

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 Kothari¹⁻², 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. Barajas¹⁻³, 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

1 **Abstract**

2 Polyketide synthase (PKS) engineering is an attractive method to generate new molecules such
3 as commodity, fine and specialty chemicals. A central challenge in PKS design is replacing a
4 partially reductive module with a fully reductive module through a reductive loop exchange,
5 thereby generating a saturated β -carbon. In this work, we sought to establish an engineering
6 strategy for reductive loop exchanges based on chemoinformatics, a field traditionally used in
7 drug discovery. We first introduced a set of donor reductive loops of diverse genetic origin and
8 chemical substrate structures into the first extension module of the lipomycin PKS (LipPKS1).
9 These results demonstrated that chemical similarity between the substrate of the donor loops
10 and recipient LipPKS1 correlated with product titers. Consequently, we identified donor loops
11 with substrates chemically similar to LipPKS1 for further reductive loop exchanges, and we
12 observed a statistically significant correlation with production. Reductive loops with the highest
13 chemical similarity resulted in production of branched, short-chain fatty acids reaching a titer of
14 165 mg/L in *Streptomyces albus* J1074. Collectively, our work formulates a new
15 chemoinformatic paradigm for *de novo* PKS biosynthesis which may accelerate the production
16 of valuable bioproducts.
17

18 **Intro**

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

36 Chemoinformatics, an interdisciplinary field blending computational chemistry,
37 molecular modeling and statistics, was initially developed for drug discovery through analysis of
38 structure-activity relationships.⁹ Recently, we suggested that a chemoinformatic approach to
39 PKS engineering could be valuable, particularly in reductive loop exchanges due to the
40 dependence of the KR and DH domains on substrate size¹. For example, due to a hydrophobic
41 catalytic tunnel,^{10,11} acyl chain length had a critical influence on dehydration in both stand-alone
42 DH¹² and full PKS module studies.⁷ Moreover, a previous study of engineered reductive loop

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

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

69 specificity, analogous to the evolution of KS domains of cis-AT¹⁹ and trans-AT PKS modules^{21,22}
70 towards substrate specificity. As KRs from KR-DH-ER modules evolved distinctly from KR-only
71 modules, we hypothesized that the KR phylogenetic distance between the donor loops and
72 acceptor module in reductive loop exchanges was unlikely to correlate with production levels.

73 **Results and Discussion**

74 To compare the importance of chemical similarity and phylogenetic distance in reductive
75 loop exchanges, we swapped diverse reductive loops into the first module of the lipomycin PKS
76 as the acceptor module. In our previous work, we introduced a heterologous thioesterase from
77 6-deoxyerythronolide (DEBS) into the C-terminus of the first module of the lipomycin PKS
78 (denoted Lip1TE); the resulting truncated PKS produced a β -hydroxy acid.²³ In this work, our
79 experimental design was based on introducing full reductive loops using conserved residues as
80 exchange sites (denoted “A”, “B” and “C”) in Lip1TE (**Scheme 1**).⁷ We selected these conserved
81 residues based on our work in reductive loop exchanges in the first module of borreledin.⁷ To
82 evaluate the effects of genetic and chemical similarity, we identified five donor reductive loops
83 (IdmO, indanomycin, *S. antibioticus*; SpnB, spinosyn, *S. spinosa*; AurB, aureothin, *S.*
84 *aureofaciens*; NanA2, nanchangamycin, *S. nanchangensis*; MAS, mycoserosic acid, *M.*
85 *marinum*) to swap into Lip1TE. A pairwise comparison of phylogenetic distance as well as direct
86 sequence identity illustrates that the KR domain of the three donor reductive loops IdmO, SpnB,
87 and AurB are the most similar to the KR in LipPKS1 (**Figure 2A**). A similar trend also holds in
88 the analysis of the KS domain (**Supplementary Figure 3**). In contrast, the NanA2 substrate is
89 the most chemically similar to LipPKS1, followed by SpnB, based on AP similarity (**Figure 2B**)
90 and MCS similarity (data not shown). With the introduction of a reductive loop swap, the
91 chimeric enzymes would programmatically produce 2,4-dimethyl pentanoic acid. As *in vitro* PKS
92 studies have shown divergence from *in vivo* results^{24,25} due to underestimation of factors
93 including limiting substrate, crowding, and solubility,²⁶ we cloned ten chimeric modules into an

