UNIVERSITY OF CALIFORNIA,
IRVINE

Decoding Polyketide Synthase Specificity through Structural Informatics and Biochemical Reconstitution

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Fanglue Ni

Dissertation Committee:
Professor Shiou-Chuan (Sheryl) Tsai, Chair
Professor Ray Luo
Professor Yilin Hu

2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Professor Sheryl Tsai for giving me the opportunity to work in her lab throughout my years at UCI. She welcomed me as an undergraduate researcher and later offered me a research assistant position during my gap year after graduation. During the most difficult period of COVID, she supported me unconditionally—from language exams to program applications—and gave me the chance to pursue my PhD in her lab. She also supported my internship during my fourth year, which became a major turning point in my career. None of the work I completed over these five years would have been possible without her guidance, trust, and support.

I would also like to sincerely thank my committee members, Professors Ray Luo and Yilin Hu. Professor Hu's lab was the first research environment I experienced as a freshman, and that opportunity was what inspired me to pursue a PhD in biology. Her advice during my preliminary exam and advancement was invaluable, and her support throughout this dissertation greatly strengthened my work. Professor Luo introduced me to computational biology and encouraged me to explore a new direction in my scientific development. From the moment I met him in his class, he welcomed me into his lab and generously shared his expertise. His mentorship—whether in MD simulations, research ideas, or career advice—has been constant and deeply appreciated. I am truly grateful to all the members of my committee for their support during my PhD journey.

I would also like to thank Dr. Jacob Wolff, Dr. Anjani Sharma, and Dr. Gabriel Moreno. Jacob and Gabe were incredible mentors during my first three years. Jacob was always there to discuss experiments, troubleshoot with me, and cover for me whenever issues arose, truly like an older brother in the lab. Gabe, calm and wise, could solve any experimental or equipment problem, and continued to offer help even after leaving the group. Anjani joined the Tsai lab at almost the same time as I did; she has been my comrade through RIP talks, advancement, lab moves, and countless challenges. Her encouragement over the past two years gave me the strength to keep going even after repeated failures. I truly could not have started or completed this thesis without them.

I am also grateful to my mentees—Bingyu (Melanie) Zou, Chi Nguyen, Joyce Chen, Rui Zhou, and Melody Lee—whose help with experiments was essential for my research. I would like to thank the Tsai lab alumni as well: Dr. David Jackson, Dr. Jesus Barajas, Dr. Ming Lee, Dr. Joel Bruegger, Dr. Jake Milligan, Dr. Tyler Korman, and Dr. Tim Valentic. Their guidance on job searching and career development greatly shaped my path and helped me secure my internship during my fourth year.

To my friends—Dr. Shiji Zhao, Dr. Haixin Wei, Dr. Lequan Wang, Master Dahai Hao, Dr. Yaoyu Cao and Dr.Mingyu Du—thank you for being with me since my first year. From weekend gatherings to troubleshooting research problems, your support made my PhD life far more enjoyable. Shiji and Haixin helped me tremendously when I was in the Luo lab, while Lequan and Dahai guided me when I began learning machine learning. I am also grateful for the

friendship of Elsa Tseng, Yifan Zhang, Wenbo Yang, Honglei Wang, Jie Di, Evelyn Pu, and Benson Wu. Thank you for all the cooking parties, snowboarding trips, and activities that kept me sane and happy during the toughest times. I also want to thank Xiaoxin Yu, the first person I met at UCI. Her support during my first year and her introduction to the Hu/Ribbe Lab were instrumental in starting my research journey.

Finally, I want to express my heartfelt gratitude to my parents and to my girlfriend and life partner, Fei Xiao. My parents supported my decision to study at UC Irvine and in the United States, and their financial and emotional support sustained me throughout these years. There is no way to repay their love, but I hope to continue making them proud. Fei has been my strongest support through every setback and doubt. Whenever experiments failed or I questioned myself, she encouraged me, comforted me, and reminded me of my worth. She cooked for me, surprised me with gifts, and made every difficult day easier. Her presence made me feel at home even when I was far from my family. Her love and support have been irreplaceable, and I am deeply grateful to have her by my side.

Vita

Education

Ph.D., Biological Sciences — Sept. 2020–Dec. 2025
University of California, Irvine

B.S., Biological Sciences — Sept. 2016-Sept. 2019
University of California, Irvine

Research Experience

Graduate Student: University of California, Irvine — July 2020-Dec. 2025
Adviser: Professor Shiou-Chuan (Sheryl) Tsai, Ph.D.
Polyketide synthase structural biology

Co-op, Purification Process Development: Sanofi — June 2024-Sept. 2024
Biologics downstream monoclonal antibody purification process development

Undergraduate Researcher**:** University of California, Irvine — Sept. 2018-June 2020
Adviser: Professor Shiou-Chuan (Sheryl) Tsai, Ph.D.
Polyketide synthase structural biology

Teaching Experience

Teaching Assistant: University of California, Irvine
- Biochemistry laboratory (7 quarters)
- Biochemistry lecture (5 quarters)
- Molecular Biology lecture (2 quarters)
- Molecular Biology laboratory (1 quarter)
- General Microbiology lecture (1 quarter)
- Protein Structure and Function (1 quarter)

Publications

Pre-publication

Wolff JT, Ni F, Tsai SC. Novel SARS-CoV-2 inhibitory candidates identified from natural product analogs and FDA-approved natural products by molecular docking. Submitted to *Journal of Medicinal Chemistry*.

Sharma A, Wolff JT, Ni F, Tsai SC. Elucidation of molecular determinants of priming unit specificity in type II PKS priming ketosynthases. (2025) *Manuscript in preparation*

Published

Zhao, S., Ni, F., Qiu, T., Wolff, J. T., Tsai, S.-C., & Luo, R. (2020). Molecular Basis for Polyketide Ketoreductase–Substrate Interactions. *International Journal of Molecular Sciences*, *21*(20), Article 20.

Certificates

Management beyond the classroom certificate program — Spring Quarter 2021
University of California, Irvine

ABSTRACT OF THE DISSERTATION

Decoding Polyketide Synthase Specificity through Structural Informatics and Biochemical Reconstitution

by

Fanglue Ni

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2025

Professor Shiou-Chuan (Sheryl) Tsai, Chair

Natural products have a long history of use in traditional and modern medicine due to their inherent bioactivity. Some medicinal activities include antibiotic, antifungal, anticancer, antiviral, antihypercholestrolemic, and immunosuppressant. One of the largest classes of bioactive natural products are polyketides, produced by polyketide synthases (PKS). PKS are closely related to fatty acid synthases (FAS), sharing core biosynthetic logic that uses simple precursor molecules, but there are key differences in PKS that lead to the incredible diversity of structure observed in nature. Polyketide biosynthesis can be grouped into initiation, extension, reduction, aromatization and cyclization, and tailoring steps. Changes at each step have the potential to produce many variations in structure and this has sparked incredible interest in studying PKS chemistry to engineer these systems to produce novel medicinal compounds. Early engineering attempts mixed PKS components which could produce new polyketides but suffered from reduced yields and low fidelity. Furthermore, it was common for some combinations to not work at all, indicating that more investigation is necessary to improve engineering success.

Three major challenges are addressed in this dissertation span natural product drug discovery and PKS biosynthesis. The first challenge is COVID-19's ongoing threat to human health. While COVID-19 does not have the case fatality rate of its predecessors, SARS and MERS, its ability to spread quickly led to global shutdowns. Currently, several effective vaccines are available but small molecule drugs, while maintaining effectiveness across multiple viral strains, are limited in scope. Therefore, there is a pressing need to find new compounds that effectively inhibit SARS-CoV-2 infection. The intrinsic bioactivity displayed by natural products makes them an excellent source for new drugs therefore, molecular docking was employed (chapter 2) to identify promising compounds for further study from a library of over 200.

The second two challenges are from polyketide biosynthesis at the initiation and reduction steps. Initiation begins polyketide biosynthesis, and one or more enzymes can be involved to select a particular chemical moiety as a polyketide's first building block. The most common enzyme is the priming ketosynthase, but the molecular factors that determine its selectivity for the first building block remain unclear. At the reduction step, the substrate is a long, poly-β-ketone chain that is prone to spontaneous cyclization. This precludes structurally characterizing the ketoreductase responsible for regiospecific reduction at the ninth carbon and C7-C12 first ring cyclization with the native substrate. To elucidate initiation selectivity, three structures of the priming ketosynthase ZhuH from the R1128 system were solved with different acyl groups bound to the active site cysteine. How these acyl chains interact with active site residues elucidated the role of ZhuH in selecting its building block (chapter 3). Substrate reactivity at the reduction step was overcome by co-crystallizing the actinorhodin ketoreductase with two generations of substrate mimic. The resulting structures provided the first insight into how the substrate is positioned in the active site (chapter 4).

Chapter 1

Natural product impact on medicine and polyketide biosynthetic logic

1.1 Natural product

Natural products (NPs) have played a central role in human society for millennia, spanning uses from spices and dyes to therapeutic agents. Written records of medicinal natural products—primarily derived from plants—can be traced back to 2600 B.C. in Mesopotamia[1]. Their importance continued throughout more recent history, exemplified by the use of *Digitalis purpurea* for dropsy in 1775, the isolation of morphine from *Papaver somniferum* in 1816, and the purification of quinine from *Cinchona spp.* in 1820 as a treatment for malaria[1,2]. Despite the long-standing reliance on traditional medicines, systematic drug discovery efforts focused specifically on isolating natural products did not take form until the mid-20th century. The discovery of penicillin by Alexander Fleming in 1928 represented a turning point, marking the first clear case of a natural product being extracted from its producing organism. Although penicillin initially received limited attention, renewed interest emerged in 1939 following the success of sulfa drugs[3]. Early demonstrations of penicillin's therapeutic efficacy in infected mice and in a critically ill patient further accelerated its development[3].

During World War II, the urgent need to scale production of *Penicillium notatum* cultures prompted major industrial involvement from companies including Merck, Pfizer, Squibb, and Lilly, ultimately enabling large-scale penicillin manufacture and dramatically reducing treatment costs[3]. This industrial success established the foundation for sustained corporate investment in natural product–based drug discovery, ushering in the "Golden Age" of natural products research from the 1940s through the 1970s[1,4,5].

This Golden Age was characterized by extensive screening campaigns in which soil samples from diverse environments were cultured, extracted, and tested for bioactivity using straightforward cell-death assays[4,5]. Compounds exhibiting activity were then purified and structurally characterized. This paradigm continued into the subsequent era of natural product discovery (1970s–1990s), which benefited from expanded biochemical assays, broader ecological sampling, assay miniaturization, and major advances in dereplication enabled by HPLC, MS, and NMR "fingerprints" that facilitated rapid recognition of known compounds[4–6]. The late 20th century also saw transformative developments in molecular biology. Advances in recombinant DNA technology enabled the first cloning of an entire biosynthetic gene cluster (BGC) from *Streptomyces coelicolor*, followed shortly by the generation of hybrid molecules through combining elements from three distinct BGCs[7,8]. Additional BGCs—including those for frenolicin, jadomycin, tetracenomycin, oxytetracycline, and doxorubicin—were subsequently cloned into more tractable hosts for heterologous expression[9–13]. These efforts culminated in the emergence of "combinatorial biosynthesis," in which modules from disparate BGCs were mixed and matched to generate novel compounds[14–24]. Although scientifically promising, industrial enthusiasm declined with the rise of combinatorial chemistry, which offered far higher throughput than traditional natural product workflows[1,4–6,25].

However, the initial promise of combinatorial chemistry proved overstated. Despite the screening of millions of synthetic compounds, only two *de novo* molecules reached commercialization between 1981 and 2019[5,26]. This limited success has been attributed largely to early library designs that lacked structural diversity and were heavily biased toward similar scaffolds[1]. Subsequent redesign of libraries, guided by natural product-like structural complexity, has improved discovery outcomes[1,27,28]. Nonetheless, these shortcomings—and the inherently low biological relevance of many synthetic compounds—have contributed to the

renewed emphasis on natural products as privileged scaffolds shaped by evolutionary selection[29]. Concurrently, dramatic reductions in DNA sequencing costs and the surge of available microbial genomes have reinvigorated interest in natural products, particularly through mining metagenomic datasets for previously unknown BGCs[29,30].

Modern genome-based discovery has accelerated rapidly with the development of computational tools capable of identifying and predicting BGC function[31]. AntiSMASH, now in its 7th major release, remains the most widely used platform for rule-based annotation of BGCs, supported by its extensive accompanying database[32,33]. Parallel advances in machine learning have produced more flexible and generalizable predictive frameworks[34]. Databases such as MIBiG, DoBISCUIT, and the large but less curated IMG-ABC collection—which contains more than 330,000 BGCs—further support genome-driven natural product discovery[35–37]. These resources expand the search space beyond the small fraction of culturable organisms, enabling systematic exploration of the genetic diversity underlying natural product biosynthesis.

Even with these advances, genome-mined BGCs represent only one of three major natural product reservoirs: (i) gene clusters that are identified computationally, (ii) BGCs from organisms whose genomes remain unsequenced, and (iii) silent or cryptic BGCs in known organisms that remain transcriptionally inactive under standard laboratory conditions[38,39]. Recent improvements in synthetic biology have enhanced access to all three reservoirs. DNA engineering approaches—including homologous recombination, full-cluster synthesis, promoter refactoring, global and pathway-specific regulator engineering, and strategic deletion of competing pathways—have enabled more robust and tunable expression of BGCs in heterologous hosts[40,41]. Advances in culturing methods, extraction workflows, and high-

throughput metabolomics further streamline compound detection, dereplication, and functional screening[6,41–44].

From 1981 to 2019, natural products and their derivatives accounted for nearly 70% of all newly approved small-molecule therapeutics[26]. These include unmodified natural products, semisynthetic derivatives, synthetic molecules inspired by natural product pharmacophores, and natural product mimics. Recently approved natural-product–derived drugs include Eribulin(Metastatic breast cancer), Trabectedin(Soft tissue sarcoma), Romidepsin(Cutaneous and peripheral T-cell lymphoma), Fingolimod(Multiple sclerosis), and Lovastatin(Hypercholesterolemia) (Table 1.1). Their therapeutic applications span antibacterial, antifungal, anticancer, antiviral, antimalarial, and immunosuppressive indications—core pillars of modern medicine.

Given their historical success, broad biological relevance, and the rapidly expanding genomic and synthetic tools available, natural products remain an essential—and continually growing—source of therapeutic innovation. With vast chemical space still unexplored, the future of natural product research is poised to deliver an even deeper impact on human health.

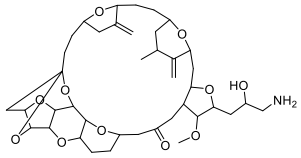
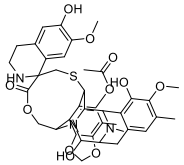
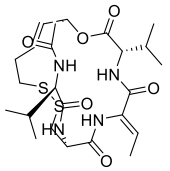
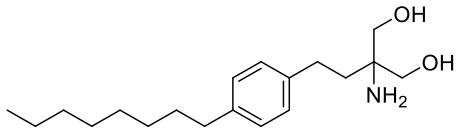
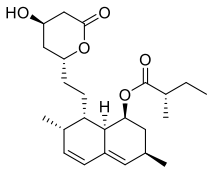
Drug	Source	Clinical Use	Structure
Eribulin	Synthetic analog of halichondrin B	Metastatic breast cancer	
Trabectedin	*Ecteinascidia turbinata*	Soft tissue sarcoma	
Romidepsin	*Chromobacterium violaceum*	Cutaneous and peripheral T-cell lymphoma	
Fingolimod	*Isaria sinclairii*	Multiple sclerosis	
Lovastatin	*Aspergillus terreus*	Hypercholesterolemia	

Table 1.1*: Recent drugs that are natural products or derived from natural products.*

1.2 PKS Architectures and Biosynthetic Overview

Polyketide synthases (PKSs) display three major architectural classes—type I, type II, and type III—each defined by distinct organizational and mechanistic features (Figure 1.1). Type I PKSs are large, multifunctional polypeptides composed of covalently linked catalytic domains. They are subdivided into modular and iterative forms. Type I iterative PKSs contain a single module in which each domain may act multiple times during chain extension, whereas type I modular PKSs contain multiple sequential modules, with each domain typically acting once per module[45–49]. In contrast, type II PKSs are dissociated systems composed of mono- or di-domain enzymes that assemble transiently during biosynthesis1[50]. The final architecture, type III PKSs, consists of a homodimeric ketosynthase (KS) capable of catalyzing chain initiation, extension, and cyclization in a single active site without the need for accessory domains[51,52].

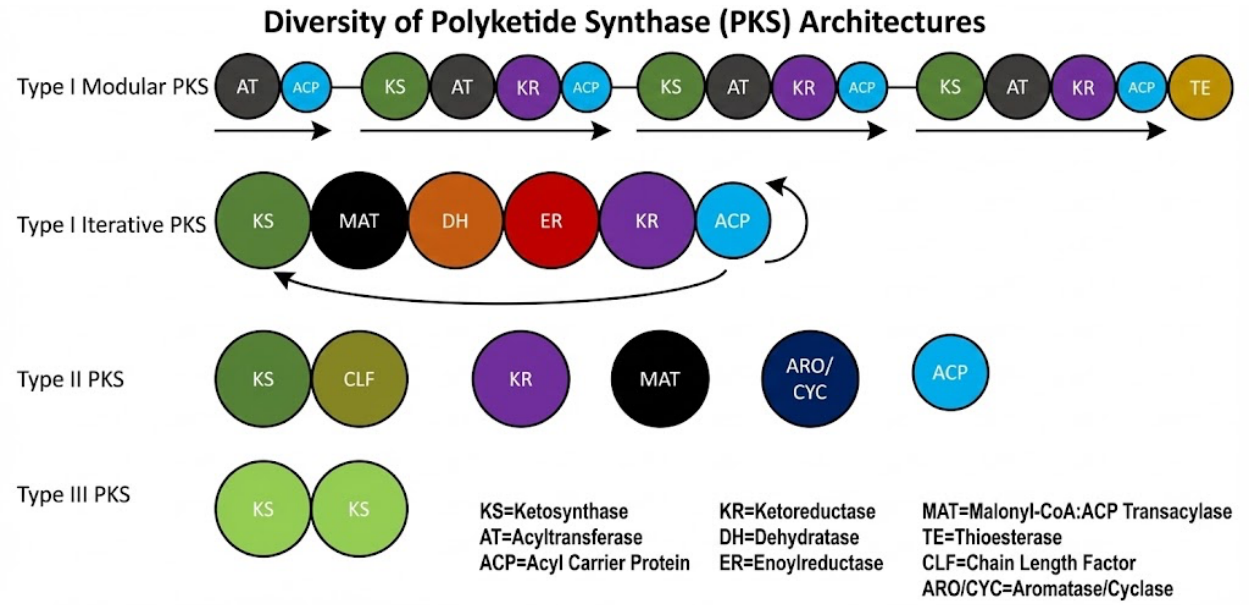


Figure 1.1. *Cartoon representation of the major polyketide synthase (PKS) architectures. PKSs are classified into three distinct architectural types: type I modular, type I iterative, type II, and type III systems. Type I PKSs consist of multiple catalytic domains located on a single polypeptide chain. In type I modular PKSs, each module contains the full set of domains required for one cycle of chain extension and modification before the intermediate is passed to the next module. In contrast, type I iterative PKSs contain domains resembling a single module, but these domains act multiple times during synthesis, and the rules governing the timing and order of domain use remain incompletely understood. Type II PKSs are composed of individual mono- or didomain proteins that assemble non-covalently into functional complexes to carry out polyketide chain initiation and elongation. Type III PKSs function as homodimeric ketosynthases, where a single active site catalyzes substrate loading, chain extension, and cyclization.*

Polyketide biosynthesis begins with initiation, a decarboxylative Claisen condensation between a priming unit and an extender unit[53–56]. While diverse starter molecules can be used, acetate is the most common priming substrate, and malonate is the predominant extender unit[55,57–59]. Malonyl loading is mechanistically conserved across PKS systems: a malonyl-CoA:ACP transferase (MAT/MCAT) or acyltransferase (AT) transfers malonyl-CoA onto a phosphopantetheine-modified acyl carrier protein (ACP), which delivers the extender unit to the KS for condensation (Figure 1.2).

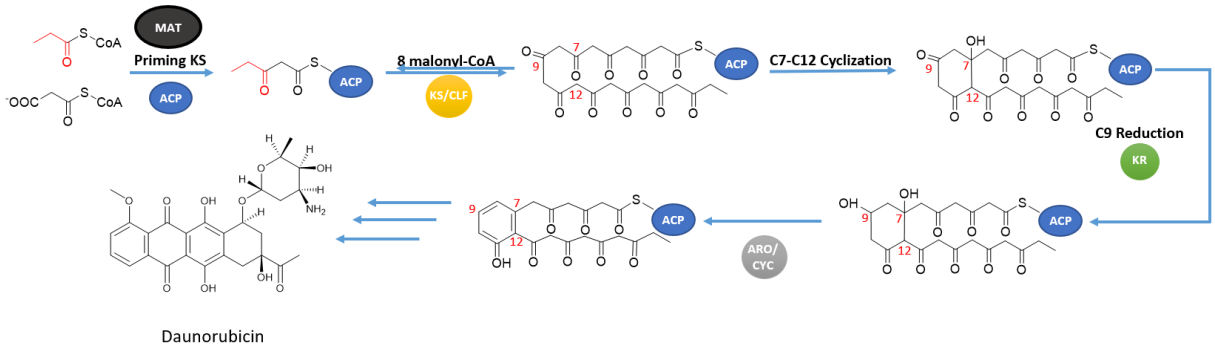


Figure 1.2. *Overview of key reactions in PKS biosynthesis with a focus on type II PKS.*

Initiation strategies vary across PKS architectures. In type II PKSs, chain initiation may be catalyzed directly by the KS/CLF heterodimer, as seen in actinorhodin and tetracenomycin biosynthesis, or by a discrete priming KS homodimer, as in frenolicin and R1128 biosynthesis, where non-acetate primers are incorporated[9,53,60,61]. Type I PKSs generally lack a dedicated priming KS. Instead, modular systems rely on a specialized loading AT/ACP to deliver the starter unit to the first module132, while type I iterative PKSs use a starter-unit:ACP transacylase (SAT) domain to select and load the priming substrate[56,62].

After initiation, the biosynthetic logic diverges according to architecture, although the core reactions—condensation, optional reduction, aromatization, and cyclization—remain conserved. Each condensation introduces a β-ketone and an α-methyl group to the growing chain. In highly reducing (HR) type I iterative PKSs, this β-ketone may undergo varying levels of reduction, depending on kinetic partitioning among KR, DH, and ER activities[63]. Type I modular PKSs also generate partially reduced intermediates, but the presence or absence of reducing domains within each module dictates the reduction pattern[46,64]. In contrast, non-reducing (NR) type I iterative and type II PKSs extend the chain repeatedly without β-keto reduction until the KS or KS/CLF defines the final chain length[26,65,66]. This produces a highly

reactive poly-β-keto intermediate that can spontaneously cyclize or rearrange, giving rise to substantial structural complexity[16,26,65,67–69]. Cyclization may involve aldol condensations, Claisen rearrangements, or even Diels–Alder cycloadditions within the growing chain[70,71].

Cyclization logic is a major determinant of structural diversity in PKS systems. Type I modular PKSs typically release and cyclize intermediates through TE-catalyzed lactonization, as exemplified by 6-deoxyerythronolide B, or through enzyme-catalyzed Diels–Alder reactions in specialized cases such as spinosyn[46,64,71]. Type I iterative PKSs employ different mechanisms depending on their reductive class: HR systems can catalyze intramolecular Diels–Alder reactions, as reported in lovastatin, equisetin, and solanapyrone, whereas NR systems typically use TE-mediated Dieckmann condensation for chain release[65,70,72–74]. NR type I iterative PKSs contain a product template (PT) domain that precisely orients the poly-β-keto intermediate to direct regio- and stereospecific cyclizations[65,75]. This is functionally analogous to the aromatase/cyclase (ARO/CYC) enzymes that govern ring formation in type II PKSs, including the actinorhodin pathway[20,76–79].

Following carbon-skeleton formation, polyketides frequently undergo extensive tailoring modifications, which expand chemical diversity and biological activity. These transformations may include glycosylation, methylation, epoxidation, oxidative rearrangement, halogenation, and further reductions[80]. Because tailoring reactions vary widely among PKS pathways, they are not discussed in detail here; numerous comprehensive reviews describe these steps extensively[81–84]. Subsequent sections will instead focus on the core biosynthetic logic of type II PKSs, emphasizing mechanistic diversity, structural determinants, and engineering potential.

1.2.1 Biosynthetic Logic: Elongation

Elongation constitutes the central chemical logic that underpins polyketide assembly, wherein malonyl-ACP is iteratively condensed with the growing acyl intermediate. In type II PKS systems, this process is mediated by the minimal PKS, composed of KS/CLF, ACP, and MAT (Figure 1.2)[85]. Successive rounds of decarboxylative Claisen condensation generate a highly reactive poly-β-keto intermediate that readily undergoes spontaneous cyclization, a phenomenon demonstrated through both in vivo and in vitro reconstitution studies[22,67–69,86,87].

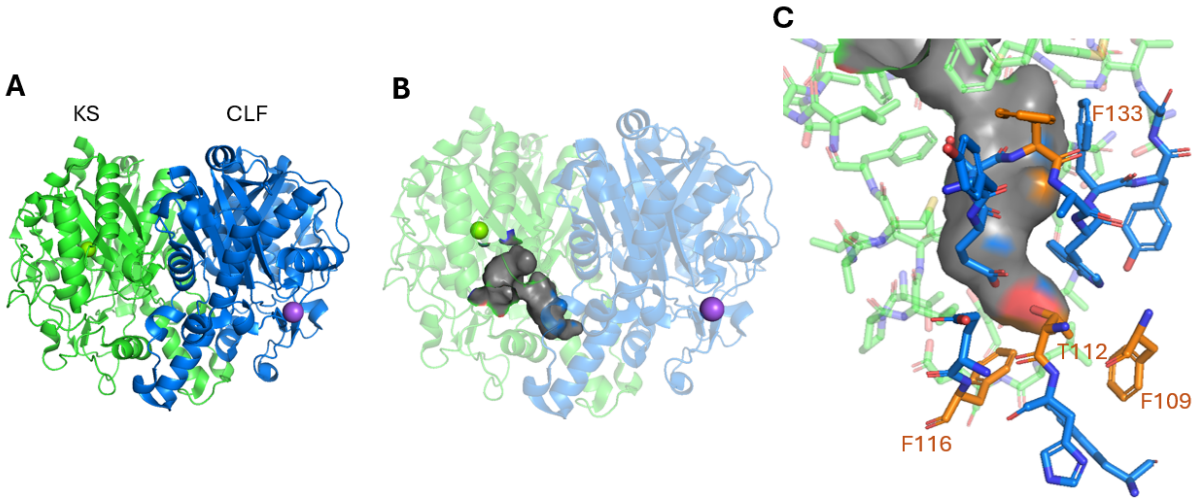


Figure 1.3. *A) Structure of the actinorhodin KS/CLF (green/marine) B) Structure of the actinorhodin KS/CLF (green/marine) with the amphipathic tunnel defining polyketide chain length in gray. C) Second view of ActKS/CLF showing the tunnel defining polyketide chain length. The chain length control residues Phe 109, Tyr 112, Phe 116 and Phe 133 shown in orange with label.*

The elongating ketosynthase is a heterodimeric KS/CLF complex **(Figure 1.3A)**, in which the KS subunit provides catalytic function while the CLF subunit, although lacking catalytic residues, is essential for structural stabilization and productive chain extension[60,88–90]. Although rare cases of KS/CLF subunit exchange have been observed, CLF partners are generally not interchangeable across systems, highlighting strict specificity at the heterodimer interface[91,92]. The KS subunit employs a Cys-His-His catalytic triad, together with a conserved

lysine residue, all of which are required for efficient chain extension[93]. Despite architectural differences relative to priming ketosynthases, the underlying chemical mechanism is conserved: decarboxylation of malonyl-ACP produces an enolate nucleophile that undergoes Claisen condensation with the KS-bound acyl intermediate[130]. Repeated condensations elongate the polyketide chain within a tunnel formed at the KS/CLF interface **(Figure 1.3B,C)**[90,94,95]. Chain-length determination is primarily governed by the CLF subunit, as shown by mutational studies capable of shifting polyketide length and by observations that extended priming units still yield products of native chain length[66,96]. Structural characterization of KS/CLF complexes remains sparse due to expression challenges; only three crystal structures have been reported, each displaying a conserved KS fold and dimeric interface but notable variation in CLF size, spanning from the 383-residue actinorhodin CLF to the 211-residue aryl-polyene CLF[90,94,95].

The ACP is a small, ~90-residue four-helix bundle bearing a conserved serine residue that is post-translationally modified with a 4′-phosphopantetheine arm [97–100]. This prosthetic group forms thioester linkages with malonyl and polyketide intermediates, enabling their controlled trafficking among PKS enzymes. ACP is indispensable for product formation, as no polyketide is generated in its absence[101]. Structural and biochemical studies have revealed a predominantly acidic ACP surface, whereas partner enzymes—including KS/CLF and MAT—present electropositive regions near their active-site entrances [45,102–106]. These observations support electrostatic complementarity as a major determinant of ACP–enzyme recognition. The partial interchangeability of ACPs across systems further supports this model, though many ACP–partner combinations remain incompatible[101,107–109]. ACP also contributes to intermediate protection.

The third component of the minimal PKS, the malonyl-CoA:ACP transacylase (MAT), catalyzes the transfer of malonyl groups onto ACP. Notably, MAT is not strictly required, as

type II ACPs can undergo self-malonylation[110–112]. Most type II PKS gene clusters do not encode a dedicated MAT, and polyketide production remains functional in heterologous hosts, indicating that the endogenous fatty acid synthase MAT typically provides malonyl transfer[85]. When cluster-encoded acyltransferases are present, they often fulfill specialized roles distinct from malonyl transfer, as exemplified by ZhuC, EncL, and GilQ, each of which displays poor malonylation activity relative to its native substrate[113–115]. A solved structure for the *S. coelicolor* MAT from the actinorhodin minimal PKS shows the canonical acyltransferase fold [116]. Attempts to alter MAT specificity toward methylmalonyl-CoA—an extender unit common in type I PKSs—have been unsuccessful, with malonyl-CoA remaining strongly preferred[110]. Given the essential role of MAT in fatty acid metabolism, engineering of MAT is further constrained by physiological limitations that likely contribute to the near-universal use of malonyl-CoA as the extender unit in type II PKSs[59].

Engineering of the elongation machinery has thus proven challenging. Mix-and-match strategies targeting KS/CLF subunits have generally failed due to interface incompatibility, while mutational modifications of elongation-tunnel residues in actinorhodin and tetracenomycin systems have produced altered chain lengths, such as decaketides in normally octaketide-producing systems[66]. However, substantial quantities of native-length products typically persist. Extensive mutagenesis of the fredericamycin KS/CLF (48 variants) similarly failed to generate new chain lengths, yielding only wild-type products[117]. These findings underscore that chain-length determination arises from the interplay of structural constraints, reaction kinetics, and ACP–KS interaction dynamics. Engineering of ACP has largely involved helix-swap studies between incompatible ACPs, such as actinorhodin ACP and E. coli ACP[228]; although replacement of helix II enabled weak binding to E. coli FabF, overall affinity was reduced seven- to ten-fold, highlighting the distributed and dynamic nature of ACP–partner

recognition[105]. MAT engineering has been attempted by various groups and was similarly unsuccessful[110]. Collectively, these studies indicate that the elongation machinery of type II PKS systems is shaped by tightly integrated structural, electrostatic, and kinetic parameters. As a consequence, rational engineering of these components remains difficult and requires consideration of multiple layers of molecular specificity beyond primary enzyme structure.

1.2.2 Biosynthetic Logic: Ketoreduction

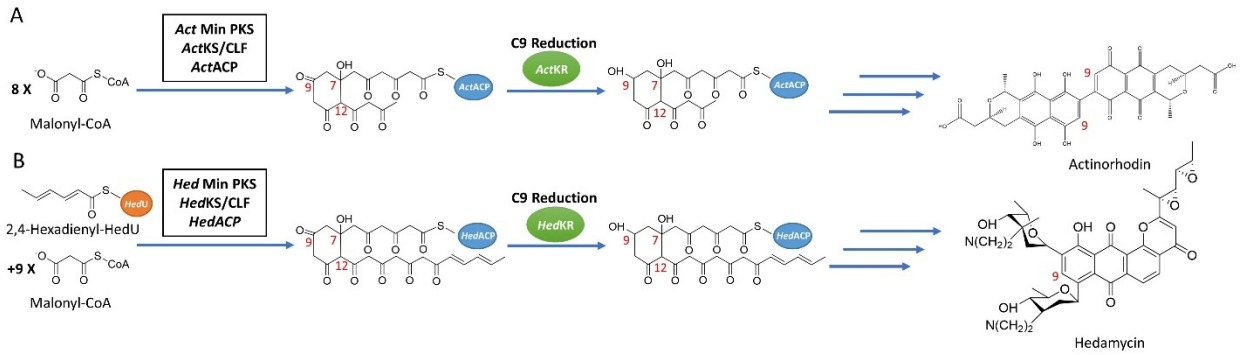


Figure 1.4. *Structure The ketoreduction pathway of type II PKSs, using actinorhodin PKS(A) and hedamycin PKS(B) as an example.*

Following full extension of the poly-β-ketone intermediate by the minimal PKS, many type II pathways recruit an NADPH-dependent ketoreductase (KR) to reduce the C9 carbonyl to an alcohol, when such an enzyme is present **(Figure 1.4AB)**[10,109,118–121]. These C9-reducing KRs adopt a classical Rossmann-fold architecture, consisting of a twisted parallel β-sheet flanked by α-helices, in which NADPH binds to a conserved TGxxxGxG motif located at the C-terminal ends of the β-strands **(Figure 1.5A, 1.5B)**[120,122–124]. The catalytic machinery comprises a Tyr-Lys-Ser-Asn tetrad that mediates hydride transfer and subsequent proton release to bulk solvent, with crystallographic waters supporting the proposed mechanism[123,125]. The active site cleft is formed by residues on helices surrounding the central β-sheet and features two α-helices (6–7) that create a lid-like structure hinged for conformational mobility **(Figure 1.5A).** Despite this

detailed structural information, the mode of substrate entry, positioning, and orientation within the KR active site remains unresolved, especially given that C9 reduction occurs only after the full-length poly-β-ketone has been generated by the minimal PKS. This obscurity also complicates understanding of KR substrate specificity, which appears highly variable among homologs. For example, the actinorhodin KR (ActKR) reduces poly-β-ketones spanning hexaketides to dodecaketides, whereas the hedamycin KR (HedKR/HedE) acts on tetra-, undeca-, and dodecaketides[14,19,78,109,126,127]. The daunorubicin C9 KR DpsE, in contrast, cannot reduce octaketides generated by the actinorhodin minimal PKS but successfully reduces decaketides produced by the daunorubicin PKS or by the tetracenomycin minimal PKS[24]. Moreover, C9 reduction is tightly associated with C7–C12 first-ring cyclization, although the mechanistic basis for this coupling remains unclear[16,128–131].

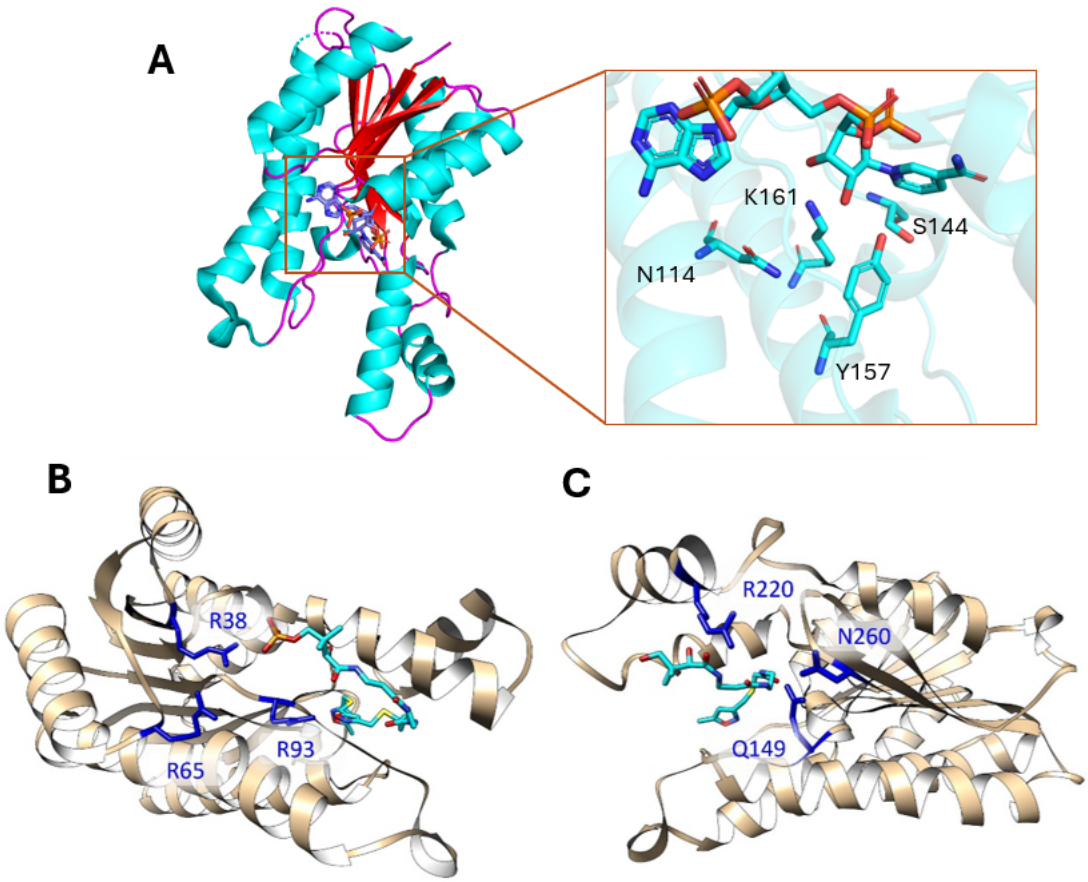


Figure 1.5. *A) Side view of the actinorhodin ketoreuctase (ActKR). α-helices are colored cyan, β-strands red, and loops purple. NADPH is represented as spheres with carbons colored gray. Asn-Ser-Tyr-Lys form the catalytic tetrad of ActKR. The open conformation is depicted in this image. (B) DM-ActKR co-crystal structure bound with the octaketide-phosphopantetheine mimic identified front-patch, formed by R38, R65, R93. (C) DM-ActKR co-crystal structure bound with the tetraketide-pantetheine mimic identified back-patch, formed by Q149, R220, N260. The mimics are displayed in cyan, front and back-patched are displayed in blue.*

Much of the current mechanistic understanding derives from studies of ActKR. ActKR displays strong preference for bicyclic substrates, with trans-1-decalone and α-tetralone serving as productive substrates *in vitro*, indicating that the initially linear poly-β-keto intermediate undergoes spontaneous cyclization prior to C9 reduction[132]. Earlier docking studies further supported this model, showing that in a C7–C12 cyclized intermediate, the C9 carbonyl is optimally positioned relative to the nicotinamide ring for hydride transfer[123].

Stereochemical preferences of ActKR and HedKR have been interrogated using α-tetralol oxidation assays, revealing a ~3:1 preference for S-(+)-tetralol in ActKR and strict S-(+)-selectivity in HedKR[120,133]. Combining these experimental data with insights from type I PKS KR stereochemical motifs, extensive mutagenesis, and in vitro reconstitution produced a sophisticated mechanistic model describing ACP docking, substrate positioning, C7–C12 cyclization, and C9 reduction[133,134]. In this model, the negatively charged ACP binds first to a positively charged patch near the active site, comprising residues R38, R65, and R93 **(Figure 1.5B)**. The linear intermediate then enters the active site beneath "gating" residues P94 and M194 and is oriented toward NADPH by hydrophobic steering residues V151, F189, and V198. Cyclization is enabled by hydrogen bonding between T145 and the C11 carbonyl, which depresses the C12 α-proton pK_a and facilitates intramolecular aldol condensation. The resulting bicyclic intermediate positions the C9 carbonyl closest to NADPH, yielding an S-configured alcohol upon hydride transfer.

In the actinorhodin pathway, the C9 alcohol is rapidly dehydrated to aromatize the first-ring system, precluding stereochemical assignment at C9. Disruption of the actinorhodin ARO/CYC enzymes prevents dehydration and yields mutactin; however, multiple analytical approaches have been unsuccessful in determining its absolute stereochemistry[133,135]. Uncertainty is further compounded by the lack of direct experimental evidence for ACP–KR binding orientation. Two competing models propose that ACP binds either to the positively charged "front" patch (R38/R65/R93) or to the "back" side, which includes residues near the lid helices or another polar cluster (Q149, R220, N260)[123,129,136]. The most comprehensive study to date combined protein–protein docking, NMR, and molecular dynamics using an ACP-bound C7–C12 cyclized intermediate, finding overwhelming preference for S-favored poses. Mutational analysis further demonstrated that perturbing front-patch residues weakened ACP

binding and altered stereochemical output toward S-(+)-tetralol, strongly supporting the front-binding model[264]. Nonetheless, the transient nature of ACP–KR interactions and the instability of native substrates continue to impede direct structural characterization.

A second major class of ketoreductases, distinct from the C9-reducing ACP-bound enzymes, acts on polyketide intermediates only after chain release and cyclization[137]. These tailoring KRs catalyze reductions at C15, C17, C19, or C21 and can be clearly distinguished from C9 KRs by sequence clustering[138–141]. Although they share the same Rossmann-fold core, these tailoring KRs possess a markedly shorter lid region [142–145]. Owing to the reduced reactivity of their substrates, co-crystal structures with substrate analogs have been obtained and reveal active-site positioning similar to that observed in emodin-bound ActKR structures[132,142,145]. These structures have facilitated rational engineering attempts aimed at altering stereospecificity. For example, iterative region-swapping between LanV (producer of 11-deoxylandomycinone) and CabV (producer of gaudimycin C) identified four regions—R1 (residues 102–118), R2 (154–164), R3 (single residue S192), and R4 (residues 198–210)—as key contributors to stereochemical outcome[138,146,147]. Among these, swaps at R3 had the strongest individual impact, while simultaneous exchange of all four regions produced the greatest product shift[143]. Although these swaps inverted the stereochemical ratio of the native products, no truly non-native stereochemical outcomes were observed, likely reflecting conformational differences caused by differential A-ring aromatization during substrate processing.

Additional mutational studies in both DpsE and ActKR further demonstrate the malleability of KR stereochemistry. In ActKR, comparison with type I KRs identified a PGG motif (residues 94–96) analogous to the LDD motif in B-type KRs[133,148]. Mutagenesis revealed that P94L shifts selectivity from a 3:1 mixture of S-(+)- to R-(-)-tetralol to exclusive S-(+)-tetralol production without impairing catalytic efficiency[134], while V151L produces strict R-(-)-

tetralol selectivity with improved catalytic efficiency[134]. Attempts to confirm these stereochemical swaps using native poly-β-ketone substrates were inconclusive due to limitations in chiral HPLC and NMR analyses[134]. Engineering of DpsE, motivated by its inability to reduce actinorhodin-derived octaketides, identified the lid region as a major determinant of substrate recognition. A lid-swap chimera and a more extensive chimera incorporating additional point mutations (G94P, A152V, L153H) both enabled mutactin formation in the actinorhodin minimal PKS system, with the lid swap contributing most strongly to functional rescue[149].

Collectively, studies of type II PKS KRs—both C9-reducing and tailoring classes—demonstrate that ketoreduction is governed by a finely tuned interplay of substrate architecture, ring-formation dynamics, ACP–enzyme interactions, and active-site conformational plasticity. While KR stereochemistry can be altered through mutagenesis and domain swapping, achieving predictable and generalizable changes remains challenging due to the dynamic, multistep nature of substrate processing within the polyketide assembly line.

1.2.3 Biosynthetic Logic in Type I Fungal PKS: Product Template Domain

The product template (PT) domain is a defining and functionally indispensable feature of fungal iterative, nonreducing type I polyketide synthases (NR-PKSs).[150] Although PT domains occupy a single, compact region within the multidomain megasynthase, they exert disproportionate control over the architecture of the final product by dictating the regioselective aldol cyclization of the highly reactive poly-β-ketone intermediate.[75] Given that NR-PKSs operate iteratively and do not follow the colinearity rules characteristic of bacterial modular PKSs[150], the PT domain serves as a major determinant of structural programming. Consequently, deciphering PT domain classification and understanding sequence–function

relationships has become a critical step toward predicting polyketide scaffolds from genomic data and enabling rational biosynthetic engineering[151].

Early biochemical and structural studies established that PT domains catalyze the first cyclization event following full chain extension[152,153]. Their characterized activities show that PTs suppress spontaneous cyclizations and enforce specific aldol condensations—typically C2–C7, C4–C9/C2–C11, or C6–C11—resulting in the canonical F-folding patterns observed in fungal aromatic polyketides[151,154]. Despite these insights, the diversity of cyclization modes and the rapidly growing number of uncharacterized NR-PKSs underscored the need for a broader, sequence-based classification framework.

