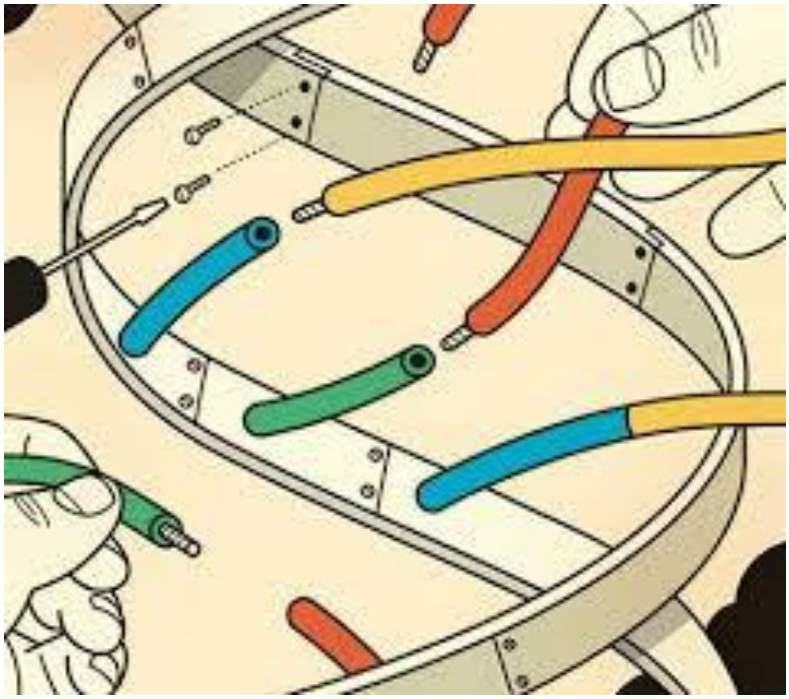

Principles of Synthetic Biology

Or how I quit worrying and learned to love the genome

Adam Arkin and Ron Weiss

Principles of Synthetic Biology, Fall 2013
Arkin and Weiss - Draft 9/5/13

Table of Contents

Preamble: History and Overview of the Field

A false first impression

Welcome

On the blackboard at his time of death in 1988, the famed physicist and raconteur Richard Feynman left written, "What I cannot create I do not understand." It may be strange to some to see these words from a key figure of the in some ways inscrutable field of quantum mechanics, but this is also a man who was involved in developing the atomic bomb and investigating the Space Shuttle Challenger disaster and helped pioneer the fields of quantum computing and nanotechnology. He was a man clearly unafraid to explore the possibilities for harnessing the basic rules of existence for human purpose, or to face the difficult scientific, technical and ethical issues that such endeavors create.

This desire to understand nature permeates all scientific disciplines. In turn, scientists can learn to predict, control and design both natural and man-made systems, with enormous opportunities to benefit society. Biology is no exception, and indeed may represent one of the oldest and most ubiquitous technologies on earth.

Many of the largest problems facing society in the future, including scarcity of treatments for chronic diseases, dwindling supplies of clean water and fertile soil for agriculture, or pollutants building up in the environment, are deeply tied to biological resources. It is increasingly clear that cellular biotechnology provides a potentially powerful tool for solving these challenging global problems. Specifically, Synthetic Biology aims to engineer organisms that can reliably, predictably, and safely produce complex biomolecules like drugs and biofuels, break down toxins, and sense and respond to environmental cues, among many other possible applications.

There have already been some successes in this field, but to achieve these goals more broadly in an efficient, scalable, and safe manner, synthetic biologists are working to develop a more codified engineering approach, as has been successful in disciplines like electrical engineering and computer science. Such a formal design framework will also help synthetic biologists reduce the uncertainties in their designs and provide a rigorous testing environment. These facets are crucial as biological engineering continues to grow in new directions, because however excited society is with the prospect of these new applications, some are also uncomfortable with the currently uncertain and unpredictable implications of introducing engineered organisms into the environment, resulting in potentially unknown ecological interactions.

All engineering fields aim to produce a desired effect via an efficient design, whether the components are electrical, structural, or biological, but trust is also an important element. The end user of the engineered product must be assured of its safety, effectiveness, and economic efficiency. In addition, other engineers who wish to benefit from the each other must trust that the parts built by others meet rigorous specifications for operation and interface to other engineered systems. By creating a formal design framework for cellular and genetic engineering, Synthetic Biology aims to increase efficiency and effect, as well as trust.

