Engineering Bacteria for Bio-Based Chemicals Production and Gut Microbiota Applications

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:39947209

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Engineering Bacteria for Bio-Based Chemicals Production and Gut Microbiota Applications

A dissertation presented

by

Marika Ziesack

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

August 2018
Engineering Bacteria for Bio-Based Chemicals Production and Gut Microbiota Applications

Abstract

Microbes have been used for production of food and beverages for thousands of years. In the last century we have begun to modify them by means of genetic engineering to produce more complex compounds and we have gained knowledge of how our own microbiome influences our body in health and disease. In my dissertation I contribute together with my colleagues and collaborators to the domestication of microbes for both production of bio-based chemicals and microbiota applications through synthetic biology. In chapter 2 I will discuss our efforts to improve production of fatty acid derived chemicals by protein engineering. We generated chimeric thioesterases with novel functions and gained mechanistic insight into their microbial productivity and chain-length specificity. In chapter 3 I will discuss our work on engineering four gut dwelling bacterial species to cooperate via metabolite cross-feeding. We showed that introducing these positive interactions in synthetic consortia generates bacterial composition that is characterized my evenness and robust towards perturbations in vitro. We further demonstrate that the engineered consortia composition is more even in the animal gut compared to the WT consortia. Lastly, in chapter 4 I will explore the genome of a natural E. coli isolate that is genetically tractable and colonizes the mouse gut well, with the aim to identify characteristics that aid in gut colonization. The field of metabolic engineering branches into two main tasks of improving production of bio-based chemicals and understanding and ultimately modifying the microbiome for health purposes. The work done in this dissertation contributes to both of these branches by applying synthetic biology approaches.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Microbes as workhorses</td>
<td>2</td>
</tr>
<tr>
<td>Engineered microbes for bio-based chemicals production</td>
<td>4</td>
</tr>
<tr>
<td>Engineering microbial consortia</td>
<td>10</td>
</tr>
<tr>
<td>Chassis Discovery</td>
<td>16</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>21</td>
</tr>
<tr>
<td>References</td>
<td>23</td>
</tr>
<tr>
<td>Chapter 1: Chimeric Fatty Acyl-ACP Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression</td>
<td>31</td>
</tr>
<tr>
<td>Preface and Abstract</td>
<td>32</td>
</tr>
<tr>
<td>Introduction</td>
<td>34</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>37</td>
</tr>
<tr>
<td>Results</td>
<td>40</td>
</tr>
<tr>
<td>Discussion</td>
<td>50</td>
</tr>
<tr>
<td>References</td>
<td>54</td>
</tr>
<tr>
<td>Chapter 2: Engineered cooperation in heterogeneous microbial consortia introduces resilience in vitro and in vivo</td>
<td>58</td>
</tr>
<tr>
<td>Preface and Abstract</td>
<td>59</td>
</tr>
<tr>
<td>Introduction</td>
<td>60</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>62</td>
</tr>
<tr>
<td>Results</td>
<td>69</td>
</tr>
<tr>
<td>Discussion</td>
<td>89</td>
</tr>
<tr>
<td>References</td>
<td>92</td>
</tr>
<tr>
<td>Chapter 3: Escherichia coli NGF-1 – a natural gut flora isolate that is genetically tractable and colonizes the animal gut well</td>
<td>96</td>
</tr>
<tr>
<td>Preface and Abstract</td>
<td>97</td>
</tr>
<tr>
<td>Main Text</td>
<td>98</td>
</tr>
<tr>
<td>References</td>
<td>104</td>
</tr>
<tr>
<td>Discussion</td>
<td>106</td>
</tr>
<tr>
<td>Appendix A: Supplemental Material for Chapter 1</td>
<td>112</td>
</tr>
<tr>
<td>Appendix B: Supplemental Material for Chapter 2</td>
<td>120</td>
</tr>
</tbody>
</table>
Acknowledgements

Many thanks to Pam Silver and Jeff Way for their mentorship and guidance. You have made my graduate school experience educational and inspiring and I am grateful that I was able to do my thesis work with you.

Thanks to Matt Mattozzi and Avi Robinson-Mosher who were my mentors when I first started working in the Silver lab as a Master student and stayed dear friends ever since.

Thanks to my lab mates especially Joe Torella and Tyler Ford for helping me with my thioesterase work and introducing me into the field of bioenergy. To David Riglar and Travis Gibson who ventured with me into the microbiome field. To Shannon Nangle – because Viriditas! Special thanks to Andy Shumaker for teaching me how to work efficiently so that there is more time for play.

Thanks to the ladies that I am proud to call students of mine: Aashna Shah, Elena Schaefer, Kriti Lall, Emily Molloy and Maya Ordonez. I hope I was able to provide you with a little bit of the mentorship that I was lucky to receive. You always made my day a bit brighter.

Thanks to Bryan and Roberta for letting me watch their cat Blinkin which was my companion for the last two weeks of thesis writing.

Thank you to my siblings Manja and Roger Ziesack. My sister Manja for being the only person in the world who picks up the phone when I call at 1am in the morning. My brother Roger who encouraged me to go far beyond what I would have trusted myself to achieve.

Thanks to my parents who help me remember how to speak German and support me in everything I do.
Introduction
Microbes as workhorses

The first uses of microbes date back to earliest civilizations (Figure I.1). The earliest evidence for the use of fermentation was found at a 9200-year old site in Sweden where fermentation was used to preserve fish (1). Other ancient applications of microbes are in production of bread, alcoholic beverages, yoghurt or pickled preserves (2). Even though humanity has used microbes for millennia we only became aware of them in the 1850s and 1860s in great part due to the work of French biologist Loui Pasteur (3).

Since then microbes are well established industrially for production of amino acids, vitamins and proteins with amino acids being the largest sector (2). At first, only WT strains were used; production of penicillin was the first example where mutational optimizations were undertaken to improve yield (4). The field of metabolic engineering kick-started after the first report of expression of foreign genes by Cohen and Boyer in 1973 (5). In the following years molecular biology and biochemical engineering merged into the field of metabolic engineering through development of methods and tools such as gel electrophoresis, restriction endonucleases, cloning methods and DNA synthesis and sequencing (6, 7) The first industrially produced protein was insulin made by Genentech and Eli Lily in 1978. In recent years biotechnology has gained more attention for environmental applications such as biofuel and bioplastics production.

From the beginning of metabolic engineering to now, we have gained knowledge and tools to produce a range of chemicals in the food, fuel, pharma and chemical sector (8). Metabolic engineering is inherently a top down approach to enhance flux through a certain pathway by genetic modifications and detailed description of metabolic pathways (9, 10). Synthetic Biology on the other hand is a bottom up approach that applies engineering principles such as quantitative prediction and use of assembly standards to build entire metabolic
pathways (10–12). A combination of both of these approaches enables production of more complex chemicals on industrial levels.

Figure I.1 Use of microbes in human history.

Most recently, research of microbial communities in and on our body has gained interest. The Human Microbiome Project (HMP) – which was started in 2008 - reports over 1300 strains that live in and on the human body (13). The gut microbiota that is all bacteria in our gut aid us in metabolism, immunity and behavior (14, 15). The microbiota can be influenced by internal factors such as genetics and age as well as environmental factors such as antibiotic treatment and the diet. A disadvantageous change in the gut microbiota can lead to diseases such as obesity, IBD or pathogen infections. Findings like those have sparked interest in improving our understanding of the gut microbiota and devise therapies to target it’s composition for preservation of health and to treat diseases.

Synthetic biology plays a key part in the microbiome field. Tools that have been developed for bio-based chemistry production can now be applied and further developed for the microbiota. The startup Synlogic is one of the pioneers in this field and has started clinical trials with engineered bacteria to treat rare metabolic diseases (16).
Engineering microbes for bio-based chemistry and microbiota applications.

Engineered microbes help us tackle environmental and human health challenges. In part 1 I will discuss metabolic engineering techniques for bio-based chemicals production. Then I will describe lessons learned from metabolic engineering to build cooperative synthetic ecosystems for gut applications. In part 3 I will explore the diversity of chassis that are available for a diverse set of tasks.

Part 1: Engineered microbes for bio-based chemicals production

In contrast to petroleum based chemicals, bio-based chemistry relies on fermentation of microbes which can be a carbon neutral process. Petro-based chemicals are part of our everyday lives in almost every way possible: from fuel and materials to cosmetics and even as food additives (17). These compounds are breakdown products of fossil oil that are generated during refinery and further derivatised by chemistry. Most – if not all – of these chemicals could be replaced by microbially produced alternatives (Figure I.2). In some cases microbes are even better suited for generating certain compounds than chemistry because enzymes operate at low temperature (reduction of cost) and have specificity such as stereo-selectivity which most chemistry lacks (18, 19). It is difficult to revolutionize such a vast industry as the petroleum industry but a bio-based economy is on the horizon and steps towards it are being made (20).
Microbially produced compounds overview

Microbes are used to produce a range of chemically relevant compounds called platform chemicals (19). Platform chemicals can be converted into several other chemical compounds or they serve directly as additives into cosmetics or materials. Currently microbially produced platform chemicals include propanediols (21), butanediols (22), succinate and malate (23). In the case of malate biological production is not just “greener” but also more efficient due to the enzymes stereo-selective production of the L-malate enantiomer (19).

A growing field of use for microbes is material production. Biopolymers can either be produced by generating monomers microbially and then polymerizing through chemical
processes as in the case of diamines (24) and lactic acid (25) or by production of ready-to-go polymers in microbial systems such as polyhydroxyalkanoates (PHAs) (26). Lactic acid has a long history in microbial fermentation and can be produced at high yields but needs to be polymerized by energy costly processes. PHAs have the advantage that the producing microbes already synthesize the polymer and it can have various compositions conferring different properties (26).

Microbial biofuel production holds great promises for a bio-based economy. The most abundant biofuel is ethanol which is largely generated in yeast by fermentation of corn starch (27). Yeast has been developed for food and beverage production for many years and hence reaches high yields. However, ethanol as a fuel has several disadvantages such as low energy density, high volatility and hygroscopy. Higher alcohols such as isobutanol are better suited as fuel. Amino acid biosynthesis can be re-directed into isobutanol production which logically led to engineering the long-term amino acid producer *C. glutamicum* to generate isobutanol instead of valine (28). Even better fuel replacements are fatty acid methyl esters (FAMEs) which are also commercially available but at much smaller scale than ethanol. Finally, there are the so called third generation biofuels such as alkanes which are largely in laboratory stage (29).

An area of microbial production that as gained traction industrially in recent years is the food and fragrance sector. The largest market share and oldest production is the flavor enhancer L-glutamate which is produced by *Corynebacterium glutamicum* (30). Many food and flavoring compounds derive from amino acid biosynthesis pathway and can hence be generated at relatively high yields harnessing previous research. For example aromatic amino acids are precursors for violacein which is a promising antiviral and antitumor drug (31). Generally, the food and fragrance sector allows for producing more expensive products at lower titers which makes entering the market easier than for commodity chemicals.
Fatty Acid derived chemicals

Fatty acid-derived compounds are characterized by long hydrocarbon chains with none, one or more functional groups; they have a wide range of applications as biofuels, biomaterials and bio-chemicals (32). While fatty acids occur in every organism as integral part of cell membranes, the yield from WT organisms is generally too low to achieve industrial scale titers. Hence, synthetic biology concepts are being applied to improve fatty acid biosynthesis.

The vision is to optimize every step in fatty acid biosynthesis to improve production of tailored fatty acids and their derivatives with specific chain lengths and arrangements (Figure I.3). An integral part of synthetic biology is development of tools for rapid and predictable manipulation of DNA. Tools for rapid DNA synthesis and assembly such as Gibson assembly (33), Golden Gate (34) as well as DNA editing techniques such as MAGE, CAGE and CRISPR (35) are helpful tools in improving fatty acid biosynthesis. On the transcriptional level promoter variations and specific regulatory RNA sequences can be used to modify gene expression (36, 37). Translational regulation can be achieved through ribosome binding sites (38) and riboswitches (39). Using these tools titers of up to 140 g/L for C12 hydroxy fatty acids have been achieved. However most fatty acid derived chemicals are still in the range of lower g/L or even less (32).
Figure I.3 Engineering options for heterogeneous biosynthesis of tailored free fatty acids in *E. coli*. Endogenous fatty acid biosynthesis in *E. coli* is initiated by condensation of two acetyl-CoA moieties into malonyl-coA by AccACBD and subsequent transfer onto acyl carrier protein (ACP) by FabD. FabH generates precursor acetoacetyl-ACP (R = CH₃) which initiates the biosynthesis cycle. Acyl-ACP is generated through reduction, dehydration and further reduction by FabG, FabA/Z and FabI respectively. Acyl-ACP is now either elongated by 2 carbon atoms through condensation with malonyl-ACP by FabB/F or the free fatty acid can be liberated from ACP through the action of thioesterases typically originating from plant species. Possible targets for engineering of chain-length specificity are highlighted in green. Legend: "Torella et al 2013 (40); Liu et al 2016 (41); Ziesack et al. 2018 (42) (also described in chapter 1 of this thesis)."
Figure I.3 (Continued).

\[
\begin{align*}
\text{CoA} & \xrightarrow{\text{AccABCD}} \text{malonyl-CoA} \\
\text{malonyl-CoA} & \xrightarrow{\text{FabD}} \text{malonyl-ACP}^b \\
\text{ACP} & \xrightarrow{\text{FabH}} \text{acetoacetyl-ACP} \\
\text{ACP} & \xrightarrow{\text{FabG}} \text{β-ketoacyl-ACP} \\
\text{ACP} & \xrightarrow{\text{FabA/Z}} \text{trans-enzyme-acyl-ACP} \\
\end{align*}
\]

\[
\begin{align*}
\text{ACP} & \xrightarrow{\text{FabB/F}^a} \text{acyl-ACP} \\
\text{ACP} & \xrightarrow{\text{FabI}} \text{thioesterase}^c \\
\text{ACP} & \xrightarrow{\text{FabI}} \text{free fatty acid}
\end{align*}
\]
Another level of pathway optimization occurs at the protein level. The field of protein engineering aids in optimization of natural protein functions and expansion of enzyme chemistry. Using the synthetic biology toolbox new metabolic pathways can be assembled into cell hosts. However, those enzymes usually originate from different organisms and are placed out of context. This leads to low enzyme activity and promiscuity (43). Protein engineering can help fine-tune the enzymatic functions to optimize yield or even generate new functions for diverse secondary metabolites. Previous examples of pathway optimization through protein engineering include, protein scaffolds (44), RNA scaffolds (45) or transporter engineering (46).

In the Silver lab we have focused on engineering the biosynthesis enzymes for production of fatty acids of specific chain-lengths and at high yields (Figure I.3). An important enzyme in the process of free fatty acid production that determines both specificity and yield is the thioesterase. In my work I used a rational design approach to elucidate thioesterase function and engineer enzymes with novel functions.

**Part 2: Engineering microbial consortia**

Synthetic Biology is advancing the field of microbial consortia engineering. One of the first reviews on engineering microbial consortia was written in Frances Arnold’s lab in 2008 and has since inspired many researchers including myself to tackle the “New Frontier in Synthetic Biology” (47). Microbial consortia are attractive because they promise enhanced microbial production and the ability to fulfill more complex tasks through the division of labor. Some successes have been reported in bio-production including an *E. coli*, yeast consortium that produces medical compounds (48) and a co-culture of *E. coli* and fungi to degrade cellulose and produce a biofuel compound (49). A major challenge in engineering microbial consortia is to be able to control and predict their behavior. Tools to achieve control include quorum sensing, toxin production and metabolic cross-feeding (50).
In nature, microbes live as consortia characterized by robustness and the ability to fulfill complex tasks (50). An example for natural microbial consortia is our own gut microbiota (Figure I.4). It can be stable over a life time and is able to influence our body in metabolism, immunity and behavior (15). The gut microbiota and the microbial world as a whole is a slew of interacting bacteria that touch, signal chemically, transfer genes horizontally, compete or cooperate and modify their surroundings. All these interactions define how a consortia is build and what kind of functions it can fulfill is defined by the genetic make-up of its inhabitants. With increasing knowledge about microbial consortia we thrive to engineer them for our own purposes.

*The human gut microbiota – an example of natural microbial consortia*

The gut microbiota are all microbes that live in our gut. Healthy adults harbor more than 1000 bacterial species most of which belong to the phyla *Bacteroidetes* and *Firmicutes* (51). The diversity of the gut microbiota increases in early years and is stable for a life time if no major disruptions occur (52). The gut microbiome is the entirety of genes that are contained in our gut microbiota. Even though bacterial species differ between adults, the microbiome and therefore metabolic functions are largely the same (13).

In a healthy individual the gut microbiota aids in metabolism, immunity and behavior. Gut microbes can produce nutrients from sources that would otherwise be inaccessible for humans (53). Specifically *Bacteroides* spp. break down dietary fiber into short chain fatty acids (SCFA) which are both nutrient source and immune regulator (54). Certain gut microbes such as non-pathogenic clostridia “train” our immune system to accept the inherent gut microbiota while rejecting pathogenic offenders (55). In rodents, it has been shown that there is a link between the gut microbiota and the brain through endocrine and neurocrine pathways which may also apply to human physiology (56).
Figure I.4 Influence of gut microbiota on human body. Gut microbes break down inaccessible nutrients into accessible ones. Molecules secreted from gut microbes influence the immune system and neural pathways. Gut microbes prevent infection by pathogens through niche occupation and by alerting the immune system. An intricate network of microbe-microbe interactions keeps the gut microbiota stable.

