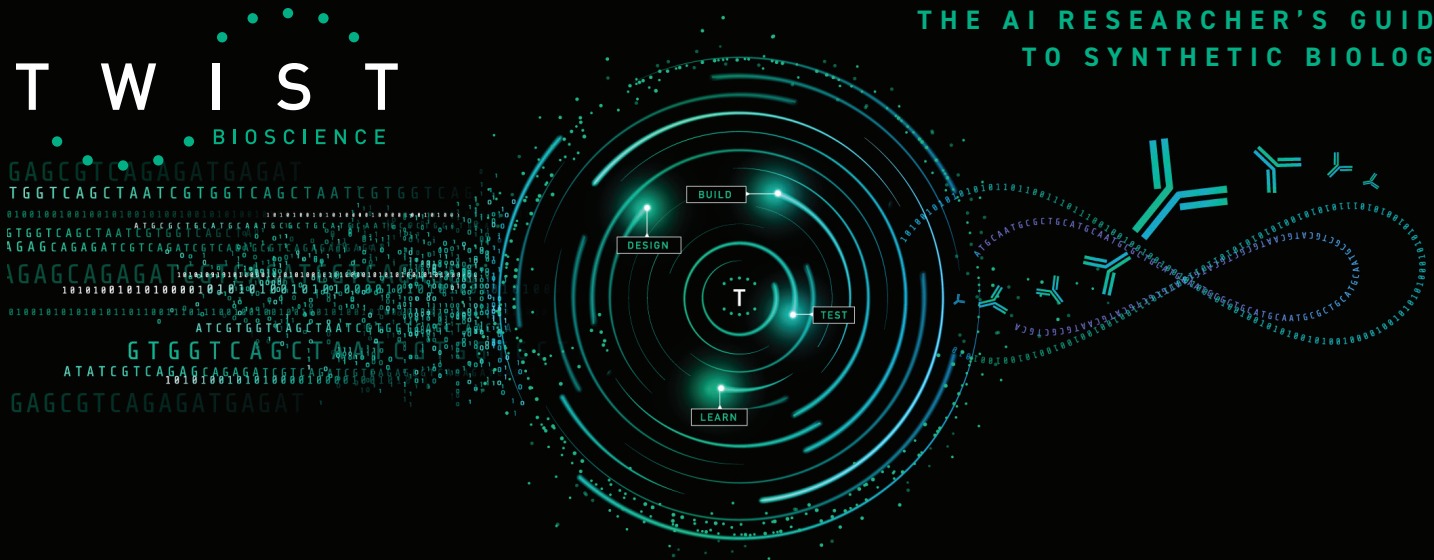
TWIST BIOSCIENCE

THE AI RESEARCHER'S GUIDE TO SYNTHETIC BIOLOGY



Unlocking the full potential of AI/ML with wet lab validation

Large language models (LLMs) are revolutionizing drug development and agricultural science, but their potential is only realized when they move beyond the computational domain. AI is a powerful tool that augments the wet lab, it can highlight where to edit a genome or how to design a therapeutic protein, but it cannot synthesize them. Acting on AI's guidance with the same level of accuracy as the digital design is the fundamental challenge researchers face today.

Even the best AI/ML models can have unexpected flaws that, if uncorrected, undermine their utility. Twist's silicon-based synthesis platform ensures that these models are validated against the exact sequences designed in silico, preventing subpar training data caused by synthesis errors. Too often, the success of models that perform well in a computational setting fails to translate into the laboratory environment. By testing model output in the wet lab, researchers can establish the validity of their model's predictions and, when needed, perform retraining to improve performance. In many cases, these studies rely on synthetic DNA, the quality of which can make or break the model.

Why is wet lab validation necessary?

In short, because no model is perfect. Consider the example of AI/ML tools designed for protein engineering. Validation is necessary for these tools because of the following:

LIMITED TRAINING DATA. AI models often make complex extrapolations from imperfect datasets, leading them to 'over-index' on single properties while underrepresenting critical constraints like stability or solubility.

EMERGENT BEHAVIOR. Proteins are governed by complex intra- and intermolecular interactions that may be absent in training data. Certain combinations of mutations can result in protein misfolding, self-aggregation, or unexpected glycosylation's that affect protein performance.

CONTEXT SPECIFICITY. Models trained on data from a specific context may not translate to another. Changes in expression systems (*E. coli*, HEK293, CHO), for example, may expose constraints that the model never "saw" in training.

BLOG POST: *AI's Future Hinges On The Wet Lab*

Design-build-test-learn

It's notable too that wet-lab testing doesn't just confirm model validity—it can feed new data back into the model (**Figure 1**). This iterative loop (design → build → test → learn) expands the model's focus to encompass relevant properties while improving downstream efficiency.