Formal design means dissecting complex specifications into hierarchically arranged modular subproblems and adhering to standards both for performance and interface. For example, it is popularly noted that the 286 processor was the last CPU that an individual person could fully understand. However, that limit to individual comprehension has not stopped us from designing the vastly more complex Xeon processors used in many of our computers today—and this is because a design framework exists that allows engineers to understand the individual modules as encapsulated functions and to design at this higher level of abstraction. In addition, computer-aided design programs backed by standardized elements (parts), interconnects, and manufacturing processes allow easy translation of high-level specification into reliable representations of the physical materials and operations necessary to implement those high-level functions.

Synthetic biologists are just now beginning to create such tools through the development of standardized parts, “tamed” host cells (chasses), automated cloning (manufacture), and excellent models of these systems. In this book, we will be covering translation of the “physical” layer of biochemistry and molecular mechanism to the “gate” layer of abstraction to digital logic or controlled analog function. In turn we will add an additional abstraction to a device layer, which represents the operation of a functional module such as controlled deployment of a metabolic pathway or mobility system.

Of course biological engineering involves phenomena particular to its substrate, which provides both advantages and disadvantages. For instance, cells can self-replicate and evolve, which on the one hand makes them easy to manufacture and "select" for required function, but on the other hand presents possible problems for containment and reliability. A real biological engineering discipline will need to provide tools to describe and strictly control these phenomena.

Despite the challenges to adhering to this engineering philosophy, we achieve the crucial understanding derived from creation to which Feynman referred. In search of most trustworthy mechanisms and design, we become masters of the medium. This Introduction to Synthetic Biology presents the first steps to this mastery and provides a possible framework for the future of the field.

can use to our advantage
w/o control

Harnessing Biology for Human Purpose: Whence Synthetic Biology?

"Oh, God help us! We're in the hands of engineers…"

Ian Malcolm, the fictional "Chaotician" from *Jurassic Park* and speaker of the quote above, was fascinated and excited by the ability to recreate saurian creatures, but also skeptical of the engineers' assurances of safety. In his story, of course, his fears were well-founded.

In reality, however, we must remember that humans have been drastically changing life to suit our needs since the onset of agriculture thousands of years ago. At that time, "engineering" consisted of controlled breeding to vastly change the form and distribution of a number of plants and animals to meet the needs of growing human civilizations. In modern times, molecular engineering approaches have made these changes swifter, more precise and perhaps a bit more predictable and safer. In addition, modern tools allow genetic manipulations that would not have before been possible. Before we discuss the benefits and potential pitfalls of our relatively recently achieved facility for extremely rapid bioengineering, it will be useful to lay out humanity's history as a shaper of our environment.

A brief history of biological engineering

Our modern biological industries derive from perhaps the longest history of human innovation on earth linking biological, mechanical, environmental/geological, and chemical engineering. Humans arguably began harnessing natural organisms to our purposes over 12,000 years ago. The first domesticated rice is thought to have originated in China around 11000 BC, and domestic wheat was found in the Levant dating from 9800 BC. Techniques for cultivation and breeding spread across the world and were applied to an increasing number of species, including cattle and other animals, as the millennia passed.

In about 4000 BC, Egyptians turned the metabolism of yeast into an engine to leaven bread, a process that about a thousand years later led to fermenting grain and fruits into alcoholic beverages. Of course, these biological innovations were accompanied by technologies that made their manufacture and deployment far easier, including storage systems such as granaries (c. 6000 BC), formal irrigated and enclosed farmlands (c. 5500 BC), and the plough for more efficient cultivation (c. 4000 BC). Later, biologically derived chemicals such as pure distilled ethanol (c. 600 AD) and sugar (c. 900 AD) began to breed new industries. The centuries that followed brought industrial processing methods for these agricultural products, such as wind-driven gristmills, irrigation pumps, factory milling, and more. These innovations enabled people to live in larger urban centers and develop larger tracts of land.