Dysbiosis, the harmful imbalance of the gut microbiota is associated with many disease states. It is a chicken-and-egg problem if a disease causes dysbiosis or if the disease is caused by dysbiosis with likely it being a little bit of both. The most prominent association between gut microbiota dysbiosis and a disease is *Clostridia difficile* infection (57). A healthy microbiota protects against infection (or outbreak) of *C. difficile* by occupying the nutritional niches but also by alerting the hosts immune system. It was shown that fecal transfer from a healthy donor can alleviate *C. difficile* infection which kick-started the idea of modulating the microbiota for
treatment of a variety of diseases (58). Diseases that have now been linked with microbiota dysbiosis are IDB, cancer, allergies, a range of metabolic disorders including obesity and diabetes, cardiovascular diseases, IBS and even behavioral disorders such as autism (59). Understanding the connections between our gut microbes and these disease states will help develop treatments in future.

The microbial interaction network keeps balance of the community in place. Gut microbes can interact positively by cross-feeding metabolites (60), instructional via quorum sensing (61) and negatively through nutrient competition or bacteriocin production (62) (Figure I.5). The community is able to withstand a range of perturbations but if the perturbation - such as drastic diet change, infection, antibiotics treatment - is too large, dysbiosis can occur (63).

*Microbial interactions*

Competition is thought to be the most prevalent form of interactions within the microbial community (62). Competitive interactions are realized through “specialized metabolites” (SM) - i.e. molecules that are not involved in central metabolism and are not essential under laboratory conditions. Bacteriocins actively kill sensitive strains to clear metabolic niches (64), siderophores scavenge iron (65) and secreted organic acids can reduce the pH of the environment (66). Physical contact can also exert a negative interaction as in the case of contact dependent inhibition which is used to introduce toxic molecules into a neighboring cell (67).
Figure I.5 Modes of bacterial communication. Bacteria produce toxins to remove competitors from nutritional niches. Generally, toxins bind to strain-specific membrane proteins and cause a signaling cascade that leads to cell death. In quorum sensing bacteria secrete a signaling molecule at a constant low rate. When the bacterial density is low, the molecule diffuses and has no effect. Upon growth, bacterial density increases, the signaling molecule reaches a threshold value and induces a concerted transcriptional response in all bacteria at once. Bacteria that lack metabolism complementary to another strain can cross-feed metabolites between each other in a cooperative way.

Quorum sensing as means of communication is neither positive nor negative but rather a transfer of instructions. Quorum is the Latin word for "of whom"; so quorum sensing means "self sensing" of bacteria. Generally, small molecules or peptides are secreted and act as transcriptional regulators (68). Bacteria produce and secrete the small molecule at a constant
low rate which does not have an effect. Once a certain abundance of bacteria is reached the amount of secreted molecules passes a threshold value at which it becomes an active transcriptional regulator to exert functions such as luminescence (69) and biofilm formation (70).

Metabolite cross-feeding is a prevalent positive interaction among bacteria. Metagenomic studies show that auxotrophies are a common feature among microbes indicating that they can source metabolites from their environment (71). In a study with \textit{S. cerevisiae} and lactic acid bacteria it was found that overflow metabolism could be a mechanism for how cross-feeding can arise (72). If \textit{S. cerevisiae} encounters a nitrogen rich environment it produces amino acids as a nitrogen sink and secretes them into the media. Lactic acid bacteria were shown to utilize those amino acids thus establishing a one-way positive interaction. Another example was provided recently for the gut microbiota (60). The authors found that certain \textit{Bacteroides} species cross-feed breakdown products of complex sugars to their neighbors.

\textit{Engineering Microbial Consortia}

With knowledge gained from studying natural occurring communication we can now engineer communication between synthetic communities. Initial studies demonstrated sharing of metabolites between pairs of \textit{E. coli} strains. For example Wintermute \textit{et al.} performed a comprehensive study on \textit{E. coli} auxotrophic strains and found synergy in 17\% of 1035 tested pairs.

Later, it was shown that such exchange can sustain multiple strains. Mee \textit{et al.} devised several multi-member \textit{E. coli} consortia that cross-fed amino acids. They found that biosynthetically costly amino acids such as L-methionine, aromatic amino acids and L-arginine create stronger cooperative interactions than cheaper amino acids. In addition to pairwise interactions they also analyzed three-member consortia which included double auxotrophs.
They found that weak pairwise growth translated into weak three member consortia if such a pair was involved. This work indicated versatility of multi-member synthetic consortia based on pairwise cross-feeding of amino acids.

We extend the use of engineered metabolite sharing to a multi-species synthetic ecology. Control of multi-species bacterial consortia could open up many possibilities in the realm of bio-based chemicals production and bacterial therapeutics. In my thesis work I contributed to these efforts by engineering four different bacterial species to cross-feed amino acids. In this system each strain is a triple auxotroph and overproduces one amino acid thus generating a network of positive interactions. We studied this system both in vitro and in vivo which provided ecological insight and a proof of concept for a synthetic heterogeneous consortium.

**Part 3: Chassis Discovery**

Finding the best chassis organism for a certain task is an important challenge to ensure optimum performance. The most important aspect of a chassis organism is its genetic tractability. Synthetic biology tools analogous to *E. coli* transcription and translation control are being developed for new chassis organisms (73). The benefit of using different chassis organisms is that their specific features can be utilized. For example some bacteria are very well suited to grow at high temperatures or are able to use carbon sources that *E. coli* or other standard microbial workhorses cannot use. An example is the chemolithotroph *Cupriavidus necator* which is able to use hydrogen as an energy source to reduce carbon dioxide into biomass (74). This organism could potentially revolutionize the bio-based chemicals industry by providing a cheap alternative to plant based feedstocks.
Most engineering was (and is still) done in *E. coli*. It is the most predictable and genetically tractable organism but classic lab strains have limitations. For instance, *E. coli* is a heterotrophic gut bacterium which means it is limited to certain temperatures and requires a comparable complex feedstock. Furthermore, most *E. coli* strains have been cultivated under laboratory conditions for many generations causing it to lose the ability to successfully colonize the animal gut. Even though many genetic tools have been developed and function well in *E. coli* it is worth the effort to domesticate more specialized bacteria to enable new ways of chemicals production and bacterial therapeutics.

Engineering efforts are expanding to different less well studied organisms with useful properties. Most of such efforts are in the field of biofuel production and bioremediation applications; an example for a chassis organism in biofuel production that has gained recent interest are solventogenic clostridia due to their ability to use a large range of substrates and their ability to produce many compounds useful as biofuels, bulk chemicals and antibiotics (75). Chassis are also being developed specifically for production of drugs such as *Streptomyces* (76) or *Mycoplasma* (77) or control of the gut microbiota (78).

*Chassis for amino acid production*

The first organism that was used for amino acid production is *Corynebacterium glutamicum* which naturally produces glutamate. From this initial discovery almost 60 years ago, arose a growing industry with production volumes of up to 2.5 million tons (79). After the initial discovery, *C. glutamicum* was further engineered for production of different amino acids. Amino acid biosynthesis is a mesh of intertwined metabolic reactions that are tightly regulated transcriptionally. Hence, a holistic approach termed systems metabolic engineering was
developed for improvement of yield. Here, omics technology is used for profiling of metabolic pathways and computational modeling aids in design of efficient cellular factories.

*E. coli* also serves as chassis for amino acid production (80). The amino acid L-threonine is primarily produced in *E. coli*. The producer strain was generated through classical mutagenesis. The benefit of *E. coli* as a chassis organism is that is can be engineered to use a variety of carbon sources efficiently. An extension of this approach is the design of a minimal chassis organism with only the required functions for optimal carbon source to product conversion.

**Chassis for plastic production**

The main bioplastics currently on the market are poly-lactic acid (PLA), poly(butylene succinate) (PBS) and polyhydroxyalkanoates (PHA). Lactic acid monomers are produced by a range of bacteria and fungi with specialized carbon utilization properties (81). BASF is producing PBS commercially and developed a specialized succinate producer named *Basti succinoproducens* (82). The most popular producer of PHA is *Cupriavidus necator* which can convert up to 80% of its biomass into ready-to-go PHA polymer stored in vesicles (26).

**Chassis for health applications**

A bacterial therapeutic needs to be highly adapted to specific environments which leads to the discovery of new chassis organisms (83). One application for bacterial therapeutics is stimulation of immune system. Most bacteria in clinical trials are intracellular bacteria *Listeria monocytogenes* (84). In its attenuated form it can grow within circulating immune cells and thus
activate tumor responses. *L. monocytogenes* is engineered to express tumor specific antigens to activate the immune system in pancreatic and ovarian cancers.

Other bacteria could be used to deliver drugs into hard to reach places such as solid tumors. Attenuated *Salmonella* spp. colonize the hypoxic tumor environment when administered systemically (85). It was engineered to not be able to reproduce in the blood stream but only in the tumor by interrupting the transport of certain metabolites. In addition to that it is engineered to produce proteins that stimulate the host immune system or potentially to deliver anticancer drugs directly into the tumor. *Salmonella* like most bacteria currently in clinical trials is easy to engineer and grows well in laboratory conditions.

*Chassis for gut applications*

Applications in the gut require organisms that are non-pathogenic but able to survive in the mammalian gut. Current bacteria that are used for gut applications include *Lactococcus lactis* – a yoghurt bacterium – and *Escherichia coli* – longtime workhorse in microbiology and also naturally found in the gut (83). An important factor for microbes as therapeutics is the ability to contain them. In *Lactococcus* this was achieved via a knockout of thymine biosynthesis (86). Thymine is present within the animal gut but not outside. Hence, a reduction of titer by 10 orders of magnitude could be achieved when the bacteria left the gut.

In our lab we have discovered a well colonizing *E. coli* that we isolated from a mouse gut. While gut colonizing engineered bacteria are not yet approved by the food and drug administration (FDA) in future it may have benefits to use long-term bacterial therapeutics rather than short-term applications. In our lab we have shown that a natural *Escherichia coli* isolate is able to colonize the mouse gut well and is easy to engineer in various ways (87, 88). We
sequenced the genome of this bacteria to learn about genetic mechanisms for gut colonization and its potential for future applications in bacterial therapeutics.
Chapter Summary

The general concepts discussed in this thesis have now been introduced. My main two messages are i) Bacteria should be applied to realize a bio-based economy and preserve both our environment and health. ii) The bacteria in and on our body (our microbiome) needs to be understood so we can modulate it to maintain health and treat diseases. The following chapters discuss specific approaches to achieve these two goals.

Chapter 1: Chimeric Fatty Acyl-ACP Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression

Fatty acids are precursors for a plethora of bio-based chemicals. One part of the biosynthesis pathway that can be modified for better yield and specificity are thioesterases. In this chapter I will describe how we generated chimeric enzymes from two plant-derived thioesterases. We obtained thioesterases with non-natural functions and were able to gain mechanistic insight into microbial productivity and chain-length specificity of these enzymes.

Chapter 2: Engineered cooperation in heterogeneous microbial consortia introduces resilience in vitro and in vivo

The gut microbiota influences the human body in both health and disease. In this chapter I will discuss our work on engineering a heterogeneous consortium that cooperates by cross-feeding metabolites. The engineered positive interactions introduced a more balanced bacterial composition and caused the consortium to persist environmental challenges in vitro and in vivo. Overall this work provided a proof of concept for introducing predictable behavior of a synthetic consortium in complex environments.
Chapter 3: Escherichia coli NGF-1 – a natural gut flora isolate that is genetically tractable and colonizes the animal gut well

*E. coli* strains have been the workhorses of microbiology for generations. In this chapter I will discuss our efforts to sequence a natural *E. coli* mouse isolate which is genetically tractable and exhibits high colonization ability. The sequence provides clues for gut colonization requirements and allows for further engineering of this strain to study the gut environment and devise bacterial therapeutics.

I will close my dissertation with a discussion about my view on the development of metabolic engineering and the microbiome field. How they merge into each other and how we can ensure continuing efforts to improve protection of our environment and health.
References


16. **Synlogic.** 2018. Synlogic Doses First Subject in Phase 1 / 2a Trial of SYNB1618 for Treatment of Phenylketonuria.


34. Engler C, Marillonnet S. 2014. DNA Cloning and Assembly Methods, Chapter 9.


64. **Donia MS, Fischbach MA.** 2015. Small molecules from the human microbiota. Science (80- ) **349**.


Chapter 1

Chimeric Fatty Acyl-ACP Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression
Preface

This work was performed in concert with Nathan Rollins as well as Aashna Shah, Brendon Dusel and Gordon Webster. Nathan performed computational modeling and analysis. Aashna and Brendon constructed plasmids and performed experiments and Gordon designed constructs. This chapter is adapted from our paper published in *Applied and Environmental Microbiology*:


Abstract

Medium-chain fatty acids are commodity chemicals. Increasing and modifying the activity of thioesterases (TEs) on medium chain fatty acid-ACP esters may enable high yield microbial production of these molecules. The plant Cuphea palustris harbors two distinct TEs: *CpFatB1* (C8-specificity, lower activity) and *CpFatB2* (C14-specificity, higher activity) with 78% sequence identity. We combined structural features from these two enzymes to create several chimeric TEs, some of which showed non-natural fatty acid production profiles as measured by an enzymatic assay and GC-MS. Notably, chimera 4 exhibited improved microbial expression and increased C8 fatty acid production. This chimera led us to identify *CpFatB2*-specific amino acids between positions 219 and 272 that may lead to better microbial expression and hence improved activity. Chimera 7 produced a broad range of fatty acids, and appeared to combine a fatty acid binding pocket with long-chain specificity and an ACP interaction site that may activate fatty acid extrusion. Using homology modeling and in silico docking with ACP, we identified a “positive patch” within amino acids 162 to 218, that may direct ACP interaction and regulate
access to short-chain fatty acids. Based on this modeling, we transplanted putative ACP interaction sequences from *CpFatB1* into *CpFatB2* and created a chimeric thioesterase that produced medium-chain as well as long-chain fatty acids. Thus, engineering chimeric enzymes and characterizing their chain-length specificity and productivity suggested mechanistic insights into TE functions, and also generated thioesterases with potentially useful properties. These observations may inform rational engineering of TEs to allow alkyl chain length control.

**Importance**

Medium chain fatty acids are important commodity chemicals. These molecules are used as plastic precursors, in shampoos and other detergents, and could be used as biofuel precursors if production economics were favorable. Hydrocarbon-based liquid fuels must be optimized to have a desired boiling point, low freezing point, low viscosity, and other physical characteristics. Similarly, the solubility and harshness of detergents and flexibility of plastic polymers can be modulated. The length and distribution of carbon chains in the alkyl chains determine these properties. Biological synthesis of cell membranes and fatty acids primarily produces chains of 16-18 carbons, which give rise to current biofuels. The ultimate goal of work presented here is to engineer metabolic pathways to produce designer molecules with the right number of carbons in a chain, so that such molecules could be used directly as specialty commodity chemicals or as fuels after minimal processing.
Introduction

Fatty acids are commodity chemicals with applications as detergents, plastic precursors, and biofuel precursors (1–6). Normally bacterial fatty acids reach lengths of 16 or 18 carbons and are incorporated as membrane components. Fatty acids of these chain lengths are typical in hard water insoluble soap; improved lathering can be achieved through the addition of shorter-chain fatty acids. Similarly, the pliability of plastic polymers can be modulated by incorporation of fatty acids with varying chain lengths. Finally, medium chain alkanes have desirable fuel properties such as volatility and low viscosity, and C₈ to C₁₂ fatty acids would not require ‘cracking’ to serve as cost-effective fuel precursors; for example by conversion to alkanes by a fatty acid reductase and an aldehyde decarbonylase (6, 7). Medium-chain fatty acids can be produced by microbial expression of plant-derived thioesterases (TEs) alone or in combination with other heterologous engineering strategies (8–10).

TEs generate free fatty acids by hydrolyzing the thioester bond that connects the acyl-chain to the acyl-chain carrier protein (ACP) (Figure 1.1). The FatB class of TEs are proteins from plant seeds that are ~ 46 kDa in size and form homodimers (11, 12). Despite differences in activity and specificity, they all share the highly conserved Hot Dog motif (11, 13). It consists of a β-sheet holding two parallel α-helices (Figure 1.1). Several hypotheses about active residues in TEs have been proposed (14–17). Structures of microbial homologs have been determined. (During the preparation of this manuscript, the first structure of a plant derived FatB enzyme with C₁₂ specificity was determined, and found to be quite similar to bacterial homologues; (18).) These structures suggest a potential Acyl-ACP cleaving site that includes the conserved positions Asp173, Asn175, His177 and Glu211 (19) (PDB: 2OWN, 2ESS, Joint Center for Structural Genomics; 4GAK, Midwest Center for Genomics; numbering according to Figure S1.1).
Figure 1.1 FatB-type thioesterase (TE) biochemistry. 

a. Synthetic biochemical pathway integrating plant-derived TEs having different chain-length specificities with bacterial metabolism. During *E. coli* fatty acid biosynthesis the hydrocarbon chains are tethered to acyl chain carrier protein (ACP). Fab enzymes elongate the fatty acid chain, through sequential acetyl-ACP condensation, reduction, dehydration and further reduction. Heterologous TEs (here *CpFatB1* and *CpFatB2*) interrupt this elongation cycle by hydrolyzing fatty acids of different chain–lengths from ACP.

b. Ribbon model of TE dimer (PDB: 2OWN). Numbering of residues corresponds to that in Figure 1.2, based on an alignment of 2OWN, *CpFatB1*, and *CpFatB2* (Figure S1.1). Each monomer (light and dark grey) exhibits the characteristic hotdog fold of two α-helices partially wrapped in a β-sheet. The proposed fatty acid pocket is indicated by electron density that was identified in the crystal structure (space-filling, orange). Side chains are shown for the putative catalytic residues Asp219, Asn221, His223 and Glu257 (orange side chains) (13), and the key positive patch residues Arg165, Arg166, Arg192, Arg195, Arg253 (green) that may be involved in ACP binding (18). The *CpFatB2*-specific amino acid deletion (red) is also highlighted.

