTCAGATAGTTGGAGACA	Con KR_2
conkr2_r	CACACCAGGTCTCACTGACCATCTTGT	Con KR_2
dedel_f	GACGCATGGCGTTACCGCAC	pSY044
dedel_r	GTCACCAGTGTTTGCAGTGGTGCTATC	pSY044

Table B.2: Primers used to investigate the generalizability of the KR domain exchange strategy.

Primer	Sequence	Target
debb_f	CACACCAGGTCTCACGTGGTAGCGGCCTGGGC	pSY044
debb_r	CACACCAGGTCTCAAACAGCCGCTGCCGTACGC	pSY044
avekr1_f	CACACCAGGTCTCAGGCTAATACCTGGACGTACCAGGAAACGTGGA	Ave KR_1
avekr1_r	CACACCAGGTCTCACTGCCTCGGCAGGCGGGGTAGGAC	Ave KR_1
erykr1_f	CACACCAGGTCTCATGTTTCTGCACTGCGTTATCGCATTGAATGG	Ery KR_1
erykr1_r	CACACCAGGTCTCACACGCGGTTCCGCAGCTGCTTGC	Ery KR_1
borkr1_f	CACACCAGGTCTCATGTTGGCGGTTGGCGTTATCGTATTCACTGG	pARH100
borkr1_r	CACACCAGGTCTCACACGACCGTCCTCTTCACCTTGGCCG	pARH100
spnkr3_f	CACACCAGGTCTCATGTTGACGGTTGGCGCTATCGTGTTGATTGG	Spn KR_3
spnkr3_r	CACACCAGGTCTCACACGCGGGCTACCACCCGGTTCTGC	Spn KR_3
ampkr1_f	CACACCAGGTCTCATGTTGATAGCTGGCGTTATCGTGTGGTGTGG	Amp KR_1
ampkr1_r	CACACCAGGTCTCACACGACCACTGGTTGCAACTGCGG	Amp KR_1
pikkr1_f	CACACCAGGTCTCATGTTGATGACTGGCGCTATCGCATTGACTGG	Pik KR_1
pikkr1_r	CACACCAGGTCTCACACGCGGATTGGCGCCCTGTGCG	Pik KR_1
erykr6_f	CACACCAGGTCTCATGTTGACTCCCGCTATCGTGTTGATTGGCG	Ery KR_6
erykr6_r	CACACCAGGTCTCACACGACCGACGGCCGGGGTCACC	Ery KR_6

B.6 Plasmids

Table B.3: Expression plasmids used in this study.

Plasmid	Description	Parent	Antibiotic resistance
pSY044	Lip KR_1	pET30b	Kanamycin
pCE001	Amp KR_2	pSY044	Kanamycin
pCE002^a	Amp DE,KR_2	pSY044	Kanamycin
pCE003	Con KR_2	pSY044	Kanamycin
pCE004^a	Con DE,KR_2	pSY044	Kanamycin
pCE006	Ave KR_1	pSY044	Kanamycin
pCE007	Ery KR_1	pSY044	Kanamycin
pCE008	Bor KR_1	pSY044	Kanamycin
pCE009	Spn KR_3	pSY044	Kanamycin
pCE010	Amp KR_1	pSY044	Kanamycin
pCE012	Pik KR_1	pSY044	Kanamycin
pCE015	Ery KR_6	pSY044	Kanamycin
pCE022	No DE (RPA to AAV), Lip KR_1	pSY044	Kanamycin

^{a} DE from the module furnishing the donor KR domain substituted for that of LipPks1.
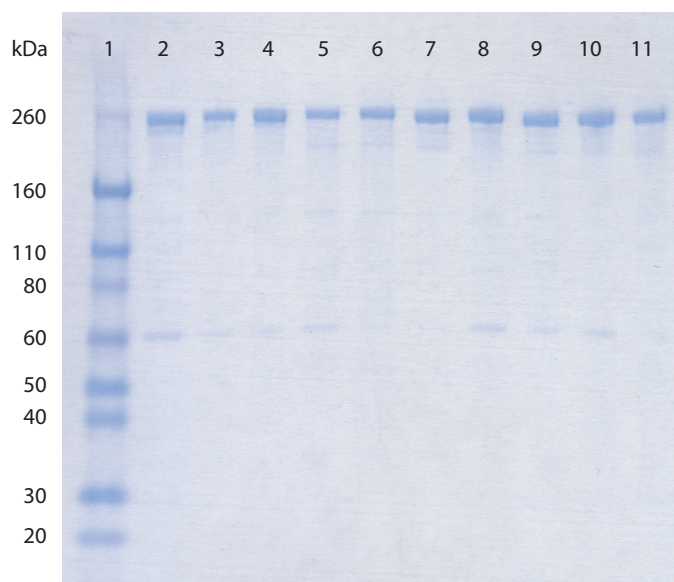
B.7 Purity of LipPks1+TE Variants

Figure B.2: SDS-PAGE analysis of recombinant LipPks1+TE (259 kDa) and variants purified to >90% purity. Proteins were resolved using a 4-15% Tris-Glycine gel and stained with Coomassie blue. Lane 1, molecular weight marker; Lanes 2-11 (in order), LipPks1+TE, (Amp KR$_2$)LipPks1+TE, (Amp DE,KR$_2$)LipPks1+TE, (Con KR$_2$)LipPks1+TE, (Con DE,KR$_2$)LipPks1+TE, (Ery KR$_1$)LipPks1+TE, (Bor KR$_1$)LipPks1+TE, (Spn KR$_3$)LipPks1+TE, (Amp KR$_1$)LipPks1+TE, (Ery KR$_6$)LipPks1+TE

.

B.8   Production of 3a-d by LipPks1+TE Variants
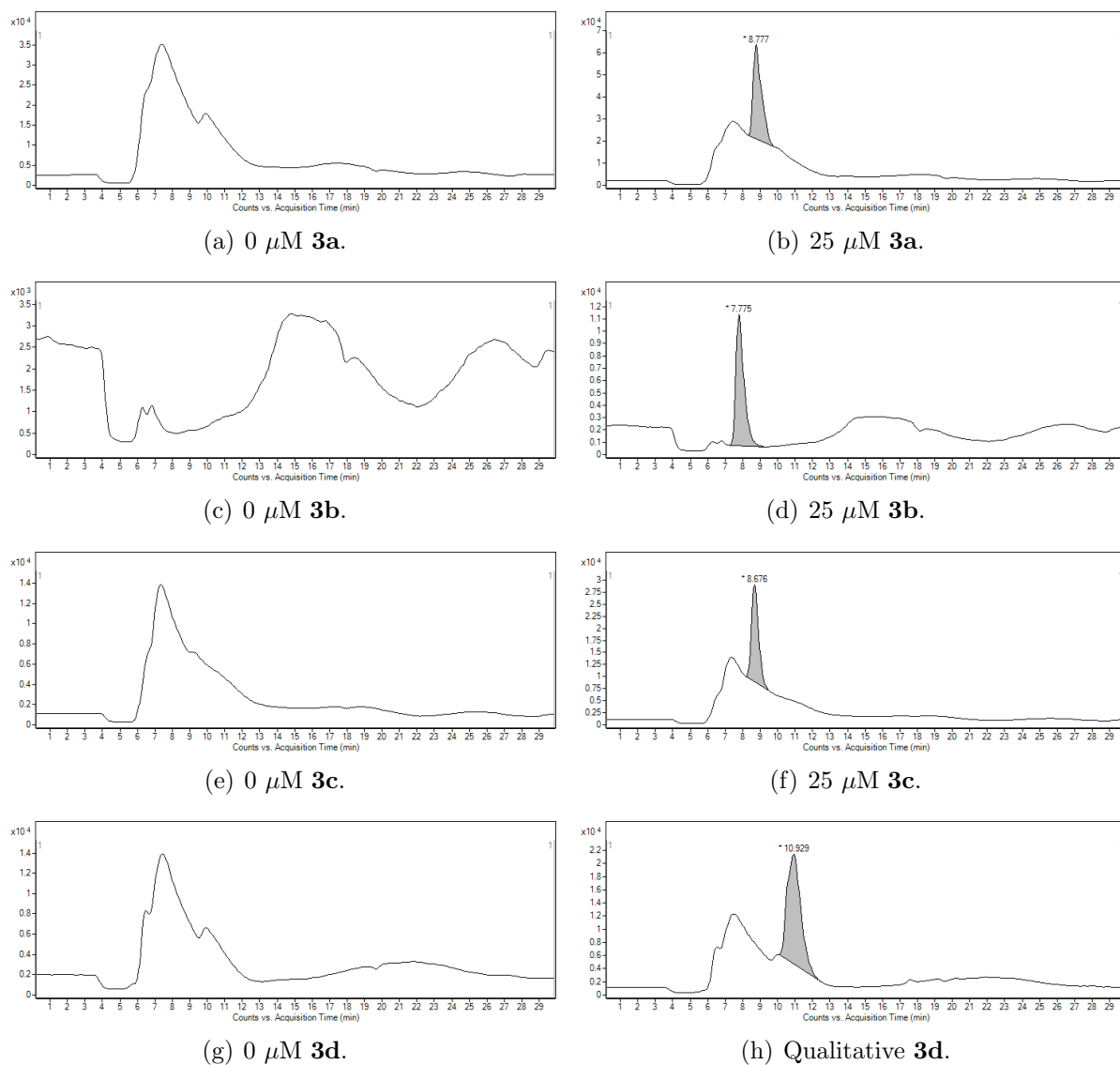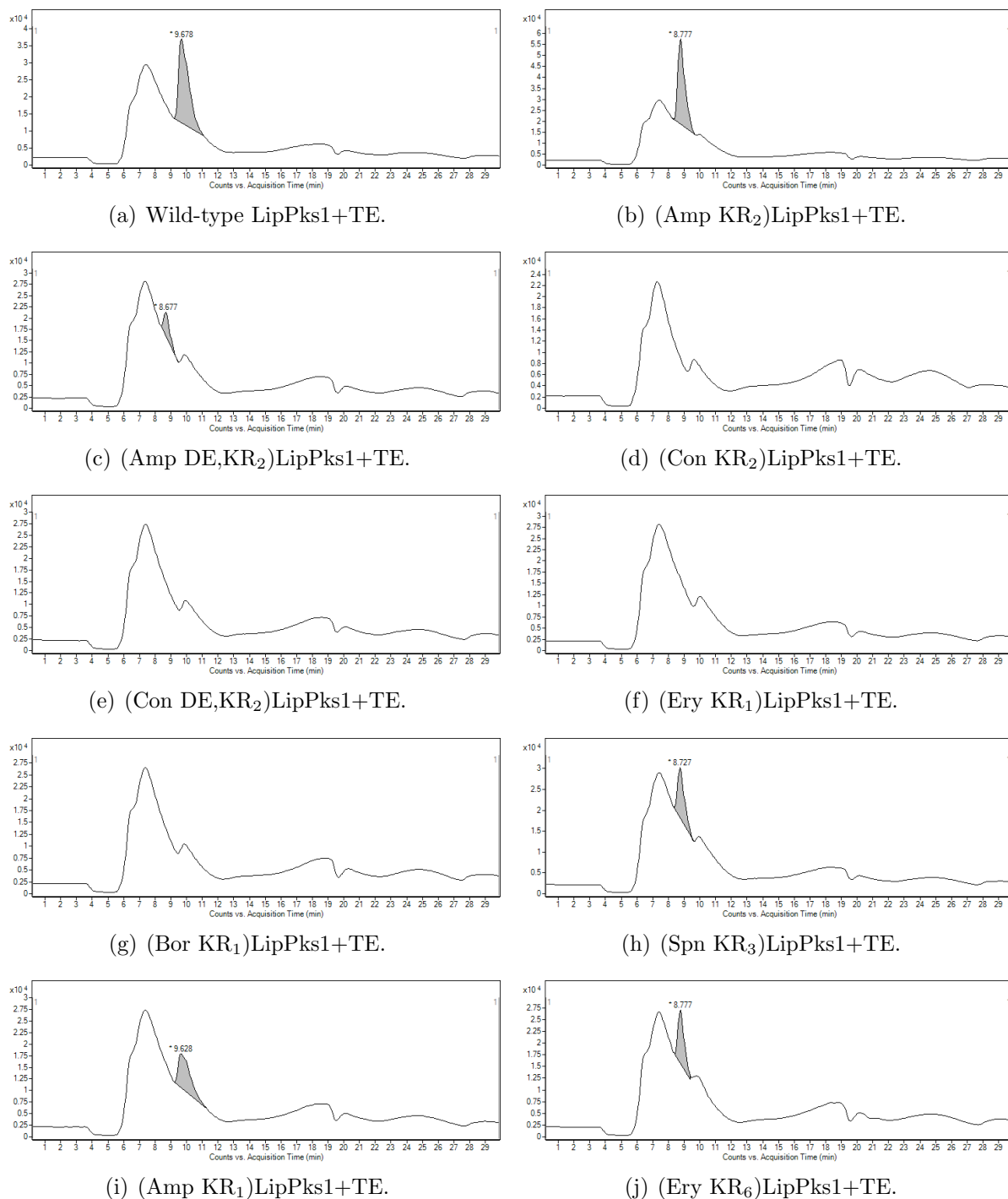

(a) 0 μM **3a**.