Over this entire period, people slowly improved their crops and livestock, but without a formal scientific basis the results were achieved by trial and error, mating organisms with desirable properties with the hope that those traits would propagate. This was clearly a powerful approach; early humans were able to produce the bloated maize plants we are familiar with today from the grass-like teosinte, and to domesticate wolves into the chaotic variety of modern dogs. However, this magnitude of change was extremely slow, time-consuming, and expensive, and until recently the causal basis of these changes were not understood - making the trust side of the equation difficult to achieve.

The Darwinian evolutionary framework (1859) began to allow people to make some sense of the breeding cycle, but it wasn't until Gregor Mendel published his famous paper on the laws of inheritance in 1866 that the mechanism behind breeding and domestication really started to come into focus. From there, it was a relatively steady march to our current understanding of genetic inheritance. Dutch botanist Hugo de Vries postulated in 1889 that inherited traits were encapsulated in particles he called "(pan)genes," and Englishman William Bateson coined the term "genetic" in 1905. In the early 1900's, scientists observed particle-associated chromosomes

assorting as Mendel had described. Shortly thereafter, Thomas Hunt Morgan proved that genes are on chromosomes, and chromosomes were mapped for the first time.

In 1919, the term "biotechnology" was coined to encapsulate the growing use of plants, animals and microbes for industrial purposes. Interestingly, the first use of the term "synthetic biology" was around 1910 by Stephane Leduc who, in his ambition to dispel any mystical phenomena from the operations of life, suggested it would be possible to synthesize life-like things from chemical and physical components.

Then, around 1944, DNA was proven as the genetic material; in 1953, DNA was shown to be a double helix structure by Watson and Crick; and the genetic code was broken from 1961-1967.

The 1970 discovery of restriction enzymes in the bacterium *Haemophilus influenza* launched modern genetic engineering, defined by the technical ability to cut and paste DNA, which is the direct progenitor of synthetic biology as we discuss it today. In 1973, Herbert Boyer and Stanley Cohen used these enzymes and a ligase to create the first organism with recombinant DNA (DNA molecules formed by laboratory methods).

The NIH, simultaneously impressed and disconcerted by this new ability, formed a Recombinant DNA Advisory Committee in 1974 to oversee recombinant genetic research. The same year, Paul Berg wrote an article in *Science* recommending a moratorium on using this technology until guidelines were established, and researchers voluntarily complied. The next year, top scientists gathered at Asilomar Conference Grounds in Pacific Grove, California to evaluate the state of the new technology and the risks associated with it and establish principles to safely conduct experiments, after which research in this area resumed.

In 1975, Waclaw Szybalski coined the first modern usage of "Synthetic Biology" and posited engineering whole genomes with standard regulatory circuits to deploy useful modules of function. At a more humble scale, in 1976 the first working synthetic gene was created and yeast genes were expressed in *E. coli*, followed in 1977 by the first human gene. Next came recombinant insulin in 1978 and human growth hormone in 1979.

In 1980, intellectual property law surrounding the field hit the big time when Cohen and Boyer were granted the patent on cloning and the Supreme Court approved the patenting of genetically engineered life forms. This started a debate which continues today.

From then on, industry has been growing. Metabolic engineering has taken off, with various companies emerging to market the recombinant biologicals listed above and more from sources

including *E. coli*, yeast, and cow's milk. The medical implications also stretch beyond the simple production of important therapeutics. For example, in 1990 a four-year-old girl suffering from an immune disorder was the subject of the first federally approved and successful gene therapy procedure.

There are many other landmarks and technologies, including DNA sequencing and PCR, for example, that emerged during this time of rapid growth, but agriculture may be the clearest marker for genetic engineering's industrial success. In 1986, tobacco was the first genetically engineered crop to be tested and approved by EPA, followed by cotton, tomatoes, corn, and soy over the next decade. More than 70% of corn and 80% of soy grown in the United States has been engineered. In 2009, genetically modified are crops grown on over 148 million hectares of land worldwide, out of the approximately 5 billion hectares of total agricultural land.