The determinants of chain length–specificity of FatB TEs are largely undefined. Previous work has demonstrated that mutations of single amino acids in the proposed N-terminal substrate binding domain changed the specificity of *Umbellularia californica* FatB1 from C₁₂ to C₁₄ (18, 20). In another study, exchange of the N-terminus of *Arabidopsis thaliana* FatA and
FatB indicated that this part of the protein is largely responsible for specificity despite not containing the active site residues (21). However, these studies focused on changing a short-chain specific TE into a long-chain specific enzyme. To our knowledge, no attempt has been made to achieve the reverse.

The interaction dynamic between ACP and fatty acid biosynthesis enzymes may be an important factor in chain-length determination (19–22). ACP itself may provide specificity in differentiating between shorter and longer fatty acid tails. Bacterial and plant chloroplast ACPs have a pocket that shelters part or all of the fatty acid and the pantetheine group from the aqueous phase (23). In molecular dynamics simulations of acyl-ACP starting with an ‘out’ conformation, longer fatty acid tails return inside the ACP less frequently, suggesting that the ratio of in/out states decreases with increasing fatty acid length (22). The interaction between ACP and fatty acid metabolizing enzymes is fleeting, but Nguyen et al. (24) solved the structure of an acyl-ACP/FabA trapped intermediate, and found an interaction with arginine side chains in a “positive patch” on FabA that might change the conformation of ACP and induce fatty acid extrusion from within ACP. Thus the in vs. out state of the fatty acid on ACP is dynamic and may be regulated by interaction with an enzymatic partner (25).

Fatty acid synthesis has many steps, and several enzymes in this process have acyl chain-length specificities designed to produce chains of about 16 carbons, which constitute the bulk of fatty acids in membrane lipids. Several groups have engineered proteins of fatty acid synthesis to have shortened chain-length specificity (8–10, 26–28). Ultimately it may be possible to tune the entire fatty acid synthesis pathway to produce designer molecules of practical interest. TEs will be a key enzyme in any such artificial pathway, so here we sought to understand the features in these enzymes that control chain length specificity, to enable metabolic engineers to efficiently create fatty acid products.
Materials and Methods

**Strain Construction.** Double stranded DNA encoding fragments A-F of *Cp*FatB1 and *Cp*FatB2 (Figure 1.2) were synthesized as gBlocks (Integrated DNA Technologies) and cloned into the pBAD-HisA (Thermo Fisher Scientific) vector by isothermal assembly (29). Amino acid mutations of the *Cp*FatB1 and *Cp*FatB2 gene were incorporated by PCR amplification with mutagenesis oligos. An empty pBAD-HisA vector without a FatB gene was included as control. Plasmids were sequence-verified and transformed into *E.coli* BL21(DE3) ∆fadD.

**Growth and Induction.** Fatty acid production experiments were performed in glass tubes. Individual colonies of fresh transformants were grown overnight in 1.5 mL LB with 1% glucose in 15 mL culture tubes at 37 °C, then diluted 1:20 into 2 mL or 5 mL M9 with 1% glycerol, grown at 30 °C in a roller drum until OD600 of 0.4-0.6 was reached, and induced with 1% arabinose. Cells were harvested after 24 h for enzymatic fatty acid analysis, GC-MS analysis and western blotting.

**OD and Enzymatic FFA Measurements.** The OD595 of 100 µL culture was measured in a 96-well flat-bottom, non-treated, sterile polystyrene plate (Corning) on a Victor3V 1420 Multilabel Counter (Perkin-Elmer) using 595/60-nm filter. Enzymatic FFA measurements were performed according to Torella et al. 2013 (10) with modifications as follows. 200 µL of each culture were separated into supernatant and pellet by centrifugation. FFAs in 5 µL supernatant were measured directly using the enzymatic Free Fatty Acid Half Micro Test (Roche) according to the manufacturer’s instructions, scaled down to 1/10 volume in a 96-well plate with octanoate as standard for quantification. FFAs in the pellet were extracted into ethyl acetate as follows: to 100 µL re-dissolved pellet in M9-glycerol with 1% NP-40, we added 10 µL of 10% NaCl, 10 µL glacial acetic acid and 40 µL ethyl acetate. The solution was mixed by vortexing and centrifuged for 10 min at 16000 rpm. 17 µL of the ethyl acetate phase were then evaporated using a
SpeedVac and re-dissolved in M9 glycerol, 5 µL of which were subjected to the enzymatic Free Fatty acid Half Micro Test as described above, using a myristate standard for quantification.

**Fatty Acid Identification and Quantification Using GC-MS.** Identification and Quantification of FFA were essentially performed according to Torella et al. 2013 (10). In brief, we separated FFA from culture by acidifying 400 µL culture with 50 µL 10% (wt/vol) NaCl and 50 µL glacial acetic acid, and extracting FFA into 200 µL ethyl acetate. FFA in 100 µL organic phase were then esterified in 900 µL of a 30:1 mixture of EtOH and 37% (vol/vol) HCl by incubating at 55 °C, and extracted into 500 µL hexane. Fatty ethyl esters were analyzed on an Agilent GC-MS 5975/7890 (Agilent Technologies) using an HP-5MS column (length: 30 m; diameter: either 0.25 or 0.5 mm; film: 0.25 um). We identified chain-lengths by GC retention times and mass spectra and quantified using an internal standard (800 mg/L pentadecanoate added to the culture before extraction procedure). We used known concentrations of C₆, C₈, C₁₀, C₁₂, C₁₄ and C₁₆ fatty acids to generate a standard curve and quantify the production of single fatty acid species.

**Computational fold prediction.** Lack of structures for these TEs or chimeric proteins required that the structural analysis rely on representative homology models. To generate these models, we first assembled initial structures using SWISS-MODEL to thread the CpFatB1 and CpFatB2 sequences to aligned coordinates in all structures of dimeric TEs within 25% sequence homology (30–33). We then used Rosetta to apply several consecutive rounds of random perturbations to each starting point provided by SWISS-MODEL, extending the sampling to thousands of neighboring conformations (34–36). The lowest energy perturbations of each homology model were optimized by iterative Monte Carlo adjustments to the protein backbone (37). The best scoring models were relaxed with spheroid sidechain approximations, finely backbone minimized, and then sidechain minimized with full-atom refinement. The final models were then clustered and selected by Rosetta score.
**In silico docking of ACP with modeled TEs.** We used Rosetta’s docking protocol (RosettaDock) to identify potential binding orientations of the ACP to \( Cp\text{FatB1}, \ Cp\text{FatB2}, \) and chimera 4 (38). To determine whether binding modes were particular to the loading state of the ACP, we docked solved structures of the \( E. \ coli \) ACP both empty and carrying 7-, 10-, and 14-carbon fatty acids (PDBIDs: 1T8K, 2FAD, 3EJB respectively) (23, 39, 40). Sampling was unbiased, obtained by randomizing the rotational orientation of ACP and testing random positions of the protein placed within a 45 Angstrom radius of the TE center of mass. 10,000 arrangements of ACPs and TE dimers were generated this way for each pairing, and sidechains at the ACP-TE contacting surfaces were optimized. The overall structures were scored, and we took the best scoring of each pair. Clustering those structures to identify the most successful binding modes, we inspected the largest clusters as representative binding orientations.

**Western blot for his-tagged thioesterases.** We directly measured thioesterase levels in expressing \( E. \ coli \) strains by western blotting. Cells were grown and induced as described in Growth and Induction. 5 mL cultures were spun down and resuspended in 50 uL Tris-HCl, pH 7.4, protein was quantified using Pierce BCA Protein Assay Kit (Thermo Scientific) and 25 ug were loaded in Novex™ WedgeWell™ 4-20% Tris-Glycine Gel (Invitrogen by Thermo Fisher Scientific) according to manufacturer instructions. Gel was run for 2.5 hr at 120V. Proteins were transferred to a nitrocellulose membrane using the iBlot Gel Transfer Stacks Nitrocellulose, Regular (Invitrogen by Thermo Fisher Scientific). The membrane was blocked with 4% milk in Tris Buffered Saline/Tween 20 buffer (TBS-T) at 4 °C overnight. The Anti-RNA polymerase beta antibody (EPR18704, abcam) was added in a 1:2000 dilution and incubated for 30 min followed by four washes for 5 min with TBS-T. The antibodies Anti-Rabbit IgG (H&L) HRP Conjugate (XG-6160HRP, ProSci Incorporated) and Anti-6X His tag antibody (HRP) (ab1187, abcam) were added at a 1:5000 dilution, incubated and washed as with primary antibody. The Blot was imaged using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific).
Results

Homology modeling: homodimer structure prediction of \textit{CpFatB1} and \textit{CpFatB2}

To understand structure-function relationships within the \textit{C. palustris} FatB TEs, we constructed homology models of \textit{C. palustris} FatB1 and FatB2 and the chimeric enzymes described below (Figure 1.2A). We based the models on the structures of known hotdog-fold TEs, then resampled and refined backbone and sidechain conformations using Rosetta (see Methods). \textit{Bacteroides thetaiotaomicron} acyl-ACP TE [2ESS] produced the best predictions, and the structures of \textit{Lactobacillus plantarum} putative oleoyl TE (2OWN) and \textit{Spirosoma linguale} acyl-ACP TE (4GAK) were also closely related. General validity of the gross predicted structures is supported by the fact that the highly conserved residues that are not part of the putative active site play key roles in formation of the hydrophobic core and dimer interface in the models, and by similarity of our models with the C12 thioesterase structure of Feng et al (PDB ID 5X04; (18)), which was published while this manuscript was in preparation.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.2.png}
\caption{Homology model and sequence alignment of TEs \textit{CpFatB1} and \textit{CpFatB2} used as parent sequences in this study to generate chimeric enzymes. \textit{a.} Models of monomers were generated with PDB: 2ESS as template using SWISS-MODEL and Rosetta (Methods). Fragments A-F, which were used}
\end{figure}
**Figure 1.2 (Continued).** to generate chimeras, are labeled and color-highlighted in \( \text{CpFatB1} \) (left). The electron density from structure 2OWN representing a putative alkyl chain was inserted by structure alignment and is shown as an orange space-filling model. Catalytic residue and key positive patch residue side chains are colored as in Figure 1.1B. **b.** Alignment of the amino acid sequences of \( \text{CpFatB1} \) and \( \text{CpFatB2} \). The fragments A-F that were used to generate chimeric enzymes are indicated above sequences and color-highlighted. Non-conserved amino acids are highlighted in bold.

**Design and Construction of Chimeric TEs**

The fatty acid specificity and sequence of many FatB-type TEs are known, including \( \text{CpFatB1} \) (\( \text{C}_8 \)) and \( \text{CpFatB2} \) (\( \text{C}_{14} \)) of *C. palustris* (Figure 1.2) (11, 41). Even though these two enzymes show distinct specificity and microbial activity levels, they share 78% sequence identity. We therefore hypothesized that only a small number of amino acids determine activity and specificity. We generated chimeric proteins constructed from six 50-60 amino acid long fragments of \( \text{CpFatB1} \) and \( \text{CpFatB2} \) (Figure 1.2, fragments A-F). Break-points for the fragments were placed at sequence regions with high identity between the two enzymes. We created the chimeric genes by systematically assembling and cloning segments into the arabinose-inducible expression vector pBAD/HisA (Thermo Fisher Scientific). Expression of hybrid TEs gave a variety of fatty acid synthesis phenotypes (Figure 1.3). To initially analyze FFA production of chimeric TEs when expressed in *E. coli*, we used the Free Fatty Acids Half Micro Test (Roche). The shorter hydrocarbon chains are amphiphilic and therefore cross membranes more easily than long-chain fatty acids. \( \text{C}_8 \) fatty acids are expected to occur in the supernatant whereas \( \text{C}_{14} \) fatty acids should be retained in cell pellet and membranes. We detected free fatty acids in both supernatant and pellet, then quantified their abundance using a standard curve with known concentrations of myristate and octanoate respectively. These initial results were validated by quantitation of specific free fatty acids in whole cultures as determined by GC-MS (Methods;
The final densities of the tested cultures were similar (Figure 1.3A), indicating that variations in fatty acid production is most likely due to differences in enzyme expression and/or activity.

**Figure 1.3** A typical experiment showing fatty acid production by *E. coli* expressing chimeric TEs.

**a.** Titers of total free fatty acids in the pellet and supernatant of *E. coli* (DE3) ∆fadD transformed with pBADHisA expressing TEs as were obtained by performing an enzymate-based colorimetric assay (Methods). Densities of the corresponding cultures are given as OD600 values. Fragments A-F of chimeras and WT enzymes are shown as colored bars. Measurements were obtained from three biological replicates for each TE, with error bars indicating standard deviation.

**b.** The same samples were subsequently analyzed for their free fatty acid species using GC-MS (Chimera 5 was omitted here but generated a fatty acid profile similar to *CpFatB1* in other experiments).

**c.** From these data, we inferred specific functions for distinct segments of the protein.

As expected, cells expressing *CpFatB1* yielded fatty acids only in the supernatant and when expressing *CpFatB2* yielded compounds only in the pellet (Figure 1.3). The chimeras
showed a variety of fatty acid production characteristics: Enzymes that resulted in fatty acid production exhibited \( Cp\text{FatB1-like} \) (chimeras 4-6), \( Cp\text{FatB2-like} \) (chimeras 1 and 8-10) or intermediate (chimera 7) specificities. Some chimeras did not show any fatty acid production (chimeras 2, 3).

**Increased productivity of chimera 4**

Chimera 4 maintains the specificity from \( Cp\text{FatB1} \) but has much higher microbial activity (Figure 1.3), which may make this construct useful in engineered synthesis of octanoic acid. Conversely, chimera 5 showed the specificity and activity of \( Cp\text{FatB1} \). The only difference between the two chimeras is fragment E (amino acids 220 – 272) which also contains the proposed active site loop and adjacent helices and loops (Figure 1.2). We inferred that this fragment contains polymorphisms responsible for the activity difference between \( Cp\text{FatB1} \) and \( Cp\text{FatB2} \). To quantify the total composition of fatty acids produced we used GC-MS analysis. The octanoate production of chimera 4 exceeded \( Cp\text{FatB1} \)’s production by about 6- to 13-fold, whereas essentially no \( C_{14} \) fatty acids were produced (Figure 1.3-1.4).

Protein expression levels are dramatically enhanced by \( Cp\text{FatB2} \) genetic polymorphisms between amino acids 219-272, which likely explains the enhanced fatty acid production by Chimera 4 (Figure 1.4c). This region has only six FatB1/FatB2 polymorphisms: S225N, E235Q, M237V, L242F, S249G, and S269A. To understand how these polymorphisms contribute to enhanced fatty acid production, we introduced the six \( Cp\text{FatB2} \) (favorable) residues into \( Cp\text{FatB1} \) and the six \( Cp\text{FatB1} \) (unfavorable) residues into Chimera 4, and measured free fatty acid production and protein expression levels via western blot (Figure 1.4). Three of the polymorphisms had a detectable effect within \( Cp\text{FatB1} \) on fatty acid production: S225N, E235Q and S249G (Fig 4b), while the other three did not (Figure 1.4b, Figure S1.3).
Figure 1.4 Dissection of fragment E's effect on FFA production. a. Chimera 4 contains fragment E and F from CpFatB2. Differences in fatty acid production levels are due to fragment E (Figure 1.3A). There are six amino acid substitutions between CpFatB1 and CpFatB2 in Fragment E. b. Amino acid substitutions were introduced into either CpFatB1 (converting to CpFatB2 sequence) or into Chimera 4 (converting to CpFatB1 sequence). Constructs were expressed and FFA in supernatant were measured using an enzymatic assay. Shown is the average of three biological replicas; error bars indicate standard deviation. (Substitutions of L/F249 and S/A269 had no effect.) c. From the same cultures as in b., TEs were visualized by western blot using an anti-His6 antibody. As a loading reference the blot was also probed with an anti-RNA polymerase beta antibody (RPβ).
For example, introduction of the change S225 to N increased fatty acid production about 4-fold, while the reciprocal change of N225 to S in Chimera 4 reduced fatty acid production about 2-fold. Exchanges of E235Q and S249G had analogous but less dramatic effects. Combining these three mutations within a CpFatB1 protein background had an additive effect, approaching the activity of Chimera 4. Western blot analysis indicated that the most active proteins, Chimera 4 and Chimera 4 (V237M) were clearly produced at higher levels than the less active proteins (Figure 1.4C). These results suggest that the folding of these plant-derived TEs may not be optimal in E. coli, with unfolded protein being degraded, and that Chimera 4 fortuitously has superior folding properties.