(b) 25 μM **3a**.

(c) 0 μM **3b**.

(d) 25 μM **3b**.

(e) 0 μM **3c**.

(f) 25 μM **3c**.

(g) 0 μM **3d**.

(h) Qualitative **3d**.

Figure B.3: Authentic hydroxyacid standards **3a-d**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR$_2$)LipPks1+TE.

(c) (Amp DE,KR$_2$)LipPks1+TE.

(d) (Con KR$_2$)LipPks1+TE.

(e) (Con DE,KR$_2$)LipPks1+TE.

(f) (Ery KR$_1$)LipPks1+TE.

(g) (Bor KR$_1$)LipPks1+TE.

(h) (Spn KR$_3$)LipPks1+TE.

(i) (Amp KR$_1$)LipPks1+TE.

(j) (Ery KR$_6$)LipPks1+TE.

Figure B.4: Overnight production of **3a**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR$_2$)LipPks1+TE.

(c) (Amp DE,KR$_2$)LipPks1+TE.

(d) (Con KR$_2$)LipPks1+TE.

(e) (Con DE,KR$_2$)LipPks1+TE.

(f) (Ery KR$_1$)LipPks1+TE.

(g) (Bor KR$_1$)LipPks1+TE.

(h) (Spn KR$_3$)LipPks1+TE.

(i) (Amp KR$_1$)LipPks1+TE.

(j) (Ery KR$_6$)LipPks1+TE.
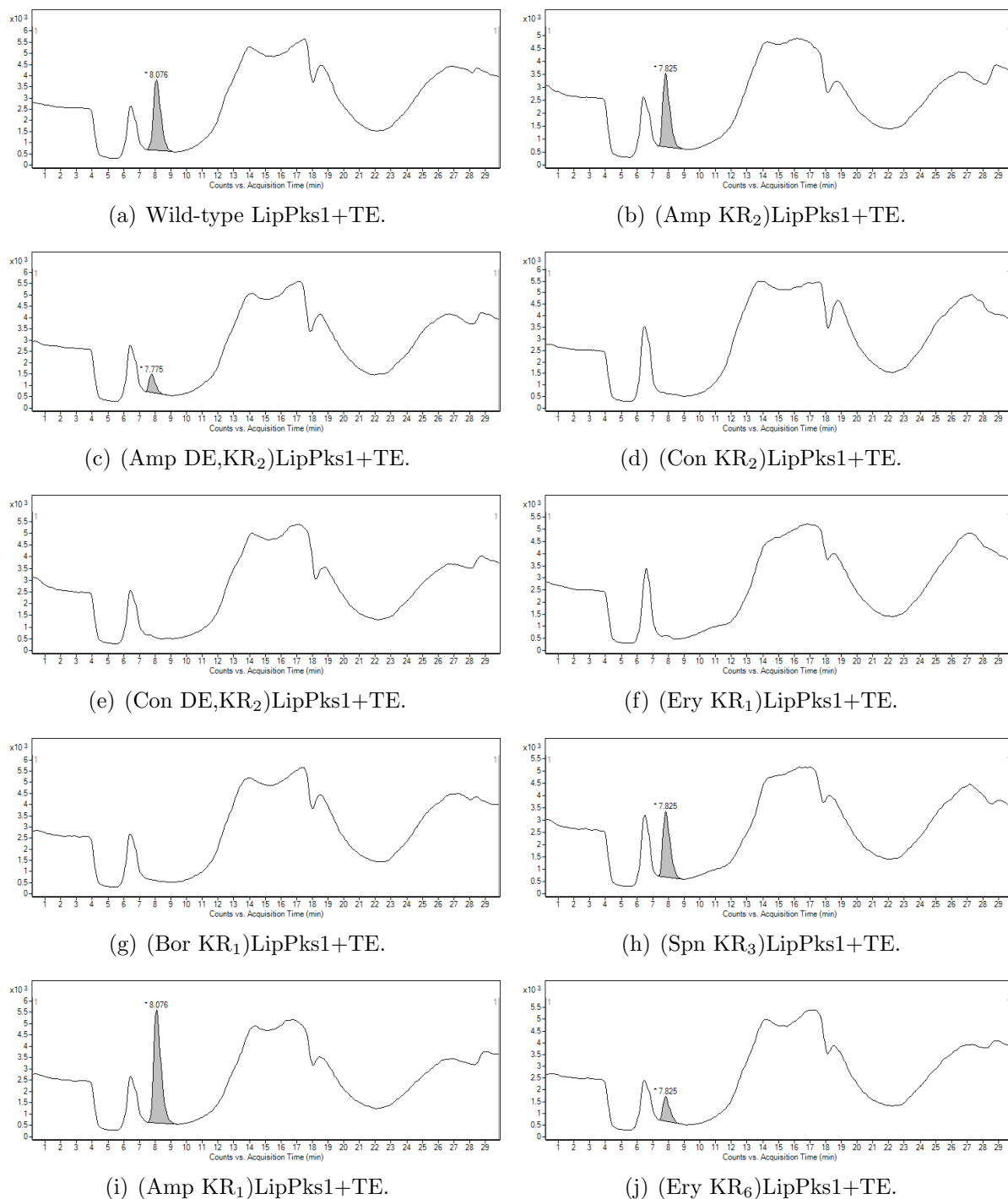
Figure B.5: Overnight production of **3b**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR_2)LipPks1+TE.

(c) (Amp DE,KR_2)LipPks1+TE.

(d) (Con KR_2)LipPks1+TE.

(e) (Con DE,KR_2)LipPks1+TE.

(f) (Ery KR_1)LipPks1+TE.

(g) (Bor KR_1)LipPks1+TE.

(h) (Spn KR_3)LipPks1+TE.

(i) (Amp KR_1)LipPks1+TE.

(j) (Ery KR_6)LipPks1+TE.
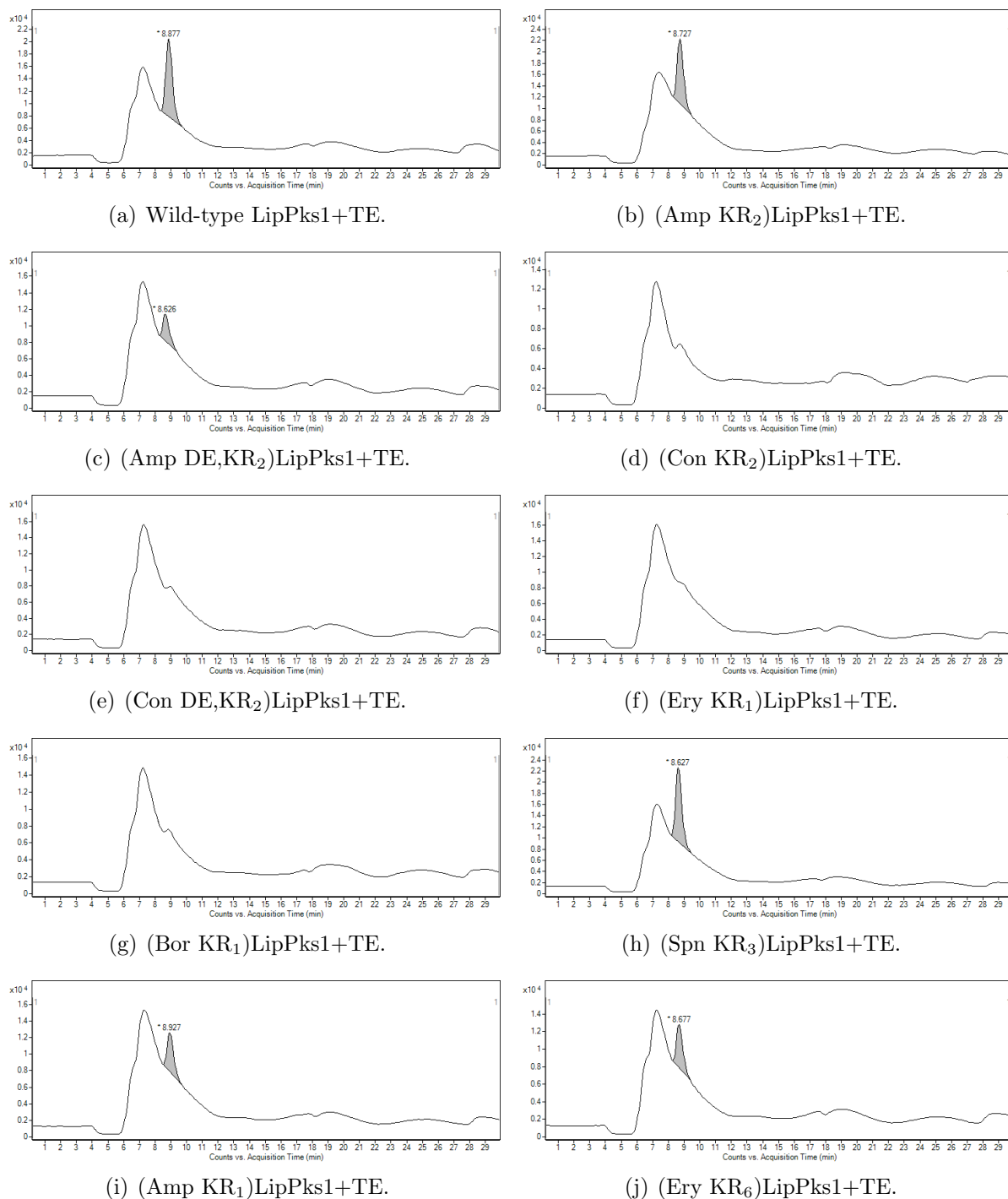
Figure B.6: Overnight production of **3c**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR$_2$)LipPks1+TE.

(c) (Amp DE,KR$_2$)LipPks1+TE.

(d) (Con KR$_2$)LipPks1+TE.

(e) (Con DE,KR$_2$)LipPks1+TE.

(f) (Ery KR$_1$)LipPks1+TE.

(g) (Bor KR$_1$)LipPks1+TE.

(h) (Spn KR$_3$)LipPks1+TE.

(i) (Amp KR$_1$)LipPks1+TE.

