Chemoinformatic-guided engineering of polyketide synthases

Amin Zargar,[1,2,4] Ravi Lal[1,2], Luis Valencia[1,2], Jessica Wang[1,2], Tyler Backman[1,2], Pablo Cruz-Morales[1,2], Ankita Kothari[1-2], Miranda Werts[1,2], Andrew R. Wong[1,2], Constance B. Bailey[1,2,4], Arthur Loubat[1,2], Yuzhong Liu[1,2], Veronica Benites[1,2,3], Samantha Chang[1,2], Amanda C. Hernández[1,2], Jesus F. Barajas[1-3], Mitchell G. Thompson[1,2], Carolina Barcelos[1,2], Rasha Anayah[1,2], Hector Garcia Martin[1-3,9], Aindrila Mukhopadhyay[1,2], Edward Baidoo[1,2,3], Leonard Katz[1,4], Jay D. Keasling[1-2,4-8]

[1] Joint BioEnergy Institute, Lawrence Berkeley National Laboratory,

Emeryville, CA 94608, United States

[2] Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[3] Department of Energy Agile BioFoundry, Emeryville, CA 94608, USA

[4] QB3 Institute, University of California-Berkeley, 5885 Hollis Street, 4th Floor, Emeryville, CA 94608, United States

[5] Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720, United States

[6] Department of Bioengineering, University of California, Berkeley, CA 94720, United States

[7] Novo Nordisk Foundation Center for Biosustainability, Technical University

Denmark, DK2970-Horsholm, Denmark

[8] Synthetic Biochemistry Center, Institute for Synthetic Biology, Shenzhen Institutes for Advanced Technologies, Shenzhen, China

[9] BCAM, Basque Center for Applied Mathematics, 48009 Bilbao, Spain

*Corresponding author

Jay D. Keasling

jdkeasling@lbl.gov

Abstract

Polyketide synthase (PKS) engineering is an attractive method to generate new molecules such as commodity, fine and specialty chemicals. A central challenge in PKS design is replacing a partially reductive module with a fully reductive module through a reductive loop exchange, thereby generating a saturated β-carbon. In this work, we sought to establish an engineering strategy for reductive loop exchanges based on chemoinformatics, a field traditionally used in drug discovery. We first introduced a set of donor reductive loops of diverse genetic origin and chemical substrate structures into the first extension module of the lipomycin PKS (LipPKS1). These results demonstrated that chemical similarity between the substrate of the donor loops and recipient LipPKS1 correlated with product titers. Consequently, we identified donor loops with substrates chemically similar to LipPKS1 for further reductive loop exchanges, and we observed a statistically significant correlation with production. Reductive loops with the highest chemical similarity resulted in production of branched, short-chain fatty acids reaching a titer of 165 mg/L in *Streptomyces albus* J1074. Collectively, our work formulizes a new chemoinformatic paradigm for *de novo* PKS biosynthesis which may accelerate the production of valuable bioproducts.

Intro

As the architecture of Type I PKSs determines molecular structure, rational reprogramming of PKS enzymes for the biosynthesis of new polyketides has been a major research thrust over the past three decades.[1–3] Like fatty acid synthases, PKSs extend the growing chain from the ketosynthase (KS) domain with a malonyl-CoA analog loaded onto the acyl carrier protein (ACP) by the acyltransferase (AT) domain through a decarboxylative Claisen condensation reaction. Unlike fatty acid synthases, which faithfully produce saturated fatty acids, PKSs have variability in β-carbonyl reduction, an attractive feature for molecular design. After chain extension, the β-carbonyl reduction state is determined by the reductive domains within the module, namely the ketoreductase (KR), dehydratase (DH), and enoylreductase (ER), which generate the β-hydroxyl, α-β alkene, or saturated β-carbons respectively, when progressively combined. As the degree of β-carbon reduction is an important feature in molecular design, multiple studies have reported the engineering of a PKS module for various oxidation states of the β-carbon.[4–8] However, design principles for introduction of reductive loop exchanges (*i.e.* KR-DH-ER domains) into partially reductive modules have not yet been developed. In this work, we compare bioinformatic and chemoinformatic approaches to guide reductive loop exchanges and formalize a new paradigm based on the chemical similarity of the substrate.