Seeking structural explanations for the phenotypes of each chimera, we constructed and examined three-dimensional homology models of CpFatB1, CpFatB2 and chimera 4. We generated initial structures from SWISS-MODEL and refined using Rosetta. Inspection of structure models of CpFatB1, CpFatB2 and Chimera 4 suggests that the most helpful mutation, S225N, may enhance formation or stability of an α-helix that extends from Asn226 to Met237. In addition, Ser/Asn225 is at the +2 position relative to the putative active site histidine, and this substitution may cause a slight conformational change in the active site and influence enzyme activity. The reasons why the polymorphisms at positions E235Q and S249G affect fatty acid production are not apparent from inspection of the structure; these variations may affect formation of incorrectly folded states that are degraded.

**Permissive and restrictive determinants of chain-length specificity**

Chimera 7 shows a novel fatty acid production phenotype that is distinct from both CpFatB1 and CpFatB2. This chimeric protein produces a range of fatty acids from C₈ to C₁₆, with C₁₂ being the most abundant species. Chimera 7 differs by 10 amino acids in fragment B (amino acids 63
– 116) from chimera 6, which produces primarily C8 fatty acids. These results suggest that the CpFatB1 polymorphisms in this region together prevent longer-chain fatty acids from entering the proposed binding pocket, while the CpFatB2 polymorphisms are more permissive to longer chains. (We also constructed numerous variant proteins with single amino acid substitutions in this region, but were unable to find single point mutations that conferred a change in specificity; Table S1). The polymorphic amino acids are in the proposed fatty acid pocket, suggesting that these polymorphisms may collectively influence the binding pocket size.

Molecular modeling and examination of homologous structures supports the notion that this region of the protein defines a fatty acid binding pocket. The related structures 2ESS and 2OWN show electron density in the putative binding pocket, consistent with this identification. Specifically, the structure 2ESS has two molecules of the crystallization agent 2-methyl-2,4-pentanediol in the pocket (Figure 1.1B), and 2OWN has unidentified electron density, modeled as a 9-carbon alkyl chain, in the same position. Taken together, these results suggest that a pocket defined amino acids Gln72, Cys79, Asn86, Val104, Arg137, Trp102, Met135 and Tyr/Trp157, Val 219 and Asn224 (in CpFatB2) binds to the alkyl chain of the fatty acid. This interpretation is supported by mutational and structural analysis of other related structures, including the recently solved C12 TE (16, 18).

Fragment C (amino acids 117 - 161) appear to define a segment of CpFatB2 that functions to restrict the substrate specificity of this protein to longer-chain fatty acyl-ACPs. Chimera 7 has a relaxed specificity and generates a range of fatty acid products, while the distribution of products from chimera 8 is much more skewed to longer products (Figure 1.3B), with C14 the main product and very little C8. Chimera 9, in which fragment D (amino acids 162 - 218) also derive from FatB2, produces a distribution of fatty acids even more skewed toward C14 and with no C8, essentially identical to that of CpFatB2 itself. Figure 1.3C summarizes these observations. We note that according to the scheme in Figure 1.3C, chimera 2, which is the
converse of chimera 7, would contain segments that restrict both long and short-chain fatty acyl substrates and would be predicted to be inactive, which it is. However, there are other reasons why this chimera might be inactive, such as improper folding. We hypothesize that features of fragments C and D may promote ACP binding in a way that, in CpFatB1, promotes extrusion of the fatty acid group, particularly for shorter chains.

Docking of ACP with CpFatB1 and CpFatB2 models. We generated ten thousand docking conformations between *E. coli* ACPs and the models of CpFatB1, CpFatB1, and chimera 4 (Methods). Out of the most energetically favorable conformations, we clustered similar structures to identify the dominant binding modes (Figure 1.5).

For example, introduction of the change S225 to N increased fatty acid production about 4-fold, while the reciprocal change of N225 to S in Chimera 4 reduced fatty acid production about 2-fold. Exchanges of E235Q and S249G had analogous but less dramatic effects. Combining these three mutations within a CpFatB1 protein background had an additive effect, approaching the activity of Chimera 4. Western blot analysis indicated that the most active proteins, Chimera 4 and Chimera 4 (V237M) were clearly produced at higher levels than the less active proteins (Figure 1.4C). These results suggest that the folding of these plant-derived TEs may not be optimal in *E. coli*, with unfolded protein being degraded, and that Chimera 4 fortuitously has superior folding properties.
Figure 1.5 ACP Docking to *CpFatB1* and *CpFatB2*. In silico docking of ACP (green) with modeled TE (blue) was done using Rosetta’s docking protocol (RosettaDock). We applied several solved structures of ACP (PDB: 1T8K, 2FAD, 2FAE, 3EFB) and sampling was unbiased. 10,000 arrangements were generated and scored, and the best 200 of each pair were clustered to identify ten representative binding orientations. The proposed fatty acid pocket and active site residues (orange) are in close proximity to ACP binding site. The expanded view shows direct interactions between “positive patch” residues K195, R255 and R256 and ACP residues. The predicted position of K195 in *CpFatB2* is superimposed in gray.

Transplantation of amino acids 162-218 (Fragment D) from FatB1 into FatB2 causes a relaxation of specificity, allowing medium-chain fatty acids to be produced (Figure 1.6). These amino acids are not part of the alkyl chain binding pocket, but do interact with ACP in the docking models. Specifically, the production of C8 and C10 fatty acids from *CpFatB2* is undetectable, while production of these fatty acids from FatB2(FatB1-Fragment D) is about 10%
of the total, and the C$_{12}$ fatty acid is only 6% of the product from $Cp$FatB2 but is 21% of the product from FatB2(FatB1-Fragment D) (Figure 1.6B). These results indicated that it is possible to modulate chain length substrate specificity without changing amino acids around the likely alkyl chain binding pocket. We hypothesize that amino acids 162-218 and also 117-161 contribute to an ACP contact site, and the polymorphisms found in $Cp$FatB1 enhance extrusion of medium-chain fatty acids from ACP.

a. Figure 1. 6 A region distinct from the fatty acid binding pocket, Fragment D, modulates FFA length specificity of $Cp$FatB1/B2. a. Chimera 7 organization. Chimera 7 includes sequences from $Cp$FatB2 that allow binding of a long-chain fatty acid (Fragment B), and sequences from $Cp$FatB1 corresponding to the ACP binding site inferred from modeling (Fragments C, D). b. Transplantation of fragment D from $Cp$FatB1 into $Cp$FatB2 widens the distribution of fatty acids produced. The entire fragment D from $Cp$FatB1 was introduced into $Cp$FatB2, proteins were expressed in *E. coli*, and fatty acid production profiles were analyzed via GC-MS. Shown are the averages of three biological replicas and their standard deviation.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$C_8$</th>
<th>$C_{10}$</th>
<th>$C_{12}$</th>
<th>$C_{14}$</th>
<th>$C_{16}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Cp$FatB1</td>
<td>24.9 ± 9.7</td>
<td>0</td>
<td>0</td>
<td>0.8 ± 1.2</td>
<td>1.9 ± 2.1</td>
</tr>
<tr>
<td>$Cp$FatB2</td>
<td>0</td>
<td>0</td>
<td>11.4 ± 0.6</td>
<td>162.9 ± 10.5</td>
<td>10.3 ± 2</td>
</tr>
<tr>
<td>Ch7</td>
<td>37.7 ± 9.4</td>
<td>15 ± 0.5</td>
<td>87.8 ± 2.4</td>
<td>34.6 ± 1.3</td>
<td>4.5 ± 1.5</td>
</tr>
<tr>
<td>B2; FragD(B1)</td>
<td>13.7 ± 9.7</td>
<td>7.3 ± 0.5</td>
<td>48.7 ± 2</td>
<td>134.8 ± 2.6</td>
<td>28.9 ± 1.8</td>
</tr>
</tbody>
</table>

b. FFA species [µg/mL]
Discussion

Plant-derived fatty acyl TEs are important tools in microbial production of fatty acids by metabolic engineering. However, the factors that determine the activity levels and alkyl chain length-specificity of these enzymes are poorly understood. Here, we used a combination of hybrid enzyme construction, site-directed mutagenesis and in silico modeling to find determinants of the expression and specificity of the C₈ and C₁₄ fatty-acyl TEs of C. palustris. These two enzymes are 78% identical at the amino acid level, yet have strikingly different substrate specificities.

One hybrid enzyme (chimera 4) has the C₈ specificity of CpFatB1 with the higher in vivo activity of CpFatB2 (Figure 1.3). This enzyme consists of amino acids 1-218 of CpFatB1 and amino acids 219-316 of CpFatB2. The enhanced activity is likely due to improved expression of the hybrid protein (Figure 1.4), which appears to be expressed in E. coli at higher levels than either CpFatB1 or CpFatB2 under identical conditions. The polymorphisms Ser/Asn225, Glu/Gln235 and Ser/Gly249 in this segment appear to additively contribute to improved activity. This hybrid enzyme may be particularly useful in microbial production of C₈ fatty acids.

The binding site for the alkyl chain of the fatty acid is likely to be a deep pocket flanked by amino acids including Cys79, Trp102, Met135, Arg137, and Tyr/Trp157. This conclusion is based on the observation that active chimeric enzymes with amino acids in this segment from CpFatB1 produce only C₈ fatty acid, while those with the amino acids from CpFatB2 can produce C₁₄ (Figure 1.3). Further evidence that this region encodes the fatty acid binding pocket comes from the structure of a bacterial TE 2ESS, which has two solvent molecules in this pocket, and the putative TE structure 2OWN, which has electron density consistent with an alkyl chain in this area. Site-directed mutagenesis studies that convert a plant C₁₂ TE to accommodate C₁₄ substrates support this interpretation (18). CpFatB1/2 polymorphisms in this region appear to allow or restrict access by long alkyl chains in the fatty acid (Figure 1.3C).
The length specificity of \( CpFatB1 \), the \( C_8 \) TE, may be explained in terms of steric exclusion of longer alkyl chains, but the mechanism by which the \( CpFatB2 \) \( C_{14} \) TE excludes shorter chains cannot be so simple. We hypothesize that \( CpFatB2 \) achieves specificity for longer chains by using the equilibrium of ACP with longer fatty acids, which favors the protruded state. ACP in bacteria and plants has a deep pocket in which an attached fatty acid preferentially resides (23). This can influence the length specificity of enzymes involved in fatty acid synthesis (8). Chan et al. (22) simulated the in-to-out and out-to-in transition rates of ACP bound to fatty acids of various lengths; they found that the longer fatty acids remain outside of the ACP for a longer period of time, and had fewer stabilizing hydrogen bonds between pantetheine and the protein when the fatty acid is in the ‘in’ state. We propose that the \( C_8 \) TE may have a mechanism for encouraging the protrusion of the fatty acid from ACP, while the \( C_{14} \) TE may rely on the natural equilibrium \( C_{14} \)-acyl-ACP to be in the ‘out’ state.

ACP is a negatively charged protein, and fatty-acyl ACPs generally interact via a “positive patch” on enzymes that process fatty acids. Nguyen et al. (24) determined the structure of the \( E. coli \) fatty acid dehydratase (FabA) with ACP and an attached fatty acid in the extruded state. In this structure, the ACP interacts directly with two arginines within a larger positive patch, and this interaction appears to modulate the structure of ACP to favor fatty acid release. The \( C_8 \) and \( C_{14} \) TEs also have a positive patch near the active site: amino acids Lys97, Arg98, Arg/Lys106, Lys132, Lys163, Lys165, Arg166, Arg169, Arg175, Arg/Gly194, Lys195, Arg255, and Arg256. Most of these amino acids lie in fragments C and D, which appear to allow or block activity on shorter-chain substrates (Figure 1.3, 1.6). The entire “positive patch” is rather large, so ACP may initially bind nonspecifically through charge-charge interactions before docking at the site that stimulates fatty acid extrusion.

Polymorphisms between \( CpFatB1 \) and \( CpFatB2 \) may change the ACP docking site to differentially promote fatty acid protrusion. Specifically, we hypothesize that the \( CpFatB2-ACP \)
interaction is less favorable than the \( Cp\text{FatB1-ACP} \) interaction, but this effect would be balanced by the greater intrinsic tendency of \( C_{14} \) acyl-ACP to remain in the ‘out’ configuration. Figure 1.7 summarizes this model. To address this hypothesis, we transplanted amino acids 162-218 (Fragment D) from \( Cp\text{FatB1} \) into \( Cp\text{FatB2} \); this segment does not contain any amino acids around the putative alkyl chain binding pocket, but does include amino acids predicted to interact with ACP. The resulting hybrid protein retains essentially full activity on \( C_{14} \) substrates, but has added activity on \( C_{8}, C_{10} \), and \( C_{12} \)-length substrates (Figure 1.6B). This observation suggests that \( Cp\text{FatB} \) TEs contain features that are permissive/restrictive for medium chain substrates and that these features are distinct from the fatty acid binding pocket and its steric restriction of long-chain substrates.

Figure 1.7 Model for structural mechanisms that determine chain-length specificity in FatB-type TEs. According to this model, \( C_{8} \) fatty acids can be cleaved by \( Cp\text{FatB1} \) (blue shape) because the positive patch (green ‘+’ signs) on this enzyme stabilizes the required ACP state (light green shape). \( Cp\text{FatB2} \) (grey shape) is unable to cleave \( C_{8} \) fatty acids because it lacks the right conformation of the
Figure 1.7 (Continued). positive patch, which is mostly guided by fragment D. CpFatB1 cannot cleave C_{14} fatty acids because its pocket is too short, causing the pantetheine group (yellow) to stick out instead of residing near the catalytic residues (orange arrow). C_{14} fatty acids are released from ACP more readily without requirement of stabilization through the positive patch. CpFatB2 contains a pocket that accepts larger fatty acids, which is governed by fragment B (dark grey). Hence, it is able to cleave C_{14} fatty acids.

In sum, we analyzed structure-function relationships that control activity and length-specificity of Cuphea palustris fatty acyl-ACP TEs. We identified sequence features of these proteins that can be modified to increase protein expression in bacteria, as well as features that may control length specificity. These observations may be useful in production of designer chemicals by metabolic engineering.
References


Chapter 2

Engineered cooperation in heterogeneous microbial consortia introduces resilience *in vitro* and *in vivo*
Preface

This work was performed in concert with Travis Gibson, John Oliver, David Riglar, Andy Shumaker, Tobias Giessen, Nick Dibenedetto, Kriti Lall and Bryan Hsu. Travis Gibson performed simulations and network inference analysis. John Oliver helped developing qPCR assay. David Riglar and Bryan Hsu advised on experimental design. Andy Shumaker and Kriti Lall helped with strain engineering. Tobias Giessen performed LC-MS. Nick Dibenedetto helped with conducting mouse experiments.

Abstract

Microbes in nature are typically members of consortia that carry out a multitude of complex functions. Cooperative interactions and feedback mechanisms between microbes are key to consortia function and robustness, yet these factors are rarely considered in the design of engineered organisms for probiotic, diagnostic or therapeutic purposes. To this end, we engineered cooperative interactions to create metabolic dependency between four gut dwelling bacterial species. Each species was made auxotrophic for three amino acids and to overproduce one amino acid to share with its neighbors. Using Bayesian network inference, we show that our engineering introduced dominant positive interactions which either reversed or neutralized existing competitive interactions in the synthetic consortia. Further, at low nutrient supplementation - when cooperation is required for growth – we demonstrate increased balance in the relative abundance, and reduced impact of growth perturbations, of engineered strains compared to wild-type consortia both in in vitro culture and in a gnotobiotic mouse model. Our findings indicate that engineered cooperation can introduce robustness in a synthetic microbial ecosystem and have implications for development of probiotic therapies.
Introduction

In nature, microbes occur as conglomerates of various species with diverse sets of genomes and metabolic capabilities, allowing for division of labor and increased robustness (1). For example, microbial consortia have been shown to withstand external perturbations such as invasion of other species, toxic compounds and nutrient sparseness (2, 3). A major driver for microbial consortia robustness is cooperative behavior through production of public goods and metabolic cross-feeding (4). Consortia robustness is correlated with balanced relative microbe abundances (community composition) and an extensive network of interactions between species (community structure with feedback) (5, 6).

Metagenomic analyses reveal nutrient auxotrophies as a prevalent feature of microbes in nature suggesting that cross-feeding might be a common mode of interaction in natural consortia (7, 8). Examples of natural microbial consortia have been described including metabolically interacting communities in soil (9) and in the mammalian gut (10). The question is why metabolic cross-feeding evolved and persists against exploitation of cheaters. It has been theorized that distributing biosynthesis of amino acids over two species – a form of functional complementarity – increases productivity of the consortium through more efficient resource utilization (11). This hypothesis was verified experimentally in a synthetic *E. coli* consortium (8). The authors demonstrated increased fitness of an amino acid cross-feeding consortium in direct comparison with a non-interacting consortium. A fitness benefit of cross-feeding consortia could explain its evolution.