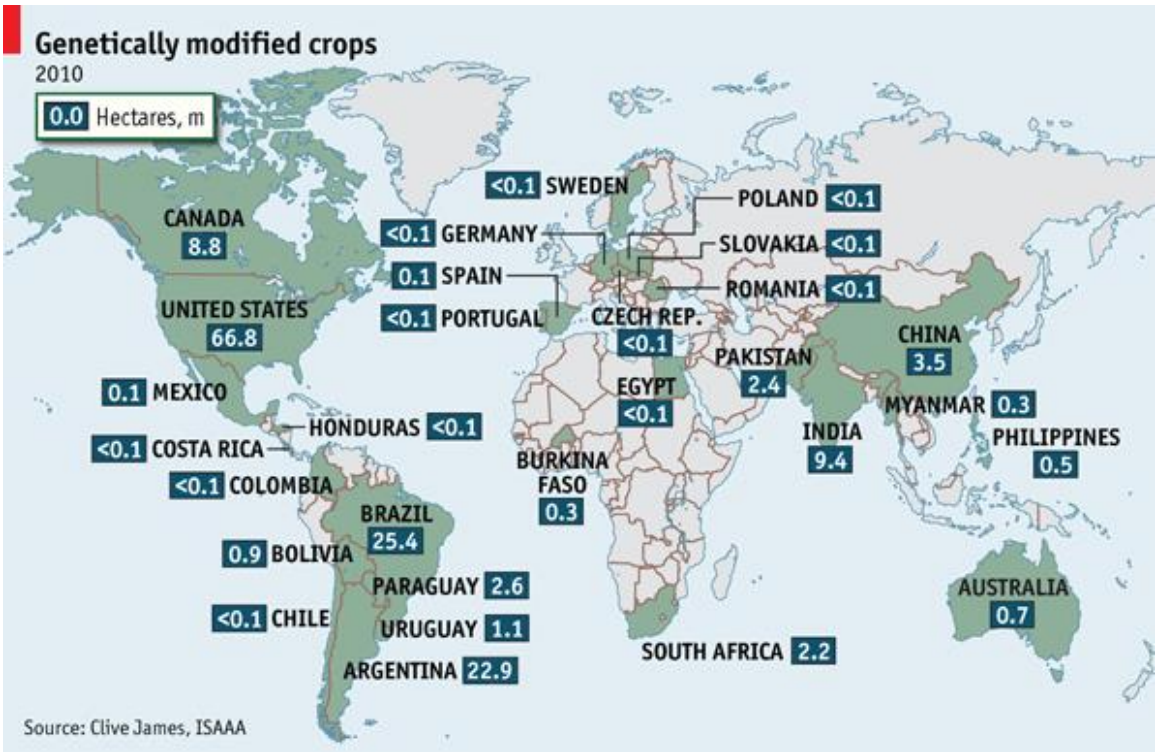

Figure 1. Hectares of genetically modified crops grown around the world in 2010.

The current state of Synthetic Biology

Most of the successes described above required very few modifications of the original organism; modifications in just one or a few genes confer a single or a few unlinked constitutive traits to an

organism or simply produce a useful protein. When Waclaw Szybalski envisioned modern "Synthetic Biology" in 1975, he was thinking beyond the relatively minor genetic tweaks that have provided us with pesticide-resistant crops and easy access to insulin.

Craig Venter and colleagues have demonstrated some successes in genome-scale engineering - in 2008, Craig Venter and colleagues completed the first synthesis of a full bacterial genome, and in 2010 his team effectively transformed a bacterium from one species to a closely related one by displacing its native genome with a synthetic one - but these approaches relied heavily on replicating existing genomes rather than designing from scratch, and therefore still do not fully achieve Szybalski's vision of engineering complete genomes based on standard parts ("circuits") to achieve specific functions.

In fact, despite the many successes of genetic engineering, synthetic biology as Szybalski envisioned has been relatively slow to develop. Specifically, the scaling technologies in biology have not yet driven a similar scaling in the design of the biological systems. The ability to sequence and synthesize DNA and characterize the resulting systems has been growing incredibly rapidly (Figure 2a). This type of scaling is defined in electronics by Moore's law, which says that the number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years and has enabled a vast increase in the complexity of engineered circuits; the growth in biological tools has kept up with, and perhaps even outpaced, Moore's law.

In other words, it has been technically possible to engineer large-scale "circuits" into various organisms, especially bacteria and yeast for many years, but in a 2009 review, Weiss found that the number of promoters (the elements that initiate the expression of genes) in designs was more or less remaining flat.

