

Enabling the Synthetic Biology Revolution

By Matthew Hill, PhD

The Team at Elegen Knows that when barriers to the fast, accurate, and low-cost synthesis of long DNA fall, biotechnology will be driven by the design-build-test-learn cycle



Matthew Hill, PhD
Founder and CEO, Elegen

DNA is the language of life, and DNA sequencing has allowed us to dive into the genetic code that evolved over the last 3.5 billion years. As sequencing has become faster and cheaper, the scale of projects and the types of questions we can answer have grown enormously. In 2003, we sequenced the first genome at a cost of about one billion dollars, and the effort took more than a decade. Now, 20 years later, we can sequence a genome in a week for \$1,000—a millionth the cost, as well as a thousandth the time and effort. This dramatic increase in efficiency has enabled a revolution that allows us to read whatever DNA we choose.

But the real genetic revolution has yet to take off. For more than a decade, thought leaders have envisioned a world powered by “synthetic biology,” where scientists harness and retool biological systems to create bio-based solutions to many of humanity’s most vexing challenges: disease, aging, hunger, and dependency on fossil fuels for power generation and chemical production. However, we will truly achieve fluency in DNA, the language of biology, only when we can both understand and write this language. This level of proficiency will remain out of reach until we increase the efficiency and reduce the cost of DNA synthesis by many orders of magnitude.

Mastering a complex system involves an iterative design strategy: developing an initial prototype, testing that prototype, analyzing its performance, learning what did and didn’t work, designing a new prototype, and iterating. This “design-build-test-learn” (DBTL) cycle, long used in classical engineering, aims to rapidly improve quality and functionality. Importantly, the learning is exponential, compounding with each iteration, like interest.

However, a rapid DBTL cycle in biology remains largely aspirational. Advances in sequencing, high-throughput screening, and machine learning have greatly accelerated the design, test, and learn phases—in some cases, they can all be accomplished in a week. Frustratingly, the build phase—physically

writing the DNA—has remained stubbornly slow, expensive, and cumbersome. To synthesize DNA quickly, inexpensively, and accurately enough to generate DNA strands long enough to constitute full genes, there are several roadblocks to overcome.

Building long DNA

The production of long, double-stranded DNA for synthetic biology involves three fundamental steps. First is the production of the initial oligonucleotides (oligos), typically 50–200 bases in length. Second is the assembly of sets of oligos into longer double-stranded DNA on the order of 1,000–1,500 base pairs (bps). Third is the isolation of pure DNA sequences. Because errors in both synthesis and assembly occur, intermediate and final fragments are often cloned to isolate and amplify sequence-perfect molecules. Production of sequences greater than 1,500 bps often involves multiple rounds of assembly and cloning.

These workflows are complex and harbor challenges that have, as of yet, resulted in the failure to deliver fast, accurate DNA of arbitrary complexity and length, at a reasonable cost, all of which is required to enable a rapid DBTL cycle. Here, we review the four main challenges.

The first challenge is the scaling and parallelization of the initial oligo synthesis. Substantial effort over several decades has been put into automation and multiplexing, resulting in DNA synthesis machines that can make hundreds of thousands of oligos in parallel, have high product consistency, and achieve lower costs through economies of scale. However, it is important to note that current approaches don’t maintain per-oligo costs with variable size batches. DNA synthesis workflows must be run at full capacity to realize low costs, much like how flow cells put a minimum cost on a sequencing run, no matter how small the sequencing job. Another challenge is the solid-phase reagent-flow nature of the chemistry, requiring large excesses of reagent, which can be very costly.

The second challenge is synthetic fidelity. Current-

ly, error rates for initial oligo synthesis are about 1:200, which is adequate for 1,000-bp oligos; however, this rate is 100- or 1,000-fold too high when the goal is to synthesize entire genes. Most oligos are produced using traditional phosphoramidite chemistry, the industry standard for four decades. Enormous efforts have been spent optimizing phosphoramidite chemistry with great success; however, further efforts are likely to yield diminishing (and even minimal) improvements. Error correction has historically provided a 2–5-fold improvement on top of that.