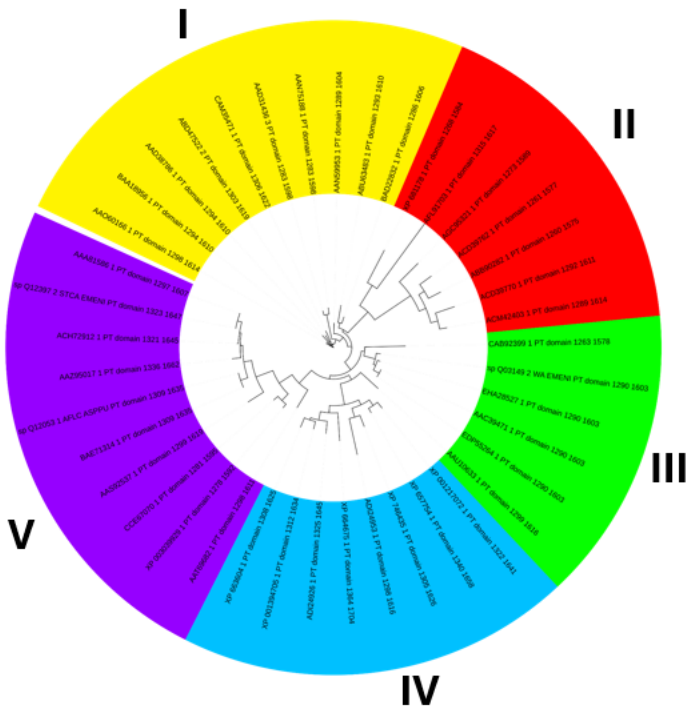


Figure 1.6. *Phylogenetic analysis of PT domains from 40 NRPKSs that have been related to known aromatic polyketides. The NRPKSs have been classified into five major groups: group I, C2-C7 monocyclic PKSs; group II, THN synthases(C2-C7 bicyclic PKSs); group III, C2-C7 multicyclic PKSs; group IV, C4-C9 PKSs; and group V, C6-C11 PKSs. The evolutionary distances are in the units of the number of amino acid substitutions per site. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.*

A major advance came from the work of Li, Xu, and Tang, who performed the first large-scale phylogenetic analysis of PT domain sequences[151]. By assembling PTs from NR-PKSs with known product structures and comparing their evolutionary relationships, they showed that PT domains cluster into five robust groups, each corresponding to a distinct cyclization pattern and product length class **(Figure 1.6)**. Groups I and II encompass PTs responsible for monocyclic and bicyclic aromatic polyketides synthesized from tetraketide and pentaketide intermediates, respectively. Group III comprises PTs involved in C2–C7 cyclization of longer polyketide chains, while Group IV contains the C4–C9/C2–C11 cyclases typified by the

aflatoxin pathway. Group V PT domains form a distinct clade associated with the C6–C11 cyclization mode and the biosynthesis of emodin-, asperthecin-, and atrochrysone-like scaffolds.

Beyond phylogenetics, Li et al. verified predictive accuracy by excising uncharacterized PT domains from native NR-PKSs and combining them with a heterologous minimal PKS to reconstruct early steps in cyclization[154]. These experiments demonstrated that the phylogenetic grouping correlates strongly with the regioselectivity observed in vitro, solidifying the use of sequence-based classification as a practical predictive tool for annotating the growing number of cryptic fungal biosynthetic gene clusters.

Subsequent research refined this classification framework by integrating structural comparisons and domain architecture analyses. A subtle sequence variation within conserved PT folds that influence active-site shape and substrate positioning was highlighted. By analyzing a broader distribution of PTs across sequenced fungal genomes, they confirmed that the five-group classification is stable even when hundreds of uncharacterized PTs are added.[154] Their work also demonstrated that PT domain evolution aligns strongly with ketide chain length, predicted folding mode, and the presence or absence of terminal off-loading domains such as TE/CLC modules. The resulting classification not only recapitulates known biochemical categories but also identifies outlier PT domains that may represent unexplored cyclization chemistries.

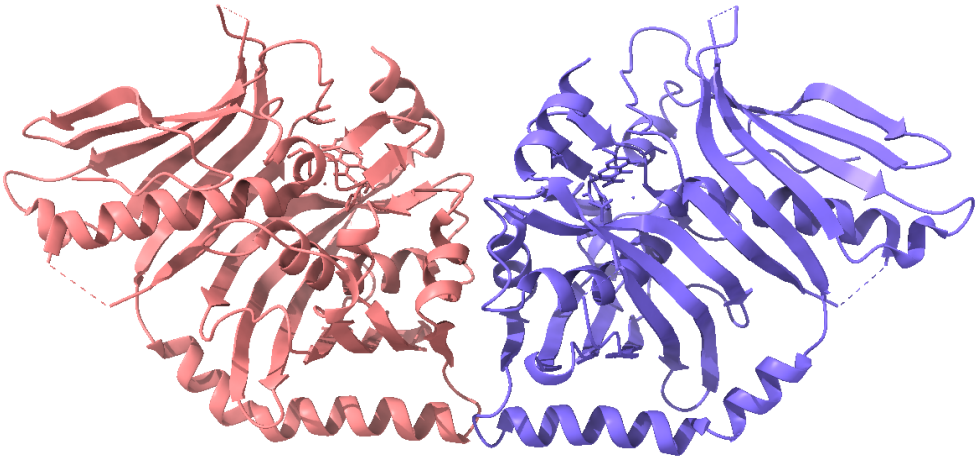


Figure 1.7. *Side view of product template domain crystal structure of PksA.*

Structural studies have greatly expanded understanding of how PT domains impose regioselectivity (**Figure 1.7**). The crystal structures of the PksA PT domain, solved in complex with palmitate, a bicyclic substrate mimic, and a bis-isoxazole heptaketide analog, revealed that PTs adopt a double hot-dog fold similar to the dehydratase domains of modular PKSs and fungal megasynthases[65,151–153,155]. In PksA, the two monomers dimerize via their C-terminal hot-dog folds rather than the N-terminal interface typical of DH domains. The resulting ~30 Å substrate tunnel is divided into three functional regions: a linear ~14 Å segment accommodating the phosphopantetheine arm; a central cyclization chamber containing the catalytic His1345–Asp1543 dyad; and a distal hydrophobic pocket that binds the hexyl starter unit. Structural analysis showed that substrate mimics bind in an extended conformation, with the pPant arm stabilized by R1623, while hydrophobic interactions with Y1492, L1508, and F1551 orient the polyketide chain such that C4 is positioned ~3 Å from His1345, enabling the regiospecific C4–C9 aldol cyclization[152,153]. These structures confirmed that PT domains use a

rigid tunnel architecture and strategically placed catalytic residues to enforce precise folding patterns, thereby controlling both chain length and cyclization regiochemistry.

More recently, the structure of the bacterial PT domain AviM revealed that not all PTs conform to the fungal canonical architecture[155]. Although AviM catalyzes a C2–C7 cyclization analogous to Group III fungal PTs, its overall fold, dimeric interface, and catalytic dyad more closely resemble those of modular PKS DH domains than fungal PTs. Phylogenetic analysis suggests that AviM represents an evolutionary intermediate linking ancestral dehydratases to the specialized PT domains of NR-PKSs. These structural comparisons indicate that regioselective aldol cyclization has arisen multiple times during polyketide evolution through repurposing and diversification of DH-like folds. The existence of this bacterial intermediate also suggests that active-site cavity geometry—not the broader domain scaffold—is the primary driver of cyclization regiochemistry.

Together, these studies have transformed the understanding of PT domain biology. The combination of phylogenetic grouping, structure-based analysis, and functional reconstitution has emerged as a powerful strategy for predicting NR-PKS product structures. As fungal genome sequencing continues to accelerate, PT domain classification provides one of the most reliable means of connecting uncharacterized NR-PKSs to putative metabolite scaffolds. This sequence-informed framework has already begun to influence genome mining efforts, enabling the prioritization of biosynthetic clusters likely to yield new aromatic polyketides and offering a rational foundation for future domain swapping and pathway engineering experiments.

1.3 Remaining Challenge

Polyketide synthase (PKS) research has progressed dramatically since the actinorhodin biosynthetic gene cluster was first cloned into a heterologous host and shown to produce actinorhodin[8]. The late 1980s and 1990s were dominated by gene-by-gene dissection of PKS clusters to define enzymatic roles, accompanied by early attempts at combinatorial biosynthesis through mixing modules from different systems. A major shift occurred in the early 2000s, when advances in recombinant DNA technology enabled a transition from predominantly in vivo studies to robust in vitro reconstitution of PKS components. Sequencing of the *Streptomyces coelicolor A3* genome revealed far more biosynthetic gene clusters than previously anticipated[38], catalyzing a deeper appreciation of the metabolic potential encoded in actinomycetes. During this period, key structural breakthroughs were achieved for the malonyl-CoA:ACP transacylase (MAT), priming ketosynthase, elongating KS/CLF heterodimer, C9-reducing ketoreductase, first-ring cyclase, and fourth-ring cyclases[76,90,116,123,129,156–159]. The first high-resolution structures of mammalian and bacterial type I fatty acid synthases further provided unprecedented insight into the architecture and conformational logic of megasynthases[48,49]. As sequencing costs fell, the number of available bacterial genomes—including those generated through metagenomic efforts—increased rapidly, driving the development of bioinformatic tools for biosynthetic gene cluster (BGC) prediction and comprehensive databases to store annotated clusters[31,32,36,37]. These resources transformed discovery pipelines by enabling researchers to identify unusual gene clusters with potentially novel enzymatic chemistry, make initial predictions about product structures, and heterologously express entire pathways in well-characterized hosts to access new natural

products. Today, the field benefits from a robust ecosystem of genetic, biochemical, structural, and computational tools that support PKS research at every level.

Despite these advances, a major challenge that remains unresolved is deciphering the molecular principles that govern carbon chain–length control within KS/CLF complexes. Although the CLF is known to set the elongation limit of the growing polyketide chain[50,57] and decades of biochemical work have established the importance of the KS/CLF tunnel in determining chain accommodation[17,109], the specific sequence and structural features that encode programmed chain length remain elusive. Existing mutational studies have produced only modest or inconsistent shifts in product length, underscoring the fact that chain-length determination arises from a complex interplay of steric constraints, electrostatic environment, acyl-chain positioning, and ACP–KS dynamics rather than from any single residue or motif. Furthermore, the small number of KS/CLF systems with experimentally validated chain lengths stands in stark contrast to the vast genomic diversity of type II PKS sequences now available, making it difficult to generalize insights gained from individual systems or manually align sequence variations with functional outcomes. As a result, efforts to rationally redesign KS/CLF pairs for custom chain lengths have been largely unsuccessful, and a comprehensive mechanistic model capable of predicting elongation limits across diverse PKS families remains beyond reach. Addressing this challenge requires methods capable of extracting high-dimensional sequence–function relationships from sparse and unevenly annotated datasets—motivating the need for machine learning approaches that transcend traditional sequence alignment and structural intuition.

A second enduring challenge lies in the fundamental problem of predicting and engineering ACP–KR compatibility in type II PKSs. Although ACPs and KRs often share high structural conservation and similar catalytic architectures, their productive interaction is

highly specific and not reliably interchangeable across pathways.[3–6] Decades of combinatorial biosynthesis demonstrate that simply transplanting a KR into a heterologous minimal PKS rarely results in efficient C9 reduction, as exemplified by the persistent inactivity of DpsKR outside its native doxorubicin context.[23,160] These failures underscore a deeper issue: the field still lacks a quantitative or predictive framework for determining which ACP–KR pairs will form productive complexes. Even with high-resolution structures of KRs and detailed biochemical characterizations of ACPs, the transient and dynamic nature of ACP–enzyme contacts have rendered systematic mapping of these interaction surfaces difficult. This limitation hampers rational pathway engineering, constrains modular swapping strategies, and restricts our ability to design type II PKSs with predictable tailoring outcomes. Overcoming this challenge will require the integration of structural modeling, big data approach, mutational scanning, and *in vitro* reconstitution to define the molecular rules that govern selective KR engagement and productive reduction.

A third major challenge is the persistent lack of a unified structural and mechanistic framework to explain how PT domains encode cyclization selectivity across the full diversity of fungal NR-PKSs. Although the PksA PT crystal structure provided the first glimpse of the double–hot-dog fold and its tripartite internal cavity[75], it remains unclear how generalizable this architecture is across the eight phylogenetic groups identified by Liu et al.[154]. Homology models generated prior to modern deep-learning methods offered valuable hypotheses regarding cavity-lining residues, loop rearrangements, and pocket volume differences[154], yet their limited accuracy has prevented definitive conclusions about how these features sculpt substrate conformations or regulate access to the catalytic Asp–His dyad. More fundamentally, the dynamic interplay between PT pocket geometry, polyketide chain length, and early folding trajectories is almost entirely uncharacterized. Most studies have inferred cyclization behavior

solely from final product structures or phylogenetic clustering[151,154] rather than through direct evaluation of linear versus pre-cyclized intermediates within accurate PT structural models. Compounding this problem is the near absence of experimental information on ACP–PT interactions, which likely influence substrate delivery and positioning but remain inaccessible due to the transient and flexible nature of ACP docking[161]. As a result, current mechanistic models of PT selectivity rely heavily on isolated case studies and cannot yet account for the full range of regioselectivities observed across fungal clades. Developing a rigorous, predictive understanding of how PT structural features encode cyclization logic will require integrating high-accuracy structural prediction, quantitative docking, and comparative modeling across large PT datasets—an effort only recently made feasible by tools such as AlphaFold3.

1.4 Dissertation Overview

Polyketides represent one of the most structurally diverse and pharmacologically valuable classes of natural products, encompassing antibiotics, anticancer agents, and other therapeutically relevant compounds. Their remarkable chemical diversity arises from the highly controlled and iterative enzymatic processes mediated by polyketide synthases (PKSs). Understanding the molecular determinants that govern chain elongation, cyclization, and tailoring of polyketide intermediates is essential for both elucidating natural biosynthetic logic and enabling rational engineering of novel bioactive molecules. This dissertation investigates three interconnected aspects of polyketide biosynthesis: chain-length specificity, ketoreductase–acyl carrier protein (KR–ACP) interactions, and cyclization regioselectivity of the product template(PT) domain, employing a combination of biochemical reconstitution, structural modeling, and machine learning approaches.

Chapter 2 focuses on the determinants of polyketide chain-length control in type II PKSs. Aromatic polyketide diversity is largely dictated by the ketosynthase/chain-length factor (KS/CLF) heterodimer, yet the specific sequence features that govern chain-length selection remain poorly understood. We employed a hybrid experimental and computational strategy, combining *in vitro* reconstitution of KS/CLF complexes with modern machine learning techniques. Classical sequence encodings and simple neural networks proved insufficient for capturing the subtle biochemical signals underlying chain-length specificity, whereas embeddings derived from the protein language model ESM-2 improved predictive accuracy. Nevertheless, unsupervised clustering and large-scale sequences analysis revealed that KS/CLF sequences are highly conserved, and that chain-length control likely depends on subtle residue-level and structural features that are not easily inferred from sequence alone. This

work highlights both the potential and the limitations of sequence-based machine learning for predicting PKS function, underscoring the need for integrative approaches incorporating structural and biochemical data.

Chapter 3 addresses the molecular basis of substrate specificity in type II PKS tailoring enzymes, focusing on the C9 ketoreductase (DpsKR) from the doxorubicin pathway. Type II KRs rely on transient, highly specific protein–protein interactions with ACPs to recognize and modify polyketide intermediates. Through in vitro reconstitution and structural modeling, we demonstrated that DpsKR activity depends critically on the structural features of the so-called "lid helices" (α6–α7) and on ACP identity. Swapping lid helices from ActKR into DpsKR restored productive C9 reduction, and ACP-swapping experiments further confirmed that compatibility between KR and ACP partners dictates enzymatic activity. These findings reveal that protein–protein interfaces, rather than active-site chemistry alone, govern substrate specificity in type II PKSs and provide a mechanistic framework for engineering combinatorial biosynthesis with improved fidelity.

Chapter 4 investigates the structural and mechanistic determinants of cyclization regioselectivity in fungal non-reducing PKSs (NR-PKSs). The product template (PT) domain directs the initial aromatization of polyketide intermediates, thereby establishing the core ring-connectivity patterns of fungal aromatic natural products. By integrating AlphaFold3-based structural predictions, extensive docking simulations, statistical modeling, and ACP–PT complex analysis, we identified three primary factors controlling regioselectivity: (i) the geometry of the catalytic dyad, (ii) pocket volume and polarity, and (iii) the orientation of a finger-like loop that modulates substrate accessibility. Moreover, PT–ACP interactions influence effective chain length during substrate delivery, highlighting the importance of protein context in determining cyclization outcomes. Collectively, these findings provide a

unified model for predicting fungal aromatic scaffolds from sequence and structure and offer insight into the evolutionary diversification of NR-PKS catalytic logic.

Taken together, the studies presented in this dissertation elucidate fundamental principles governing polyketide chain elongation, tailoring, and cyclization. By integrating biochemical, structural, and computational approaches, this work advances our understanding of PKS specificity at multiple levels, laying a foundation for rational engineering of polyketides with tailored structures and enhanced therapeutic potential.

1.5 References

(1) Cragg, G. M.; Newman, D. J. Biodiversity: A Continuing Source of Novel Drug Leads. *Pure Appl. Chem.* **2005**, *77*, 7–24.

(2) Buenz, E. J.; Verpoorte, R.; Bauer, B. A. The Ethnopharmacologic Contribution to Bioprospecting Natural Products. *Annu. Rev. Pharmacol. Toxicol.* **2018**, *58*, 509–530.

(3) Gaynes, R. The Discovery of Penicillin—New Insights after More than 75 Years of Clinical Use. *Emerg Infect Dis* **2017**, *23*, 849–853.

(4) Beutler, J. A. Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol.* **2019**, *86, e67*.

(5) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176.

(6) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

(7) Hopwood, D. A. Production of 'hybrid' Antibiotics by Genetic Engineering. *Nature* **1985**, *314*, 642–644.

(8) Malpartida, F.; Hopwood, D. A. Molecular Cloning of the Whole Biosynthetic Pathway of a Streptomyces Antibiotic and Its Expression in a Heterologous Host. *Nature* **1984**, *309* (5967), 462–464. https://doi.org/10.1038/309462a0.

(9) Bibb, M. J.; Sherman, D. H.; Ōmura, S.; Hopwood, D. A. Cloning, Sequencing and Deduced Functions of a Cluster of Streptomyces Genes Probably Encoding Biosynthesis of the Polyketide Antibiotic Frenolicin. *Gene* **1994**, *142*, 31–39.

(10) Han, L.; Yang, K.; Ramalingam, E.; Mosher, R. H.; Vining, L. C. Cloning and characterization of polyketide synthase genes for jadomycin b biosynthesis in streptomyces venezuelaeISP5230. *Microbiology* **1994**, *140*, 3379–3389.

(11) Motamedi, H.; Hutchinson, C. R. Cloning and Heterologous Expression of a Gene Cluster for the Biosynthesis of Tetracenomycin C, the Anthracycline Antitumor Antibiotic of Streptomyces Glaucescens. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4445–4449.

(12) Kim, E.-S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. Sequences of the Oxytetracycline Polyketide Synthase-Encoding Otc Genes from Streptomyces Rimosus. *Gene* **1994**, *141* (1), 141–142. https://doi.org/10.1016/0378-1119(94)90144-9.

(13) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. Characterization of the streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **1994**, *151*, 1–10.

(14) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Manipulation and Analysis of an Aromatic Polyketide Synthase with Unproved Catalytic Specificities. *J. Am. Chem. Soc.* **1993**, *115*, 11671–11675.

(15) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. The tcmVI Region of the Tetracenomycin c Biosynthetic Gene Cluster of Streptomyces Glaucescens Encodes the Tetracenomycin F1 Monooxygenase, Tetracenomycin F2 Cyclase, and, Most Likely, a Second Cyclase. *J. Bacteriol.* **1993**, *175*, 7571–7580.

(16) Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Stereochemical Course of Two Reactions Catalyzed by a Polyketide Synthase. *Biochemistry* **1994**, *33*, 9321–9326.

(17) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic

Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91* (24), 11542–11546. https://doi.org/10.1073/pnas.91.24.11542.
(18) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91*, 11542–11546.
(19) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Rational Design of Aromatic Polyketide Natural Products by Recombinant Assembly of Enzymatic Subunits. *Nature* **1995**, *375* (6532), 549–554. https://doi.org/10.1038/375549a0.
(20) Shen, B.; Summers, R. G.; Wendt-Pienkowski, E.; Hutchinson, C. R. The Streptomyces Glaucescens tcmKL Polyketide Synthase and tcmN Polyketide Cyclase Genes Govern the Size and Shape of Aromatic Polyketides. *J Am Chem Soc* **1995**, *117*, 6811–6821.
(21) Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. Engineered Biosynthesis of Novel Polyketides: Properties of the whiE Aromatase/Cyclase. *Nat Biotechnol* **1996**, *14*, 335–338.
(22) Kramer, P. J. Rational Design and Engineered Biosynthesis of a Novel 18-Carbon Aromatic Polyketide. *J Am Chem Soc* **1997**, *119*, 635–639.
(23) Meurer, G. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4*, 433–443.
(24) McDaniel, R. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel 'Unnatural' Natural Products. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 1846–1851.
(25) Harvey, A. L. Natural Products in Drug Discovery. *Drug Discov. Today* **2008**, *13*, 894–901.
(26) Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A. Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale for Catalytic Reprogramming. *J Am Chem Soc* **2014**, *136*, 7348–7362.
(27) Connor, C. J. O.; Beckmann, H. S. G.; Spring, D. R. Diversity-Oriented Synthesis: Producing Chemical Tools for Dissecting Biology. *Chem Soc Rev* **2012**, *41*, 4444–4456.
(28) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for Bad Bugs: Confronting the Challenges of Antibacterial Discovery. *Nat Rev Drug Discov* **2007**, *6*, 29–40.
(29) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat Rev Drug Discov* **2015**, *14*, 111–129.
(30) Schloss, P. D.; Handelsman, J. Metagenomics for Studying Unculturable Microorganisms: Cutting the Gordian Knot. *Genome Biol.* **2005**, *6*, 229.
(31) Medema, H.; M. The Year 2020 in Natural Product Bioinformatics: An Overview of the Latest Tools and Databases. *Nat. Prod. Rep.* **2021**, *38*, 301–306.
(32) Blin, K. antiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation. *Nucleic Acids Res.* **2023**, *51*, 46–50.
(33) Blin, K.; Shaw, S.; Kautsar, S. A.; Medema, M. H.; Weber, T. The antiSMASH Database Version 3: Increased Taxonomic Coverage and New Query Features for Modular Enzymes. *Nucleic Acids Res.* **2021**, *49*, 639–643.
(34) Hannigan, G. D. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res* **2019**, *47*, 110–110.

(35) Ichikawa, N. DoBISCUIT: A Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2013**, *41*, 408–414.
(36) Kautsar, S. A. MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* **2020**, *48*, 454–458.
(37) Palaniappan, K. IMG-ABC v.5.0: An Update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res.* **2020**, *48*, 422–430.
(38) Bentley, S. D. Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2. *Nature* **2002**, *417*, 141–147.
(39) Nett, M.; Ikeda, H.; Moore, B. S. Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. *Nat Prod Rep* **2009**, *26*, 1362–1384.
(40) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating Natural Product Combinatorial Biosynthesis with Synthetic Biology. *Nat Chem Biol* **2015**, *11*, 649–659.
(41) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8), 509–523. https://doi.org/10.1038/nrmicro3496.
(42) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat Rev Drug Discov* **2021**, *20*, 200–216.
(43) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem* **2021**, *90*, 763–788.
(44) Xu, F. A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat Chem Biol* **2019**, *15*, 161–168.
(45) Herbst, D. A.; Townsend, C. A.; Maier, T. The Architectures of Iterative Type i PKS and FAS. *Nat Prod Rep* **2018**, *35*, 1046–1069.
(46) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
(47) Dutta, S. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510*, 512–517.
(48) Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. *Science* **2007**, *316*, 288–290.
(49) Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315–1322.
(50) Staunton, J.; J. Weissman, K. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416. https://doi.org/10.1039/A909079G.
(51) Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Nat Prod Rep* **2003**, *20*, 79–110.
(52) Ferrer, J.-L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Structure of Chalcone Synthase and the Molecular Basis of Plant Polyketide Biosynthesis. *Nat Struct Mol Biol* **1999**, *6*, 775–784.
(53) Bisang, C. A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* **1999**, *401*, 502–505.
(54) Jackowski, S.; Rock, C. O. Acetoacetyl-Acyl Carrier Protein Synthase, a Potential Regulator of Fatty Acid Biosynthesis in Bacteria. *J. Biol. Chem.* **1987**, *262*, 7927–7931.
(55) Moore, B. S.; Hertweck, C. Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units. *Nat Prod Rep* **2002**, *19*, 70–99.
(56) Crawford, J. M.; Dancy, B. C. R.; Hill, E. A.; Udwary, D. W.; Townsend, C. A. Identification of a Starter Unit Acyl-Carrier Protein Transacylase Domain in an Iterative Type I Polyketide Synthase. *Proc. Natl. Acad. Sci.* **2006**, *103* (45), 16728–16733. https://doi.org/10.1073/pnas.0604112103.

(57) Tsai, S.-C. (Sheryl). The Structural Enzymology of Iterative Aromatic Polyketide Synthases: A Critical Comparison with Fatty Acid Synthases. *Annu. Rev. Biochem.* **2018**, *87* (1), 503–531. https://doi.org/10.1146/annurev-biochem-063011-164509.
(58) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat Prod Rep* **2016**, *33*, 150–161.
(59) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **2009**, *26*, 90–114.
(60) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* **1998**, *37* (22), 8132–8138. https://doi.org/10.1021/bi980466i.
(61) Meadows, E. S.; Khosla, C. In Vitro Reconstitution and Analysis of the Chain Initiating Enzymes of the R1128 Polyketide Synthase. *Biochemistry* **2001**, *40*, 14855–14861.
(62) Winter, J. M. Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, *137*, 9885–9893.
(63) Cacho, R. A. Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J Am Chem Soc* **2015**, *137*, 15688–15691.
(64) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide b Synthase. *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
(65) Crawford, J. M. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320*, 243–246.
(66) Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709. https://doi.org/10.1021/ja0378759.
(67) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Dissection of the Catalytic Specificity of the Act Ketoreductase. *J. Am. Chem. Soc.* **1994**, *116*, 4166–4170.
(68) Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Evidence for Temporal, but Not Regiospecific, Control of Cyclization of an Aromatic Polyketide Precursor. *Chem. Biol.* **1994**, *1*, 205–210.
(69) Shen, Y. Ectopic Expression of the Minimal whiE Polyketide Synthase Generates a Library of Aromatic Polyketides of Diverse Sizes and Shapes. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 3622–3627.
(70) Auclair, K. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels–alder Reaction of a Substrate Analogue. *J Am Chem Soc* **2000**, *122*, 11519–11520.
(71) Kim, H. J.; Ruszczycky, M. W.; Choi, S.; Liu, Y.; Liu, H. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn a. *Nature* **2011**, *473*, 109–112.
(72) Kato, N. A New Enzyme Involved in the Control of the Stereochemistry in the Decalin Formation during Equisetin Biosynthesis. *Biochem. Biophys. Res. Commun.* **2015**, *460*, 210–215.
(73) Kasahara, K. Solanapyrone Synthase, a Possible Diels–Alderase and Iterative Type i Polyketide Synthase Encoded in a Biosynthetic Gene Cluster from Alternaria Solani. *ChemBioChem* **2010**, *11*, 1245–1252.
(74) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of Claisen Cyclase Domain in Fungal Polyketide Synthase WA, a Naphthopyrone Synthase of Aspergillus Nidulans. *Chem. Biol.* **2001**, *8*, 189–197.
(75) Crawford, J. M. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461*, 1139–1143.

(76) Ames, B. D. Crystal Structure and Functional Analysis of Tetracenomycin ARO/CYC: Implications for Cyclization Specificity of Aromatic Polyketides. *PNAS* **2008**, *105*, 5349–5354.
(77) Ames, B. D. Structural and Biochemical Characterization of ZhuI Aromatase/Cyclase from the R1128 Polyketide Pathway. *Biochemistry* **2011**, *50*, 8392–8406.
(78) Yu, T.-W. Engineered biosynthesis of novel polyketides from streptomyces spore pigment polyketide synthases. *J Am Chem Soc* **1998**, *120*, 7749–7759.
(79) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively. *J. Am. Chem. Soc.* **1994**, *116*, 10855–10859.
(80) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of Post-PKS Tailoring Steps through Combinatorial Biosynthesis. *Nat Prod Rep* **2002**, *19*, 542–580.
(81) Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. Cytochromes P450 for Natural Product Biosynthesis in Streptomyces: Sequence, Structure, and Function. *Nat Prod Rep* **2017**, *34*, 1141–1172.
(82) Chung, W.; Vanderwal, C. D. Stereoselective Halogenation in Natural Product Synthesis. *Angew. Chem. Int. Ed.* **2016**, *55*, 4396–4434.
(83) Singh, S., Jr; P., G. N.; Thorson, J. S. The Structural Biology of Enzymes Involved in Natural Product Glycosylation. *Nat Prod Rep* **2012**, *29*, 1201–1237.
(84) Podust, L. M.; Sherman, D. H. Diversity of P450 Enzymes in the Biosynthesis of Natural Products. *Nat Prod Rep* **2012**, *29*, 1251–1266.
(85) Carreras, C. W.; Pieper, R.; Khosla, C. Efficient Synthesis of Aromatic Polyketides in Vitro by the Actinorhodin Polyketide Synthase. *J Am Chem Soc* **1996**, *118*, 5158–5159.
(86) Shen, B.; Hutchinson, C. R. Deciphering the Mechanism for the Assembly of Aromatic Polyketides by a Bacterial Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1996**, *93*, 6600–6604.
(87) Harris, T. M.; Harris, C. M. Biomimetic Syntheses of Aromatic Polyketide Metabolites. *Pure Appl Chem* **1986**, *58*, 283–294.
(88) Sherman, D. h. Structure and Deduced Function of the Granaticin-Producing Polyketide Synthase Gene Cluster of Streptomyces Violaceoruber Tü22. *EMBO J.* **1989**, *8*, 2717–2725.
(89) Kim, E. S.; Cramer, K. D.; Shreve, A. L.; Sherman, D. H. Heterologous Expression of an Engineered Biosynthetic Pathway: Functional Dissection of Type II Polyketide Synthase Components in Streptomyces Species. *J Bacteriol* **1995**, *177*, 1202–1207.
(90) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11* (9), 888–893. https://doi.org/10.1038/nsmb808.
(91) Sherman, D. H.; Kim, E. S.; Bibb, M. J.; Hopwood, D. A. Functional Replacement of Genes for Individual Polyketide Synthase Components in Streptomyces Coelicolor A3(2) by Heterologous Genes from a Different Polyketide Pathway. *J. Bacteriol.* **1992**, *174*, 6184–6190.
(92) Kim, E. S.; Hopwood, D. A.; Sherman, D. H. Analysis of Type II Polyketide Beta-Ketoacyl Synthase Specificity in Streptomyces Coelicolor A3(2) by Trans Complementation of Actinorhodin Synthase Mutants. *J Bacteriol* **1994**, *176*, 1801–1804.
(93) Dreier, J.; Khosla, C. Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the β-Ketoacyl Synthase–chain Length Factor Heterodimer. *Biochemistry* **2000**, *39*, 2088–2095.

(94) Grammbitter, G. L. C. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. *J Am Chem Soc* **2019**, *141*, 16615–16623.
(95) Bräuer, A. Structural Snapshots of the Minimal PKS System Responsible for Octaketide Biosynthesis. *Nat Chem* **2020**, *12*, 755–763.
(96) Nicholson, T. P. First in Vitro Directed Biosynthesis of New Compounds by a Minimal Type II Polyketide Synthase: Evidence for the Mechanism of Chain Length Determination. *Chem Commun* **2003**, *686–687*. https://doi.org/10.1039/B300847A.
(97) Kim, Y.; Prestegard, J. H. Refinement of the NMR Structures for Acyl Carrier Protein with Scalar Coupling Data. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 377–385.
(98) Evans, S. E. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* **2008**, *9*, 2424–2432.
(99) Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein. *Biochemistry* **2003**, *42*, 4648–4657.
(100) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. Solution Structure of the Actinorhodin Polyketide Synthase Acyl Carrier Protein from Streptomyces Coelicolor A3(2),. *Biochemistry* **1997**, *36* (20), 6000–6008. https://doi.org/10.1021/bi970006+.
(101) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. Targeted Gene Replacements in a Streptomyces Polyketide Synthase Gene Cluster: Role for the Acyl Carrier Protein. *Mol. Microbiol.* **1992**, *6*, 3237–3249.
(102) Zhang, Y.-M.; Wu, B.; Zheng, J.; Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction*. *J. Biol. Chem.* **2003**, *278*, 52935–52943.
(103) Arthur, C. J. Structure and Malonyl CoA-ACP Transacylase Binding of Streptomyces Coelicolor Fatty Acid Synthase Acyl Carrier Protein. *ACS Chem Biol* **2009**, *4*, 625–636.
(104) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-Based Analysis of the Molecular Interactions between Acyltransferase and Acyl Carrier Protein in Vicenistatin Biosynthesis. *PNAS* **2016**, *113*, 1802–1807.
(105) Bartholow, T. G. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4*, 1–10.
(106) Cho, Y. I. Engineered Chimeras Unveil Swappable Modular Features of Fatty Acid and Polyketide Synthase Acyl Carrier Proteins. *Biochemistry* **2022**, *61*, 217–227.
(107) Khosla, C. Genetic Construction and Functional Analysis of Hybrid Polyketide Synthases Containing Heterologous Acyl Carrier Proteins. *J. Bacteriol.* **1993**, *175*, 2197–2204.
(108) Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C. Ketosynthases in the Initiation and Elongation Modules of Aromatic Polyketide Synthases Have Orthogonal Acyl Carrier Protein Specificity†, 2003. https://doi.org/10.1021/BI0341962.
(109) Das, A.; Khosla, C. In Vivo and in Vitro Analysis of the Hedamycin Polyketide Synthase. *Chem. Biol.* **2009**, *16*, 1197–1207.
(110) Koppisch, A. T.; Khosla, C. Structure-Based Mutagenesis of the Malonyl-CoA:Acyl Carrier Protein Transacylase from Streptomyces Coelicolor. *Biochemistry* **2003**, *42*, 11057–11064.
(111) Beltran-Alvarez, P.; Cox, R. J.; Crosby, J.; Simpson, T. J. Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* **2007**, *46*, 14672–14681.

(112) Arthur, C. J. Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* **2005**, *44*, 15414–15421.
(113) Kalaitzis, J. A. Policing Starter Unit Selection of the Enterocin Type II Polyketide Synthase by the Type II Thioesterase EncL. *Bioorg. Med. Chem.* **2011**, *19*, 6633–6638.
(114) Shepherd, M. D.; Kharel, M. K.; Zhu, L. L.; Lanen, S. G. van; Rohr, J. Delineating the Earliest Steps of Gilvocarcin Biosynthesis: Role of GilP and GilQ in Starter Unit Specificity. *Org Biomol Chem* **2010**, *8*, 3851–3856.
(115) Tang, Y.; Koppisch, A. T.; Khosla, C. The Acyltransferase Homologue from the Initiation Module of the R1128 Polyketide Synthase Is an Acyl-ACP Thioesterase That Edits Acetyl Primer Units. *Biochemistry* **2004**, *43* (29), 9546–9555. https://doi.org/10.1021/bi049157k.
(116) Keatinge-Clay, A. T. Catalysis, Specificity, and ACP Docking Site of Streptomyces Coelicolor Malonyl-CoA:ACP Transacylase. *Structure* **2003**, *11*, 147–154.
(117) Szu, P.-H. Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **2011**, *18*, 1021–1031.
(118) Bartel, P. L. Biosynthesis of Anthraquinones by Interspecies Cloning of Actinorhodin Biosynthesis Genes in Streptomycetes: Clarification of Actinorhodin Gene Functions. *J. Bacteriol.* **1990**, *172*, 4816–4826.
(119) Rajgarhia, V. B.; Strohl, W. R. Minimal streptomyces sp. Strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis. *J Bacteriol* **1997**, *179*, 2690–2696.
(120) Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C. Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKR): Molecular Basis of Stereo- and Regiospecificities. *Biochemistry* **2011**, *50*, 7426–7439.
(121) Piel, J. Cloning, Sequencing and Analysis of the Enterocin Biosynthesis Gene Cluster from the Marine Isolate 'Streptomyces Maritimus': Evidence for the Derailment of an Aromatic Polyketide Synthase. *Chem. Biol.* **2000**, *7*, 943–955.
(122) Price, A. C.; Zhang, Y. M.; Rock, C. O.; White, S. W. Structure of β-Ketoacyl-[Acyl Carrier Protein] Reductase from Escherichia Coli: Negative Cooperativity and Its Structural Basis. *Biochemistry* **2001**, *40*, 12772–12781.
(123) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43* (46), 14529–14538. https://doi.org/10.1021/bi048133a.
(124) Oppermann, U. Short-Chain Dehydrogenases/Reductases (SDR): The 2002 Update. *Chem. Biol. Interact.* **2003**, *143–144*, 247–253.
(125) Filling, C. Critical Residues for Structure and Catalysis in Short-Chain Dehydrogenases/Reductases*. *J. Biol. Chem.* **2002**, *277*, 25677–25684.
(126) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **1993**, *262* (5139), 1546–1550.
(127) Kalaitzis, J. A.; Moore, B. S. Heterologous Biosynthesis of Truncated Hexaketides Derived from the Actinorhodin Polyketide Synthase. *J. Nat. Prod.* **2004**, *67*, 1419–1422.
(128) Zawada, R. J.; Khosla, C. Heterologous Expression, Purification, Reconstitution and Kinetic Analysis of an Extended Type II Polyketide Synthase. *Chem. Biol.* **1999**, *6* (9), 607–615. https://doi.org/10.1016/s1074-5521(99)80112-4.
(129) Hadfield, A. T. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12*, 1865–1875.

(130) Kantola, J. Folding of the Polyketide Chain Is Not Dictated by Minimal Polyketide Synthase in the Biosynthesis of Mithramycin and Anthracycline. *Chem. Biol.* **1997**, *4*, 751–755.
(131) Kunnari, T. Incorrectly Folded Aromatic Polyketides from Polyketide Reductase Deficient Mutants. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2639–2642.
(132) Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C. Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase,. *Biochemistry* **2008**, *47* (7), 1837–1847. https://doi.org/10.1021/bi7016427.
(133) Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C. C. Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochemistry* **2011**, *50*, 4638–4649.
(134) Javidpour, P. The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chem. Biol.* **2013**, *20*, 1225–1234.
(135) Zhang, H. L. Mutactin, a Novel Polyketide from Streptomyces Coelicolor. Structure and Biosynthetic Relationship to Actinorhodin. *J Org Chem* **1990**, *55*, 1682–1684.
(136) Zhao, S. Molecular Basis for Polyketide Ketoreductase–Substrate Interactions. *Int. J. Mol. Sci.* **2020**, *21*, 7562.
(137) Kharel, M. K. Elucidation of post-PKS tailoring steps involved in landomycin biosynthesis. *Org Biomol Chem* **2012**, *10*, 4256–4265.
(138) Mayer, A. LanV, a Bifunctional Enzyme: Aromatase and Ketoreductase during Landomycin a Biosynthesis. *ChemBioChem* **2005**, *6*, 2312–2315.
(139) Dickens, M. L.; Ye, J.; Strohl, W. R. Cloning, Sequencing, and Analysis of Aklaviketone Reductase from Streptomyces Sp. Strain C5. *J. Bacteriol.* **1996**, *178*, 3384–3388.
(140) Lackner, G. Biosynthesis of Pentangular Polyphenols: Deductions from the Benastatin and Griseorhodin Pathways. *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
(141) Torkkell, S. The Entire Nogalamycin Biosynthetic Gene Cluster of Streptomyces Nogalater: Characterization of a 20-Kb DNA Region and Generation of Hybrid Structures. *Mol Gen Genomics* **2001**, *266*, 276–288.
(142) Paananen, P. Structural and Functional Analysis of Angucycline C-6 Ketoreductase LanV Involved in Landomycin Biosynthesis. *Biochemistry* **2013**, *52*, 5304–5314.
(143) Patrikainen, P. Structure-Based Engineering of Angucyclinone 6-Ketoreductases. *Chem. Biol.* **2014**, *21*, 1381–1391.
(144) Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C. Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (12), 3421–3430. https://doi.org/10.1021/acschembio.6b00658.
(145) Xiao, X. Functional and Structural Insights into a Novel Promiscuous Ketoreductase of the Lugdunomycin Biosynthetic Pathway. *ACS Chem Biol* **2020**, *15*, 2529–2538.
(146) Kallio, P. Flavoprotein Hydroxylase PgaE Catalyzes Two Consecutive Oxygen-Dependent Tailoring Reactions in Angucycline Biosynthesis. *Biochemistry* **2011**, *50*, 5535–5543.
(147) Patrikainen, P. Tailoring Enzymes Involved in the Biosynthesis of Angucyclines Contain Latent Context-Dependent Catalytic Activities. *Chem. Biol.* **2012**, *19*, 647–655.
(148) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **2003**, *4*, 654–657.
(149) Moreno, G. O. *Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type i and Type II Polyketide Synthases Using Atom-Replaced Probes*; UC Irvine, 2018.

(150) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(151) Li 然 李嫣, Y.; Xu 伟徐, W.; Tang 奕唐, Y. Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains*. *J. Biol. Chem.* **2010**, *285* (30), 22764–22773. https://doi.org/10.1074/jbc.M110.128504.
(152) Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C. A. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461* (7267), 1139–1143. https://doi.org/10.1038/nature08475.
(153) Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. *Proc. Natl. Acad. Sci.* **2017**, *114* (21), E4142–E4148. https://doi.org/10.1073/pnas.1609001114.
(154) Liu, L.; Zhang, Z.; Shao, C.-L.; Wang, J.-L.; Bai, H.; Wang, C.-Y. Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Sci. Rep.* **2015**, *5* (1), 10463. https://doi.org/10.1038/srep10463.
(155) Feng, Y.; Yang, X.; Ji, H.; Deng, Z.; Lin, S.; Zheng, J. The Streptomyces Viridochromogenes Product Template Domain Represents an Evolutionary Intermediate between Dehydratase and Aldol Cyclase of Type I Polyketide Synthases. *Commun. Biol.* **2022**, *5* (1), 508. https://doi.org/10.1038/s42003-022-03477-8.
(156) Pan, H. Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. *Structure* **2002**, *10*, 1559–1568.
(157) Thompson, T. B.; Katayama, K.; Watanabe, K.; Hutchinson, C. R.; Rayment, I. Structural and Functional Analysis of Tetracenomycin F2 Cyclase from Streptomyces Glaucescens: A TYPE II POLYKETIDE CYCLASE*. *J. Biol. Chem.* **2004**, *279*, 37956–37963.
(158) Sultana, A. Structure of the Polyketide Cyclase SnoaL Reveals a Novel Mechanism for Enzymatic Aldol Condensation. *EMBO J.* **2004**, *23*, 1911–1921.
(159) Kallio, P.; Sultana, A.; Niemi, J.; Mäntsälä, P.; Schneider, G. Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity. *J. Mol. Biol.* **2006**, *357*, 210–220.
(160) Hutchinson, C. R. Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem. Rev.* **1997**, *97* (7), 2525–2536. https://doi.org/10.1021/cr960022x.
(161) Ahuja, M.; Chiang, Y.-M.; Chang, S.-L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H.-C.; Yeh, H.-H.; Oakley, B. R.; Wang, C. C. C. Illuminating the Diversity of Aromatic Polyketide Synthases in Aspergillus Nidulans. *J. Am. Chem. Soc.* **2012**, *134* (19), 8212–8221. https://doi.org/10.1021/ja3016395.
(162) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97* (7), 2465–2498. https://doi.org/10.1021/cr960034i.
(163) Shen, B. Polyketide Biosynthesis beyond the Type I, II and III Polyketide Synthase Paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6.
(164) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639. https://doi.org/10.1021/ar8002249.
(165) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A.

Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130. https://doi.org/10.1126/science.ade2574.

(166) Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. Structural Basis for Selectivity in a Highly Reducing Type II Polyketide Synthase. *Nat. Chem. Biol.* **2020**, *16* (7), 776–782. https://doi.org/10.1038/s41589-020-0530-0.

(167) Yadav, G.; Gokhale, R. S.; Mohanty, D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J. Mol. Biol.* **2003**, *328* (2), 335–363. https://doi.org/10.1016/S0022-2836(03)00232-8.

(168) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci.* **2021**, *118* (15), e2016239118. https://doi.org/10.1073/pnas.2016239118.

(169) Capela, J.; Zimmermann-Kogadeeva, M.; Dijk, A. D. J. van; de Ridder, D.; Dias, O.; Rocha, M. Comparative Assessment of Protein Large Language Models for Enzyme Commission Number Prediction. *BMC Bioinformatics* **2025**, *26* (1), 68. https://doi.org/10.1186/s12859-025-06081-9.

(170) Jacques, M.-A.; Durand, K.; Orgeur, G.; Balidas, S.; Fricot, C.; Bonneau, S.; Quillévéré, A.; Audusseau, C.; Olivier, V.; Grimault, V.; Mathis, R. Phylogenetic Analysis and Polyphasic Characterization of Clavibacter Michiganensis Strains Isolated from Tomato Seeds Reveal That Nonpathogenic Strains Are Distinct from C. Michiganensis Subsp. Michiganensis. *Appl. Environ. Microbiol.* **2012**, *78* (23), 8388–8402. https://doi.org/10.1128/AEM.02158-12.

(171) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat. Chem. Biol.* **2019**, *15* (7), 669–671. https://doi.org/10.1038/s41589-019-0301-y.

(172) Ellis, B. D.; Milligan, J. C.; White, A. R.; Duong, V.; Altman, P. X.; Mohammed, L. Y.; Crump, M. P.; Crosby, J.; Luo, R.; Vanderwal, C. D.; Tsai, S.-C. An Oxetane-Based Polyketide Surrogate To Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140* (15), 4961–4964. https://doi.org/10.1021/jacs.7b11793.

