Previous work

Since our receipt of the first round of constructs from the DOE JGI DNA Synthesis grant in Nov 2018 we have submitted four scientific articles[1-4] (2 accepted to peer-reviewed journals, 1 submitted, 1 available on pre-print servers) and two reviews[5,6] on the topic of the *oleABCD* pathway for the biosynthesis of β-lactone natural products and hydrocarbons. A major focus of our research, based on the initial grant proposal to the DOE JGI, has been investigating OleA, the first enzyme in the biosynthetic pathway. OleA catalyzes the 'head-to-head' condensation of fatty acids and represents the first committed step in hydrocarbon biosynthesis. Importantly, the substrate specificity of OleA 'sets' the chemical composition of downstream products and therefore represents a critical bottleneck in the biosynthesis of hydrocarbons with desired branching and chain length.

OleA catalysis proceeds through a two-step mechanism of transesterification followed by carbon-carbon bond formation via Claisen condensation.[7] We established a high-throughput screening method to assess natural variation of OleA substrate specificity.[2] This screen is based on our discovery that OleA enzymes react with activated *p*-nitrophenyl esters (*p*NP) through transesterification, releasing the *p*-nitrophenolate chromophore.

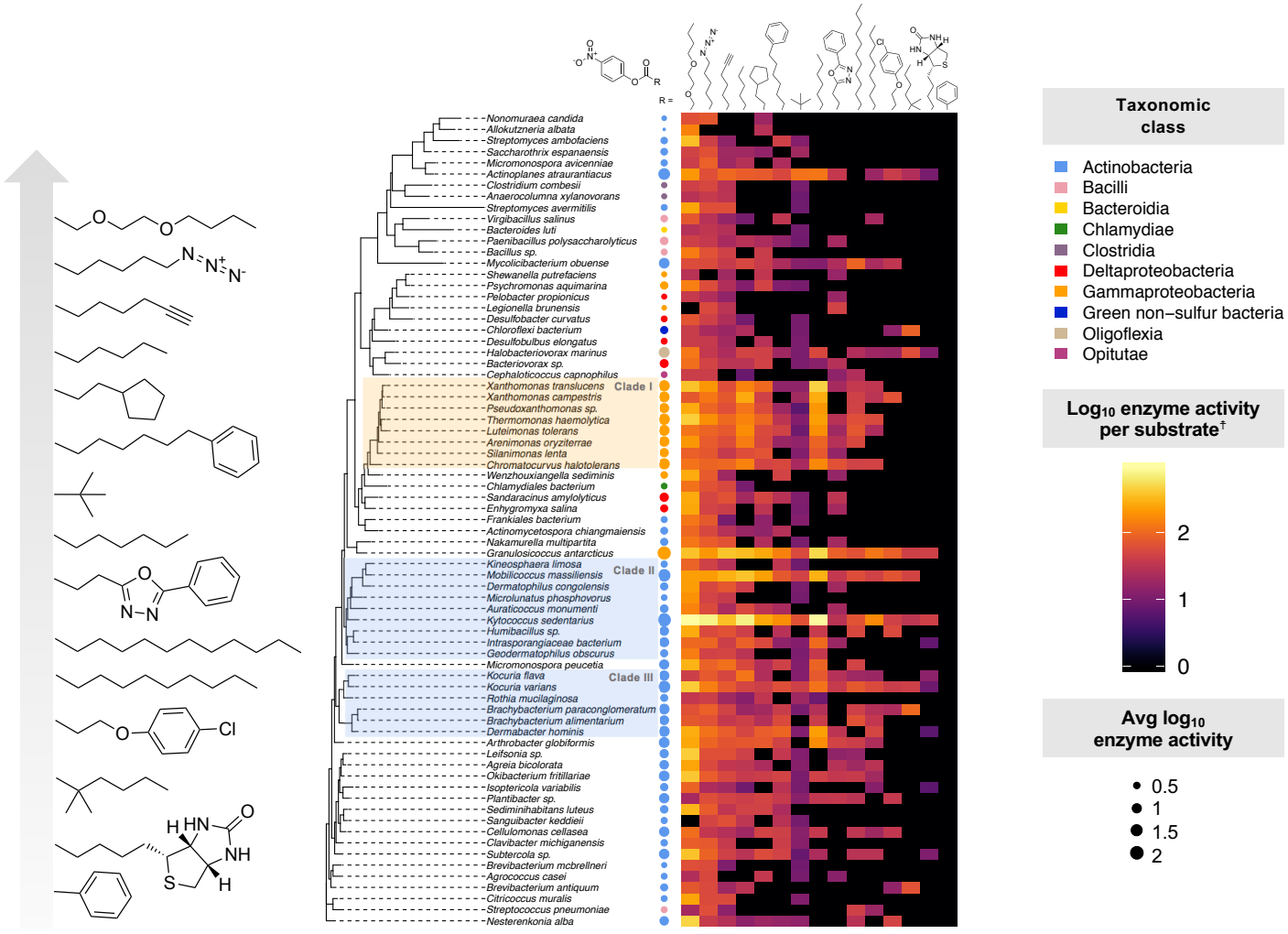

Figure 1. Approximate maximum-likelihood phylogeny of amino acid sequences of 73 OleA enzymes paired with heatmap of enzyme activity across 15 different *p*NP substrates. †Enzyme activity is measured by $\log_{10}$ of nmol *p*-nitrophenolate produced over the course of one hour by an *E. coli* BL21 culture with an OD of 1.0 per 200 µL of cells heterologously expressing OleA. *Average enzyme activity is across three biological replicates for each substrate screened. Circle sizes are scaled to average enzyme activity across all substrates. **NOTE: THIS FIGURE IS NOT FOR PUBLIC DISSEMINATION OR USE (adapted from manuscript currently submitted by us to *Synthetic Biology).***