Amino acid cross-feeding is an attractive way to introduce cooperation to synthetic microbial consortia. *E. coli* for example readily secretes certain amino acids upon starvation whereas other metabolites are more likely to be retained (12–14). Amino acids have been shown to play an important role in inter-species communication in natural systems (15). Ponomarova and colleagues showed that amino acids are used by *S. cerivisiae* to regulate
nitrogen overflow which leads to natural cross-feeding of amino acids to lactic acid bacteria (16). Numerous studies have engineered pairwise amino acid cross-feeding \textit{E. coli} pairs and generated quantitative models to describe their behavior (8, 17–20). Synthetic \textit{Escherichia coli} consortia have been built by removing biosynthesis genes for single amino acids in each strain and co-culturing with a complementing strain. Metabolically costly amino acids such as L-methionine, L-arginine and aromatics are well suited for building cooperative interactions within more complex communities due to the growth burden from their synthesis. However, most of these studies have been restricted to single species, especially \textit{E. coli}, with higher order multi-species cross-feeding yet to be achieved.

Here we constructed a synthetic consortium of four different bacterial species that interacts by cross-feeding four amino acids. We observe cooperative behavior experimentally and we computationally infer a network of dominant positive interactions between the engineered species. The consortium is characterized by a balanced relative microbe abundance that is stable when subjected to perturbations \textit{in vitro} and when introduced into the mouse gut.
Materials and Methods

Auxotroph engineering

For auxotroph generation in the *E. coli* NGF strain we introduced multiple knockouts using sequential P1 transduction (21) from the Keio knockout collection (22). Flip-out of kanamycin cassettes was done using pCP20 (23). In brief, for P1 transduction we prepared phage by diluting an overnight culture of the donor strain 1:100 LB with 0.2% glucose, 5 mM CaCl$_2$ and 25 mM MgCl$_2$ and incubate for 1-2 hours at 37 °C until slightly turbid. We then added 40 µL P1 lysate and continued growth for 1-3 h at 37 °C while shaking until lysed. Lysate was then filtered with 20 µm sterile filter and stored in the fridge. For transduction we harvested 2 mL overnight culture of recipient strain and re-suspended in 2 mL LB with 5 mM CaCl$_2$ and 100 mM MgSO$_4$. We then mixed 100 µL donor lysate with 100 µL recipient, incubated 30 min at 37 °C and added 200 µL sodium citrate (1 M, pH 5.5) and 1 mL LB and incubated for another 1 hr at 37 °C. Cells were harvested, re-suspended in 100 µL LB with 100 mM sodium citrate and plated on LB Kan plates (75 µg/mL). The transduced kanamycin cassette was then removed using pCP20 according to protocol. In brief, we transformed pCP20 via electroporation and transformants were selected on LB agar plates supplemented with 100 µg/mL carbenicillin grown at 30 °C. Single colonies were re-streaked on LB without drugs and incubated for 10 hours at 42 °C. From there single colonies were re-streaked on LB plates without drugs and grown overnight at 37 °C. Colonies were checked for Carbenicillin, and Kanamycin sensitivity and further confirmed via PCR at respective loci. This procedure was repeated until all knockouts were introduced.

Engineering of *S. Typhimurium* LT2 required generation of single knockout strains in *S. typhimurium* that contained pKD46 integrated into its genome which allowed for linear DNA integration using lambda red recombination (23). We then introduced the knockouts into the *S. Typhimurium* strain through sequential P22 transduction and pCP20 flipout analogous to *E. coli*.
engineering. Single knockout strains were generated by PCR amplifying a Kanamycin resistance cassette from pKD13 generating linear fragments that contained upstream and downstream homology to the gene of interest and the kanamycin cassette with FRT sequences. Fragments were introduced via electroporation and selected on LB agar plates supplemented with 50 µg/mL Kanamycin. Sequential P22 transduction and pCP20 flip-out was essentially performed as described above for P1 transduction but lysis was done overnight.

For knockout generation of both *B. theta* and *B. fragilis* we used pExchange KO vectors as described (24). Briefly, we introduced 750 bp flanking regions for genes of interest adjacent to each other into the vector. The vector contains erythromycin resistance positive marker and a thymidine kinase as counter selection marker. Cloning was done in pir+ *E. coli* strains and vectors were transferred to MFDpir for conjugation (25). Conjugation was done according to protocol with minor changes. In brief, five drops of overnight culture of *E. coli* donor was inoculated in LB supplemented with 300 µM Diamino pimelic acid (DAP) and five drops of recipient overnight culture was inoculated in 50 mL basal media. Both cultures were grown for about 2 hr (*E. coli* aerobically, *Bacteroides* spp. anaerobically) until *E. coli* culture was well turbid and *Bacteroides* culture just slightly turbid. Subsequently, 9 mL recipient and 3 mL donor were combined and spun down for 10 min at 4000 rpm together. The pellet was re-suspended in 100 µL fresh basal media with 300 µM DAP and pipetted on basal media agar plates without cysteine and supplemented with 300 µM DAP. The cells were incubated at 37 °C aerobically face up for up to 20 hr, scraped off and re-suspended in 10% glycerol. Dilutions were plated on basal-agar plates supplemented with 10 µg/mL erythromycin and incubated at 37 °C anaerobically for 2-3 days. Single colonies were re-streaked in the presence of erythromycin and grown for another 2 days. 10 single colonies were inoculated in basal media without drug and grown overnight. 500 µL of each culture was mixed, spun down and re-suspended in 10% glycerol. We then plated different dilutions on basal media plates supplemented with 5-fluoro-2-
deoxy-uridine (FuDR) (200 µg/mL) and incubated at 37 °C anaerobically for 3 days. Knockouts were verified via PCR. This procedure was repeated multiple times to obtain the multiple auxotroph strains.

**Overproducer selection**

Overproducers were generated by selecting for mutants that could grow on minimal media agar plates supplemented with anti-metabolites (E. coli: 5 mg/mL Norleucine for Met overproduction; S. Typhimurium: > 0.7 mg/mL beta-(2-thiazolyl)-DL-alanine for His overproduction; B. theta: 50 µg/mL 4-methyl tryptophan for Trp overproduction; B. fragilis: 80 µg/mL Canavinine for Arg overproduction). Single colonies that showed halos were re-streaked and overproduction was measured using a bioassay. In brief, for screening of overproducing mutants the isolated strains were grown overnight at 37 °C shaking aerobically (for E. coli and S. typhimurium) or anaerobically without agitation (for Bacteroides spp.). Supernatant was harvested, diluted 1:1 with fresh media and E. coli auxotrophs were inoculated and their growth was recorded after 24 hr. For E. coli NGF overproducers we used a S. Typhimurium auxotroph instead since its colicin production prevented the E. coli biosensor to grow. Confirmed overproducers were further quantified using LC-MS.

**LC-MS for Overproduction Measurements**

To quantitate amino acid levels in overproducer supernatants, a standard curve was recorded using freshly prepared amino acid standards dissolved in growth media (1mM, 500 uM, 100 uM, 50 uM, 10 uM of L. Methionine, L/Histidine, L-Tryptophan, L-Arginine each). To prepare for HPLC-MS analysis 0.5 mL sample or standard were added to 1.5 mL ice-cold methanol and incubated on ice for 10 min. The mixture was centrifuged for 5 min at 15,000 rpm
and 500 µL supernatant was vacuum concentrated and re-suspended in 50 µL methanol. Samples were kept on ice or at 4°C. HPLC-MS analysis of standards and extracts was carried out using an Agilent 1260 Infinity HPLC system equipped with an Agilent Eclipse Plus C18 (100 × 4.6 mm, particle size 3.5 mm, flow rate: 0.3 mL/min, solvent A: dd.H2O/0.1% (v/v) formic acid, solvent B: acetonitrile, injection volume: 4 mL) connected to an Agilent 6530 Accurate-Mass Q-TOF instrument. The following gradient was used (time/min, %B): 0, 0; 0.5, 0; 14, 100; 19, 100; 20, 0, 25, 0. The mass spectrometer was operated in positive mode and the autosampler was kept at 4°C. After HPLC-MS analysis, extracted ion current (EIC) peaks were automatically integrated using the MassHunter Workstation Software (version: B.07.00). A plot of peak area versus amino acid concentration was used to generate a linear fit.

**Sequencing**

Bacterial cultures were prepared in rich media (basal for *Bacteroides* spp. and LB for *E. coli* and S. Typhimurium). Genomic DNA (gDNA) extraction was performed using the Wizard Genomic DNA Purification Kit (Promega) according to protocol. Briefly, pellets from 1 mL bacterial overnight cultures were re-suspended in 600 µL Nucleo Lysis Solution and incubated at 80 °C for 5 min. After cooling to room temperature, 3 µL RNase Solution was added, mixed by inverting and incubated at 37 °C for 60 min. Then 200 µL Protein Precipitation Solution was added, mixed by vortexing and incubated on ice for 5 min. The lysate was centrifuged 16,000 x g for 3 min and the supernatant was transferred into 600 µL of room temperature isopropanol. The mixture was centrifuged at 16,000 x g for 2 min, supernatant was poured off and the pellet was washed with 600 µL of 70% ethanol. Ethanol was removed and the pellet was allowed to air-dry for 15 min before re-hydration with 100 µL of DNA Rehydration Solution during incubation at 65 °C for 1 hr. The extracted gDNA was sheared using Covaris DNA Shearing and the library was prepared using Kapa Biosystem DNA Hyper Prep NGS Library (Dana Faber Core MBCFL
Genomics). Sequencing was done using Illumina MiSeq Sequencing 150 bp paired End (PE150). Sequences were analyzed for SNPs using Geneious software and published genome sequences (E. coli: CP016007.1; S. typhimurium: NC_003197; B. theta: AE015928; B. fragilis: NC_016776).

**Growth and Media Conditions**

All basal media and co-culture media was pre-incubated for at least 24 hr anaerobically before use. Bacteroides spp. were inoculated from glycerol stock into basal media, grown overnight and 400 µL was inoculated in 5 mL basal and grown 2 hr anaerobically. Cells were spun down, washed twice in PBS and diluted in growth media as described for each experiment. E. coli and Salmonella were inoculated from glycerol stock into LB and grown overnight at 37 °C while shaking. 100 µL of culture was then inoculated into pre-incubated LB and grown anaerobically for 2 hr, diluted, washed in PBS and diluted into co-culture media as described. Co-culture media consisted of modified M9 salts (0.2 g/L Na₂HPO₄, 90 mg/L KH₂PO₄, 30 mg/L NH₄Cl, 15 mg/L NaCl), 1 mM MgSO₄, 10 µg/mL heme, 0.1 mM CaCl₂, 1 µg/mL Niacinamide, vitamin B12 and thiamine, 400 µg/mL L-cysteine, 0.3% bicarbonate buffer, 2.5 ng/mL vitamin K, 2 µg/mL FeSO₄·7H₂O and carbon sources and amino acid supplementation as indicated in experiments.

**Multiplex qPCR**

We designed strain specific primer/probe-fluorophore pairs according to IDT protocol (Figure S2.1). We chose strain specific genes by multiple genome alignment between the strain of interest, the other consortia members and closely related strains using Geneious. Multiplex qPCR was used to quantify each strain in co-culture by using a standard curve obtained by
plating late log phase cultures grown in rich media. In brief, each strain was grown from overnight culture for ~5 hours until about OD of 1. Cells were then counted by plating. Cultures were mixed, diluted and frozen at -80 °C for use as standard curve. Samples were diluted 1:10 in ddH$_2$O and snap-frozen in liquid nitrogen and stored at least overnight at -80 °C. Growth curve and sample were both thawed together and prepared in a 5 µL Primetime Mastermix (IDT) with 1 µL Primer/Probe mixture (final concentrations: 100 nM for primers and 50 nM for probes). The qPCR was run with the following program: 20 min at 98 µC (to boil the cells and denature gDNA), followed by 40 cycles of 60 °C and 98 °C.

**Pairwise Supernatant Experiment**

Overproducers and WT strains were grown overnight anaerobically at 37 °C in co-culture media and supernatant was harvested and sterile filtered. Auxotrophs were prepared as described in **Media and Growth conditions** and inoculated in media that contained 50% fresh co-culture media and 50% spent supernatant. OD600 was measured after 24 hr growth anaerobically at 37 °C.

**In vivo testing of Cooperation**

We used 6-8 week old male Swiss Webster germ free mice that were bred in house. For the low amino acid abundance model we fed mice low-protein diet (3% custom diet, Envigo) for 10 days prior to the experiment. There is evidence that low protein diet induces depletion of amino acids in the mouse gut (26). To prepare bacteria for gavage we grew them to mid-log phase, plated for counting and snap-froze aliquots. For gavage, aliquots were thawed, spun down and re-suspended in 200 µL for gavage. Per mouse we orally gavaged $10^7$ bacteria/mouse of each strain. After gavage, mice were transferred to opti mice cages for 10
days. Fecal samples were collected before gavage and after 10 days and snap-frozen for storage at -80 °C.

DNA was extracted using Zymobiomic 96 DNA Kit with following modification: We omitted the silicon-ATM-HRC wash. Cells were lysed in a bead beater at speed 20 for 10 min, plates were turned and lysed for another 10 min at the same speed. We added an additional 3 min incubation step for Binding Buffer and additional 5 min incubation steps when transferring to Zymo-Spin-I-96-Z plates. Elution was done in 50 µL ZymoBIOMIC DNase/RNase Free Water.

Direct multiplex probe-based qPCR was done on extracted DNA samples as described above. For standard curves we used plated overnight cultures spiked into germfree fecal samples and extracted them like the samples.
Results

Cooperative Consortia Design and Engineering

It was previously reported that cooperative interactions can introduce robustness to microbial consortia (4). Robustness in this context is defined as the ability to withstand a certain level of disturbance. To inform about our synthetic consortia design we simulated behavior of non-interacting consortia and consortia that were linked by positive interactions (Figure 2.1A). We found that the community of bacteria that is not interacting is susceptible to external disturbance which can drive community members to extinction. As evident upon simulation, unlinked consortia will tend towards a certain composition in which all bacteria are present if no disturbance is applied. However, disturbance causes one of the bacteria to die out thus changing the composition. When we linked the consortia together through as few as six positive interactions between some of the bacterial species (not all), then the composition persisted at higher levels of disturbance. However, when the disturbance becomes too large, then the single strain also died out (data not shown). Thus our simulation suggests that generating a network of positive interactions within a bacterial consortium introduces robustness towards environmental disturbances.

For our synthetic consortia design we chose four genetically tractable bacterial species that were able to survive in the animal gut in diverse niches and have varied relative abundances. Those characteristics maximize potential for downstream applications (for example therapeutic delivery or diagnosis). The first member of the synthetic consortia is \textit{Escherichia coli} NGF which was isolated from BALB/c mice; It stably colonize the mouse gut and can be engineered with standard genetic tools (27, 28). The second member is \textit{Salmonella enterica subsp enterica serovar} Typhimurium (hereafter S. Typhimurium) LT2 that was further attenuated by removing the pathogenicity islands SPI1 and SPI2 and thus did not cause any disease phenotype when administered to germ-free mice. We chose the two \textit{Bacteroides}
species Bacteroides thetaiotaomicron (hereafter B. theta) and Bacteroides fragilis (hereafter B. fragilis); both are commensals that can achieve high abundance in the mouse gut.

Figure 2. 1 Engineering positive interactions in a microbial consortia introduces resilience towards disturbance. A: Simulations show increased resilience in consortia with positive interactions. In unlinked consortia disturbance can lead to a change in the consortia composition; here one of the strains dies out. Linking the consortia strains via positive interactions keeps it stable when disturbance occurs. B: We engineered strains to positively interact by cross-feeding metabolites. Each strain was knocked
Figure 2.1 (Continued). Out for three amino acid biosynthesis pathways and mutated to overproduce one amino acid. Through this the four amino acids L-methionine (Met), L-histidine (His), L-tryptophan (Trp) and L-arginine (Arg) are cross-fed between the four strains.

We engineered each of the constituent species to depend on the other three by cross-feeding of the four metabolites L-methionine, L-histidine, L-tryptophan and L-arginine (hereafter referred to as Met, His, Trp and Arg) (Figure 2.1B). Auxotrophies for three of these amino acids were generated in each strain (E. coli: His, Trp and Arg; S. Typhimurium: Met, Trp, Arg; B. theta: Met, His, Arg; B. fragilis: Met, His, Trp), along with the ability to overproduce one amino acid in each strain (E. coli: Met; S. Typhimurium: His; B. theta: Trp; B. fragilis: Arg) (Table 2.1). E. coli and S. Typhimurium were engineered by sequential phage transduction from three single auxotroph strains. E. coli was transduced with genome fragments from BW25113 that contained insertions in argA, trpC, hisA (see methods) and S. Typhimurium with genome fragments of the same parent strain with insertions in argA, trpC, metA. Bacteroides spp. triple knockout generation utilized the pExchange-tdk vector to precisely delete metA, hisG and argF in B. theta and metA, hisG and trpC in B. fragilis.