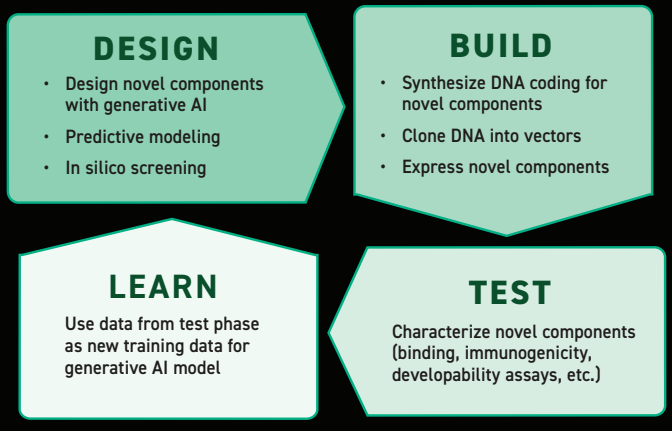


Figure 1. Illustration showing the Design-Build-Test-Learn cycle that enables rapid improvement of AI-model performance.

Case study: Lab-in-the-Loop

Researchers at a leading biotechnology company recently developed a new ML-driven process for antibody development. The model is a closed loop system (referred to as lab-in-the-loop, or LitL) in which generative ML designed antibody variants which are then ranked according to predictive modeling, synthesized using an industry leading platform, and subject to *in vitro* characterization (**Figure 2**). That data is then used for adaptive model retraining, thereby enabling rapid model improvement.

Over four iterative rounds, LitL generated 1,800 unique antibody variants across four target antigens and identified several antibodies with 10x-100x improved binding affinity.

Part of this success was due to the teams ability to faithfully synthesize ML-designed antibody variants.

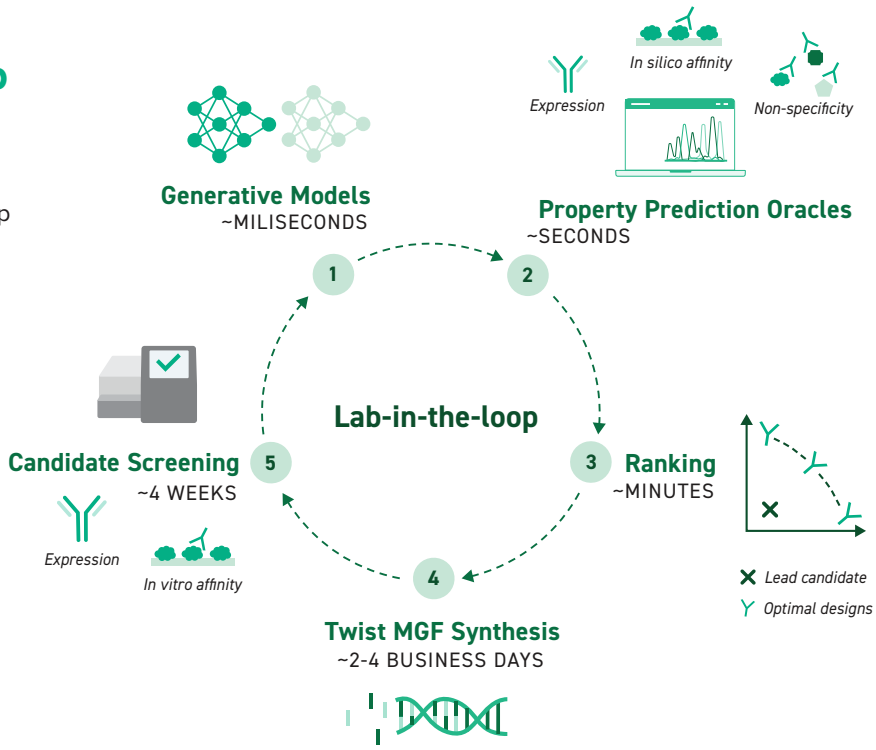


Figure 2. The lab-in-the-loop (LitL) workflow orchestrates generative AI design, property prediction, and high-fidelity Twist DNA synthesis to transform static predictions into an active learning loop that rapidly optimizes antibody performance.

READ MORE:
DNA tools for generating a "Lab-in-the-loop" system for antibody design

The importance of synthetic DNA

AI models are often asked to make complex extrapolations from imperfect training data, which can lead to models 'over-indexing' on a single property like affinity while ignoring critical constraints like stability or solubility. To overcome this, researchers must establish a wet-lab feedback loop where AI predictions are put to the test and the resulting data is used to refine training, transforming design from a static task into an active learning problem.