This chromophore absorbs at a max wavelength of 410nm with a signal clearly detectable above the background of permeabilized *E. coli* cells, allowing for a whole-cell readout of the OleA transesterification reaction.

We synthesized a library of chemically-diverse *p*NPs and screened them against 73 different *oleA*s provided by the DOE JGI to map the reactivity of 1,095 distinct enzyme-substrate pairs (Figure 1).[4] By encoding the physicochemical properties of enzyme active site residues and substrates, we used machine learning to quantitatively model and predict OleA-substrate reactivity. We are now well-positioned to use these techniques to comprehensively investigate the biological relevance of OleA substrate specificity in marine ecosystems.

Proposed work

The world's oceans are the richest resource for microbially-produced polyunsaturated fatty acids and hydrocarbons. Computational predictions of the microbial biosynthetic potential for hydrocarbons in the oceans is now becoming available thanks to large-scale -omics surveys. The most comprehensive marine dataset assembled to date was collected through the TARA Oceans initiative, offering an unprecedented window into the biological and geochemical composition of marine ecosystems at 126 globally-distributed sites. In total, over 187 metatranscriptomes and 370 metagenomes were collected.[8] At the ETH Zürich, Dr. Shinichi Sunagawa and colleagues have now assembled 26,293 high-quality metagenome-assembled genomes (MAGs). The MAGs have been run through bioinformatics pipelines, including antiSMASH,[9] to predict candidate biosynthetic clusters within these MAGs. **We have formed a collaboration with Dr. Sunagawa and are currently mining this dataset.** The antiSMASH algorithm identifies all classes of polyketide synthase and fatty acid synthases including gene clusters for polyunsaturated fatty acid (PUFA) biosynthesis. PUFAs are high-value metabolites sold as dietary supplements with proven benefits for cardiovascular health. Long-chain polyunsaturated hydrocarbons (PUHCs) produced by the condensation of PUFAs also have commercial potential as **fuel additives** with relevance to the mission and current initiatives of the U.S. Department of Energy in bioenergy.