Chemoinformatics, an interdisciplinary field blending computational chemistry, molecular modeling and statistics, was initially developed for drug discovery through analysis of structure-activity relationships.[9] Recently, we suggested that a chemoinformatic approach to PKS engineering could be valuable, particularly in reductive loop exchanges due to the dependence of the KR and DH domains on substrate size [1]. For example, due to a hydrophobic catalytic tunnel,[10,11] acyl chain length had a critical influence on dehydration in both stand-alone DH[12] and full PKS module studies.[7] Moreover, a previous study of engineered reductive loop

swaps resulted in a correlation between production and substrate size similarity of the donor reductive loops and the recipient module.[13] Chemoinformatic methods such as atom pair (AP) similarity and maximum common substructure (MCS) similarity could be used to describe the substrate profiles for catalysis by these domains. AP similarity characterizes atom pairs (*e.g.* length of bond path, number of π electrons), and MCS similarity is based on identifying the largest common substructure between two molecules. Both similarity methods can be translated to a Tanimoto coefficient with a range of 0 (least similar) to 1 (most similar).[14] Based on the substrate-dependence of the reductive domains, we hypothesized that chemosimilarity between the substrates of donor and acceptor modules in reductive loop exchanges would correlate with production levels.

Bioinformatic studies of PKS evolution have guided engineering efforts in closely related biosynthetic gene clusters (BGCs).[15,16] We therefore undertook a phylogenetic analysis of the reductive domain common to all reductive loops, the ketoreductase (KR). The KR not only reduces the β-keto group to a β-hydroxyl, but also sets the stereochemistry of the β-group and, if a branched extender is used, sets the □-carbon stereochemistry resulting in subtypes A1, A2, B1, B2 (**Figure 1A**). We generated a phylogenetic tree of every manually curated ketoreductase and ketosynthase in ClusterCAD, a database for Type I PKSs, totaling 72 biosynthetic gene clusters (BGCs) and 1077 modules (**Figure 1B**).[17] This evolutionary reconstruction revealed that KR-only B1 subtypes split from a common ancestor of fatty acid synthases and iterative PKSs.[18] As in previous investigations, we found that KR-only B1 subtypes later resulted in the addition of DH and DH/ER domains,[19] likely through recombination.[20] We extend this finding to note that the KR-only B1 subtype branch diverged to produce the other KR-only subtypes (*i.e.* A1, A2 and B2) (**Figure 1B, Supplementary Figure 1**). While KR domains cluster by the presence of a DH or DH-ER domains, KS domains do not phylogenetically cluster by the type of reductive domains active in the module (**Supplementary Figure 2**).[19] The KRs generally grouped by their product types, and this suggests a link between their evolution and product

specificity, analogous to the evolution of KS domains of cis-AT[19] and trans-AT PKS modules[21,22] towards substrate specificity. As KRs from KR-DH-ER modules evolved distinctly from KR-only modules, we hypothesized that the KR phylogenetic distance between the donor loops and acceptor module in reductive loop exchanges was unlikely to correlate with production levels.