(173) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(174) Zdouc, M. M.; Blin, K.; Louwen, N. L. L.; Navarro, J.; Loureiro, C.; Bader, C. D.; Bailey, C. B.; Barra, L.; Booth, T. J.; Bozhüyük, K. A. J.; Cediel-Becerra, J. D. D.; Charlop-Powers, Z.; Chevrette, M. G.; Chooi, Y. H.; D'Agostino, P. M.; de Rond, T.; Del Pup, E.; Duncan, K. R.; Gu, W.; Hanif, N.; Helfrich, E. J. N.; Jenner, M.; Katsuyama, Y.; Korenskaia, A.; Krug, D.; Libis, V.; Lund, G. A.; Mantri, S.; Morgan, K. D.; Owen, C.; Phan, C.-S.; Philmus, B.; Reitz, Z. L.; Robinson, S. L.; Singh, K. S.; Teufel, R.; Tong, Y.; Tugizimana, F.; Ulanova, D.;

Winter, J. M.; Aguilar, C.; Akiyama, D. Y.; Al-Salihi, S. A. A.; Alanjary, M.; Alberti, F.; Aleti, G.; Alharthi, S. A.; Rojo, M. Y. A.; Arishi, A. A.; Augustijn, H. E.; Avalon, N. E.; Avelar-Rivas, J. A.; Axt, K. K.; Barbieri, H. B.; Barbosa, J. C. J.; Barboza Segato, L. G.; Barrett, S. E.; Baunach, M.; Beemelmanns, C.; Beqaj, D.; Berger, T.; Bernaldo-Agüero, J.; Bettenbühl, S. M.; Bielinski, V. A.; Biermann, F.; Borges, R. M.; Borriss, R.; Breitenbach, M.; Bretscher, K. M.; Brigham, M. W.; Buedenbender, L.; Bulcock, B. W.; Cano-Prieto, C.; Capela, J.; Carrion, V. J.; Carter, R. S.; Castelo-Branco, R.; Castro-Falcón, G.; Chagas, F. O.; Charria-Girón, E.; Chaudhri, A. A.; Chaudhry, V.; Choi, H.; Choi, Y.; Choupannejad, R.; Chromy, J.; Donahey, M. S. C.; Collemare, J.; Connolly, J. A.; Creamer, K. E.; Crüsemann, M.; Cruz, A. A.; Cumsille, A.; Dallery, J.-F.; Damas-Ramos, L. C.; Damiani, T.; de Kruijff, M.; Martín, B. D.; Sala, G. D.; Dillen, J.; Doering, D. T.; Dommaraju, S. R.; Durusu, S.; Egbert, S.; Ellerhorst, M.; Faussurier, B.; Fetter, A.; Feuermann, M.; Fewer, D. P.; Foldi, J.; Frediansyah, A.; Garza, E. A.; Gavriilidou, A.; Gentile, A.; Gerke, J.; Gerstmans, H.; Gomez-Escribano, J. P.; González-Salazar, L. A.; Grayson, N. E.; Greco, C.; Gomez, J. E. G.; Guerra, S.; Flores, S. G.; Gurevich, A.; Gutiérrez-García, K.; Hart, L.; Haslinger, K.; He, B.; Hebra, T.; Hemmann, J. L.; Hindra, H.; Höing, L.; Holland, D. C.; Holme, J. E.; Horch, T.; Hrab, P.; Hu, J.; Huynh, T.-H.; Hwang, J.-Y.; Iacovelli, R.; Iftime, D.; Iorio, M.; Jayachandran, S.; Jeong, E.; Jing, J.; Jung, J. J.; Kakumu, Y.; Kalkreuter, E.; Kang, K. B.; Kang, S.; Kim, W.; Kim, G. J.; Kim, H.; Kim, H. U.; Klapper, M.; Koetsier, R. A.; Kollten, C.; Kovács, Á. T.; Kriukova, Y.; Kubach, N.; Kunjapur, A. M.; Kushnareva, A. K.; Kust, A.; Lamber, J.; Larralde, M.; Larsen, N. J.; Launay, A. P.; Le, N.-T.-H.; Lebeer, S.; Lee, B. T.; Lee, K.; Lev, K. L.; Li, S.-M.; Li, Y.-X.; Licona-Cassani, C.; Lien, A.; Liu, J.; Lopez, J. A. V.; Machushynets, N. V.; Macias, M. I.; Mahmud, T.; Maleckis, M.; Martinez-Martinez, A. M.; Mast, Y.; Maximo, M. F.; McBride, C. M.; McLellan, R. M.; Bhatt, K. M.; Melkonian, C.; Merrild, A.; Metsä-Ketelä, M.; Mitchell, D. A.; Müller, A. V.; Nguyen, G.-S.; Nguyen, H. T.; Niedermeyer, T. H. J.; O'Hare, J. H.; Ossowicki, A.; Ostash, B. O.; Otani, H.; Padva, L.; Paliyal, S.; Pan, X.; Panghal, M.; Parade, D. S.; Park, J.; Parra, J.; Rubio, M. P.; Pham, H. T.; Pidot, S. J.; Piel, J.; Pourmohsenin, B.; Rakhmanov, M.; Ramesh, S.; Rasmussen, M. H.; Rego, A.; Reher, R.; Rice, A. J.; Rigolet, A.; Romero-Otero, A.; Rosas-Becerra, L. R.; Rosiles, P. Y.; Rutz, A.; Ryu, B.; Sahadeo, L.-A.; Saldanha, M.; Salvi, L.; Sánchez-Carvajal, E.; Santos-Medellin, C.; Sbaraini, N.; Schoellhorn, S. M.; Schumm, C.; Sehnal, L.; Selem, N.; Shah, A. D.; Shishido, T. K.; Sieber, S.; Silviani, V.; Singh, G.; Singh, H.; Sokolova, N.; Sonnenschein, E. C.; Sosio, M.; Sowa, S. T.; Steffen, K.; Stegmann, E.; Streiff, A. B.; Strüder, A.; Surup, F.; Svenningsen, T.; Sweeney, D.; Szenei, J.; Tagirdzhanov, A.; Tan, B.; Tarnowski, M. J.; Terlouw, B. R.; Rey, T.; Thome, N. U.; Torres Ortega, L. R.; Tørring, T.; Trindade, M.; Truman, A. W.; Tvilum, M.; Udwary, D. W.; Ulbricht, C.; Vader, L.; van Wezel, G. P.; Walmsley, M.; Warnasinghe, R.; Weddeling, H. G.; Weir, A. N. M.; Williams, K.; Williams, S. E.; Witte, T. E.; Rocca, S. M. W.; Yamada, K.; Yang, D.; Yang, D.; Yu, J.; Zhou, Z.; Ziemert, N.; Zimmer, L.; Zimmermann, A.; Zimmermann, C.; van der Hooft, J. J. J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 4.0: Advancing Biosynthetic Gene Cluster Curation through Global Collaboration. *Nucleic Acids Res.* **2025**, *53* (D1), D678–D690. https://doi.org/10.1093/nar/gkae1115.

(175) Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S. J.; Webb, C.; Wu, D.; Varghese, N. J.; Reddy, T. B. K.; Mukherjee, S.; Ovchinnikova, G.; Nolan, M.; Seshadri, R.; Roux, S.; Visel, A.; Woyke, T.; Eloe-Fadrosh, E. A.; Kyrpides, N. C.; Ivanova, N. N. The IMG/M Data Management and Analysis System

v.7: Content Updates and New Features. *Nucleic Acids Res.* **2023**, *51* (D1), D723–D732. https://doi.org/10.1093/nar/gkac976.
(176) Fernández-Moreno, M. A.; Martínez, E.; Caballero, J. L.; Ichinose, K.; Hopwood, D. A.; Malpartida, F. DNA Sequence and Functions of the actVI Region of the Actinorhodin Biosynthetic Gene Cluster of Streptomyces Coelicolor A3(2). *J. Biol. Chem.* **1994**, *269* (40), 24854–24863. https://doi.org/10.1016/S0021-9258(17)31469-2.
(177) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Richard Hutchinson, C. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4* (6), 433–443. https://doi.org/10.1016/S1074-5521(97)90195-2.
(178) Moreno, G. O. Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type I and Type II Polyketide Synthases Using Atom-Replaced Probes, UC Irvine, 2018. https://escholarship.org/uc/item/1tx0518k.
(179) Serapian, S. A.; Crosby, J.; Crump, M. P.; van der Kamp, M. W. Path to Actinorhodin: Regio- and Stereoselective Ketone Reduction by a Type II Polyketide Ketoreductase Revealed in Atomistic Detail. *JACS Au* **2022**, *2* (4), 972–984. https://doi.org/10.1021/jacsau.2c00086.
(180) Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941–953. https://doi.org/10.1016/j.jmb.2008.09.084.
(181) Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12* (10), 1865–1875. https://doi.org/10.1016/j.str.2004.08.002.
(182) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat. Prod. Rep.* **2016**, *33* (2), 203–230. https://doi.org/10.1039/C5NP00109A.
(183) Chooi, Y.-H.; Tang, Y. Navigating the Fungal Polyketide Chemical Space: From Genes to Molecules. *J. Org. Chem.* **2012**, *77* (22), 9933–9953. https://doi.org/10.1021/jo301592k.
(184) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(185) Barajas, J. F.; Finzel, K.; Valentic, T. R.; Shakya, G.; Gamarra, N.; Martinez, D.; Meier, J. L.; Vagstad, A. L.; Newman, A. G.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Structural and Biochemical Analysis of Protein–Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew. Chem. Int. Ed.* **2016**, *55* (42), 13005–13009. https://doi.org/10.1002/anie.201605401.
(186) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(187) Barajas, J. F. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. In *Proceedings of the National Academy of Sciences*; 2017; Vol. 114, pp 4142–4148.
(188) Milligan, J. C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat Chem Biol* **2019**, *15*, 669–671.

(189) Ellis, B. D. An Oxetane-Based Polyketide Surrogate to Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140*, 4961–4964.
(190) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43*, 14529–14538.

Chapter 2

Machine Learning-Driven Exploration of KS/CLF Chain-Length Specificity: Insights and Limitations from Protein Language Models

2.1 Summary

Aromatic polyketides derive their structural diversity and biological activity from precise control over carbon chain length during biosynthesis by type II polyketide synthases (PKSs). Although the ketosynthase/chain-length factor (KS/CLF) heterodimer is known to dictate the length of the poly-β-ketone intermediate, the sequence determinants underlying this specificity remain poorly understood, limiting efforts to rationally engineer novel polyketide scaffolds. Here, we combined biochemical reconstitution with machine learning (ML) and large-scale bioinformatics to evaluate whether KS/CLF chain-length specificity can be predicted directly from primary sequence. Act and Dps KS/CLF complexes were successfully expressed, purified, and reconstituted in vitro; however, structural studies remained challenging due to poor crystallizability, motivating a computational approach. Initial classification models built using one-hot encodings and simple neural networks performed poorly, reflecting the inadequacy of classical representations for capturing subtle functional signals. Transformer architectures modestly improved accuracy, but the most significant gains were achieved using ESM-2, a protein language model trained on tens of millions of sequences. ESM-2 embeddings increased prediction accuracy to ~70% (training) and ~50% (validation), demonstrating that modern language models capture meaningful biochemical features relevant to chain-length control.

To assess whether functional groupings could be discovered without labels, we collected 2,315 additional KS/CLF sequences and applied unsupervised clustering using ESM-2 embeddings, with and without Partial Least Squares (PLS)–derived functional latent spaces. However, clustering consistently failed to separate sequences by chain-length class. Together, these outcomes reveal that KS/CLF sequences are exceptionally conserved and that chain-length control is governed by extremely subtle residue-level and structural variations not easily recoverable from global sequence patterns.

Overall, this work demonstrates both the promise and the limitations of sequence-based ML for predicting KS/CLF chain-length specificity. While protein language models offer substantial advantages over classical encodings, sequence information alone does not fully capture the fine structural determinants that regulate polyketide chain elongation. Future integration of structural modeling, coevolutionary analysis, and high-resolution biochemical data will be essential for building predictive frameworks capable of guiding rational engineering of aromatic polyketides with customized chain lengths and improved therapeutic potential.

2.2 Introduction

Polyketides constitute one of the most structurally diverse and therapeutically valuable classes of natural products, giving rise to numerous clinically important antibiotics, anticancer agents, antifungal agents, and antiviral compounds.[8,50] Their biosynthesis is carried out by polyketide synthases (PKSs), large multidomain enzyme systems that assemble complex molecules from simple acyl-CoA precursors. PKSs share a deep evolutionary relationship with fatty acid synthases (FASs) and utilize similar chemistries and domain architectures to elongate carbon chains and introduce tailored structural modifications.[162] Based on their organization and catalytic logic, PKSs are classified into three groups: type I, type II, and type III.[163] Among these, the iterative type II PKSs—found predominantly in bacteria—produce aromatic polyketides through repeated cycles of condensation, reduction, and cyclization reactions.[57]

A defining and mechanistically crucial step in type II PKS biosynthesis is the determination of carbon chain length. The minimal PKS, consisting of the ketosynthase (KS), chain length factor (CLF), and acyl carrier protein (ACP), initiates the construction of the poly-β-ketone intermediate by combining a starter unit with malonyl-ACP and elongating the chain through multiple rounds of decarboxylative condensation.[17,164] The KS/CLF heterodimer acts as the central catalytic hub in this process, and the CLF is known to be the primary determinant of the final chain length.[50,57] Variations in chain length lead to structurally distinct polyketides—for example, doxorubicin, steffimycin, and BE-7585A all originate from 20-carbon decaketides, whereas WhiE pigments derive from 24-carbon dodecaketides.[57] Despite the fundamental importance of chain length control, the molecular features within KS/CLF that dictate substrate accommodation and termination remain poorly understood. Sequence–

function relationships are particularly obscure, and rational engineering aimed at producing polyketides of custom lengths has been largely unsuccessful.

A major barrier to resolving chain length specificity is the scarcity of structural information and the lack of predictable sequence patterns that correlate KS/CLF variants to defined chain lengths. Traditional approaches—such as structure-guided mutagenesis and comparative sequence analysis—have offered valuable insights but have not yielded a generalizable model capable of predicting chain length across diverse PKS systems. This challenge is amplified by the vast number of KS and CLF sequences available in genomic databases versus the relatively small number of experimentally characterized chain lengths. The resulting imbalance makes chain-length prediction a problem uniquely suited for modern computational and machine learning approaches.

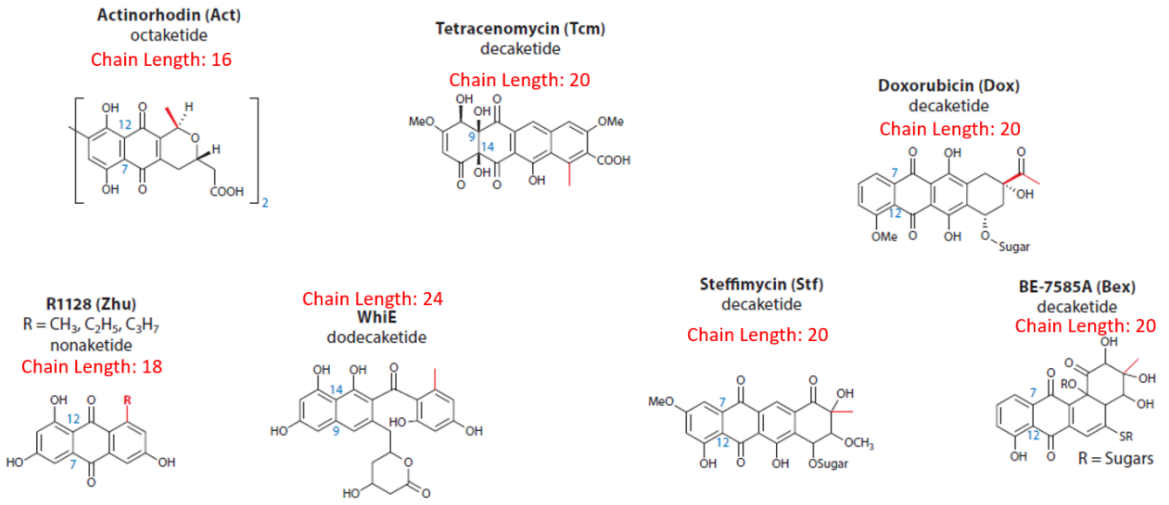


Figure 2.1. *Examples of polyketides of type II PKS with various chain length, showing the starter unit (red), first-ring cyclization (numbered), and its corresponding PKS (parentheses).*

Recent advances in protein language models, particularly transformer-based architectures such as ESM-2, provide a powerful opportunity to address this knowledge gap.[165] These models learn rich, evolutionarily informed representations of protein sequences from massive databases without requirement of structural or functional annotations. Such

embeddings capture subtle biochemical and structural properties that may not be apparent from sequence alignment alone. Leveraging these models, it is now possible to build predictive frameworks that map KS/CLF sequence information to polyketide chain length with unprecedented accuracy. This approach has the potential to uncover previously hidden determinants of substrate accommodation within the KS active site and the CLF chain-length tunnel, thereby enabling systematic engineering of chain-length control for the first time.

This chapter focuses on elucidating the molecular basis of carbon chain length control in KS/CLF complexes through an integrated machine learning and structural biology strategy. To achieve this goal, a transformer-based model leveraging ESM-2 embeddings will be trained on curated KS/CLF sequences paired with experimentally validated chain-length outputs, thus enabling a predictive link between sequence features and programmed elongation limits. This approach will facilitate the identification of key residues, co-evolving motifs, and structural determinants that govern chain-length specificity. Because experimentally annotated datasets remain limited, a complementary unsupervised workflow will be implemented in which unlabelled CLF sequences are embedded using the same ESM-2 representation and subsequently clustered via K-means. This analysis will allow classification of CLF homologs into chain-length groups and help infer their likely product profiles based on evolutionary and structural proximity. Together, these supervised and unsupervised efforts will establish the first data-driven and mechanistically informed framework for understanding, predicting, and ultimately engineering polyketide chain-length control—advancing the rational design of aromatic polyketides with novel or improved therapeutic properties.

2.3 Results

2.3.1 Expression, Purification, and Reconstitution of KS/CLF Complexes

Both actinorhodin(Act) and daunorubicin(Dps) KS/CLF complexes were successfully overexpressed and purified using Ni-NTA affinity chromatography followed by size-exclusion chromatography. SDS-PAGE analysis confirmed the expected band patterns corresponding to the KS and CLF subunits, demonstrating high purity of both complexes. In vitro reconstitution assays further verified the functional assembly of the heterodimers. Despite the successful purification, obtaining sufficient quantities required approximately one month per construct due to slow expression and the need for repeated optimization cycles. Crystallization trials of Dps KS/CLF were performed extensively under a variety of screening conditions; however, no diffracting crystals were obtained, and most trials yielded only amorphous precipitate or poorly formed microcrystals. These results support the need for alternative structural approaches, such as computation-guided analysis, to investigate chain-length determinants in KS/CLF systems.

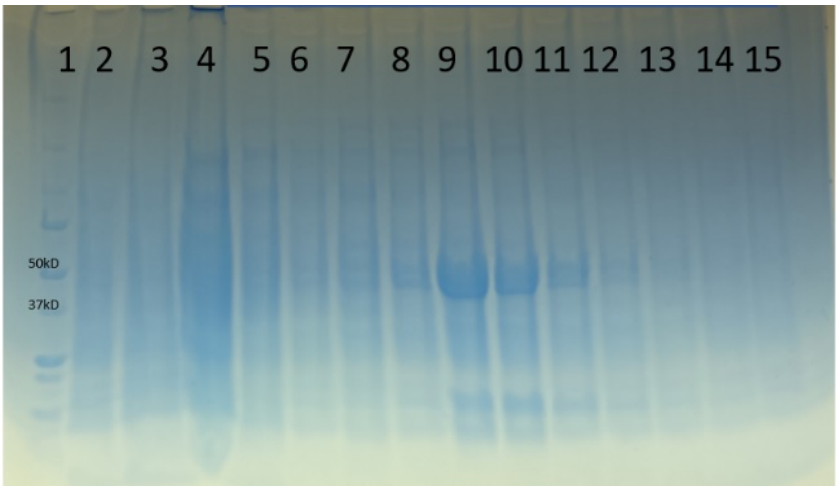


Figure 2.2. *Dps KS/CLF fractions after purification with nickel resin*

2.3.2 Initial Machine Learning Classification Using One-Hot Encoding

To establish a computational framework for predicting polyketide chain length, KS/CLF sequences with known product lengths were collected from the MIBiG database and encoded using one-hot representation. A simple fully connected neural network was trained on this dataset as an initial baseline model. This architecture performed poorly, achieving only ~50% accuracy on the training set and 0% accuracy on the validation set, with a high cross-entropy loss exceeding 5 after training. These results indicated substantial overfitting and suggested that one-hot encoding fails to capture the complex evolutionary and biochemical relationships embedded in KS/CLF sequences.

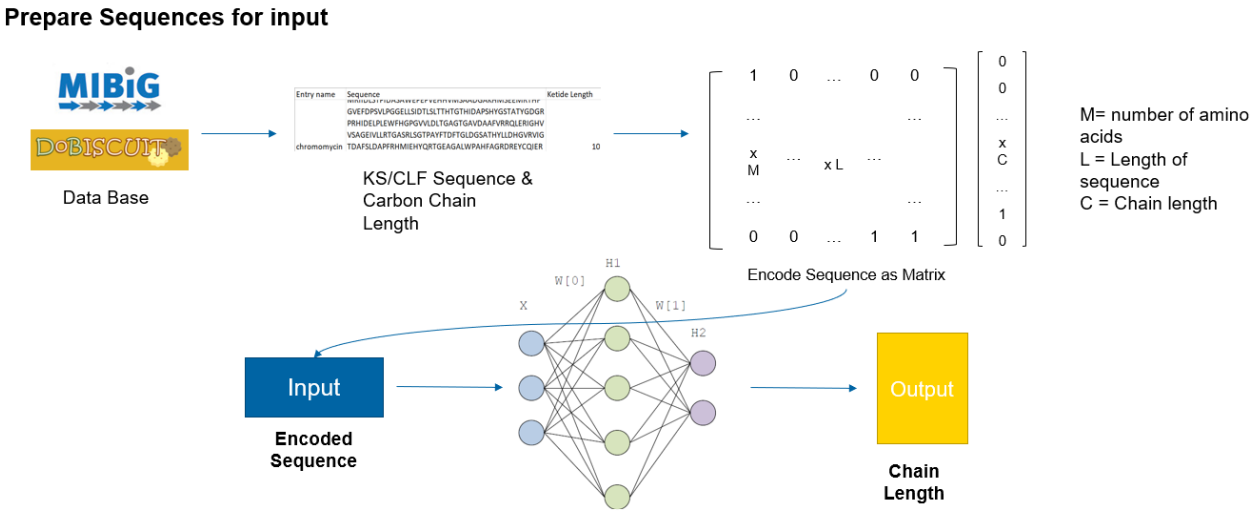


Figure 2.3. *Diagram of workflow to collect KS/CLF sequence data and encoding it to matrix. Sequence and chain length data were obtained from MIBiG and the sequences are encoded with one-hot encode.*

2.3.3 Improved Performance Using Transformer-Based Representations

To better capture long-range dependencies and subtle sequence features, a transformer-based sequence model was next applied. Using a transformer trained directly on the one-hot sequences, training accuracy modestly improved to ~50%, while validation accuracy increased to ~30%. However, the loss remained high (approximately 3), reflecting limited ability of the

model to generalize. These results suggested that while transformers provide a more expressive architecture than simple fully connected networks, meaningful improvements would require a more informative sequence embedding.

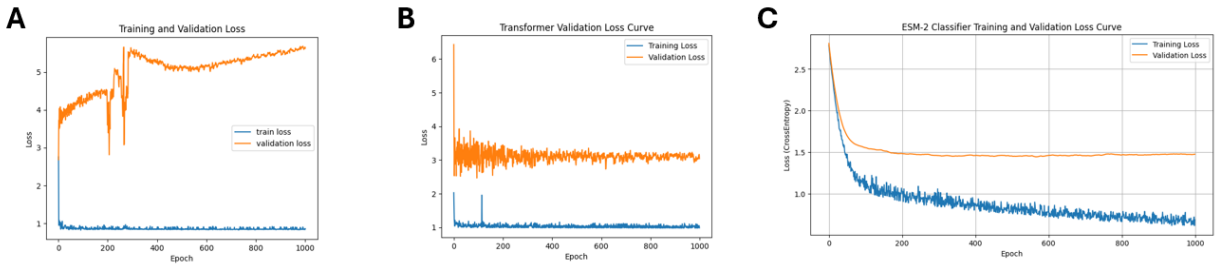


Figure 2.4. *Train and validation loss curve of A) fully connected model, B)one-hot transformer model and C) ESM-2 embedded transformer model. Each model was run with 33 training squences and 10 validation sequences. The epochs = 1000 and batch size =8.*

2.3.4 Application of ESM-2 Embeddings

Substantial performance improvement was achieved by using the ESM-2 model, a state-of-the-art protein language model developed by Meta AI. ESM-2 is trained on hundreds of millions of protein sequences in a masked language modeling framework, enabling it to learn biochemical, structural, and evolutionary features without alignment or supervision. Embedding KS/CLF sequences with ESM-2 before classification significantly enhanced predictive performance: training accuracy increased to ~70%, and validation accuracy reached ~50%. Despite this improvement, the performance plateau suggested that the limited size of the labeled dataset and the high sequence similarity among KS/CLF homologs constrained the discriminative power of the model.

2.3.5 Unsupervised Clustering of 2315 Unlabeled KS/CLF Sequences

To expand the sequence space and explore whether evolutionary clustering could reveal natural chain-length groupings, an additional 2315 unlabeled KS/CLF sequences were collected from the NCBI database. These sequences were embedded using the same ESM-2 representation and subjected to various clustering approaches. Initial direct K-means clustering on the raw embeddings failed to produce meaningful separation corresponding to known chain lengths. Visualizations and cluster metrics indicated that sequences clustered primarily by phylogenetic origin rather than functional properties.

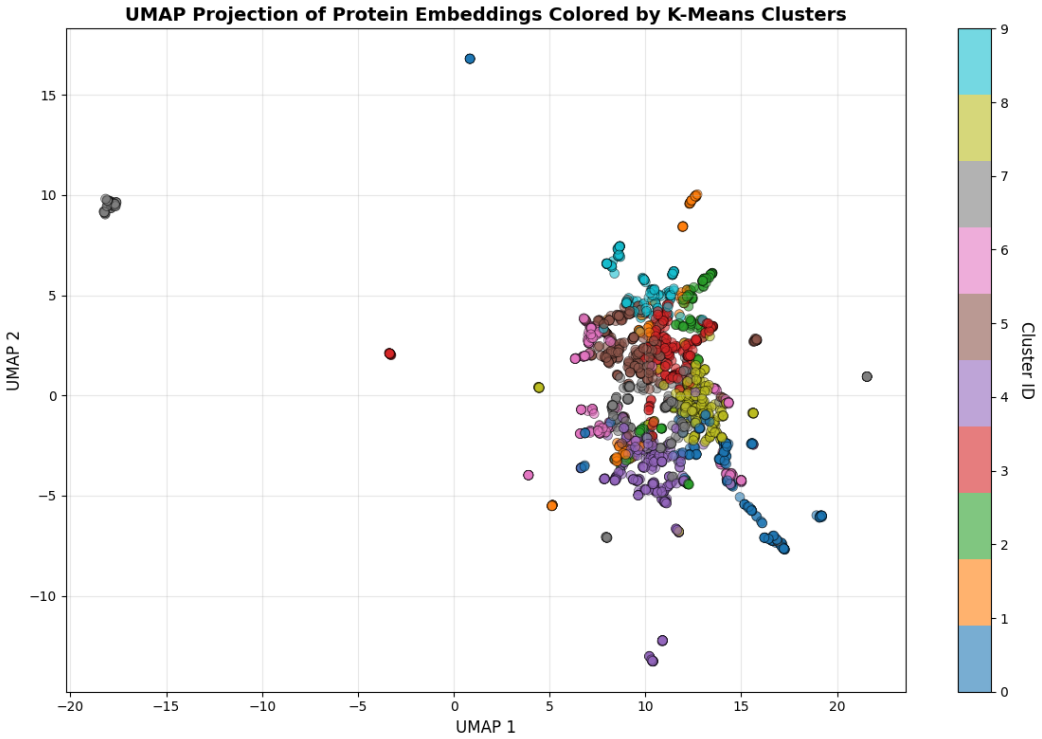


Figure 2.5 UMAP projection of KS/CLF protein sequences clusters. *Two-dimensional UMAP embedding of 2315 KS/CLF protein sequences based on embeddings generated ESM-2 with chain-length-relevant latent space created by PLS Points are colored by cluster ID.*

2.3.6 Partial Least Squares (PLS) Regression to Create a Chain-Length-Relevant Latent Space

To shift clustering toward functional features rather than general evolutionary similarity, Partial Least Squares (PLS) regression was applied to the labeled ESM-2 embeddings. PLS constructs a new latent space by identifying components that maximize covariance between the sequence embedding and chain-length output. Using ten PLS components, the model successfully emphasized patterns previously linked to chain-length differences while reducing irrelevant noise within the embedding.

The unlabeled dataset was then projected into this PLS-derived latent space and clustered using K-means. However, the resulting clusters exhibited substantial overlap among known chain-length classes. Silhouette scores and PCA/UMAP visualizations confirmed that no strong functional grouping emerged even in the PLS-filtered space.

2.3.7 Overall Assessment of Clustering Performance

Across all supervised and unsupervised attempts, clustering KS/CLF sequences into distinct chain-length classes remained challenging. The primary limitation appears to be the extremely high sequence conservation within the KS/CLF family, especially at functionally critical positions. Many KS/CLF pairs differ by only a few residues, and chain-length specificity often arises from subtle local features rather than domain-level or evolutionary separation. As a result, the sequence space lacks the clear functional boundaries required for robust clustering, even when using advanced embeddings and latent-space techniques.

2.4 Discussion:

Understanding how type II ketosynthase/chain-length factor (KS/CLF) complexes control polyketide chain length remains one of the central unresolved questions in aromatic polyketide biosynthesis. Although decades of biochemical and structural studies have established that chain length is dictated primarily by the geometry of the KS–CLF active-site tunnel,[66,90,166] translating primary sequence into predictable chain-length output remains fundamentally challenging. The present study sought to address this gap by developing machine-learning approaches capable of predicting chain-length control directly from KS/CLF sequence, while also probing whether unsupervised clustering of large unlabeled datasets can reveal new functional groupings. Collectively, the results reveal both the promise and limitations of sequence-based prediction for KS/CLF function, ultimately underscoring the exceptionally high evolutionary conservation within the minimal PKS and the structural subtlety underlying chain-length determination.

Our initial attempts using one-hot encoded sequences and a fully connected neural network yielded extremely poor performance, with high loss values and near-zero validation accuracy. This result is consistent with the well-known limitation that one-hot encodings do not capture long-range evolutionary constraints or structural context in enzyme families whose functional residues are sparse and highly interdependent.[167] Replacing this approach with a transformer-based sequence model modestly improved accuracy, suggesting that contextual embeddings better represent the distributed sequence signals governing chain length.

The most substantial improvement came from applying ESM-2, a state-of-the-art transformer model trained on ~65 million protein sequences.[165,168] ESM-2 embeddings increased

training accuracy to ~70% and validation accuracy to ~50%, demonstrating that large protein language models indeed capture richer structural and evolutionary information relevant to KS/CLF function. Comparable improvements in functional annotation using ESM-2 have been reported for enzyme commission prediction and protein engineering,[169] supporting its utility for biosynthetic enzyme prediction.

Nevertheless, despite these advances, performance remained insufficient for reliable annotation of unknown KS/CLFs. These limitations likely stem from the underlying biology rather than the model architecture. Structural work has repeatedly shown that chain-length control is dictated by very subtle geometric differences—often involving only a small set of residues at the "floor" of the acyl-binding tunnel or steric gatekeepers in CLF.[66,90,166] Such minimal sequence variation produces significant functional consequences, but these differences can be easily overwhelmed in large alignments dominated by highly conserved residues. Thus, even highly expressive language models may struggle to detect signals that are not strongly encoded at the sequence level.

To expand the searchable sequence space, we incorporated an additional [2,3]15 unlabeled KS/CLF sequences from NCBI and used ESM-2 embeddings followed by Partial Least Squares (PLS) regression to construct a latent space emphasizing chain-length features before clustering. However, clustering again failed to produce meaningful segregation by chain length. This outcome mirrors observations from evolutionary analyses showing that KS/CLF sequences across taxa share remarkably high identity, and that chain-length variation is often driven by only one or two residue changes.[126,170]

This result aligns with experimental exceptions such as the fdm minimal PKS, which unexpectedly produces both 11- and 12-unit chains despite being almost indistinguishable in sequence from related systems.[164] Collectively, these findings reinforce the idea that KS/CLF

function is not easily recoverable from global sequence similarity, and that clustering is inherently limited in enzyme families where the functional signal is carried by only a handful of residues.

A major obstacle to building accurate predictive models is the lack of high-quality labeled data. Experimental assignment of chain length is difficult because poly-β-ketone intermediates are chemically unstable and undergo spontaneous cyclization[171], making direct biochemical validation of elongation limits difficult. Historically, many mechanistic studies relied on indirect or truncated substrate analogs, such as diketide SNACs or isoxazole-based probes [172], which differ structurally from natural intermediates. Only recently have oxetane-based isosteres enabled more faithful stabilization of polyketide intermediates [172], but these tools have not yet been widely applied to large-scale annotation of KS/CLFs. As a result, training datasets remain small, skewed, and occasionally inconsistent.

The strong conservation of KS/CLF sequences reflects their position at the center of aromatic polyketide biosynthesis. KS and CLF must form an obligate heterodimer capable of binding ACP, stabilizing the growing chain, preventing premature cyclization, and enabling precise iteration of Claisen condensations[163,166]. Any mutations that interfere with dimerization or ACP binding are strongly selected against. Thus, evolution tolerates only minimal variation in the regions surrounding the active site, which in turn means that chain-length differences arise from extremely fine geometric modulations rather than broad sequence divergence.

These constraints help explain why our machine-learning and clustering approaches produced only modest predictive power: the functional signal is present but faint, encoded in only a small number of residues that collectively contribute to tunnel depth, shape, and electrostatics. Ultimately, the interplay between protein-protein interactions between KS/CLF-

ACP and substrate pocket shape will both contribute to the observed chain lengths in type II PKSs.

Although our models did not achieve high predictive accuracy, the study reveals important insights guiding future engineering efforts. First, ESM-2 embeddings clearly outperform classical encodings for KS/CLF prediction tasks, suggesting that future models trained on domain-specific datasets—perhaps incorporating structure-based embeddings or tunnel-shape descriptors—could further improve accuracy. Structural prediction tools such as AlphaFold3[173] and RoseTTAFold may enable generation of predicted KS/CLF tunnels that could be quantitatively analyzed using cavity-volume metrics that have proven predictive for chain length in modular PKSs.[167]

Second, combining ML with high-resolution structural biology—including cryo-EM of ACP–KS–CLF complexes or co-crystals with oxetane mimics—may provide the missing geometric labels needed to build more powerful predictive models. Recent advances in cryo-EM resolution for PKS systems make these approaches increasingly feasible.

Finally, more accurate prediction of chain-length control would have substantial implications for natural product engineering. Tailored manipulation of chain length has already yielded novel anthracyclines and polycyclic antibiotics with improved biological properties[66,163]. Improved predictive models could accelerate discovery of new polyketides with enhanced or diversified therapeutic activity, including derivatives of clinically important molecules such as doxorubicin and tetracycline.

2.5 Conclusion

In summary, our study highlights both the promise and the current limitations of sequence-based machine learning approaches for predicting KS/CLF chain-length specificity. Although progressively more expressive models—from simple fully connected networks to transformer architectures and finally ESM-2—yielded stepwise improvements in predictive performance, accuracy remained insufficient for confident functional assignment. Unsupervised clustering of >2,300 unlabeled KS/CLFs, even after embedding with ESM-2 and projection into a PLS-defined latent space, similarly failed to resolve chain-length–specific groups, underscoring the profound evolutionary conservation and restricted sequence divergence within this enzyme family. Together, these findings emphasize that sequence information alone captures only part of the determinants underlying KS/CLF specificity and that catalytic behavior is likely governed by subtle structural features, dynamic protein–protein interactions, and ACP-mediated substrate presentation. Moving forward, integrating structural modeling, coevolutionary analysis, and experimental biochemistry will be essential for building predictive frameworks that bridge the gap between sequence and function in aromatic polyketide biosynthesis.

2.6 Method

2.6.1 Expression and Purification of KS/CLF Protein

Act KS/CLF and Dps KS/CLF was expressed in Streptomyces coelicolor CH999 harboring the pRJC006 expression vector, following previously described methods with minor modifications (REF). Fresh spores were used to inoculate 50 mL of Super YEME medium supplemented with 50 μg/mL kanamycin. Seed cultures were incubated for 3 days at 30 °C with shaking at 250 rpm, then transferred into 2.5 L baffled flasks containing 500 mL of Super YEME with 50 μg/mL kanamycin. Cultures were grown for an additional 2 days under the same conditions before induction with 10 μg/mL thiostrepton.

Cells were harvested 24 h after induction by centrifugation (20 min, 5100 rpm), and pellets were stored at −80 °C. For purification, pellets from 2 L of culture were resuspended in 200 mL of lysis buffer (100 mM potassium phosphate, 15% glycerol, 300 mM NaCl, pH 7.5) supplemented with two tablets of EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonication on ice (10 cycles × 1 min). Insoluble material was removed by centrifugation (1 h, 14,000 rpm). Streptomycin sulfate was added to the clarified lysate to a final concentration of 2% (w/v) to precipitate nucleic acids, followed by centrifugation (45 min, 14,000 rpm). The resulting supernatant was filtered through a 0.45 μm membrane.

Protein was precipitated by adding ammonium sulfate to 30–50% saturation and incubating overnight at 4 °C. The precipitated protein was collected by centrifugation (30 min, 14,000 rpm) and resuspended in Ni-binding buffer (50 mM potassium phosphate, 10% glycerol, 500 mM NaCl, 5 mM imidazole, pH 7.5). The solution was filtered again (0.45 μm) and incubated with 2 mL of Ni-IMAC resin (Bio-Rad) pre-equilibrated in binding buffer, with gentle stirring at 4 °C for 1 h.

Bound protein was eluted using binding buffer containing increasing concentrations of imidazole. Fractions (2.5 mL) containing 375–750 mM imidazole were pooled and individually buffer-exchanged into storage buffer (100 mM potassium phosphate, 10% glycerol, pH 7.5) using PD-10 desalting columns (GE Healthcare). The purified protein was flash frozen and stored at −80 °C until use.

2.6.2 Sequence Encoding

The gene sequences in this study were from the MIBIG[174] and Integrated Microbial Genomes & Microbiomes system (IMG/M)[175]; their accession numbers or the source organisms are listed in SI Tables 1 and 2. Amino acid sequences were numerically encoded using a fixed vocabulary of 22 tokens consisting of the 20 canonical amino acids, a gap character ("–"), and a stop character ("*"). Each character was mapped to a unique integer index, and variable-length sequences were converted to integer tensors after padding. The encoding procedure was implemented in a custom function (pep2num).

Encoded and padded sequences were stored in a PyTorch Dataset subclass (FeatureDataset), which returned integer-encoded tensors paired with their corresponding chain-length labels. Data loading during training was handled by PyTorch DataLoader objects with shuffling enabled for training batches.

2.6.3 Transformer-Based Classification Model

Model Architecture

A transformer encoder architecture was implemented using PyTorch to classify KS/CLF sequences by predicted chain-length specificity. The model consisted of the following components:

1. Input Embedding Layer:

   Each integer-encoded residue token was mapped to a 512-dimensional dense vector

using a learnable embedding layer. The embeddings were scaled by the square root of the model dimension, following the original transformer formulation.

2. Positional Encoding:

   Because amino acid sequences lack intrinsic positional representation, sinusoidal positional encodings (Vaswani et al., 2017) were added to the token embeddings. These encodings used alternating sine and cosine functions across dimensions and were registered as non-trainable buffers.

3. Transformer Encoder:

   The classifier used a stack of three transformer encoder layers (n_layers = 3), each consisting of:

   - Multi-head self-attention with eight heads (n_head = 8)
   - A position-wise feedforward network of dimension 128
   - Dropout (0.1) between attention and feedforward blocks

4. Sequence Pooling:

   The encoder produced contextual embeddings for each residue position. These were averaged across sequence length (mean-pooling) to produce a single fixed-length representation per protein.

5. Output Layer:

   A fully connected linear layer projected the pooled embedding into trg_vocab_size dimensions, corresponding to the number of chain-length classes + 1. Predictions were obtained via Softmax during training through the cross-entropy loss.

Hyperparameters

- Embedding dimension: 512
- Number of transformer layers: 3
- Attention heads: 8
- Feedforward dimension: 128
- Dropout: 0.1
- Batch size: 3
- Learning rate: 0.001
- Number of epochs: 1000
- Loss function: Cross-entropy
- Optimizer: Adam
- Training device: NVIDIA GPU (CUDA), with fallback to CPU

2.6.4 Training Procedure

The model was trained for up to 1000 epochs using batches of integer-encoded sequences. For each batch:

1. Input tensors were reshaped to match transformer input format (sequence length × batch size).
2. Predictions were computed via the forward pass of the transformer classifier.
3. Cross-entropy loss between predicted chain-length class and ground-truth label was calculated.

4. Gradients were backpropagated, and model parameters were updated using Adam.

To reduce memory usage, GPU cache clearing (torch.cuda.empty_cache()) and Python garbage collection (gc.collect()) were performed once per epoch.

Training and validation losses were recorded for each epoch and stored in lists (losslist_tr and losslist_val). Loss curves were plotted using Matplotlib and exported as PNG files.

A minimal early-stopping logic was used to track the best-performing epoch: the model with the lowest validation loss was copied and stored in memory via copy.deepcopy().

2.6.5 Model Evaluation

After completing training, classification accuracy was computed on both the training and independent test datasets. Accuracy was calculated by comparing the argmax of the model logits to the true chain-length labels over all samples.

2.6.6 ESM-2 Model for Protein Representation

We employed the ESM-2 protein language model (ESM2-T33-650M-UR50D), a transformer-based model trained on 65 million unique protein sequences, which has demonstrated state-of-the-art performance in zero-shot prediction and structural representation tasks (Lin et al., 2023; Rives et al., 2021). The model was loaded in PyTorch and executed on GPU when available, with a batch converter provided by the ESM-2 library for formatting sequences into tokenized inputs.

Embedding Extraction

Protein sequences were converted into fixed-length numerical embeddings by mean-pooling the per-residue representations from layer 33 of the transformer. Specifically:

1. Sequences were batched (batch size = 16) for efficient GPU computation.

2. Each batch was passed through the ESM-2 model to obtain per-residue embeddings.

3. For each sequence, embeddings of non-padding tokens were averaged to generate a single 650-dimensional vector representing the full protein sequence.

This procedure was applied to both the labeled reference sequences and the unlabeled sequences.

2.6.7 PLS Regression for Chain-Length-Aware Latent Space

To incorporate chain-length information from the reference dataset, Partial Least Squares (PLS) Regression was used. PLS finds a latent space that maximally correlates the input features with output variables (Wold et al., 2001). The reference embeddings and their corresponding chain-length labels were used to fit a PLS model with 10 components, which emphasized patterns in the embedding space most relevant to chain-length determination. The fitted PLS model was then used to transform the embeddings of the unlabeled sequences into this latent space.

Unsupervised Clustering and Visualization

K-Means Clustering

The latent embeddings were clustered using K-Means (k = 10) to group sequences with similar chain-length-associated features. Cluster assignments were appended to the unlabeled dataset for subsequent analysis. Cluster sizes were visualized as a bar chart to evaluate the distribution of sequences among clusters.

2.6.8 UMAP Dimensionality Reduction

To visualize high-dimensional relationships, Uniform Manifold Approximation and Projection (UMAP) was applied to the PLS-transformed embeddings, reducing them to two dimensions while preserving local structure (McInnes et al., 2018). The resulting 2D coordinates were plotted, with points colored by K-Means cluster assignment, providing an interpretable representation of sequence similarity and clustering patterns.

2.6.9 Software and Libraries

- PyTorch (v1.13) for model inference and tensor operations
- ESM library for pretrained transformer models
- scikit-learn for PLS regression and K-Means clustering
- UMAP-learn for dimensionality reduction
- Matplotlib / Seaborn for visualization

2.7 References

(1) Cragg, G. M.; Newman, D. J. Biodiversity: A Continuing Source of Novel Drug Leads. *Pure Appl. Chem.* **2005**, *77*, 7–24.

(2) Buenz, E. J.; Verpoorte, R.; Bauer, B. A. The Ethnopharmacologic Contribution to Bioprospecting Natural Products. *Annu. Rev. Pharmacol. Toxicol.* **2018**, *58*, 509–530.

(3) Gaynes, R. The Discovery of Penicillin—New Insights after More than 75 Years of Clinical Use. *Emerg Infect Dis* **2017**, *23*, 849–853.

(4) Beutler, J. A. Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol.* **2019**, *86, e67*.

(5) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176.

(6) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

(7) Hopwood, D. A. Production of 'hybrid' Antibiotics by Genetic Engineering. *Nature* **1985**, *314*, 642–644.

(8) Malpartida, F.; Hopwood, D. A. Molecular Cloning of the Whole Biosynthetic Pathway of a Streptomyces Antibiotic and Its Expression in a Heterologous Host. *Nature* **1984**, *309* (5967), 462–464. https://doi.org/10.1038/309462a0.

(9) Bibb, M. J.; Sherman, D. H.; Ōmura, S.; Hopwood, D. A. Cloning, Sequencing and Deduced Functions of a Cluster of Streptomyces Genes Probably Encoding Biosynthesis of the Polyketide Antibiotic Frenolicin. *Gene* **1994**, *142*, 31–39.

