TWIST BIOSCIENCE

APPLICATION NOTE

Protecting the data well: DNA tools for generating a "Lab-in-the-loop" system for antibody design

INTRODUCTION

Therapeutic antibodies have risen to prominence in recent years as one of modern medicine's most promising tools, with more than 200 antibody therapeutics currently in the market and more than 1,400 in clinical trials.[1] As large protein complexes, antibodies possess distinct pharmacological properties, including greater specificity, selective distribution, and longer half-lives.[2] Yet, these physical properties make the design of safe and effective antibodies a considerable challenge.

Antibody behavior is dictated by complex intramolecular interactions. Altering the protein's amino acid sequence may help optimize for one pharmacologically relevant property at the expense of another. Therefore, designing clinically viable therapeutic antibodies requires researchers to find a balance across multiple competing properties, iteratively searching a vast and sparsely functional sequence space in search of the rare harmonizing sequence. The time and expense required for such a process greatly reduces productivity.

To address this, scientists behind a recent publication developed a new approach—known as *Lab-in-the-loop* (LitL)—that leverages generative machine learning, property prediction, experimental validation, and adaptive model retraining to rapidly hone antibody designs. This collection of tools, along with a unique closed-loop system that utilizes Twist Bioscience's Gene Fragment synthesis, allows researchers to rationally and efficiently navigate a vast protein sequence space.

> **"The linear DNA based expression workflow trims multiple cloning steps, saving significant time and cost while also being highly amenable to automation."**
>
> Frey, Nathan C, et al.

THE CHALLENGES OF ANTIBODY DESIGN—AND WHY MACHINE LEARNING ALONE ISN'T ENOUGH

Among the many challenges of designing a successful clinical antibody is the complexity of optimizing multiple, sometimes conflicting properties (such as affinity, specificity, expression, viscosity, solubility, or immunogenicity), and doing so without deviating from non-immunogenic frameworks. Traditional approaches have relied on brute force screening of antibody variant libraries and subsequent iterative refinement of potential hits in order to find a single viable candidate. Even when rationally designed libraries are used, this approach makes evaluating and ranking candidates across multiple properties a difficult, costly, and time consuming endeavor.

Machine learning (ML) has emerged as a promising tool to alleviate this pain and help the productive navigation of sequence space *in silico*. Capable of identifying complex multidimensional patterns in datasets that human observers overlook, ML design tools learn the language of proteins and design libraries that are more likely to deliver the candidates with the desired properties.

However, limitations of training datasets result in most models optimizing for individual properties (such as binding affinity) and ignoring therapeutic constraints. As a result, ML-generated antibodies still require considerable optimization investments that dull the technology's impact on development speed and efficiency.[3]

Researchers from a leading biotechnology corporation set out to overcome these limitations by integrating ML tools into a closed-loop, design-test-learn system. By adding experimental feedback in ML training data, the antibody design process transforms from a static prediction task into an active learning problem where each round of testing informs the next. This empowers the system to rapidly hone antibodies for multiple design properties.[3]

For Research Use Only. Not for use in diagnostic procedures.

BUILDING A CLOSED-LOOP DESIGN SYSTEM WITH TWIST DNA

Typically, antibody development, even with ML, is carried out through open-loop systems. Therein, ML technology is used to design vast libraries of antibody sequences that are then screened and optimized, all without feeding results back into ML training data sets. A closed-loop system uses wet-lab results to refine ML, ultimately resulting in an enriched pool of high value candidates.

One benefit of a closed loop system is that it allows researchers to improve computational antibody design and greatly reduce the number of candidates that need to be synthesized before a potential hit is discovered. Rather than synthesize billions of antibodies, researchers can reduce their antibody pool to thousands, significantly cutting down on experimental costs.

To do this, researchers need the ability to faithfully connect an antibody's sequence to its performance in wet-lab experiments. This means reliably synthesizing diverse antibody variants—each with precise user-defined complementarity-determining regions (CDRs).

The researchers utilized Gene Fragments from Twist Bioscience to encode vH and vL sequences used in system validation. Twist rapidly delivers high-accuracy, high volume synthesis of linear DNA, which the team used to encode their custom CDR regions. Subsequently these sequences are directly assembled into linear expression cassettes via Gibson assembly. The value of this approach is outlined in the article: "The linear DNA based expression workflow trims multiple cloning steps, saving significant time and cost while also being highly amenable to automation." In turn, the researchers could express designed antibody variants in HEK293 cells at a 1 mL scale, ensuring the parallel generation of hundreds of high quality datapoints in a 4-6 week timescale (**Figure 1**).

Since LitL retrains its predictive models and generative engines on each new round of assay data, any mismatch between intended and actual DNA sequence could mislead the system. Twist's platform helped ensure that what the ML model designed was faithfully represented in the wet lab, safeguarding the integrity of the loop.

