

# Profluent's \$2.25B Lilly Deal and Why Treating Proteins as a Language Modeling Problem Is a Bigger Story Than the Headline Suggests: Scaling Laws, Synthetic Biology, and the Compute Substrate Thesis

APR 30, 2026 • PAID



Abstract

## The deal at a glance:

- Profluent and Eli Lilly announced a multi-program collaboration on April 2 2026, valued at up to \$2.25B in upfronts, milestones, and royalties
- Focus area: AI-designed enzymes for genetic medicines, spanning gene editing, delivery enzymes, and modulators
- Profluent: founded 2022 by Ali Madani, formerly of Salesforce Research; ~\$100M raised across seed and Series B before this deal
- Backers: Altimeter, Bezos Expeditions, Insight Partners, Air Street, plus too many partners like Integrated DNA Technologies
- Prior validation includes the ProGen3 foundation model and OpenCRISPR, a fully AI-generated gene editor with wet-lab confirmed activity in human cells

## What it actually represents:

- Not another “AI biotech raises money” headline
- A platform-level bet that scaling laws extend from natural language to protein function

- The first large pharma endorsement of treating biology as a generative comp problem rather than a screening problem
- Comparable in shape, though not yet in outcome, to early Recursion or Insil mega-deals, except more “model is the product” and less “screen is the prod

## **Five things to track from here:**

1. Whether design abundance compresses development timelines or just relocates the bottleneck to delivery
2. How regulators handle non-natural proteins on immunogenicity and off-target biology
3. Whether pharma deal structures shift toward platform fees versus per-asset milestones
4. Whether closed-loop training (design, synthesize, test, retrain) produces a real flywheel or a marginal improvement
5. Whether biology consolidates into a few foundation-model winners, the way frontier AI did

## **Table of Contents**

The deal in one paragraph

What Profluent actually is, stripped of marketing

The founding insight that explains the speed

The technical stack pharma actually pays for

Three years to a mega-deal, broken down

Why Lilly wrote the check

Scaling laws crossing into biology

Breaking out of evolution's local maxima

Discovery gets cheap, delivery gets expensive

Big pharma is quietly re-platforming

The closed-loop flywheel and data moats that actually compound

Inference economics arrives in biotech

Where to be skeptical

The category, not the company

## **The deal in one paragraph**

Reuters broke it this morning. Profluent and Eli Lilly signed a multi-program collaboration worth up to \$2.25B in biobucks, focused on AI-designed enzymes and genetic medicines. The headline number is the kind of figure that always needs a translator, since most of it is milestones strung across discovery, preclinical, clinical and approval, plus royalties on whatever survives. The interesting parts are not the dollars. It is the structure (multi-program, platform-flavored), the focus (enzyme editing and delivery, not therapeutic antibodies), and the timeline. Profluent went from Salesforce Research spinout to top-five pharma deal in about three years or roughly \$150M raised. For a category that usually takes a decade to earn this kind of validation, that compression is the actual story. The dollar value is the bait. The reason a strategic like Lilly is willing to sign a platform-style agreement with a 200-person company is the dish.

## **What Profluent actually is, stripped of marketing**

Profluent gets called an “AI drug discovery company,” which is true the way calling Stripe a “payments company” is true. The label is correct and tells you almost nothing. The accurate version is that Profluent treats proteins as a language modeling problem and trains generative foundation models to design new ones. The framing is close

GPT for amino acid sequences than to a structure-based drug discovery shop run on better docking software. Where most of the AI-for-bio cohort uses machine learning as a discriminator (predict folding, predict binding, predict toxicity), Profluent uses it as a generator. The system writes proteins that have never existed in nature. It synthesizes and tests them. Then it folds the results back into training. That loop is the entire company. Everything else, the partnerships, the press releases, the data, the names, is scaffolding around that one idea.

The technical stake matters because it changes the unit of value creation. A discriminative model predicts which existing molecules are most likely to work. A generative model designs new ones. The first is a better filter. The second is a different search space. Pharma has been getting better filters for thirty years. It has not really had a different search space since recombinant DNA, and arguably maybe not before. If Profluent is right, it has one now.





Continue reading this post for free, courtesy of Thoughts of  
Healthcare.

Claim my free post

Or purchase a paid subscription.

← Previous

© 2026 Healthcare Markets & Technology · [Privacy](#) · [Terms](#) · [Collection notice](#)  
[Substack](#) is the home for great culture