(j) (Ery KR$_6$)LipPks1+TE.
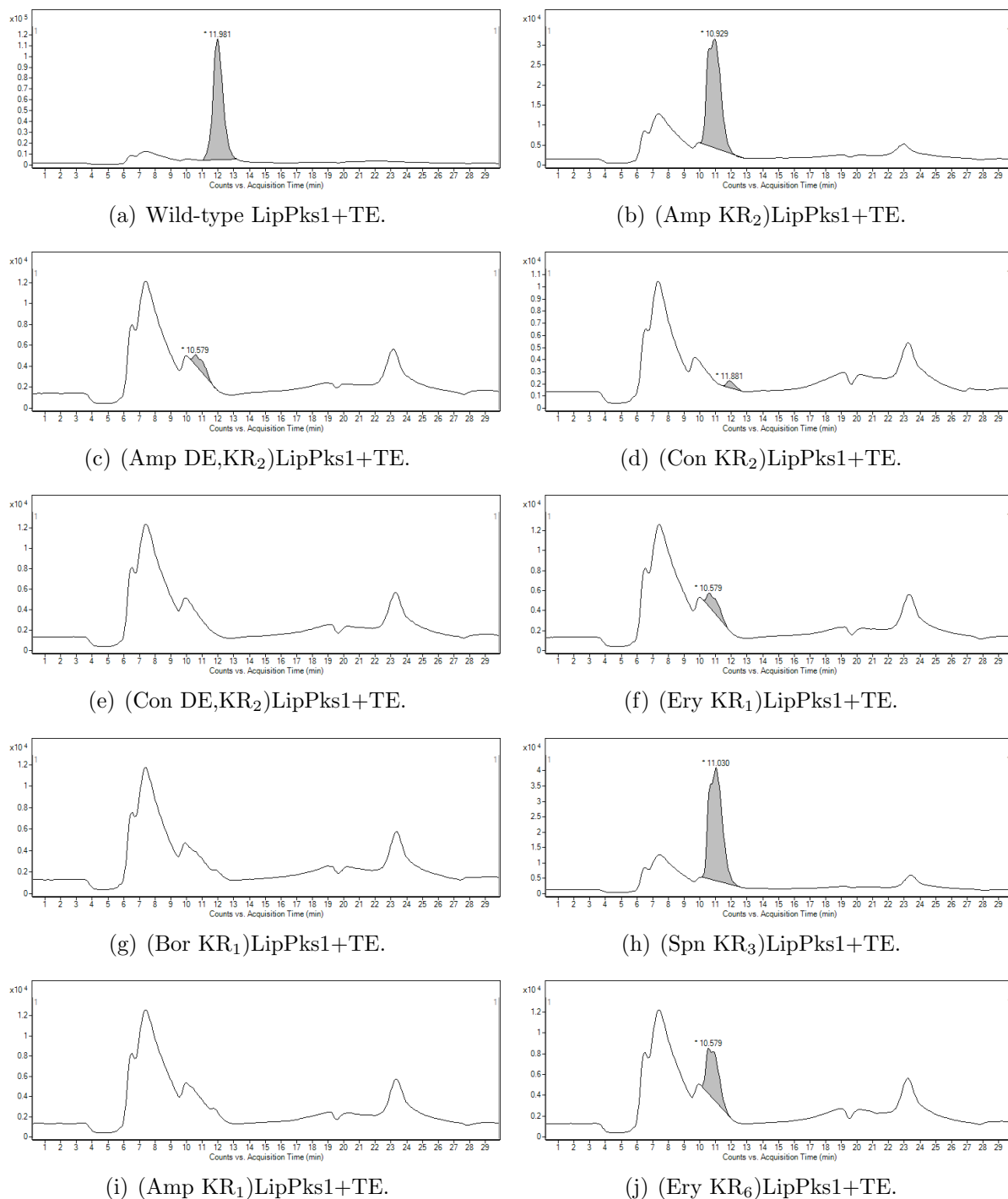
Figure B.7: Overnight production of **3d**.

B.9 Production of 2a-d by LipPks1+TE Variants

(a) 0 μM **2a**.

(b) 25 μM **2a**.

(c) 0 μM **2b**.

(d) 25 μM **2b**.

(e) 0 μM **2c**.

(f) 25 μM **2c**.

(g) 0 μM **2d**.

(h) 25 μM **2d**.
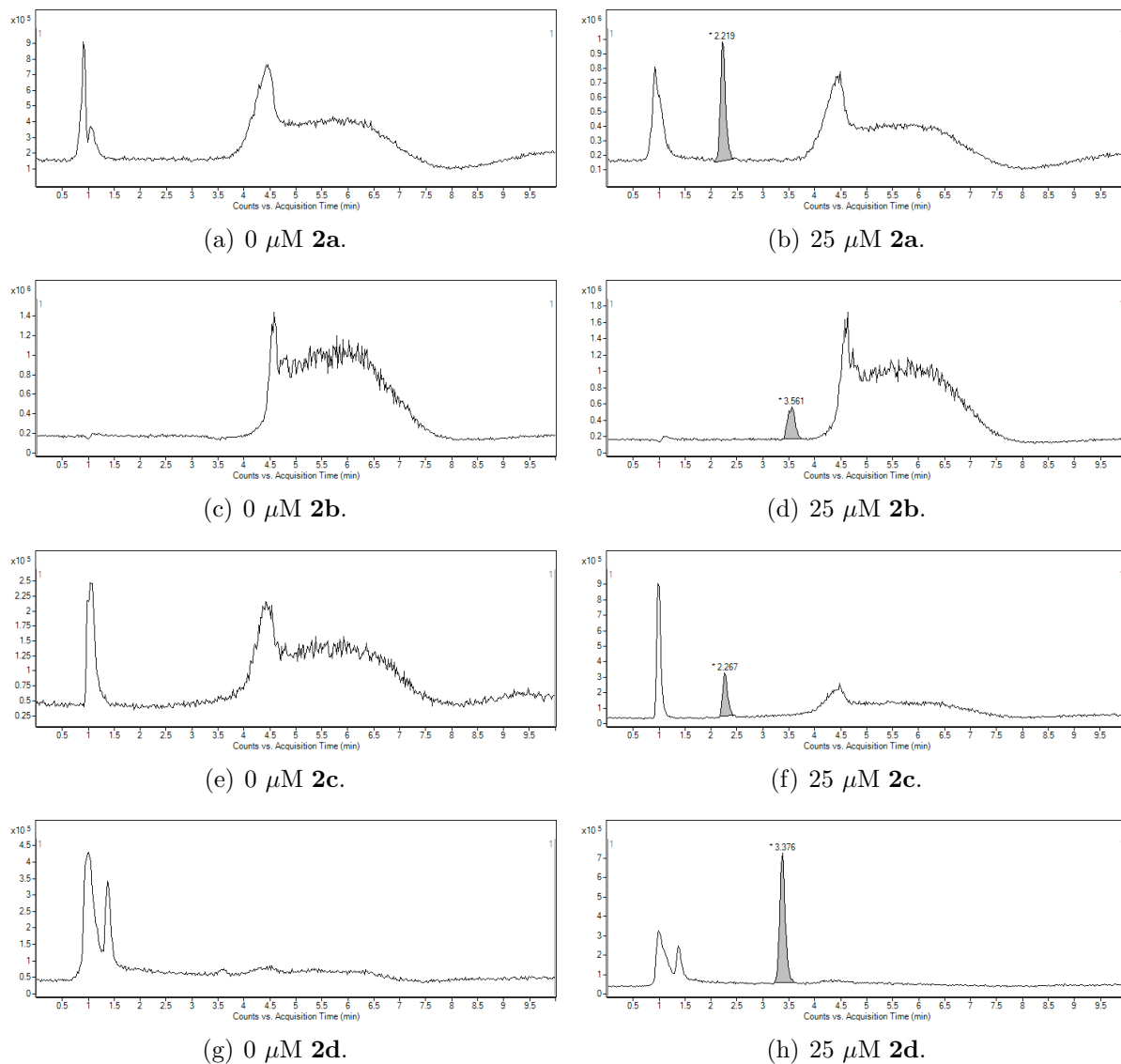
Figure B.8: Authentic ketone standards **2a-d**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR$_2$)LipPks1+TE.

(c) (Amp DE,KR$_2$)LipPks1+TE.

(d) (Con KR$_2$)LipPks1+TE.

(e) (Con DE,KR$_2$)LipPks1+TE.

(f) (Ery KR$_1$)LipPks1+TE.

(g) (Bor KR$_1$)LipPks1+TE.

(h) (Spn KR$_3$)LipPks1+TE.

(i) (Amp KR$_1$)LipPks1+TE.

(j) (Ery KR$_6$)LipPks1+TE.
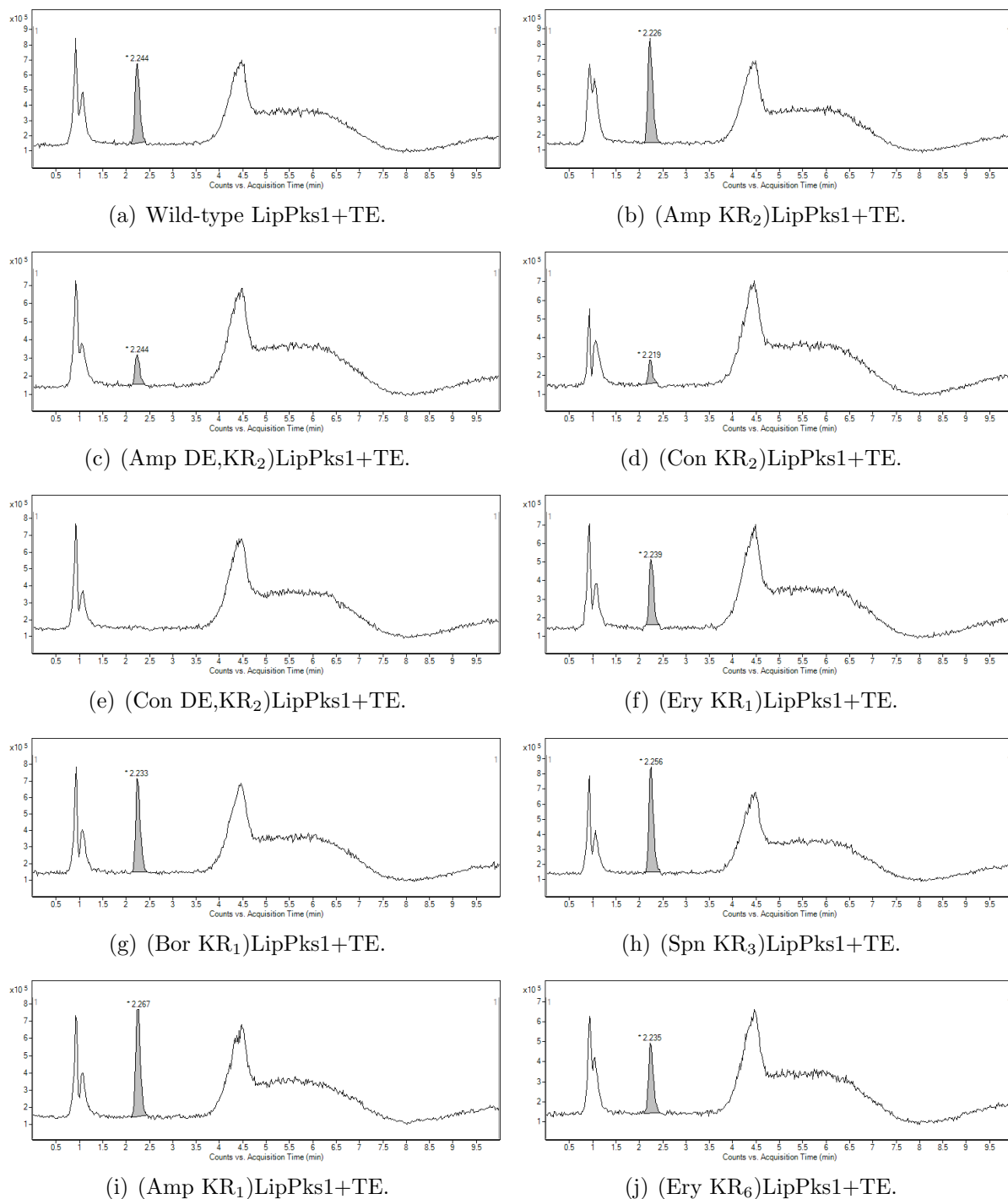
Figure B.9: Overnight production of **2a**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR$_2$)LipPks1+TE.

(c) (Amp DE,KR$_2$)LipPks1+TE.

(d) (Con KR$_2$)LipPks1+TE.

(e) (Con DE,KR$_2$)LipPks1+TE.

(f) (Ery KR$_1$)LipPks1+TE.

(g) (Bor KR$_1$)LipPks1+TE.

(h) (Spn KR$_3$)LipPks1+TE.

(i) (Amp KR$_1$)LipPks1+TE.