Results and Discussion

To compare the importance of chemical similarity and phylogenetic distance in reductive loop exchanges, we swapped diverse reductive loops into the first module of the lipomycin PKS as the acceptor module. In our previous work, we introduced a heterologous thioesterase from 6-deoxyerythronolide (DEBS) into the C-terminus of the first module of the lipomycin PKS (denoted Lip1TE); the resulting truncated PKS produced a β-hydroxy acid.[23] In this work, our experimental design was based on introducing full reductive loops using conserved residues as exchange sites (denoted “A”, “B” and “C”) in Lip1TE (**Scheme 1**).[7] We selected these conserved residues based on our work in reductive loop exchanges in the first module of borreledin.[7] To evaluate the effects of genetic and chemical similarity, we identified five donor reductive loops (IdmO, indanomycin, *S. antibioticus*; SpnB, spinosyn, *S. spinosa*; AurB, aureothin, *S. aureofaciens*; NanA2, nanchangamycin, *S. nanchangensis*; MAS, mycoserosic acid, *M. marinum*) to swap into Lip1TE. A pairwise comparison of phylogenetic distance as well as direct sequence identity illustrates that the KR domain of the three donor reductive loops IdmO, SpnB, and AurB are the most similar to the KR in LipPKS1 (**Figure 2A**). A similar trend also holds in the analysis of the KS domain (**Supplementary Figure 3**). In contrast, the NanA2 substrate is the most chemically similar to LipPKS1, followed by SpnB, based on AP similarity (**Figure 2B**) and MCS similarity (data not shown). With the introduction of a reductive loop swap, the chimeric enzymes would programmatically produce 2,4-dimethyl pentanoic acid. As *in vitro* PKS studies have shown divergence from *in vivo* results[24,25] due to underestimation of factors including limiting substrate, crowding, and solubility,[26] we cloned ten chimeric modules into an

E. coli -Streptomyces albus shuttle vector and conjugated it into *Streptomyces albus* J1074 (**Table S1**).[27] Following ten-day production runs in a rich medium, cultures of *Streptomyces albus* harboring each of the constructs were harvested and the supernatants were analyzed with LC-MS for product levels.

Consistent with our hypothesis, we found a strong correlation between production titers of the desired product and the AP and MCS chemosimilarities of the donor and LipPKS1 module substrates (AP Spearman Rank Correlation of R_s of 0.99 and $p < 0.01$; MCS R_s of 0.90 and $p = 0.04$) (**Figure 2C**). On the other hand, no correlation between product titer and phylogenetic distance or sequence similarity of the KS or KR domains was found. Based on our bioinformatic analysis, this was not surprising as the lipomycin KR is an A2-type, evolving separately from a KR with a full reductive loop. This trend held with either junction A or B, although generally junction B chimeras resulted in higher levels of production, as demonstrated in a previous study of reductive loop exchanges.[7] We found that substituting the donor loop most chemically similar to LipPKS1, NanA2, resulted in the highest titers of the desired product, 2,4-dimethyl pentanoic acid, reaching 165 mg/L. Low titers of the intermediate 2,5-dimethyl-3-hydroxypentanoic acid were produced, which we hypothesize is due to a comparatively lower rate of turnover at the energetically intensive DH domain,[28] resulting in premature cleavage of the stalled product by hydration or by the thioesterase. As in our previous study of *in vitro* production of adipic acid, we did not detect alkene or keto acid stalled products [7]. This is not surprising as non-functional KRs produce short chain β-keto acids that spontaneously decarboxylate to form ketones, whereas ERs have been generally shown to rapidly reduce *trans* double bonds.

Based on these results, we took a chemoinformatic approach to further test our hypothesis that chemosimilarity is a critical factor in PKS engineering. We searched the ClusterCAD[17] database for PKS modules with full reductive loops and substrates of high chemical similiarity to that of the KR of LipPKS1. The closest matches identified were PKS

modules from laidlomycin and monensin, which used the same substrate as nanchangamycin. (**Figure 3A**). As junction B resulted in levels of production superior to junction A, we cloned the reductive loops of LaidSII and MonA2 into junction B of lipomycin. The chimeric PKSs containing reductive loops with substrates of similar chemical structure (NanA2, LaidSII, and MonA2) produced higher titers of the desired fully reduced product than less chemically similar reductive loops. We determined a Spearman rank correlation between AP Tanimoto chemosimilarity and production to have an R_s of 0.89 and a p-value of less than 0.01 compared to a correlation of R_s of 0.85 and p value of 0.01 for MCS chemosimilarity. With divergent methods of chemical similarity calculations (AP and MCS), we found a statistically significant correlation between substrate similarity and product titer.