(10) Han, L.; Yang, K.; Ramalingam, E.; Mosher, R. H.; Vining, L. C. Cloning and characterization of polyketide synthase genes for jadomycin b biosynthesis in streptomyces venezuelaeISP5230. *Microbiology* **1994**, *140*, 3379–3389.

(11) Motamedi, H.; Hutchinson, C. R. Cloning and Heterologous Expression of a Gene Cluster for the Biosynthesis of Tetracenomycin C, the Anthracycline Antitumor Antibiotic of Streptomyces Glaucescens. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4445–4449.

(12) Kim, E.-S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. Sequences of the Oxytetracycline Polyketide Synthase-Encoding Otc Genes from Streptomyces Rimosus. *Gene* **1994**, *141* (1), 141–142. https://doi.org/10.1016/0378-1119(94)90144-9.

(13) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. Characterization of the streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **1994**, *151*, 1–10.

(14) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Manipulation and Analysis of an Aromatic Polyketide Synthase with Unproved Catalytic Specificities. *J. Am. Chem. Soc.* **1993**, *115*, 11671–11675.

(15) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. The tcmVI Region of the Tetracenomycin c Biosynthetic Gene Cluster of Streptomyces Glaucescens Encodes the Tetracenomycin F1 Monooxygenase, Tetracenomycin F2 Cyclase, and, Most Likely, a Second Cyclase. *J. Bacteriol.* **1993**, *175*, 7571–7580.

(16) Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Stereochemical Course of Two Reactions Catalyzed by a Polyketide Synthase. *Biochemistry* **1994**, *33*, 9321–9326.

(17) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91* (24), 11542–11546. https://doi.org/10.1073/pnas.91.24.11542.

(18) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91*, 11542–11546.
(19) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Rational Design of Aromatic Polyketide Natural Products by Recombinant Assembly of Enzymatic Subunits. *Nature* **1995**, *375* (6532), 549–554. https://doi.org/10.1038/375549a0.
(20) Shen, B.; Summers, R. G.; Wendt-Pienkowski, E.; Hutchinson, C. R. The Streptomyces Glaucescens tcmKL Polyketide Synthase and tcmN Polyketide Cyclase Genes Govern the Size and Shape of Aromatic Polyketides. *J Am Chem Soc* **1995**, *117*, 6811–6821.
(21) Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. Engineered Biosynthesis of Novel Polyketides: Properties of the whiE Aromatase/Cyclase. *Nat Biotechnol* **1996**, *14*, 335–338.
(22) Kramer, P. J. Rational Design and Engineered Biosynthesis of a Novel 18-Carbon Aromatic Polyketide. *J Am Chem Soc* **1997**, *119*, 635–639.
(23) Meurer, G. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4*, 433–443.
(24) McDaniel, R. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel 'Unnatural' Natural Products. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 1846–1851.
(25) Harvey, A. L. Natural Products in Drug Discovery. *Drug Discov. Today* **2008**, *13*, 894–901.
(26) Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A. Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale for Catalytic Reprogramming. *J Am Chem Soc* **2014**, *136*, 7348–7362.
(27) Connor, C. J. O.; Beckmann, H. S. G.; Spring, D. R. Diversity-Oriented Synthesis: Producing Chemical Tools for Dissecting Biology. *Chem Soc Rev* **2012**, *41*, 4444–4456.
(28) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for Bad Bugs: Confronting the Challenges of Antibacterial Discovery. *Nat Rev Drug Discov* **2007**, *6*, 29–40.
(29) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat Rev Drug Discov* **2015**, *14*, 111–129.
(30) Schloss, P. D.; Handelsman, J. Metagenomics for Studying Unculturable Microorganisms: Cutting the Gordian Knot. *Genome Biol.* **2005**, *6*, 229.
(31) Medema, H.; M. The Year 2020 in Natural Product Bioinformatics: An Overview of the Latest Tools and Databases. *Nat. Prod. Rep.* **2021**, *38*, 301–306.
(32) Blin, K. antiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation. *Nucleic Acids Res.* **2023**, *51*, 46–50.
(33) Blin, K.; Shaw, S.; Kautsar, S. A.; Medema, M. H.; Weber, T. The antiSMASH Database Version 3: Increased Taxonomic Coverage and New Query Features for Modular Enzymes. *Nucleic Acids Res.* **2021**, *49*, 639–643.
(34) Hannigan, G. D. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res* **2019**, *47*, 110–110.
(35) Ichikawa, N. DoBISCUIT: A Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2013**, *41*, 408–414.

(36) Kautsar, S. A. MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* **2020**, *48*, 454–458.
(37) Palaniappan, K. IMG-ABC v.5.0: An Update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res.* **2020**, *48*, 422–430.
(38) Bentley, S. D. Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2. *Nature* **2002**, *417*, 141–147.
(39) Nett, M.; Ikeda, H.; Moore, B. S. Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. *Nat Prod Rep* **2009**, *26*, 1362–1384.
(40) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating Natural Product Combinatorial Biosynthesis with Synthetic Biology. *Nat Chem Biol* **2015**, *11*, 649–659.
(41) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8), 509–523. https://doi.org/10.1038/nrmicro3496.
(42) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat Rev Drug Discov* **2021**, *20*, 200–216.
(43) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem* **2021**, *90*, 763–788.
(44) Xu, F. A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat Chem Biol* **2019**, *15*, 161–168.
(45) Herbst, D. A.; Townsend, C. A.; Maier, T. The Architectures of Iterative Type i PKS and FAS. *Nat Prod Rep* **2018**, *35*, 1046–1069.
(46) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
(47) Dutta, S. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510*, 512–517.
(48) Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. *Science* **2007**, *316*, 288–290.
(49) Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315–1322.
(50) Staunton, J.; J. Weissman, K. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416. https://doi.org/10.1039/A909079G.
(51) Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Nat Prod Rep* **2003**, *20*, 79–110.
(52) Ferrer, J.-L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Structure of Chalcone Synthase and the Molecular Basis of Plant Polyketide Biosynthesis. *Nat Struct Mol Biol* **1999**, *6*, 775–784.
(53) Bisang, C. A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* **1999**, *401*, 502–505.
(54) Jackowski, S.; Rock, C. O. Acetoacetyl-Acyl Carrier Protein Synthase, a Potential Regulator of Fatty Acid Biosynthesis in Bacteria. *J. Biol. Chem.* **1987**, *262*, 7927–7931.
(55) Moore, B. S.; Hertweck, C. Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units. *Nat Prod Rep* **2002**, *19*, 70–99.
(56) Crawford, J. M.; Dancy, B. C. R.; Hill, E. A.; Udwary, D. W.; Townsend, C. A. Identification of a Starter Unit Acyl-Carrier Protein Transacylase Domain in an Iterative Type I Polyketide Synthase. *Proc. Natl. Acad. Sci.* **2006**, *103* (45), 16728–16733. https://doi.org/10.1073/pnas.0604112103.

(57) Tsai, S.-C. (Sheryl). The Structural Enzymology of Iterative Aromatic Polyketide Synthases: A Critical Comparison with Fatty Acid Synthases. *Annu. Rev. Biochem.* **2018**, *87* (1), 503–531. https://doi.org/10.1146/annurev-biochem-063011-164509.
(58) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat Prod Rep* **2016**, *33*, 150–161.
(59) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **2009**, *26*, 90–114.
(60) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* **1998**, *37* (22), 8132–8138. https://doi.org/10.1021/bi980466i.
(61) Meadows, E. S.; Khosla, C. In Vitro Reconstitution and Analysis of the Chain Initiating Enzymes of the R1128 Polyketide Synthase. *Biochemistry* **2001**, *40*, 14855–14861.
(62) Winter, J. M. Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, *137*, 9885–9893.
(63) Cacho, R. A. Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J Am Chem Soc* **2015**, *137*, 15688–15691.
(64) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide b Synthase. *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
(65) Crawford, J. M. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320*, 243–246.
(66) Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709. https://doi.org/10.1021/ja0378759.
(67) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Dissection of the Catalytic Specificity of the Act Ketoreductase. *J. Am. Chem. Soc.* **1994**, *116*, 4166–4170.
(68) Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Evidence for Temporal, but Not Regiospecific, Control of Cyclization of an Aromatic Polyketide Precursor. *Chem. Biol.* **1994**, *1*, 205–210.
(69) Shen, Y. Ectopic Expression of the Minimal whiE Polyketide Synthase Generates a Library of Aromatic Polyketides of Diverse Sizes and Shapes. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 3622–3627.
(70) Auclair, K. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels–alder Reaction of a Substrate Analogue. *J Am Chem Soc* **2000**, *122*, 11519–11520.
(71) Kim, H. J.; Ruszczycky, M. W.; Choi, S.; Liu, Y.; Liu, H. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn a. *Nature* **2011**, *473*, 109–112.
(72) Kato, N. A New Enzyme Involved in the Control of the Stereochemistry in the Decalin Formation during Equisetin Biosynthesis. *Biochem. Biophys. Res. Commun.* **2015**, *460*, 210–215.
(73) Kasahara, K. Solanapyrone Synthase, a Possible Diels–Alderase and Iterative Type i Polyketide Synthase Encoded in a Biosynthetic Gene Cluster from Alternaria Solani. *ChemBioChem* **2010**, *11*, 1245–1252.
(74) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of Claisen Cyclase Domain in Fungal Polyketide Synthase WA, a Naphthopyrone Synthase of Aspergillus Nidulans. *Chem. Biol.* **2001**, *8*, 189–197.
(75) Crawford, J. M. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461*, 1139–1143.

(76) Ames, B. D. Crystal Structure and Functional Analysis of Tetracenomycin ARO/CYC: Implications for Cyclization Specificity of Aromatic Polyketides. *PNAS* **2008**, *105*, 5349–5354.
(77) Ames, B. D. Structural and Biochemical Characterization of ZhuI Aromatase/Cyclase from the R1128 Polyketide Pathway. *Biochemistry* **2011**, *50*, 8392–8406.
(78) Yu, T.-W. Engineered biosynthesis of novel polyketides from streptomyces spore pigment polyketide synthases. *J Am Chem Soc* **1998**, *120*, 7749–7759.
(79) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively. *J. Am. Chem. Soc.* **1994**, *116*, 10855–10859.
(80) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of Post-PKS Tailoring Steps through Combinatorial Biosynthesis. *Nat Prod Rep* **2002**, *19*, 542–580.
(81) Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. Cytochromes P450 for Natural Product Biosynthesis in Streptomyces: Sequence, Structure, and Function. *Nat Prod Rep* **2017**, *34*, 1141–1172.
(82) Chung, W.; Vanderwal, C. D. Stereoselective Halogenation in Natural Product Synthesis. *Angew. Chem. Int. Ed.* **2016**, *55*, 4396–4434.
(83) Singh, S., Jr; P., G. N.; Thorson, J. S. The Structural Biology of Enzymes Involved in Natural Product Glycosylation. *Nat Prod Rep* **2012**, *29*, 1201–1237.
(84) Podust, L. M.; Sherman, D. H. Diversity of P450 Enzymes in the Biosynthesis of Natural Products. *Nat Prod Rep* **2012**, *29*, 1251–1266.
(85) Carreras, C. W.; Pieper, R.; Khosla, C. Efficient Synthesis of Aromatic Polyketides in Vitro by the Actinorhodin Polyketide Synthase. *J Am Chem Soc* **1996**, *118*, 5158–5159.
(86) Shen, B.; Hutchinson, C. R. Deciphering the Mechanism for the Assembly of Aromatic Polyketides by a Bacterial Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1996**, *93*, 6600–6604.
(87) Harris, T. M.; Harris, C. M. Biomimetic Syntheses of Aromatic Polyketide Metabolites. *Pure Appl Chem* **1986**, *58*, 283–294.
(88) Sherman, D. h. Structure and Deduced Function of the Granaticin-Producing Polyketide Synthase Gene Cluster of Streptomyces Violaceoruber Tü22. *EMBO J.* **1989**, *8*, 2717–2725.
(89) Kim, E. S.; Cramer, K. D.; Shreve, A. L.; Sherman, D. H. Heterologous Expression of an Engineered Biosynthetic Pathway: Functional Dissection of Type II Polyketide Synthase Components in Streptomyces Species. *J Bacteriol* **1995**, *177*, 1202–1207.
(90) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11* (9), 888–893. https://doi.org/10.1038/nsmb808.
(91) Sherman, D. H.; Kim, E. S.; Bibb, M. J.; Hopwood, D. A. Functional Replacement of Genes for Individual Polyketide Synthase Components in Streptomyces Coelicolor A3(2) by Heterologous Genes from a Different Polyketide Pathway. *J. Bacteriol.* **1992**, *174*, 6184–6190.
(92) Kim, E. S.; Hopwood, D. A.; Sherman, D. H. Analysis of Type II Polyketide Beta-Ketoacyl Synthase Specificity in Streptomyces Coelicolor A3(2) by Trans Complementation of Actinorhodin Synthase Mutants. *J Bacteriol* **1994**, *176*, 1801–1804.
(93) Dreier, J.; Khosla, C. Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the β-Ketoacyl Synthase–chain Length Factor Heterodimer. *Biochemistry* **2000**, *39*, 2088–2095.

(94) Grammbitter, G. L. C. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. *J Am Chem Soc* **2019**, *141*, 16615–16623.
(95) Bräuer, A. Structural Snapshots of the Minimal PKS System Responsible for Octaketide Biosynthesis. *Nat Chem* **2020**, *12*, 755–763.
(96) Nicholson, T. P. First in Vitro Directed Biosynthesis of New Compounds by a Minimal Type II Polyketide Synthase: Evidence for the Mechanism of Chain Length Determination. *Chem Commun* **2003**, *686–687*. https://doi.org/10.1039/B300847A.
(97) Kim, Y.; Prestegard, J. H. Refinement of the NMR Structures for Acyl Carrier Protein with Scalar Coupling Data. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 377–385.
(98) Evans, S. E. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* **2008**, *9*, 2424–2432.
(99) Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein. *Biochemistry* **2003**, *42*, 4648–4657.
(100) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. Solution Structure of the Actinorhodin Polyketide Synthase Acyl Carrier Protein from Streptomyces Coelicolor A3(2),. *Biochemistry* **1997**, *36* (20), 6000–6008. https://doi.org/10.1021/bi970006+.
(101) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. Targeted Gene Replacements in a Streptomyces Polyketide Synthase Gene Cluster: Role for the Acyl Carrier Protein. *Mol. Microbiol.* **1992**, *6*, 3237–3249.
(102) Zhang, Y.-M.; Wu, B.; Zheng, J.; Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction*. *J. Biol. Chem.* **2003**, *278*, 52935–52943.
(103) Arthur, C. J. Structure and Malonyl CoA-ACP Transacylase Binding of Streptomyces Coelicolor Fatty Acid Synthase Acyl Carrier Protein. *ACS Chem Biol* **2009**, *4*, 625–636.
(104) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-Based Analysis of the Molecular Interactions between Acyltransferase and Acyl Carrier Protein in Vicenistatin Biosynthesis. *PNAS* **2016**, *113*, 1802–1807.
(105) Bartholow, T. G. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4*, 1–10.
(106) Cho, Y. I. Engineered Chimeras Unveil Swappable Modular Features of Fatty Acid and Polyketide Synthase Acyl Carrier Proteins. *Biochemistry* **2022**, *61*, 217–227.
(107) Khosla, C. Genetic Construction and Functional Analysis of Hybrid Polyketide Synthases Containing Heterologous Acyl Carrier Proteins. *J. Bacteriol.* **1993**, *175*, 2197–2204.
(108) Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C. Ketosynthases in the Initiation and Elongation Modules of Aromatic Polyketide Synthases Have Orthogonal Acyl Carrier Protein Specificity†, 2003. https://doi.org/10.1021/BI0341962.
(109) Das, A.; Khosla, C. In Vivo and in Vitro Analysis of the Hedamycin Polyketide Synthase. *Chem. Biol.* **2009**, *16*, 1197–1207.
(110) Koppisch, A. T.; Khosla, C. Structure-Based Mutagenesis of the Malonyl-CoA:Acyl Carrier Protein Transacylase from Streptomyces Coelicolor. *Biochemistry* **2003**, *42*, 11057–11064.
(111) Beltran-Alvarez, P.; Cox, R. J.; Crosby, J.; Simpson, T. J. Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* **2007**, *46*, 14672–14681.

(112) Arthur, C. J. Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* **2005**, *44*, 15414–15421.
(113) Kalaitzis, J. A. Policing Starter Unit Selection of the Enterocin Type II Polyketide Synthase by the Type II Thioesterase EncL. *Bioorg. Med. Chem.* **2011**, *19*, 6633–6638.
(114) Shepherd, M. D.; Kharel, M. K.; Zhu, L. L.; Lanen, S. G. van; Rohr, J. Delineating the Earliest Steps of Gilvocarcin Biosynthesis: Role of GilP and GilQ in Starter Unit Specificity. *Org Biomol Chem* **2010**, *8*, 3851–3856.
(115) Tang, Y.; Koppisch, A. T.; Khosla, C. The Acyltransferase Homologue from the Initiation Module of the R1128 Polyketide Synthase Is an Acyl-ACP Thioesterase That Edits Acetyl Primer Units. *Biochemistry* **2004**, *43* (29), 9546–9555. https://doi.org/10.1021/bi049157k.
(116) Keatinge-Clay, A. T. Catalysis, Specificity, and ACP Docking Site of Streptomyces Coelicolor Malonyl-CoA:ACP Transacylase. *Structure* **2003**, *11*, 147–154.
(117) Szu, P.-H. Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **2011**, *18*, 1021–1031.
(118) Bartel, P. L. Biosynthesis of Anthraquinones by Interspecies Cloning of Actinorhodin Biosynthesis Genes in Streptomycetes: Clarification of Actinorhodin Gene Functions. *J. Bacteriol.* **1990**, *172*, 4816–4826.
(119) Rajgarhia, V. B.; Strohl, W. R. Minimal streptomyces sp. Strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis. *J Bacteriol* **1997**, *179*, 2690–2696.
(120) Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C. Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKR): Molecular Basis of Stereo- and Regiospecificities. *Biochemistry* **2011**, *50*, 7426–7439.
(121) Piel, J. Cloning, Sequencing and Analysis of the Enterocin Biosynthesis Gene Cluster from the Marine Isolate 'Streptomyces Maritimus': Evidence for the Derailment of an Aromatic Polyketide Synthase. *Chem. Biol.* **2000**, *7*, 943–955.
(122) Price, A. C.; Zhang, Y. M.; Rock, C. O.; White, S. W. Structure of β-Ketoacyl-[Acyl Carrier Protein] Reductase from Escherichia Coli: Negative Cooperativity and Its Structural Basis. *Biochemistry* **2001**, *40*, 12772–12781.
(123) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43* (46), 14529–14538. https://doi.org/10.1021/bi048133a.
(124) Oppermann, U. Short-Chain Dehydrogenases/Reductases (SDR): The 2002 Update. *Chem. Biol. Interact.* **2003**, *143–144*, 247–253.
(125) Filling, C. Critical Residues for Structure and Catalysis in Short-Chain Dehydrogenases/Reductases*. *J. Biol. Chem.* **2002**, *277*, 25677–25684.
(126) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **1993**, *262* (5139), 1546–1550.
(127) Kalaitzis, J. A.; Moore, B. S. Heterologous Biosynthesis of Truncated Hexaketides Derived from the Actinorhodin Polyketide Synthase. *J. Nat. Prod.* **2004**, *67*, 1419–1422.
(128) Zawada, R. J.; Khosla, C. Heterologous Expression, Purification, Reconstitution and Kinetic Analysis of an Extended Type II Polyketide Synthase. *Chem. Biol.* **1999**, *6* (9), 607–615. https://doi.org/10.1016/s1074-5521(99)80112-4.
(129) Hadfield, A. T. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12*, 1865–1875.

(130) Kantola, J. Folding of the Polyketide Chain Is Not Dictated by Minimal Polyketide Synthase in the Biosynthesis of Mithramycin and Anthracycline. *Chem. Biol.* **1997**, *4*, 751–755.
(131) Kunnari, T. Incorrectly Folded Aromatic Polyketides from Polyketide Reductase Deficient Mutants. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2639–2642.
(132) Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C. Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase,. *Biochemistry* **2008**, *47* (7), 1837–1847. https://doi.org/10.1021/bi7016427.
(133) Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C. C. Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochemistry* **2011**, *50*, 4638–4649.
(134) Javidpour, P. The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chem. Biol.* **2013**, *20*, 1225–1234.
(135) Zhang, H. L. Mutactin, a Novel Polyketide from Streptomyces Coelicolor. Structure and Biosynthetic Relationship to Actinorhodin. *J Org Chem* **1990**, *55*, 1682–1684.
(136) Zhao, S. Molecular Basis for Polyketide Ketoreductase–Substrate Interactions. *Int. J. Mol. Sci.* **2020**, *21*, 7562.
(137) Kharel, M. K. Elucidation of post-PKS tailoring steps involved in landomycin biosynthesis. *Org Biomol Chem* **2012**, *10*, 4256–4265.
(138) Mayer, A. LanV, a Bifunctional Enzyme: Aromatase and Ketoreductase during Landomycin a Biosynthesis. *ChemBioChem* **2005**, *6*, 2312–2315.
(139) Dickens, M. L.; Ye, J.; Strohl, W. R. Cloning, Sequencing, and Analysis of Aklaviketone Reductase from Streptomyces Sp. Strain C5. *J. Bacteriol.* **1996**, *178*, 3384–3388.
(140) Lackner, G. Biosynthesis of Pentangular Polyphenols: Deductions from the Benastatin and Griseorhodin Pathways. *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
(141) Torkkell, S. The Entire Nogalamycin Biosynthetic Gene Cluster of Streptomyces Nogalater: Characterization of a 20-Kb DNA Region and Generation of Hybrid Structures. *Mol Gen Genomics* **2001**, *266*, 276–288.
(142) Paananen, P. Structural and Functional Analysis of Angucycline C-6 Ketoreductase LanV Involved in Landomycin Biosynthesis. *Biochemistry* **2013**, *52*, 5304–5314.
(143) Patrikainen, P. Structure-Based Engineering of Angucyclinone 6-Ketoreductases. *Chem. Biol.* **2014**, *21*, 1381–1391.
(144) Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C. Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (12), 3421–3430. https://doi.org/10.1021/acschembio.6b00658.
(145) Xiao, X. Functional and Structural Insights into a Novel Promiscuous Ketoreductase of the Lugdunomycin Biosynthetic Pathway. *ACS Chem Biol* **2020**, *15*, 2529–2538.
(146) Kallio, P. Flavoprotein Hydroxylase PgaE Catalyzes Two Consecutive Oxygen-Dependent Tailoring Reactions in Angucycline Biosynthesis. *Biochemistry* **2011**, *50*, 5535–5543.
(147) Patrikainen, P. Tailoring Enzymes Involved in the Biosynthesis of Angucyclines Contain Latent Context-Dependent Catalytic Activities. *Chem. Biol.* **2012**, *19*, 647–655.
(148) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **2003**, *4*, 654–657.
(149) Moreno, G. O. *Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type i and Type II Polyketide Synthases Using Atom-Replaced Probes*; UC Irvine, 2018.

(150) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(151) Li 然 李嫣, Y.; Xu 伟徐, W.; Tang 奕唐, Y. Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains*. *J. Biol. Chem.* **2010**, *285* (30), 22764–22773. https://doi.org/10.1074/jbc.M110.128504.
(152) Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C. A. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461* (7267), 1139–1143. https://doi.org/10.1038/nature08475.
(153) Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. *Proc. Natl. Acad. Sci.* **2017**, *114* (21), E4142–E4148. https://doi.org/10.1073/pnas.1609001114.
(154) Liu, L.; Zhang, Z.; Shao, C.-L.; Wang, J.-L.; Bai, H.; Wang, C.-Y. Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Sci. Rep.* **2015**, *5* (1), 10463. https://doi.org/10.1038/srep10463.
(155) Feng, Y.; Yang, X.; Ji, H.; Deng, Z.; Lin, S.; Zheng, J. The Streptomyces Viridochromogenes Product Template Domain Represents an Evolutionary Intermediate between Dehydratase and Aldol Cyclase of Type I Polyketide Synthases. *Commun. Biol.* **2022**, *5* (1), 508. https://doi.org/10.1038/s42003-022-03477-8.
(156) Pan, H. Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. *Structure* **2002**, *10*, 1559–1568.
(157) Thompson, T. B.; Katayama, K.; Watanabe, K.; Hutchinson, C. R.; Rayment, I. Structural and Functional Analysis of Tetracenomycin F2 Cyclase from Streptomyces Glaucescens: A TYPE II POLYKETIDE CYCLASE*. *J. Biol. Chem.* **2004**, *279*, 37956–37963.
(158) Sultana, A. Structure of the Polyketide Cyclase SnoaL Reveals a Novel Mechanism for Enzymatic Aldol Condensation. *EMBO J.* **2004**, *23*, 1911–1921.
(159) Kallio, P.; Sultana, A.; Niemi, J.; Mäntsälä, P.; Schneider, G. Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity. *J. Mol. Biol.* **2006**, *357*, 210–220.
(160) Hutchinson, C. R. Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem. Rev.* **1997**, *97* (7), 2525–2536. https://doi.org/10.1021/cr960022x.
(161) Ahuja, M.; Chiang, Y.-M.; Chang, S.-L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H.-C.; Yeh, H.-H.; Oakley, B. R.; Wang, C. C. C. Illuminating the Diversity of Aromatic Polyketide Synthases in Aspergillus Nidulans. *J. Am. Chem. Soc.* **2012**, *134* (19), 8212–8221. https://doi.org/10.1021/ja3016395.
(162) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97* (7), 2465–2498. https://doi.org/10.1021/cr960034i.
(163) Shen, B. Polyketide Biosynthesis beyond the Type I, II and III Polyketide Synthase Paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6.
(164) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639. https://doi.org/10.1021/ar8002249.
(165) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A.

Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130. https://doi.org/10.1126/science.ade2574.
(166) Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. Structural Basis for Selectivity in a Highly Reducing Type II Polyketide Synthase. *Nat. Chem. Biol.* **2020**, *16* (7), 776–782. https://doi.org/10.1038/s41589-020-0530-0.
(167) Yadav, G.; Gokhale, R. S.; Mohanty, D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J. Mol. Biol.* **2003**, *328* (2), 335–363. https://doi.org/10.1016/S0022-2836(03)00232-8.
(168) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci.* **2021**, *118* (15), e2016239118. https://doi.org/10.1073/pnas.2016239118.
(169) Capela, J.; Zimmermann-Kogadeeva, M.; Dijk, A. D. J. van; de Ridder, D.; Dias, O.; Rocha, M. Comparative Assessment of Protein Large Language Models for Enzyme Commission Number Prediction. *BMC Bioinformatics* **2025**, *26* (1), 68. https://doi.org/10.1186/s12859-025-06081-9.
(170) Jacques, M.-A.; Durand, K.; Orgeur, G.; Balidas, S.; Fricot, C.; Bonneau, S.; Quillévéré, A.; Audusseau, C.; Olivier, V.; Grimault, V.; Mathis, R. Phylogenetic Analysis and Polyphasic Characterization of Clavibacter Michiganensis Strains Isolated from Tomato Seeds Reveal That Nonpathogenic Strains Are Distinct from C. Michiganensis Subsp. Michiganensis. *Appl. Environ. Microbiol.* **2012**, *78* (23), 8388–8402. https://doi.org/10.1128/AEM.02158-12.
(171) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat. Chem. Biol.* **2019**, *15* (7), 669–671. https://doi.org/10.1038/s41589-019-0301-y.
(172) Ellis, B. D.; Milligan, J. C.; White, A. R.; Duong, V.; Altman, P. X.; Mohammed, L. Y.; Crump, M. P.; Crosby, J.; Luo, R.; Vanderwal, C. D.; Tsai, S.-C. An Oxetane-Based Polyketide Surrogate To Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140* (15), 4961–4964. https://doi.org/10.1021/jacs.7b11793.
(173) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.
(174) Zdouc, M. M.; Blin, K.; Louwen, N. L. L.; Navarro, J.; Loureiro, C.; Bader, C. D.; Bailey, C. B.; Barra, L.; Booth, T. J.; Bozhüyük, K. A. J.; Cediel-Becerra, J. D. D.; Charlop-Powers, Z.; Chevrette, M. G.; Chooi, Y. H.; D'Agostino, P. M.; de Rond, T.; Del Pup, E.; Duncan, K. R.; Gu, W.; Hanif, N.; Helfrich, E. J. N.; Jenner, M.; Katsuyama, Y.; Korenskaia, A.; Krug, D.; Libis, V.; Lund, G. A.; Mantri, S.; Morgan, K. D.; Owen, C.; Phan, C.-S.; Philmus, B.; Reitz, Z. L.; Robinson, S. L.; Singh, K. S.; Teufel, R.; Tong, Y.; Tugizimana, F.; Ulanova, D.;

Winter, J. M.; Aguilar, C.; Akiyama, D. Y.; Al-Salihi, S. A. A.; Alanjary, M.; Alberti, F.; Aleti, G.; Alharthi, S. A.; Rojo, M. Y. A.; Arishi, A. A.; Augustijn, H. E.; Avalon, N. E.; Avelar-Rivas, J. A.; Axt, K. K.; Barbieri, H. B.; Barbosa, J. C. J.; Barboza Segato, L. G.; Barrett, S. E.; Baunach, M.; Beemelmanns, C.; Beqaj, D.; Berger, T.; Bernaldo-Agüero, J.; Bettenbühl, S. M.; Bielinski, V. A.; Biermann, F.; Borges, R. M.; Borriss, R.; Breitenbach, M.; Bretscher, K. M.; Brigham, M. W.; Buedenbender, L.; Bulcock, B. W.; Cano-Prieto, C.; Capela, J.; Carrion, V. J.; Carter, R. S.; Castelo-Branco, R.; Castro-Falcón, G.; Chagas, F. O.; Charria-Girón, E.; Chaudhri, A. A.; Chaudhry, V.; Choi, H.; Choi, Y.; Choupannejad, R.; Chromy, J.; Donahey, M. S. C.; Collemare, J.; Connolly, J. A.; Creamer, K. E.; Crüsemann, M.; Cruz, A. A.; Cumsille, A.; Dallery, J.-F.; Damas-Ramos, L. C.; Damiani, T.; de Kruijff, M.; Martín, B. D.; Sala, G. D.; Dillen, J.; Doering, D. T.; Dommaraju, S. R.; Durusu, S.; Egbert, S.; Ellerhorst, M.; Faussurier, B.; Fetter, A.; Feuermann, M.; Fewer, D. P.; Foldi, J.; Frediansyah, A.; Garza, E. A.; Gavriilidou, A.; Gentile, A.; Gerke, J.; Gerstmans, H.; Gomez-Escribano, J. P.; González-Salazar, L. A.; Grayson, N. E.; Greco, C.; Gomez, J. E. G.; Guerra, S.; Flores, S. G.; Gurevich, A.; Gutiérrez-García, K.; Hart, L.; Haslinger, K.; He, B.; Hebra, T.; Hemmann, J. L.; Hindra, H.; Höing, L.; Holland, D. C.; Holme, J. E.; Horch, T.; Hrab, P.; Hu, J.; Huynh, T.-H.; Hwang, J.-Y.; Iacovelli, R.; Iftime, D.; Iorio, M.; Jayachandran, S.; Jeong, E.; Jing, J.; Jung, J. J.; Kakumu, Y.; Kalkreuter, E.; Kang, K. B.; Kang, S.; Kim, W.; Kim, G. J.; Kim, H.; Kim, H. U.; Klapper, M.; Koetsier, R. A.; Kollten, C.; Kovács, Á. T.; Kriukova, Y.; Kubach, N.; Kunjapur, A. M.; Kushnareva, A. K.; Kust, A.; Lamber, J.; Larralde, M.; Larsen, N. J.; Launay, A. P.; Le, N.-T.-H.; Lebeer, S.; Lee, B. T.; Lee, K.; Lev, K. L.; Li, S.-M.; Li, Y.-X.; Licona-Cassani, C.; Lien, A.; Liu, J.; Lopez, J. A. V.; Machushynets, N. V.; Macias, M. I.; Mahmud, T.; Maleckis, M.; Martinez-Martinez, A. M.; Mast, Y.; Maximo, M. F.; McBride, C. M.; McLellan, R. M.; Bhatt, K. M.; Melkonian, C.; Merrild, A.; Metsä-Ketelä, M.; Mitchell, D. A.; Müller, A. V.; Nguyen, G.-S.; Nguyen, H. T.; Niedermeyer, T. H. J.; O'Hare, J. H.; Ossowicki, A.; Ostash, B. O.; Otani, H.; Padva, L.; Paliyal, S.; Pan, X.; Panghal, M.; Parade, D. S.; Park, J.; Parra, J.; Rubio, M. P.; Pham, H. T.; Pidot, S. J.; Piel, J.; Pourmohsenin, B.; Rakhmanov, M.; Ramesh, S.; Rasmussen, M. H.; Rego, A.; Reher, R.; Rice, A. J.; Rigolet, A.; Romero-Otero, A.; Rosas-Becerra, L. R.; Rosiles, P. Y.; Rutz, A.; Ryu, B.; Sahadeo, L.-A.; Saldanha, M.; Salvi, L.; Sánchez-Carvajal, E.; Santos-Medellin, C.; Sbaraini, N.; Schoellhorn, S. M.; Schumm, C.; Sehnal, L.; Selem, N.; Shah, A. D.; Shishido, T. K.; Sieber, S.; Silviani, V.; Singh, G.; Singh, H.; Sokolova, N.; Sonnenschein, E. C.; Sosio, M.; Sowa, S. T.; Steffen, K.; Stegmann, E.; Streiff, A. B.; Strüder, A.; Surup, F.; Svenningsen, T.; Sweeney, D.; Szenei, J.; Tagirdzhanov, A.; Tan, B.; Tarnowski, M. J.; Terlouw, B. R.; Rey, T.; Thome, N. U.; Torres Ortega, L. R.; Tørring, T.; Trindade, M.; Truman, A. W.; Tvilum, M.; Udwary, D. W.; Ulbricht, C.; Vader, L.; van Wezel, G. P.; Walmsley, M.; Warnasinghe, R.; Weddeling, H. G.; Weir, A. N. M.; Williams, K.; Williams, S. E.; Witte, T. E.; Rocca, S. M. W.; Yamada, K.; Yang, D.; Yang, D.; Yu, J.; Zhou, Z.; Ziemert, N.; Zimmer, L.; Zimmermann, A.; Zimmermann, C.; van der Hooft, J. J. J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 4.0: Advancing Biosynthetic Gene Cluster Curation through Global Collaboration. *Nucleic Acids Res.* **2025**, *53* (D1), D678–D690. https://doi.org/10.1093/nar/gkae1115.

(175) Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S. J.; Webb, C.; Wu, D.; Varghese, N. J.; Reddy, T. B. K.; Mukherjee, S.; Ovchinnikova, G.; Nolan, M.; Seshadri, R.; Roux, S.; Visel, A.; Woyke, T.; Eloe-Fadrosh, E. A.; Kyrpides, N. C.; Ivanova, N. N. The IMG/M Data Management and Analysis System

v.7: Content Updates and New Features. *Nucleic Acids Res.* **2023**, *51* (D1), D723–D732. https://doi.org/10.1093/nar/gkac976.
(176) Fernández-Moreno, M. A.; Martínez, E.; Caballero, J. L.; Ichinose, K.; Hopwood, D. A.; Malpartida, F. DNA Sequence and Functions of the actVI Region of the Actinorhodin Biosynthetic Gene Cluster of Streptomyces Coelicolor A3(2). *J. Biol. Chem.* **1994**, *269* (40), 24854–24863. https://doi.org/10.1016/S0021-9258(17)31469-2.
(177) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Richard Hutchinson, C. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4* (6), 433–443. https://doi.org/10.1016/S1074-5521(97)90195-2.
(178) Moreno, G. O. Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type I and Type II Polyketide Synthases Using Atom-Replaced Probes, UC Irvine, 2018. https://escholarship.org/uc/item/1tx0518k.
(179) Serapian, S. A.; Crosby, J.; Crump, M. P.; van der Kamp, M. W. Path to Actinorhodin: Regio- and Stereoselective Ketone Reduction by a Type II Polyketide Ketoreductase Revealed in Atomistic Detail. *JACS Au* **2022**, *2* (4), 972–984. https://doi.org/10.1021/jacsau.2c00086.
(180) Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941–953. https://doi.org/10.1016/j.jmb.2008.09.084.
(181) Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12* (10), 1865–1875. https://doi.org/10.1016/j.str.2004.08.002.
(182) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat. Prod. Rep.* **2016**, *33* (2), 203–230. https://doi.org/10.1039/C5NP00109A.
(183) Chooi, Y.-H.; Tang, Y. Navigating the Fungal Polyketide Chemical Space: From Genes to Molecules. *J. Org. Chem.* **2012**, *77* (22), 9933–9953. https://doi.org/10.1021/jo301592k.
(184) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(185) Barajas, J. F.; Finzel, K.; Valentic, T. R.; Shakya, G.; Gamarra, N.; Martinez, D.; Meier, J. L.; Vagstad, A. L.; Newman, A. G.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Structural and Biochemical Analysis of Protein–Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew. Chem. Int. Ed.* **2016**, *55* (42), 13005–13009. https://doi.org/10.1002/anie.201605401.
(186) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(187) Barajas, J. F. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. In *Proceedings of the National Academy of Sciences*; 2017; Vol. 114, pp 4142–4148.
(188) Milligan, J. C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat Chem Biol* **2019**, *15*, 669–671.

(189) Ellis, B. D. An Oxetane-Based Polyketide Surrogate to Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140*, 4961–4964.
(190) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43*, 14529–14538.

Chapter 3

Lid Helices and ACP Identity as Determinants of Ketoreductase Specificity in Type II Polyketide Synthases

3.1 Summary

Type II polyketide synthases (PKSs) rely on highly specific protein–protein interactions to control the regiospecific tailoring of their polyketide intermediates. In this chapter, we investigate the C9 ketoreductase (DpsKR) from the doxorubicin PKS and uncover how structural features of the lid helices and the identity of the acyl carrier protein (ACP) govern substrate reduction. In vitro reconstitution assays showed that wild-type DpsKR failed to catalyze C9 reduction in the actinorhodin minimal PKS, whereas a chimeric DpsKR containing the ActKR lid helices restored its activity. ACP-swapping experiments further revealed that productive reduction requires compatibility between KR and ACP partners. Structural modeling of wild-type and chimeric KRs demonstrated that residues within the lid region (α6–α7) establish critical ACP–KR contacts that define docking orientation and substrate positioning. Together, these results highlight the lid helices as key molecular determinants of KR–ACP compatibility, providing a structural basis for how protein–protein interactions shape specificity in type II PKSs.

3.2 Introduction

Polyketide synthases (PKSs) are a rich source of bioactive natural products, including many clinically important antibiotics and anticancer agents. Type II PKSs, in particular, are responsible for the biosynthesis of aromatic polyketides such as doxorubicin, anthracyclines, tetracyclines, and actinorhodin. A hallmark of type II PKSs is their distinct enzymology: rather than being fused into a large multidomain protein, as in type I PKSs, the individual catalytic functions are encoded on discrete polypeptides that rely on specific protein–protein interactions with the acyl carrier protein (ACP).[176] The ACP tethers and shuttles the growing polyketide intermediate to the catalytic partners that tailor its structure.[100] Despite decades of studies, the rules governing how ACPs achieve selective recognition of partner enzymes remain poorly understood.

The ketoreductases (KRs) are central tailoring enzymes in type II PKSs. These NADPH-dependent enzymes reduce specific carbonyl groups on the polyketide backbone, and in many cases, they also influence the regiospecific cyclization of the nascent chain. Among the best studied are the C9-reducing KRs of the anthracycline pathways(Figure 3.1): the ActKR from the actinorhodin PKS[123] and the DpsKR from the doxorubicin PKS[160]. While ActKR has been shown to catalyze productive C9 reduction *in vitro*, DpsKR exhibits a puzzling lack of activity in heterologous reconstitutions, often yielding only unreduced products.[177] This raises fundamental questions about what determines KR activity and substrate specificity in type II systems.

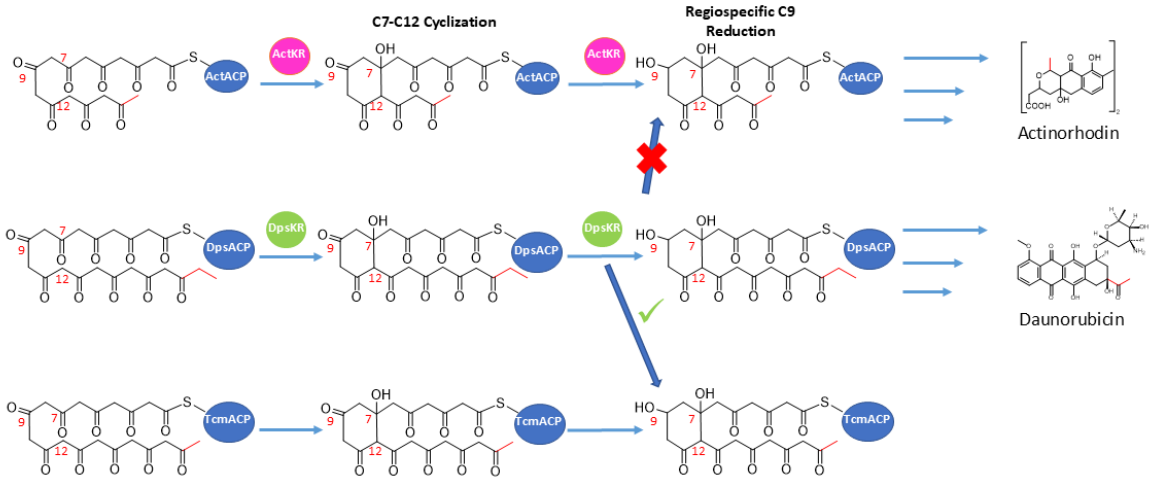


Figure 3.1. *Early stages of actinorhodin and daunorubicin biosynthesis. Both the ActKR and DpsKR function in the same role in their respective pathways, reducing the C9 carbonyl and facilitating first ring cyclization. Importantly, the DpsKR has been shown, both in vivo and in vitro, to be unable to replace the ActKR in the Act PKS, while the reverse is possible (ActKR can replace DpsKR in the Dps PKS). But DpsKR can replace TcmKR in TcmPKS.*

Previous structural studies of type II KRs have established that the overall fold and catalytic residues are conserved across families, suggesting that factors beyond active-site chemistry must account for differences in activity.[123,178] One candidate determinant is the so-called "lid region," formed by helices α6 and α7. In ActKR, this lid has been implicated in both substrate positioning and interactions with the ACP. However, whether the lid helices also dictate ACP compatibility in other type II KRs has remained unclear.

In this work, we dissect the contribution of the lid helices and ACP identity to KR activity using a combination of *in vitro* reconstitution and structural modeling. By systematically testing wild-type and chimeric KRs in minimal PKS systems, we show that lid swapping between DpsKR and ActKR reprograms reduction activity, highlighting the critical role of this structural element. Complementary ACP-swapping experiments further reveal that ACP identity strongly influences whether a KR can productively

engage and reduce its substrate. Structural predictions provide mechanistic insight, demonstrating that residues within the lid region mediate ACP docking and define the extent of productive binding interfaces.

Together, these findings establish the lid helices as molecular determinants of KR–ACP compatibility and illuminate how protein–protein interactions, rather than active-site chemistry alone, govern substrate reduction in type II PKSs. These insights help explain why combinatorial biosynthesis often fails when KRs are exchanged between pathways and underscore the importance of engineering protein–protein interfaces in future efforts to generate novel polyketide derivatives.

3.3 Results

3.3.1 Role of Lid Helices and ACP Identity in DpsKR Activity

To investigate the role of the lid helices in DpsKR activity, the Act minimal PKS was reconstituted with WT DpsKR. As shown in Figure3.2 A, only the unreduced products Sek4 and Sek4b were detected, indicating that WT DpsKR is unable to perform C9 ketoreduction in this context. In contrast, when the lid region of DpsKR was replaced with that of WT ActKR, the resulting chimeric KR enabled the production of reduced product (Figure3.2 B), demonstrating that the lid helices are critical for productive reduction.

We next asked whether the ACP partner contributes to reduction. When the Act minimal PKS was reconstituted with DpsACP and WT ActKR, the reduced product mutactin was formed (Figure3.2 B). Similarly, the Dps minimal PKS containing ActACP and WT DpsKR also produced mutactin (Figure3.2 C). These results indicate that ACP identity plays a key role in enabling KR activity. However, when the Act minimal PKS was assembled with DpsACP and WT DpsKR, only Sek4 and Sek4b were observed (Figure3.2 B). Together, these data show that both the lid helices and ACP partner identity contribute to KR-dependent reduction, underscoring the importance of protein–protein interactions in type II PKS biosynthesis.