(j) (Ery KR$_6$)LipPks1+TE.
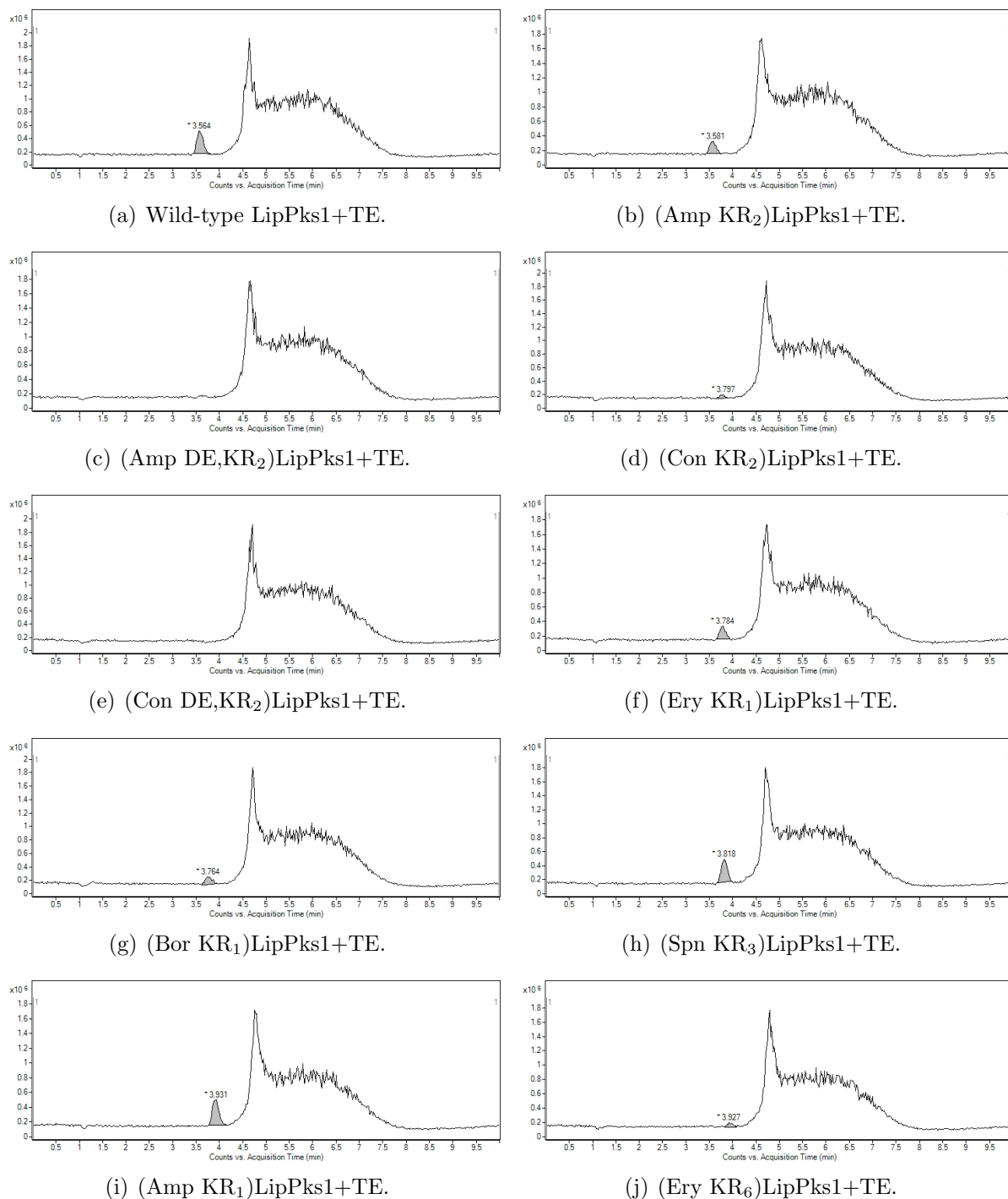
Figure B.10: Overnight production of **2b**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR_2)LipPks1+TE.

(c) (Amp DE,KR_2)LipPks1+TE.

(d) (Con KR_2)LipPks1+TE.

(e) (Con DE,KR_2)LipPks1+TE.

(f) (Ery KR_1)LipPks1+TE.

(g) (Bor KR_1)LipPks1+TE.

(h) (Spn KR_3)LipPks1+TE.

(i) (Amp KR_1)LipPks1+TE.

(j) (Ery KR_6)LipPks1+TE.
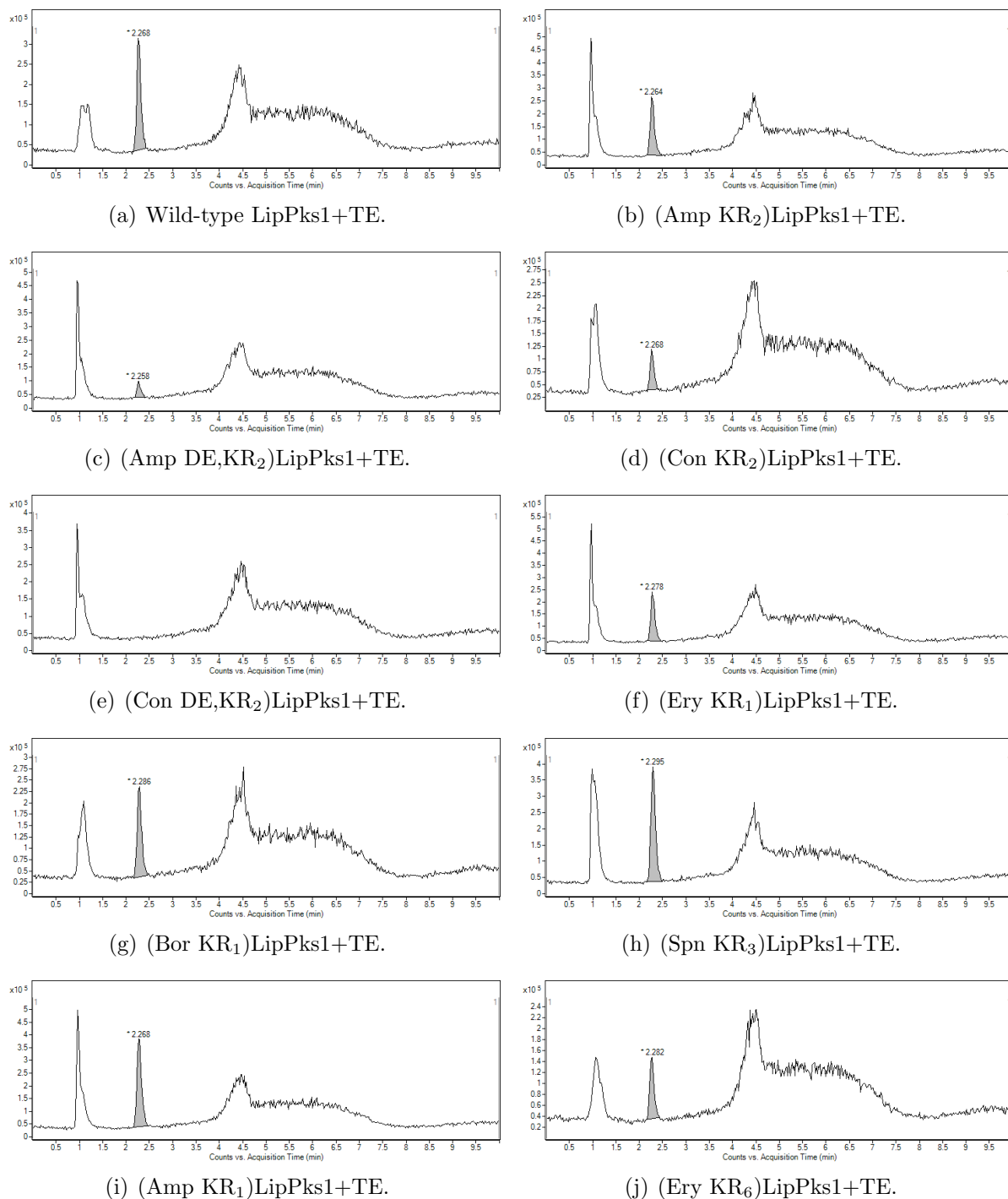
Figure B.11: Overnight production of **2c**.

(a) Wild-type LipPks1+TE.

(b) (Amp KR_2)LipPks1+TE.

(c) (Amp DE,KR_2)LipPks1+TE.

(d) (Con KR_2)LipPks1+TE.

(e) (Con DE,KR_2)LipPks1+TE.

(f) (Ery KR_1)LipPks1+TE.

(g) (Bor KR_1)LipPks1+TE.

(h) (Spn KR_3)LipPks1+TE.

(i) (Amp KR_1)LipPks1+TE.

(j) (Ery KR_6)LipPks1+TE.
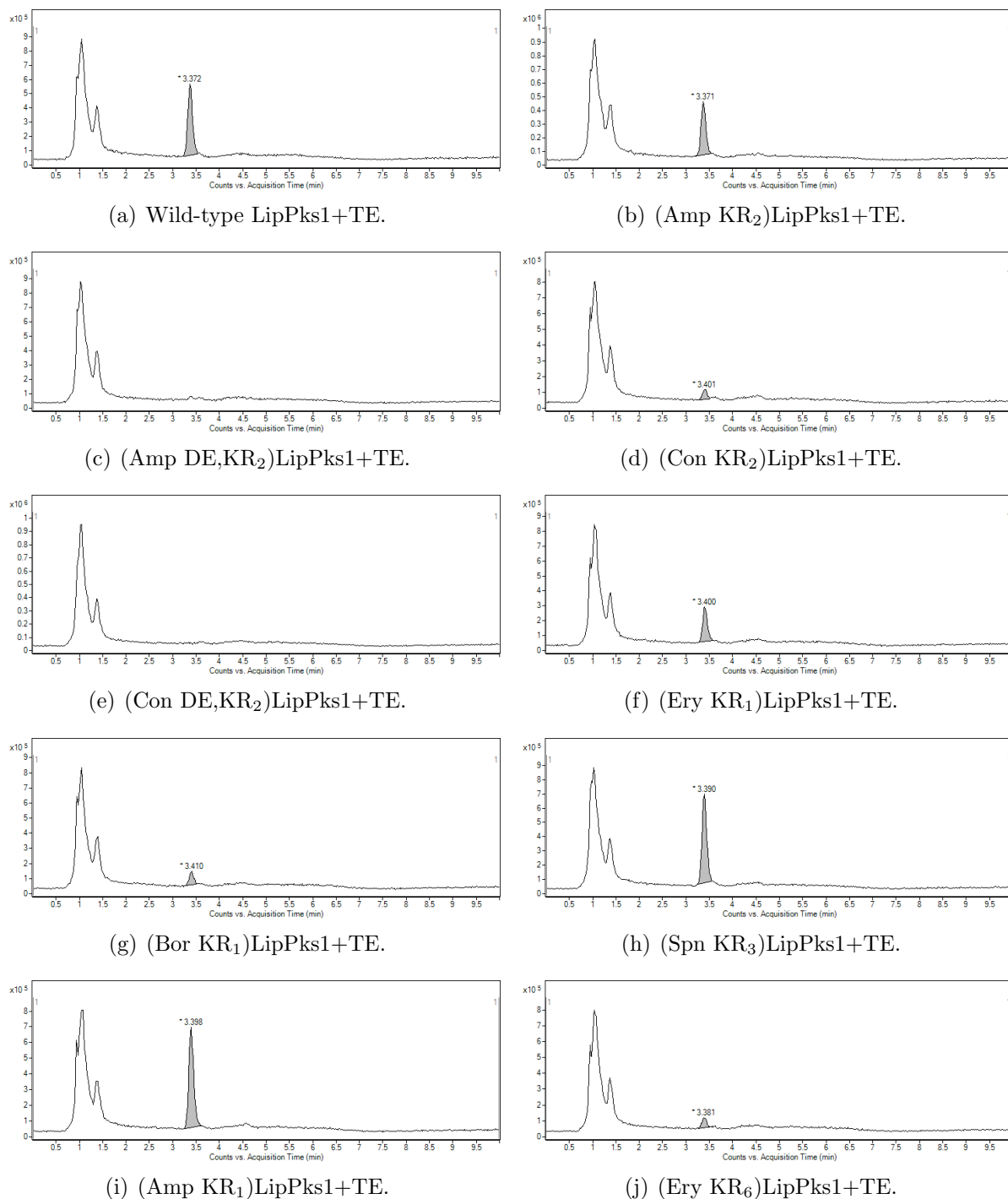
Figure B.12: Overnight production of **2d**.

