Final Technical Report

Project Title: Accelerating polyketide synthase engineering for high TRY production of biofuels and bioproducts

EERE award number: DE-EE0008926
Report number: DOE-UCB-08926-1
Recipient: University of California, Berkeley
Principal investigator: Jay Keasling, Tyler Backman (co-PI)
Consortium/teaming members:

- Lawrence Berkeley National Laboratory: Tyler Backman, Bob Haushalter, Christopher Petzold, Hector Garcia Martin
- Pacific Northwest National Laboratory: Young-Mo Kim, Jon Magnuson (retired from lab)
- National Renewable Energy Laboratory: Christopher W. Johnson
- Argonne National Laboratory: Philip Laible

Acknowledgment
This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Bioenergy Technology Office Award Number DE-EE0008926.

Executive Summary

Polyketide synthase (PKS) enzymes have a modular, deterministic logic that holds the potential to act as a flexible chemical factory for the biological production of a huge diversity of valuable small molecule compounds. However, engineering a custom PKS to produce a specific desired product currently requires years of trial and error, for reasons that remain poorly understood. In this project, we have developed a rapid, high throughput, Design-Build-Test-Learn (DBTL) cycle for polyketide synthases (PKSs) and demonstrate its utility for production of materials precursors. The objectives are 1) to develop a rapid, high-throughput (HT) DBTL cycle for PKSs that will enable production of a large number of unnatural, organic molecules on demand at high titer, rate, and yield (TRY); 2) to demonstrate the utility of the PKS DBTL cycle to produce three molecules: one commodity chemical (caprolactam or valerolactam) and two novel materials precursors (caprolactam or valerolactam derivatives); and 3) to demonstrate the utility of the PKS DBTL cycle to increase the TRY of one molecule (caprolactam or valerolactam). In this project, we have successfully demonstrated our high throughput PKS DBTL pipeline, and have biologically produced valerolactam and several other novel nylon monomers.

Goals and Objectives

End of Project Goal	**Outcome**	**Met/Not Met**
Production of caprolactam or valerolactam at ≥ 5 g/L, 25% of theoretical yield, & 0.1 g/l/hr from cellulosic biomass	~3 mg/L of valerolactam	Not Met
Production of 2 caprolactam or valerolactam derivatives at ≥ 0.5 g/L, 2.5% theoretical yield, & 0.01 g/l/hr from cellulosic biomass	~ 10 mg/L of methylvalerolactam ~ 2 mg/L of ethylvalerolactam ~ 1 mg/L of isobutyrylvalerolactam	Not Met
One round of DBTL cycle requires 2 ≤ months to complete and have a throughput of 500 PKSs designed, built, tested, and learned from per cycle (500 PKSs constructed & transformed into P. putida, the engineered. P. putida grown & tested for PKS function, the data analyzed, and learnings input into design software)	Few days for PKS DESIGN 3 weeks for PKS DNA BUILD 1 week for strain BUILD 1 week for strain culture & TEST 1 week for LEARN Overall, 2 ≤ months to complete one round of DBTL cycle	Met

Summary of Project Activities

Task 2.0: Design: Enhancement of ClusterCAD software

ClusterCAD is an online database of natural Polyketide Synthase (PKS) enzymes, for reuse in building novel chimeric PKSs. As part of this project, we substantially updated ClusterCAD to include a larger number of natural product gene clusters, and nonribosomal peptide synthetases (NRPSs). The latter work in a similar manner to PKSs but incorporate amino acids as extender substrates, creating peptides instead of polyketides. This is valuable in the context of this project, as amino acid building blocks from NRPSs and PKS/NRPS hybrids, necessary for creating lactams.

ClusterCAD and its associated software tools were used to construct maximally diverse libraries of PKS parts for DNA synthesis in building PKS designs used in this project. Our PKS retrobiosynthesis algorithm, also suggested a number of possible biosynthetic routes to build lactams with PKSs.