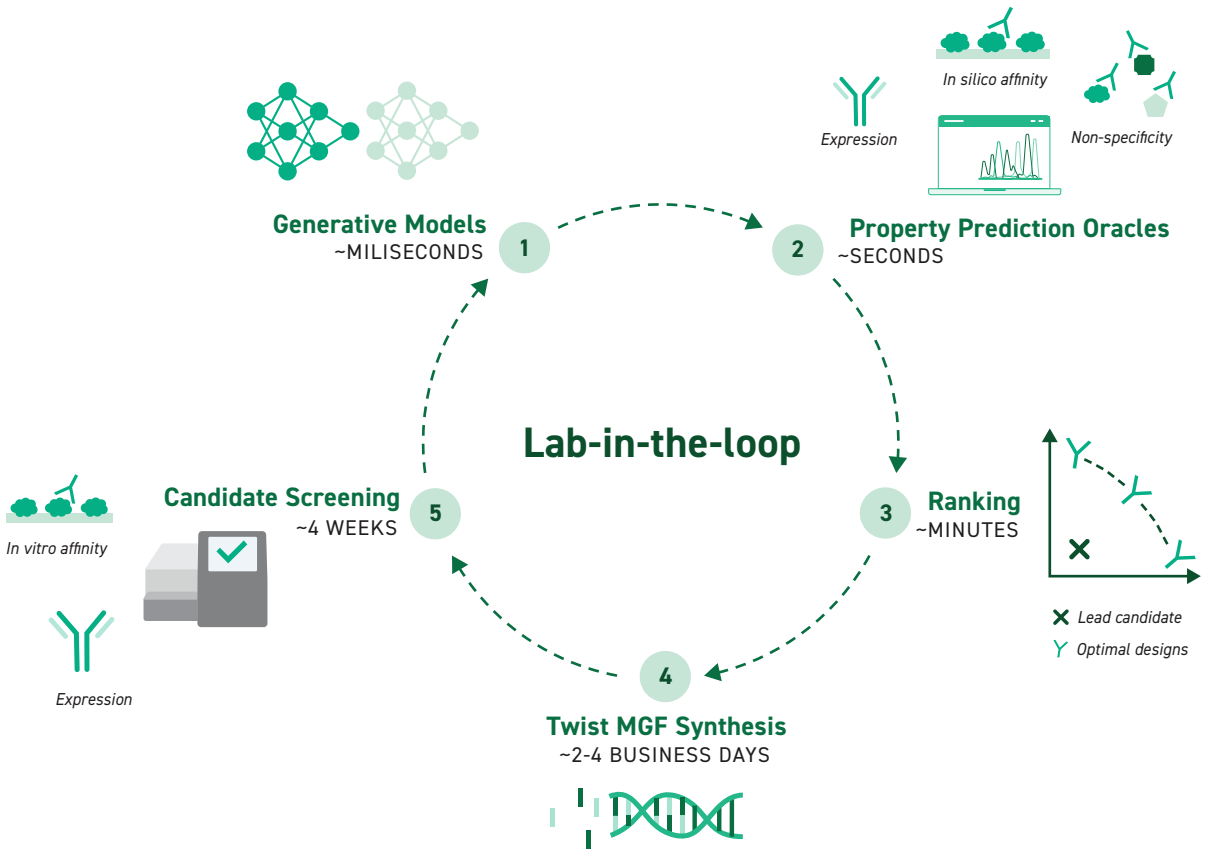

Figure 1: Lab-in-the-loop workflow. Step 1: generative machine learning models build thousands of candidate sequences based on an original lead candidate. Step 2: Generated designs are computationally modeled to predict various properties, including expression, non-specificity, and affinity. Step 3: Ranking based on model predictions is used to identify designs that fall in the Pareto frontier (dotted line), representing maximal improvement across multiple properties compared to the original lead candidate.Step 4: Complementarity determining regions (CDRs) are ordered as Twist Gene Fragments for each optimal design. Step 5: CDRs are assembled into linear cassettes and used to express antibody candidates in HEK293T cells. Binding affinity of expressed antibodies is evaluated with surface plasmon resonance, data from which is used to retrain generative models.[3]

THE LitL PIPELINE: DESIGN, BUILD, TEST, RETRAIN

The LitL system begins with a lead antibody—typically one discovered through animal immunization. A suite of generative ML models, both guided and unguided, then produce thousands of candidate sequence variants. Guided methods optimize explicitly for multiple properties (e.g., affinity, expression, developability), while unguided approaches emphasize diversity and exploration of the sequence space.

The resulting variants are ranked using a set of predictive "oracle" models that estimate performance across several axes. The highest ranked candidates are then synthesized and expressed for *in vitro* testing, specifically via surface plasmon resonance (SPR). This allowed the researchers to assess how each candidate performed with antigen-binding kinetics. The resulting data is then fed back into the ML models, which use the information to iterate and design new, improved screening libraries.

Over the course of four iterative rounds, LitL generated more than 1,800 unique variants against four antigen targets—EGFR, IL-6, HER2, and OSM. It identified at least one antibody with 3x improved binding affinity for each target, and several with improvements of 10x-100x. Experimental structure determination confirmed that these optimized antibodies retained their original binding geometries while incorporating beneficial mutations that enhanced affinity and minimized developability risk.

CLOSING THE LOOP ON THE FUTURE OF ANTIBODY DESIGN

The LitL platform represents a major step forward in ML-guided therapeutic design. By integrating predictive modeling, high-throughput screening, and automated retraining in a single pipeline, it achieves what many earlier ML approaches could not: genuine iterative learning.

As AI tools continue to expand the frontier of antibody design, Twist's synthesis capabilities will remain a critical bridge between theoretical potential and experimental reality—helping researchers transform digital predictions into clinical candidates with the necessary speed and precision.

REFERENCES

1. Crescioli, Silvia, et al. "Antibodies to Watch in 2025." MAbs, vol. 17, no. 1, 22 Dec. 2024, https://doi.org/10.1080/19420862.2024.2443538.

2. Ovacik, Meric, and Kedan Lin. "Tutorial on Monoclonal Antibody Pharmacokinetics and Its Considerations in Early Development." Clinical and Translational Science, vol. 11, no. 6, 7 Aug. 2018, pp. 540–552, https://doi.org/10.1111/cts.12567.

3. Frey, Nathan C, et al. "Lab-In-The-Loop Therapeutic Antibody Design with Deep Learning." BioRxiv (Cold Spring Harbor Laboratory), 24 Feb. 2025, https://doi.org/10.1101/2025.02.19.639050.