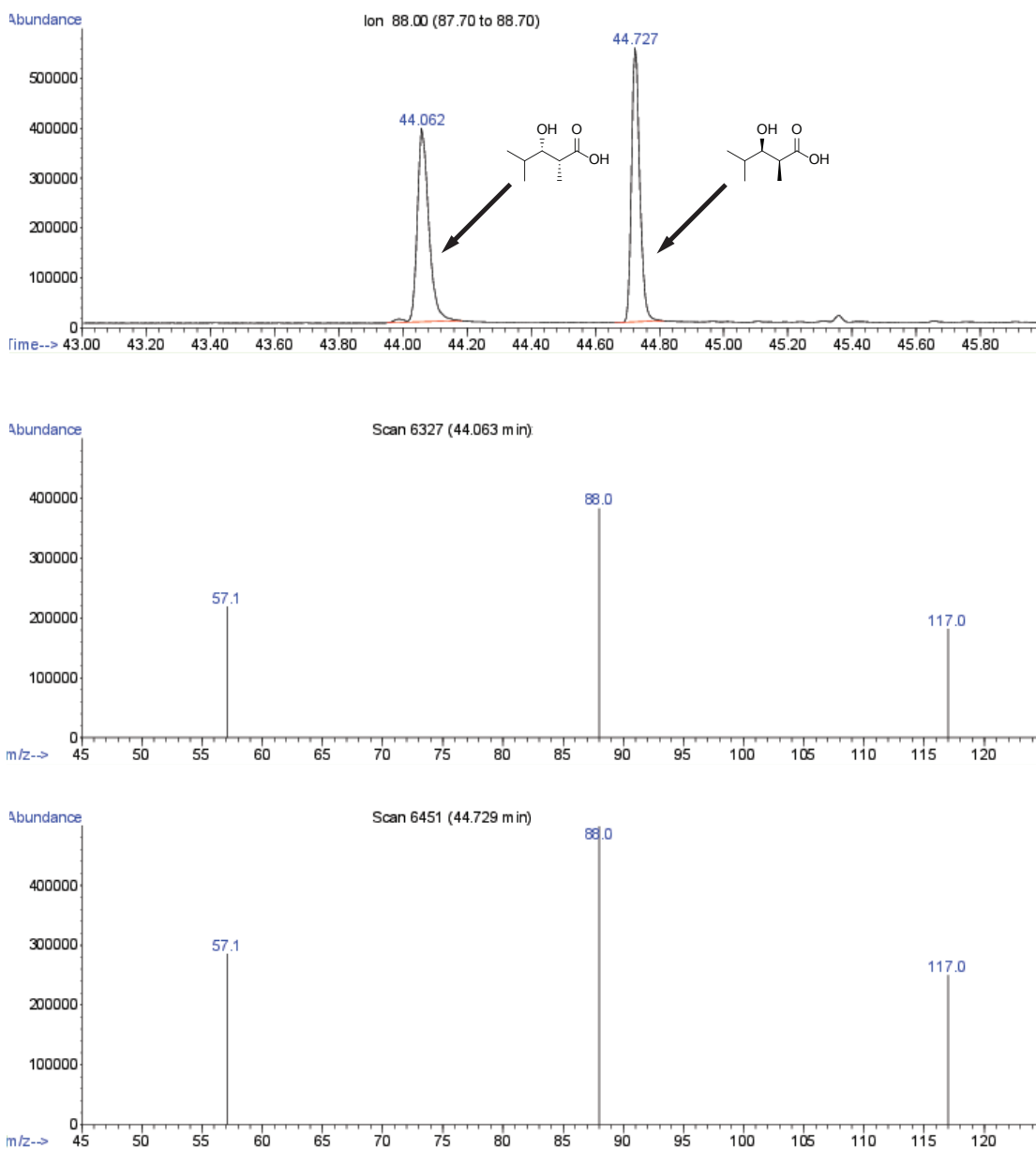# B.10 Stereochemical Purity of 3a Produced by A-type LipPks1+TE Variants

Figure B.13: Enantiomeric mixture of synthetic *syn* standards, (*2R,3S*)-**3a** and (*2S,3R*)-**3a**, corresponding to A1 and B2 type products.
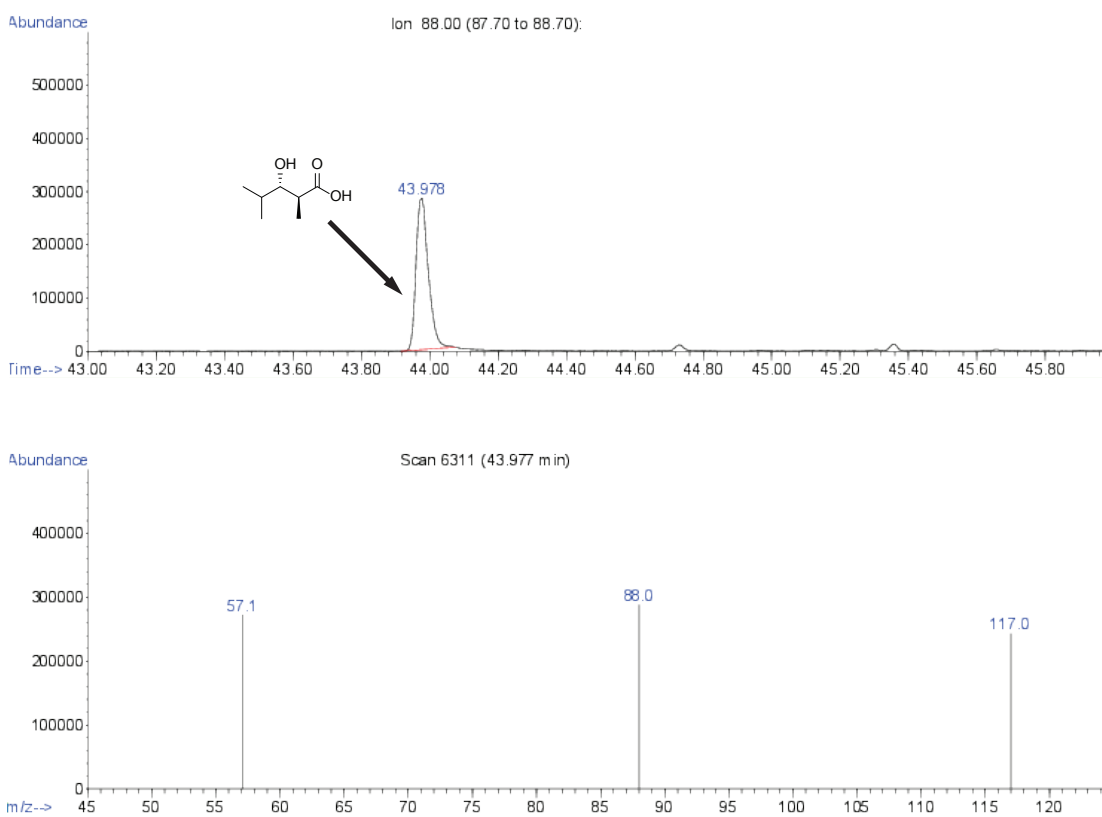
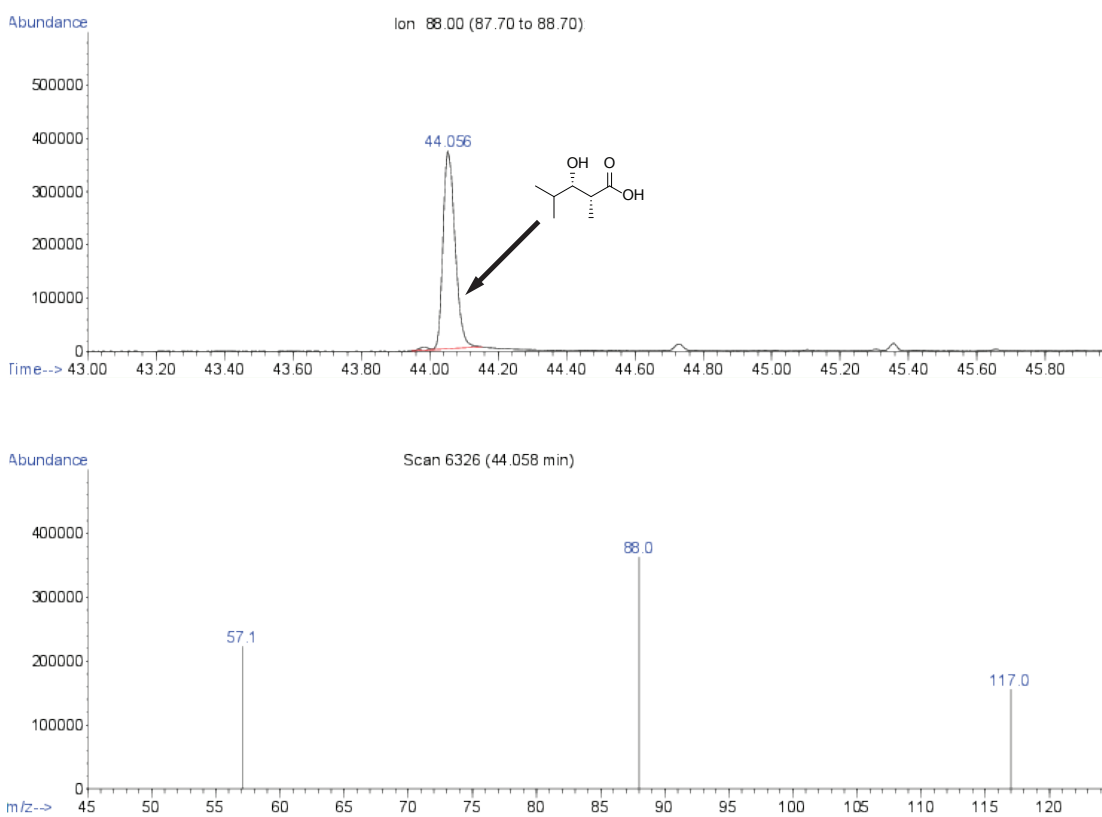Figure B.14: Synthetic (*2S,3S*)-**3a** standard corresponding to A2 type product.

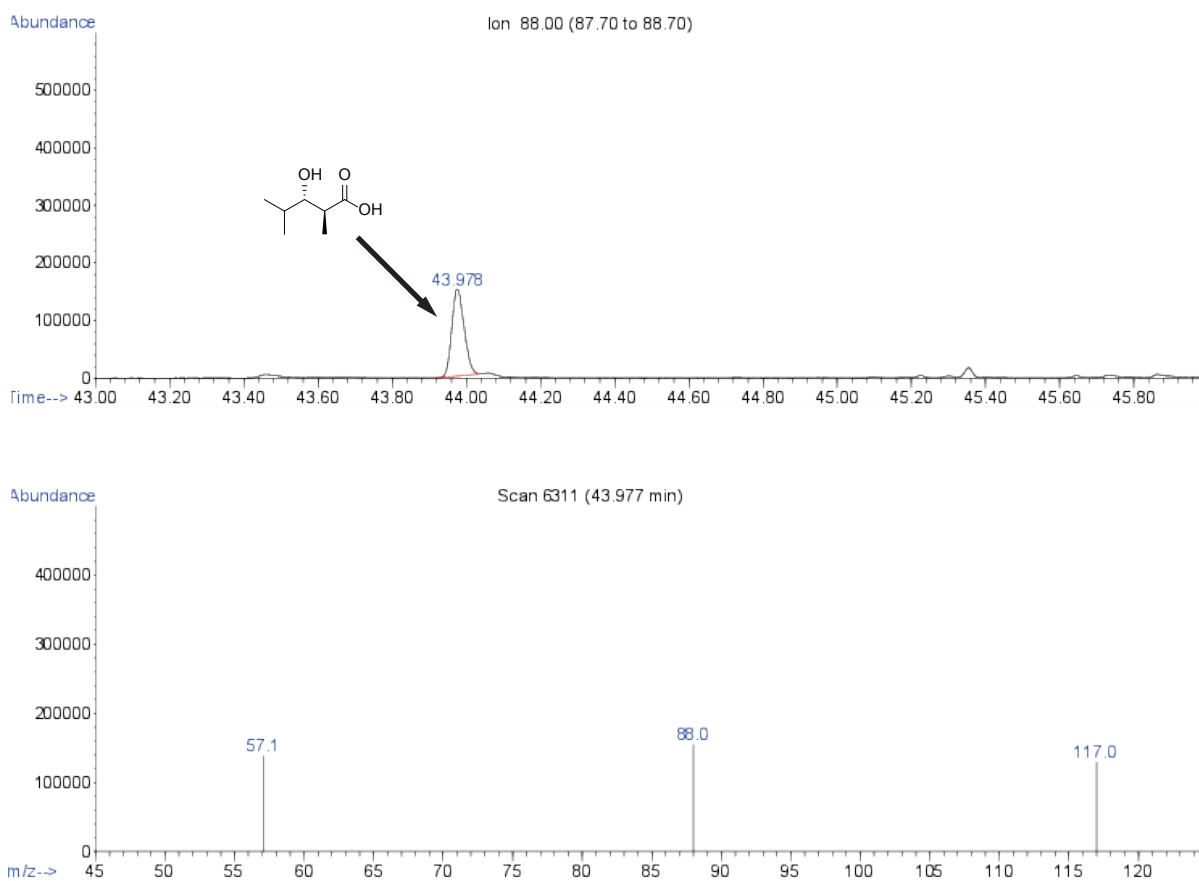Figure B.15: Synthetic (*2R,3S*)-**3a** standard corresponding to A1 type product.