In this work, we have undertaken a bioinformatic and chemoinformatic analysis of reductive loop exchanges. Through a phylogenetic reconstruction, we suggested that the evolutionary history of KR-only modules does not reveal useful information for predicting production rates in reductive loop swaps; in fact, phylogenetic distance and sequence similarity between donor KRs of full reductive loops and recipient KRs of partial loops did not correlate to production. Highlighting previous literature regarding the importance of substrate size in reductive domains, we hypothesized that the field of chemoinformatics, traditionally used to study structure-activity relationships in drug discovery, could be applied to PKS engineering to better predict production results. Using different reductive loops of varying phylogenetic and chemical similarity, we determined that chemosimilarity had a strong correlation with product titers. Based on these findings, we selected two more reductive loops with the most chemically similar substrates to LipPKS1 and found higher levels of production. The analysis of our results and previous experiments formalize a new paradigm in PKS engineering based on the chemosimilarity of the substrate between the donor and recipient modules. These design principles may fast-track the combinatorial approach currently taken for *de novo* biosynthesis and develop a framework to more rapidly produce valuable biochemicals.

Acknowledgements

This work was funded by the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy, the National Institute of Health Awards F32GM125179, F32GM125166 and the Agile Biofoundry sponsored by the U.S. DOE Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies and Vehicle Technologies Offices, under Contract DEAC02-05CH11231 between DOE and Lawrence Berkeley National Laboratory. H.G.M. was also supported by the Basque Government through the BERC 2018-2021 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017- 0718.

Competing Financial Interest

J.D.K. has a financial interest in Amyris, Lygos, Demetrix, Napigen, Maple Bio, Berkeley Brewing Sciences, Ansa Biotech and Apertor Labs.

Bibliography

(1) Barajas, J. F.; Blake-Hedges, J. M.; Bailey, C. B.; Curran, S.; Keasling, J. D. *Synthetic and Systems Biotechnology* **2017**, *2*, 147–166.

(2) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. *Biochemistry* **2014**, *53*, 2875–2883.

(3) Yuzawa, S.; Zargar, A.; Pang, B.; Katz, L.; Keasling, J. D. *Commodity chemicals from engineered modular type I polyketide synthases.*; 2018; Vol. 608, pp. 393–415.

(4) Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R. *Biochemistry* **2003**, *42*, 72–79.

(5) Keatinge-Clay, A. *J. Mol. Biol.* **2008**, *384*, 941–953.

(6) Kellenberger, L.; Galloway, I. S.; Sauter, G.; Böhm, G.; Hanefeld, U.; Cortés, J.; Staunton, J.; Leadlay, P. F. *Chembiochem* **2008**, *9*, 2740–2749.

(7) Hagen, A.; Poust, S.; Rond, T. de; Fortman, J. L.; Katz, L.; Petzold, C. J.; Keasling, J. D. *ACS Synth. Biol.* **2016**, *5*, 21–27.

(8) Gaisser, S.; Kellenberger, L.; Kaja, A. L.; Weston, A. J.; Lill, R. E.; Wirtz, G.; Kendrew, S. G.; Low, L.; Sheridan, R. M.; Wilkinson, B.; Galloway, I. S.; Stutzman-Engwall, K.; McArthur, H. A.; Staunton, J.; Leadlay, P. F. *Org. Biomol. Chem.* **2003**, *1*, 2840–2847.

(9) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B.-T. *Mol Divers* **2006**, *10*, 39–79.

(10) Herbst, D. A.; Jakob, R. P.; Zähringer, F.; Maier, T. *Nature* **2016**, *531*, 533–537.