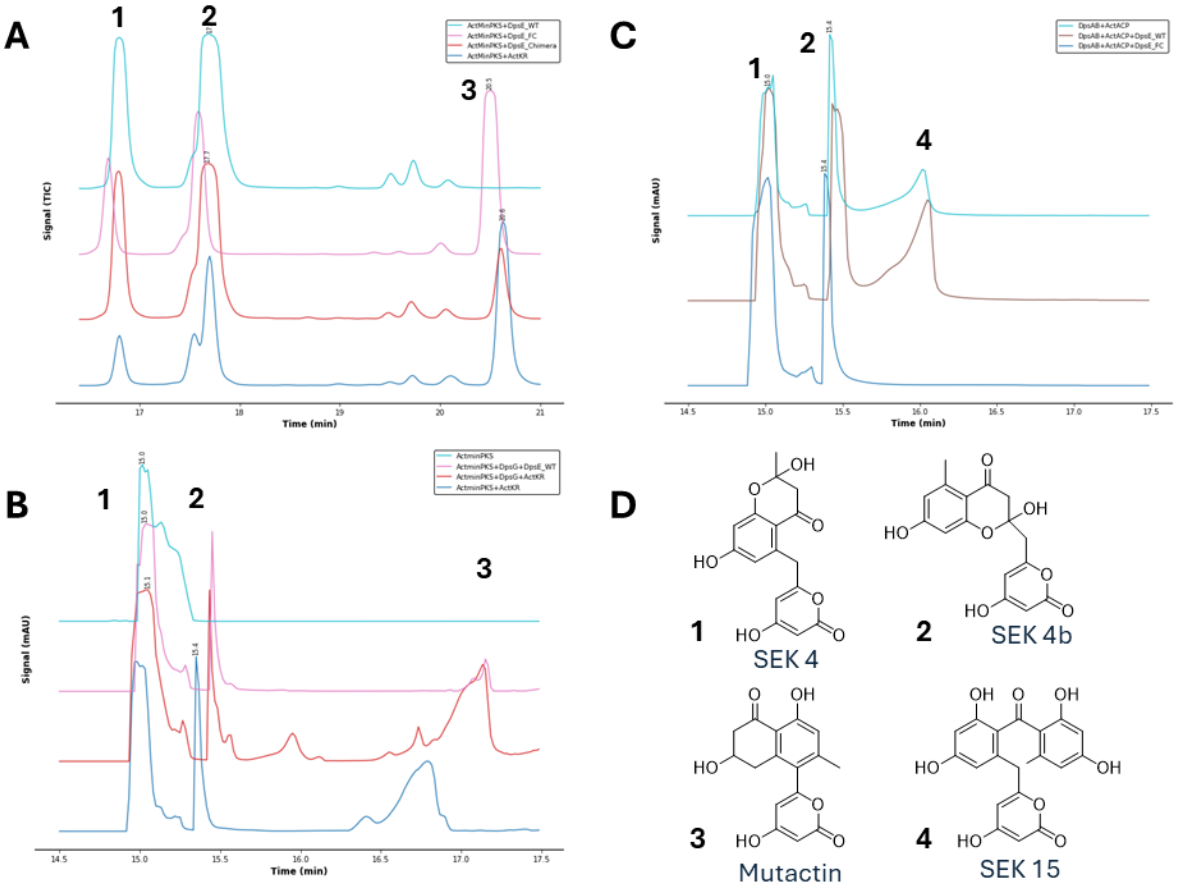


Figure 3.2. *HPLC product profile of the Act minimal PKS (ActminPKS) and Dps minimal PKS (Dps minPKS) when reconstituted with the ActKR, DpsKR, DpsActKR Lid Chimera and DpsActKR Full Pocket Chimera. A)ActminPKS with ActACP and ActKR, DpsKR, DpsActKR Lid Chimera and DpsActKR Full Pocket Chimera as legend shows. B)ActminPKS with DpsACP and ActKR, DpsKR, DpsActKR Lid Chimera and DpsActKR Full Pocket Chimera as legend shows. C) DpsminPKS with ActACP and ActKR, DpsKR, DpsActKR Lid Chimera and DpsActKR Full Pocket Chimera as legend shows. D) The products peaks are: 1) sek4, 2) sek4b, 3) mutactin and 4) SEK15. Product peaks were identified by typical published retention times.*

3.3.2 Structural Predictions Reveal Lid Region Residues Mediate ACP-KR Interactions

To further investigate the molecular basis of substrate specificity, structural predictions of the chimeric KRs were generated. The models suggest that key residues

within the lid region (residues 194–221) play a central role in ACP binding and substrate positioning.

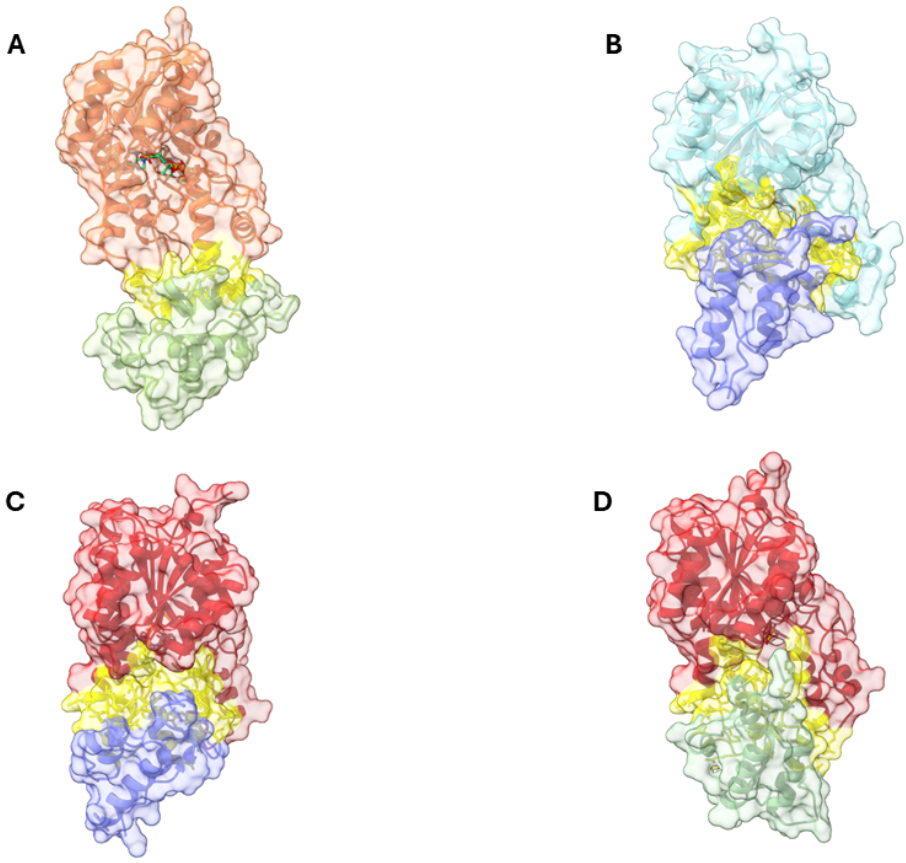


Figure 3.3. *Predicted ACP-KR structures with Alphafold3 A) DpsACP and DpsKR WT, B)ActACP and ActKR WT C)DpsACP and DpsKR Chimera and D)ActACP and DpsKR Chimera*

The predicted positions of DpsACP and ActACP relative to their partner KRs mirrored the *in vitro* reconstitution results. Specifically, DpsACP aligned similarly to ActACP when paired with the full DpsKR–ActKR lid chimera (Figure 3.3A–B). By contrast, the orientation of DpsACP was markedly different when bound to WT DpsKR (Figure 3.3C), consistent with the inability of WT DpsKR to reduce substrates in the Act minimal PKS system.

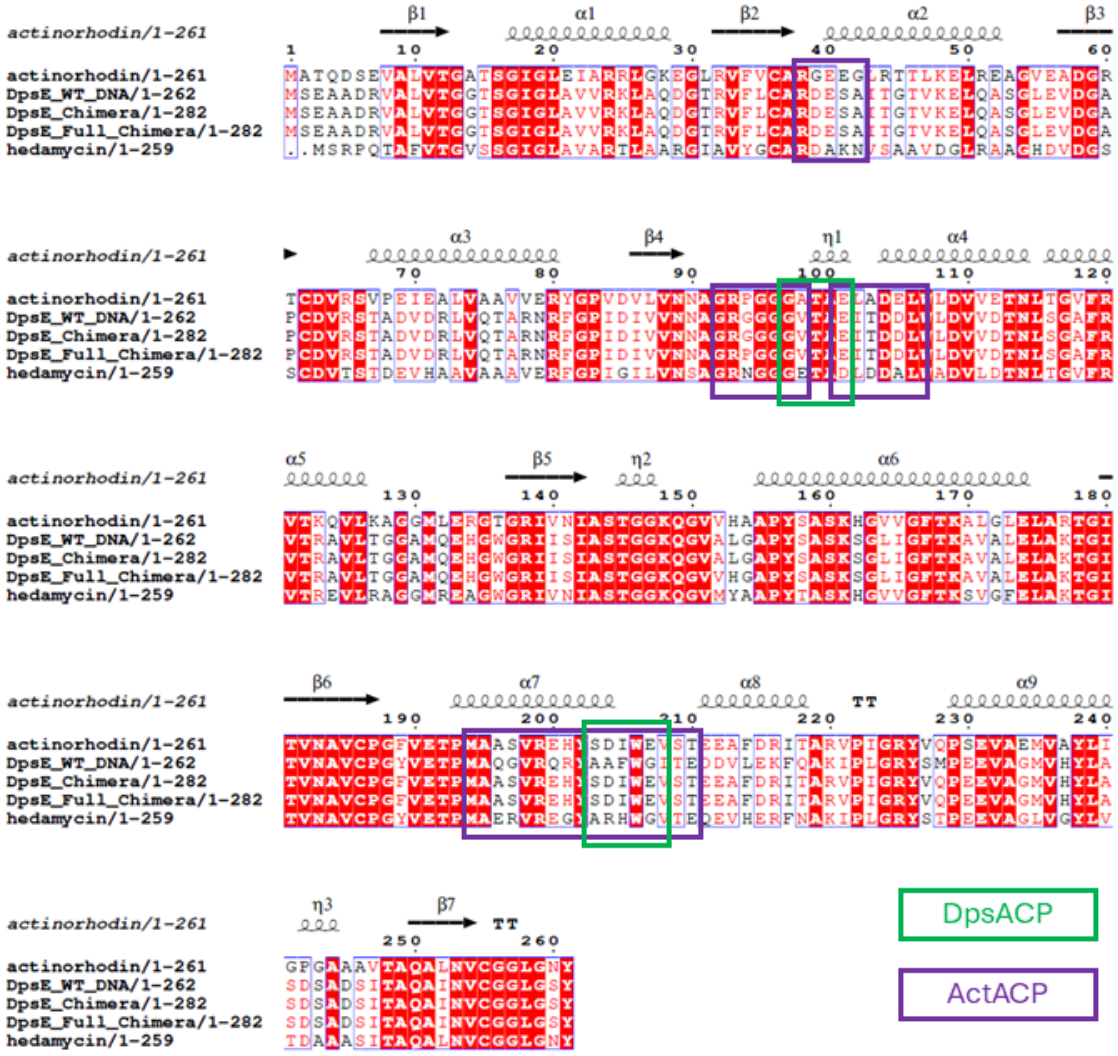


Figure 3.4. *Sequence alignment of ActKR, DpsKR WT, DpsKR Chimera, DpsKR Full Chimera and Hedamycin KR with the highlight of ACP binding sites.*

Residue mapping further highlighted differences in ACP–KR contacts. DpsACP was predicted to bind WT DpsKR primarily through residues V98, E101, A203, A204, F205, W206, G207, and I208. In contrast, DpsACP bound the full DpsKR–ActKR chimera at a broader interface, including R38, D39, R93, P94, G95–G97, V98, E101, I102, T103,

D104–D105, L106, H153, P193, S197, V198, H201, D204, I205, and E207 (Figure 3.4). Notably, this binding pattern strongly resembled the interaction of ActACP with WT ActKR, which involved R38, G39, E40, G42, R65, G92, R93, P94, G95–G97, E101, L102, A103, D104, E105, L106, P193, M194, S197, V198, E200, H201, and Y202 (Figure 3.4).

The largest differences between WT DpsKR and the chimera–ActKR systems localized to the lid region (residues 194–221). These structural distinctions correlate with the functional results of the *in vitro* reconstitution assays, demonstrating that ACP binding is modulated by the lid helices and, in turn, alters KR substrate specificity.

3.4 Discussion:

The results presented here demonstrate that the activity and substrate specificity of type II KRs are determined not only by active-site chemistry but also by structural features within the lid helices that mediate productive ACP binding. In vitro reconstitution assays showed that WT DpsKR cannot catalyze C9 reduction in the Act minimal PKS, yielding only unreduced products. Strikingly, replacing the DpsKR lid with that of ActKR restored reduction, underscoring the critical role of the lid region in enabling catalysis. This functional dependence on lid identity was reinforced by ACP-swapping experiments, which revealed that ACP identity strongly influences whether reduced product is formed.

Structural modeling provided a mechanistic rationale for these observations. The predicted structures indicated that the lid region (residues 194–221) defines ACP docking orientation and the interaction network. DpsACP-bound WT DpsKR through a narrow set of residues, whereas the DpsKR–ActKR lid chimera engaged DpsACP across an expanded interface that closely resembled the ActACP–ActKR interaction. These differences were localized to the lid helices and correlated precisely with the outcomes of the reconstitution assays. Together, these data suggest that the lid helices govern KR–ACP compatibility and, by shaping ACP docking, dictate the likelihood of substrate reduction.

Our docking results are consistent with the integrative models of ActKR reported by Serapian et al. [179], who combined BUDE docking, MD simulations, and NMR titrations to identify productive ACP binding modes. In their study, the most plausible models (notably from the M14 family) positioned the actACP α3 helix across the KR surface,

engaging an arginine patch (R38/R65/R93) that directs PPant entry into the KR channel. Similarly, we find that the DpsKR–ActKR lid chimera establishes an expanded interface overlapping this ActKR contact network, incorporating residues such as R38, R93, and the G95–G97 motif. By contrast, WT DpsKR presents only a restricted, non-productive surface lacking PPant-patch engagement. Thus, while ActKR achieves catalysis through arginine patch interactions and docking plasticity, our results show that lid swapping reconfigures DpsKR to recapitulate these critical contacts. This demonstrates that lid-mediated interface compatibility is a key determinant of productive ACP docking and substrate reduction.

These findings align with and extend previous work on type II PKSs. Crystal structures of ActKR and related KRs have revealed overall conservation of fold and active-site residues, but the contribution of ACP binding to regiospecificity has remained unclear.[123,180,181] Our results suggest that lid helices act as molecular determinants of ACP recognition, providing a structural explanation for why KR swapping in combinatorial biosynthesis often fails: mismatched lid–ACP interfaces can block productive reduction even when catalytic residues are conserved.[182]

More broadly, this study underscores the importance of protein–protein interactions in type II PKSs. Unlike type I systems, where domains are covalently linked, type II PKSs rely on transient yet highly specific contacts between discrete enzymes and ACPs. Our findings indicate that KR specificity arises from the interplay between lid helices, conserved docking motifs such as the arginine patch, and the ACP surface. Future structural studies, including co-crystallization or cryo-EM of ACP–KR complexes, will be

critical to validate these predicted binding modes. Ultimately, understanding and engineering lid–ACP compatibility may enable the rational design of type II PKS pathways to generate novel anthracycline derivatives with improved therapeutic potential.

3.5 Conclusion

Together, our biochemical reconstitution assays and structural predictions demonstrate that substrate reduction by type II ketoreductases is governed not only by active-site chemistry but by precise protein–protein interactions mediated through the lid helices and ACP surface. WT DpsKR was incapable of reducing substrates in the *Act* minimal PKS, yet introduction of the ActKR lid region restored full activity, revealing that the lid helices dictate KR–ACP compatibility. ACP swapping further showed that ACP identity is equally critical, as productive reduction occurred only when the ACP docking geometry matched the preferred lid configuration. Structural modeling supported these observations by showing that the chimeric KR establishes a binding interface with DpsACP that closely resembles the natural ActACP–ActKR complex, while WT DpsKR aligns ActACP in a nonproductive orientation. These findings establish the lid region as a key molecular determinant that reshapes ACP binding surfaces, thereby controlling substrate positioning and reaction outcome. More broadly, this work highlights that specificity in type II PKSs emerges from cooperative structural features rather than isolated catalytic motifs, offering a mechanistic basis for the historical difficulty of engineering functional KR swaps. Understanding and manipulating lid–ACP compatibility will be essential for future efforts to reprogram type II PKS pathways for anthracycline diversification and natural product engineering.

3.6 Method

3.6.1 Expression and purification of DpsKR

The recombinant N-terminal histidine tagged Act/DpsKR chimeric proteins were expressed, grown, and harvested in BL21(DE3) E. coli cells. The E. coli cells containing the KR plasmids were grown to OD600 0.6 at 37°C in LB medium, 50 μg/mL of kanamycin, and induced at 18°C with 1mM IPTG.

The cell culture was first incubated for 16 hours at 18°C. Then it was harvested with centrifugation at 5,525 r.c.f. for 15 minutes. The cell pellets were resuspended in the lysis buffer of 50 mM Tris-HCl, pH 7.5, 10% glycerol, 10 mM imidazole, and 300 mM NaCl. Resuspended cells were sonicated, and cell debris were separated by centrifugation at 21,036 r.c.f. for 1 hour. The collected supernatant was combined with HisPurTM Nickel Resin (Thermo Scientific) following the manufacturer's instructions to purify the desired his-tagged Act/DpsKR chimeric protein with imidazole step-gradient. The purified proteins were collected and verified through SDS-PAGE. The purified protein was concentrated to 7 mg/mL and stored in the dialysis buffer of 50 mM Tris-HCl, pH 7.5, 10% glycerol, 100 mM NaCl, and water.

3.6.2 Crystallization and data collection

Utilizing Hampton Research Crystallization kits, the Act/DpsKR chimeric purified protein were set up in 24-well VDX plates and Cryschem plates. The purified proteins were plated and screened with Crystal Screens from Hampton Research and 3 mM NADPH at both room temperature and 4 °C conditions. Diffraction data for quality

protein crystals were optimized and seeded for better growth. The optimized protein crystals were washed and flash frozen in liquid nitrogen, with diffraction data collected at UC Berkeley Advanced Light Source using monochromatic X-rays (1.0000 Å)

3.6.3 Actinorhodin and daunorubicin KSCLF expression and purification

ActKS/CLF was expressed in Streptomyces coelicolor CH999 containing the pRJC006 expression vector. Fresh spores were prepared by spreading 100 ul spore stock onto a mannitol soy agar plate containing 50 ug/ml kanamycin. This plate was incubated at 30 C for 5-7days. The fresh spores were then collected and used to inoculate 50 mL of Super YEME containing 50 μg/mL kanamycin. The 50 mL cultures were grown for 3 days at 30 °C with shaking at 250 rpm. The 50 mL cultures were transferred to 2.5 L baffled flasks containing 500 ml of Super YEME and 50 μg/mL kanamycin. The cultures were grown for an additional 2 days under the same conditions. Protein expression was induced with 10 μg/mL thiostrepton. Cells were harvested 24 hours after induction by centrifugation (20 minutes at 5100 rpm) and the cell pellets were stored at -80 °C. Cell pellets from 2 L of cell culture were resuspended in 200 mL of lysis buffer (100 mM KPi, 15 % glycerol, 300 mM NaCl, pH 7.5, with 2 EDTA-free protease inhibitor cocktail (Roche) tablets). The cell suspension was lysed by sonication on ice (10 x 1-minute cycles). Cell debris was pelleted by centrifugation (1 hour at 14000 rpm) and DNA was precipitated using streptomycin sulfate (2 % final concentration) followed by centrifugation (45 min at 14000 rpm). The supernatant was filtered using a 0.45 μM filter and 30 % – 50 % (NH4)2SO4 was used to precipitate ActKS/CLF overnight. Precipitated protein was

pelleted by centrifugation (30 min at 14000 rpm) and redissolved in Ni-binding buffer (50 mM KPi, 10 % glycerol, 500 mM NaCl, 5 mM imidazole, pH 7.5). After resuspension, the solution was filtered using a 0.45 μM filter then bound to 2 ml of Ni-IMAC resin (Bio-Rad) pre-equilibrated with Ni-binding buffer and stirred at 4 °C for 1 hour. ActKS/CLF was eluted using Ni-binding buffer containing increasing amounts of imidazole. Fractions (2.5 mL each) containing between 375 mM and 750 mM imidazole were individually buffer exchanged into storage buffer(100 mM KPi, 10 % glycerol, pH 7.5) using a PD-10 column (GE Healthcare) and flash frozen at -80 °C.

3.6.4 *In vitro* reconstitution assays

DpsKR and DpsKR chimera activity were measured relative to the ActKR using an in vitro assay based on the actinorhodin minimal PKS. Minimal PKS activity is reconstituted in vitro by combining the KS/CLF (7.5 μM) (responsible for chain extension), ActACP/DpsG (40 μM) (responsible for shuttling of the intermediates), and MCAT (5 μM) (responsible for extension unit selection). MatB (20 μM) is added in order to generate the activated extension unit, malonyl-CoA, from malonate and CoA. The reaction volume was 250 μL and contained 30 μM KR, 2.6 mM NADPH (Sigma) (when KR is present), 100 mM malonate, 7.5 mM CoA, 10 mM ATP, and 5 mM MgCl2 in 100 mM Tris, pH 7.5. The reactions were carried out at room temperature for 24 hours. The reaction mixtures were extracted three times with 300 μL of 94: 5: 1 of ethyl acetate: methanol : acetic acid and the organic layer was dried using Speed Vac Concentrator. The residual oil was resuspended in DMSO (Sigma) and subject to reverse-phase HPLC analysis using a

Synergi Hydro-RP column (Phenomenex). The HPLC gradient was from 5 to 50 % MeCN in a H2O/0.1 % formic acid mixture over 15 minutes followed by 50 to 95 % MeCN in a H2O/0.1 % formic acid mixture over 5 min. SEK4, SEK4B, and mutactin were identified using relative retention times.

3.7 References

(1) Cragg, G. M.; Newman, D. J. Biodiversity: A Continuing Source of Novel Drug Leads. *Pure Appl. Chem.* **2005**, *77*, 7–24.

(2) Buenz, E. J.; Verpoorte, R.; Bauer, B. A. The Ethnopharmacologic Contribution to Bioprospecting Natural Products. *Annu. Rev. Pharmacol. Toxicol.* **2018**, *58*, 509–530.

(3) Gaynes, R. The Discovery of Penicillin—New Insights after More than 75 Years of Clinical Use. *Emerg Infect Dis* **2017**, *23*, 849–853.

(4) Beutler, J. A. Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol.* **2019**, *86, e67*.

(5) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176.

(6) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

(7) Hopwood, D. A. Production of 'hybrid' Antibiotics by Genetic Engineering. *Nature* **1985**, *314*, 642–644.

(8) Malpartida, F.; Hopwood, D. A. Molecular Cloning of the Whole Biosynthetic Pathway of a Streptomyces Antibiotic and Its Expression in a Heterologous Host. *Nature* **1984**, *309* (5967), 462–464. https://doi.org/10.1038/309462a0.

(9) Bibb, M. J.; Sherman, D. H.; Ōmura, S.; Hopwood, D. A. Cloning, Sequencing and Deduced Functions of a Cluster of Streptomyces Genes Probably Encoding Biosynthesis of the Polyketide Antibiotic Frenolicin. *Gene* **1994**, *142*, 31–39.

(10) Han, L.; Yang, K.; Ramalingam, E.; Mosher, R. H.; Vining, L. C. Cloning and characterization of polyketide synthase genes for jadomycin b biosynthesis in streptomyces venezuelaeISP5230. *Microbiology* **1994**, *140*, 3379–3389.

(11) Motamedi, H.; Hutchinson, C. R. Cloning and Heterologous Expression of a Gene Cluster for the Biosynthesis of Tetracenomycin C, the Anthracycline Antitumor Antibiotic of Streptomyces Glaucescens. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4445–4449.

(12) Kim, E.-S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. Sequences of the Oxytetracycline Polyketide Synthase-Encoding Otc Genes from Streptomyces Rimosus. *Gene* **1994**, *141* (1), 141–142. https://doi.org/10.1016/0378-1119(94)90144-9.

(13) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. Characterization of the streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **1994**, *151*, 1–10.

(14) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Manipulation and Analysis of an Aromatic Polyketide Synthase with Unproved Catalytic Specificities. *J. Am. Chem. Soc.* **1993**, *115*, 11671–11675.

(15) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. The tcmVI Region of the Tetracenomycin c Biosynthetic Gene Cluster of Streptomyces

Glaucescens Encodes the Tetracenomycin F1 Monooxygenase, Tetracenomycin F2 Cyclase, and, Most Likely, a Second Cyclase. *J. Bacteriol.* **1993**, *175*, 7571–7580.
(16) Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Stereochemical Course of Two Reactions Catalyzed by a Polyketide Synthase. *Biochemistry* **1994**, *33*, 9321–9326.
(17) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91* (24), 11542–11546. https://doi.org/10.1073/pnas.91.24.11542.
(18) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91*, 11542–11546.
(19) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Rational Design of Aromatic Polyketide Natural Products by Recombinant Assembly of Enzymatic Subunits. *Nature* **1995**, *375* (6532), 549–554. https://doi.org/10.1038/375549a0.
(20) Shen, B.; Summers, R. G.; Wendt-Pienkowski, E.; Hutchinson, C. R. The Streptomyces Glaucescens tcmKL Polyketide Synthase and tcmN Polyketide Cyclase Genes Govern the Size and Shape of Aromatic Polyketides. *J Am Chem Soc* **1995**, *117*, 6811–6821.
(21) Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. Engineered Biosynthesis of Novel Polyketides: Properties of the whiE Aromatase/Cyclase. *Nat Biotechnol* **1996**, *14*, 335–338.
(22) Kramer, P. J. Rational Design and Engineered Biosynthesis of a Novel 18-Carbon Aromatic Polyketide. *J Am Chem Soc* **1997**, *119*, 635–639.
(23) Meurer, G. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4*, 433–443.
(24) McDaniel, R. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel 'Unnatural' Natural Products. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 1846–1851.
(25) Harvey, A. L. Natural Products in Drug Discovery. *Drug Discov. Today* **2008**, *13*, 894–901.
(26) Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A. Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale for Catalytic Reprogramming. *J Am Chem Soc* **2014**, *136*, 7348–7362.
(27) Connor, C. J. O.; Beckmann, H. S. G.; Spring, D. R. Diversity-Oriented Synthesis: Producing Chemical Tools for Dissecting Biology. *Chem Soc Rev* **2012**, *41*, 4444–4456.
(28) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for Bad Bugs: Confronting the Challenges of Antibacterial Discovery. *Nat Rev Drug Discov* **2007**, *6*, 29–40.

(29) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat Rev Drug Discov* **2015**, *14*, 111–129.
(30) Schloss, P. D.; Handelsman, J. Metagenomics for Studying Unculturable Microorganisms: Cutting the Gordian Knot. *Genome Biol.* **2005**, *6*, 229.
(31) Medema, H.; M. The Year 2020 in Natural Product Bioinformatics: An Overview of the Latest Tools and Databases. *Nat. Prod. Rep.* **2021**, *38*, 301–306.
(32) Blin, K. antiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation. *Nucleic Acids Res.* **2023**, *51*, 46–50.
(33) Blin, K.; Shaw, S.; Kautsar, S. A.; Medema, M. H.; Weber, T. The antiSMASH Database Version 3: Increased Taxonomic Coverage and New Query Features for Modular Enzymes. *Nucleic Acids Res.* **2021**, *49*, 639–643.
(34) Hannigan, G. D. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res* **2019**, *47*, 110–110.
(35) Ichikawa, N. DoBISCUIT: A Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2013**, *41*, 408–414.
(36) Kautsar, S. A. MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* **2020**, *48*, 454–458.
(37) Palaniappan, K. IMG-ABC v.5.0: An Update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res.* **2020**, *48*, 422–430.
(38) Bentley, S. D. Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2. *Nature* **2002**, *417*, 141–147.
(39) Nett, M.; Ikeda, H.; Moore, B. S. Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. *Nat Prod Rep* **2009**, *26*, 1362–1384.
(40) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating Natural Product Combinatorial Biosynthesis with Synthetic Biology. *Nat Chem Biol* **2015**, *11*, 649–659.
(41) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8), 509–523. https://doi.org/10.1038/nrmicro3496.
(42) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat Rev Drug Discov* **2021**, *20*, 200–216.
(43) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem* **2021**, *90*, 763–788.
(44) Xu, F. A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat Chem Biol* **2019**, *15*, 161–168.
(45) Herbst, D. A.; Townsend, C. A.; Maier, T. The Architectures of Iterative Type i PKS and FAS. *Nat Prod Rep* **2018**, *35*, 1046–1069.
(46) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
(47) Dutta, S. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510*, 512–517.
(48) Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. *Science* **2007**, *316*, 288–290.

(49) Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315–1322.
(50) Staunton, J.; J. Weissman, K. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416. https://doi.org/10.1039/A909079G.
(51) Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Nat Prod Rep* **2003**, *20*, 79–110.
(52) Ferrer, J.-L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Structure of Chalcone Synthase and the Molecular Basis of Plant Polyketide Biosynthesis. *Nat Struct Mol Biol* **1999**, *6*, 775–784.
(53) Bisang, C. A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* **1999**, *401*, 502–505.
(54) Jackowski, S.; Rock, C. O. Acetoacetyl-Acyl Carrier Protein Synthase, a Potential Regulator of Fatty Acid Biosynthesis in Bacteria. *J. Biol. Chem.* **1987**, *262*, 7927–7931.
(55) Moore, B. S.; Hertweck, C. Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units. *Nat Prod Rep* **2002**, *19*, 70–99.
(56) Crawford, J. M.; Dancy, B. C. R.; Hill, E. A.; Udwary, D. W.; Townsend, C. A. Identification of a Starter Unit Acyl-Carrier Protein Transacylase Domain in an Iterative Type I Polyketide Synthase. *Proc. Natl. Acad. Sci.* **2006**, *103* (45), 16728–16733. https://doi.org/10.1073/pnas.0604112103.
(57) Tsai, S.-C. (Sheryl). The Structural Enzymology of Iterative Aromatic Polyketide Synthases: A Critical Comparison with Fatty Acid Synthases. *Annu. Rev. Biochem.* **2018**, *87* (1), 503–531. https://doi.org/10.1146/annurev-biochem-063011-164509.
(58) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat Prod Rep* **2016**, *33*, 150–161.
(59) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **2009**, *26*, 90–114.
(60) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* **1998**, *37* (22), 8132–8138. https://doi.org/10.1021/bi980466i.
(61) Meadows, E. S.; Khosla, C. In Vitro Reconstitution and Analysis of the Chain Initiating Enzymes of the R1128 Polyketide Synthase. *Biochemistry* **2001**, *40*, 14855–14861.
(62) Winter, J. M. Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, *137*, 9885–9893.
(63) Cacho, R. A. Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J Am Chem Soc* **2015**, *137*, 15688–15691.
(64) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide b Synthase. *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
(65) Crawford, J. M. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320*, 243–246.

(66) Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709. https://doi.org/10.1021/ja0378759.
(67) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Dissection of the Catalytic Specificity of the Act Ketoreductase. *J. Am. Chem. Soc.* **1994**, *116*, 4166–4170.
(68) Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Evidence for Temporal, but Not Regiospecific, Control of Cyclization of an Aromatic Polyketide Precursor. *Chem. Biol.* **1994**, *1*, 205–210.
(69) Shen, Y. Ectopic Expression of the Minimal whiE Polyketide Synthase Generates a Library of Aromatic Polyketides of Diverse Sizes and Shapes. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 3622–3627.
(70) Auclair, K. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels–alder Reaction of a Substrate Analogue. *J Am Chem Soc* **2000**, *122*, 11519–11520.
(71) Kim, H. J.; Ruszczycky, M. W.; Choi, S.; Liu, Y.; Liu, H. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn a. *Nature* **2011**, *473*, 109–112.
(72) Kato, N. A New Enzyme Involved in the Control of the Stereochemistry in the Decalin Formation during Equisetin Biosynthesis. *Biochem. Biophys. Res. Commun.* **2015**, *460*, 210–215.
(73) Kasahara, K. Solanapyrone Synthase, a Possible Diels–Alderase and Iterative Type i Polyketide Synthase Encoded in a Biosynthetic Gene Cluster from Alternaria Solani. *ChemBioChem* **2010**, *11*, 1245–1252.
(74) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of Claisen Cyclase Domain in Fungal Polyketide Synthase WA, a Naphthopyrone Synthase of Aspergillus Nidulans. *Chem. Biol.* **2001**, *8*, 189–197.
(75) Crawford, J. M. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461*, 1139–1143.
(76) Ames, B. D. Crystal Structure and Functional Analysis of Tetracenomycin ARO/CYC: Implications for Cyclization Specificity of Aromatic Polyketides. *PNAS* **2008**, *105*, 5349–5354.
(77) Ames, B. D. Structural and Biochemical Characterization of ZhuI Aromatase/Cyclase from the R1128 Polyketide Pathway. *Biochemistry* **2011**, *50*, 8392–8406.
(78) Yu, T.-W. Engineered biosynthesis of novel polyketides from streptomyces spore pigment polyketide synthases. *J Am Chem Soc* **1998**, *120*, 7749–7759.
(79) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively. *J. Am. Chem. Soc.* **1994**, *116*, 10855–10859.
(80) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of Post-PKS Tailoring Steps through Combinatorial Biosynthesis. *Nat Prod Rep* **2002**, *19*, 542–580.

(81) Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. Cytochromes P450 for Natural Product Biosynthesis in Streptomyces: Sequence, Structure, and Function. *Nat Prod Rep* **2017**, *34*, 1141–1172.
(82) Chung, W.; Vanderwal, C. D. Stereoselective Halogenation in Natural Product Synthesis. *Angew. Chem. Int. Ed.* **2016**, *55*, 4396–4434.
(83) Singh, S., Jr; P., G. N.; Thorson, J. S. The Structural Biology of Enzymes Involved in Natural Product Glycosylation. *Nat Prod Rep* **2012**, *29*, 1201–1237.
(84) Podust, L. M.; Sherman, D. H. Diversity of P450 Enzymes in the Biosynthesis of Natural Products. *Nat Prod Rep* **2012**, *29*, 1251–1266.
(85) Carreras, C. W.; Pieper, R.; Khosla, C. Efficient Synthesis of Aromatic Polyketides in Vitro by the Actinorhodin Polyketide Synthase. *J Am Chem Soc* **1996**, *118*, 5158–5159.
(86) Shen, B.; Hutchinson, C. R. Deciphering the Mechanism for the Assembly of Aromatic Polyketides by a Bacterial Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1996**, *93*, 6600–6604.
(87) Harris, T. M.; Harris, C. M. Biomimetic Syntheses of Aromatic Polyketide Metabolites. *Pure Appl Chem* **1986**, *58*, 283–294.
(88) Sherman, D. h. Structure and Deduced Function of the Granaticin-Producing Polyketide Synthase Gene Cluster of Streptomyces Violaceoruber Tü22. *EMBO J.* **1989**, *8*, 2717–2725.
(89) Kim, E. S.; Cramer, K. D.; Shreve, A. L.; Sherman, D. H. Heterologous Expression of an Engineered Biosynthetic Pathway: Functional Dissection of Type II Polyketide Synthase Components in Streptomyces Species. *J Bacteriol* **1995**, *177*, 1202–1207.
(90) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11* (9), 888–893. https://doi.org/10.1038/nsmb808.
(91) Sherman, D. H.; Kim, E. S.; Bibb, M. J.; Hopwood, D. A. Functional Replacement of Genes for Individual Polyketide Synthase Components in Streptomyces Coelicolor A3(2) by Heterologous Genes from a Different Polyketide Pathway. *J. Bacteriol.* **1992**, *174*, 6184–6190.
(92) Kim, E. S.; Hopwood, D. A.; Sherman, D. H. Analysis of Type II Polyketide Beta-Ketoacyl Synthase Specificity in Streptomyces Coelicolor A3(2) by Trans Complementation of Actinorhodin Synthase Mutants. *J Bacteriol* **1994**, *176*, 1801–1804.
(93) Dreier, J.; Khosla, C. Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the β-Ketoacyl Synthase–chain Length Factor Heterodimer. *Biochemistry* **2000**, *39*, 2088–2095.
(94) Grammbitter, G. L. C. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. *J Am Chem Soc* **2019**, *141*, 16615–16623.
(95) Bräuer, A. Structural Snapshots of the Minimal PKS System Responsible for Octaketide Biosynthesis. *Nat Chem* **2020**, *12*, 755–763.
(96) Nicholson, T. P. First in Vitro Directed Biosynthesis of New Compounds by a Minimal Type II Polyketide Synthase: Evidence for the Mechanism of Chain Length Determination. *Chem Commun* **2003**, *686–687*. https://doi.org/10.1039/B300847A.

(97) Kim, Y.; Prestegard, J. H. Refinement of the NMR Structures for Acyl Carrier Protein with Scalar Coupling Data. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 377–385.
(98) Evans, S. E. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* **2008**, *9*, 2424–2432.
(99) Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein. *Biochemistry* **2003**, *42*, 4648–4657.
(100) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. Solution Structure of the Actinorhodin Polyketide Synthase Acyl Carrier Protein from Streptomyces Coelicolor A3(2),. *Biochemistry* **1997**, *36* (20), 6000–6008. https://doi.org/10.1021/bi970006+.
(101) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. Targeted Gene Replacements in a Streptomyces Polyketide Synthase Gene Cluster: Role for the Acyl Carrier Protein. *Mol. Microbiol.* **1992**, *6*, 3237–3249.
(102) Zhang, Y.-M.; Wu, B.; Zheng, J.; Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction*. *J. Biol. Chem.* **2003**, *278*, 52935–52943.
(103) Arthur, C. J. Structure and Malonyl CoA-ACP Transacylase Binding of Streptomyces Coelicolor Fatty Acid Synthase Acyl Carrier Protein. *ACS Chem Biol* **2009**, *4*, 625–636.
(104) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-Based Analysis of the Molecular Interactions between Acyltransferase and Acyl Carrier Protein in Vicenistatin Biosynthesis. *PNAS* **2016**, *113*, 1802–1807.
(105) Bartholow, T. G. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4*, 1–10.
(106) Cho, Y. I. Engineered Chimeras Unveil Swappable Modular Features of Fatty Acid and Polyketide Synthase Acyl Carrier Proteins. *Biochemistry* **2022**, *61*, 217–227.
(107) Khosla, C. Genetic Construction and Functional Analysis of Hybrid Polyketide Synthases Containing Heterologous Acyl Carrier Proteins. *J. Bacteriol.* **1993**, *175*, 2197–2204.
(108) Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C. Ketosynthases in the Initiation and Elongation Modules of Aromatic Polyketide Synthases Have Orthogonal Acyl Carrier Protein Specificity†, 2003. https://doi.org/10.1021/BI0341962.
(109) Das, A.; Khosla, C. In Vivo and in Vitro Analysis of the Hedamycin Polyketide Synthase. *Chem. Biol.* **2009**, *16*, 1197–1207.
(110) Koppisch, A. T.; Khosla, C. Structure-Based Mutagenesis of the Malonyl-CoA:Acyl Carrier Protein Transacylase from Streptomyces Coelicolor. *Biochemistry* **2003**, *42*, 11057–11064.
(111) Beltran-Alvarez, P.; Cox, R. J.; Crosby, J.; Simpson, T. J. Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* **2007**, *46*, 14672–14681.

(112) Arthur, C. J. Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* **2005**, *44*, 15414–15421.
(113) Kalaitzis, J. A. Policing Starter Unit Selection of the Enterocin Type II Polyketide Synthase by the Type II Thioesterase EncL. *Bioorg. Med. Chem.* **2011**, *19*, 6633–6638.
(114) Shepherd, M. D.; Kharel, M. K.; Zhu, L. L.; Lanen, S. G. van; Rohr, J. Delineating the Earliest Steps of Gilvocarcin Biosynthesis: Role of GilP and GilQ in Starter Unit Specificity. *Org Biomol Chem* **2010**, *8*, 3851–3856.
(115) Tang, Y.; Koppisch, A. T.; Khosla, C. The Acyltransferase Homologue from the Initiation Module of the R1128 Polyketide Synthase Is an Acyl-ACP Thioesterase That Edits Acetyl Primer Units. *Biochemistry* **2004**, *43* (29), 9546–9555. https://doi.org/10.1021/bi049157k.
(116) Keatinge-Clay, A. T. Catalysis, Specificity, and ACP Docking Site of Streptomyces Coelicolor Malonyl-CoA:ACP Transacylase. *Structure* **2003**, *11*, 147–154.
(117) Szu, P.-H. Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **2011**, *18*, 1021–1031.
(118) Bartel, P. L. Biosynthesis of Anthraquinones by Interspecies Cloning of Actinorhodin Biosynthesis Genes in Streptomycetes: Clarification of Actinorhodin Gene Functions. *J. Bacteriol.* **1990**, *172*, 4816–4826.
(119) Rajgarhia, V. B.; Strohl, W. R. Minimal streptomyces sp. Strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis. *J Bacteriol* **1997**, *179*, 2690–2696.
(120) Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C. Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKR): Molecular Basis of Stereo- and Regiospecificities. *Biochemistry* **2011**, *50*, 7426–7439.
(121) Piel, J. Cloning, Sequencing and Analysis of the Enterocin Biosynthesis Gene Cluster from the Marine Isolate 'Streptomyces Maritimus': Evidence for the Derailment of an Aromatic Polyketide Synthase. *Chem. Biol.* **2000**, *7*, 943–955.
(122) Price, A. C.; Zhang, Y. M.; Rock, C. O.; White, S. W. Structure of β-Ketoacyl-[Acyl Carrier Protein] Reductase from Escherichia Coli: Negative Cooperativity and Its Structural Basis. *Biochemistry* **2001**, *40*, 12772–12781.
(123) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43* (46), 14529–14538. https://doi.org/10.1021/bi048133a.
(124) Oppermann, U. Short-Chain Dehydrogenases/Reductases (SDR): The 2002 Update. *Chem. Biol. Interact.* **2003**, *143–144*, 247–253.
(125) Filling, C. Critical Residues for Structure and Catalysis in Short-Chain Dehydrogenases/Reductases*. *J. Biol. Chem.* **2002**, *277*, 25677–25684.
(126) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **1993**, *262* (5139), 1546–1550.
(127) Kalaitzis, J. A.; Moore, B. S. Heterologous Biosynthesis of Truncated Hexaketides Derived from the Actinorhodin Polyketide Synthase. *J. Nat. Prod.* **2004**, *67*, 1419–1422.

(128) Zawada, R. J.; Khosla, C. Heterologous Expression, Purification, Reconstitution and Kinetic Analysis of an Extended Type II Polyketide Synthase. *Chem. Biol.* **1999**, *6* (9), 607–615. https://doi.org/10.1016/s1074-5521(99)80112-4.
(129) Hadfield, A. T. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12*, 1865–1875.
(130) Kantola, J. Folding of the Polyketide Chain Is Not Dictated by Minimal Polyketide Synthase in the Biosynthesis of Mithramycin and Anthracycline. *Chem. Biol.* **1997**, *4*, 751–755.
(131) Kunnari, T. Incorrectly Folded Aromatic Polyketides from Polyketide Reductase Deficient Mutants. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2639–2642.
(132) Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C. Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase,. *Biochemistry* **2008**, *47* (7), 1837–1847. https://doi.org/10.1021/bi7016427.
(133) Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C. C. Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochemistry* **2011**, *50*, 4638–4649.
(134) Javidpour, P. The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chem. Biol.* **2013**, *20*, 1225–1234.
(135) Zhang, H. L. Mutactin, a Novel Polyketide from Streptomyces Coelicolor. Structure and Biosynthetic Relationship to Actinorhodin. *J Org Chem* **1990**, *55*, 1682–1684.
(136) Zhao, S. Molecular Basis for Polyketide Ketoreductase–Substrate Interactions. *Int. J. Mol. Sci.* **2020**, *21*, 7562.
(137) Kharel, M. K. Elucidation of post-PKS tailoring steps involved in landomycin biosynthesis. *Org Biomol Chem* **2012**, *10*, 4256–4265.
(138) Mayer, A. LanV, a Bifunctional Enzyme: Aromatase and Ketoreductase during Landomycin a Biosynthesis. *ChemBioChem* **2005**, *6*, 2312–2315.
(139) Dickens, M. L.; Ye, J.; Strohl, W. R. Cloning, Sequencing, and Analysis of Aklaviketone Reductase from Streptomyces Sp. Strain C5. *J. Bacteriol.* **1996**, *178*, 3384–3388.
(140) Lackner, G. Biosynthesis of Pentangular Polyphenols: Deductions from the Benastatin and Griseorhodin Pathways. *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
(141) Torkkell, S. The Entire Nogalamycin Biosynthetic Gene Cluster of Streptomyces Nogalater: Characterization of a 20-Kb DNA Region and Generation of Hybrid Structures. *Mol Gen Genomics* **2001**, *266*, 276–288.
(142) Paananen, P. Structural and Functional Analysis of Angucycline C-6 Ketoreductase LanV Involved in Landomycin Biosynthesis. *Biochemistry* **2013**, *52*, 5304–5314.
(143) Patrikainen, P. Structure-Based Engineering of Angucyclinone 6-Ketoreductases. *Chem. Biol.* **2014**, *21*, 1381–1391.
(144) Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C. Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (12), 3421–3430. https://doi.org/10.1021/acschembio.6b00658.
(145) Xiao, X. Functional and Structural Insights into a Novel Promiscuous Ketoreductase of the Lugdunomycin Biosynthetic Pathway. *ACS Chem Biol* **2020**, *15*, 2529–2538.

(146) Kallio, P. Flavoprotein Hydroxylase PgaE Catalyzes Two Consecutive Oxygen-Dependent Tailoring Reactions in Angucycline Biosynthesis. *Biochemistry* **2011**, *50*, 5535–5543.
(147) Patrikainen, P. Tailoring Enzymes Involved in the Biosynthesis of Angucyclines Contain Latent Context-Dependent Catalytic Activities. *Chem. Biol.* **2012**, *19*, 647–655.
(148) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **2003**, *4*, 654–657.
(149) Moreno, G. O. *Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type i and Type II Polyketide Synthases Using Atom-Replaced Probes*; UC Irvine, 2018.
(150) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(151) Li 然 李嫣, Y.; Xu 伟徐, W.; Tang 奕唐, Y. Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains*. *J. Biol. Chem.* **2010**, *285* (30), 22764–22773. https://doi.org/10.1074/jbc.M110.128504.
(152) Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C. A. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461* (7267), 1139–1143. https://doi.org/10.1038/nature08475.
(153) Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. *Proc. Natl. Acad. Sci.* **2017**, *114* (21), E4142–E4148. https://doi.org/10.1073/pnas.1609001114.
(154) Liu, L.; Zhang, Z.; Shao, C.-L.; Wang, J.-L.; Bai, H.; Wang, C.-Y. Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Sci. Rep.* **2015**, *5* (1), 10463. https://doi.org/10.1038/srep10463.
(155) Feng, Y.; Yang, X.; Ji, H.; Deng, Z.; Lin, S.; Zheng, J. The Streptomyces Viridochromogenes Product Template Domain Represents an Evolutionary Intermediate between Dehydratase and Aldol Cyclase of Type I Polyketide Synthases. *Commun. Biol.* **2022**, *5* (1), 508. https://doi.org/10.1038/s42003-022-03477-8.
(156) Pan, H. Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. *Structure* **2002**, *10*, 1559–1568.
(157) Thompson, T. B.; Katayama, K.; Watanabe, K.; Hutchinson, C. R.; Rayment, I. Structural and Functional Analysis of Tetracenomycin F2 Cyclase from Streptomyces Glaucescens: A TYPE II POLYKETIDE CYCLASE*. *J. Biol. Chem.* **2004**, *279*, 37956–37963.