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

98 Consistent with our hypothesis, we found a strong correlation between production titers
99 of the desired product and the AP and MCS chemosimilarities of the donor and LipPKS1
100 module substrates (AP Spearman Rank Correlation of R_s of 0.99 and $p < 0.01$; MCS R_s of 0.90
101 and $p = 0.04$) (**Figure 2C**). On the other hand, no correlation between product titer and
102 phylogenetic distance or sequence similarity of the KS or KR domains was found. Based on our
103 bioinformatic analysis, this was not surprising as the lipomycin KR is an A2-type, evolving
104 separately from a KR with a full reductive loop. This trend held with either junction A or B,
105 although generally junction B chimeras resulted in higher levels of production, as demonstrated
106 in a previous study of reductive loop exchanges.⁷ We found that substituting the donor loop
107 most chemically similar to LipPKS1, NanA2, resulted in the highest titers of the desired product,
108 2,4-dimethyl pentanoic acid, reaching 165 mg/L. Low titers of the intermediate 2,5-dimethyl-3-
109 hydroxypentanoic acid were produced, which we hypothesize is due to a comparatively lower
110 rate of turnover at the energetically intensive DH domain,²⁸ resulting in premature cleavage of
111 the stalled product by hydration or by the thioesterase. As in our previous study of *in vitro*
112 production of adipic acid, we did not detect alkene or keto acid stalled products⁷. This is not
113 surprising as non-functional KRs produce short chain β -keto acids that spontaneously
114 decarboxylate to form ketones, whereas ERs have been generally shown to rapidly reduce *trans*
115 double bonds.

116 Based on these results, we took a chemoinformatic approach to further test our
117 hypothesis that chemosimilarity is a critical factor in PKS engineering. We searched the
118 ClusterCAD¹⁷ database for PKS modules with full reductive loops and substrates of high
119 chemical similarity to that of the KR of LipPKS1. The closest matches identified were PKS

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

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

146

147 **Acknowledgements**

148 This work was funded by the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by
149 the U.S. Department of Energy, Office of Science, Office of Biological and Environmental
150 Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National
151 Laboratory and the U.S. Department of Energy, the National Institute of Health Awards
152 F32GM125179, F32GM125166 and the Agile Biofoundry sponsored by the U.S. DOE Office of
153 Energy Efficiency and Renewable Energy, Bioenergy Technologies and Vehicle Technologies
154 Offices, under Contract DEAC02-05CH11231 between DOE and Lawrence Berkeley National
155 Laboratory. H.G.M. was also supported by the Basque Government through the BERC 2018-
156 2021 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM
157 Severo Ochoa excellence accreditation SEV-2017- 0718.