We also built a comprehensive pathway for high throughput automated design of large combinatorial PKS experiments, consisting of hundreds or thousands of designs. This includes the PKS Junction Builder Python software library which automatically applies chimeric junctions to new PKS parts, based on a multiple sequence alignment, as well as a DIVA design creation tool, which can create DIVA designs in an automated way, removing the need to manually design chimeric PKSs one at a time for large experiments. Importantly, these tools pipeline together in an automated way, that also enables the high throughput computation of machine learning features, as necessary for Task 6.

Task 3.0: Build: Develop a HT, automated platform for PKS gene construction

We were successful in building a high-throughput automated platform for PKS gene construction. This process involves PCR amplification of genes or gene fragments of interest, assembly of amplified genes and fragments into circular plasmids such that the genes are intact and in frame, and verification of the correct sequence with next generation sequencing (NGS). Lessons learned from this Task were a) the importance of performing a diagnostic colony PCR step to confirm the presence of all expected DNA fragments before submitting for NGS, and b) the importance of purifying PCR products by agarose gel extraction before plasmid assembly to remove primer dimers.

Our first PKS build campaign in this project sought to assemble 1,300 unique PKS constructs, and we were only able to confirm proper assembly of about 2% of the target plasmids. We soon found that if colonies are picked at random after plasmid assembly and transformation, a very high proportion of colonies (>95%) do not carry plasmids with the exact target sequence.

The first step we took to improve our success rate was to implement a diagnostic colony PCR step prior to NGS analysis. The NGS analysis is one of the most time and cost intensive stages of the DBTL cycle, so it is vital that we make an effort to only sequence colonies that have a high probability of success. To ensure this, colonies are pre-screened to verify that the constructs are approximately the correct length in base pairs before we sequencing.

The second step to improve our success rate was to extract and purify all PCR fragments from an agarose gel prior to plasmid assembly. We found that colonies picked at random will result in selection of many that are missing an entire target fragments. This suggests that undesired small primer dimers, which have the proper sites for recombination with the other PCR fragments in the assembly reaction, are recombining more efficiently than desired full length PCR products. Therefore, we adopted the strategy of gel excising all the full-length PCR products, thus removing primer dimers and increasing the proportion of colonies with the correct target sequence.

Ultimately, we improved the success rate of full assembly of PKS designs from less than 3% in Build 1 to 81% by Build 4 as compared to the number of initial designs made. We improved the success rate of clean constructs from less than 3% in Build 1 to 73% by Build 4. At present, a single researcher can assemble hundreds of PKSs within a 3 week timeframe.

Task 4.0: Build: Develop *Pseudomonas putida* as a host for engineered PKSs

We performed several genetic modifications to *Pseudomonas putida* at the outset of this project to prepare for the production of lactams using PKSs. We started our work with strain AG5577 from the lab of Dr. Adam Guss at Oak Ridge National Lab. This strain comprises nine distinct attB sites for serine integrase enzymes spread across three genetic loci. This allows us to quickly and easily integrate large DNA components site-specifically into the genome of the host. *P. putida* is naturally capable of catabolizing many of the small lactams we are seeking to produce, so we removed the genes *oplAB*, which catalyze hydrolysis of lactam rings to form linear amino acids. We additionally removed the gene *PP_5182* encoding a transaminase which facilitates further catabolism of linear amino acids resulting from lactam hydrolysis. Finally, we removed the genes *davAB* which can catalyze the production of valerolactam from L-lysine to ensure that any valerolactam we detect is resulting from PKS activity.

Task 5.0: Test: Grow engineered strains and assay for PKS and host function

We developed a high-throughput method to integrate chimeric PKS constructs into *P. putida* strains. After confirming the integration, we test for soluble protein expression of PKS constructs, and small molecule production titer *in vivo*.