The major bacterial pathway for production of long-chain hydrocarbons including PUHCs is catalyzed by enzymes encoded in the *oleABCD* operon. A study on *oleABCD*-produced hydrocarbons from common plant-associated microbes such as *Xanthomonas* and *Stenotrophomonas* revealed up to 15 different C_{27}-C_{31} linear and branched olefin products identifiable by GC-MS.[10] The common precursor fatty acids used for hydrocarbon production in these organisms are saturated, monounsaturated and branched chain produced by type II fatty acid synthases. In contrast, many marine organisms produce a single species of PUHC, hentriacontanonaene (*n*-$C_{31:9}$), This compound was first detected in extracts of bacterial isolates from Antarctic sea ice and was proposed to be a distinctive biomarker for PUFA-producing bacteria in polar environmental samples.[11] Our group later demonstrated the same single hentriacontanonaene product is produced by common marine Gammaproteobacterial genera including *Shewanella, Chloroflexus, Colwellia, Geobacter,* and *Planctomyces.*[10]

The OleA enzyme in PUFA-producing organisms appears to be highly selective for hexadeca-4,7,10,13-tetraenoic acid despite the presence of Type II fatty acid synthase genes and intracellular concentrations of other saturated and monounsaturated long-chain fatty acids available as potential substrates. Our hypothesis that OleA specificity determines the chemical composition of pathway hydrocarbons is supported by *in vivo* evidence generated in our lab. Previously, we demonstrated that substituting the 'narrow-specificity' *Shewanella oleA* producing a single PUHC species with the 'broad-specificity' *oleA* from *Stenotrophomonas* shifted the GC-MS product profile of *Shewanella* to match that of *Stenotrophomonas* resulting in the production of >10 different chain-length hydrocarbons.[10,12] However, direct *in vitro* evidence of PUFA-associated OleA substrate specificity is lacking. Here we propose to investigate the substrate specificity of OleA enzymes from the TARA Oceans expedition with respect to co-occurrence with PUFA clusters and cold adaptation.

The majority of microorganisms with genes for PUFA and PUHC production thrive in polar marine ecosystems. Numerous studies have demonstrated that PUFAs and PUHC production increases with decreasing temperature as a result of a cellular need for increased membrane flexibility.[12,13] However, it remains unknown whether these pathway enzymes, and particularly OleA, bear molecular signatures of cold-adaptation. Cold adapted enzymes typically tend to have a high level of flexibility allowing for molecular motion in low energy environments with

higher water viscosity.[14] They also have low conformational stability which makes them more thermolabile. Some classes with cold-adapted enzymes have lower arginine-to-lysine ratio, since the guanidine group of arginines form more ionic interactions (2 salt bridges and 5 H bonds) thereby conferring higher thermostability compared to lysines.[15] We hypothesize cold-adapted, PUFA-associated OleAs have higher enzyme flexibility and a lower arginine-to-lysine ratio that may contribute to selectivity for polyunsaturated substrates.

The TARA oceans dataset described above provides the most comprehensive glimpse of the 'biosynthetic potential' of the world's oceans reported to date. Previous publications from TARA oceans data have primarily been observational rather than experimental, and have focused on taxonomic distribution. Here we propose to take a functional metagenomics approach to examine the molecular basis for substrate specificity of OleA enzymes for polyunsaturated vs. monounsaturated fatty acids in marine systems. Furthermore, we will use computational and biochemical methods to probe the relationship between substrate preferences and cold adaptation. In tandem, we will obtain important new information about the taxonomic and geospatial distribution of this hydrocarbon biosynthetic machinery in the world's oceans.