158 **Competing Financial Interest**

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

161

162

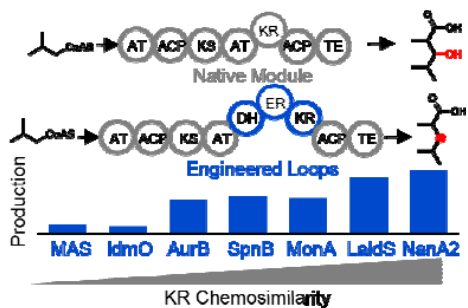
163

164

165 Bibliography

- 166 (1) Barajas, J. F.; Blake-Hedges, J. M.; Bailey, C. B.; Curran, S.; Keasling, J. D. *Synthetic*
167 *and Systems Biotechnology* **2017**, *2*, 147–166.
- 168 (2) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. *Biochemistry* **2014**, *53*, 2875–2883.
- 169 (3) Yuzawa, S.; Zargar, A.; Pang, B.; Katz, L.; Keasling, J. D. *Commodity chemicals from*
170 *engineered modular type I polyketide synthases.*; 2018; Vol. 608, pp. 393–415.
- 171 (4) Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D.
172 V.; Hutchinson, C. R.; McDaniel, R. *Biochemistry* **2003**, *42*, 72–79.
- 173 (5) Keatinge-Clay, A. *J. Mol. Biol.* **2008**, *384*, 941–953.
- 174 (6) Kellenberger, L.; Galloway, I. S.; Sauter, G.; Böhm, G.; Hanefeld, U.; Cortés, J.;
175 Staunton, J.; Leadlay, P. F. *Chembiochem* **2008**, *9*, 2740–2749.
- 176 (7) Hagen, A.; Poust, S.; Rond, T. de; Fortman, J. L.; Katz, L.; Petzold, C. J.; Keasling, J. D.
177 *ACS Synth. Biol.* **2016**, *5*, 21–27.
- 178 (8) Gaisser, S.; Kellenberger, L.; Kaja, A. L.; Weston, A. J.; Lill, R. E.; Wirtz, G.; Kendrew, S.
179 G.; Low, L.; Sheridan, R. M.; Wilkinson, B.; Galloway, I. S.; Stutzman-Engwall, K.;
180 McArthur, H. A.; Staunton, J.; Leadlay, P. F. *Org. Biomol. Chem.* **2003**, *1*, 2840–2847.
- 181 (9) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B.-T. *Mol Divers* **2006**, *10*, 39–79.
- 182 (10) Herbst, D. A.; Jakob, R. P.; Zähringer, F.; Maier, T. *Nature* **2016**, *531*, 533–537.
- 183 (11) Barajas, J. F.; McAndrew, R. P.; Thompson, M. G.; Backman, T. W. H.; Pang, B.; de
184 Rond, T.; Pereira, J. H.; Benites, V. T.; Martín, H. G.; Baidoo, E. E. K.; Hillson, N. J.;
185 Adams, P. D.; Keasling, J. D. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 1225–1235.
- 186 (12) Faille, A.; Gavalda, S.; Slama, N.; Lherbet, C.; Maveyraud, L.; Guillet, V.; Laval, F.;
187 Quémar, A.; Mourey, L.; Pedelacq, J.-D. *J. Mol. Biol.* **2017**, *429*, 1554–1569.
- 188 (13) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. *Proc.*
189 *Natl. Acad. Sci. USA* **1999**, *96*, 1846–1851.
- 190 (14) Chen, X.; Reynolds, C. H. *J Chem Inf Comput Sci* **2002**, *42*, 1407–1414.
- 191 (15) Peng, H.; Ishida, K.; Sugimoto, Y.; Jenke-Kodama, H.; Hertweck, C. *Nat. Commun.*
192 **2019**, *10*, 3918.
- 193 (16) Awakawa, T.; Fujioka, T.; Zhang, L.; Hoshino, S.; Hu, Z.; Hashimoto, J.; Kozone, I.;
194 Ikeda, H.; Shin-Ya, K.; Liu, W.; Abe, I. *Nat. Commun.* **2018**, *9*, 3534.
- 195 (17) Eng, C. H.; Backman, T. W. H.; Bailey, C. B.; Magnan, C.; García Martín, H.; Katz, L.;
196 Baldi, P.; Keasling, J. D. *Nucleic Acids Res.* **2018**, *46*, D509–D515.
- 197 (18) Keatinge-Clay, A. T. *Chem. Biol.* **2007**, *14*, 898–908.
- 198 (19) Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, I.; Kawahara, T.; Okada, M.;
199 Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.;
200 Ikeda, H.; Shin-Ya, K.; Abe, I. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 1740–1745.
- 201 (20) Jenke-Kodama, H.; Börner, T.; Dittmann, E. *PLoS Comput. Biol.* **2006**, *2*, e132.
- 202 (21) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.;
203 Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008**, *26*, 225–233.
- 204 (22) Vander Wood, D. A.; Keatinge-Clay, A. T. *Proteins* **2018**, *86*, 664–675.
- 205 (23) Yuzawa, S.; Eng, C. H.; Katz, L.; Keasling, J. D. *Biochemistry* **2013**, *52*, 3791–3793.
- 206 (24) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. *Annu. Rev. Biochem.*
207 **2007**, *76*, 195–221.
- 208 (25) Yan, J.; Hazzard, C.; Bonnett, S. A.; Reynolds, K. A. *Biochemistry* **2012**, *51*, 9333–9341.
- 209 (26) Zotter, A.; Bäuerle, F.; Dey, D.; Kiss, V.; Schreiber, G. *J. Biol. Chem.* **2017**, *292*, 15838–
210 15848.
- 211 (27) Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.;
212 Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D. *ACS Synth. Biol.*
213 **2017**, *6*, 159–166.