Research into alternative chemistries has promised to deliver high-fidelity long DNA at lower cost. For example, enzymatic DNA synthesis, a promising candidate, has attracted tens of millions in investment dollars. Although it is faster than phosphoramidite chemistry and avoids using harsh chemicals during synthesis, it has not yet improved on the cost or accuracy issues. Fundamentally new approaches are needed to bring DNA error rates to the range where routine gene synthesis is possible.

The third challenge is the complexity of oligomer assembly, particularly in the synthesis of sequences with high guanine-cytosine content, homopolymers, or sequence repeats. This challenge is important to address because many organisms, plants in particular, rely on these difficult sequences. Although no single innovation is likely to address this issue, the more these challenges are ironed out, the more flexibility biologists will have in how DNA is segmented and subsequently assembled.

The fourth challenge—and primary impediment to the production of rapid, long synthetic DNA—is cloning, which has remained largely unchanged for decades. The procedure is performed to overcome the errors that accumulate in the synthesis and assembly steps. In cloning, transformed cells are plated onto agar; successful transformants—which harbor a single original copy of the DNA—grow clonally. However, the need to plate clonal cells, along with the requirement to pick and sequence the mul-

tiplicity of colonies, introduces significant barriers to a truly high-throughput and low-cost workflow. This is further complicated when building DNA longer than 3,000 bps, which requires sequential iterations of assembly and cloning. Cloning accounts for the majority of both the cost and time in long DNA synthesis and is ripe for disruptive innovation. Without improvements in cloning, it will be hard to deliver on the promise of a rapid DBTL cycle.

Because of these challenges, today's synthetic DNA supply chain remains complex, fragmented, and slow. As such, nearly every life sciences company maintains a biofoundry, sometimes very large, employing highly skilled labor to manipulate DNA. But if the challenges we describe above are overcome, biofoundries could become a thing of the past. Widespread availability of long, fast, accurate DNA would finally unlock the DBTL cycle and accelerate advances across the entire life sciences.

The future of synthetic biology

The synthetic biology market has the potential to be valued at more than a trillion dollars and is projected to achieve double digit growth for the foreseeable future. The demand for synthetic DNA, a key part of that market, is shifting from short oligos to much longer, gene-length sequences. Antibody therapeutics companies already use large libraries of synthetic genes to perform affinity maturation of antibodies. In oncology, engineered chimeric antigen receptor

(CAR) T cells have been approved to treat leukemia and lymphoma, and require engineering of longer sequences yet. Outside of healthcare, companies are using synthetically manufactured DNA to generate high-value enzymes, chemicals, nutraceuticals, plant-based meats, new materials, and fertilizers.

The availability of on-demand, inexpensive, high-fidelity long DNA will democratize the ability to engineer organisms and biologic products to provide materials, energy, food, and therapeutics to the world. Although this shift may take decades to realize, it will redefine how people's needs are met, and thus how the economy functions. Much like 3D printing promises to revolutionize the production of shaped solids, rapid long DNA synthesis promises to revolutionize the production of organic materials and the engineering of organisms, allowing the onshoring of many supply chains and the local sourcing of key materials, thus reducing dependence on global supply chain stability.

Humanity will reach a critical milestone when we can rapidly reprogram or reengineer biological systems. Transforming biology into engineering requires innovation that directly addresses the inefficiencies in the DNA manufacturing process that have held the field back. Fully unlocking synthetic biology's promise will require rapid and substantial advancements in the ability to accurately write DNA of any length and complexity in days, at scale, and at a reasonable price. **GEN**



To improve the synthetic DNA supply chain, Elegen has developed a DNA production system that can deliver high-accuracy linear DNA longer than 7 kilobases. Moreover the DNA can be shipped in 6–8 business days. The company asserts that widespread availability of accurate, long, and fast DNA will unlock the design-build-test-learn cycle and accelerate advances across the life sciences.