(158) Sultana, A. Structure of the Polyketide Cyclase SnoaL Reveals a Novel Mechanism for Enzymatic Aldol Condensation. *EMBO J.* **2004**, *23*, 1911–1921.
(159) Kallio, P.; Sultana, A.; Niemi, J.; Mäntsälä, P.; Schneider, G. Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity. *J. Mol. Biol.* **2006**, *357*, 210–220.
(160) Hutchinson, C. R. Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem. Rev.* **1997**, *97* (7), 2525–2536. https://doi.org/10.1021/cr960022x.
(161) Ahuja, M.; Chiang, Y.-M.; Chang, S.-L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H.-C.; Yeh, H.-H.; Oakley, B. R.; Wang, C. C. C. Illuminating the Diversity of Aromatic Polyketide Synthases in Aspergillus Nidulans. *J. Am. Chem. Soc.* **2012**, *134* (19), 8212–8221. https://doi.org/10.1021/ja3016395.
(162) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97* (7), 2465–2498. https://doi.org/10.1021/cr960034i.
(163) Shen, B. Polyketide Biosynthesis beyond the Type I, II and III Polyketide Synthase Paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6.
(164) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639. https://doi.org/10.1021/ar8002249.
(165) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130. https://doi.org/10.1126/science.ade2574.
(166) Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. Structural Basis for Selectivity in a Highly Reducing Type II Polyketide Synthase. *Nat. Chem. Biol.* **2020**, *16* (7), 776–782. https://doi.org/10.1038/s41589-020-0530-0.
(167) Yadav, G.; Gokhale, R. S.; Mohanty, D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J. Mol. Biol.* **2003**, *328* (2), 335–363. https://doi.org/10.1016/S0022-2836(03)00232-8.
(168) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci.* **2021**, *118* (15), e2016239118. https://doi.org/10.1073/pnas.2016239118.
(169) Capela, J.; Zimmermann-Kogadeeva, M.; Dijk, A. D. J. van; de Ridder, D.; Dias, O.; Rocha, M. Comparative Assessment of Protein Large Language Models for Enzyme Commission Number Prediction. *BMC Bioinformatics* **2025**, *26* (1), 68. https://doi.org/10.1186/s12859-025-06081-9.
(170) Jacques, M.-A.; Durand, K.; Orgeur, G.; Balidas, S.; Fricot, C.; Bonneau, S.; Quillévéré, A.; Audusseau, C.; Olivier, V.; Grimault, V.; Mathis, R. Phylogenetic Analysis and Polyphasic Characterization of Clavibacter Michiganensis Strains Isolated from Tomato Seeds Reveal That Nonpathogenic Strains Are Distinct from

C. Michiganensis Subsp. Michiganensis. *Appl. Environ. Microbiol.* **2012**, *78* (23), 8388–8402. https://doi.org/10.1128/AEM.02158-12.

(171) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat. Chem. Biol.* **2019**, *15* (7), 669–671. https://doi.org/10.1038/s41589-019-0301-y.

(172) Ellis, B. D.; Milligan, J. C.; White, A. R.; Duong, V.; Altman, P. X.; Mohammed, L. Y.; Crump, M. P.; Crosby, J.; Luo, R.; Vanderwal, C. D.; Tsai, S.-C. An Oxetane-Based Polyketide Surrogate To Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140* (15), 4961–4964. https://doi.org/10.1021/jacs.7b11793.

(173) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(174) Zdouc, M. M.; Blin, K.; Louwen, N. L. L.; Navarro, J.; Loureiro, C.; Bader, C. D.; Bailey, C. B.; Barra, L.; Booth, T. J.; Bozhüyük, K. A. J.; Cediel-Becerra, J. D. D.; Charlop-Powers, Z.; Chevrette, M. G.; Chooi, Y. H.; D'Agostino, P. M.; de Rond, T.; Del Pup, E.; Duncan, K. R.; Gu, W.; Hanif, N.; Helfrich, E. J. N.; Jenner, M.; Katsuyama, Y.; Korenskaia, A.; Krug, D.; Libis, V.; Lund, G. A.; Mantri, S.; Morgan, K. D.; Owen, C.; Phan, C.-S.; Philmus, B.; Reitz, Z. L.; Robinson, S. L.; Singh, K. S.; Teufel, R.; Tong, Y.; Tugizimana, F.; Ulanova, D.; Winter, J. M.; Aguilar, C.; Akiyama, D. Y.; Al-Salihi, S. A. A.; Alanjary, M.; Alberti, F.; Aleti, G.; Alharthi, S. A.; Rojo, M. Y. A.; Arishi, A. A.; Augustijn, H. E.; Avalon, N. E.; Avelar-Rivas, J. A.; Axt, K. K.; Barbieri, H. B.; Barbosa, J. C. J.; Barboza Segato, L. G.; Barrett, S. E.; Baunach, M.; Beemelmanns, C.; Beqaj, D.; Berger, T.; Bernaldo-Agüero, J.; Bettenbühl, S. M.; Bielinski, V. A.; Biermann, F.; Borges, R. M.; Borriss, R.; Breitenbach, M.; Bretscher, K. M.; Brigham, M. W.; Buedenbender, L.; Bulcock, B. W.; Cano-Prieto, C.; Capela, J.; Carrion, V. J.; Carter, R. S.; Castelo-Branco, R.; Castro-Falcón, G.; Chagas, F. O.; Charria-Girón, E.; Chaudhri, A. A.; Chaudhry, V.; Choi, H.; Choi, Y.; Choupannejad, R.; Chromy, J.; Donahey, M. S. C.; Collemare, J.; Connolly, J. A.; Creamer, K. E.; Crüsemann, M.; Cruz, A. A.; Cumsille, A.; Dallery, J.-F.; Damas-Ramos, L. C.; Damiani, T.; de Kruijff, M.; Martín, B. D.; Sala, G. D.; Dillen, J.; Doering, D. T.; Dommaraju, S. R.; Durusu, S.; Egbert, S.; Ellerhorst, M.; Faussurier, B.; Fetter, A.; Feuermann, M.; Fewer, D. P.; Foldi, J.; Frediansyah, A.; Garza, E. A.; Gavriilidou, A.; Gentile, A.; Gerke, J.; Gerstmans, H.; Gomez-Escribano, J. P.; González-Salazar, L. A.; Grayson, N. E.; Greco, C.; Gomez, J. E. G.; Guerra, S.; Flores, S. G.; Gurevich, A.; Gutiérrez-García, K.; Hart, L.; Haslinger, K.; He, B.; Hebra, T.; Hemmann, J. L.; Hindra, H.; Höing, L.; Holland, D. C.; Holme, J. E.; Horch, T.; Hrab, P.; Hu, J.;

Huynh, T.-H.; Hwang, J.-Y.; Iacovelli, R.; Iftime, D.; Iorio, M.; Jayachandran, S.; Jeong, E.; Jing, J.; Jung, J. J.; Kakumu, Y.; Kalkreuter, E.; Kang, K. B.; Kang, S.; Kim, W.; Kim, G. J.; Kim, H.; Kim, H. U.; Klapper, M.; Koetsier, R. A.; Kollten, C.; Kovács, Á. T.; Kriukova, Y.; Kubach, N.; Kunjapur, A. M.; Kushnareva, A. K.; Kust, A.; Lamber, J.; Larralde, M.; Larsen, N. J.; Launay, A. P.; Le, N.-T.-H.; Lebeer, S.; Lee, B. T.; Lee, K.; Lev, K. L.; Li, S.-M.; Li, Y.-X.; Licona-Cassani, C.; Lien, A.; Liu, J.; Lopez, J. A. V.; Machushynets, N. V.; Macias, M. I.; Mahmud, T.; Maleckis, M.; Martinez-Martinez, A. M.; Mast, Y.; Maximo, M. F.; McBride, C. M.; McLellan, R. M.; Bhatt, K. M.; Melkonian, C.; Merrild, A.; Metsä-Ketelä, M.; Mitchell, D. A.; Müller, A. V.; Nguyen, G.-S.; Nguyen, H. T.; Niedermeyer, T. H. J.; O'Hare, J. H.; Ossowicki, A.; Ostash, B. O.; Otani, H.; Padva, L.; Paliyal, S.; Pan, X.; Panghal, M.; Parade, D. S.; Park, J.; Parra, J.; Rubio, M. P.; Pham, H. T.; Pidot, S. J.; Piel, J.; Pourmohsenin, B.; Rakhmanov, M.; Ramesh, S.; Rasmussen, M. H.; Rego, A.; Reher, R.; Rice, A. J.; Rigolet, A.; Romero-Otero, A.; Rosas-Becerra, L. R.; Rosiles, P. Y.; Rutz, A.; Ryu, B.; Sahadeo, L.-A.; Saldanha, M.; Salvi, L.; Sánchez-Carvajal, E.; Santos-Medellin, C.; Sbaraini, N.; Schoellhorn, S. M.; Schumm, C.; Sehnal, L.; Selem, N.; Shah, A. D.; Shishido, T. K.; Sieber, S.; Silviani, V.; Singh, G.; Singh, H.; Sokolova, N.; Sonnenschein, E. C.; Sosio, M.; Sowa, S. T.; Steffen, K.; Stegmann, E.; Streiff, A. B.; Strüder, A.; Surup, F.; Svenningsen, T.; Sweeney, D.; Szenei, J.; Tagirdzhanov, A.; Tan, B.; Tarnowski, M. J.; Terlouw, B. R.; Rey, T.; Thome, N. U.; Torres Ortega, L. R.; Tørring, T.; Trindade, M.; Truman, A. W.; Tvilum, M.; Udwary, D. W.; Ulbricht, C.; Vader, L.; van Wezel, G. P.; Walmsley, M.; Warnasinghe, R.; Weddeling, H. G.; Weir, A. N. M.; Williams, K.; Williams, S. E.; Witte, T. E.; Rocca, S. M. W.; Yamada, K.; Yang, D.; Yang, D.; Yu, J.; Zhou, Z.; Ziemert, N.; Zimmer, L.; Zimmermann, A.; Zimmermann, C.; van der Hooft, J. J. J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 4.0: Advancing Biosynthetic Gene Cluster Curation through Global Collaboration. *Nucleic Acids Res.* **2025**, *53* (D1), D678–D690. https://doi.org/10.1093/nar/gkae1115.

(175) Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S. J.; Webb, C.; Wu, D.; Varghese, N. J.; Reddy, T. B. K.; Mukherjee, S.; Ovchinnikova, G.; Nolan, M.; Seshadri, R.; Roux, S.; Visel, A.; Woyke, T.; Eloe-Fadrosh, E. A.; Kyrpides, N. C.; Ivanova, N. N. The IMG/M Data Management and Analysis System v.7: Content Updates and New Features. *Nucleic Acids Res.* **2023**, *51* (D1), D723–D732. https://doi.org/10.1093/nar/gkac976.

(176) Fernández-Moreno, M. A.; Martínez, E.; Caballero, J. L.; Ichinose, K.; Hopwood, D. A.; Malpartida, F. DNA Sequence and Functions of the actVI Region of the Actinorhodin Biosynthetic Gene Cluster of Streptomyces Coelicolor A3(2). *J. Biol. Chem.* **1994**, *269* (40), 24854–24863. https://doi.org/10.1016/S0021-9258(17)31469-2.

(177) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Richard Hutchinson, C. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4* (6), 433–443. https://doi.org/10.1016/S1074-5521(97)90195-2.

(178) Moreno, G. O. Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type I and Type II Polyketide Synthases Using Atom-Replaced Probes, UC Irvine, 2018. https://escholarship.org/uc/item/1tx0518k.
(179) Serapian, S. A.; Crosby, J.; Crump, M. P.; van der Kamp, M. W. Path to Actinorhodin: Regio- and Stereoselective Ketone Reduction by a Type II Polyketide Ketoreductase Revealed in Atomistic Detail. *JACS Au* **2022**, *2* (4), 972–984. https://doi.org/10.1021/jacsau.2c00086.
(180) Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941–953. https://doi.org/10.1016/j.jmb.2008.09.084.
(181) Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12* (10), 1865–1875. https://doi.org/10.1016/j.str.2004.08.002.
(182) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat. Prod. Rep.* **2016**, *33* (2), 203–230. https://doi.org/10.1039/C5NP00109A.
(183) Chooi, Y.-H.; Tang, Y. Navigating the Fungal Polyketide Chemical Space: From Genes to Molecules. *J. Org. Chem.* **2012**, *77* (22), 9933–9953. https://doi.org/10.1021/jo301592k.
(184) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(185) Barajas, J. F.; Finzel, K.; Valentic, T. R.; Shakya, G.; Gamarra, N.; Martinez, D.; Meier, J. L.; Vagstad, A. L.; Newman, A. G.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Structural and Biochemical Analysis of Protein–Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew. Chem. Int. Ed.* **2016**, *55* (42), 13005–13009. https://doi.org/10.1002/anie.201605401.
(186) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(187) Barajas, J. F. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. In *Proceedings of the National Academy of Sciences*; 2017; Vol. 114, pp 4142–4148.
(188) Milligan, J. C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat Chem Biol* **2019**, *15*, 669–671.
(189) Ellis, B. D. An Oxetane-Based Polyketide Surrogate to Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140*, 4961–4964.
(190) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43*, 14529–14538.

Chapter 4

Structural Informatics of Product Template Domain Regioselectivity in Fungal NR-PKSs

4.1 Summary

Fungal non-reducing polyketide synthases (NR-PKSs) rely on the product template (PT) domain to control the first aromatization step of polyketide biosynthesis, thereby establishing the core ring-connectivity patterns that define fungal aromatic natural products. Although PT domains have long been recognized as key determinants of cyclization regioselectivity, the structural basis by which they program C2–C7, C4–C9, and C6–C11 aldol condensations remains poorly understood. In this study, we integrate high-accuracy AlphaFold3 structural predictions, extensive docking of linear and pre-cyclized polyketides, ANOVA-based statistical modeling, and ACP–PT complex simulations to elucidate the mechanistic foundations of PT specificity. Across forty PT domains representing five phylogenetic clades, docking energies were significantly shaped by PT clade, ketide chain length, and their interaction, thus revealing clade-specific substrate-length preferences. Structural analysis identified three key determinants of regioselectivity: (i) catalytic dyad geometry, which correlates with cyclization mode; (ii) pocket volume and polarity, which govern substrate accommodation and chain-length compatibility; and (iii) the orientation of a finger-like loop that modulates pocket accessibility and folding trajectories. ACP–PT multimer modeling further showed that substrate delivery can constrain effective chain length, particularly in PTs with deep but poorly aligned cavities. Together, these findings support a unified model in which regioselectivity emerges from the interplay between catalytic dyad positioning, pocket architecture, and ACP-mediated substrate presentation. This framework enables improved prediction of fungal

aromatic scaffolds from PT sequence or structure and offers new mechanistic insight into the evolution and functional diversification of fungal NR-PKS biosynthetic logic.

4.2 Introduction

Fungal non-reducing polyketide synthases (NR-PKSs) are iterative megasynthases responsible for producing a vast array of aromatic natural products with roles in pigmentation, pathogenicity, chemical defense, and pharmaceutical activity.[183] These multidomain enzymes build highly reactive poly-β-keto chains that must be precisely folded to generate chemically diverse scaffolds such as anthraquinones, naphthalenes, resorcinols, and pyran derivatives. Because these linear intermediates are unstable and prone to spontaneous cyclization, NR-PKSs exert strict control over the first aromatization step—the F-mode aldol condensation—typically occurring in a C2–C7, C4–C9, or C6–C11 manner.[184] This early folding event dictates the global architecture of the final metabolite.

Central to this control is the product template (PT) domain, located between the AT and ACP domains of the PKS megasynthase. The PT domain selectively folds the nascent polyketide and catalyzes the regiospecific aldol cyclization that forms the first aromatic ring, thereby programming the fundamental ring-connectivity pattern of fungal aromatic scaffolds.[151] Genetic dissection of systems such as PKS4 and PksA demonstrated that removing the PT domain abolishes directed cyclization, producing aberrantly folded products, whereas providing a PT in trans restores the correct C2–C7 or C4–C9/C2–C11 regioselectivity.[151] These findings established the PT domain as the "cyclization logic gate" for NR-PKSs.

Phylogenetic analysis has shown that PT domains can be classified into several functional clades that correlate strongly with cyclization modes and product sizes.[151] Li et al. (2010) demonstrated that PT domains from NR-PKSs of known function fall into five clades, each encoding C2–C7, C4–C9, or C6–C11 regioselectivity, and experimentally validated that Group V PTs catalyze C6–C11/C4–C13 cyclizations by transplanting them into the PKS4

minimal system.[151] This work established that cyclization specificity is embedded within PT sequence and can be predicted phylogenetically.

However, despite their central biosynthetic role, the structural and mechanistic basis of PT selectivity remains poorly understood. To date, only one fungal PT crystal structure (PksA-PT) has been solved, revealing a "double hot-dog" fold with a multi-region internal cavity consisting of a phosphopantetheine-binding region, a cyclization chamber, and a hydrophobic hexyl-binding region.[152] The catalytic Asp–His dyad resides within the cyclization chamber, suggesting that the shape and orientation of the pocket govern substrate folding trajectories.[152] Yet most PT domains lack experimental structures, and the generality of this architectural framework across fungal clades remains unclear.

Liu et al. (2015) expanded the PT phylogeny to 661 NR-PKSs and identified eight groups with distinct structural features and cyclization modes (C2–C7, C4–C9, C6–C11).[154] Their modeling showed that differences in cavity volume, cavity-lining residues (CLRs), catalytic dyad positioning, and the presence of finger-like loops correlate with the length of the accommodated polyketide and the preferred aldol condensation mode. These analyses suggested that subtle evolutionary modifications in pocket geometry and CLR composition can reprogram cyclization specificity.[154] However, their conclusions were limited by the accuracy of homology models available prior to modern structure prediction tools.

Meanwhile, genome sequencing has revealed hundreds of cryptic NR-PKS clusters, many of which produce unknown metabolites.[161] Because full PKS reconstitution is experimentally challenging, computational prediction of PT regioselectivity has become vital for natural product discovery and genome mining. Yet no study has systematically investigated PT domains using high-accuracy structural predictions (e.g., AlphaFold3) combined with quantitative ligand docking or explored how PT pocket geometry interacts with chain length,

pre-cyclized versus post-cyclized intermediates, or ACP–PT protein–protein interactions, which may influence substrate delivery. This knowledge gap has prevented effective directed biosynthesis of new fungal polyketides via the engineering of PT domains.

In this chapter, we address these knowledge gaps by analyzing 40 fungal PT domains across five phylogenetic groups. Using AlphaFold3 structural modeling, docking simulations of linear and pre-cyclized polyketides across multiple chain lengths, ANOVA-based statistical modeling, and ACP–PT complex prediction, we investigate:

1. How structural features of PT domains—catalytic dyad orientation, pocket depth/volume, and loop architecture—correlate with cyclization selectivity;
2. How substrate length and pre-cyclized conformation influence PT binding energetics across different clades;
3. How ACP–PT interactions may contribute to substrate positioning in clades with atypical or restrictive chain-length preferences.

Together, this work provides new mechanistic insight into the basis of fungal aromatic ring formation and offers an updated framework for predicting PT cyclization specificity from sequence or structure.

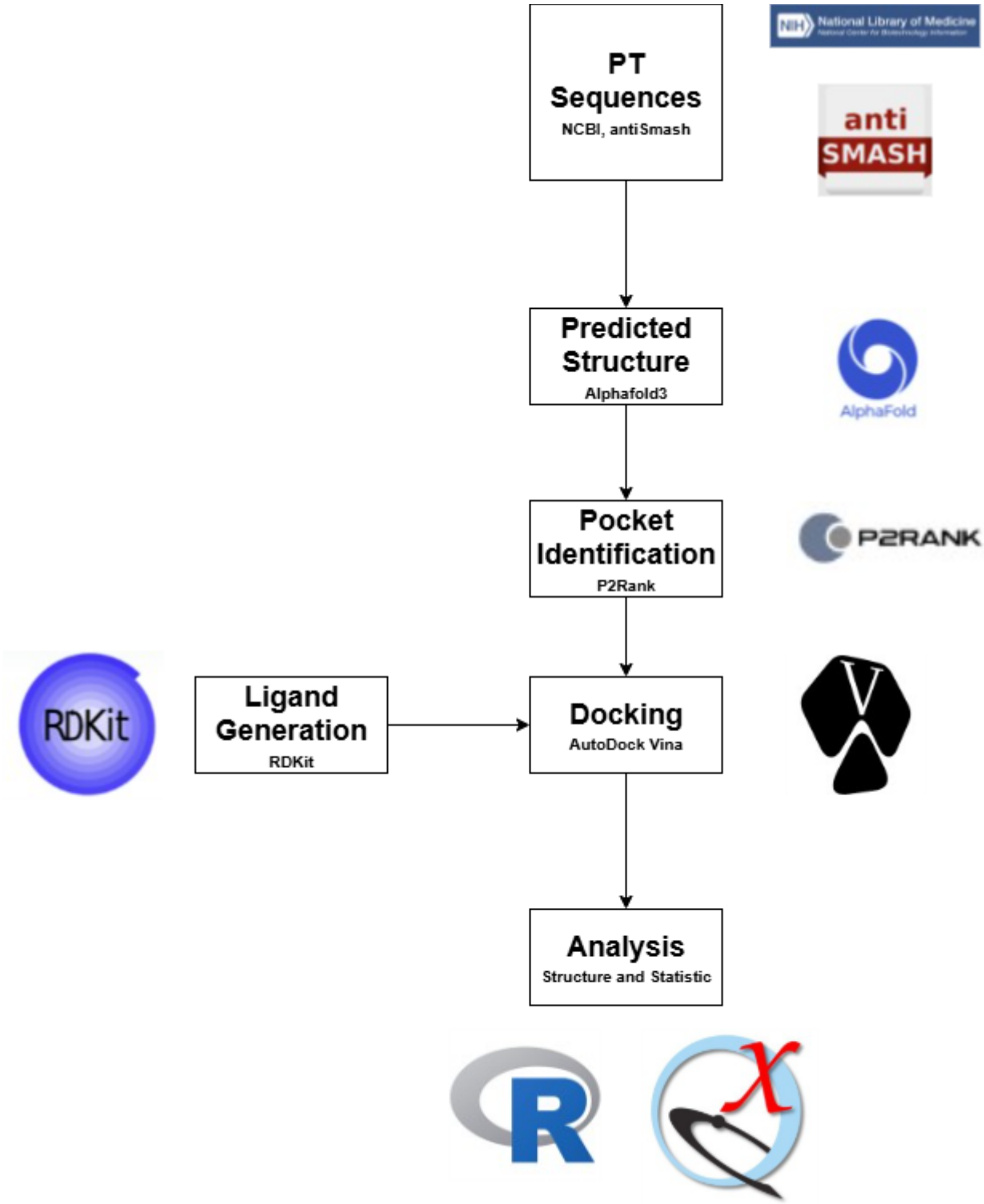


Figure 4.1. *Structural Informatic workflow of Type I Iterative PKS PT.*

4.3 Results

4.3.1 Statistical Analysis of Docking Energies

Analysis of variance (ANOVA) revealed that docking energies were significantly influenced by PT domain category, ketide chain length, and cyclization type ($p < 0.01$ for all factors)(Table 4.1). The interaction between category and ketide length was also significant, indicating that the effect of chain elongation on binding affinity varied among PT clades. In contrast, the influence of cyclization pattern (C2–C7, C4–C9, and C6–C11) was generally consistent across all groups, suggesting that PTs within different clades share similar energetic preferences toward specific cyclization modes(Figure 4.1).

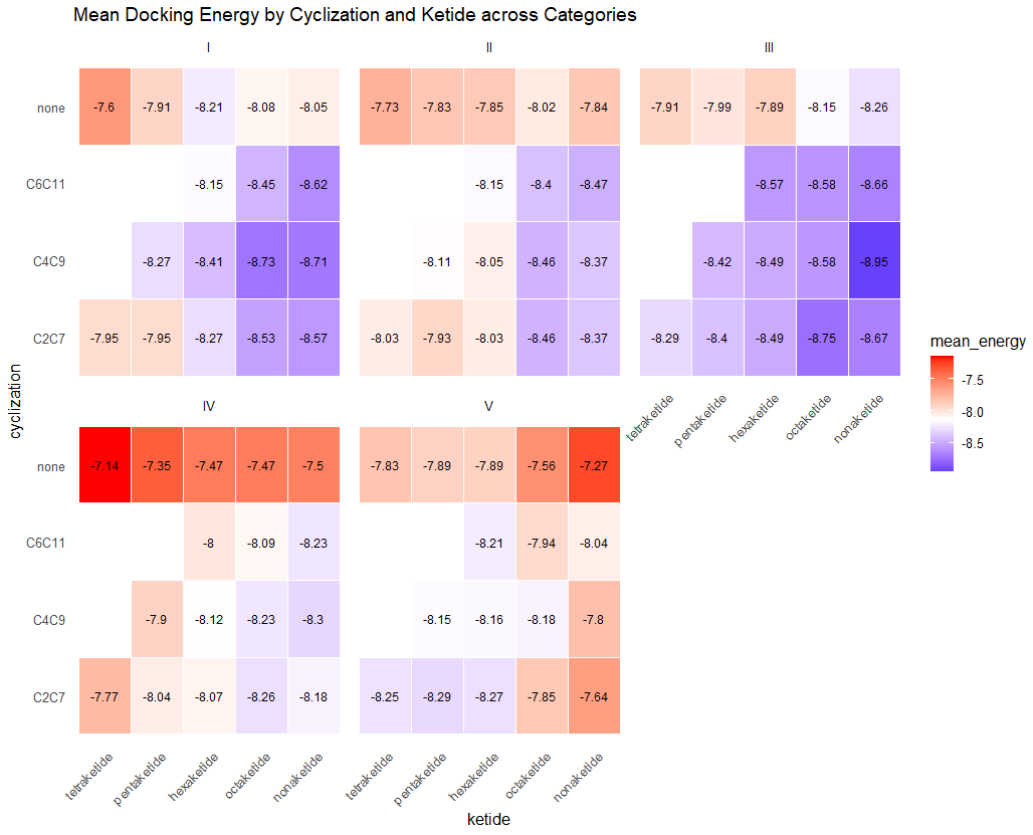


Figure 4.2. *Mean Docking Energy by Cyclization and Ketide across Categories. Rows represent individual ligand (None=linear, CXCX=cyclization pattern), and columns represent length of polyketides. Color intensity indicates the binding energy of ligand and receptor. Red indicates lower binding energy, while blue indicates higher binding energy.*

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
category	4	20.1	5.015	5.351	0.000308 ***
ketide_length	1	7.2	7.167	7.647	0.005866 **
cyclization	3	26.9	8.966	9.567	3.55e-06 ***
category:ketide_length	4	11.9	2.972	3.171	0.013567 *
category:cyclization	12	2.9	0.24	0.256	0.994829
ketide_length:cyclization	3	0.7	0.218	0.233	0.873739
category:ketide_length:cyclization	12	0.6	0.054	0.057	0.999998
Residuals	589	552	0.937		

Table 4.1. *One-way ANOVA analysis of Docking Energy treated with Cyclization, Ketide length and Categories. Data are presented as mean square (MS), degrees of freedom (df), and F-values. Significant differences were determined using a one-way ANOVA followed by Tukey's HSD post-hoc test for multiple comparisons. *P < 0.05, **P < 0.01, ***P < 0.001 vs. control group.*

4.3.2 Structural and Docking Characteristics Across Clades

Clade I PTs displayed a deep and extended binding pocket but showed favorable docking energies only with tetraketide substrates(Figure 4.3A). The narrow substrate range, despite the large pocket size, suggests that ACP–PT interactions may play a major role in substrate positioning. AlphaFold3-based ACP–PT docking further supported this hypothesis: the ACP docked correctly at the pocket entrance, but the tethered ligand remained distant from the catalytic center, indicating that ACP orientation and protein dynamics likely influences substrate delivery and positioning.

Clades II and III(Figure 4.3BC), both associated with C2–C7 cyclization, exhibited comparable overall architecture but distinct pocket depths. Clade III PTs possessed deeper catalytic cavities, consistent with the ability to accommodate longer ketide intermediates. Docking analyses supported this observation: Clade III PTs showed stronger binding affinities

for extended substrates such as nonaketides, while Clade II PTs preferred shorter intermediates.

In contrast, Clade IV PTs demonstrated lower average binding energies across all ligand types. Structural examination revealed the presence of a non-polar region near the pocket entrance, consistent with their known utilization of hexanoyl starter units in corresponding PKSs(Figure 4.3D). This hydrophobic feature likely stabilizes early intermediates and contributes to the observed lower docking energies.

Finally, Clade V PTs possessed the smallest and most restricted binding pockets, resulting in weak binding affinities for longer ketides (e.g., nonaketides). The pocket geometry suggests a preference for shorter intermediates and early cyclization events, in line with their proposed functional role in compact aromatic ring formation(Figure 4.3E).

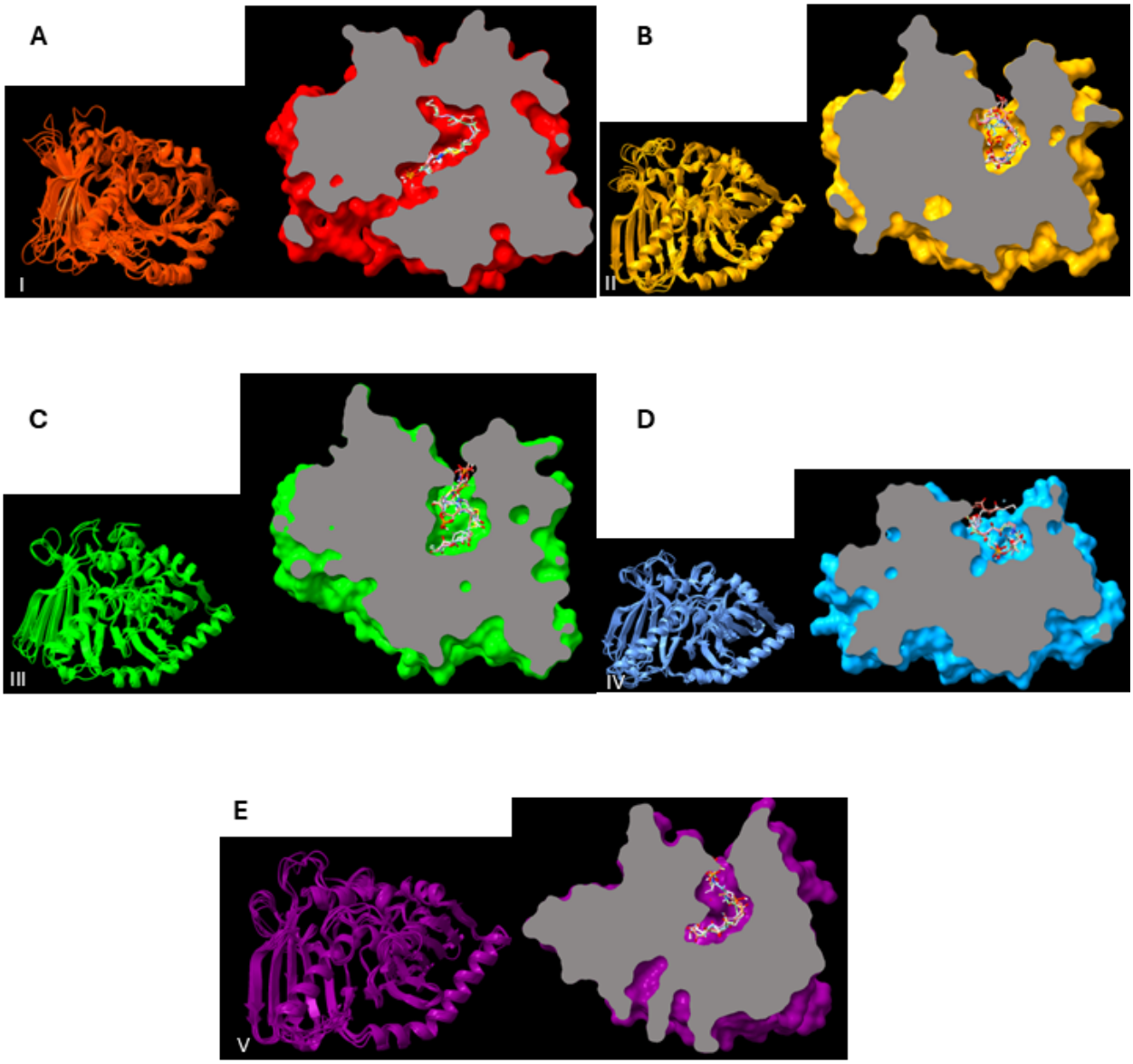


Figure 4.3. *Predicted PT structures of Clade I-V(A-E) with side view of docked pocket. Ligands are tetraketide, pentaketide, hexaketide, octaketide and nonaketide.*

4.3.3 Structural Correlation Analysis

To identify structural determinants underlying the observed docking and energetic trends, all predicted PT domain models were analyzed for variations in catalytic dyad geometry, pocket volume, and finger-like loop orientation.

Across clades, the distance and relative orientation between the catalytic histidine and aspartate residues were found to correlate with cyclization regioselectivity. PTs favoring C2–C7 cyclization (Clades II and III) displayed a compact catalytic dyad (average His–Asp distance of ~3.2 Å), which positions the reactive carbonyl groups for intramolecular aldol condensation near the C2 and C7 atoms. In contrast, C4–C9 and C6–C11-type PTs (Clades IV and V) showed more open dyad geometries (3.8–4.2 Å), potentially accommodating different folding trajectories of the polyketide chain before ring closure.

Pocket volume was another strong predictor of substrate length preference. Larger cavities in Clades I and III correlated with stronger binding to longer intermediates (up to nonaketides), whereas smaller pockets in Clade V constrained the accommodation of extended chains, explaining their weaker affinities for long substrates. These structural differences are consistent with functional specialization toward polyketide products of varying chain lengths.

Finally, the finger-like loop—a flexible structural motif near the pocket entrance—exhibited distinct orientations among clades. In Clades II and III, the loop formed a partially enclosed lid that may stabilize the substrate conformation required for C2–C7 cyclization. By contrast, in Clades IV and V, the loop was rotated outward, leaving the pocket more open and accessible, a configuration that may favor alternative cyclization modes such as C4–C9 or C6–C11.

Together, these analyses highlight that cyclization regioselectivity is encoded by subtle structural variations in the PT active site architecture, integrating catalytic dyad geometry, pocket topology, and loop flexibility to control substrate folding and ring formation.

4.4 Discussion:

The goal of this study was to understand how fungal product template (PT) domains encode cyclization regioselectivity, and how differences in active-site geometry shape preferences for C2–C7, C4–C9, and C6–C11 aldol cyclizations. By combining AlphaFold3 structural predictions, extensive ligand docking, statistical modeling, and ACP–PT complex simulations, the results provide new mechanistic insights that refine and extend the major conclusions from previous PT domain studies. Because the ACP-PT docking for Clade I PT is distinctly different from those for Clades II-V, here we focus on comparisons between Clades II-V.

4.4.1 Catalytic Dyad Geometry and Cyclization Outcome

One of the clearest trends emerging from both previous studies and our modeling is the importance of the His–Asp catalytic dyad in defining cyclization regioselectivity. Liu et al. (2015) reported that PTs catalyzing C2–C7 cyclization typically possess a shallow pocket with the dyad positioned near the base[154], whereas C4–C9 and C6–C11 PTs contain elongated pockets with the dyad located deeper within the cavity. Our AlphaFold3 predictions corroborate this structural organization: C2–C7 PTs (Clades II and III) exhibit compact, bottom-positioned dyads that constrain the substrate into an early folding trajectory, while C4–C9 and C6–C11 PTs (Clades IV and V) display mid-pocket dyads situated in a more extended cavity, consistent with their ability to accommodate longer polyketide chains before cyclization. The convergence of our results with prior structural analyses supports the view that dyad placement is a central architectural feature governing regioselectivity.

4.4.2 Pocket size, shape and polarity

The differences in pocket volume across clades strongly correlate with substrate-length preferences observed in docking experiments. Clade III PTs contain relatively deep cavities that accommodate longer chains and show improved affinity for nonaketides, whereas Clade II PTs possess smaller pockets more compatible with pentaketide intermediates. Clade V PTs, although associated with C6–C11 cyclization, contain pockets that are elongated but tightly contoured, limiting their capacity to bind very long chains such as nonaketides. These findings align with Liu et al.'s conclusion that PT cavity size is a major determinant of final product chain length.

Pocket polarity also contributes to substrate behavior. In Clade IV, the uniformly low docking energies observed for hexaketide-like intermediates correspond with the hydrophobic "hexyl-binding" region identified in earlier structural models. This niche is optimized for hexanoyl starter units, thus stabilizing early intermediates but potentially reducing affinity for substrates lacking hydrophobic substituents. This adds a second layer of specificity—beyond chain length—by linking pocket composition to starter-unit recognition.

4.4.3 Role of Finger-Like Loop Orientation in Substrate Positioning

Another structural element influencing PT function is the finger-like loop near the pocket entrance. Prior work identified these loops as present in PTs that catalyze later-folding mechanisms (C4–C9 and C6–C11) and absent or compressed in early-folding C2–C7 systems. Our models reproduce these features and reveal a functional consequence: outward-facing loops in Clades IV and V create an expanded entrance that allows longer polyketides to align properly near the catalytic dyad, whereas inward or minimal loops in Clades II and III produce

a narrower channel, restricting the substrate and favoring early folding. This suggests that the loop acts as a flexible gating mechanism controlling substrate trajectory.

4.4.4 Influence of ACP-PT Interactions on Substrate Delivery

The ACP–PT interface adds a critical layer of control beyond static pocket architecture. Our observation that Clade I PTs have very deep cavities yet strongly favor only tetraketide substrates finds a compelling mechanistic explanation in light of the research from our previous member Dr. Barajas.[185] In our earlier study, we used mechanism-based crosslinkers, in vitro assays, in silico docking, and structure-directed mutagenesis to map the transient interactions between ACP and PT in fungal NR-PKSs. We identified specific surface residues of ACP and PT that mediate recognition, and showed that enhancing compatibility at this interface (even with non-cognate partners) improves functional coupling.[185] These data suggest that productive cyclization depends not only on a suitable pocket but also on optimal ACP docking geometry. In our Clade I models, ACP docking placed the phosphopantetheinyl-tethered ligand far from the dyad, consistent with a suboptimal orientation. This misalignment could effectively shorten the "functional" substrate length in catalytically relevant contexts — even when the cavity is physically large. In other words, poor ACP positioning can act as a gatekeeper, limiting the chain length that is actually presented to the PT active site. Clearly, extensive protein dynamics is involved when PT and ACP is docked with the optimal orientation for C2-C7 cyclization to take place. Alteranatively, C2-C7 cyclization may also occur spontaneously for the tetraketide at the entrance of the PT pocket (as opposed to the catalytic dyad).

4.4.5 Evolutionary Interpretation of Structural Divergence

The phylogenetic patterns described by Liu et al. identify cavity-lining residues (CLRs) as major contributors to PT evolution.[154] Our structural comparisons support this model: clades with larger or more specialized pocket shapes (IV and V) display the highest number of CLR substitutions, whereas C2–C7 clades preserve most conserved residues, corresponding with their narrower substrate and cyclization profiles. The presence or absence of finger-like regions also appears to stem from small residue changes or short insertions, illustrating how modest structural modifications allow PT domains to diversify cyclization outcomes while retaining a conserved catalytic core.

4.4.6 Methodological Considerations

While docking captures energetic preferences and steric compatibility, it cannot fully model the dynamic substrate movement inside PT pockets, ACP-driven conformational changes, or transition-state stabilization during aldol condensation.

Nonetheless, the strong agreement between structural modeling, docking ANOVA, and known biochemical trends suggests that our computational pipeline captures core mechanistic features with high reliability.

Future molecular dynamics and experimental validation will be valuable for refining these mechanistic models.

4.5 Conclusion

This chapter provides a comprehensive structural and mechanistic analysis of fungal product template (PT) domains, revealing how variations in catalytic geometry, pocket architecture, and ACP-mediated substrate delivery collectively program cyclization regioselectivity in non-reducing polyketide synthases (NR-PKSs). By integrating AlphaFold3 structural predictions with systematic docking of linear and pre-cyclized intermediates, statistical modeling, and ACP–PT interaction analysis, we uncover a set of architectural principles that unify and extend previous models of PT function.

Our results demonstrate that three structural elements—catalytic dyad placement, pocket volume and polarity, and finger-like loop orientation—consistently correlate with the observed preferences for C2–C7, C4–C9, and C6–C11 aldol cyclizations across PT clades. Compact dyad geometries in C2–C7-type PTs constrain substrates into early-folding trajectories, whereas more open dyads in C4–C9 and C6–C11 PTs allow deeper substrate penetration before cyclization. Differences in pocket depth, contour, and hydrophobicity further define each clade's preferred substrate chain length and starter-unit compatibility. The finger-like loop acts as a dynamic regulatory feature, gating substrate access and guiding folding paths in a manner consistent with the required cyclization mode.

Importantly, ACP–PT multimer modeling highlights that regioselectivity cannot be attributed solely to static pocket architecture. In clades such as Clade I, misalignment at the ACP–PT interface limits the effective chain length delivered to the catalytic center, offering a mechanistic explanation for narrow substrate preferences despite large cavity dimensions. This finding emphasizes that PT domains operate as part of a coordinated catalytic module, in

which interdomain orientation and substrate channeling are integral to programming final product outcomes.

Together, these analyses support a refined conceptual framework in which PT evolution is driven by subtle modifications in cavity-lining residues, loop insertions, and dyad positioning—structural changes that allow PTs to diversify aldehyde condensation modes while retaining a conserved catalytic core. By combining modern structural prediction tools with quantitative energetic analysis, this chapter establishes new foundations for predicting fungal aromatic scaffolds from PT sequence or structure and advances our understanding of how NR-PKSs encode the logic of aromatic ring formation. This work also provides a roadmap for future studies incorporating molecular dynamics, engineered ACP–PT interfaces, and experimental reconstitution to further dissect the dynamics of polyketide cyclization.

4.6 Method

4.6.1 Selection and Classification of PT Domains

A total of forty product template (PT) domain sequences were collected from representative fungal non-reducing polyketide synthases (NR-PKSs) spanning five phylogenetically and functionally distinct categories. Sequence classification was based on previously reported NR-PKS domain architectures and cyclization modes (C2–C7, C4–C9, and C6–C11). Each PT sequence was curated to ensure the presence of conserved catalytic residues (typically His–Asp dyad) and verified for completeness of the domain boundary.

4.6.2 Structural Prediction of PT Domains

Three-dimensional structures of all PT domains were predicted using AlphaFold3 via the web server.[173] Default model parameters were used, and the highest-confidence models (based on pLDDT and predicted TM-score) were selected for downstream analysis. Structural quality was assessed by visual inspection.

4.6.3 Ligand Preparation and Docking Setup

Model substrates representing tetraketide, pentaketide, hexaketide, octaketide, and nonaketide intermediates were constructed based on known NR-PKS polyketide elongation intermediates. For each chain length, both linear forms and pre-cyclized variants corresponding to C2–C7, C4–C9, and C6–C11 cyclization modes were included.

4.6.4 Docking Simulations

Docking simulations were performed with AutoDock Vina to predict PT–substrate binding conformations.[186] For each PT domain, all ligand types (five chain lengths × three cyclization modes) were docked independently using an exhaustiveness value of 32 and num_modes = 10.

Docking grids were centered on the catalytic cavity predicted by the P2Rank package, which identifies potential ligand-binding pockets based on geometric and physicochemical features. The grid box dimensions were adjusted to fully encompass the predicted cavity and the surrounding substrate-binding channel. The lowest binding energy pose for each ligand–PT pair was used for subsequent statistical analyses.

4.6.5 Statistical Analysis

Binding energy data were analyzed using ANOVA to assess the effects of (1) PT category, (2) ketide chain length, and (3) cyclization pattern on docking affinity. Post-hoc comparisons were conducted using Tukey's HSD test to determine significant pairwise differences among categories. Statistical analyses were performed in R (v4.3.0), and results were visualized with ggplot2.

4.6.6 ACP-PT Interaction Modeling

To assess potential domain–domain interfaces, ACP–PT complex models were generated for representative Category I PTs using AlphaFold3 multimer prediction mode. Models were evaluated for interface confidence scores (ipTM) and analyzed for orientation consistency with previously characterized PKS ACP–PT complexes. The predicted interface residues were further mapped onto PT pocket structures to examine possible gating or substrate-channeling effects.

4.7 References

(1) Cragg, G. M.; Newman, D. J. Biodiversity: A Continuing Source of Novel Drug Leads. *Pure Appl. Chem.* **2005**, *77*, 7–24.

(2) Buenz, E. J.; Verpoorte, R.; Bauer, B. A. The Ethnopharmacologic Contribution to Bioprospecting Natural Products. *Annu. Rev. Pharmacol. Toxicol.* **2018**, *58*, 509–530.