Key research questions

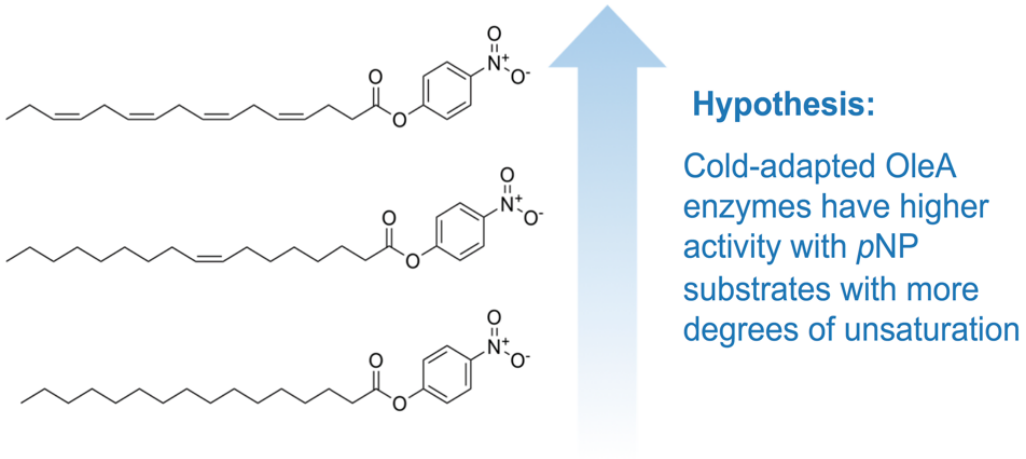

Figure 2. Graphical hypothesis

1. **What is the molecular basis for substrate selectivity in PUFA-associated OleAs?**

*p*NP substrates with increasing degrees of unsaturation will be synthesized in our lab using methods developed in our previous study (Figure 2).[4] Substrate specificity of PUFA- and non-PUFA associated OleA genes will be assessed using high-throughput *p*NP activity assay developed in our lab.[2] Random forest regression and other machine learning techniques will be evaluated for their prediction accuracy in assessing enzyme activity levels based on physicochemical protein and substrate features. Importantly, this analysis will provide quantitative insights which amino acid residues and protein properties are important in determining enzyme-substrate activity profiles. We hypothesize that higher protein flexibility and structural disorder will correlate with increasing degrees of substrate unsaturation. Hypotheses will be further investigated through molecular docking studies and site-directed mutagenesis of important residues to mechanistically examine their contribution to enzyme-substrate activity.

2. **Do OleAs from polar bacteria possess common protein signatures of cold adaptation? Is there a relationship between cold-adaptive properties and preference for polyunsaturated fatty acid substrates?**

Thermolability of purified OleAs from this study will be measured by circular dichroism. We hypothesize that OleA enzymes from polar regions will be more thermolabile as a result of increased flexibility and altered amino acid composition which can assessed through intensive homology modeling and computational methods. We hypothesize that cold-adapted OleAs will have a lower composition of order-promoting residues (V, L, I, M, F, W, and Y), which typically form the core of globular folded proteins. We predict instead cold-adapted OleAs will have higher content of polar and charged residues which tend to promote disorder (Q, S, P, E, K). We will calculate the arginine-lysine ratio to test the hypothesis that this ratio is lower in OleAs from polar regions compared to those from temperate and tropical sampling locations.

Taken together, these questions probe the relationship between cold-adaptation of OleA enzymes and their preferences for degrees of unsaturation in long-chain fatty acid substrates (Figure 2). Ultimately, we seek to understand the molecular features governing OleA substrate specificity for the biosynthesis of specialized, polyunsaturated hydrocarbons in polar environments.

Proposed workflow and current status

We have already identified over 1,967 MAGs that belong to taxonomic classes commonly possessing PUFA clusters and identified the corresponding *pfaABCDE* genes encoding the Pfa synthase within each of these groups. A more comprehensive search of uncultured bacteria from more distant phyla not known to possess PUFA clusters will also be conducted, which we anticipate may expand the number of taxa known to produce PUFAs and PUHCs. We are currently in the process of identifying *oleABCD* clusters in the TARA oceans MAGs and mapping *pfaABCDE – oleABCD* co-occurrence. From these candidate MAGs, we propose to select 45 *oleAs* which are *pfaABCDE-associated* and 45 *oleAs* which are from non-PUFA-producing organisms. Genes will be selected to maximize taxonomic diversity and cover a complete gradient of temperatures and environmental conditions in polar and nonpolar regions.