Table 2.1 Engineered strains and genotypes. a) prevents feedback inhibition (29), b) decouples from histidine feedback inhibition (30); c) trpE, removes feedback inhibition (31); d) arginine repressor, nonfunctional (32)

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Auxotroph Genotype</th>
<th>Other Genotype</th>
<th>Overproduction Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>NGF</td>
<td>ΔargA, ΔtrpC, ΔhisA</td>
<td>ΔthiE</td>
<td>metA(I296S)</td>
</tr>
<tr>
<td>S. Typhimurium</td>
<td>LT2</td>
<td>ΔargA, ΔtrpC, ΔmetA</td>
<td>ΔthiE, ΔSPI1, ΔSPI2</td>
<td>hisG(E271K)</td>
</tr>
<tr>
<td>B. theta</td>
<td>VPI5482</td>
<td>ΔmetA, ΔhisG, ΔargF, ΔtrpC</td>
<td>ΔthiSEG, Δtdk</td>
<td>BT_0532 (A306V, N63D)</td>
</tr>
<tr>
<td>B. fragilis</td>
<td>368R</td>
<td>ΔmetA, ΔhisG, ΔtrpC</td>
<td>ΔthiSEG, Δtdk</td>
<td>BF638R_0532 (L26R)</td>
</tr>
</tbody>
</table>
Characterization of Auxotrophies and Overproduction

To assess the auxotrophic strains’ amino acid requirements we measured growth on varying concentrations of each metabolite in the presence of non-limiting concentrations of all the other metabolites (Figure 2.2A). Each strain had a requirement for specific levels of the different amino acids. Previous work has demonstrated that cross-feeding of amino acids can allow for mutual growth between two auxotrophs (7, 17). However, in our case a single strain had to feed not only one but three other bacterial species. We therefore engineered each strain to overproduce one amino acid that can be cross-fed to the other consortia members. To that end we selected for bacterial strains that showed resistance to specific antimetabolites (Table 2.1 and Figure S2.2). Overproduction was measured in comparison to a defined amino acid standard using LC-MS (Figure 2.2A, horizontal bars). In order to compare overproduction with each strain’s specific requirement we fitted a sigmoidal curve to the growth response data (Figure 2.2A). Comparison of requirement levels and overproduction levels can be an indication for stronger or weaker interactions between strains. Based on production levels we can infer to what OD600 the respective strains should grow based on their requirements.
Figure 2.2 Characterization of auxotrophs and overproducers in 4-species consortia. a. Growth response curves. Each auxotroph was grown in media supplemented with varying concentrations of one amino acid and saturating concentration of the two others. Depicted is the average of three biological replicates; error bars indicate standard deviation. A sigmoidal curve was fitted using GraphPad Prism 7.
**Figure 2.2 (Continued).** Horizontal lines indicate average overproduction values of respective strains measured via LC-MS. b. Cross-feeding capabilities of each strain were assessed by testing for rescue of auxotrophs in supernatants obtained from overproducers. Each overproducer was grown in media supplemented with all but one metabolite at 37 °C anaerobically and supernatant was collected after 24 hr. The three auxotrophs were grown in media that consisted of 50% fresh media supplemented with all but one metabolite and 50% supernatant from overproducer. Auxotrophs were then grown for 24 hr and growth was assessed as OD600. c. Growth of Overproducers after 24 hr at which point supernatant was collected for cross-feeding experiment. d. Growth of auxotrophs with and without amino acid supplementation and in supernatant of engineered overproducers and WT equivalents. Shown are three biological replicates with median indicated as horizontal lines.

*B. fragilis* was the highest overproducer at 362 µM which should allow for growth of *E. coli* to OD 0.144, *S. Typhimurium* to OD 0.166 and *B. theta* to OD 0.154. *B. theta* overproduced 34 µM which should allow for *E. coli* growth to OD 0.411, *S. Typhimurium* to OD 0.156 and *B. fragilis* to OD 0.128. The two *Bacteroides* spp. WT also produced detectable amounts of amino acids which would allow for some growth in the case of *B. theta* but not *B. fragilis*. *E. coli* overproduction was 5.3 µM L-methionine. This could allow for growth of *S. Typhimurium* to OD 0.032, *B. theta* to 0.017 and *B. fragilis* to OD 0.027. We could not detect any L-methionine in *E. coli* WT supernatant. These findings indicate strong cooperation from *Bacteroides* spp to other strains and moderate cooperation from *E. coli* to other strains. *S. Typhimurium* overproduces 16 µM L-histidine which should allow for growth of *E. coli* to about OD 0.028. However, *Bacteroides* spp growth was only detected at concentrations higher than 100 uM. Note, these experiments only indicate the potential for cross-feeding. In order to validate each overproducer’s cross-feeding capability we performed the following supernatant experiments.
Bacterial amino acid overproducers can rescue growth defects of the corresponding auxotroph.

Having established amino acid dependency and overproduction of each strain, we tested operation of pairwise cross-feeding using culture supernatants (Figure 2.2B). This test was used to verify occurrence and strength of one-way interactions between consortia members. WT strains and overproducers were grown for 24 hr before supernatant was collected (Figure 2.2C). Notably, three out of four overproducers did not show any growth defect compared to the WT. B. theta growth was decreased by 3-fold. Supernatant from WT strains served as control as these strains were able to produce but not overproduce the respective amino acids. Extent of rescue was determined by OD600 values after 24 hr of growth in supernatant that was diluted 1:1 with fresh media lacking the tested amino acid (Figure 2.2D). As another comparator, we grew the auxotrophic strains without amino acid supplementation and with full supplementation. As expected, we detected no growth in any of the auxotrophs when no amino acid was supplied.

Consistent with our design, and data from growth requirement and overproduction, E. coli grew well in supernatant from engineered B. theta (180% of fully supplemented growth) and B. fragilis (130% of fully supplemented) and somewhat in supernatant from engineered S. Typhimurium (13% of fully supplemented). E. coli grows better in Bacteroides spp. supernatant than in supplemented media indicating that the bacteria may produce other beneficial metabolites. S. Typhimurium’s growth profile looks analogous to the E. coli profile (Figure 2.2D, purple panel) with slightly improved growth (88% in E. coli supernatant; 330% in B. theta supernatant and 227% in B. fragilis supernatant). As expected, growth of B. theta is only barely rescued by engineered E. coli (3%). Engineered B. fragilis rescues B.theta growth only moderately (12%) which is in contradiction with our growth response assessment. It may indicate that B. fragilis secretes molecules that are harmful for B. theta but not E. coli or S. Typhimurium. There does
not appear to be growth by \textit{B. theta} in \textit{S. Typhimurium} supernatant. \textit{B. fragilis} growth was not well rescued by any of the strains (Figure 2.2D, blue panel). Note, that co-culture conditions differ from this experiment in that the bacteria can be in direct contact and local concentrations could be higher thus allowing for growth. Overall these experiments indicate an abundance of cross-feeding events between the engineered bacterial strains. Next, in order to analyze the effect on the consortium in co-culture context we performed full assemblage experiments.

\textbf{Cooperation and evenness of species composition is driven by amino acid abundance}

We used monocultures and co-cultures to test our design, and found that cross-feeding of amino acids from overproducers to auxotrophs has a positive effect on each consortia member in co-culture. We grew monocultures and co-cultures of WT and engineered consortia in a medium, specifically designed to accommodate the four bacterial species in a single batch culture, without amino acid supplementation and estimated bacterial abundance (cfu/mL) via qPCR after 24 hr (Figure 2.3A,B). After 24 hr, WT strains grew to about 2 orders of magnitude higher abundance. Together with data from pairwise co-culture, this result confirms that engineered strains allow for moderate growth but cannot restore WT growth in full.
Figure 2.3 Engineered Cooperation is linked to balanced relative abundances. a. Mono cultures and co-culture of engineered strains were inoculated in media without L-methionine, L-histidine, L-tryptophan.
Figure 2.3 (Continued). and L-arginine supplementation and analyzed via qPCR after 24 hr. Growth of engineered (upper panel) and WT (lower panel) mono cultures and co-culture grown without supplementation was recorded after 24 hr as qPCR cfu/mL estimates. Shown are three replicates with median indicated. For engineered strains co-culture reaches higher cfu/mL than sum of monocultures quantified by cooperation factor. *sum in co-culture in engineered strains is significantly larger than sum in monocultures (p-value: 0.0071)

d. Supplementation titration experiment. Engineered consortia co-culture and mono cultures was subjected to a range of amino acid supplementation. Growth of each strain was analyzed after 24 hr via qPCR. Cooperation and Evenness co-vary. Upper panel: sum of monocultures growth (empty circles) and co-culture growth (filled circles) after 24 hr via qPCR and calculated cooperation factor (grey bars). Shown are three replicas with median indicated as black horizontal line. Cooperation increases with decreasing supplementation. Lower Panel: Population ratios and standardized evenness of co-cultures as function of supplementation. Shown is the mean of three replicates. At high supplementation, evenness is also high, then drops at intermediate levels and rises again at lower levels.

e. Monocultures and co-culture growth in cfu/mL after 24 hr for each single strain.

We next compared growth of each strain in monoculture and co-culture. In the WT consortia, each of the strains grows better in monoculture than in co-culture (Figure 2.3B) suggesting that there exists extensive competition for nutrients and potentially production of toxic compounds between WT strains. As a result, the sum of growth of all monocultures is larger than the overall growth of the co-culture. We quantified this by defining a cooperation factor as overall growth of co-culture divided by sum of growth of monocultures. According to this definition a cooperation factor of less than 1 indicates competitive behavior whereas a cooperation factor of more than 1 indicates cooperation. In the case of the WT consortia this factor is 0.14. For the engineered consortia, *E. coli* and *S. Typhimurium* growth is improved in co-culture compared to monoculture while *B. theta* and *B. fragilis* are unchanged, leading to a cooperation factor of 1.18. We deduce that our engineering introduced net positive effects of co-culture growth on *E.*
coli and S. Typhimurium whereas effects on B. theta and B. fragilis may have been neutralized compared to the WT scenario.

Relative abundance of specific amino acids affects the extent of cooperation of the consortium. The goal of this work is to engineer a platform consortia for applications in the mammalian gut. In the previous experiment we have established cooperation in a condition without any supplementation of the cross-fed amino acids. However, the mouse gut likely contains amino acids from food sources. In order to study the effect of amino acid supplementation on cooperation we subjected the engineered bacteria in monoculture and co-culture to different concentrations of relative amino acid supplementation (Figure 2.3C). We measured bacterial abundance (cfu/mL) via qPCR after 24 hr and calculated cooperation factors for each condition. Both monoculture and co-culture growth decreases with decreasing supplementation. At high supplementation, monocultures growth exceeds co-culture growth resembling WT conditions that are characterized by competitive interactions. By as low as 3 µM supplementation the co-culture growth exceeds monoculture growth indicating occurrence of cooperative interactions. Consistently, the cooperation factor increases with decreasing supplementation and reaches a value >1 at supplementation of 3 µM. We conclude that cross-feeding only occurs in nutrient sparse regimes which is consistent with previous findings describing the occurrence of cross-feeding in natural consortia during nutrient scarcity (33).

Relative abundances of each consortia member are regulated by amino acid supplementation. In order to elucidate the effect of amino acid supplementation on the abundance of individual strains in our consortium, we quantified relative abundances of each strain (Figure 2.3C, lower panel). At low supplementation (10 µM and lower) – when cooperation occurs – the culture assumes a population state that is characterized by approximately balanced abundance of all strains with moderate dominance of S. Typhimurium. S. Typhimurium benefits most from growing in co-culture which is consistent with our previous results (Figure 2.3D).
We quantified the co-culture evenness using normalized entropy of the consortium composition (also called the evenness index). Using this measure, a completely even (balanced) community would have a relative entropy of one. At the highest supplementation (1000 µM) the co-culture shows the highest evenness. This is likely because all the required amino acids are supplied and there is no competition which would result in low evenness. When supplementation is decreased to intermediate levels (30 and 100 µM) the relative abundances of bacteria become less balanced, i.e. entropy decreases. Specifically, *Bacteroides* species abundance decrease dramatically and *E. coli* and *S. Typhimurium* dominate the culture. Both *S. Typhimurium* and *E. coli* have the lowest amino acid requirements (Figure 2.2). In intermediate supplementation levels they might still be fully supplemented whereas *Bacteroides* spp. lack enough amino acid supplementation to thrive. When supplementation is reduced further, evenness increases as cooperation becomes the prevalent mode of interaction.

**Engineered consortia are more robust to variations in starting conditions.** Evenness is an indication of robustness of the co-culture. Here we define robustness as the ability to resume similar stationary phase relative abundances when starting from different inocula ratios. We inoculated both WT and engineered consortia at five different starting conditions each (Figure 2.4). In condition 1 all bacteria species were inoculated at the same ratios and in conditions 2-5 we reduced one of the species inocula by a factor of 10. We then recorded growth of each strain in co-culture over 12 hours (Figure 2.4A, B). For each condition the overall biomass of WT consortium grows similarly to about $5 \times 10^7$ cfu/mL. However, evolution of the relative bacteria abundance is different between conditions resulting in distinct stationary phase states of relative microbe abundance. Specifically, the knocked down species in each condition is unable to recover to values observed in condition 1.
Figure 2.4 Engineered consortia is characterized by net positive interactions and improved robustness. a. We inoculated WT and engineered consortia in media without amino acid supplementation and followed growth over time: for the first 6 hr we collected samples every 30 min, then every hour up to 12 hr. Starting inocula varied to mimic perturbations. Each strain was inoculated at the same ratio (condition 1) and then each one dropped down by 1:10 (condition 2-5). We recorded growth
**Figure 2.4 (Continued).** Trajectories and population ratios over time for WT and engineered strains. Total bacterial abundance (black dots) and relative abundances (colored bars) for each starting condition for WT and engineered strains in sequence. **c.** Evenness of WT and engineered strains after 12 hr growth. **d.** Discrete Hellinger distance between consortia composition of condition 1 and the other four conditions calculated for WT and engineered consortia. **e.** The obtained data serve as input for the Bayesian inference algorithm to calculate estimates for Bayes factors, growth rates and microbe-microbe interactions. With the inferred parameters we predicted directed microbe-microbe interaction networks with edge probabilities for WT and engineered consortia. **f.** The renormalized difference between WT and engineered consortia identifies 5 net positive, 7 neutral interactions and their likelihood (color coded).

The overall evenness between WT and engineered consortia is similar (Figure 2.4C). However, the population states after 12 hr are more similar in engineered consortia across the different initial condition (Figure 2.4D).

**Growth dynamics of the engineered consortia reveal a positive interaction network *in vitro***

To comprehensively interrogate our engineered bacterial interaction network, we measured its dynamics under a variety of starting conditions, and used a tailored machine learning model to infer an underlying predictive dynamical system model. The resulting data-derived model is phenomenological, abstracting various types of possible biological interactions (e.g., competition for nutrients, bacteriocins production, syntrophy, etc.) into quantitative pairwise interaction coefficients. This model thus allows us to objectively assess the net interactions between species in the consortium. Key to the ability to infer such a model is sufficiently rich time series data (Figure 2.4). Densely sampled time series of microbial biomass were used as input to our inference algorithm for both WT and engineered consortia and inference was
performed on each separately to infer growth rates and microbe-microbe interaction strengths (Figure 2.4E,F). The model is Bayesian (measurement error and model uncertainty are propagated throughout the model) with indicator variables added to the microbe-microbe interactions so as to help identify only those interactions which the model is confident are needed to explain the data. Only those interactions with Bayes factors above 3 are considered to be meaningful. The carrying capacity and self-limiting terms of the WT and engineered consortia are at different scales, and thus we normalized them before determining if engineering caused a net increase of interaction strength. The rescaling that we chose to perform was akin to comparing the steady state Jacobians. Comparing the steady state Jacobians for the WT and engineered consortia (and keeping only interactions with Bayes factor >3) we find 5 net positive and 7 neutral interactions. This confirms the ability of the engineered interactions to promote cooperation on top of the intrinsic interactions already present in the WT consortia. Of the 5 net positive changes only one interaction, *B. fragilis* to *S. Typhimurium*, was strong enough to overcome the strong competitive interactions present in WT and produce a true positive interaction. This is consistent with the co-culture experiments in Figure 2.3C, where engineered *S. Typhimurium* (and *E. coli*, but to a lesser extent) had increased growth compared to monoculture.

**Systematic analysis of consortia robustness**

In Figure 2.1 we illustrated how the introduction of positive interactions can increase the robustness of a microbial consortia to disturbances as well as produce a more even steady state as compared to a collection of microbes without interactions. We now study consortia robustness in a more systematic fashion. We begin by parameterizing a Lotka-Volterra dynamical system of four species with identical growth rates $\rho$, self-interaction $\delta$ (which is negative), and identical interactions coefficients $\alpha$, see Figure 2.5A. We then study the
robustness of this system with increasing realism (as it relates to what we observe in our engineered system) under four different conditions.

**Figure 2.5 Parameter exploration with regard to robustness for inferred networks.**

a. Lotka-Volterra dynamical system parameterized by constant growth $\rho$, self-interaction $\delta$ (which is negative), and identical interactions coefficients $\alpha$, for all species. b. Scenario 1: growth rate and self-interaction terms are kept constant and only the positive interaction term is increased. c. Scenario 2: growth rate held constant while the magnitude of the self-interaction term is increased as the interaction coefficient is increased. d. Scenario: interaction coefficient is increased as the magnitude of the self-limiting term is increased and the growth rate is decreased. e. Scenario 4: identical to Scenario 3, but instead of the interaction coefficient starting at zero, it begins with a negative value.