(3) Gaynes, R. The Discovery of Penicillin—New Insights after More than 75 Years of Clinical Use. *Emerg Infect Dis* **2017**, *23*, 849–853.

(4) Beutler, J. A. Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol.* **2019**, *86, e67*.

(5) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176.

(6) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

(7) Hopwood, D. A. Production of 'hybrid' Antibiotics by Genetic Engineering. *Nature* **1985**, *314*, 642–644.

(8) Malpartida, F.; Hopwood, D. A. Molecular Cloning of the Whole Biosynthetic Pathway of a Streptomyces Antibiotic and Its Expression in a Heterologous Host. *Nature* **1984**, *309* (5967), 462–464. https://doi.org/10.1038/309462a0.

(9) Bibb, M. J.; Sherman, D. H.; Ōmura, S.; Hopwood, D. A. Cloning, Sequencing and Deduced Functions of a Cluster of Streptomyces Genes Probably Encoding Biosynthesis of the Polyketide Antibiotic Frenolicin. *Gene* **1994**, *142*, 31–39.

(10) Han, L.; Yang, K.; Ramalingam, E.; Mosher, R. H.; Vining, L. C. Cloning and characterization of polyketide synthase genes for jadomycin b biosynthesis in streptomyces venezuelaeISP5230. *Microbiology* **1994**, *140*, 3379–3389.

(11) Motamedi, H.; Hutchinson, C. R. Cloning and Heterologous Expression of a Gene Cluster for the Biosynthesis of Tetracenomycin C, the Anthracycline Antitumor Antibiotic of Streptomyces Glaucescens. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4445–4449.

(12) Kim, E.-S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. Sequences of the Oxytetracycline Polyketide Synthase-Encoding Otc Genes from Streptomyces Rimosus. *Gene* **1994**, *141* (1), 141–142. https://doi.org/10.1016/0378-1119(94)90144-9.

(13) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. Characterization of the streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **1994**, *151*, 1–10.

(14) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Manipulation and Analysis of an Aromatic Polyketide Synthase with Unproved Catalytic Specificities. *J. Am. Chem. Soc.* **1993**, *115*, 11671–11675.

(15) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. The tcmVI Region of the Tetracenomycin c Biosynthetic Gene Cluster of Streptomyces Glaucescens Encodes the Tetracenomycin F1 Monooxygenase, Tetracenomycin F2 Cyclase, and, Most Likely, a Second Cyclase. *J. Bacteriol.* **1993**, *175*, 7571–7580.

(16) Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Stereochemical Course of Two Reactions Catalyzed by a Polyketide Synthase. *Biochemistry* **1994**, *33*, 9321–9326.

(17) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91* (24), 11542–11546. https://doi.org/10.1073/pnas.91.24.11542.

(18) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91*, 11542–11546.
(19) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Rational Design of Aromatic Polyketide Natural Products by Recombinant Assembly of Enzymatic Subunits. *Nature* **1995**, *375* (6532), 549–554. https://doi.org/10.1038/375549a0.
(20) Shen, B.; Summers, R. G.; Wendt-Pienkowski, E.; Hutchinson, C. R. The Streptomyces Glaucescens tcmKL Polyketide Synthase and tcmN Polyketide Cyclase Genes Govern the Size and Shape of Aromatic Polyketides. *J Am Chem Soc* **1995**, *117*, 6811–6821.
(21) Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. Engineered Biosynthesis of Novel Polyketides: Properties of the whiE Aromatase/Cyclase. *Nat Biotechnol* **1996**, *14*, 335–338.
(22) Kramer, P. J. Rational Design and Engineered Biosynthesis of a Novel 18-Carbon Aromatic Polyketide. *J Am Chem Soc* **1997**, *119*, 635–639.
(23) Meurer, G. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4*, 433–443.
(24) McDaniel, R. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel 'Unnatural' Natural Products. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 1846–1851.
(25) Harvey, A. L. Natural Products in Drug Discovery. *Drug Discov. Today* **2008**, *13*, 894–901.
(26) Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A. Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale for Catalytic Reprogramming. *J Am Chem Soc* **2014**, *136*, 7348–7362.
(27) Connor, C. J. O.; Beckmann, H. S. G.; Spring, D. R. Diversity-Oriented Synthesis: Producing Chemical Tools for Dissecting Biology. *Chem Soc Rev* **2012**, *41*, 4444–4456.
(28) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for Bad Bugs: Confronting the Challenges of Antibacterial Discovery. *Nat Rev Drug Discov* **2007**, *6*, 29–40.
(29) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat Rev Drug Discov* **2015**, *14*, 111–129.
(30) Schloss, P. D.; Handelsman, J. Metagenomics for Studying Unculturable Microorganisms: Cutting the Gordian Knot. *Genome Biol.* **2005**, *6*, 229.
(31) Medema, H.; M. The Year 2020 in Natural Product Bioinformatics: An Overview of the Latest Tools and Databases. *Nat. Prod. Rep.* **2021**, *38*, 301–306.
(32) Blin, K. antiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation. *Nucleic Acids Res.* **2023**, *51*, 46–50.
(33) Blin, K.; Shaw, S.; Kautsar, S. A.; Medema, M. H.; Weber, T. The antiSMASH Database Version 3: Increased Taxonomic Coverage and New Query Features for Modular Enzymes. *Nucleic Acids Res.* **2021**, *49*, 639–643.
(34) Hannigan, G. D. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res* **2019**, *47*, 110–110.
(35) Ichikawa, N. DoBISCUIT: A Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2013**, *41*, 408–414.

(36) Kautsar, S. A. MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* **2020**, *48*, 454–458.
(37) Palaniappan, K. IMG-ABC v.5.0: An Update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res.* **2020**, *48*, 422–430.
(38) Bentley, S. D. Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2. *Nature* **2002**, *417*, 141–147.
(39) Nett, M.; Ikeda, H.; Moore, B. S. Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. *Nat Prod Rep* **2009**, *26*, 1362–1384.
(40) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating Natural Product Combinatorial Biosynthesis with Synthetic Biology. *Nat Chem Biol* **2015**, *11*, 649–659.
(41) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8), 509–523. https://doi.org/10.1038/nrmicro3496.
(42) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat Rev Drug Discov* **2021**, *20*, 200–216.
(43) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem* **2021**, *90*, 763–788.
(44) Xu, F. A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat Chem Biol* **2019**, *15*, 161–168.
(45) Herbst, D. A.; Townsend, C. A.; Maier, T. The Architectures of Iterative Type i PKS and FAS. *Nat Prod Rep* **2018**, *35*, 1046–1069.
(46) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
(47) Dutta, S. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510*, 512–517.
(48) Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. *Science* **2007**, *316*, 288–290.
(49) Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315–1322.
(50) Staunton, J.; J. Weissman, K. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416. https://doi.org/10.1039/A909079G.
(51) Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Nat Prod Rep* **2003**, *20*, 79–110.
(52) Ferrer, J.-L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Structure of Chalcone Synthase and the Molecular Basis of Plant Polyketide Biosynthesis. *Nat Struct Mol Biol* **1999**, *6*, 775–784.
(53) Bisang, C. A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* **1999**, *401*, 502–505.
(54) Jackowski, S.; Rock, C. O. Acetoacetyl-Acyl Carrier Protein Synthase, a Potential Regulator of Fatty Acid Biosynthesis in Bacteria. *J. Biol. Chem.* **1987**, *262*, 7927–7931.
(55) Moore, B. S.; Hertweck, C. Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units. *Nat Prod Rep* **2002**, *19*, 70–99.
(56) Crawford, J. M.; Dancy, B. C. R.; Hill, E. A.; Udwary, D. W.; Townsend, C. A. Identification of a Starter Unit Acyl-Carrier Protein Transacylase Domain in an Iterative Type I Polyketide Synthase. *Proc. Natl. Acad. Sci.* **2006**, *103* (45), 16728–16733. https://doi.org/10.1073/pnas.0604112103.

(57) Tsai, S.-C. (Sheryl). The Structural Enzymology of Iterative Aromatic Polyketide Synthases: A Critical Comparison with Fatty Acid Synthases. *Annu. Rev. Biochem.* **2018**, *87* (1), 503–531. https://doi.org/10.1146/annurev-biochem-063011-164509.
(58) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat Prod Rep* **2016**, *33*, 150–161.
(59) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **2009**, *26*, 90–114.
(60) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* **1998**, *37* (22), 8132–8138. https://doi.org/10.1021/bi980466i.
(61) Meadows, E. S.; Khosla, C. In Vitro Reconstitution and Analysis of the Chain Initiating Enzymes of the R1128 Polyketide Synthase. *Biochemistry* **2001**, *40*, 14855–14861.
(62) Winter, J. M. Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, *137*, 9885–9893.
(63) Cacho, R. A. Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J Am Chem Soc* **2015**, *137*, 15688–15691.
(64) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide b Synthase. *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
(65) Crawford, J. M. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320*, 243–246.
(66) Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709. https://doi.org/10.1021/ja0378759.
(67) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Dissection of the Catalytic Specificity of the Act Ketoreductase. *J. Am. Chem. Soc.* **1994**, *116*, 4166–4170.
(68) Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Evidence for Temporal, but Not Regiospecific, Control of Cyclization of an Aromatic Polyketide Precursor. *Chem. Biol.* **1994**, *1*, 205–210.
(69) Shen, Y. Ectopic Expression of the Minimal whiE Polyketide Synthase Generates a Library of Aromatic Polyketides of Diverse Sizes and Shapes. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 3622–3627.
(70) Auclair, K. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels–alder Reaction of a Substrate Analogue. *J Am Chem Soc* **2000**, *122*, 11519–11520.
(71) Kim, H. J.; Ruszczycky, M. W.; Choi, S.; Liu, Y.; Liu, H. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn a. *Nature* **2011**, *473*, 109–112.
(72) Kato, N. A New Enzyme Involved in the Control of the Stereochemistry in the Decalin Formation during Equisetin Biosynthesis. *Biochem. Biophys. Res. Commun.* **2015**, *460*, 210–215.
(73) Kasahara, K. Solanapyrone Synthase, a Possible Diels–Alderase and Iterative Type i Polyketide Synthase Encoded in a Biosynthetic Gene Cluster from Alternaria Solani. *ChemBioChem* **2010**, *11*, 1245–1252.
(74) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of Claisen Cyclase Domain in Fungal Polyketide Synthase WA, a Naphthopyrone Synthase of Aspergillus Nidulans. *Chem. Biol.* **2001**, *8*, 189–197.
(75) Crawford, J. M. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461*, 1139–1143.

(76) Ames, B. D. Crystal Structure and Functional Analysis of Tetracenomycin ARO/CYC: Implications for Cyclization Specificity of Aromatic Polyketides. *PNAS* **2008**, *105*, 5349–5354.
(77) Ames, B. D. Structural and Biochemical Characterization of ZhuI Aromatase/Cyclase from the R1128 Polyketide Pathway. *Biochemistry* **2011**, *50*, 8392–8406.
(78) Yu, T.-W. Engineered biosynthesis of novel polyketides from streptomyces spore pigment polyketide synthases. *J Am Chem Soc* **1998**, *120*, 7749–7759.
(79) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively. *J. Am. Chem. Soc.* **1994**, *116*, 10855–10859.
(80) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of Post-PKS Tailoring Steps through Combinatorial Biosynthesis. *Nat Prod Rep* **2002**, *19*, 542–580.
(81) Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. Cytochromes P450 for Natural Product Biosynthesis in Streptomyces: Sequence, Structure, and Function. *Nat Prod Rep* **2017**, *34*, 1141–1172.
(82) Chung, W.; Vanderwal, C. D. Stereoselective Halogenation in Natural Product Synthesis. *Angew. Chem. Int. Ed.* **2016**, *55*, 4396–4434.
(83) Singh, S., Jr; P., G. N.; Thorson, J. S. The Structural Biology of Enzymes Involved in Natural Product Glycosylation. *Nat Prod Rep* **2012**, *29*, 1201–1237.
(84) Podust, L. M.; Sherman, D. H. Diversity of P450 Enzymes in the Biosynthesis of Natural Products. *Nat Prod Rep* **2012**, *29*, 1251–1266.
(85) Carreras, C. W.; Pieper, R.; Khosla, C. Efficient Synthesis of Aromatic Polyketides in Vitro by the Actinorhodin Polyketide Synthase. *J Am Chem Soc* **1996**, *118*, 5158–5159.
(86) Shen, B.; Hutchinson, C. R. Deciphering the Mechanism for the Assembly of Aromatic Polyketides by a Bacterial Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1996**, *93*, 6600–6604.
(87) Harris, T. M.; Harris, C. M. Biomimetic Syntheses of Aromatic Polyketide Metabolites. *Pure Appl Chem* **1986**, *58*, 283–294.
(88) Sherman, D. h. Structure and Deduced Function of the Granaticin-Producing Polyketide Synthase Gene Cluster of Streptomyces Violaceoruber Tü22. *EMBO J.* **1989**, *8*, 2717–2725.
(89) Kim, E. S.; Cramer, K. D.; Shreve, A. L.; Sherman, D. H. Heterologous Expression of an Engineered Biosynthetic Pathway: Functional Dissection of Type II Polyketide Synthase Components in Streptomyces Species. *J Bacteriol* **1995**, *177*, 1202–1207.
(90) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11* (9), 888–893. https://doi.org/10.1038/nsmb808.
(91) Sherman, D. H.; Kim, E. S.; Bibb, M. J.; Hopwood, D. A. Functional Replacement of Genes for Individual Polyketide Synthase Components in Streptomyces Coelicolor A3(2) by Heterologous Genes from a Different Polyketide Pathway. *J. Bacteriol.* **1992**, *174*, 6184–6190.
(92) Kim, E. S.; Hopwood, D. A.; Sherman, D. H. Analysis of Type II Polyketide Beta-Ketoacyl Synthase Specificity in Streptomyces Coelicolor A3(2) by Trans Complementation of Actinorhodin Synthase Mutants. *J Bacteriol* **1994**, *176*, 1801–1804.
(93) Dreier, J.; Khosla, C. Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the β-Ketoacyl Synthase–chain Length Factor Heterodimer. *Biochemistry* **2000**, *39*, 2088–2095.

(94) Grammbitter, G. L. C. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. *J Am Chem Soc* **2019**, *141*, 16615–16623.
(95) Bräuer, A. Structural Snapshots of the Minimal PKS System Responsible for Octaketide Biosynthesis. *Nat Chem* **2020**, *12*, 755–763.
(96) Nicholson, T. P. First in Vitro Directed Biosynthesis of New Compounds by a Minimal Type II Polyketide Synthase: Evidence for the Mechanism of Chain Length Determination. *Chem Commun* **2003**, *686–687*. https://doi.org/10.1039/B300847A.
(97) Kim, Y.; Prestegard, J. H. Refinement of the NMR Structures for Acyl Carrier Protein with Scalar Coupling Data. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 377–385.
(98) Evans, S. E. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* **2008**, *9*, 2424–2432.
(99) Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein. *Biochemistry* **2003**, *42*, 4648–4657.
(100) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. Solution Structure of the Actinorhodin Polyketide Synthase Acyl Carrier Protein from Streptomyces Coelicolor A3(2),. *Biochemistry* **1997**, *36* (20), 6000–6008. https://doi.org/10.1021/bi970006+.
(101) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. Targeted Gene Replacements in a Streptomyces Polyketide Synthase Gene Cluster: Role for the Acyl Carrier Protein. *Mol. Microbiol.* **1992**, *6*, 3237–3249.
(102) Zhang, Y.-M.; Wu, B.; Zheng, J.; Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction*. *J. Biol. Chem.* **2003**, *278*, 52935–52943.
(103) Arthur, C. J. Structure and Malonyl CoA-ACP Transacylase Binding of Streptomyces Coelicolor Fatty Acid Synthase Acyl Carrier Protein. *ACS Chem Biol* **2009**, *4*, 625–636.
(104) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-Based Analysis of the Molecular Interactions between Acyltransferase and Acyl Carrier Protein in Vicenistatin Biosynthesis. *PNAS* **2016**, *113*, 1802–1807.
(105) Bartholow, T. G. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4*, 1–10.
(106) Cho, Y. I. Engineered Chimeras Unveil Swappable Modular Features of Fatty Acid and Polyketide Synthase Acyl Carrier Proteins. *Biochemistry* **2022**, *61*, 217–227.
(107) Khosla, C. Genetic Construction and Functional Analysis of Hybrid Polyketide Synthases Containing Heterologous Acyl Carrier Proteins. *J. Bacteriol.* **1993**, *175*, 2197–2204.
(108) Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C. Ketosynthases in the Initiation and Elongation Modules of Aromatic Polyketide Synthases Have Orthogonal Acyl Carrier Protein Specificity†, 2003. https://doi.org/10.1021/BI0341962.
(109) Das, A.; Khosla, C. In Vivo and in Vitro Analysis of the Hedamycin Polyketide Synthase. *Chem. Biol.* **2009**, *16*, 1197–1207.
(110) Koppisch, A. T.; Khosla, C. Structure-Based Mutagenesis of the Malonyl-CoA:Acyl Carrier Protein Transacylase from Streptomyces Coelicolor. *Biochemistry* **2003**, *42*, 11057–11064.
(111) Beltran-Alvarez, P.; Cox, R. J.; Crosby, J.; Simpson, T. J. Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* **2007**, *46*, 14672–14681.

(112) Arthur, C. J. Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* **2005**, *44*, 15414–15421.
(113) Kalaitzis, J. A. Policing Starter Unit Selection of the Enterocin Type II Polyketide Synthase by the Type II Thioesterase EncL. *Bioorg. Med. Chem.* **2011**, *19*, 6633–6638.
(114) Shepherd, M. D.; Kharel, M. K.; Zhu, L. L.; Lanen, S. G. van; Rohr, J. Delineating the Earliest Steps of Gilvocarcin Biosynthesis: Role of GilP and GilQ in Starter Unit Specificity. *Org Biomol Chem* **2010**, *8*, 3851–3856.
(115) Tang, Y.; Koppisch, A. T.; Khosla, C. The Acyltransferase Homologue from the Initiation Module of the R1128 Polyketide Synthase Is an Acyl-ACP Thioesterase That Edits Acetyl Primer Units. *Biochemistry* **2004**, *43* (29), 9546–9555. https://doi.org/10.1021/bi049157k.
(116) Keatinge-Clay, A. T. Catalysis, Specificity, and ACP Docking Site of Streptomyces Coelicolor Malonyl-CoA:ACP Transacylase. *Structure* **2003**, *11*, 147–154.
(117) Szu, P.-H. Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **2011**, *18*, 1021–1031.
(118) Bartel, P. L. Biosynthesis of Anthraquinones by Interspecies Cloning of Actinorhodin Biosynthesis Genes in Streptomycetes: Clarification of Actinorhodin Gene Functions. *J. Bacteriol.* **1990**, *172*, 4816–4826.
(119) Rajgarhia, V. B.; Strohl, W. R. Minimal streptomyces sp. Strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis. *J Bacteriol* **1997**, *179*, 2690–2696.
(120) Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C. Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKR): Molecular Basis of Stereo- and Regiospecificities. *Biochemistry* **2011**, *50*, 7426–7439.
(121) Piel, J. Cloning, Sequencing and Analysis of the Enterocin Biosynthesis Gene Cluster from the Marine Isolate 'Streptomyces Maritimus': Evidence for the Derailment of an Aromatic Polyketide Synthase. *Chem. Biol.* **2000**, *7*, 943–955.
(122) Price, A. C.; Zhang, Y. M.; Rock, C. O.; White, S. W. Structure of β-Ketoacyl-[Acyl Carrier Protein] Reductase from Escherichia Coli: Negative Cooperativity and Its Structural Basis. *Biochemistry* **2001**, *40*, 12772–12781.
(123) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43* (46), 14529–14538. https://doi.org/10.1021/bi048133a.
(124) Oppermann, U. Short-Chain Dehydrogenases/Reductases (SDR): The 2002 Update. *Chem. Biol. Interact.* **2003**, *143–144*, 247–253.
(125) Filling, C. Critical Residues for Structure and Catalysis in Short-Chain Dehydrogenases/Reductases*. *J. Biol. Chem.* **2002**, *277*, 25677–25684.
(126) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **1993**, *262* (5139), 1546–1550.
(127) Kalaitzis, J. A.; Moore, B. S. Heterologous Biosynthesis of Truncated Hexaketides Derived from the Actinorhodin Polyketide Synthase. *J. Nat. Prod.* **2004**, *67*, 1419–1422.
(128) Zawada, R. J.; Khosla, C. Heterologous Expression, Purification, Reconstitution and Kinetic Analysis of an Extended Type II Polyketide Synthase. *Chem. Biol.* **1999**, *6* (9), 607–615. https://doi.org/10.1016/s1074-5521(99)80112-4.
(129) Hadfield, A. T. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12*, 1865–1875.

(130) Kantola, J. Folding of the Polyketide Chain Is Not Dictated by Minimal Polyketide Synthase in the Biosynthesis of Mithramycin and Anthracycline. *Chem. Biol.* **1997**, *4*, 751–755.
(131) Kunnari, T. Incorrectly Folded Aromatic Polyketides from Polyketide Reductase Deficient Mutants. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2639–2642.
(132) Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C. Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase,. *Biochemistry* **2008**, *47* (7), 1837–1847. https://doi.org/10.1021/bi7016427.
(133) Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C. C. Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochemistry* **2011**, *50*, 4638–4649.
(134) Javidpour, P. The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chem. Biol.* **2013**, *20*, 1225–1234.
(135) Zhang, H. L. Mutactin, a Novel Polyketide from Streptomyces Coelicolor. Structure and Biosynthetic Relationship to Actinorhodin. *J Org Chem* **1990**, *55*, 1682–1684.
(136) Zhao, S. Molecular Basis for Polyketide Ketoreductase–Substrate Interactions. *Int. J. Mol. Sci.* **2020**, *21*, 7562.
(137) Kharel, M. K. Elucidation of post-PKS tailoring steps involved in landomycin biosynthesis. *Org Biomol Chem* **2012**, *10*, 4256–4265.
(138) Mayer, A. LanV, a Bifunctional Enzyme: Aromatase and Ketoreductase during Landomycin a Biosynthesis. *ChemBioChem* **2005**, *6*, 2312–2315.
(139) Dickens, M. L.; Ye, J.; Strohl, W. R. Cloning, Sequencing, and Analysis of Aklaviketone Reductase from Streptomyces Sp. Strain C5. *J. Bacteriol.* **1996**, *178*, 3384–3388.
(140) Lackner, G. Biosynthesis of Pentangular Polyphenols: Deductions from the Benastatin and Griseorhodin Pathways. *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
(141) Torkkell, S. The Entire Nogalamycin Biosynthetic Gene Cluster of Streptomyces Nogalater: Characterization of a 20-Kb DNA Region and Generation of Hybrid Structures. *Mol Gen Genomics* **2001**, *266*, 276–288.
(142) Paananen, P. Structural and Functional Analysis of Angucycline C-6 Ketoreductase LanV Involved in Landomycin Biosynthesis. *Biochemistry* **2013**, *52*, 5304–5314.
(143) Patrikainen, P. Structure-Based Engineering of Angucyclinone 6-Ketoreductases. *Chem. Biol.* **2014**, *21*, 1381–1391.
(144) Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C. Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (12), 3421–3430. https://doi.org/10.1021/acschembio.6b00658.
(145) Xiao, X. Functional and Structural Insights into a Novel Promiscuous Ketoreductase of the Lugdunomycin Biosynthetic Pathway. *ACS Chem Biol* **2020**, *15*, 2529–2538.
(146) Kallio, P. Flavoprotein Hydroxylase PgaE Catalyzes Two Consecutive Oxygen-Dependent Tailoring Reactions in Angucycline Biosynthesis. *Biochemistry* **2011**, *50*, 5535–5543.
(147) Patrikainen, P. Tailoring Enzymes Involved in the Biosynthesis of Angucyclines Contain Latent Context-Dependent Catalytic Activities. *Chem. Biol.* **2012**, *19*, 647–655.
(148) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **2003**, *4*, 654–657.
(149) Moreno, G. O. *Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type i and Type II Polyketide Synthases Using Atom-Replaced Probes*; UC Irvine, 2018.

(150) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(151) Li 然 李嫣, Y.; Xu 伟徐, W.; Tang 奕唐, Y. Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains*. *J. Biol. Chem.* **2010**, *285* (30), 22764–22773. https://doi.org/10.1074/jbc.M110.128504.
(152) Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C. A. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461* (7267), 1139–1143. https://doi.org/10.1038/nature08475.
(153) Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. *Proc. Natl. Acad. Sci.* **2017**, *114* (21), E4142–E4148. https://doi.org/10.1073/pnas.1609001114.
(154) Liu, L.; Zhang, Z.; Shao, C.-L.; Wang, J.-L.; Bai, H.; Wang, C.-Y. Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Sci. Rep.* **2015**, *5* (1), 10463. https://doi.org/10.1038/srep10463.
(155) Feng, Y.; Yang, X.; Ji, H.; Deng, Z.; Lin, S.; Zheng, J. The Streptomyces Viridochromogenes Product Template Domain Represents an Evolutionary Intermediate between Dehydratase and Aldol Cyclase of Type I Polyketide Synthases. *Commun. Biol.* **2022**, *5* (1), 508. https://doi.org/10.1038/s42003-022-03477-8.
(156) Pan, H. Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. *Structure* **2002**, *10*, 1559–1568.
(157) Thompson, T. B.; Katayama, K.; Watanabe, K.; Hutchinson, C. R.; Rayment, I. Structural and Functional Analysis of Tetracenomycin F2 Cyclase from Streptomyces Glaucescens: A TYPE II POLYKETIDE CYCLASE*. *J. Biol. Chem.* **2004**, *279*, 37956–37963.
(158) Sultana, A. Structure of the Polyketide Cyclase SnoaL Reveals a Novel Mechanism for Enzymatic Aldol Condensation. *EMBO J.* **2004**, *23*, 1911–1921.
(159) Kallio, P.; Sultana, A.; Niemi, J.; Mäntsälä, P.; Schneider, G. Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity. *J. Mol. Biol.* **2006**, *357*, 210–220.
(160) Hutchinson, C. R. Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem. Rev.* **1997**, *97* (7), 2525–2536. https://doi.org/10.1021/cr960022x.
(161) Ahuja, M.; Chiang, Y.-M.; Chang, S.-L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H.-C.; Yeh, H.-H.; Oakley, B. R.; Wang, C. C. C. Illuminating the Diversity of Aromatic Polyketide Synthases in Aspergillus Nidulans. *J. Am. Chem. Soc.* **2012**, *134* (19), 8212–8221. https://doi.org/10.1021/ja3016395.
(162) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97* (7), 2465–2498. https://doi.org/10.1021/cr960034i.
(163) Shen, B. Polyketide Biosynthesis beyond the Type I, II and III Polyketide Synthase Paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6.
(164) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639. https://doi.org/10.1021/ar8002249.
(165) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A.

Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130. https://doi.org/10.1126/science.ade2574.
(166) Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. Structural Basis for Selectivity in a Highly Reducing Type II Polyketide Synthase. *Nat. Chem. Biol.* **2020**, *16* (7), 776–782. https://doi.org/10.1038/s41589-020-0530-0.
(167) Yadav, G.; Gokhale, R. S.; Mohanty, D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J. Mol. Biol.* **2003**, *328* (2), 335–363. https://doi.org/10.1016/S0022-2836(03)00232-8.
(168) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci.* **2021**, *118* (15), e2016239118. https://doi.org/10.1073/pnas.2016239118.
(169) Capela, J.; Zimmermann-Kogadeeva, M.; Dijk, A. D. J. van; de Ridder, D.; Dias, O.; Rocha, M. Comparative Assessment of Protein Large Language Models for Enzyme Commission Number Prediction. *BMC Bioinformatics* **2025**, *26* (1), 68. https://doi.org/10.1186/s12859-025-06081-9.
(170) Jacques, M.-A.; Durand, K.; Orgeur, G.; Balidas, S.; Fricot, C.; Bonneau, S.; Quillévéré, A.; Audusseau, C.; Olivier, V.; Grimault, V.; Mathis, R. Phylogenetic Analysis and Polyphasic Characterization of Clavibacter Michiganensis Strains Isolated from Tomato Seeds Reveal That Nonpathogenic Strains Are Distinct from C. Michiganensis Subsp. Michiganensis. *Appl. Environ. Microbiol.* **2012**, *78* (23), 8388–8402. https://doi.org/10.1128/AEM.02158-12.
(171) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat. Chem. Biol.* **2019**, *15* (7), 669–671. https://doi.org/10.1038/s41589-019-0301-y.
(172) Ellis, B. D.; Milligan, J. C.; White, A. R.; Duong, V.; Altman, P. X.; Mohammed, L. Y.; Crump, M. P.; Crosby, J.; Luo, R.; Vanderwal, C. D.; Tsai, S.-C. An Oxetane-Based Polyketide Surrogate To Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140* (15), 4961–4964. https://doi.org/10.1021/jacs.7b11793.
(173) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.
(174) Zdouc, M. M.; Blin, K.; Louwen, N. L. L.; Navarro, J.; Loureiro, C.; Bader, C. D.; Bailey, C. B.; Barra, L.; Booth, T. J.; Bozhüyük, K. A. J.; Cediel-Becerra, J. D. D.; Charlop-Powers, Z.; Chevrette, M. G.; Chooi, Y. H.; D'Agostino, P. M.; de Rond, T.; Del Pup, E.; Duncan, K. R.; Gu, W.; Hanif, N.; Helfrich, E. J. N.; Jenner, M.; Katsuyama, Y.; Korenskaia, A.; Krug, D.; Libis, V.; Lund, G. A.; Mantri, S.; Morgan, K. D.; Owen, C.; Phan, C.-S.; Philmus, B.; Reitz, Z. L.; Robinson, S. L.; Singh, K. S.; Teufel, R.; Tong, Y.; Tugizimana, F.; Ulanova, D.;

Winter, J. M.; Aguilar, C.; Akiyama, D. Y.; Al-Salihi, S. A. A.; Alanjary, M.; Alberti, F.; Aleti, G.; Alharthi, S. A.; Rojo, M. Y. A.; Arishi, A. A.; Augustijn, H. E.; Avalon, N. E.; Avelar-Rivas, J. A.; Axt, K. K.; Barbieri, H. B.; Barbosa, J. C. J.; Barboza Segato, L. G.; Barrett, S. E.; Baunach, M.; Beemelmanns, C.; Beqaj, D.; Berger, T.; Bernaldo-Agüero, J.; Bettenbühl, S. M.; Bielinski, V. A.; Biermann, F.; Borges, R. M.; Borriss, R.; Breitenbach, M.; Bretscher, K. M.; Brigham, M. W.; Buedenbender, L.; Bulcock, B. W.; Cano-Prieto, C.; Capela, J.; Carrion, V. J.; Carter, R. S.; Castelo-Branco, R.; Castro-Falcón, G.; Chagas, F. O.; Charria-Girón, E.; Chaudhri, A. A.; Chaudhry, V.; Choi, H.; Choi, Y.; Choupannejad, R.; Chromy, J.; Donahey, M. S. C.; Collemare, J.; Connolly, J. A.; Creamer, K. E.; Crüsemann, M.; Cruz, A. A.; Cumsille, A.; Dallery, J.-F.; Damas-Ramos, L. C.; Damiani, T.; de Kruijff, M.; Martín, B. D.; Sala, G. D.; Dillen, J.; Doering, D. T.; Dommaraju, S. R.; Durusu, S.; Egbert, S.; Ellerhorst, M.; Faussurier, B.; Fetter, A.; Feuermann, M.; Fewer, D. P.; Foldi, J.; Frediansyah, A.; Garza, E. A.; Gavriilidou, A.; Gentile, A.; Gerke, J.; Gerstmans, H.; Gomez-Escribano, J. P.; González-Salazar, L. A.; Grayson, N. E.; Greco, C.; Gomez, J. E. G.; Guerra, S.; Flores, S. G.; Gurevich, A.; Gutiérrez-García, K.; Hart, L.; Haslinger, K.; He, B.; Hebra, T.; Hemmann, J. L.; Hindra, H.; Höing, L.; Holland, D. C.; Holme, J. E.; Horch, T.; Hrab, P.; Hu, J.; Huynh, T.-H.; Hwang, J.-Y.; Iacovelli, R.; Iftime, D.; Iorio, M.; Jayachandran, S.; Jeong, E.; Jing, J.; Jung, J. J.; Kakumu, Y.; Kalkreuter, E.; Kang, K. B.; Kang, S.; Kim, W.; Kim, G. J.; Kim, H.; Kim, H. U.; Klapper, M.; Koetsier, R. A.; Kollten, C.; Kovács, Á. T.; Kriukova, Y.; Kubach, N.; Kunjapur, A. M.; Kushnareva, A. K.; Kust, A.; Lamber, J.; Larralde, M.; Larsen, N. J.; Launay, A. P.; Le, N.-T.-H.; Lebeer, S.; Lee, B. T.; Lee, K.; Lev, K. L.; Li, S.-M.; Li, Y.-X.; Licona-Cassani, C.; Lien, A.; Liu, J.; Lopez, J. A. V.; Machushynets, N. V.; Macias, M. I.; Mahmud, T.; Maleckis, M.; Martinez-Martinez, A. M.; Mast, Y.; Maximo, M. F.; McBride, C. M.; McLellan, R. M.; Bhatt, K. M.; Melkonian, C.; Merrild, A.; Metsä-Ketelä, M.; Mitchell, D. A.; Müller, A. V.; Nguyen, G.-S.; Nguyen, H. T.; Niedermeyer, T. H. J.; O'Hare, J. H.; Ossowicki, A.; Ostash, B. O.; Otani, H.; Padva, L.; Paliyal, S.; Pan, X.; Panghal, M.; Parade, D. S.; Park, J.; Parra, J.; Rubio, M. P.; Pham, H. T.; Pidot, S. J.; Piel, J.; Pourmohsenin, B.; Rakhmanov, M.; Ramesh, S.; Rasmussen, M. H.; Rego, A.; Reher, R.; Rice, A. J.; Rigolet, A.; Romero-Otero, A.; Rosas-Becerra, L. R.; Rosiles, P. Y.; Rutz, A.; Ryu, B.; Sahadeo, L.-A.; Saldanha, M.; Salvi, L.; Sánchez-Carvajal, E.; Santos-Medellin, C.; Sbaraini, N.; Schoellhorn, S. M.; Schumm, C.; Sehnal, L.; Selem, N.; Shah, A. D.; Shishido, T. K.; Sieber, S.; Silviani, V.; Singh, G.; Singh, H.; Sokolova, N.; Sonnenschein, E. C.; Sosio, M.; Sowa, S. T.; Steffen, K.; Stegmann, E.; Streiff, A. B.; Strüder, A.; Surup, F.; Svenningsen, T.; Sweeney, D.; Szenei, J.; Tagirdzhanov, A.; Tan, B.; Tarnowski, M. J.; Terlouw, B. R.; Rey, T.; Thome, N. U.; Torres Ortega, L. R.; Tørring, T.; Trindade, M.; Truman, A. W.; Tvilum, M.; Udwary, D. W.; Ulbricht, C.; Vader, L.; van Wezel, G. P.; Walmsley, M.; Warnasinghe, R.; Weddeling, H. G.; Weir, A. N. M.; Williams, K.; Williams, S. E.; Witte, T. E.; Rocca, S. M. W.; Yamada, K.; Yang, D.; Yang, D.; Yu, J.; Zhou, Z.; Ziemert, N.; Zimmer, L.; Zimmermann, A.; Zimmermann, C.; van der Hooft, J. J. J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 4.0: Advancing Biosynthetic Gene Cluster Curation through Global Collaboration. *Nucleic Acids Res.* **2025**, *53* (D1), D678–D690. https://doi.org/10.1093/nar/gkae1115.

(175) Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S. J.; Webb, C.; Wu, D.; Varghese, N. J.; Reddy, T. B. K.; Mukherjee, S.; Ovchinnikova, G.; Nolan, M.; Seshadri, R.; Roux, S.; Visel, A.; Woyke, T.; Eloe-Fadrosh, E. A.; Kyrpides, N. C.; Ivanova, N. N. The IMG/M Data Management and Analysis System

v.7: Content Updates and New Features. *Nucleic Acids Res.* **2023**, *51* (D1), D723–D732. https://doi.org/10.1093/nar/gkac976.
(176) Fernández-Moreno, M. A.; Martínez, E.; Caballero, J. L.; Ichinose, K.; Hopwood, D. A.; Malpartida, F. DNA Sequence and Functions of the actVI Region of the Actinorhodin Biosynthetic Gene Cluster of Streptomyces Coelicolor A3(2). *J. Biol. Chem.* **1994**, *269* (40), 24854–24863. https://doi.org/10.1016/S0021-9258(17)31469-2.
(177) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Richard Hutchinson, C. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4* (6), 433–443. https://doi.org/10.1016/S1074-5521(97)90195-2.
(178) Moreno, G. O. Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type I and Type II Polyketide Synthases Using Atom-Replaced Probes, UC Irvine, 2018. https://escholarship.org/uc/item/1tx0518k.
(179) Serapian, S. A.; Crosby, J.; Crump, M. P.; van der Kamp, M. W. Path to Actinorhodin: Regio- and Stereoselective Ketone Reduction by a Type II Polyketide Ketoreductase Revealed in Atomistic Detail. *JACS Au* **2022**, *2* (4), 972–984. https://doi.org/10.1021/jacsau.2c00086.
(180) Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941–953. https://doi.org/10.1016/j.jmb.2008.09.084.
(181) Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12* (10), 1865–1875. https://doi.org/10.1016/j.str.2004.08.002.
(182) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat. Prod. Rep.* **2016**, *33* (2), 203–230. https://doi.org/10.1039/C5NP00109A.
(183) Chooi, Y.-H.; Tang, Y. Navigating the Fungal Polyketide Chemical Space: From Genes to Molecules. *J. Org. Chem.* **2012**, *77* (22), 9933–9953. https://doi.org/10.1021/jo301592k.
(184) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(185) Barajas, J. F.; Finzel, K.; Valentic, T. R.; Shakya, G.; Gamarra, N.; Martinez, D.; Meier, J. L.; Vagstad, A. L.; Newman, A. G.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Structural and Biochemical Analysis of Protein–Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew. Chem. Int. Ed.* **2016**, *55* (42), 13005–13009. https://doi.org/10.1002/anie.201605401.
(186) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(187) Barajas, J. F. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. In *Proceedings of the National Academy of Sciences*; 2017; Vol. 114, pp 4142–4148.
(188) Milligan, J. C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat Chem Biol* **2019**, *15*, 669–671.

(189) Ellis, B. D. An Oxetane-Based Polyketide Surrogate to Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140*, 4961–4964.
(190) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43*, 14529–14538.

Chapter 5

Conclusions and future directions

5.1 Conclusions

Polyketide synthases (PKSs) exhibit three major architectural classes—type I, type II, and type III—distinguished by their organizational and mechanistic features. Type I PKSs are multifunctional megasynthases composed of covalently linked catalytic domains and subdivided into modular and iterative forms, with modular systems employing each domain once per extension cycle and iterative systems reusing a single module multiple times[45–49]. Type II PKSs are dissociated systems of mono- or didomain enzymes that assemble transiently during biosynthesis[65], whereas type III PKSs are homodimeric ketosynthases capable of priming, chain extension, and cyclization within a single active site[51,52]. Polyketide biosynthesis begins with chain initiation via a decarboxylative Claisen condensation, typically between acetate and malonyl-derived extender units[53–59], with initiation strategies varying among PKS types[9,53,56,60–62,132]. Following initiation, core reactions of chain extension, β-keto reduction, aromatization, and cyclization generate structural diversity, with reducing and nonreducing type I iterative PKSs, type I modular PKSs, and type II PKSs employing distinct mechanisms[26,46,63–71,76–79,152]. Cyclization fidelity is enforced by product template (PT) domains in fungal nonreducing type I iterative PKSs[65,151,152,152,154,155,187] and by discrete aromatase/cyclase enzymes in type II systems[20,65,76–79], while tailoring reactions such as glycosylation, methylation, halogenation, and oxidative modifications further diversify polyketide scaffolds[80–84]. Structural and biochemical studies of type II minimal PKS components, including KS/CLF, ACP, and MAT, along with C9-reducing and tailoring KRs, reveal that chain-length determination, stereochemistry, and substrate specificity are governed by tightly integrated structural,

electrostatic, and kinetic parameters[66,85–117,118–149]. Similarly, PT domain phylogenetics and structural analyses demonstrate that regioselective cyclization arises from precise active-site geometry, controlling both folding and product architecture[151–155]. Together, these studies highlight that the mechanistic logic of PKS elongation, reduction, and cyclization underpins the remarkable structural diversity of polyketides and provides a foundation for future predictive and engineering efforts.

Despite the tremendous efforts to identify genes and several crystal structures available for different type II PKS systems, there is still a serious knowledge gap of how chain-length control is molecularly encoded within KS/CLF complexes, as existing studies have shown that although the CLF sets the elongation limit[50,57] and the KS/CLF tunnel influences chain accommodation[17,109], the specific sequence and structural features that program control carbon-chain length remain unclear. Mutational studies across KS/CLF systems have produced only modest or inconsistent shifts in chain length, demonstrating that programmed elongation arises from a complex interplay of steric constraints, electrostatics, acyl-chain positioning, and ACP–KS dynamics rather than any single residue or motif, and this challenge is intensified by the small number of KS/CLF systems with experimentally validated chain lengths relative to the vast genomic diversity of type II PKSs now available. A second major knowledge gap lies in the inability to predict or engineer ACP–KR compatibility, as decades of combinatorial biosynthesis have shown that even structurally similar ACPs and KRs are not reliably interchangeable[3–6], exemplified by the persistent inactivity of DpsKR outside its native doxorubicin context[23,160]. The transient and dynamic nature of ACP–KR interactions has prevented the establishment of quantitative rules governing productive docking, severely limiting rational pathway engineering and constraining modular swapping strategies. Together, these unresolved problems—deciphering how KS/CLF complexes encode chain-

length determination and defining the molecular logic governing ACP–KR specificity—constitute the core mechanistic knowledge gaps that this thesis aims to address.

In this thesis, Chapter 2 establishes that sequence-based machine learning models, even at their most advanced, capture only a fraction of the determinants governing KS/CLF chain-length specificity, highlighting the necessity of integrating structural and biochemical information for accurate functional prediction. Chapter 3 demonstrates that KR activity and ACP compatibility are dictated by cooperative structural features—particularly the KR lid helices—that shape protein–protein interfaces and govern productive substrate positioning. Chapter 4 reveals that PT-catalyzed cyclization specificity in fungal NR-PKSs arises from coordinated variations in catalytic geometry, pocket architecture, and ACP-mediated substrate delivery, providing a unified framework for predicting aromatic scaffold formation. Together, conclusions of these chapters show that the logic of polyketide biosynthesis is encoded not by isolated motifs but by distributed, dynamically coordinated structural features, and they collectively highlight the need for integrated informatic, structural, and biochemical approaches to decipher and engineer PKS enzymatic specificity.

5.1.1 Machine Learning–Driven Exploration of KS/CLF Chain-Length Specificity: Insights and Limitations from Protein Language Models

Chapter 2 addresses one of the central unresolved questions in aromatic polyketide biosynthesis: how type II KS/CLF complexes encode and enforce programmed chain-length specificity. Despite decades of biochemical and structural work showing that elongation limits arise from subtle geometric features within the KS–CLF tunnel[66,90,166], translating primary sequence into predictable chain-length output remains challenging. By applying machine-learning methods ranging from fully connected networks to transformer architectures and

finally ESM-2 embeddings[165,168], this chapter demonstrates both the promise and the limitations of sequence-based inference. Although progressively more expressive models improved performance—consistent with reports that transformer embeddings capture evolutionarily and structurally relevant features[167,169]—accuracy remained insufficient for confident annotation, reflecting the exceptionally high conservation of KS/CLF sequences and the minimal sequence differences that underlie functional divergence. Unsupervised clustering of >2,300 additional unlabeled KS/CLFs, even after projection into a PLS-derived latent space, likewise failed to resolve chain-length groups, reinforcing findings from evolutionary studies showing that KS/CLFs often differ at only one or two residues despite functional differences[126,164,170].

A major recurring theme in this chapter is the constraint imposed by limited high-quality training data. Because poly-β-ketone intermediates are chemically unstable and prone to spontaneous cyclization[188], experimental chain-length determination has historically relied on imperfect substrate analogs such as diketide SNACs or isoxazole probes[189]. Although recent oxetane-based polyketide isosteres offer more faithful stabilization[18], these tools have not yet enabled large-scale annotation. As a result, datasets remain small and uneven, making it difficult for machine-learning models to detect the faint sequence signals that contribute to tunnel shape, steric gating, and ACP-mediated chain positioning.

Together, these findings emphasize that sequence alone captures only a fraction of the determinants governing KS/CLF specificity, and that predictive frameworks must integrate structural modeling, coevolutionary analysis, and experimental biochemistry. Advances in structural prediction[173], along with emerging cryo-EM studies of PKS complexes present promising opportunities to incorporate tunnel geometries, ACP–KS–CLF interfaces, and stabilized intermediates into future datasets. Ultimately, the insights from Chapter 2 highlight both the constraints and the potential of informatics-driven approaches, offering a foundation

for next-generation models that more effectively bridge sequence to function and support rational engineering of chain-length control in aromatic polyketide biosynthesis[66,163].

5.1.2 Lid Helices and ACP Identity as Determinants of Ketoreductase Specificity in Type II Polyketide Synthases

Chapter 3 investigates the longstanding question of how type II ketoreductases (KRs) achieve substrate specificity and regiospecific reduction within minimal PKS systems, revealing that catalytic outcome is governed not solely by active-site chemistry but by structurally encoded protein–protein interactions mediated through the KR lid helices and the ACP surface. Through in vitro reconstitution assays, this chapter demonstrates that WT DpsKR is unable to perform C9 reduction in the Act minimal PKS, producing only unreduced products, whereas replacing the DpsKR lid with that of ActKR restores full reductive activity. This striking functional rescue, reinforced by ACP-swapping experiments, establishes that both lid identity and ACP identity are critical determinants of productive catalysis. Structural modeling provides a mechanistic foundation for these observations by showing that the lid region (residues 194–221) dictates ACP docking orientation and interaction networks, with the DpsKR–ActKR lid chimera engaging DpsACP through an expanded interface that closely mirrors the natural ActACP–ActKR complex, whereas WT DpsKR binds DpsACP in a restricted and nonproductive orientation.