Translating precise digital designs into physical molecules is not trivial. Traditional synthesis often relies on stitching small 150-300bp fragments, an error-prone process that can misrepresent sequences and mislead researchers. Twist's silicon-based platform enables the direct and rapid DNA synthesis, ensuring physical oligos faithfully represent AI intent and preventing subpar training data from hindering performance.

Twist solutions for AI research

SYNTHETIC DNA

Gene Fragments: Shipped in as fast as 2 business days* to enable rapid "Design-Build-Test" iterations.

Gene Pools: Large-scale pooled DNA up to 1.8 kb, enabling the interrogation of massive sequence diversity for complex library validation.

Multiplexed Gene Fragments: Arrayed or Pooled dsDNA up to 500 bp. Precisely designed to span full antibody CDRs or protein domains for massive parallel screening.

Variant Libraries: Site-saturation or combinatorial variant libraries starting at 300 bp. Precisely explore a protein's affinity landscape with rational control over every mutation.

Oligo Pools: Enable ultra-high–throughput functional interrogation for millions of variants up to 350 nt in parallel.

Advancing discovery with end-to-end antibody solutions

For laboratories indexed on computational power rather than wet-lab infrastructure, Twist provides a seamless transition from sequence to data. We leverage automation-scale infrastructure to produce large panels of antibody candidates and validate them across binding, affinity, and developability assays. By delivering high-quality, standardized datasets, we remove the onus of wet-lab validation and empower researchers to focus on design.

This end-to-end service bridge connects in silico design, spanning antibody/antigen sequences and structural modeling, to physical validation. The workflow utilizes AI-based tools for sequence generation and candidate scoring, followed by high-throughput screening and experimental validation to deliver optimized lead sequences (**Figure 3**). By integrating Twist's custom data delivery, researchers receive model-ready datasets from standard controls in as few as 20 days, enabling a continuous retraining loop for AI models.

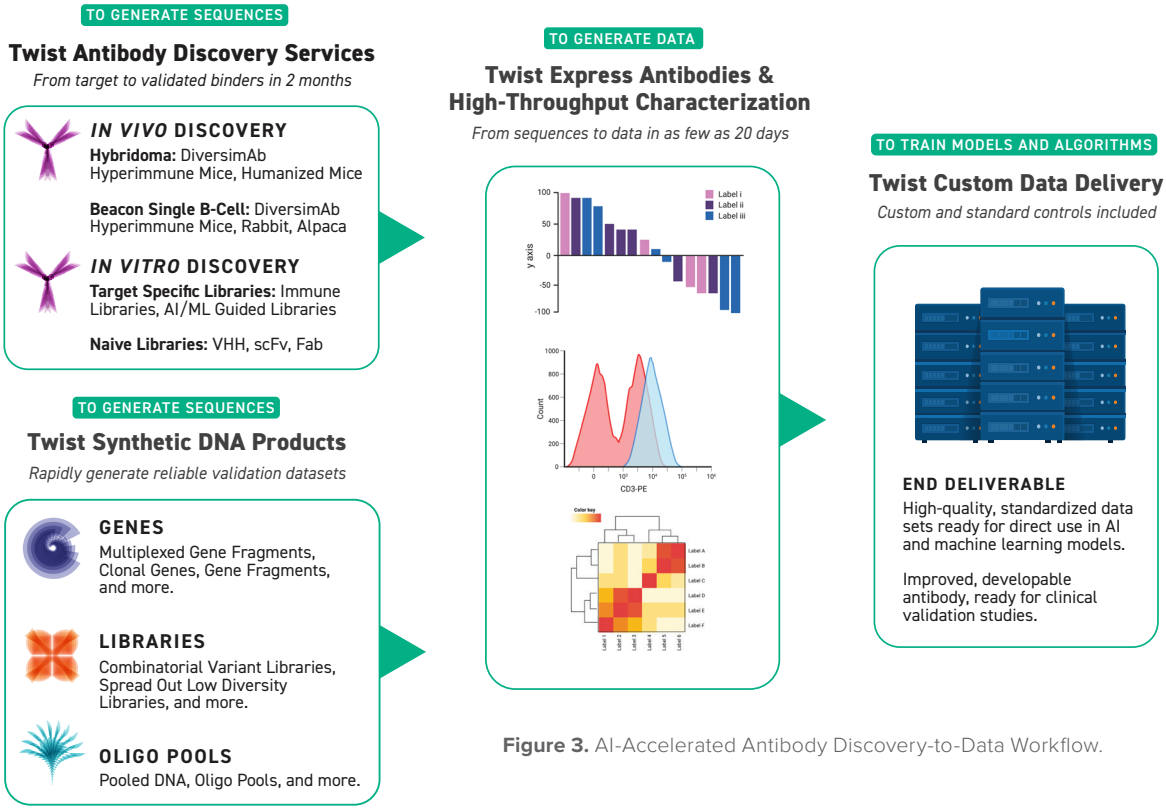


Figure 3. AI-Accelerated Antibody Discovery-to-Data Workflow.

