DNA and Protein Tools for AI-Driven Biologics Design

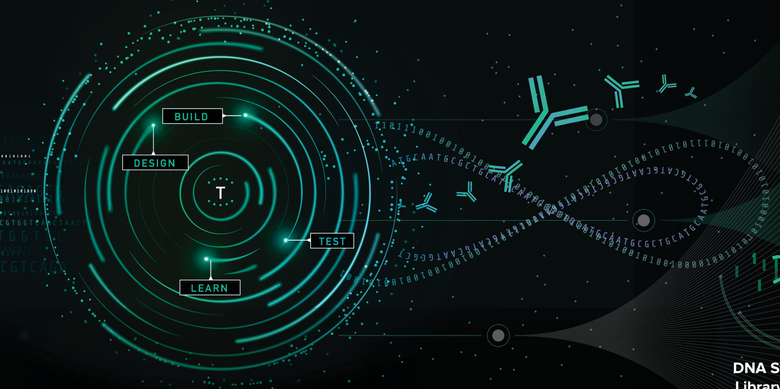
At Twist Bioscience, we work in service of customers who are changing the world for the better. In fields such as medicine, agriculture, industrial chemicals and defense, by using our synthetic DNA tools, our customers are developing ways to better lives and improve the sustainability of the planet. The faster our customers succeed, the better for all of us, and Twist Bioscience is uniquely positioned to help accelerate their efforts.

Our innovative silicon-based DNA Synthesis Platform provides precision at a scale that is otherwise unavailable to our customers. Our platform technologies overcome inefficiencies and enable cost-effective, rapid, precise, high-throughput synthesis, sequencing and therapeutics discovery, providing both the quality and quantity of the tools they need to most rapidly realize the opportunity ahead.

Great models demand great data
Wherever you are in model development, excellence begins with the right DNA and protein tools.

TWIST BIOSCIENCE

DNA and Protein Tools for AI-Driven Biologics Design

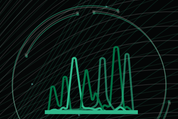


In 2020, the prediction of protein structure from amino acid sequence was largely solved by Google DeepMind's machine learning (ML) model, AlphaFold.[1] Now, scientists are putting artificial intelligence (AI) to the test on the hunt for the next holy grail: the *de novo* design of functional proteins.[2] Achieving this goal promises to transform research by accelerating target discovery, optimizing candidate design, and reducing development time and cost.[3] AI/ML models will do so by helping unpick and predict sequence-function relationships, but this approach will require high-quality, purpose-built datasets to train AI to design new candidates, predict sequence performance, and uncover novel therapeutic possibilities. Service providers that can provide solutions from DNA to data, such as Twist Bioscience, are needed to meet the AI community's experimental needs wherever they are in the process.

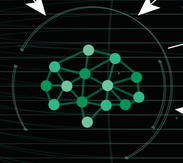
References

1. Senior AW, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;577(7792):706-710.
2. Kortemme T. *De novo* protein design—From new structures to programmable functions. *Cell*. 2024;187(3):526-544.
3. Mirakhori F, Niazi SK. Harnessing the AI/ML in drug and biological products discovery and development: The regulatory perspective. *Pharmaceuticals*. 2025;18(1):47.

Model Development



Primary Screening and Data Collection



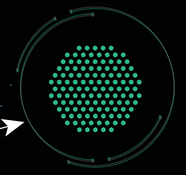
Model Development and Validation

DNA Synthesis and Library Generation

- **High sequence accuracy and customization** to establish reliable links between sequence and function
- **Sufficient library depth** to support effective model learning while matching assay constraints
- **Speed and production readiness** to enable rapid design–test–learn iteration
- **Focused experimental validation** that prioritizes high-value candidates, reducing time and resource use

Refinement

Selection



Initial Candidate Pool

Train Model on Characterization Data

Secondary Characterization Through Biological Assays

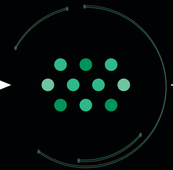
Additional Sequence Questions

- **Wet-lab data is required for ground truth,** translating AI-proposed sequences into measured performance
- **Traditional labs are not optimized** for the scale, diversity, and structure of datasets needed for AI/ML workflows
- **Specialized providers enable scale**, synthesizing large sequence panels and generating standardized assay data
- **Consistent experimental feedback** supports model refinement, candidate triage, and reliable hit selection

Outcomes

Clinical Trials

Candidate Selection



Lead Candidates

Production

Preclinical Data Collection

- **Deeper developability insight** through stability, aggregation, specificity, and expression data
- **Reduced wet-lab bottlenecks** by combining AI-driven candidate prioritization with rapid DNA, protein production, and characterization
- **Earlier elimination of weak candidates** using large, standardized, protein-level datasets
- **More efficient discovery progression** enabled by precise DNA construction and scalable, high-quality experimental data that supports model refinement