214 (28) Weber, A. L. *J. Mol. Evol.* **1991**, 32, 93–100.
215
216



217

218

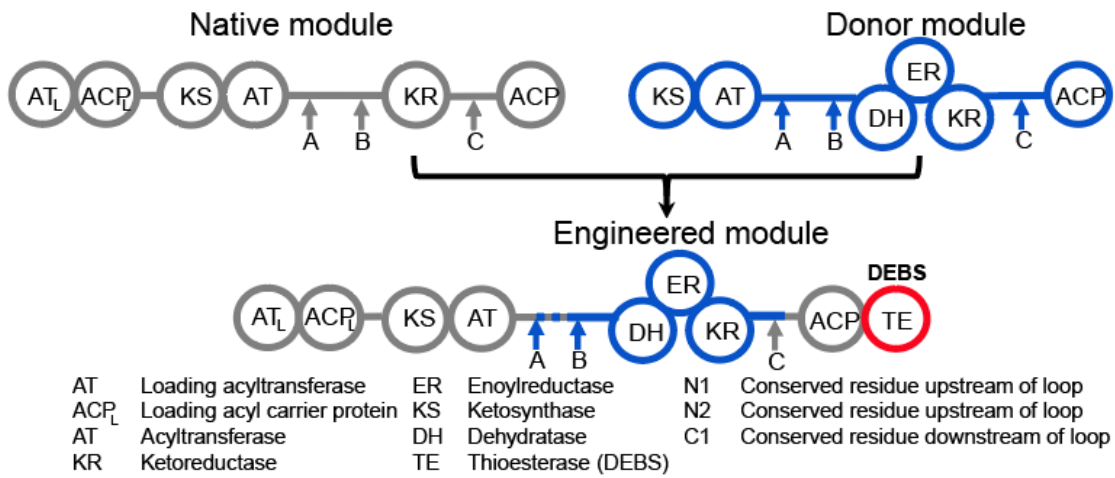
Graphical abstract

219

220

221

222



223

224

Scheme 1. Experimental design of PKS reductive loop swaps. Conserved residues are

225

identified through multiple sequence alignment surrounding the reductive domains (“A”, “B” and