We developed a high-throughput (HTP) Ppant ejection assay capable of analyzing over 1600 PKS samples in ten days. This assay is coupled to modular automated sample preparation procedures to enable high throughput. Protocols for the Ppant ejection assay and the sample preparation protocols are shared:
Sample preparation:
https://www.protocols.io/file manager/A76D5603EFBF11EB9EC30A58A9FEAC02#

Proteomics method for Ppant ejection assay:
https://www.protocols.io/view/targeted-proteomic-lc-ms-ms-analysis-bf9xjr7n
Demonstration data in the EDD (Experimental Data Depot):
https://public-edd.agilebiofoundry.org/s/proteomics-data-for-caprolactam-pkss-in-p-putida-b/#

Task 6.0: Develop machine learning tools for PKS design

Using the Automatic Recommendation Tool (ART) synthetic biology machine learning tool, we successfully trained and cross-validated a PKS machine learning algorithm that can predict activity of new PKS designs for production of valerolactam. This machine learning algorithm works by predicting the catalytic activity of PKS designs (as represented by experimentally measured titre) from features, computed by our design pipeline, which describe each design. We currently utilize 4 distinct features to describe each PKS design to the machine learning algorithm.

Training with the final valerocaptam data produced in Task 8 from this project from 140 different designs, we were able to predict production titre in cross-validation with R^2=0.63, indicating a strong ability to prioritize which designs will function, before building them. In the future we intend to investigate how translatable this prediction power is to other PKS systems and parts.

Task 8.0: Engineer PKS and *P. putida* to produce caprolactam, or valerolactam

We successfully engineered PKSs to produce valerolactam. To achieve this, we generated PKS enzymes where the N-terminus is comprised of the loading ACPL and KS domains from the fluviricin PKS and the AT, DH, ER, KR, and ACP domains are sourced from alternate PKS modules with substrate specificity for malonyl-CoA. We used our HTP PKS assembly pipeline to assemble over 140 unique PKS enzymes for valerolactam where we thoroughly investigated the optimal position in the protein sequences to fuse the fluviricin PKS to malonyl-CoA dependent PKSs. In addition to the engineered PKS enzyme, we also integrated the genes encoding the enzymes responsible for loading L-aspartate onto the carrier protein, FlvL; decarboxylating bound aspartate; adding a protecting group to the free amine with L-alanine; and transfer of this intermediate onto the N-terminal ACPL of the PKS. When we combine these two components in the same cell line, it is able to produce valerolactam. Our best valerolactam producing strain generated 1.4 mg/L of valerolactam.

We also pursued caprolactam production in conjunction with valerolactam. We attempted to engineer the existing fluvirucin-based PKS to accept L-glutamate rather than L-aspartate. We built and tested 21 variants, but ultimately pursued a whole PKS redesign based on mediomycin PKS. Several AT domains, PKS modules, and junctions were tested. Analytics showed promising detection of an intermediate within caprolactam biosynthesis, but further engineering will be required to make a PKS that can produce caprolactam as well as the valerolactam PKS.

Cell pellets and spent medium samples were collected from the P. putida strains at each engineering step and they were analyzed for having global proteomics and metabolomics data at PNNL. We found that nitrogen metabolism shifts and detected deficiency of amino acids in the

media during the production period of valerolactam. We proved that supplementation of amino acids helped to get better production yield.

Task 9.0: Engineer PKS and *P. putida* to produce two caprolactam or valerolactam derivatives

We successfully produced several derivatives of valerolactam, including lactams with methyl-, ethyl-, and isobutyryl- groups at the alpha-position. The first and last extension modules in the fluvirucin cluster are specific for ethylmalonyl-CoA. Therefore, when we constructed PKS enzymes where the N-terminal and C-terminal components of the fluvirucin PKS were fused, the resulting cell line produced ethylvalerolactam. Ethylmalonyl-CoA is not a native metabolite in *P. putida,* so we integrated a third cassette comprising the crotonyl-CoA carboxylase (*ccr*) gene from *Methylobacterium extorquens* AM1. Proteomics analysis revealed poor expression of the enzyme, FlvJ. Addition of a second copy of *flvJ* improved FlvJ expression and raised ethylvalerolactam titers from 1 mg/L to 3 mg/L.