(11) Barajas, J. F.; McAndrew, R. P.; Thompson, M. G.; Backman, T. W. H.; Pang, B.; de Rond, T.; Pereira, J. H.; Benites, V. T.; Martín, H. G.; Baidoo, E. E. K.; Hillson, N. J.; Adams, P. D.; Keasling, J. D. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 1225–1235.

(12) Faille, A.; Gavalda, S.; Slama, N.; Lherbet, C.; Maveyraud, L.; Guillet, V.; Laval, F.; Quémard, A.; Mourey, L.; Pedelacq, J.-D. *J. Mol. Biol.* **2017**, *429*, 1554–1569.

(13) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1846–1851.

(14) Chen, X.; Reynolds, C. H. *J Chem Inf Comput Sci* **2002**, *42*, 1407–1414.

(15) Peng, H.; Ishida, K.; Sugimoto, Y.; Jenke-Kodama, H.; Hertweck, C. *Nat. Commun.* **2019**, *10*, 3918.

(16) Awakawa, T.; Fujioka, T.; Zhang, L.; Hoshino, S.; Hu, Z.; Hashimoto, J.; Kozone, I.; Ikeda, H.; Shin-Ya, K.; Liu, W.; Abe, I. *Nat. Commun.* **2018**, *9*, 3534.

(17) Eng, C. H.; Backman, T. W. H.; Bailey, C. B.; Magnan, C.; García Martín, H.; Katz, L.; Baldi, P.; Keasling, J. D. *Nucleic Acids Res.* **2018**, *46*, D509–D515.

(18) Keatinge-Clay, A. T. *Chem. Biol.* **2007**, *14*, 898–908.

(19) Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.; Ikeda, H.; Shin-Ya, K.; Abe, I. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 1740–1745.

(20) Jenke-Kodama, H.; Börner, T.; Dittmann, E. *PLoS Comput. Biol.* **2006**, *2*, e132.

(21) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008**, *26*, 225–233.

(22) Vander Wood, D. A.; Keatinge-Clay, A. T. *Proteins* **2018**, *86*, 664–675.

(23) Yuzawa, S.; Eng, C. H.; Katz, L.; Keasling, J. D. *Biochemistry* **2013**, *52*, 3791–3793.

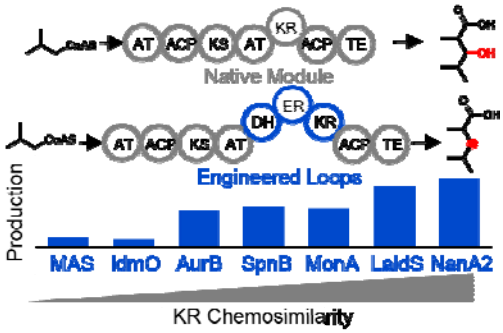
(24) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. *Annu. Rev. Biochem.* **2007**, *76*, 195–221.

(25) Yan, J.; Hazzard, C.; Bonnett, S. A.; Reynolds, K. A. *Biochemistry* **2012**, *51*, 9333–9341.