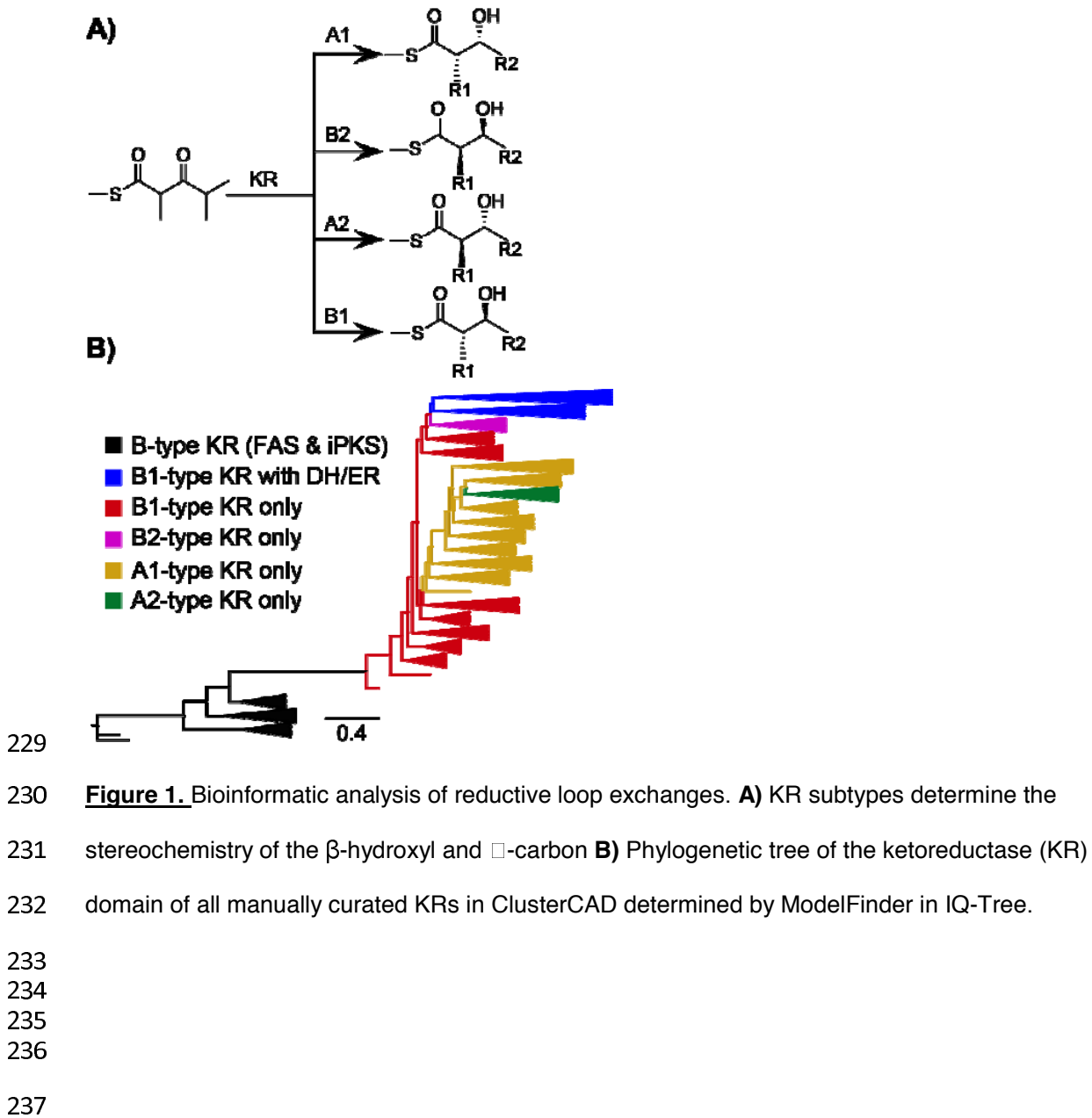
226

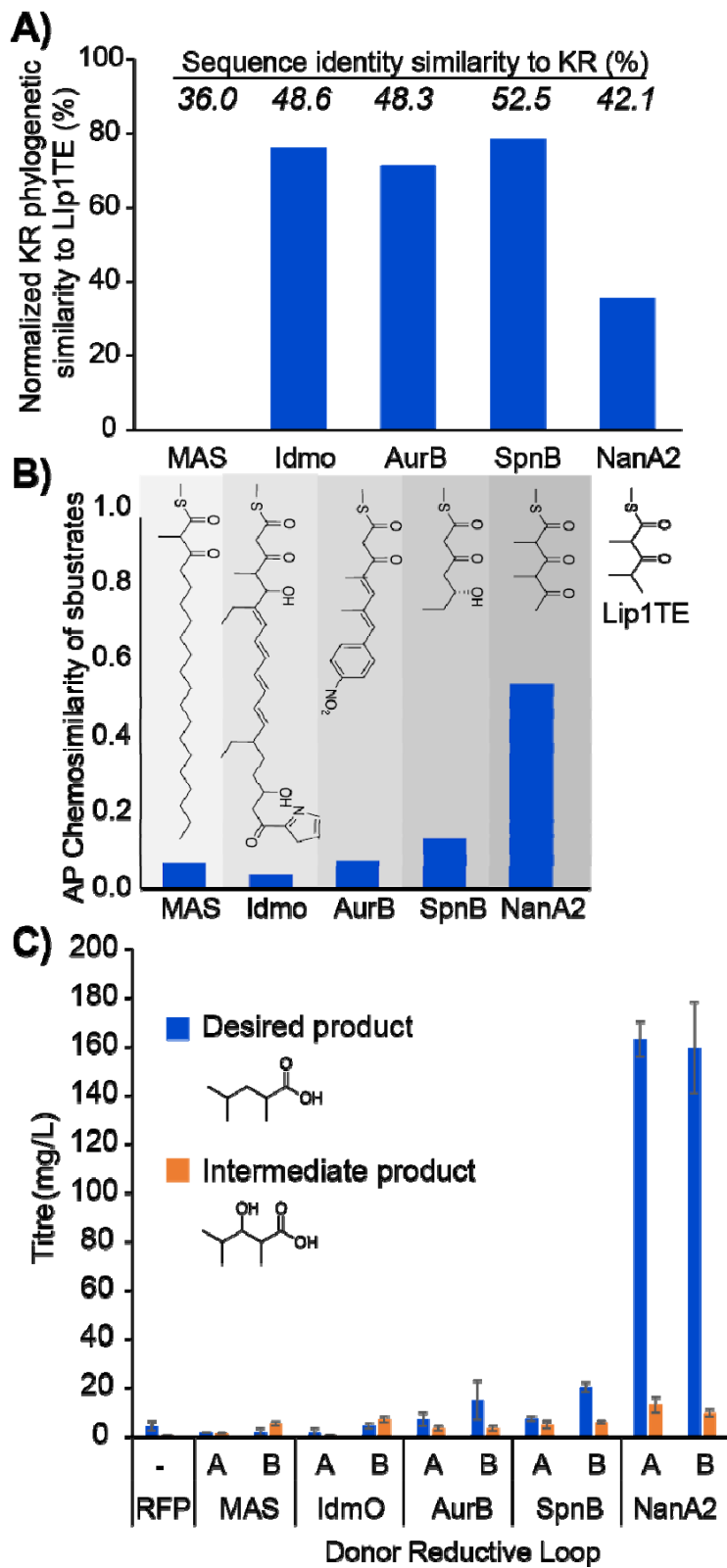
“C”). Donor reductive loops are inserted into the native lipomycin module, and the DEBS