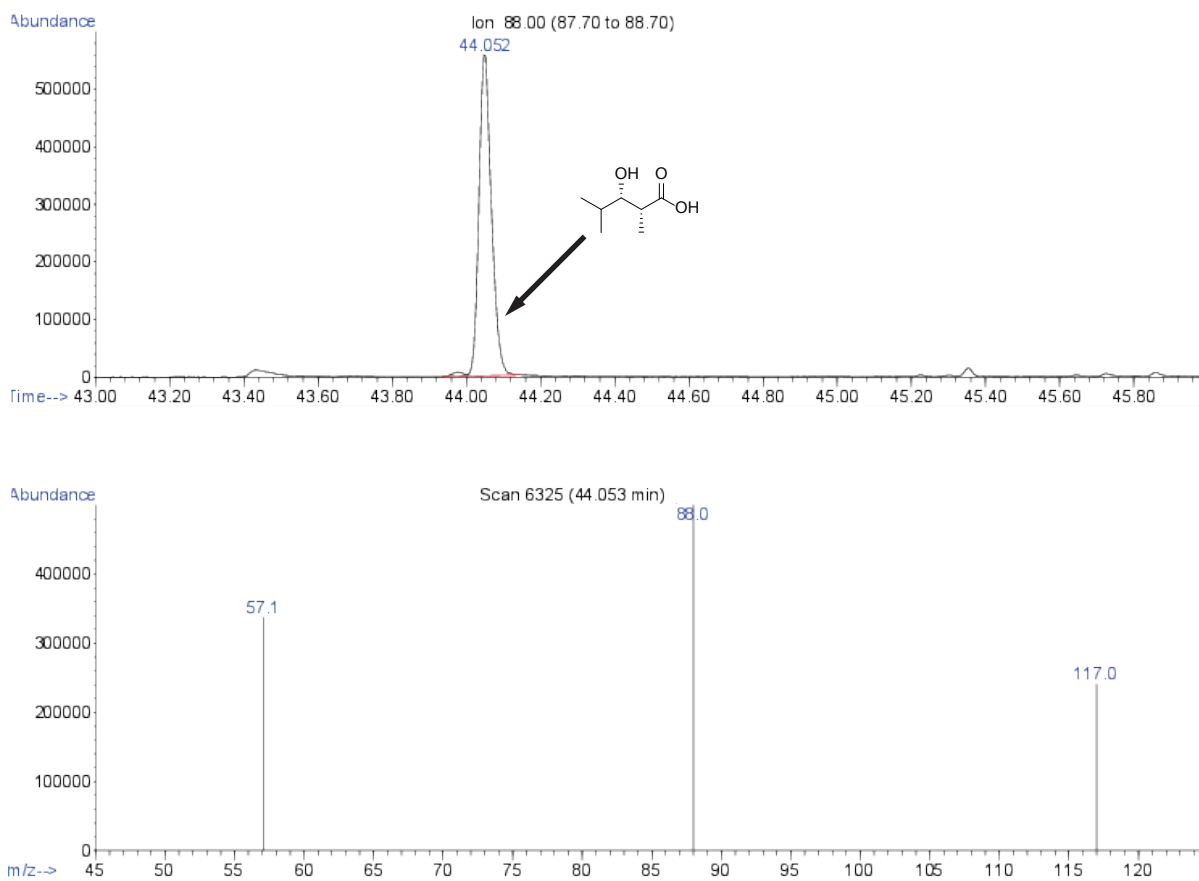Figure B.16: Stereochemical purity of **3a** produced by wild-type LipPks1+TE.

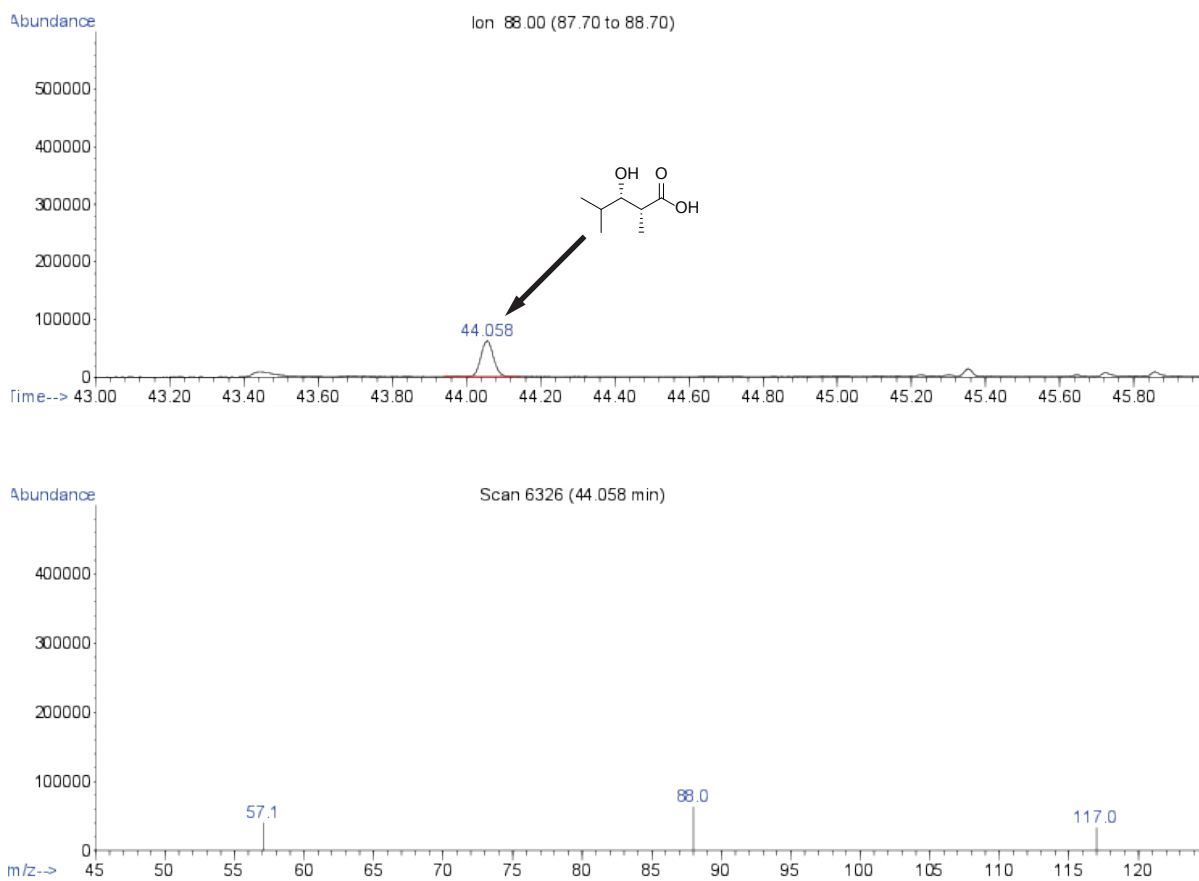Figure B.17: Stereochemical purity of **3a** produced by (Amp KR$_2$)LipPks1+TE.

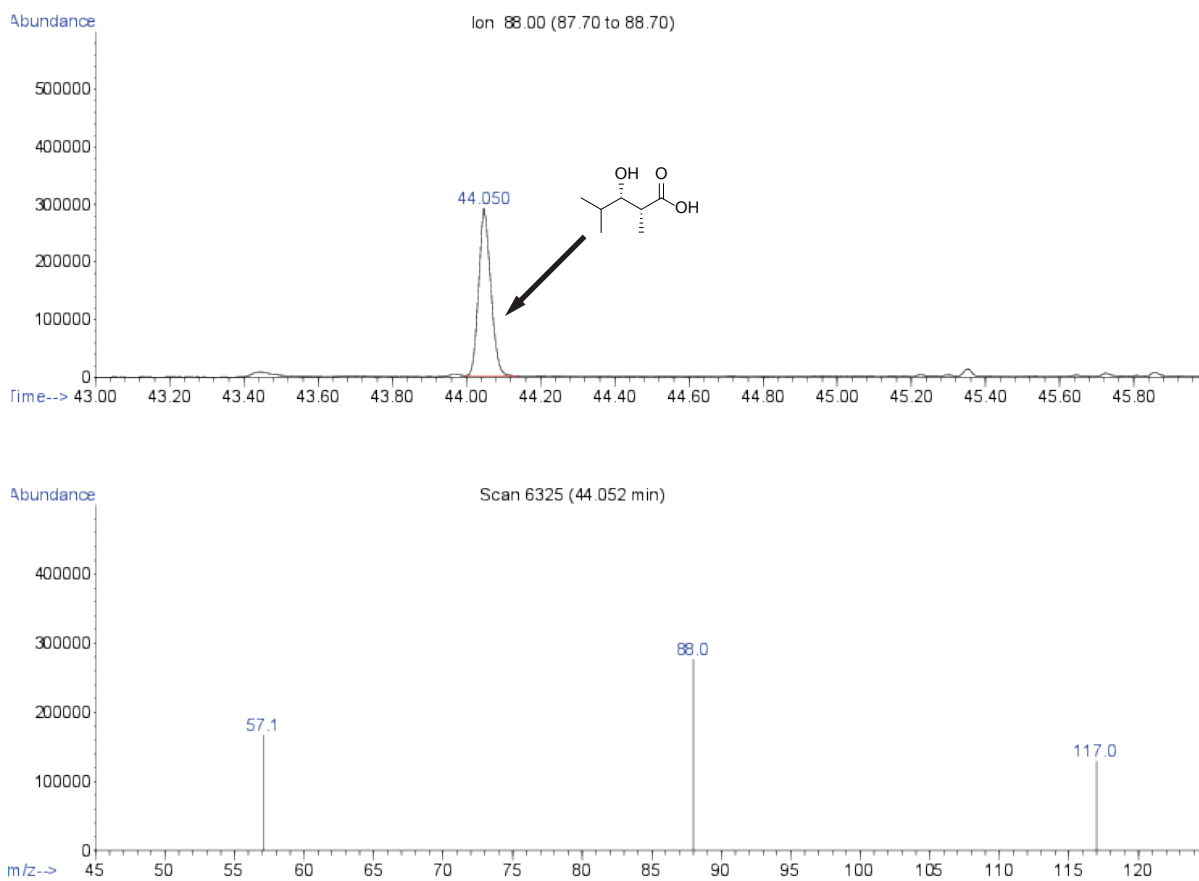Figure B.18: Stereochemical purity of **3a** produced by (Amp DE,KR$_2$)LipPks1+TE.