Proposed construct design and quantity

90 constructs with N-terminal His tags in pET28b+ vectors (NdeI/Xho1 cut-sites) will be designed as described above. The average length of an *oleA* gene with homology arms is 1.1 kb comes to a total of approximately **100 kb of DNA.**

Proposed Timeline

March 2020: Gene construct proposal is submitted to the JGI
Summer – Fall 2020: Gene constructs are received, experiments are conducted in the Wackett lab
Winter 2020 – Spring 2021: Data are processed manuscript is prepared and submitted for publication

Acknowledgements

We are grateful for the continued support of the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under Contract No. DE-AC02-05CH11231.

References

1. Robinson SL, Christenson JK, Richman JE, Jenkins DJ, Neres J, Fonseca DR, Aldrich CC, Wackett LP. (2019). Mechanism of a standalone β-lactone synthetase: New continuous assay for a widespread ANL superfamily enzyme. *ChemBiochem* 20(13):1701-1711.

2. Smith MD, Robinson SL, Molomjamts MM, Wackett LP. (2020). Rapid whole cell screen identifying microbial OleA thiolase enzymes involved in olefin and β-lactone biosynthesis. Accepted to *mBIO.*
3. Robinson SL, Terlouw BR, Smith MD, Pidot SJ, Stinear TP, Medema MH, Wackett LP. (2020). Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic *Nocardia*. *bioRxiv*. doi: https://doi.org/10.1101/856955.
4. Robinson SL, Smith MD, Richman JE, Aukema KG, Wackett LP. (2020). Machine learning-based prediction of activity and substrate scope for OleA enzymes in the thiolase superfamily. Submitted to *Synthetic Biology*.
5. Robinson SL, Christenson JK, Wackett LP. (2018). Biosynthesis and chemical diversity of β-lactone natural products. *Natural Product Reports.*
6. Robinson SL, Wackett LP. (2018) Rings of power: Enzymatic routes to β-lactones. Comprehensive Natural Products III: Chemistry and Biology, Enzymes and Enzyme Mechanisms: Elsevier.
7. Frias JA, Richman JE, Erickson JS, Wackett LP. (2011). Purification and characterization of OleA from *Xanthomonas campestris* and demonstration of a non-decarboxylative Claisen condensation reaction. *Journal of Biological Chemistry.* 286(13):10930-10938.
8. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, Cuenca M, Field CM, Coelho LP, Cruaud C, Engelen S, …. Sunagawa, S. (2019). Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* 179(5):1068-1083.
9. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids research*.
10. Sukovich DJ, Seffernick JL, Richman JE, Gralnick JA, Wackett LP. (2010). Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Applied and Environmental Microbiology* 76(12):3850-3862.
11. Nichols DS, Nichols PD, McMeekin TA. (1995). A new n-C-31/9 polyene hydrocarbon from Antarctic bacteria. *FEMS Microbiology Letters* 125(2-3):281-285.
12. Sukovich DJ, Seffernick JL, Richman JE, Hunt KA, Gralnick JA, Wackett LP. (2010). Structure, function, and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1. *Applied and Environmental Microbiology* 76(12):3842-3849.
13. Allemann MN, Shulse CN, Allen EE. (2019). Linkage of marine bacteria polyunsaturated fatty acid and long-chain hydrocarbon biosynthesis. *Frontiers in Microbiology* 10:10.
14. Socan J, Isaksen GV, Brandsdal BO, Aqvist J. (2019). Towards rational computational engineering of psychrophilic enzymes. *Scientific Reports* 9:10.
15. Siddiqui KS, Poljak A, Guilhaus M, De Francisci D, Curmi PMG, Feller G, D'Amico S, Gerday C, Uversky VN, Cavicchioli R. (2006). Role of lysine versus arginine in enzyme cold-adaptation: Modifying lysine to homo-arginine stabilizes the cold-adapted alpha-amylase from *Pseudoalteramonas haloplanktis*. *Proteins-Structure Function and Bioinformatics* 64(2):486-501.