In Scenario 1 (Figure 2.5B) the growth rate and self-interaction terms are kept constant and only the positive interaction term is increased. Under this condition the carrying capacity and robustness of the system increases as the interaction strengths increase, up until the system becomes unstable. This is however a rather fictitious scenario as the engineering of interactions would not allow for the carrying capacity of a system to be arbitrarily increased, i.e. “no free
lunch" theorem applies. Slightly more realistic is Scenario 2 in Figure 2.5C, where the magnitude of the self-interaction term is increased as the interaction coefficient is increased. This results in a constant carrying capacity. Under this condition, with increased interaction strength the robustness margin of the system increases. Still however, this is not completely realistic, as the engineering that was performed in this study removed internal metabolic features from the microbes and simultaneously required them to overproduce other metabolites. Both scenarios would have a negative impact on the intrinsic growth rate of the microbes. In Scenario 3 (Figure 2.5D) while the interaction coefficient is increased the magnitude of the self-limiting term is increased and the growth rate is decreased. Under this scenario the robustness margin cannot be arbitrarily increased, and an optimal condition arises. Note that even while the carrying capacity of the system is decreasing the robustness margin can still increase. Scenario 4 (Figure 2.5E) is identical to Scenario 3, but instead of the interaction coefficient stating at zero, it begins with a negative value. This scenario matches more closely what was observed in our WT consortia. Without engineering our four microbes had intrinsic competition. Under this scenario, once again, the robustness of the system cannot be arbitrarily increased, instead an optimal value arises. This illustrates that even if the carrying capacity of the system is decreasing (because of our engineering) its robustness to external perturbations targeting a single species can be increased (up to a point).

**Engineered consortia maintains greater evenness of composition in the animal gut.**

Our ultimate aim is to design consortia that function in natural environments, particularly in the mammalian gut, where consortia may be used as therapeutics to treat human diseases. We thus investigated the behavior of our consortium in the mammalian gut, using gnotobiotic mice as a controlled yet sufficiently complex environment for evaluation. As we demonstrated *in vitro*, amino acid supplementation levels drive the composition of the engineered consortium. To
evaluate this effect in vivo, we altered amino acid levels in the gut by changing the animal’s diet (34), comparing the composition of the engineered and WT consortia on both standard chow and amino acid depleted (low protein) diets.

Groups of five germfree mice were fed standard or low protein (3%) chow and gavaged with either the WT or engineered consortium (Figure 2.6). The consortia were allowed to colonize for 10 days, and then stool samples were collected and interrogated via qPCR with species-specific primers. The engineered consortium consistently exhibited greater compositional evenness in mice that were fed low protein diet compared to the three other groups (p-values: 0.02, 0.03, 0.02).

The diet influences absolute bacterial abundances in both engineered and WT consortia and each strain is affected to a different extent. For the engineered consortia in mice that were fed standard diet growth increases by a factor of 3 for E. coli, 8 for S. Typhimurium, 16 for B. theta and 11 for B. fragilis compared to strains in mice that were fed low protein diet. This difference in increase of growth is reflected in decreased evenness of the engineered consortium in mice that were fed standard diet. In the case of WT consortia, S. Typhimurium, B. theta and B. fragilis are similarly affected like their engineered counterparts (foldchanges: 10, 22, 13 respectively). However, E. coli growth changes by a factor of 51 indicating that its growth is strongly affected by the diet. Our engineering may have alleviated this effect by linking it to the other consortia members.
Figure 2.6 Engineered consortia is more even in gnotobiotic mice. Three groups of 5 germfree male Swiss Webster mice were either fed low protein diet or standard diet and gavaged with $10^7$ cfu of either engineered or WT bacteria. Fecal samples were taken 10 days after gavage and analyzed via qPCR. Small panels show absolute bacterial abundances as cfu/g feces of each strain. Standardized evenness index was calculated and is plotted below for each mouse. Bar indicates Median. Mann-Whitney test showed significantly increased evenness of engineered consortia in mice that were fed low protein diet compared to the consortia in the three other groups (p-values: 0.024, 0.032, 0.015). Legend: Ec: E. coli; St: S. Typhimurium; Bt: B. theta; Bf: B. fragilis.
Notably, engineered strains in mice that were fed low protein diet exhibited similar growth to WT strains in mice that were fed low protein diet with the exception of *S.* Typhimurium which grew about 8-fold better in engineered consortia compared to WT. This finding is consistent with our *in vitro* results which indicated that engineered *S.* Typhimurium benefits most from growing in co-culture. The same effect can be observed in mice that were fed standard diet albeit to a lesser extent. These *in vivo* data are consistent with our *in vitro* findings showing that our engineering introduced an even bacterial composition in the animal gut.
Discussion

We developed four gut-like bacterial species to cross-feed metabolites thus building a cooperative heterogeneous synthetic consortia. The synthetic consortia is characterized by even relative abundance of bacteria that is stable to perturbance in initial abundance of single species and net positive interactions between the bacteria as revealed by computational analysis. We further showed the power of this approach for use in controlling probiotic delivery, demonstrating the improved evenness of the engineered consortium in the complex mouse gut environment.

Cooperation of the engineered consortia manifests in increased growth of the co-culture compared to monocultures. This bacterial teamwork is a function of amino acid supplementation; it occurs at low or no external metabolite supplementation. While microbial cooperation is found in many natural habitats (10, 35, 36), and there are many ecological theories that attempt to explain its evolution, there is a dearth of experimental systems to test such hypotheses. We have created a model ecosystem that allows us to study the effects that cooperation has on microbial consortia. In our system, cooperation leads to improved growth of the overall consortia, but it also promotes continuing survival of each single species through improved evenness of bacterial abundances. These two characteristics of our system make it an attractive test-bed to address ecological questions about the evolution of cooperation, which serves both the consortia as a whole and each single species’ survival.

We inferred five net positive and 7 neutral interactions between WT and our engineered consortia using a Bayesian Inference Algorithm (Figure 2.4). This results in an interaction network which partially resembles our design and is largely consistent with experimental results. For example the strongest positive interaction is inferred from \textit{B. fragilis} to \textit{S. Typhimurium}. Concordantly, we found that \textit{B. fragilis} overproduces the highest amount of its cross-fed metabolite (Arg, Figure 2.2) and \textit{S. Typhimurium} is the strain that benefits most in co-culture
(Figure 2.3). Secondly, *B. fragilis* does not show improved growth in co-culture compared to monocultures and we also do not infer any positive interactions to this species. No positive interactions are inferred coming from *B. theta*. This might reflect the fact that the WT strain already produces a certain amount of Trp (Figure 2.2) and hence might exhibit positive interactions. As our inference extracts the net interactions between WT and engineered consortia it is possible that this interaction was not interpreted as net positive. We inferred moderate positive interactions from *E. coli* to *S. Typhimurium* and *B. theta* and also that is consistent with pairwise supernatant experiments (Figure 2.2). Interestingly, *S. Typhimurium* is inferred to exert positive interactions to *E. coli* and *B. theta*. Experimental results can confirm the positive interaction towards *E. coli* (Figure 2.2) but not *B. theta*. Additionally, experimental results would have suggested positive interactions from *B. fragilis* to *E. coli* which was not represented in the inferred network. Note that we do not expect the inferred model to match our engineering design exactly for two reasons. First, we use a machine learning approach that is driven by experimental data, which is inherently noisy and so some interactions may not be seen. Second, and more importantly, the species in our consortium naturally colonize the gut and thus may already have positive or negative interactions, so there’s no guarantee that the engineered interactions will dominate. The fact that the inferred model showed dominant positive interactions in the engineered consortium thus confirms the utility of our design that uses amino acids to induce a cooperative network.

In order to demonstrate robustness to complex environments and potential for clinical use of our engineered cooperative consortia we subjected it to the animal gut. We show that the engineered consortia exhibits increased evenness of bacterial species abundance in mice that were fed low protein diet (Figure 2.6). This diet causes a depletion of amino acids in the animal gut mirroring our *in vitro* conditions of decreased amino acid supplementation. Under these
conditions in vitro we also found increased evenness and cooperation. Hence we engineered a synthetic heterogenous microbial consortia with predictable behavior both in vitro and in vivo.

We have demonstrated a design, build and test cycle for gut microbiota engineering that provides a test bed for ecological questions and has implications for bacterial therapeutics engineering.


Chapter 3

*Escherichia coli* NGF-1 – a natural gut flora isolate that is genetically tractable and colonizes the animal gut well
Preface

This research was conducted in concert with Michiel Karrenbelt, Johannes Bues and Elena Schaefer. Johannes prepared the strain for sequencing and helped with PCR for manual curation. Michiel assembled the genome and helped with phylogenetic and metagenomics analysis. Elena Schaefer helped with phenotypical specification.

Abstract

The human gut microbiota plays an important role in health and disease. *Escherichia coli* is a genetically tractable organism that has been engineered with functions to aid in disease treatment or health enhancement and could replace drug treatments in the future. A murine *E. coli* strain named NGF-1 (Natural Gut Flora) was previously isolated from a BALB/c mouse and exhibits good colonization qualities and is genetically tractable. Whole genome sequencing was performed. The genome of NGF-1 consists of a circular 5 Mbp chromosome and two circular plasmids 8.5 kb and 40 kb in length. It is auxotrophic for niacinamide but prototrophic for amino acids. This genome sequence provides a basis for further study of molecular basics of gut colonization and will facilitate applications of NGF-1 in research and treatment of diseases.
The gut microbiome plays a key role in the health of all organisms with an intestinal tract, from nematodes to humans (1, 2). The mouse is a common model system for studying all aspects of human health, including interactions in the gut, and several groups have developed 'gnotobiotic' mice, in which germ-free mice are colonized with bacteria isolated from humans, often including *E. coli* (3). However, *E. coli* K12 and human-derived *E. coli* strains in general may not be well suited to survive in the mouse and may not represent natural interactions.

*E. coli* NGF-1 is of particular interest because it was isolated from a healthy BALB/c mouse from Charles River Labs, colonizes mice efficiently, and can be engineered with complex genetic circuits (Table 3.1). Kotula et al. isolated NGF-1 and showed that a genetic memory device consisting of a tetracycline-inducible trigger element placed into the ara operon and a memory element positioned in the lac operon could both be transduced by phage P1 from *E. coli* K12 into *E. coli* NGF-1 and functioned identically in K12 and NGF-1 both in vitro and in the mouse gut, responding to a transient tetracycline treatment by permanently changing the epigenetic state of the memory element (4). These authors also showed that the NGF-1 strain, once established in the gut, was stably maintained while K12 was lost within about 1 week in mice with competing bacteria. In addition, while a streptomycin-resistance version of NGF-1 could be established in the mouse gut by inoculation of $10^7$ CFU via oral gavage, inoculation of $10^8$ CFU could establish this strain in the gut of BALB/c mice without streptomycin treatment.

Subsequently, Riglar et al. engineered the NGF-1 strain with a tetrathionate-inducible trigger element and the same memory element, and showed that the strain could sense tetrathionate in the gut that was produced by either *Salmonella* infection or a pro-inflammatory knockout of the host IL-10 gene (5). These authors also showed that the NGF-1 strain was stably maintained in C57Bl/6 mice for over 6 months, and that upon re-isolation, testing and genome sequencing, the memory system retained its function and no mutations were identified in either the engineered
elements or in the rest of the bacterial chromosome. These observations indicate that the NGF-1 is well adapted to colonization of mice.

Table 3.1 NGF-1 Applications.

<table>
<thead>
<tr>
<th>Application</th>
<th>Engineering</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gut sensor (molecule)</td>
<td>• ATC-inducible trigger element</td>
<td>Kotula et al. 2014</td>
</tr>
<tr>
<td></td>
<td>• Memory element</td>
<td></td>
</tr>
<tr>
<td>Gut sensor (pathogen)</td>
<td>• Tetrahionate-inducible trigger element</td>
<td>Riglar et al. 2016</td>
</tr>
<tr>
<td></td>
<td>• Memory element</td>
<td></td>
</tr>
<tr>
<td>Chronic infection sensor</td>
<td>• ATC-inducible trigger element</td>
<td>Certain et al. 2017</td>
</tr>
<tr>
<td></td>
<td>• Memory element</td>
<td></td>
</tr>
<tr>
<td>Interspecies quorum sensing</td>
<td>• ATC-inducible signaling element</td>
<td>Kim et al., manuscript in review</td>
</tr>
<tr>
<td></td>
<td>• Lux-triggered memory element</td>
<td></td>
</tr>
<tr>
<td>Metabolite cross-feeding consortia member in the gut</td>
<td>• Tripple KJ of amino acid biosynthetic pathways</td>
<td>Ziesack et al., manuscript in preparation</td>
</tr>
<tr>
<td></td>
<td>• Methionine overproduction through antimetabolite selection</td>
<td></td>
</tr>
</tbody>
</table>

Certain et al., showed that the NGF-1 strain could be used in a mouse model of persistent infection associated with surgical implants (6). In this model, mice are implanted with a biocompatible plastic pin into the femur, along with $10^6$ *E. coli* NGF-1 carrying the tetracycline-inducible memory system. A fraction of the mice developed a persistent infection, and treatment of the mice with tetracycline could modulate the trigger-memory system. These results indicate that *E. coli* NGF-1 can survive near a surgical implant, making it useful for studies of persistent infection and the accessibility of drugs around poorly vascularized regions.

Because of the importance of *E. coli* NGF-1 as a model organism well-adapted to the mouse, we sequenced the genome to determine potential advantageous characteristics that allow it to colonize the mouse intestine. Genomic DNA (gDNA) was extracted from overnight bacterial culture using the QIAGEN DNeasy Blood and Tissue kit. Concentration of the gDNA was quantified using Life Technologie QuantiT PicoGreen dsDNA Assay Kit. The genome of
NGF-1 was sequenced using Illumina MiSeq Reagent Kit V2 2x250 and the paired end reads were assembled de novo with Velvet (7) and SPAdes (8) into 122 and 64 contigs, respectively. After filtering of contaminants from the contig sequences, 93 and 51 contigs remained, and had an average read coverage of 70-fold and 22-fold.

Contig ends were joined by two methods. First, some contig ends were identified that had ends with identical segments that fell below the alignment threshold of the joining software. These joinings were validated by alignment of the joined sequences with other E. coli strains. Second, in cases such as where identical rRNA genes prevented inferring continuity between sequences on either side of the repeated element, we hypothesized associations based on other E. coli sequences, and then confirmed the association by PCR using unique-sequence flanking primers and observation of a DNA fragment of predicted size. One case of sequence ambiguity was attributed to an inverting phase variation-type element.

Genome annotation was performed using Rapid Annotation using Subsystem Technology (RAST) (9) and the Pan-Genomes Analysis Pipeline (PGAP) (10), and was subsequently manually curated through function- and sequence-based comparisons.

NGF-1 contains a 5,026,105 bp chromosome with a GC content of 50.7% and encoding 5,218 genes (Figure 3.1A). Additionally, we found multiple contigs with significantly increased coverage that we assembled into two plasmid sequences, pNGF-1-CROD2 (40,158bp; GC content: 41%) and pNGF-1-colY (8,556bp; GC content: 47.5%), encoding an additional 57 and 10 genes, respectively (Figure 3.1B,C). We found 23 rRNA and 85 tRNA genes in the chromosome. We were able to classify 3555 of the identified genes to functional categories based on clusters of orthologous genes (COG) designation (11). Comparison of COGs from NGF-1 with K12 MG1655 shows that all subsystems in the cellular processes and signaling category are overrepresented in NGF-1 compared to K12. Most strongly overrepresented are cell motility, intracellular trafficking and secretion, defense mechanisms, and cell
wall/membrane/envelop biogenesis. Genes categorized in these groups may contribute to better colonization ability of NGF-11.

NGF-1 is a niacinamide auxotroph. The auxotroph can be rescued by nicotinic acid, nicotinamide and beta-nicotinamide adenine dinucleotide. This indicates that NAD salvage pathway is intact whereas de novo synthesis from tryptophan is not. Specifically nadC appears to be non-functional. Sequence analysis indicates a single non-silent mutation in position 103 which is a serine to proline conversion. Auxotrophs in certain vitamins may be the result of living in the gut environment where these are highly abundant. We did not find any other auxotrophies
using the Biolog Phenotype Microarray PM5 and the strain grows in M9 glucose media supplemented with niacinamide (Biolog).

Phylogenetic analysis revealed that NGF-1 is closely related to MP1, which was previously isolated from the same mouse breed and analyzed for its capability to colonize the animal gut (12). To identify genes that aid NGF-1 in colonization we performed comparative genomic analysis between NGF-1, MP1 and three closely related strains (Figure 3.2).

Figure 3.2 Metagenomic Analysis. a. Evolutionary phylogeny of *E. coli* and *Shigella*, using *Salmonella* Typhimurium LT2 as an outgroup. Strains used to generate Venn diagram are highlighted in color. b. Venn diagram of the distribution of clusters of orthologous genes between *E. coli* NGF-1, K12 MG1655, PMV-1, LF82 and MP1 (color coding according to a.). 33 COGs that are unique to NGF-1 are highlighted in bold and with an arrow.
NGF-1 contained 33 COGs not found in the other strains which we analyzed in further detail. We found that most of the clusters contain genes genomic islands which may be phages or originated from horizontal gene transfer. These genes include cell wall synthesis, regulation of virulence and motility and biofilm formation. Genes that were not located in genomic islands were involved in sugar uptake, membrane stability and several proteins that contained the DUF351 uncharacterized protein domain. All of these genes may have contributed to better colonization capability of NGF-1 and have therefore been selected throughout its evolution. Another important factor for colonization is the colicin-expression plasmid pNGF-1-coliY as it allows for clearing of nutritional niches (13).