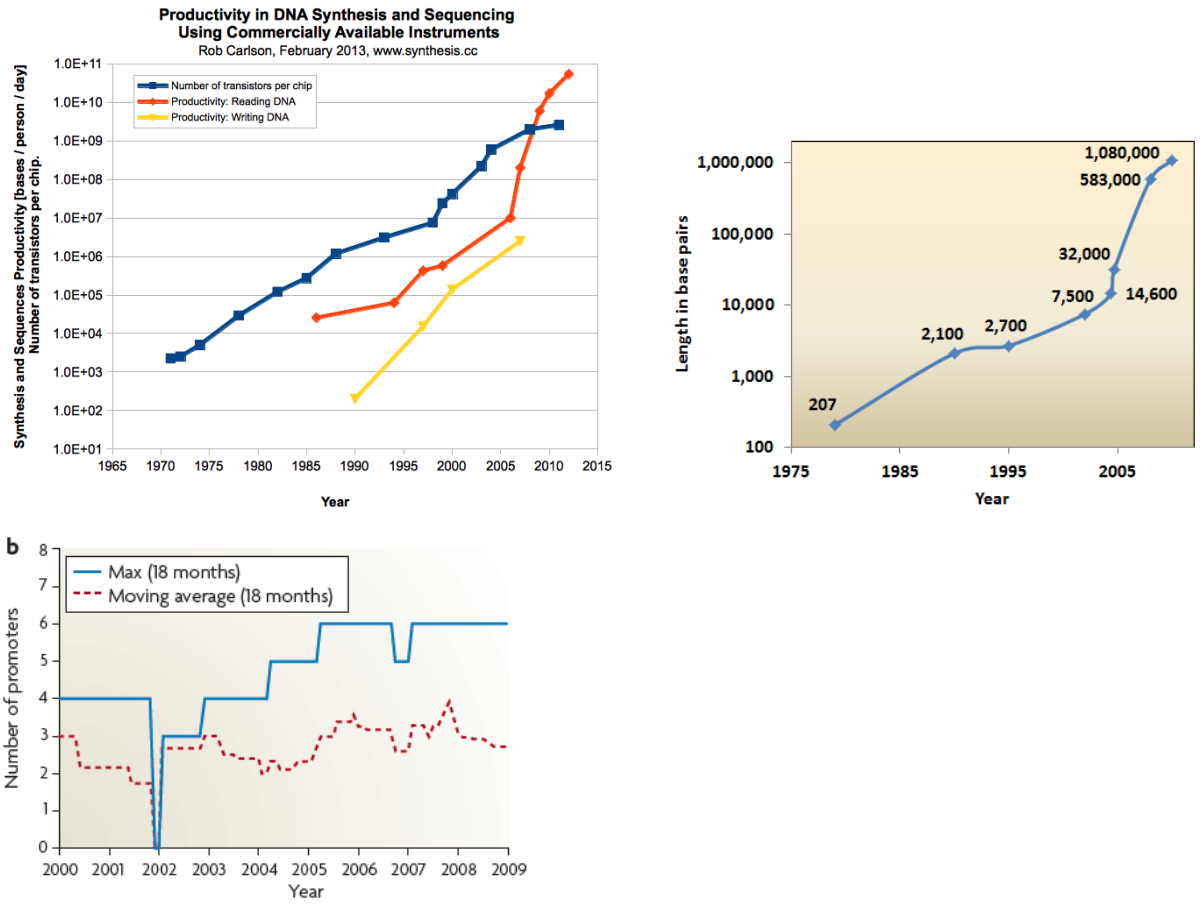

Figure 2. (a) Rob Carlson. (b) Purnick, P.E.M. and R. Weiss, The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol, 2009. 10(6): p. 410-22.

A number of factors may explain why improved technologies haven't led to better or more complex design. There are two classes of challenges: technical and economic. Technologically, until recently there were high barriers in the synthesis, assembly and transformation of DNA into target cells. These barriers have been significantly lowered as implied above. However, an important challenge is predicting how assembled DNA will function in the context of that cell and its environment. In addition, there is not yet a database of sufficiently characterized biological "components" of sufficient functional diversity to enable the scalable design of synthetic circuits. Underlying both of these is the lack of experimental protocols for characterizing these components and their interactions with each other and in the variable contexts in which they are expressed.