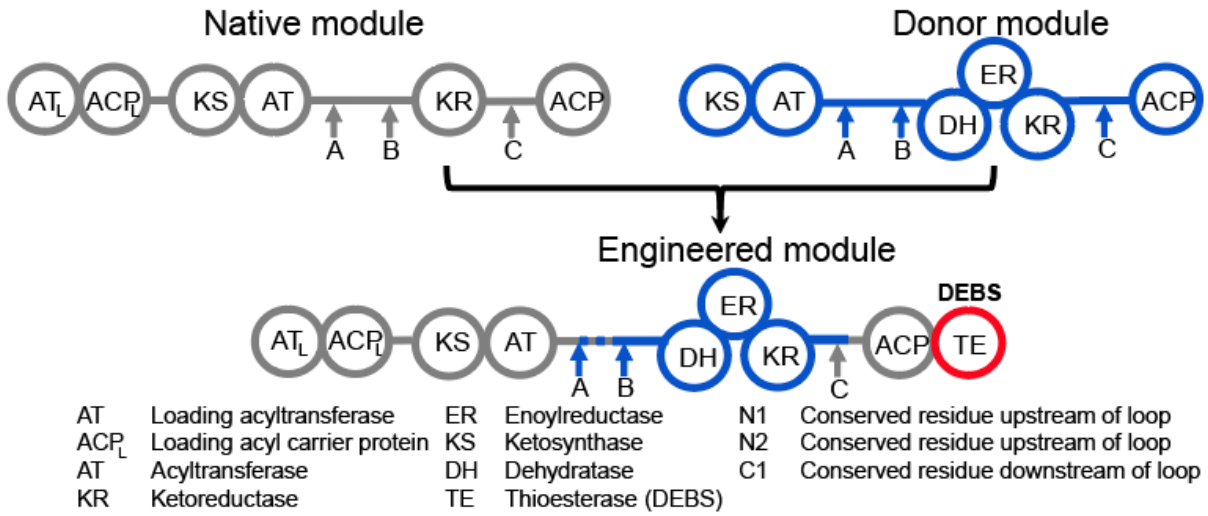
(26) Zotter, A.; Bäuerle, F.; Dey, D.; Kiss, V.; Schreiber, G. *J. Biol. Chem.* **2017**, *292*, 15838–15848.

(27) Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D. *ACS Synth. Biol.* **2017**, *6*, 159–166.

(28) Weber, A. L. *J. Mol. Evol.* **1991**, *32*, 93–100.



Graphical abstract



Scheme 1. Experimental design of PKS reductive loop swaps. Conserved residues are identified through multiple sequence alignment surrounding the reductive domains (“A”, “B” and “C”). Donor reductive loops are inserted into the native lipomycin module, and the DEBS thioesterase cleaves the product.

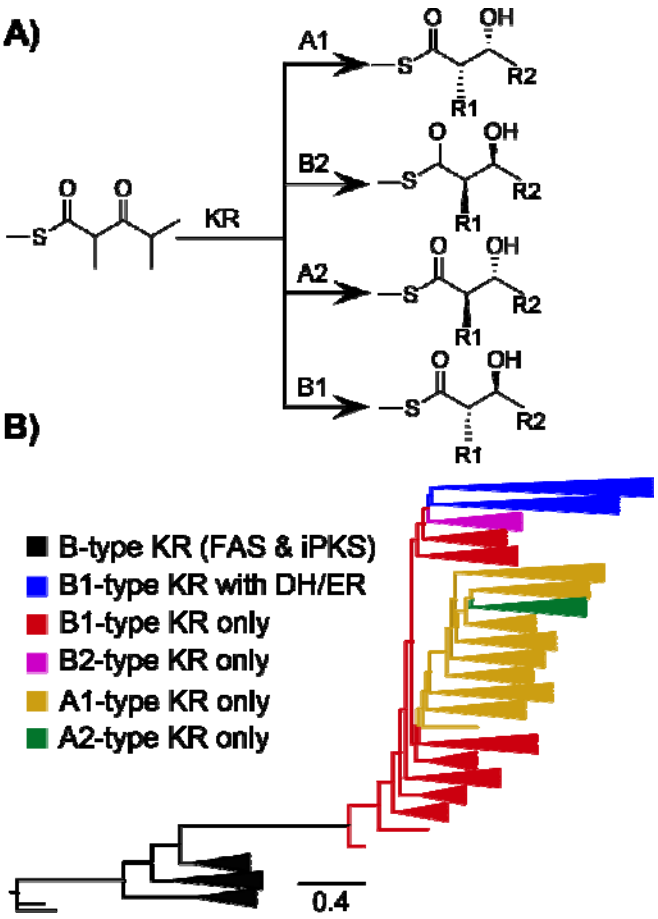


Figure 1. Bioinformatic analysis of reductive loop exchanges. **A)** KR subtypes determine the stereochemistry of the β-hydroxyl and □-carbon **B)** Phylogenetic tree of the ketoreductase (KR) domain of all manually curated KRs in ClusterCAD determined by ModelFinder in IQ-Tree.