These results resonate with integrative models of ActKR from Serapian et al.[179], in which productive ACP binding involves interaction with the KR arginine patch (R38/R65/R93) and α3 helix positioning that guides PPant entry into the active-site channel. The DpsKR–ActKR chimera reproduces these critical contacts, while WT DpsKR does not—linking activity directly to lid-mediated interface compatibility. This chapter further contextualizes these

findings within prior structural work on type II KRs[129,180,190], highlighting that while catalytic residues and global folds are conserved, ACP binding has remained a major unexplained contributor to regiospecificity. The data here provide a structural rationale for the frequent failure of KR swaps in combinatorial biosynthesis[182]: mismatched lid–ACP interfaces prevent productive reduction even when catalytic motifs are preserved.

More broadly, the conclusions of Chapter 3 underscore the central role of protein–protein interactions in type II PKSs, where discrete enzymes must form precise and transient interactions with ACP to enable functional substrate delivery. The collective evidence demonstrates that specificity arises from coordinated structural features—including lid-helix architecture, conserved KR docking motifs, and ACP surface chemistry—rather than isolated catalytic residues. As such, engineered modifications that do not account for lid–ACP complementarity will likely fail to produce desired outcomes. Future structural studies, including co-crystallization or cryo-EM of ACP–KR complexes, will be essential to validate these predicted interfaces and guide rational engineering. Ultimately, the insights from this chapter establish the lid helices as key molecular determinants of KR–ACP compatibility and lay the foundation for reprogramming type II PKS pathways toward the generation of novel anthracycline derivatives with improved therapeutic potential.

5.1.3 Structural Informatics of Product Template Domain Regioselectivity in Fungal NR-PKSs

Chapter 4 examined how fungal product template (PT) domains encode cyclization regioselectivity in non-reducing polyketide synthases (NR-PKSs). By integrating AlphaFold3 structural models with systematic docking, statistical analyses, and ACP–PT complex simulations, this chapter established a unified mechanistic framework that explains how PT

architecture programs C2–C7, C4–C9, and C6–C11 aldol cyclizations. Three major conclusions emerged.

Firstly, the spatial organization of the His–Asp catalytic dyad is a principal determinant of cyclization outcome, consistent with earlier structural work by Liu et al.[154] PTs that catalyze C2–C7 cyclization contain compact, bottom-positioned dyads that enforce early substrate folding, whereas PTs responsible for C4–C9 and C6–C11 reactions feature dyads located deeper within more elongated cavities, allowing later folding events. The correspondence between our structural predictions and previous crystallographic analyses highlights dyad placement as a conserved architectural pivot controlling the timing of aldol condensation.

Secondly, pocket architecture, including volume, contour, and polarity—provides an additional layer of substrate and starter-unit specificity. Variations in cavity depth and shape correlate strongly with chain-length preferences observed in docking, in agreement with earlier analyses of PT pocket dimensions.[154] Hydrophobic patches such as the previously described "hexyl-binding" region help explain the uniformly favorable docking energies of hexaketide intermediates in certain clades. Finger-like loop orientation also emerged as a consistent structural discriminator: outward-facing loops in later-folding PTs (C4–C9 and C6–C11) expand the pocket entrance and facilitate deeper substrate penetration, whereas compressed loops in C2–C7 PTs create a narrow channel that favors early folding trajectories.

Thirdly, the geometry of the ACP–PT interface strongly influences effective substrate delivery, extending insights from the ACP–PT interaction study of Barajas et al.[7] Even when the intrinsic pocket is large, misalignment between ACP and PT can position the phosphopantetheinyl-tethered substrate too far from the catalytic dyad, effectively shortening the chain length available for productive cyclization. This mechanism helps explain why some PT clades (e.g., Clade I) exhibit narrow substrate preferences despite possessing deep cavities.

These results show that regioselectivity is not dictated solely by static pocket architecture but also by dynamic interdomain interactions that modulate substrate presentation.

Together, the findings from Chapter 4 demonstrate that regioselectivity in fungal PT domains emerges from the integrated effects of catalytic dyad geometry, cavity architecture, loop-mediated gating, and ACP-guided substrate delivery. Structural variation across PT clades reflects evolutionary tuning of key cavity-lining residues, small loop insertions, and dyad positioning—modifications that diversify cyclization modes while retaining a conserved catalytic core. The principles defined in this chapter establish a predictive framework for inferring cyclization specificity from PT sequence or structure and provide a foundation for future engineering of PT domains with redesigned regioselectivity.

5.2 References

(1) Cragg, G. M.; Newman, D. J. Biodiversity: A Continuing Source of Novel Drug Leads. *Pure Appl. Chem.* **2005**, *77*, 7–24.

(2) Buenz, E. J.; Verpoorte, R.; Bauer, B. A. The Ethnopharmacologic Contribution to Bioprospecting Natural Products. *Annu. Rev. Pharmacol. Toxicol.* **2018**, *58*, 509–530.

(3) Gaynes, R. The Discovery of Penicillin—New Insights after More than 75 Years of Clinical Use. *Emerg Infect Dis* **2017**, *23*, 849–853.

(4) Beutler, J. A. Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol.* **2019**, *86, e67*.

(5) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176.

(6) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336.

(7) Hopwood, D. A. Production of 'hybrid' Antibiotics by Genetic Engineering. *Nature* **1985**, *314*, 642–644.

(8) Malpartida, F.; Hopwood, D. A. Molecular Cloning of the Whole Biosynthetic Pathway of a Streptomyces Antibiotic and Its Expression in a Heterologous Host. *Nature* **1984**, *309* (5967), 462–464. https://doi.org/10.1038/309462a0.

(9) Bibb, M. J.; Sherman, D. H.; Ōmura, S.; Hopwood, D. A. Cloning, Sequencing and Deduced Functions of a Cluster of Streptomyces Genes Probably Encoding Biosynthesis of the Polyketide Antibiotic Frenolicin. *Gene* **1994**, *142*, 31–39.

(10) Han, L.; Yang, K.; Ramalingam, E.; Mosher, R. H.; Vining, L. C. Cloning and characterization of polyketide synthase genes for jadomycin b biosynthesis in streptomyces venezuelaeISP5230. *Microbiology* **1994**, *140*, 3379–3389.

(11) Motamedi, H.; Hutchinson, C. R. Cloning and Heterologous Expression of a Gene Cluster for the Biosynthesis of Tetracenomycin C, the Anthracycline Antitumor Antibiotic of Streptomyces Glaucescens. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4445–4449.

(12) Kim, E.-S.; Bibb, M. J.; Butler, M. J.; Hopwood, D. A.; Sherman, D. H. Sequences of the Oxytetracycline Polyketide Synthase-Encoding Otc Genes from Streptomyces Rimosus. *Gene* **1994**, *141* (1), 141–142. https://doi.org/10.1016/0378-1119(94)90144-9.

(13) Grimm, A.; Madduri, K.; Ali, A.; Hutchinson, C. R. Characterization of the streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **1994**, *151*, 1–10.

(14) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Manipulation and Analysis of an Aromatic Polyketide Synthase with Unproved Catalytic Specificities. *J. Am. Chem. Soc.* **1993**, *115*, 11671–11675.

(15) Summers, R. G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C. R. The tcmVI Region of the Tetracenomycin c Biosynthetic Gene Cluster of Streptomyces Glaucescens Encodes the Tetracenomycin F1 Monooxygenase, Tetracenomycin F2 Cyclase, and, Most Likely, a Second Cyclase. *J. Bacteriol.* **1993**, *175*, 7571–7580.

(16) Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Stereochemical Course of Two Reactions Catalyzed by a Polyketide Synthase. *Biochemistry* **1994**, *33*, 9321–9326.

(17) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic

Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91* (24), 11542–11546. https://doi.org/10.1073/pnas.91.24.11542.
(18) McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1994**, *91*, 11542–11546.
(19) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Rational Design of Aromatic Polyketide Natural Products by Recombinant Assembly of Enzymatic Subunits. *Nature* **1995**, *375* (6532), 549–554. https://doi.org/10.1038/375549a0.
(20) Shen, B.; Summers, R. G.; Wendt-Pienkowski, E.; Hutchinson, C. R. The Streptomyces Glaucescens tcmKL Polyketide Synthase and tcmN Polyketide Cyclase Genes Govern the Size and Shape of Aromatic Polyketides. *J Am Chem Soc* **1995**, *117*, 6811–6821.
(21) Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. Engineered Biosynthesis of Novel Polyketides: Properties of the whiE Aromatase/Cyclase. *Nat Biotechnol* **1996**, *14*, 335–338.
(22) Kramer, P. J. Rational Design and Engineered Biosynthesis of a Novel 18-Carbon Aromatic Polyketide. *J Am Chem Soc* **1997**, *119*, 635–639.
(23) Meurer, G. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4*, 433–443.
(24) McDaniel, R. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel 'Unnatural' Natural Products. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 1846–1851.
(25) Harvey, A. L. Natural Products in Drug Discovery. *Drug Discov. Today* **2008**, *13*, 894–901.
(26) Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A. Systematic Domain Swaps of Iterative, Nonreducing Polyketide Synthases Provide a Mechanistic Understanding and Rationale for Catalytic Reprogramming. *J Am Chem Soc* **2014**, *136*, 7348–7362.
(27) Connor, C. J. O.; Beckmann, H. S. G.; Spring, D. R. Diversity-Oriented Synthesis: Producing Chemical Tools for Dissecting Biology. *Chem Soc Rev* **2012**, *41*, 4444–4456.
(28) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. Drugs for Bad Bugs: Confronting the Challenges of Antibacterial Discovery. *Nat Rev Drug Discov* **2007**, *6*, 29–40.
(29) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat Rev Drug Discov* **2015**, *14*, 111–129.
(30) Schloss, P. D.; Handelsman, J. Metagenomics for Studying Unculturable Microorganisms: Cutting the Gordian Knot. *Genome Biol.* **2005**, *6*, 229.
(31) Medema, H.; M. The Year 2020 in Natural Product Bioinformatics: An Overview of the Latest Tools and Databases. *Nat. Prod. Rep.* **2021**, *38*, 301–306.
(32) Blin, K. antiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation. *Nucleic Acids Res.* **2023**, *51*, 46–50.
(33) Blin, K.; Shaw, S.; Kautsar, S. A.; Medema, M. H.; Weber, T. The antiSMASH Database Version 3: Increased Taxonomic Coverage and New Query Features for Modular Enzymes. *Nucleic Acids Res.* **2021**, *49*, 639–643.
(34) Hannigan, G. D. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res* **2019**, *47*, 110–110.

(35) Ichikawa, N. DoBISCUIT: A Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2013**, *41*, 408–414.
(36) Kautsar, S. A. MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res.* **2020**, *48*, 454–458.
(37) Palaniappan, K. IMG-ABC v.5.0: An Update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res.* **2020**, *48*, 422–430.
(38) Bentley, S. D. Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2. *Nature* **2002**, *417*, 141–147.
(39) Nett, M.; Ikeda, H.; Moore, B. S. Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. *Nat Prod Rep* **2009**, *26*, 1362–1384.
(40) Kim, E.; Moore, B. S.; Yoon, Y. J. Reinvigorating Natural Product Combinatorial Biosynthesis with Synthetic Biology. *Nat Chem Biol* **2015**, *11*, 649–659.
(41) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8), 509–523. https://doi.org/10.1038/nrmicro3496.
(42) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat Rev Drug Discov* **2021**, *20*, 200–216.
(43) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem* **2021**, *90*, 763–788.
(44) Xu, F. A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat Chem Biol* **2019**, *15*, 161–168.
(45) Herbst, D. A.; Townsend, C. A.; Maier, T. The Architectures of Iterative Type i PKS and FAS. *Nat Prod Rep* **2018**, *35*, 1046–1069.
(46) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
(47) Dutta, S. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510*, 512–517.
(48) Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. *Science* **2007**, *316*, 288–290.
(49) Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315–1322.
(50) Staunton, J.; J. Weissman, K. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416. https://doi.org/10.1039/A909079G.
(51) Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Nat Prod Rep* **2003**, *20*, 79–110.
(52) Ferrer, J.-L.; Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Structure of Chalcone Synthase and the Molecular Basis of Plant Polyketide Biosynthesis. *Nat Struct Mol Biol* **1999**, *6*, 775–784.
(53) Bisang, C. A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* **1999**, *401*, 502–505.
(54) Jackowski, S.; Rock, C. O. Acetoacetyl-Acyl Carrier Protein Synthase, a Potential Regulator of Fatty Acid Biosynthesis in Bacteria. *J. Biol. Chem.* **1987**, *262*, 7927–7931.
(55) Moore, B. S.; Hertweck, C. Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units. *Nat Prod Rep* **2002**, *19*, 70–99.
(56) Crawford, J. M.; Dancy, B. C. R.; Hill, E. A.; Udwary, D. W.; Townsend, C. A. Identification of a Starter Unit Acyl-Carrier Protein Transacylase Domain in an Iterative Type I Polyketide Synthase. *Proc. Natl. Acad. Sci.* **2006**, *103* (45), 16728–16733. https://doi.org/10.1073/pnas.0604112103.

(57) Tsai, S.-C. (Sheryl). The Structural Enzymology of Iterative Aromatic Polyketide Synthases: A Critical Comparison with Fatty Acid Synthases. *Annu. Rev. Biochem.* **2018**, *87* (1), 503–531. https://doi.org/10.1146/annurev-biochem-063011-164509.
(58) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat Prod Rep* **2016**, *33*, 150–161.
(59) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **2009**, *26*, 90–114.
(60) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* **1998**, *37* (22), 8132–8138. https://doi.org/10.1021/bi980466i.
(61) Meadows, E. S.; Khosla, C. In Vitro Reconstitution and Analysis of the Chain Initiating Enzymes of the R1128 Polyketide Synthase. *Biochemistry* **2001**, *40*, 14855–14861.
(62) Winter, J. M. Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, *137*, 9885–9893.
(63) Cacho, R. A. Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J Am Chem Soc* **2015**, *137*, 15688–15691.
(64) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide b Synthase. *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
(65) Crawford, J. M. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320*, 243–246.
(66) Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709. https://doi.org/10.1021/ja0378759.
(67) Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Dissection of the Catalytic Specificity of the Act Ketoreductase. *J. Am. Chem. Soc.* **1994**, *116*, 4166–4170.
(68) Fu, H.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides: Evidence for Temporal, but Not Regiospecific, Control of Cyclization of an Aromatic Polyketide Precursor. *Chem. Biol.* **1994**, *1*, 205–210.
(69) Shen, Y. Ectopic Expression of the Minimal whiE Polyketide Synthase Generates a Library of Aromatic Polyketides of Diverse Sizes and Shapes. In *Proceedings of the National Academy of Sciences*; 1999; Vol. 96, pp 3622–3627.
(70) Auclair, K. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels–alder Reaction of a Substrate Analogue. *J Am Chem Soc* **2000**, *122*, 11519–11520.
(71) Kim, H. J.; Ruszczycky, M. W.; Choi, S.; Liu, Y.; Liu, H. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn a. *Nature* **2011**, *473*, 109–112.
(72) Kato, N. A New Enzyme Involved in the Control of the Stereochemistry in the Decalin Formation during Equisetin Biosynthesis. *Biochem. Biophys. Res. Commun.* **2015**, *460*, 210–215.
(73) Kasahara, K. Solanapyrone Synthase, a Possible Diels–Alderase and Iterative Type i Polyketide Synthase Encoded in a Biosynthetic Gene Cluster from Alternaria Solani. *ChemBioChem* **2010**, *11*, 1245–1252.
(74) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of Claisen Cyclase Domain in Fungal Polyketide Synthase WA, a Naphthopyrone Synthase of Aspergillus Nidulans. *Chem. Biol.* **2001**, *8*, 189–197.
(75) Crawford, J. M. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461*, 1139–1143.

(76) Ames, B. D. Crystal Structure and Functional Analysis of Tetracenomycin ARO/CYC: Implications for Cyclization Specificity of Aromatic Polyketides. *PNAS* **2008**, *105*, 5349–5354.
(77) Ames, B. D. Structural and Biochemical Characterization of ZhuI Aromatase/Cyclase from the R1128 Polyketide Pathway. *Biochemistry* **2011**, *50*, 8392–8406.
(78) Yu, T.-W. Engineered biosynthesis of novel polyketides from streptomyces spore pigment polyketide synthases. *J Am Chem Soc* **1998**, *120*, 7749–7759.
(79) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively. *J. Am. Chem. Soc.* **1994**, *116*, 10855–10859.
(80) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of Post-PKS Tailoring Steps through Combinatorial Biosynthesis. *Nat Prod Rep* **2002**, *19*, 542–580.
(81) Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. Cytochromes P450 for Natural Product Biosynthesis in Streptomyces: Sequence, Structure, and Function. *Nat Prod Rep* **2017**, *34*, 1141–1172.
(82) Chung, W.; Vanderwal, C. D. Stereoselective Halogenation in Natural Product Synthesis. *Angew. Chem. Int. Ed.* **2016**, *55*, 4396–4434.
(83) Singh, S., Jr; P., G. N.; Thorson, J. S. The Structural Biology of Enzymes Involved in Natural Product Glycosylation. *Nat Prod Rep* **2012**, *29*, 1201–1237.
(84) Podust, L. M.; Sherman, D. H. Diversity of P450 Enzymes in the Biosynthesis of Natural Products. *Nat Prod Rep* **2012**, *29*, 1251–1266.
(85) Carreras, C. W.; Pieper, R.; Khosla, C. Efficient Synthesis of Aromatic Polyketides in Vitro by the Actinorhodin Polyketide Synthase. *J Am Chem Soc* **1996**, *118*, 5158–5159.
(86) Shen, B.; Hutchinson, C. R. Deciphering the Mechanism for the Assembly of Aromatic Polyketides by a Bacterial Polyketide Synthase. *Proc. Natl. Acad. Sci.* **1996**, *93*, 6600–6604.
(87) Harris, T. M.; Harris, C. M. Biomimetic Syntheses of Aromatic Polyketide Metabolites. *Pure Appl Chem* **1986**, *58*, 283–294.
(88) Sherman, D. h. Structure and Deduced Function of the Granaticin-Producing Polyketide Synthase Gene Cluster of Streptomyces Violaceoruber Tü22. *EMBO J.* **1989**, *8*, 2717–2725.
(89) Kim, E. S.; Cramer, K. D.; Shreve, A. L.; Sherman, D. H. Heterologous Expression of an Engineered Biosynthetic Pathway: Functional Dissection of Type II Polyketide Synthase Components in Streptomyces Species. *J Bacteriol* **1995**, *177*, 1202–1207.
(90) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11* (9), 888–893. https://doi.org/10.1038/nsmb808.
(91) Sherman, D. H.; Kim, E. S.; Bibb, M. J.; Hopwood, D. A. Functional Replacement of Genes for Individual Polyketide Synthase Components in Streptomyces Coelicolor A3(2) by Heterologous Genes from a Different Polyketide Pathway. *J. Bacteriol.* **1992**, *174*, 6184–6190.
(92) Kim, E. S.; Hopwood, D. A.; Sherman, D. H. Analysis of Type II Polyketide Beta-Ketoacyl Synthase Specificity in Streptomyces Coelicolor A3(2) by Trans Complementation of Actinorhodin Synthase Mutants. *J Bacteriol* **1994**, *176*, 1801–1804.
(93) Dreier, J.; Khosla, C. Mechanistic Analysis of a Type II Polyketide Synthase. Role of Conserved Residues in the β-Ketoacyl Synthase–chain Length Factor Heterodimer. *Biochemistry* **2000**, *39*, 2088–2095.

(94) Grammbitter, G. L. C. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. *J Am Chem Soc* **2019**, *141*, 16615–16623.
(95) Bräuer, A. Structural Snapshots of the Minimal PKS System Responsible for Octaketide Biosynthesis. *Nat Chem* **2020**, *12*, 755–763.
(96) Nicholson, T. P. First in Vitro Directed Biosynthesis of New Compounds by a Minimal Type II Polyketide Synthase: Evidence for the Mechanism of Chain Length Determination. *Chem Commun* **2003**, *686–687*. https://doi.org/10.1039/B300847A.
(97) Kim, Y.; Prestegard, J. H. Refinement of the NMR Structures for Acyl Carrier Protein with Scalar Coupling Data. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 377–385.
(98) Evans, S. E. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* **2008**, *9*, 2424–2432.
(99) Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein. *Biochemistry* **2003**, *42*, 4648–4657.
(100) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. Solution Structure of the Actinorhodin Polyketide Synthase Acyl Carrier Protein from Streptomyces Coelicolor A3(2),. *Biochemistry* **1997**, *36* (20), 6000–6008. https://doi.org/10.1021/bi970006+.
(101) Khosla, C.; Ebert-Khosla, S.; Hopwood, D. A. Targeted Gene Replacements in a Streptomyces Polyketide Synthase Gene Cluster: Role for the Acyl Carrier Protein. *Mol. Microbiol.* **1992**, *6*, 3237–3249.
(102) Zhang, Y.-M.; Wu, B.; Zheng, J.; Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction*. *J. Biol. Chem.* **2003**, *278*, 52935–52943.
(103) Arthur, C. J. Structure and Malonyl CoA-ACP Transacylase Binding of Streptomyces Coelicolor Fatty Acid Synthase Acyl Carrier Protein. *ACS Chem Biol* **2009**, *4*, 625–636.
(104) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-Based Analysis of the Molecular Interactions between Acyltransferase and Acyl Carrier Protein in Vicenistatin Biosynthesis. *PNAS* **2016**, *113*, 1802–1807.
(105) Bartholow, T. G. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4*, 1–10.
(106) Cho, Y. I. Engineered Chimeras Unveil Swappable Modular Features of Fatty Acid and Polyketide Synthase Acyl Carrier Proteins. *Biochemistry* **2022**, *61*, 217–227.
(107) Khosla, C. Genetic Construction and Functional Analysis of Hybrid Polyketide Synthases Containing Heterologous Acyl Carrier Proteins. *J. Bacteriol.* **1993**, *175*, 2197–2204.
(108) Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C. Ketosynthases in the Initiation and Elongation Modules of Aromatic Polyketide Synthases Have Orthogonal Acyl Carrier Protein Specificity†, 2003. https://doi.org/10.1021/BI0341962.
(109) Das, A.; Khosla, C. In Vivo and in Vitro Analysis of the Hedamycin Polyketide Synthase. *Chem. Biol.* **2009**, *16*, 1197–1207.
(110) Koppisch, A. T.; Khosla, C. Structure-Based Mutagenesis of the Malonyl-CoA:Acyl Carrier Protein Transacylase from Streptomyces Coelicolor. *Biochemistry* **2003**, *42*, 11057–11064.
(111) Beltran-Alvarez, P.; Cox, R. J.; Crosby, J.; Simpson, T. J. Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase. *Biochemistry* **2007**, *46*, 14672–14681.

(112) Arthur, C. J. Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* **2005**, *44*, 15414–15421.
(113) Kalaitzis, J. A. Policing Starter Unit Selection of the Enterocin Type II Polyketide Synthase by the Type II Thioesterase EncL. *Bioorg. Med. Chem.* **2011**, *19*, 6633–6638.
(114) Shepherd, M. D.; Kharel, M. K.; Zhu, L. L.; Lanen, S. G. van; Rohr, J. Delineating the Earliest Steps of Gilvocarcin Biosynthesis: Role of GilP and GilQ in Starter Unit Specificity. *Org Biomol Chem* **2010**, *8*, 3851–3856.
(115) Tang, Y.; Koppisch, A. T.; Khosla, C. The Acyltransferase Homologue from the Initiation Module of the R1128 Polyketide Synthase Is an Acyl-ACP Thioesterase That Edits Acetyl Primer Units. *Biochemistry* **2004**, *43* (29), 9546–9555. https://doi.org/10.1021/bi049157k.
(116) Keatinge-Clay, A. T. Catalysis, Specificity, and ACP Docking Site of Streptomyces Coelicolor Malonyl-CoA:ACP Transacylase. *Structure* **2003**, *11*, 147–154.
(117) Szu, P.-H. Analysis of the Ketosynthase-Chain Length Factor Heterodimer from the Fredericamycin Polyketide Synthase. *Chem. Biol.* **2011**, *18*, 1021–1031.
(118) Bartel, P. L. Biosynthesis of Anthraquinones by Interspecies Cloning of Actinorhodin Biosynthesis Genes in Streptomycetes: Clarification of Actinorhodin Gene Functions. *J. Bacteriol.* **1990**, *172*, 4816–4826.
(119) Rajgarhia, V. B.; Strohl, W. R. Minimal streptomyces sp. Strain C5 daunorubicin polyketide biosynthesis genes required for aklanonic acid biosynthesis. *J Bacteriol* **1997**, *179*, 2690–2696.
(120) Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C. Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKR): Molecular Basis of Stereo- and Regiospecificities. *Biochemistry* **2011**, *50*, 7426–7439.
(121) Piel, J. Cloning, Sequencing and Analysis of the Enterocin Biosynthesis Gene Cluster from the Marine Isolate 'Streptomyces Maritimus': Evidence for the Derailment of an Aromatic Polyketide Synthase. *Chem. Biol.* **2000**, *7*, 943–955.
(122) Price, A. C.; Zhang, Y. M.; Rock, C. O.; White, S. W. Structure of β-Ketoacyl-[Acyl Carrier Protein] Reductase from Escherichia Coli: Negative Cooperativity and Its Structural Basis. *Biochemistry* **2001**, *40*, 12772–12781.
(123) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43* (46), 14529–14538. https://doi.org/10.1021/bi048133a.
(124) Oppermann, U. Short-Chain Dehydrogenases/Reductases (SDR): The 2002 Update. *Chem. Biol. Interact.* **2003**, *143–144*, 247–253.
(125) Filling, C. Critical Residues for Structure and Catalysis in Short-Chain Dehydrogenases/Reductases*. *J. Biol. Chem.* **2002**, *277*, 25677–25684.
(126) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **1993**, *262* (5139), 1546–1550.
(127) Kalaitzis, J. A.; Moore, B. S. Heterologous Biosynthesis of Truncated Hexaketides Derived from the Actinorhodin Polyketide Synthase. *J. Nat. Prod.* **2004**, *67*, 1419–1422.
(128) Zawada, R. J.; Khosla, C. Heterologous Expression, Purification, Reconstitution and Kinetic Analysis of an Extended Type II Polyketide Synthase. *Chem. Biol.* **1999**, *6* (9), 607–615. https://doi.org/10.1016/s1074-5521(99)80112-4.
(129) Hadfield, A. T. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12*, 1865–1875.

(130) Kantola, J. Folding of the Polyketide Chain Is Not Dictated by Minimal Polyketide Synthase in the Biosynthesis of Mithramycin and Anthracycline. *Chem. Biol.* **1997**, *4*, 751–755.
(131) Kunnari, T. Incorrectly Folded Aromatic Polyketides from Polyketide Reductase Deficient Mutants. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2639–2642.
(132) Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C. Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase,. *Biochemistry* **2008**, *47* (7), 1837–1847. https://doi.org/10.1021/bi7016427.
(133) Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C. C. Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochemistry* **2011**, *50*, 4638–4649.
(134) Javidpour, P. The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chem. Biol.* **2013**, *20*, 1225–1234.
(135) Zhang, H. L. Mutactin, a Novel Polyketide from Streptomyces Coelicolor. Structure and Biosynthetic Relationship to Actinorhodin. *J Org Chem* **1990**, *55*, 1682–1684.
(136) Zhao, S. Molecular Basis for Polyketide Ketoreductase–Substrate Interactions. *Int. J. Mol. Sci.* **2020**, *21*, 7562.
(137) Kharel, M. K. Elucidation of post-PKS tailoring steps involved in landomycin biosynthesis. *Org Biomol Chem* **2012**, *10*, 4256–4265.
(138) Mayer, A. LanV, a Bifunctional Enzyme: Aromatase and Ketoreductase during Landomycin a Biosynthesis. *ChemBioChem* **2005**, *6*, 2312–2315.
(139) Dickens, M. L.; Ye, J.; Strohl, W. R. Cloning, Sequencing, and Analysis of Aklaviketone Reductase from Streptomyces Sp. Strain C5. *J. Bacteriol.* **1996**, *178*, 3384–3388.
(140) Lackner, G. Biosynthesis of Pentangular Polyphenols: Deductions from the Benastatin and Griseorhodin Pathways. *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
(141) Torkkell, S. The Entire Nogalamycin Biosynthetic Gene Cluster of Streptomyces Nogalater: Characterization of a 20-Kb DNA Region and Generation of Hybrid Structures. *Mol Gen Genomics* **2001**, *266*, 276–288.
(142) Paananen, P. Structural and Functional Analysis of Angucycline C-6 Ketoreductase LanV Involved in Landomycin Biosynthesis. *Biochemistry* **2013**, *52*, 5304–5314.
(143) Patrikainen, P. Structure-Based Engineering of Angucyclinone 6-Ketoreductases. *Chem. Biol.* **2014**, *21*, 1381–1391.
(144) Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C. Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (12), 3421–3430. https://doi.org/10.1021/acschembio.6b00658.
(145) Xiao, X. Functional and Structural Insights into a Novel Promiscuous Ketoreductase of the Lugdunomycin Biosynthetic Pathway. *ACS Chem Biol* **2020**, *15*, 2529–2538.
(146) Kallio, P. Flavoprotein Hydroxylase PgaE Catalyzes Two Consecutive Oxygen-Dependent Tailoring Reactions in Angucycline Biosynthesis. *Biochemistry* **2011**, *50*, 5535–5543.
(147) Patrikainen, P. Tailoring Enzymes Involved in the Biosynthesis of Angucyclines Contain Latent Context-Dependent Catalytic Activities. *Chem. Biol.* **2012**, *19*, 647–655.
(148) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **2003**, *4*, 654–657.
(149) Moreno, G. O. *Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type i and Type II Polyketide Synthases Using Atom-Replaced Probes*; UC Irvine, 2018.

(150) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(151) Li 然 李嫣, Y.; Xu 伟徐, W.; Tang 奕唐, Y. Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains*. *J. Biol. Chem.* **2010**, *285* (30), 22764–22773. https://doi.org/10.1074/jbc.M110.128504.
(152) Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C. A. Structural Basis for Biosynthetic Programming of Fungal Aromatic Polyketide Cyclization. *Nature* **2009**, *461* (7267), 1139–1143. https://doi.org/10.1038/nature08475.
(153) Barajas, J. F.; Shakya, G.; Moreno, G.; Rivera, H.; Jackson, D. R.; Topper, C. L.; Vagstad, A. L.; La Clair, J. J.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. *Proc. Natl. Acad. Sci.* **2017**, *114* (21), E4142–E4148. https://doi.org/10.1073/pnas.1609001114.
(154) Liu, L.; Zhang, Z.; Shao, C.-L.; Wang, J.-L.; Bai, H.; Wang, C.-Y. Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Sci. Rep.* **2015**, *5* (1), 10463. https://doi.org/10.1038/srep10463.
(155) Feng, Y.; Yang, X.; Ji, H.; Deng, Z.; Lin, S.; Zheng, J. The Streptomyces Viridochromogenes Product Template Domain Represents an Evolutionary Intermediate between Dehydratase and Aldol Cyclase of Type I Polyketide Synthases. *Commun. Biol.* **2022**, *5* (1), 508. https://doi.org/10.1038/s42003-022-03477-8.
(156) Pan, H. Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. *Structure* **2002**, *10*, 1559–1568.
(157) Thompson, T. B.; Katayama, K.; Watanabe, K.; Hutchinson, C. R.; Rayment, I. Structural and Functional Analysis of Tetracenomycin F2 Cyclase from Streptomyces Glaucescens: A TYPE II POLYKETIDE CYCLASE*. *J. Biol. Chem.* **2004**, *279*, 37956–37963.
(158) Sultana, A. Structure of the Polyketide Cyclase SnoaL Reveals a Novel Mechanism for Enzymatic Aldol Condensation. *EMBO J.* **2004**, *23*, 1911–1921.
(159) Kallio, P.; Sultana, A.; Niemi, J.; Mäntsälä, P.; Schneider, G. Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity. *J. Mol. Biol.* **2006**, *357*, 210–220.
(160) Hutchinson, C. R. Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem. Rev.* **1997**, *97* (7), 2525–2536. https://doi.org/10.1021/cr960022x.
(161) Ahuja, M.; Chiang, Y.-M.; Chang, S.-L.; Praseuth, M. B.; Entwistle, R.; Sanchez, J. F.; Lo, H.-C.; Yeh, H.-H.; Oakley, B. R.; Wang, C. C. C. Illuminating the Diversity of Aromatic Polyketide Synthases in Aspergillus Nidulans. *J. Am. Chem. Soc.* **2012**, *134* (19), 8212–8221. https://doi.org/10.1021/ja3016395.
(162) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97* (7), 2465–2498. https://doi.org/10.1021/cr960034i.
(163) Shen, B. Polyketide Biosynthesis beyond the Type I, II and III Polyketide Synthase Paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6.
(164) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639. https://doi.org/10.1021/ar8002249.
(165) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A.

Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130. https://doi.org/10.1126/science.ade2574.
(166) Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y. Structural Basis for Selectivity in a Highly Reducing Type II Polyketide Synthase. *Nat. Chem. Biol.* **2020**, *16* (7), 776–782. https://doi.org/10.1038/s41589-020-0530-0.
(167) Yadav, G.; Gokhale, R. S.; Mohanty, D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J. Mol. Biol.* **2003**, *328* (2), 335–363. https://doi.org/10.1016/S0022-2836(03)00232-8.
(168) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci.* **2021**, *118* (15), e2016239118. https://doi.org/10.1073/pnas.2016239118.
(169) Capela, J.; Zimmermann-Kogadeeva, M.; Dijk, A. D. J. van; de Ridder, D.; Dias, O.; Rocha, M. Comparative Assessment of Protein Large Language Models for Enzyme Commission Number Prediction. *BMC Bioinformatics* **2025**, *26* (1), 68. https://doi.org/10.1186/s12859-025-06081-9.
(170) Jacques, M.-A.; Durand, K.; Orgeur, G.; Balidas, S.; Fricot, C.; Bonneau, S.; Quillévéré, A.; Audusseau, C.; Olivier, V.; Grimault, V.; Mathis, R. Phylogenetic Analysis and Polyphasic Characterization of Clavibacter Michiganensis Strains Isolated from Tomato Seeds Reveal That Nonpathogenic Strains Are Distinct from C. Michiganensis Subsp. Michiganensis. *Appl. Environ. Microbiol.* **2012**, *78* (23), 8388–8402. https://doi.org/10.1128/AEM.02158-12.
(171) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat. Chem. Biol.* **2019**, *15* (7), 669–671. https://doi.org/10.1038/s41589-019-0301-y.
(172) Ellis, B. D.; Milligan, J. C.; White, A. R.; Duong, V.; Altman, P. X.; Mohammed, L. Y.; Crump, M. P.; Crosby, J.; Luo, R.; Vanderwal, C. D.; Tsai, S.-C. An Oxetane-Based Polyketide Surrogate To Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140* (15), 4961–4964. https://doi.org/10.1021/jacs.7b11793.
(173) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.
(174) Zdouc, M. M.; Blin, K.; Louwen, N. L. L.; Navarro, J.; Loureiro, C.; Bader, C. D.; Bailey, C. B.; Barra, L.; Booth, T. J.; Bozhüyük, K. A. J.; Cediel-Becerra, J. D. D.; Charlop-Powers, Z.; Chevrette, M. G.; Chooi, Y. H.; D'Agostino, P. M.; de Rond, T.; Del Pup, E.; Duncan, K. R.; Gu, W.; Hanif, N.; Helfrich, E. J. N.; Jenner, M.; Katsuyama, Y.; Korenskaia, A.; Krug, D.; Libis, V.; Lund, G. A.; Mantri, S.; Morgan, K. D.; Owen, C.; Phan, C.-S.; Philmus, B.; Reitz, Z. L.; Robinson, S. L.; Singh, K. S.; Teufel, R.; Tong, Y.; Tugizimana, F.; Ulanova, D.;

Winter, J. M.; Aguilar, C.; Akiyama, D. Y.; Al-Salihi, S. A. A.; Alanjary, M.; Alberti, F.; Aleti, G.; Alharthi, S. A.; Rojo, M. Y. A.; Arishi, A. A.; Augustijn, H. E.; Avalon, N. E.; Avelar-Rivas, J. A.; Axt, K. K.; Barbieri, H. B.; Barbosa, J. C. J.; Barboza Segato, L. G.; Barrett, S. E.; Baunach, M.; Beemelmanns, C.; Beqaj, D.; Berger, T.; Bernaldo-Agüero, J.; Bettenbühl, S. M.; Bielinski, V. A.; Biermann, F.; Borges, R. M.; Borriss, R.; Breitenbach, M.; Bretscher, K. M.; Brigham, M. W.; Buedenbender, L.; Bulcock, B. W.; Cano-Prieto, C.; Capela, J.; Carrion, V. J.; Carter, R. S.; Castelo-Branco, R.; Castro-Falcón, G.; Chagas, F. O.; Charria-Girón, E.; Chaudhri, A. A.; Chaudhry, V.; Choi, H.; Choi, Y.; Choupannejad, R.; Chromy, J.; Donahey, M. S. C.; Collemare, J.; Connolly, J. A.; Creamer, K. E.; Crüsemann, M.; Cruz, A. A.; Cumsille, A.; Dallery, J.-F.; Damas-Ramos, L. C.; Damiani, T.; de Kruijff, M.; Martín, B. D.; Sala, G. D.; Dillen, J.; Doering, D. T.; Dommaraju, S. R.; Durusu, S.; Egbert, S.; Ellerhorst, M.; Faussurier, B.; Fetter, A.; Feuermann, M.; Fewer, D. P.; Foldi, J.; Frediansyah, A.; Garza, E. A.; Gavriilidou, A.; Gentile, A.; Gerke, J.; Gerstmans, H.; Gomez-Escribano, J. P.; González-Salazar, L. A.; Grayson, N. E.; Greco, C.; Gomez, J. E. G.; Guerra, S.; Flores, S. G.; Gurevich, A.; Gutiérrez-García, K.; Hart, L.; Haslinger, K.; He, B.; Hebra, T.; Hemmann, J. L.; Hindra, H.; Höing, L.; Holland, D. C.; Holme, J. E.; Horch, T.; Hrab, P.; Hu, J.; Huynh, T.-H.; Hwang, J.-Y.; Iacovelli, R.; Iftime, D.; Iorio, M.; Jayachandran, S.; Jeong, E.; Jing, J.; Jung, J. J.; Kakumu, Y.; Kalkreuter, E.; Kang, K. B.; Kang, S.; Kim, W.; Kim, G. J.; Kim, H.; Kim, H. U.; Klapper, M.; Koetsier, R. A.; Kollten, C.; Kovács, Á. T.; Kriukova, Y.; Kubach, N.; Kunjapur, A. M.; Kushnareva, A. K.; Kust, A.; Lamber, J.; Larralde, M.; Larsen, N. J.; Launay, A. P.; Le, N.-T.-H.; Lebeer, S.; Lee, B. T.; Lee, K.; Lev, K. L.; Li, S.-M.; Li, Y.-X.; Licona-Cassani, C.; Lien, A.; Liu, J.; Lopez, J. A. V.; Machushynets, N. V.; Macias, M. I.; Mahmud, T.; Maleckis, M.; Martinez-Martinez, A. M.; Mast, Y.; Maximo, M. F.; McBride, C. M.; McLellan, R. M.; Bhatt, K. M.; Melkonian, C.; Merrild, A.; Metsä-Ketelä, M.; Mitchell, D. A.; Müller, A. V.; Nguyen, G.-S.; Nguyen, H. T.; Niedermeyer, T. H. J.; O'Hare, J. H.; Ossowicki, A.; Ostash, B. O.; Otani, H.; Padva, L.; Paliyal, S.; Pan, X.; Panghal, M.; Parade, D. S.; Park, J.; Parra, J.; Rubio, M. P.; Pham, H. T.; Pidot, S. J.; Piel, J.; Pourmohsenin, B.; Rakhmanov, M.; Ramesh, S.; Rasmussen, M. H.; Rego, A.; Reher, R.; Rice, A. J.; Rigolet, A.; Romero-Otero, A.; Rosas-Becerra, L. R.; Rosiles, P. Y.; Rutz, A.; Ryu, B.; Sahadeo, L.-A.; Saldanha, M.; Salvi, L.; Sánchez-Carvajal, E.; Santos-Medellin, C.; Sbaraini, N.; Schoellhorn, S. M.; Schumm, C.; Sehnal, L.; Selem, N.; Shah, A. D.; Shishido, T. K.; Sieber, S.; Silviani, V.; Singh, G.; Singh, H.; Sokolova, N.; Sonnenschein, E. C.; Sosio, M.; Sowa, S. T.; Steffen, K.; Stegmann, E.; Streiff, A. B.; Strüder, A.; Surup, F.; Svenningsen, T.; Sweeney, D.; Szenei, J.; Tagirdzhanov, A.; Tan, B.; Tarnowski, M. J.; Terlouw, B. R.; Rey, T.; Thome, N. U.; Torres Ortega, L. R.; Tørring, T.; Trindade, M.; Truman, A. W.; Tvilum, M.; Udwary, D. W.; Ulbricht, C.; Vader, L.; van Wezel, G. P.; Walmsley, M.; Warnasinghe, R.; Weddeling, H. G.; Weir, A. N. M.; Williams, K.; Williams, S. E.; Witte, T. E.; Rocca, S. M. W.; Yamada, K.; Yang, D.; Yang, D.; Yu, J.; Zhou, Z.; Ziemert, N.; Zimmer, L.; Zimmermann, A.; Zimmermann, C.; van der Hooft, J. J. J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 4.0: Advancing Biosynthetic Gene Cluster Curation through Global Collaboration. *Nucleic Acids Res.* **2025**, *53* (D1), D678–D690. https://doi.org/10.1093/nar/gkae1115.

(175) Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S. J.; Webb, C.; Wu, D.; Varghese, N. J.; Reddy, T. B. K.; Mukherjee, S.; Ovchinnikova, G.; Nolan, M.; Seshadri, R.; Roux, S.; Visel, A.; Woyke, T.; Eloe-Fadrosh, E. A.; Kyrpides, N. C.; Ivanova, N. N. The IMG/M Data Management and Analysis System

v.7: Content Updates and New Features. *Nucleic Acids Res.* **2023**, *51* (D1), D723–D732. https://doi.org/10.1093/nar/gkac976.
(176) Fernández-Moreno, M. A.; Martínez, E.; Caballero, J. L.; Ichinose, K.; Hopwood, D. A.; Malpartida, F. DNA Sequence and Functions of the actVI Region of the Actinorhodin Biosynthetic Gene Cluster of Streptomyces Coelicolor A3(2). *J. Biol. Chem.* **1994**, *269* (40), 24854–24863. https://doi.org/10.1016/S0021-9258(17)31469-2.
(177) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Richard Hutchinson, C. Iterative Type II Polyketide Synthases, Cyclases and Ketoreductases Exhibit Context-Dependent Behavior in the Biosynthesis of Linear and Angular Decapolyketides. *Chem. Biol.* **1997**, *4* (6), 433–443. https://doi.org/10.1016/S1074-5521(97)90195-2.
(178) Moreno, G. O. Structural and Functional Studies of Enzymes from Non-Reducing Iterative Type I and Type II Polyketide Synthases Using Atom-Replaced Probes, UC Irvine, 2018. https://escholarship.org/uc/item/1tx0518k.
(179) Serapian, S. A.; Crosby, J.; Crump, M. P.; van der Kamp, M. W. Path to Actinorhodin: Regio- and Stereoselective Ketone Reduction by a Type II Polyketide Ketoreductase Revealed in Atomistic Detail. *JACS Au* **2022**, *2* (4), 972–984. https://doi.org/10.1021/jacsau.2c00086.
(180) Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941–953. https://doi.org/10.1016/j.jmb.2008.09.084.
(181) Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P. The Crystal Structure of the actIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Structure* **2004**, *12* (10), 1865–1875. https://doi.org/10.1016/j.str.2004.08.002.
(182) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat. Prod. Rep.* **2016**, *33* (2), 203–230. https://doi.org/10.1039/C5NP00109A.
(183) Chooi, Y.-H.; Tang, Y. Navigating the Fungal Polyketide Chemical Space: From Genes to Molecules. *J. Org. Chem.* **2012**, *77* (22), 9933–9953. https://doi.org/10.1021/jo301592k.
(184) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A. Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science* **2008**, *320* (5873), 243–246. https://doi.org/10.1126/science.1154711.
(185) Barajas, J. F.; Finzel, K.; Valentic, T. R.; Shakya, G.; Gamarra, N.; Martinez, D.; Meier, J. L.; Vagstad, A. L.; Newman, A. G.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Structural and Biochemical Analysis of Protein–Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew. Chem. Int. Ed.* **2016**, *55* (42), 13005–13009. https://doi.org/10.1002/anie.201605401.
(186) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(187) Barajas, J. F. Polyketide Mimetics Yield Structural and Mechanistic Insights into Product Template Domain Function in Nonreducing Polyketide Synthases. In *Proceedings of the National Academy of Sciences*; 2017; Vol. 114, pp 4142–4148.
(188) Milligan, J. C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat Chem Biol* **2019**, *15*, 669–671.

(189) Ellis, B. D. An Oxetane-Based Polyketide Surrogate to Probe Substrate Binding in a Polyketide Synthase. *J. Am. Chem. Soc.* **2018**, *140*, 4961–4964.
(190) Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C. Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **2004**, *43*, 14529–14538.