



## Four Years of Generate: The Leading Edge of a Revolution

September 21, 2022

By **Molly Gibson, Ph.D.**

Co-Founder and Chief  
Strategy and Innovation

Officer, Generate

Biomedicines, Senior

Principal, Flagship Pioneering

**Gevorg Grigoryan, Ph.D.**

Co-founder and Chief

Technology Officer,

Generate Biomedicines

[generatebiomedicines.com](http://generatebiomedicines.com)

When Generate was founded four years ago, we promised a digital revolution in biology. Our plan was to apply machine learning (ML) at scale to uncover general principles of biology directly from observations and to use that information to generate novel proteins that could become breakthrough therapeutics.

At the time, this sounded like science fiction. In general, scientists in the biopharma industry were skeptical: they believed that proteins were just too complicated and available datasets too small to enable data-driven learning of fundamental principles.

Four years after launching Generate, that skepticism has faded. Our advances in ML have proven that instead of just discovering biologic drugs, we could generate them digitally. Our ideas have been broadly embraced, and we find ourselves leading a revolution in drug development. We have shown we have the tools and infrastructure to bring these new possibilities to life.

### Leading the way

What we proposed four years ago was to change the prevailing dogma of how proteins and their properties are understood. The long-standing “bottoms-up” approach used imperfectly known atomistic principles to simulate and predict protein behaviors. We proposed a “top-down” approach that inferred emergent, general principles directly from data. This enabled us not only to predict protein behaviors, but also to build novel proteins with desired functions. We’ve been able to generate data, repeatedly and systematically, that support our approach.

First, we have found sufficient data to discover the generalizable principles that govern a protein’s sequence-structure relationships. Using those principles, we have been able to predict the emergent properties and higher-order patterns that result when amino acids are linked together in a protein sequence.

Moreover, we have shown that general principles inferred from one set of proteins can be used to make predictions about a completely different set of proteins, including new therapeutic modalities. This leads to a novel and exciting finding: proteins, rather than being too complex to characterize precisely, are fully programmable.

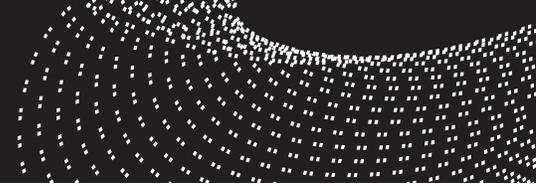
Using that insight, and the iterative learning that comes from applying our digital platform, we are able to produce novel sequences to design any

desired structure – and we’re now extending that capability to yield any desired function. We’ve generated entirely novel proteins with specific, pre-determined functions, based on models built from data alone.

In parallel to our own work, we are energized by the recent acceleration in the application of ML to proteins. A salient example is DeepMind’s winning of the Critical Assessment of Structure Prediction (CASP) competition in 2020, where the performance of their ML method at predicting structures of natural proteins far surpassed any previous approaches. This result and other successes in the application of ML to protein science mean that our initial idea of generalizations from data is catching on in the wider scientific community. It’s no longer science fiction; it’s becoming mainstream science.

### Stepping out into the white space

The skeptics asked one more question: maybe you can build novel proteins that have never existed before, but will they really work as biomedicines? The answer: yes, they can, and they do. While our computer-generated proteins have never existed before, they still function according to the same natural principles as those that nature itself has found through evolution. In short, our data reveal there is nothing about these computer-



## Four Years of Generate: The Leading Edge of a Revolution

September 21, 2022

By **Molly Gibson, Ph.D.**

Co-Founder and Chief Strategy and Innovation

Officer, Generate

Biomedicines, Senior

Principal, Flagship Pioneering

**Gevorg Grigoryan, Ph.D.**

Co-founder and Chief

Technology Officer,

Generate Biomedicines

[generatebiomedicines.com](http://generatebiomedicines.com)

designed proteins that is inherently weird, unnatural, or biologically incompatible.

Biologics are generally discovered by exploring the proteins that have already evolved in nature to perform certain physiological functions – and this approach has led to many spectacular successes in new medicines. However, the number of possible protein sequences is vast (proteins can be hundreds of amino acids long, and there are 20 different amino-acid types), and nature could only have sampled a tiny fraction of that space – far less than the equivalent of one atom in the entire universe. Unsurprisingly, we have found that the vast “white space” untapped by nature is brimming with novel and useful protein hypotheses that are biologically viable and relevant to human health.

For example, our immune system has evolved to create antibodies that are perfectly designed to ward off infections of many different types. With our computational platform and integrated experimental research, we have discovered improved antibodies that target and bind to parts of antigen molecules that the immune system doesn’t find. By applying the generalizable principles that govern the immune system’s reactions to proteins, we’ve been able to create “stealth” proteins that aren’t attacked by the immune system’s defenses but will bind to antigens and have important therapeutic effects.

The novel antibodies we have created meet the same, and often improved, metrics for developability and function as antibodies naturally produced in the human body or discovered in phage-display or other libraries.

### Setting aside the filters

The new generative-biology approach we are pioneering will drive a shift in how the biotechs of tomorrow must operate and think about discovery. Traditional drug development proceeds in a series of steps, where at each step the focus is on molecules that successfully

pass a certain filter (e.g., a screening assay) and not the molecules that fail. But for Generate, working from first principles means setting aside the traditional filters and seeing every experiment as a data-generation opportunity, regardless of the success or failure of the molecule itself.

Consequently, we think differently about computational capabilities, experimental standards, and pipeline building. Our approach leads to a different paradigm, one where the discovery step is highly efficient and overcomes the challenges of economies of scale that traditional pharmaceutical drug development has faced.

We are building a company to do all this, not just with technology, but with people who share our mindset and our vision. We seek people from many scientific and engineering disciplines who question the status quo and prevailing assumptions, and people who are excited by working at the frontiers of biology and digital technology. We are building an organization that goes beyond conventional business models and pushes the limits of what can be done in this space.

We stand on the leading edge of a revolution in how new therapeutics are developed. Soon, we expect to see new drugs entering the clinic that are completely computer-generated. Those medicines will transform the entire value chain from the clinic, to regulatory approval, to the patient’s bedside.

We believe that generative biology has the power to create medicines faster, more precisely, and at a fraction of the cost of traditional approaches. This will increase patients’ access to innovative, transformative, and novel therapeutics. The revolution has just begun – we can’t wait to see where we are in four more years!