Economically, one has to prove that engineered biological routes of solution are better than others and that the engineered solution is efficient and safe. The economic challenges are not trivial. Technologies other than biology might be more effective and currently more acceptable to society. There also might be naturally occurring organisms that provide a solution. If it is an engineered organism, on the other hand, it must be more effective than the natural organism. In some cases it is possible to modify a naturally occurring organism only slightly to achieve the desired goal, lessening the need for larger-scale engineering. In any case, the biological solution must also be a protectable technology for companies to be able to make a profit on it. As an instructive historical example, when the Supreme Court ruled that genetically engineered life could be patented, it was because Ananda Chakraborty engineered a strain of pseudomonas bacteria to be used to remediate oil spills; the engineered organisms broke down oil more efficiently than the four other microbes known to do so.

Though the researchers won the patent case, in the long run they lost out to nature and fear. Exxon thought of deploying these bacteria to the Exxon Valdez spill in 1989, but decided in the end that this approach was too controversial and chose not to. Chakraborty then suggested using the biologically produced surfactants to break up the oil, but these products were also considered controversial despite previous work showing them to be less toxic than those commonly used at the time. More recently, after the 2010 Deep Horizon Spill in the Gulf of Mexico, natural microbial communities broke down a great deal of the oil, so there was just no need for the synthetic organisms.

Perhaps this was a lucky case, but it provides fuel for the argument that the incredible diversity of existing life on the planet can often provide the solutions we need, for now at least. There is a great deal of merit in finding an organism that almost does what one wants and only slightly modifying it by breeding or direct genetic manipulation. This approach harnesses the billions of years of evolution behind the original function of the organisms - but there is also the significant downside that evolution was likely not optimizing for human purpose.

In fact, identification of an existing organism and often inadvertent “laboratory” evolution of improved function has been responsible for a good fraction of industrial biological chemical production, including the organisms that produce our ethanol, many of our antibiotics, and many of our “feedstocks” such as certain amino acids.

In some cases the naturally occurring organism does not produce economically viable amounts of a chemical, or does so only at great expense. The antimalarial drug Artemisinin provides an example of synthetic biology solving these issues by engineering pathways for efficient Artemisinin production in *E. coli* and yeast; ultimately yeast proved to be the preferred host. The

drug, which is one of the last bastions for treatment of malaria, was historically isolated from the wormwood *Artemesia annua*, but this process was very expensive. In addition, most of the needful population are poor and in developing nations, compounding the problem.

Although the active compound was originally found and isolated from a plant, synthetic biologists felt that moving the production pathway into a different organism would provide a production advantage, which could be leveraged to produce the compound more efficiently and offer it at a lower price; specifically, into an organism that grows rapidly, is easily manipulable, where more cellular resources can be diverted to producing the desired output, and for which there are industrial processes in place that allow scaling.

In 2006, Jay Keasling and his team created yeast capable of synthesizing this compound. The team needed to make many different changes, including introducing a crucial gene from *A. annua* into the yeast, controlling the flux of related biosynthetic pathways, and creating a novel enzyme. These changes required a sizable number of modifications of the organism and at first resulted in a low titer of the drug, which has subsequently been greatly improved. As the output is further increased, it is likely to reduce the cost of Artemisinin by more than ten fold. However, the optimization of Artemisinin production required a great deal of trial-and-error, many person-years and tens of millions of dollars because of the technical challenges noted above and others.

The way forward

While the artemisinin case provides an impressive and truly useful example of biological engineering's potential power, it represents a single, relatively simplistic type of application: metabolic engineering, in which feed is transformed step-by-step into a final, desired product. These applications involve pathways with many genes, but the design is usually not complex, and certainly not genome-scale, since only a few of the pathways involved are heterologous transplants. Instead of facing serious design challenges, metabolic engineers' largest challenges are in finding the appropriate enzymes to perform the necessary chemistry, conserving the proper co-factors and redox chemistries, and maximizing yield. The principles of flux optimization and reduction of toxic intermediates make the circuit design a little less trivial than simply expressing a pipeline of enzymes, but there is little behavioral regulation, and the desired phenotype - maximization of productivity - is ultimately a relatively simple one.