227

thioesterase cleaves the product.

228

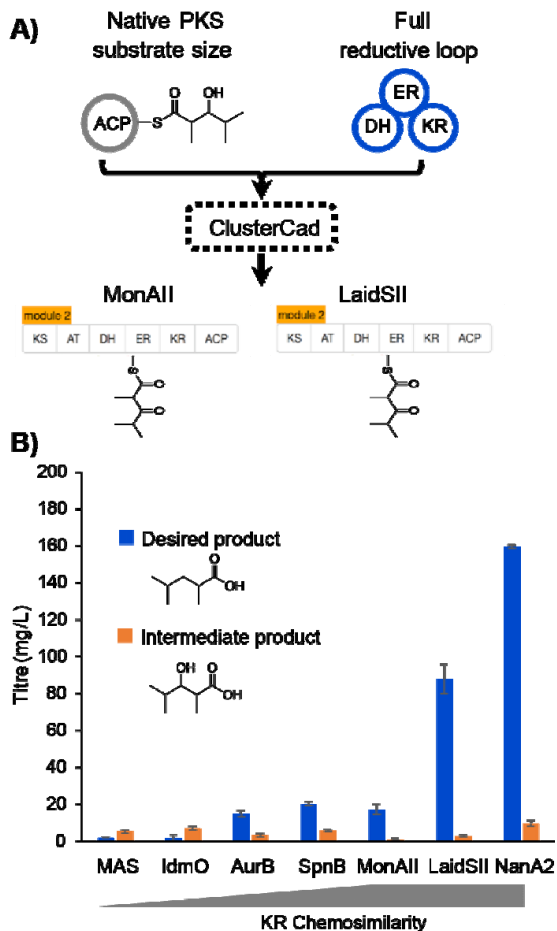




239 **Figure 2.** Phylogenetic and chemical similarity effects on reductive loop exchanges. **A)**
240 Phylogenetic similarity of the native Lip1 KR-only A2 subtype domain to each donor KR B1
241 subtype containing a DH and ER, normalized to the most similar and least similar KR domain in
242 ClusterCad. The value above each bar denotes the sequence identity comparison. **B)** AP
243 chemical similarity between the native Lip1 KR domain and each of the donor KR domains in
244 this study. Chemical structures display native KR substrate in each module **C)** Polyketide
245 production of engineered PKSs at both junction “A” and junction “B”.
246

247

248



249

250

251

252

253

254

255

256

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 (MonAll, LaidSII and NanA2) compared to donor loops with less KR

substrate similarity (MAS, IdmO, AurB, SpnB).