NGF-1 sequence analysis and comparative genomics provided us with genetic insight in its ability to colonize the gut and other physiologically relevant sites. E. coli NGF-1 is both genetically tractable and able to colonize the animal gut well. It is used to facilitate applications as bacterial therapeutic and diagnostic agent.

Nucleotide sequence accession numbers: CP016007.1
References


Discussion
Humans have domesticated animals for thousands of years and domestication of the microbial world is the next logic step. The three major uses for bacteria are in bio-based chemicals production, bioremediation and more recently as bacterial therapeutics. In my thesis I contributed to bio-based chemicals production by means of protein engineering and to development of bacterial therapeutics through consortia engineering and chassis discovery. Bacteria harbor a vast genetic pool that contains functions and biosynthetic pathways which we can harness to preserve the environment and our own health. We are able to manipulate genes, proteins and entire consortia of bacteria with increasing skill. The question however is, why do we not yet make more use of bacteria in our everyday life. Why is it that most inventions remain in the laboratory and have no impact on the “real world”?

Research needs to become more interdisciplinary early on. The value of interdisciplinary research has been recognized for a while (1). If researchers isolate themselves in their field the questions that they tend to ask are either narrow and focused on their specific niche or impossible to implement. Specialists of course are required but they need to be better connected with each other. In our educational system (Bachelor, Master, PhD) there is almost no overlap with different programs and departments are usually physically separated. So, scientists in training are rarely exposed to different disciplines. Programs such as the Harvard University Consortium for Energy and Environment try to bridge the gap between disciplines and bring students together that are focused on similar goals but tackle them from different angles. I believe that early exposure of specialists to different disciplines will improve the quality of research in general.

Synthetic Biology invites interdisciplinary research (2). One important contribution of synthetic biology as a field is that it attracts researchers from different disciplines such as chemistry, physics, computer science and more. The Bionic Leaf is a success story in which a synthetic biology laboratory collaborated together with a chemistry laboratory to build a system
with implications for the bio-based chemicals industry (3). The two disciplines together were able to address a question that required problems in both microbiology and electrochemistry to be solved.

I believe that an important reason for why microbes are not applied more than they are at the moment is a gap between the scientist and the industrial environment. I want to make the distinction between applied and basic research here. I believe that basic research is absolutely necessary and it should be entirely unbiased and disconnected from any industrial funding. However, we likewise need research that is done with applications in mind. In academia, however, there is a stigma on research that is done this way even though some disciplines call themselves “applied”. I believe that applied researchers need to embrace this concept and communicate more with the outside world in order to produce results that will actually find their way to consumers. Disciplines such as mechanical engineering and computer science already have PhDs that can be done together with industry or in industry and I believe applied biological sciences need to do the same (4). Programs such as iCORE are attempting to bridge the gap between academic “applied research” and industry by incentivizing scientists before startup to interview potential customers in order to find out if their technology is viable. I believe that it is too late to do that at this point. If we would bring young scientists in contact with industry or medicine before they start their research then they may approach it with constraints and demand in mind. If we want to call a field “applied” science then it should focus more on what the market actually wants and not what the researcher thinks the market “should” want.

In society microbes especially bacteria are still mostly associated with infections. This image can be polished by better educating the public to distinguish between “good” and “bad” microbes. This will be of special importance to gain acceptance of bacterial therapeutics.

An example of largely “good” bacteria is our own gut microbiota. Even though the term microbiota is not new intense study was enabled only recently through the -omics technologies
(5). Most gut bacteria are still hard or un-culturable and we therefore have to rely on sequencing or mass spectrometry to study them. In the past 15-20 years we have come from initial and mostly descriptive research to providing first mechanisms and directions for treatment. The human microbiota – due to its vastness and intricate connection to our own body – has the potential to revolutionize many aspects of medicine (6).

We can learn from some of the problems in bio-based chemicals production for the development of bacterial therapeutics. For my thesis chapter 2 I worked closely together with a medical doctor. His group uses a network inference software to predict interactions and devise therapies by modulating those interactions. Our synthetic consortia served as a test bed to confirm these algorithms in a defined setting and provided clues for further improvements. With more accurate predictions, therapies that modulate the microbiota will become more successful. We were only able to contribute in this way because of our close collaboration throughout the project. This example shows that it is greatly beneficial to work closely together with collaborators in medicine when devising bacterial therapeutics.

In the gut microbiome field we are in a unique position where we can learn and engineer at the same time. In my thesis work we aimed to engineer interdependent consortia as platform for bacterial therapeutics. Our design was based on the notion that if we generate dependency between each of the bacteria they will grow together. However, we did not anticipate that because this was a multi-species consortium, many negative interactions pre-existed. We overcame most of these by intensive media engineering but for some interactions we were only able to neutralize them through engineering efforts. Our engineering however provided insight into consortia interactions and the effect of cooperation on each single species and the consortia as a whole. So by engineering the system we were able to gain insight into its biology.

Given that humankind is only aware of microbe’s existence for about 200 years, we have come a long way to domesticate them. The bio-based chemicals market size was at $6.5 billion
in 2016 and projected to grow by over 16% in the next 5-10 years (7). Policies are being made
to replace petro-based chemicals with renewable and biodegradable options as society
becomes more aware of the detrimental effects on our environment (8). At the same time our
knowledge of the microbiota in and around us increases steadily; first clinical trials and startups
exist (Synlogic, Kaleido, Indigo). These trends show that microbial domestication is under way
and I am proud to have contributed to it during my thesis work.
References


Appendix A: Supplemental Material for Chapter 1

Figure S1.1 Sequence alignment of CpFatB1, CpFatB2 and 2OWN.
Figure S1.2 Sequences of all chimeras.
Figure S1.2 (Continued).

Chimera 1:
LILTAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
VKGRTEWRPKNAGAKGAILTGTK

Chimera 2:
LILTAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
VKGRTEWRPKNAGAKGAILTGTK

Chimera 3:
LILTAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
VKGRTEWRPKNAGAKGAILTGTK

Chimera 4:
LILTAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
VKGRTEWRPKNAGAKGAILTGTK

Chimera 5:
LILTAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
VKGRTEWRPKNAGAKGAILTGTK

Chimera 6:
LILSAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
IMKGRTEWRPKNAGATNGAISTGTK

Chimera 7:
LILSAITTVFVAFERKFWTMLRKRPLRPNLVMDSFGLERVQGLVERQSFIRSYIEYCARTASIETLMNNFQE
TSNHCIGIIGLLNDGFGRTPcemKDRILWVMVKMIENRYFTWGETIEVNTWANSISGKHGMGRDLWLDSCHT
GEILRATSVAMMNQKTRKSLPRKPEVQIEIEPQFVDSAIFPIVDDKFRKLRDDKLTGSDICNGHTPRWTLDDL
NQHVSVNKIGWILQSVPETFVFETQECGLTLLEYRRECGRDSVLESVTAMDSKEDGSPLQHLRLLEDGADI
IMKGRTEWRPKNAGATNGAISTGTK
Figure S1.2 (Continued).
Figure S1.3 Complete mutational analysis of fragment E’s effect on FFA production. All six amino acids in fragment E of CpFatB1 were converted to CpFatB2 equivalents (a.) and corresponding mutations were made in chimera 4 to revert those residues to CpFatB1 (b.). Constructs were expressed in E. coli and FFA in supernatant was measured using an enzymatic assay. Shown is the average of three biological replicas, error bars indicate standard deviation. Table under graph shows OD600 of cultures at time of harvest.
Figure S1.4 Thioesterase expression visualized by Westernblot. Shown is the un-cut gel of the westernblot shown in figure 5. We loaded 25 µg of protein from cultures transformed and induced as described in Methods. Proteins were stained with a primary rabbit-anti-RPβ antibody and a secondary anti-rabbit antibody that was conjugated to horseradish peroxidase (HRP). The same blot was stained with a primary anti-his tag antibody that was conjugated to HRP. Proteins were determined by size and antibody stain. Two bands underneath RPβ and the band on the bottom in all lanes is likely breakdown products of RPβ as they occur in membranes stained with only that antibody but not with His-tag stained membranes. The strong band above and less intense bands underneath TE are non-specific binding events of the his-tag antibody (perhaps proteins with several histidines). The TE band is the only one that does not show in the empty vector control.
Figure S1.5 Certain sub-fragments of fragment D do not confer length specificity. We introduced three subarts of fragment D from CpFatB1 into CpFatB2 to test if these amino acids were essential for chain-length specificity. Frag172-201 includes 13 aa changes and comprises the entire loop; introducing this fragment form CpFatB1 into CpFatB2 reduced activity drastically indicating that the protein was no longer functional. Frag190-197 changes 5 aa and introduces an additional asparagine but does not alter the chain length specificity of the enzyme. Frag172-181 introduces 4 aa changes and may alter the specificity slightly but by far not as drastically as the entire fragment D. We conclude that there are either different parts of fragment D that cause the effect or the entire part is required for shorter chain-length specificity.
## Table S1.1 Fatty Acid Production of engineered thioesterases

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Sequence origin</th>
<th>total FFA [ug/mL]</th>
<th>FFA species [ug/mL]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CpFatB1</td>
<td>CpFatB2</td>
</tr>
<tr>
<td>CpfatB1</td>
<td>1...316</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>CpfatB2</td>
<td>-</td>
<td>380</td>
<td>3</td>
</tr>
<tr>
<td><strong>Rational Approach</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CpfatB2_FatB1pocket</td>
<td>62...162</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>CpfatB1_FatB2pocket</td>
<td>1...61, 163...316</td>
<td>561</td>
<td>12</td>
</tr>
<tr>
<td>CpfatB2_I151V</td>
<td>V151</td>
<td>589</td>
<td>14</td>
</tr>
<tr>
<td>CpfatB2_F71V</td>
<td>V71</td>
<td>260</td>
<td>10</td>
</tr>
<tr>
<td>CpfatB2_L67V</td>
<td>V67</td>
<td>569</td>
<td>11</td>
</tr>
<tr>
<td>CpfatB2_tripleV</td>
<td>V151, V71, V67</td>
<td>177</td>
<td>6</td>
</tr>
<tr>
<td>CpfatB2_pocket10</td>
<td>H70, V71, S81, D86, M110, L127, Q129, Q129, Q129, Y157</td>
<td>316</td>
<td>17.9</td>
</tr>
<tr>
<td>CpfatB2_M70H</td>
<td>H70</td>
<td>333</td>
<td>4</td>
</tr>
<tr>
<td>CpfatB2_H78Q</td>
<td>Q78</td>
<td>54</td>
<td>4</td>
</tr>
<tr>
<td>CpfatB2_B15S</td>
<td>S81</td>
<td>331</td>
<td>4</td>
</tr>
<tr>
<td>CpfatB2_N86D</td>
<td>D86</td>
<td>336</td>
<td>2</td>
</tr>
<tr>
<td>CpfatB2_T92S</td>
<td>S92</td>
<td>344</td>
<td>3</td>
</tr>
<tr>
<td>CpfatB2_Q108K</td>
<td>K108</td>
<td>315</td>
<td>8</td>
</tr>
<tr>
<td>CpfatB2_E110M</td>
<td>M110</td>
<td>327</td>
<td>4</td>
</tr>
<tr>
<td>CpfatB2_N124S</td>
<td>S124</td>
<td>329</td>
<td>8</td>
</tr>
<tr>
<td>CpfatB2_V127L</td>
<td>L127</td>
<td>186</td>
<td>5</td>
</tr>
<tr>
<td>CpfatB2_A129Q</td>
<td>Q129</td>
<td>319</td>
<td>4</td>
</tr>
<tr>
<td><strong>Chimera Approach</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chim1</td>
<td>1...61</td>
<td>62...316</td>
<td>175</td>
</tr>
<tr>
<td>Chim2</td>
<td>1...116</td>
<td>117...316</td>
<td>0</td>
</tr>
<tr>
<td>Chim3</td>
<td>1...162</td>
<td>163...316</td>
<td>0</td>
</tr>
<tr>
<td>Chim4</td>
<td>1...219</td>
<td>220...316</td>
<td>31</td>
</tr>
<tr>
<td>Chim5</td>
<td>1...271</td>
<td>272...316</td>
<td>0</td>
</tr>
<tr>
<td>Chim6</td>
<td>62...316</td>
<td>1...61</td>
<td>0</td>
</tr>
<tr>
<td>Chim7</td>
<td>117...316</td>
<td>1...116</td>
<td>337</td>
</tr>
<tr>
<td>Chim8</td>
<td>163...316</td>
<td>1...162</td>
<td>125</td>
</tr>
<tr>
<td>Chim9</td>
<td>220...316</td>
<td>1...219</td>
<td>210</td>
</tr>
<tr>
<td>Chim10</td>
<td>272...316</td>
<td>1...271</td>
<td>284</td>
</tr>
</tbody>
</table>
Table S2.1 qPCR probes and primers.

<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
<th>Dye</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. fragilis</td>
<td>Forward Primer</td>
<td>GCTTGCTTCCAGTCTGCTAT</td>
<td>Hex</td>
</tr>
<tr>
<td></td>
<td>Probe</td>
<td>AGGCAGATTGCAAGAATGGAAGGCTATCAGTAATG</td>
<td>Texas Red</td>
</tr>
<tr>
<td></td>
<td>Reverse Primer</td>
<td>ATGGCATATTGTCATCCACAAAA</td>
<td></td>
</tr>
<tr>
<td>B. theta</td>
<td>Forward Primer</td>
<td>GCGAACCACGCTAACAATTAC</td>
<td>Cy5.5</td>
</tr>
<tr>
<td></td>
<td>Probe</td>
<td>CCATACCGATCGTACAGTCAGGCTATCAGTAATG</td>
<td>Cy5.5</td>
</tr>
<tr>
<td></td>
<td>Reverse Primer</td>
<td>GTTCATCGGAGAGTAGATGATCA</td>
<td></td>
</tr>
<tr>
<td>S. Typhimurium</td>
<td>Forward Primer</td>
<td>CTGGCGAAGCTCGGTCAA</td>
<td>Cy5.5</td>
</tr>
<tr>
<td></td>
<td>Probe</td>
<td>TCATTGATAGATCGGTCAATTAC</td>
<td>Cy5.5</td>
</tr>
<tr>
<td></td>
<td>Reverse Primer</td>
<td>CGAATTGATGGATCGGTCAATTAC</td>
<td>Cy5.5</td>
</tr>
<tr>
<td>E. coli</td>
<td>Forward Primer</td>
<td>CTATCCAGATTGCTTCGCTCTTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Probe</td>
<td>TCCTGGCTCTCGATGATGCTTAAGT</td>
<td>Cy5.5</td>
</tr>
<tr>
<td></td>
<td>Reverse Primer</td>
<td>CCATTGGTCAGGCTTGCTAA</td>
<td></td>
</tr>
</tbody>
</table>
### Table S2.2 Overproduction Mutations.

<table>
<thead>
<tr>
<th>Process</th>
<th>gene</th>
<th>SNP</th>
<th>protein level</th>
<th>outcome</th>
<th>literature</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td><strong>NGF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis</td>
<td>metA</td>
<td>ATC -&gt; AGC</td>
<td>I256S</td>
<td>prevents feedback inhibition --&gt; increases production, prevents norleucine uptake</td>
<td>Laird et al. 2014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>metH</td>
<td>GCG -&gt; GGG</td>
<td>A1191G</td>
<td>could affect production</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aspC</td>
<td>T -&gt; G</td>
<td>promoter -8</td>
<td>makes precursors aspartate so could also improve production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>melP</td>
<td></td>
<td></td>
<td>could be mutated</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>S. Typhimurium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis</td>
<td>hisG</td>
<td>GAA -&gt; AAA</td>
<td>E271K</td>
<td>decoupled from histidine feedback inhibition</td>
<td>Malyn et al. 2018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hisI</td>
<td>CGC -&gt; CGA</td>
<td>silent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>hisA</td>
<td>DTG -&gt; GGG</td>
<td>V112G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degradation</td>
<td>hisH</td>
<td>TTG -&gt; CTG</td>
<td>silent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>hisG</td>
<td>TGG -&gt; GGG</td>
<td>W208G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAT -&gt; GAT</td>
<td>N210D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTG -&gt; GGG</td>
<td>V214G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAG -&gt; GGG</td>
<td>E215G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAC -&gt; GCC</td>
<td>D218G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>hisJ</td>
<td>T -&gt; G</td>
<td>promoter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>hisQ</td>
<td>5 mutations between -151 and -177</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>hisP</td>
<td>GGT -&gt; GGG</td>
<td>silent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. theta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis</td>
<td>BT_0531</td>
<td>GCG -&gt; GTC</td>
<td>A306V</td>
<td>one or both of these could remove feedback inhibition</td>
<td>Fang et al. 2015</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAT -&gt; GAT</td>
<td>N93D</td>
<td>different mutations have been described for E. coli</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATT -&gt; AGT</td>
<td>W85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGC -&gt; GGC</td>
<td>C202G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAA -&gt; GGA</td>
<td>E223G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCC -&gt; ACC</td>
<td>S226T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. fragilis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis</td>
<td>BRF038R_0532</td>
<td>CTC -&gt; CCG</td>
<td>L26R</td>
<td>arginine repressor, possibly made unfunctional (interface, arginine binding)</td>
<td>Ginasy et al. 2015</td>
<td>could not find arginine transport system, could also be mutated</td>
</tr>
</tbody>
</table>