Biological engineers frequently use a screening approach, so they can simultaneously evaluate various expression elements and enzymes to create pathway variations that can then be selected

for the desired property. This approach, however, is difficult or impossible when designing more complex phenotypes. Consider designing a gene therapy to restore a metabolic function, a bacterium to identify and destroy cancerous cells, or stem cells to produce a balanced number of pancreatic beta cells for in vivo insulin production. These phenotypes are far more complex than the creation of a single useful biomolecule, and it would be difficult to find existing organisms or cells that precisely accomplish these design specifications. Even if a designer found something close, screening and selecting them for function is impractical, since the phenotypes are complex and don't lend themselves easily to rapid scoring. In addition, in many cases the real selective environment - say, a human host - simply isn't appropriate or safe for library-based selection approaches. Also many of these applications require complex, conditional behaviors dependent on environmental sensing.

Such "programs" require more circuit elements, which if not broken into separately testable and designed modules would make creating screenable libraries impossible simply due to size. In these cases, we would like to rely on predictive design served by a large number of standard elements for sensing, controlling, and actuating behaviors in diverse hosts for diverse applications. This mirrors what has happened in other engineering industries, where every application is a mix of well-understood and modeled commodity parts and specialty pieces. We have not yet reached this point in synthetic biology, even though there are standard plasmids, promoters, enzymes and other elements that *are* common to nearly every design in a given host. We need to scale this along with the other biotechnologies above.

We can summarize much of the argument above in a single diagram. Current cellular and molecular biotechnology can be placed conceptually in a two-dimensional space. One axis represents increasing rational design, ranging from discovering the function you need in nature to designing that function de novo. The second axis represents increasing complexity of the function. At the low end would be what we call "pipes," or simple chemical conversions from one compound to another. At the high end would be "programs" that must, for example, sense multiple factors in the environment and regulate complex responses to achieve the goal. Examples of each of these are shown in the figure below.