We produced methylvalerolactam and isobutyrylvalerolactam using similar strategies. We used the same loading module cassette in all strains, since the beta-alanine starter unit is consistent in all the lactam pathways tested. For each valerolactam variant, we generated one or more PKS enzymes with a non-native acyltransferase (AT) domain that specifically incorporates the target acyl-CoA (methylmalonyl-CoA for methylvalerolactam and isobutyrylmalonyl-CoA for isobutyrylvalerolactam). Each of these acyl-CoAs is also non-native to *P. putida*, so each valerolactam derivative required a heterologous pathway. In the case of methylmalonyl-CoA, we integrated mutase and epimerase encoding genes (*mcm* and *epi*, respectively), which catalyze isomerization of succinyl-CoA into methylmalonyl-CoA. As expected, we found that the lactam products were only observed in cell lines where both the engineered PKS and the heterologous acyl-CoA pathways were present. FlvJ overexpression raised the titers of methyl- and isobutyrylvalerolactam similarly to ethylvalerolactam. The titers of methylvalerolactam and isobutyrylvalerolactam in our best strains were 10 mg/L and 0.6 mg/L, respectively. Curiously, the methylvalerolactam strain produced approximately 1 g/L of malonate in batch plate-based cultures and approximately 10 g/L in fed-batch bioreactors cultured at 1 L scale. This reproducible observation is due to the methylmalonyl-CoA pathway in conjunction with the PikM7 AT domain, providing a very promising strain background for malonate-derived products.

We originally started this project with the intent to make strains producing allyl-, benzyl-, and aminovalerolactams. Strains incorporating heterologous genes capable of producing each corresponding acyl-CoA were very sick. We hypothesized a lack of PKS capable of accepting the functionalized acyl-CoA created a toxic environment due to accumulation of a “dead-end” CoA. This is supported by observations in independent experiments in a different organism within the Keasling lab. Lowering the promoter strength driving expression of the acyl-CoA pathways helped reduce toxicity. Additionally, new PKS designs were made based on the already successful designs for valerolactam and its derivatives, while incorporating an AT domain with a flexible substrate specificity. Upon introduction of the new PKS designs, in conjunction with the lower expression acyl-CoA pathways, we detected promising peaks of benzylvalerolactam and quantified allylvalerolactam production at 0.013 mg/L. Across all extension modules, we found successful builds exclusively contained a CoA motif. The two pathways encoding for ACP-bound extender units were unsuccessful. This provides insight into which extenders will be

successfully incorporated into a polyketide while simultaneously providing a new engineering target to improve the versatility of extenders a customized PKS can use.

Throughout the design process, we learned that gene expression, accumulation of toxic intermediates, codon optimization, and media supplementation all played a part in successful strain engineering.

Products Developed Under the Award

Publications

- ClusterCAD 2.0: an updated computational platform for chimeric type I polyketide synthase and nonribosomal peptide synthetase design. Tao XB, LaFrance S, Xing Y, Nava AA, Martin HG, Keasling JD, Backman TWH. *Nucleic Acids Res.* 2023 Jan 6;51(D1):D532-D538. doi: 10.1093/nar/gkac1075.
- Maximizing Heterologous Expression of Engineered Type I Polyketide Synthases: Investigating Codon Optimization Strategies. Schmidt M, Lee N, Zhan C, Roberts J, Nava A, Keiser L, Vilchez A, Chen Y, Petzold C, Haushalter R, Blank L, Keasling J. *ACS Synth. Biol*. 2023, 12, 11, 3366–3380. doi: https://doi.org/10.1021/acssynbio.3c00367.

Technologies/Techniques

- Automated PKS build pipeline (as described in this document)

Inventions/Patent Applications

- N. Lee, M. Schmidt, C. Li, R. W. Haushalter, J. D. Keasling, “PKS based biosynthesis of lactam and its derivatives”, U.S. Provisional Application No: 63/504,514

Other products

- Prototype PKS junction building software tool, along with DIVA export abilities (not yet prepared for release/publication)
- ClusterCAD online PKS design database https://clustercad.jbei.org/