Twist solutions for AI research

ANTIBODY DISCOVERY

Antibody Production: High-throughput expression of VHH, scFv, or IgG formats in HEK or CHO cells. We offer small to medium scale purification (up to 5mg) with a rapid 2-week building TAT in the US and EMEA**.

Binding Assessment: Generate standardized, model-ready datasets for rapid model retraining. This includes high-throughput SPR (Carterra LSA) kinetics, affinity, and epitope binning with a 1-week characterization TAT***.

Custom Characterization: Tailored experimental designs to assess biological activity and safety profiles, including internalization, immunogenicity, cytotoxicity, and target-specific cross-reactivity. Our in-house expertise enables accurate assay setup and data delivery for non-standard, complex services.

Developability Assays: Identify the most promising lead candidates early using high-resolution data from AC-SINS self-interaction, aSEC monomeric purity, and nanoDSF thermal stability assays.

Custom Data Delivery: We deliver high-quality, standardized datasets ready for direct use in AI and machine learning models. All data is provided in machine-readable, unified, and mapped formats to ensure seamless integration with your specific AI platform or LIMS.

Explore AI-driven breakthroughs

ANTIBODY DISCOVERY & PROTEIN ENGINEERING

Accelerating Recombinase Reprogramming with Machine Learning: Explore data-driven evolution of recombinases for high-performance gene editing with SeamlessTx.

Using AI and Twist's Multiplexed Gene Fragments in the Hunt for the "Holy Grail" in De Novo Antibody Design: See how the Baker Lab (University of Washington) tested 36,000 unique VHH designs with atomic-level accuracy using MGFs.

Better AI Begins at the Bench: Read how A-Alpha Bio predicts binding up to 20 mutations away via massive wet-lab datasets generated by AlphaSeq.

From Sequences to Systems: Designing a Next-Gen Wet Lab: Discover how Cradle and Twist Bioscience build "ML-ready" workflows to reduce experimental cycle times to three weeks.

Lab-in-the-Loop: Smarter AI for Antibody Design: Learn how Genentech reduces experimental burden and improves candidate quality through active learning.

Powering AI-Based Protein Engineering With Scalable Data: Explore how A-Alpha Bio utilizes synthetic biology to build data foundations for next-generation discovery.

Using AI to design proteins that inhibit bacterial heme-piracy: Learn how AI-designed proteins and Clonal Genes are used by The University of Melbourne to inhibit bacterial cofactors.

CRISPR & AGRICULTURAL BIOLOGY

A CRE.AI.TIVE application of AI: Engineering a more resilient global food supply: See how the CRE.AI.TIVE platform from Phytoform Labs identified 4-fold gene expression increases in drought-resistant crops.

Combining Massive Oligo Pools and AI to Predict Guide RNA Efficiency for Prime Editing Screens: Learn how massive 300mer oligo pools trained models at Twist Bioscience to predict pegRNA performance.

Resurrecting Plant Defense Mechanisms to Avoid Crop Pathogens: How Resurrect Bio uses Clonal Genes to move from in silico prediction to in planta validation.

FUNCTIONAL GENOMICS & MOLECULAR ENGINEERING

Building Better AI With High-Resolution Functional Genomics Data: Learn how Xaira approaches generating large-scale, single-cell perturbation datasets for biological foundation models.

Gene to Chemicals: An Adventure of Molecular Engineering Across Scales: Exploring how A*STAR IMCB synthetic biology activates silent biosynthetic gene clusters for metabolite production.

Genome Modeling and Design: From the Molecular to Genome Scale: Brian Hie at Stanford discusses training Evo2, a genomic foundation model, with 9.3 trillion nucleotides from over 128,000 genomes.

TERMS AND CONDITIONS

*Turnaround time for Express Genes starts at 4-7 business days and increases to 8-10 business days for 10 µg - 100 µg and 100 µg - 1 mg DNA prep scales.

** Turnaround time starts at 10-15 business days for 1mL antibody expressed in HEK293 and 13 - 18 business days for 1mL antibody expressed in CHO. 8 mL expression volumes take an additional business day.

***Characterization Turn-Around Time (TAT) varies based on the number of samples.

READY TO BUILD YOUR BRIDGE TO THE WET LAB?

Contact a Specialist at **sales@twistbioscience.com**. Or visit **twistbioscience.com/applications/ai-drug-discovery**