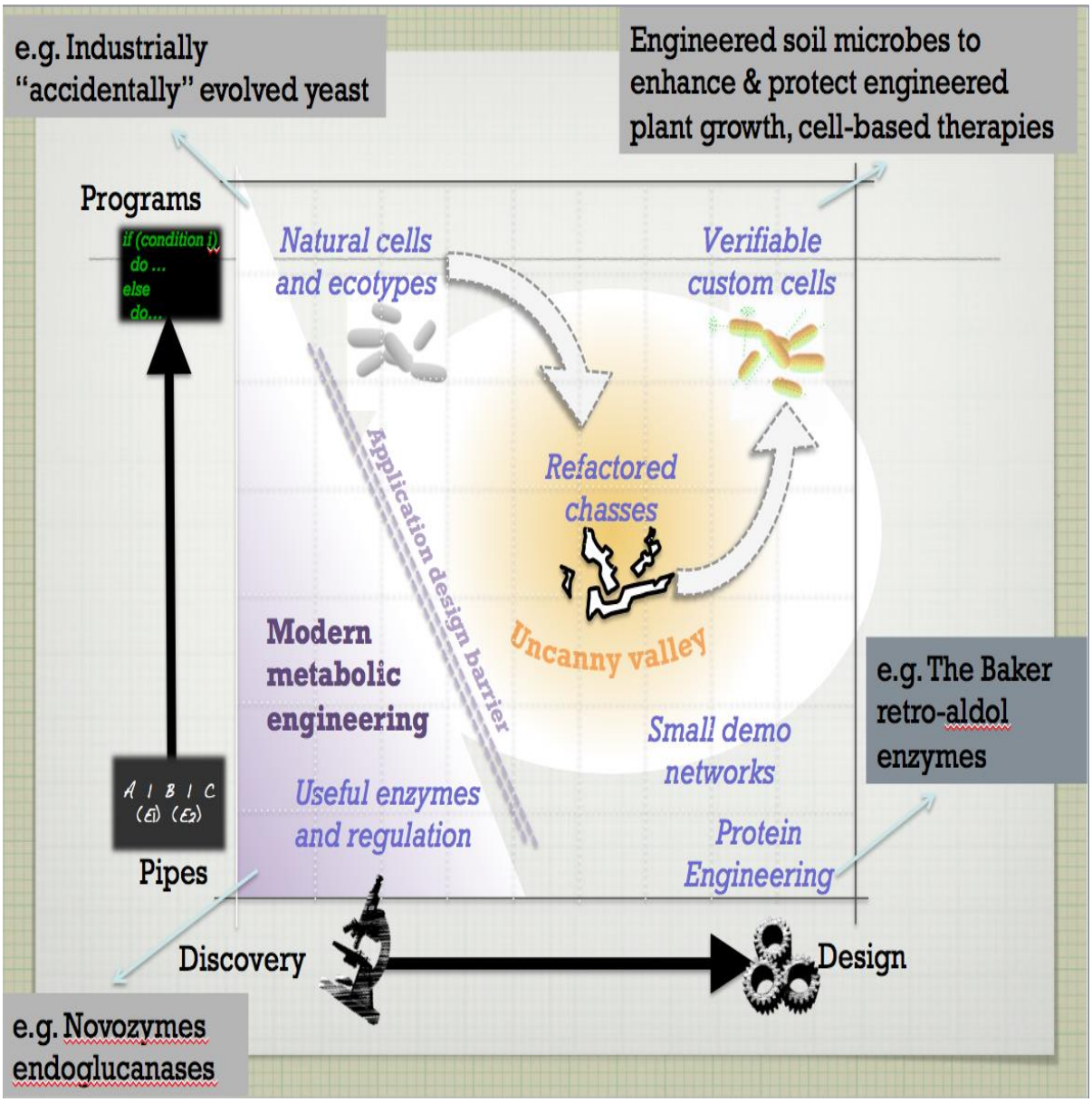

Figure 3. Dynamics in the biological engineering application space. The term 'pipe' refers to the biological transformation of one organic compound to another through an enzymatic reaction, whereas 'program' refers to sophisticated logical systems, such as those that underlie microbial homeostasis in complex environments. Current technological and scientific barriers limit the complexity of the programs that can be reliably designed and proven to be stable and safe after deployment. The systematic development of standard genetic components coupled with emerging methods of multiplexed directed evolution should provide the knowledge and tools necessary to bootstrap a virtuous cycle of genome refactoring and functional expansions. A sustained endeavor will be necessary to cross the uncanny valley of imperfect designs, till the community gathers enough tools and knowledge to enable efficient, robust and verifiable engineering of desired phenotypic behaviors. Cambray, G., Mutalik, V.K., Arkin, A.P. (2012) Towards Rational Design of Bacterial Genomes. Curr. Op. Microbiology, Volume 14, Issue 5, October 2011, Pages 624–630

The design barrier represents a combination of factors. Most industry, for example, sits on the left side of that barrier. Current chemical and pharmaceutical industry focuses on relatively simple single chemical products. The industrial processes for fermentation and production are well worn and supported by decades of experience in the chemical engineering and natural products industries. We are also as a society more comfortable with these highly contained applications where the engineered cells and organisms are safely contained in closed reactors from which the nonliving chemicals are extracted.

Agricultural applications, on the other hand, exist out in the environment, and there has been much controversy about the containment and safety of these designs. In fact, most issues in agricultural and environmental engineering using biology arise from natural organisms displaced into new environments—invasive species including plants and animals, such as cane toads and rabbits in Australia. However, these are still relatively contained and simply engineered systems.

Finally, the size of a designer's possible applications is limited by technology to some extent. As discussed earlier, the ability to synthesize, assemble, insert and modify DNA in some organisms has all but been eliminated as a barrier. Still, development and validation of common design elements—the libraries of standard, reliable, predictable sensors, regulators and actuators—has lagged, making design of complex applications exceptionally difficult.

To truly break the design barrier above we need to develop a synthetic biology design discipline, served by such libraries and computational tools and taking advantage our ability to tune our designs by growth and evolution, by which we can create new industries to support the complex applications that are largely beyond the bioreactor. We must prove that we can predictably produce safe, reliable function to cross the uncanny valley above and reach the true genome-scale applications Szybalski envisioned.

Thus we define synthetic biology as that discipline that creates methods for making the engineering of biological systems vastly more efficient, reliable, scalable, transparent and safe. Of course, it is also aims to be that discipline which engineers effective biological solutions to societal needs.

Bibliography

Purnick, P.E.M. and R. Weiss, The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol, 2009. 10(6): p. 410-22.

Cambray, G., Mutalik, V.K., Arkin, A.P. (2012) Towards Rational Design of Bacterial Genomes. Curr. Op. Microbiology, Volume 14, Issue 5, October 2011, Pages 624–630