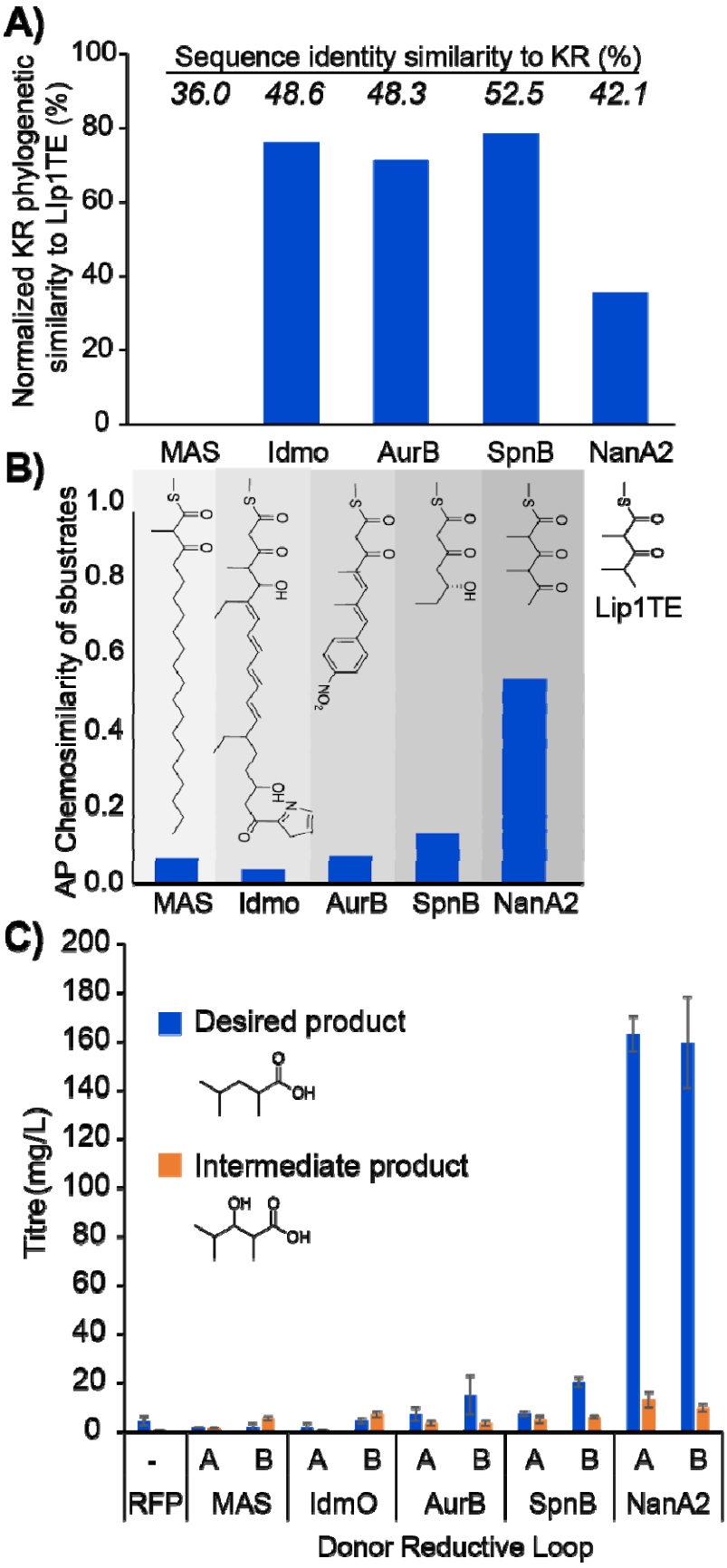

A)
Sequence identity similarity to KR (%)
36.0 48.6 48.3 52.5 42.1
Normalized KR phylogenetic similarity to Llp1TE (%)
100
80
60
40
20
0
MAS
Idmo
AurB
SpnB
NanA2
B)
AP Chemosimilarity of sbustrates
1.0
0.8
0.6
0.4
0.2
0.0
Lip1TE
MAS
Idmo
AurB
SpnB
NanA2
C)
Titre (mg/L)
200
180
160
140
120
100
80
60
40
20
0
Desired product
Intermediate product
-
RFP
A
B
MAS
A
B
IdmO
A
B
AurB
A
B
SpnB
A
B
NanA2
Donor Reductive Loop

Figure 2. Phylogenetic and chemical similarity effects on reductive loop exchanges. **A)** Phylogenetic similarity of the native Lip1 KR-only A2 subtype domain to each donor KR B1 subtype containing a DH and ER, normalized to the most similar and least similar KR domain in ClusterCad. The value above