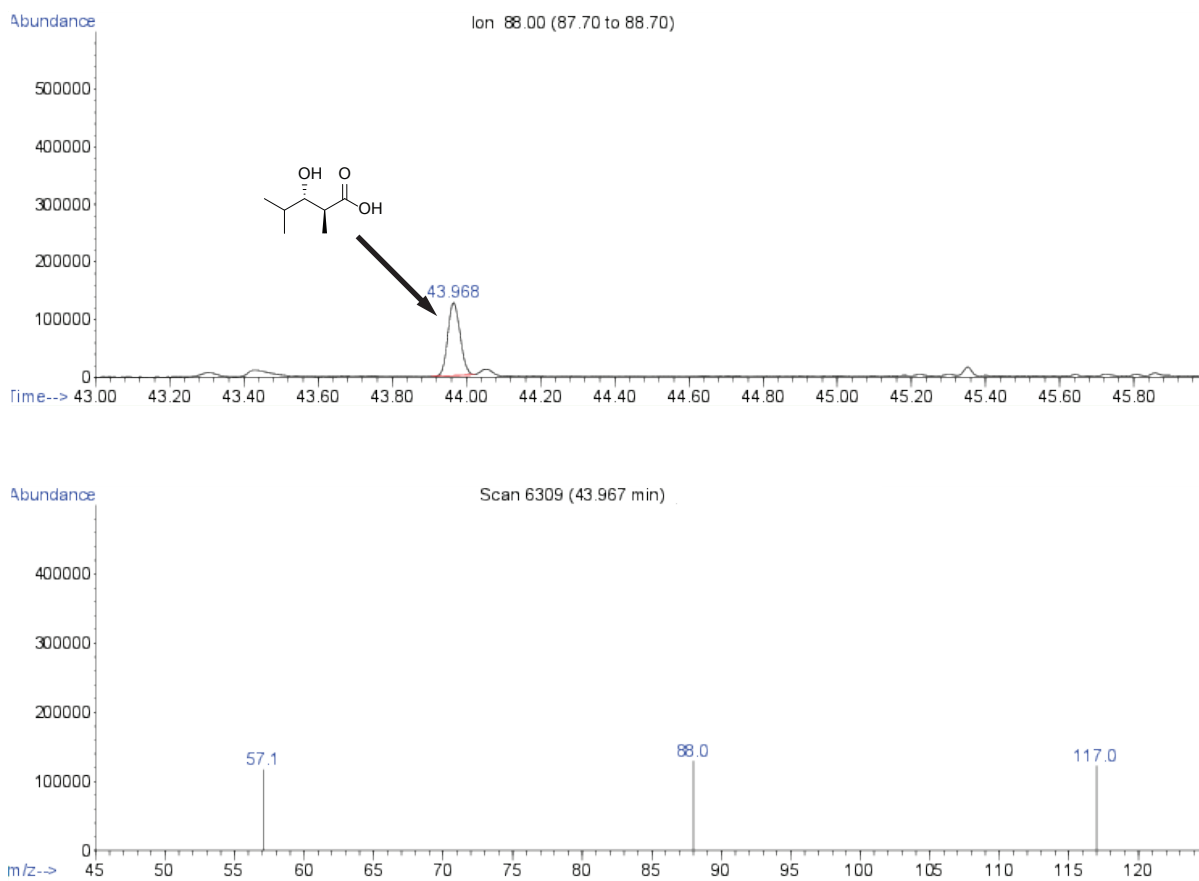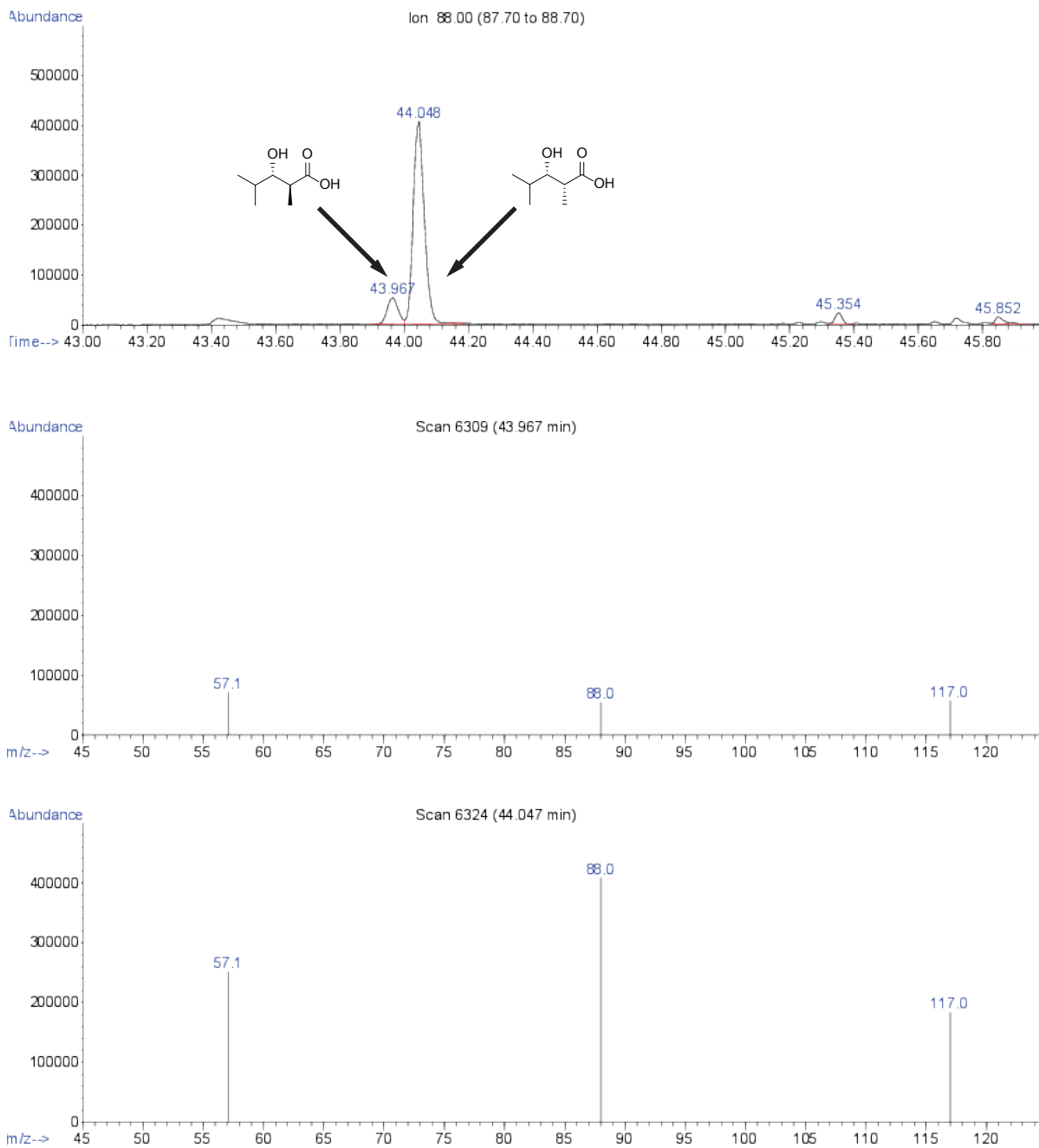Figure B.19: Stereochemical purity of **3a** produced by (Spn KR$_3$)LipPks1+TE.

Figure B.20: Stereochemical purity of **3a** produced by (Amp KR$_1$)LipPks1+TE.

Figure B.21: Stereochemical purity of **3a** produced by (Ery KR$_6$)LipPks1+TE.

B.11 Activity of LipPks1+TE Variants Using 1d as a Starter Substrate

Table B.4: Concentration of ketone **2d** produced by LipPks1+TE variants after overnight incubation with substrate **1d** in micromoles per liter, with n.d. used to indicate when no product was detected.

Substrate		1d
Product		2d
Lip KR$_1$	A2	33.6 ± 4.1
Amp KR$_2$	A1	25.5 ± 2.3
Amp KR$_2$^{a}	A1	n.d.
Con KR$_2$	B1	5.0 ± 0.7
Con KR$_2$^{a}	B1	n.d.
Ery KR$_1$	B2	18.0 ± 2.3
Bor KR$_1$	B	6.5 ± 0.3
Spn KR$_3$	A1	37.4 ± 3.3
Amp KR$_1$	A2	35.3 ± 3.3
Ery KR$_6$	A1	5.0 ± 0.7

^{a} DE from the module furnishing the donor KR domain substituted for that of LipPks1.

Table B.5: Stereochemistry of hydroxyacid **3d** produced by A-type LipPks1+TE variants as evaluated by LC-MS, with n.d. used to indicate when no product was detected.

Substrate		1d	
Stereochemistry		A1/B2	A2/B1
Lip KR$_1$	A2		●
Amp KR$_2$	A1	●	
Amp KR$_2$^{a}	A1	●	
Spn KR$_3$	A1	●	
Amp KR$_1$	A2		n.d.
Ery KR$_6$	A1	●	

^{a} DE from the module furnishing the donor KR domain substituted for that of LipPks1.

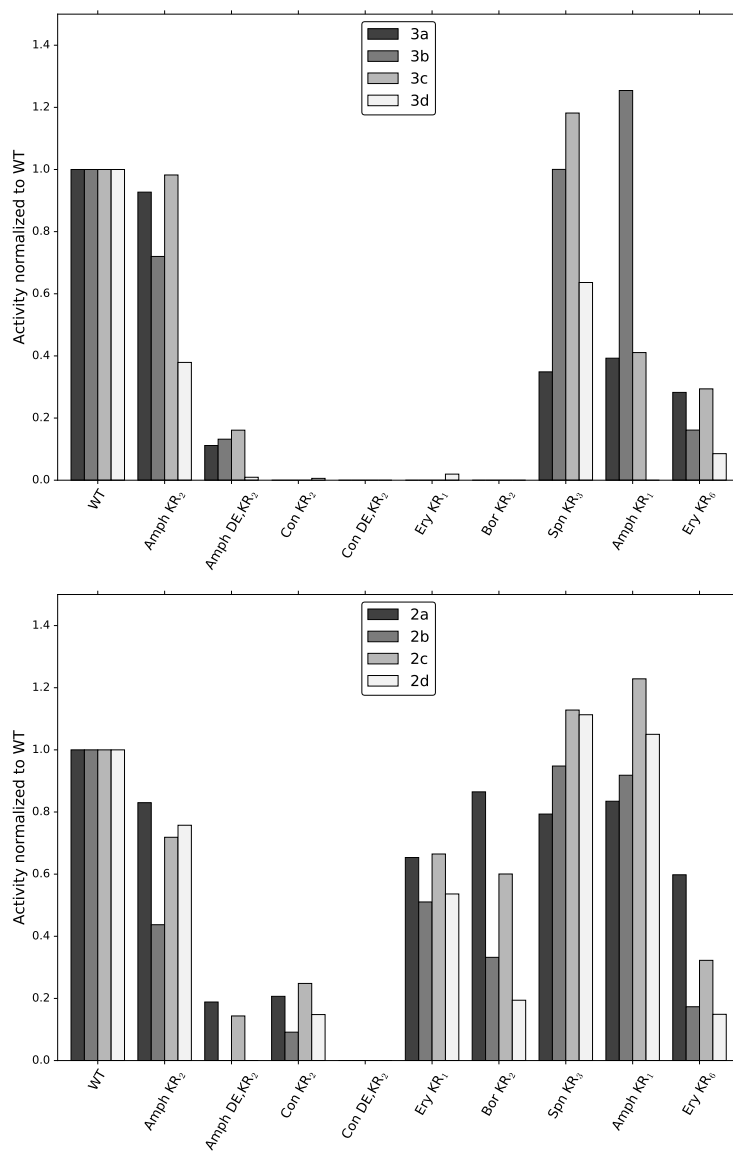# B.12 Relative Production of 3a-d and 2a-d by LipPks1+TE Variants

Figure B.22: Production of hydroxyacids **3a-d** (top) and ketones **2a-d** (bottom) by LipPks1+TE variants normalized to production by wild-type LipPks1+TE. Production is calculated as the mean of experiments performed in triplicate.

Bibliography

[1] V. Y. Alekseyev et al. "Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase". *Protein Sci.* 16, 2093–107, 2007.

[2] T. Annaval et al. "Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide Stereochemistry". *ChemBioChem* 16, 1357–1364, 2015.

[3] Abel Baerga-Ortiz et al. "Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase". *Chem. Biol.* 13, 277–285, 2006.

[4] C. Bihlmaier et al. "Biosynthetic gene cluster for the polyenoyltetramic acid alpha-lipomycin". *Antimicrob. Agents Chemother.* 50, 2113–21, 2006.

[5] R. W. Burg et al. "Avermectins, new family of potent anthelmintic agents: producing organism and fermentation". *Antimicrob. Agents Chemother.* 15, 361–367, 1979.

[6] M. A. Burlingame, E. Mendoza, and G. W. Ashley. "*N*-Acyl-2-benzoxazolinones in titanium-mediated aldol reactions". *Tetrahedron Lett.* 45, 2961–2964, 2004.

[7] P. Caffrey. "Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases". *ChemBioChem* 4, 654–657, 2003.

[8] C. Camacho et al. "BLAST+: architecture and applications". *BMC Bioinform* 10, 421, 2009.

[9] D. E. Cane et al. "Biosynthetic origin of the carbon skeleton and oxygen atoms of the avermectins". *J. Am. Chem. Soc.* 105, 4110–4112, 1983.

[10] D. E. Cane et al. "Macrolide biosynthesis. 3. Stereochemistry of the chain-elongation steps of erythromycin biosynthesis". *J. Am. Chem. Soc.* 108, 4957–4964, 1986.

[11] R. E. Carhart, D. H. Smith, and R. Venkataraghavan. "Atom pairs as molecular features in structure-activity studies: definition and applications". *J. Chem. Inf. Model.* 25, 64–73, 1985.

[12] C. Chang et al. "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.* 137, 4183–4190, 2015.

[13] J. Cheng et al. "SCRATCH: a protein structure and structural feature prediction server". *Nucleic Acids Res.* 33, W72–W76, 2005.

[14] K. R. Conway and C. N. Boddy. "ClusterMine360: a database of microbial PKS/NRPS biosynthesis". *Nucleic Acids Res* 41, D402–7, 2013.

[15] B. J. Dunn and C. Khosla. "Engineering the acyltransferase substrate specificity of assembly line polyketide synthases". *J. R. Soc. Interface* 10, 20130297, 2013.

[16] S. Dutta et al. "Structure of a modular polyketide synthase". *Nature* 510, 512–7, 2014.

[17] Christopher J Dutton et al. "Novel avermectins produced by mutational biosynthesis". *J. Antibiot.* 44, 357–365, 1991.

[18] C. H. Eng et al. "Alteration of polyketide stereochemistry from *anti* to *syn* by a ketoreductase domain exchange in a type I modular polyketide synthase subunit". *Biochemistry* 55, 1677–1680, 2016.

[19] C. Engler, R. Kandzia, and S. Marillonnet. "A one pot one step precision cloning method with high throughput capability". *PLoS ONE* 3, 2008.

[20] C. Engler et al. "Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes". *PLoS ONE* 4, 2009.

[21] D. A. Evans, T. C. Britton, and J. A. Ellman. "Contrasteric Carboximide Hydrolysis with Lithium Hydroperoxide". *Tetrahedron Lett.* 28, 6141–6144, 1987.

[22] A. Faille et al. "Insights into Substrate Modification by Dehydratases from Type I Polyketide Synthases". *J. Mol. Biol.* 429, 1554–1569, 2017.

[23] S. Gaisser et al. "Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of Streptomyces avermitilis ATCC31272". *Org. Biomol. Chem.* 1, 2840–7, 2003.

[24] A. Garg, C. Khosla, and D. E. Cane. "Coupled methyl group epimerization and reduction by polyketide synthase ketoreductase domains. Ketoreductase-catalyzed equilibrium isotope exchange". *J. Am. Chem. Soc.* 135, 16324–16327, 2013.

[25] A. Garg et al. "Elucidation of the cryptic epimerase activity of redox-inactive ketoreductase domains from modular polyketide synthases by tandem equilibrium isotope exchange". *J. Am. Chem. Soc.* 136, 10190–10193, 2014.

[26] M. Hackh, M. Muller, and S. Ludeke. "Substrate-dependent stereospecificity of Tyl-KR1: an isolated polyketide synthase ketoreductase domain from Streptomyces fradiae". *Chemistry* 19, 8922–8928, 2013.

[27] E. W. Hafner et al. "Branched-chain fatty acid requirement for avermetin production by a mutant of *Streptomyces avermitilis* lacking branched-chain 2-oxo acid dehydrogenase activity". *J. Antibiot.* 44, 349–356, 1991.

[28] A. Hagen et al. "*In vitro* analysis of carboxyacyl substrate tolerance in the loading and first extension modules of borrelidin polyketide synthase". *Biochemistry* 53, 5976–5977, 2014.

[29] A. Hagen et al. "Engineering a polyketide synthase for *in vitro* production of adipic acid". *ACS Synth. Biol.* 5, 21–27, 2015.

[30] M. Hans et al. "Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules". *J. Am. Chem. Soc.* 125, 5366–5374, 2003.

[31] O. Hartmann and M. Kalesse. "The Structure Elucidation and Total Synthesis of beta-Lipomycin". *Angew. Chem. Int. Ed. Engl.* 53, 7335–8, 2014.

[32] N. J. Hillson, R. D. Rosengarten, and J. D. Keasling. "j5 DNA assembly design automation software". *ACS Synth Biol* 1, 14–21, 2012.

[33] M. L. Hofferberth and R. Bruckner. "α- and β-Lipomycin: Total Syntheses by Sequential Stille Couplings and Assignment of the Absolute Configuration of All Stereogenic Centers". *Angew. Chem. Int. Ed. Engl.* 53, 7328–34, 2014.

[34] H. Ikeda et al. "Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*". *Proc. Natl Acad. Sci. U.S.A.* 96, 9509–9514, 1999.

[35] C. A. James, D. Weininger, and J. Delany. "Daylight theory manual daylight version 4.82. Daylight Chemical Information Systems". 2003.

[36] C. M. Kao et al. "Alcohol stereochemistry in polyketide backbones is controlled by the β-ketoreductase domains of modular polyketide synthases". *J. Am. Chem. Soc.* 120, 2478–2479, 1998.

[37] C. M. Kao et al. "Gain of function mutagenesis of the erythromycin polyketide synthase. 2. Engineered biosynthesis of eight-membered ring tetraketide lactone". *J. Am. Chem. Soc.* 119, 11339–11340, 1997.

[38] K. Katoh and D. M. Standley. "MAFFT multiple sequence alignment software version 7: improvements in performance and usability". *Mol. Biol. Evol.* 30, 772–80, 2013.

[39] A. T. Keatinge-Clay. "A tylosin ketoreductase reveals how chirality is determined in polyketides". *Chem. Biol.* 14, 898–908, 2007.

[40] A. T. Keatinge-Clay. "Stereocontrol within polyketide assembly lines". *Nat. Prod. Rep.* 33, 141–149, 2016.

[41] A. T. Keatinge-Clay. "The structures of type I polyketide synthases". *Nat. Prod. Rep.* 29, 1050–1073, 2012.

[42] A. T. Keatinge-Clay and R. M. Stroud. "The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases". *Structure* 14, 737–748, 2006.

[43] L. Kellenberger et al. "A polylinker approach to reductive loop swaps in modular polyketide synthases". *ChemBioChem* 9, 2740–9, 2008.

[44] Shradha Khater et al. "SBSPKSv2: structure-based sequence analysis of polyketide synthases and non-ribosomal peptide synthetases". *Nucleic Acids Res.* 45, W72–W79, 2017.

[45] C. Khosla et al. "Structure and mechanism of the 6-deoxyerythronolide B synthase". *Annu. Rev. Biochem.* 76, 195–221, 2007.

[46] B. Kunze et al. "Metabolic products of microorganisms. 3 lipomycins. I. Isolation, characterization and first studies of the structure and the mechanism of action". *Arch. Microbiol.* 86, 147, 1972.

[47] D. H. Kwan et al. "Insights into the stereospecificity of ketoreduction in a modular polyketide synthase". *Org. Biomol. Chem.* 9, 2053–6, 2011.

[48] G. Landrum. *RDKit: Open-source cheminformatics.* URL: http://www.rdkit.org.

[49] G. F. Liou et al. "Quantitative analysis of loading and extender acyltransferases of modular polyketide synthases". *Biochemistry* 42, 200–207, 2003.

[50] C. Magnan and P. Baldi. "SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity". *Bioinformatics* 30, 2592–2597, 2014.

[51] A. F. Marsden et al. "Stereospecific acyl transfers on the erythromycin-producing polyketide synthase". *Science* 263, 378–380, 1994.

[52] R. McDaniel et al. "Engineered intermodular and intramodular polyketide synthase fusions". *Chem. Biol.* 4, 667–674, 1997.

[53] R. McDaniel et al. "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products". *Proc. Natl Acad. Sci. U.S.A.* 96, 1846–51, 1999.

[54] M. H. Medema et al. "Minimum Information about a Biosynthetic Gene cluster". *Nat. Chem. Biol.* 11, 625–631, 2015.

[55] S. Murli et al. "Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production". *J. Ind. Microbiol. Biotechnol.* 30, 500–9, 2003.

[56] A. C. Murphy et al. "Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation". *Chem. Comm.* 52, 8373–8376, 2016.

[57] S. K. Piasecki et al. "Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks". *Chem. Biol.* 18, 1331–1340, 2011.

[58] J. W. Raymond, E. J. Gardiner, and P. Willett. "Heuristics for Similarity Searching of Chemical Graphs Using a Maximum Common Edge Subgraph Algorithm". *J. Chem. Inf. Comput. Sci.* 42, 305–316, 2002.

[59] R. Reid et al. "A model of structure and catalysis for ketoreductase domains in modular polyketide synthases". *Biochemistry* 42, 72–79, 2003.

[60] K. Schabacher and A. Zeeck. "Lipomycine, II die konstitution von α-und β-lipomycin". *Tetrahedron Lett.* 14, 2691–2694, 1973.

[61] M. D. Schulman, D. Valentino, and O. Hensens. "Biosynthesis of the avermectins by *Streptomyces avermitilis* incorporation of labeled precursors". *J. Antibiot.* 39, 541–549, 1986.

[62] A. P. Siskos et al. "Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases". *Chem. Biol.* 12, 1145–1153, 2005.

[63] J. Staunton et al. "Evidence for a double-helical structure for modular polyketide synthases". *Nat. Struct. Mol. Biol* 3, 188–192, 1996.

[64] S. Sundaram and C. Hertweck. "On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems". *Curr. Opin. Chem. Biol.* 31, 82–94, 2016.

[65] Y. Tang et al. "The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase". *Proc. Natl. Acad. Sci. U.S.A.* 103, 11124–11129, 2006.

[66] S. Tsai et al. "Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel". *Proc. Natl. Acad. Sci. U.S.A.* 98, 14808–14813, 2001.

[67] C. R. Valenzano et al. "The biochemical basis for stereochemical control in polyketide biosynthesis". *J. Am. Chem. Soc.* 131, 18501–18511, 2009.

[68] N. A. Van Draanen et al. "Acyclic Stereoselection. 53. Protocols for the Preparation of Each of the 4 Possible Stereoisomeric α-Alkyl-β-Hydroxy Carboxylic-Acids from a Single Chiral Aldol Reagent". *Journal of Organic Chemistry* 56, 2499–2506, 1991.

[69] L. Weber. "JChem Base - ChemAxon". *Chemistry World* 5, 65–66, 2008.

[70] Tilmann Weber et al. "antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters". *Nucleic Acids Res.* 43, W237–W243, 2015.

[71] K. J. Weissman. "The structural biology of biosynthetic megaenzymes". *Nat. Chem. Biol.* 11, 660–670, 2014.

[72] K. J. Weissman and P.F. Leadlay. "Combinatorial biosynthesis of reduced polyketides". *Nat. Rev. Microbiol.* 3, 925–936, 2005.

[73] K. J. Weissman et al. "The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase". *Biochemistry* 36, 13849–13855, 1997.

[74] J. R. Whicher et al. "Structural rearrangements of a polyketide synthase module during its catalytic cycle". *Nature* 510, 560–4, 2014.

[75] M. C. Wilson and B. S. Moore. "Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity". *Nat. Prod. Rep.* 29, 72–86, 2012.

[76] J. Wu et al. "Biochemical analysis of the substrate specificity of the β-ketoacyl-acyl carrier protein synthase domain of module 2 of the erythromycin polyketide synthase". *Biochemistry* 43, 16301–16310, 2004.

[77] N. Wu et al. "Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules". *J. Am. Chem. Soc.* 123, 6465–6474, 2001.

[78] S. Yue et al. "Macrolide Biosynthesis - Tylactone Formation Involves the Processive Addition of 3 Carbon Units". *J. Am. Chem. Soc.* 109, 1253–1255, 1987.

[79] S. Yuzawa, J. D. Keasling, and L. Katz. "Bio-based production of fuels and industrial chemicals by repurposing antibiotic-producing type I modular polyketide synthases: opportunities and challenges". *J. Antibiot.* 70, 378–385, 2017.

[80] S. Yuzawa et al. "Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase". *Biochemistry* 52, 3791–3, 2013.

[81] S. Yuzawa et al. "Comprehensive *in Vitro* Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production". *ACS Synth. Biol.* 7, 139–147, 2016.

[82] S. Yuzawa et al. "Enzyme analysis of the polyketide synthase leads to the discovery of a novel analog of the antibiotic α-lipomycin". *J. Antibiot.* 67, 199–201, 2014.

[83] A. Zeeck. "Lipomycine, III. Isolierung und Zuordnung der Methyl-2,6-didesoxy-D-ribo-hexoside". *Eur. J. Org. Chem.* 1975, 2079–2088, 1975.

[84] J. Zheng and A. T. Keatinge-Clay. "Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases". *J. Mol. Biol.* 410, 105–17, 2011.

[85] J. Zheng and A. T. Keatinge-Clay. "The status of type I polyketide synthase ketoreductases". *Med. Chem. Commun.* 4, 34–40, 2013.

[86] J. Zheng, S. K. Piasecki, and A. T. Keatinge-Clay. "Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling α-substituent stereochemistry". *ACS Chem. Biol.* 8, 1964–1971, 2013.

[87] J. Zheng et al. "Divergence of multimodular polyketide synthases revealed by a didomain structure". *Nat. Chem. Biol.* 8, 615–21, 2012.

[88] J. Zheng et al. "Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase". *Structure* 18, 914–922, 2010.

[89] P. F. Zierep et al. "SeMPI: a genome-based secondary metabolite prediction and identification web server". *Nucleic Acids Res.* 45, W64–W71, 2017.