each bar denotes the sequence identity comparison. **B)** AP chemical similarity between the native Lip1 KR domain and each of the donor KR domains in this study. Chemical structures display native KR substrate in each module **C)** Polyketide production of engineered PKSs at both junction “A” and junction “B”.

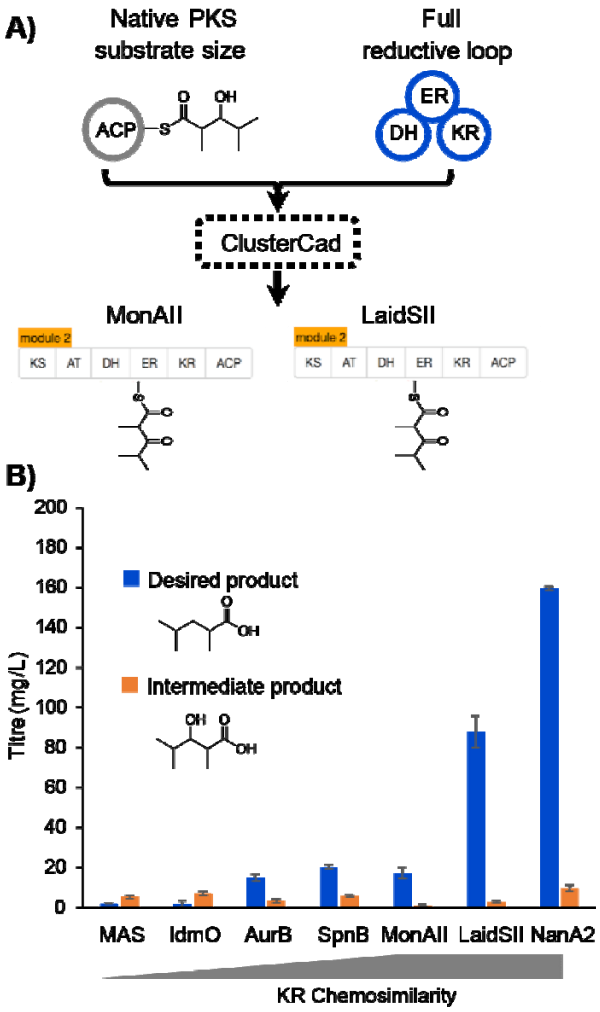


Figure 3. A chemoinformatic approach to reductive loop exchanges. **A)** A search in ClusterCad revealed the closest substrates to the first module of lipomycin containing a full reductive loop **B)** Production levels of reductive loop exchanges with the closest KR substrate similarity to LipPKS1 (MonAII, LaidSII and NanA2) compared to donor loops with less KR substrate similarity (MAS, IdmO, AurB, SpnB).