

Deep Origin explores combining AI and physics to improve virtual screening in drug discovery

August 25, 2026 | Opinion | By ernest.teck@mmactiv.com

Co-founder and Chief Scientific Officer Garegin Papoian discusses why conventional virtual screening struggles with novel targets and how physics-based AI could improve hit finding and experimental efficiency.



Artificial intelligence is increasingly being applied across early-stage drug discovery, but translating computational predictions into experimentally validated drug candidates remains one of the field's central challenges.

Virtual screening has existed for decades, yet its use as a primary hit-finding strategy remains limited. Deep Origin is developing an approach that combines generative AI with physics-based molecular modelling, with the aim of improving predictive reliability when screening novel biological targets that differ substantially from the datasets used to train conventional AI models.

Garegin (Garyk) Papoian, Co-founder and Chief Scientific Officer of Deep Origin, explores the technical limitations that are holding back the future of virtual screening, and the role of prospective wet-lab validation. We discuss how AI and physics could reshape hit discovery in the foreseeable future.

Virtual screening has long been viewed as a promising approach for accelerating drug discovery, yet industry adoption has remained limited. What have been the biggest technical barriers preventing it from becoming a mainstream hit-finding tool?

Although it's been around for 40 years, virtual screening presumably accounts for only about 1% of primary hit-finding campaigns. The challenge is reliability: virtual screening predictions have not reliably borne out in wet lab experiments. Classical docking tools relied on simplified force fields and rigid protein grids, which struggle with high false-positive rates. The main challenge is that modern co-folding models do not perform well on novel biology. They do well on targets that are similar to the data training set, but their accuracy degrades significantly the further away you get from the training data. This "memorization trap" is the top technical barrier we are overcoming by building a virtual screening tool that combines AI

with physics.

Your recent research combines generative AI with physics-based modelling to improve prediction accuracy. How does this approach address the limitations of conventional AI models, particularly when working with novel biological targets?

The AI-only models operate in large part by memorizing the training material. They might have accuracy around 80% on targets similar to the training data, but accuracy drops to less than 25% on novel targets. Without physical rules, AI-based methods have been known to produce physically impossible solutions, such as atoms colliding in the same space, molecular joints bent at unnatural angles or chemical rings twisted into unstable shapes. And, while co-folding models are good at identifying binding pockets (see SI Fig. 3 in our paper), they fail at predicting precise 3D ligand pose placement and ranking under zero-leakage conditions. This is why we need to add physics to improve accuracy – it helps find physically plausible structures.

Our model DODock uses a generative model to propose a set of possible 3D poses. Then our physics engine (DOFast) refines those candidates by calculating molecular forces, including hydrophobic and hydrogen bonding interactions. Once the candidates have been physically refined, a ranking neural network model (DOScore) ranks the best ones. That physics-based refinement step is what allows DODock to be more successful on novel targets that are unlike the training dataset. On OpenBind, a benchmark designed around novel biological targets, DODock achieves 80% pose accuracy, compared with 4% to 28% for leading co-folding models. On Runs N' Poses, it maintains more than 50% accuracy on the most dissimilar targets, where leading co-folding models fall below 25%.

The study reports significant improvements in hit rates and binding pose prediction across prospective wet-lab validation. From a drug discovery perspective, what impact could these gains have on reducing development timelines, costs, and experimental workload?

In silico early-stage drug discovery is expected to significantly cut down development timelines, costs as well as produce superior outcomes. While there are some data and estimates circulating in the field, exactly how much so is unclear since there is no clear evidence that virtual screening has led to any approved drug. What I can say is that, at this moment, based on our prospective wet-lab validations, virtual screening can already achieve another important goal in drug development, which is drugging challenging targets, possibly even targets that scientists have considered to be nearly undruggable. For example, in our prospective screen of CD37, we achieved a 30% hit rate when a prior screen achieved only 0.3%. Another strong example is our prospective test of IL-17A. This is a protein-protein (PPI) target that is considered extremely challenging. We discovered six active compounds among 194 tested. Virtual screening technology that we have today can open new frontiers in drug discovery. Highly predictive computational filtering would allow industry teams to focus wet-lab resources on high-confidence, novel chemical scaffolds. This has the potential to reduce hit-finding timelines from months to weeks and open promising therapeutic avenues.

As AI becomes more deeply integrated into drug discovery, how important is experimental validation in building confidence among pharmaceutical companies, and what do you see as the key factors driving broader industry adoption?

Prospective screens are the most important measurement. Benchmarks have limitations and may even be misleading due to the problem of “data leakage,” which leaves near-identical proteins and near-twin chemical scaffolds sitting in the training set. This can make standard AI models appear more successful than they are, because they will work well on targets similar to the training set but not well on more novel targets. It's the novel targets that often matter most in drug discovery. Virtual screening models must have proof that they are consistently predictive, and the best way to demonstrate this is through wet-lab experiments. This is what industry teams evaluating tool options need to ask for.

Looking ahead, how do you expect AI-powered virtual screening to evolve over the next five years, and what role will it play alongside traditional medicinal chemistry and laboratory-based drug discovery?

It will continue to be a transformational decade for drug discovery work. This is a fast-moving field, and every few months we make notable progress. In fact, our internal models are already performing 20% to 30% better on many dimensions than what was shown in our latest paper. This rapid advancement is why preprint publishing is so important – it's the only

way to keep up, and even it's too slow. In the next year or two more drug discovery teams will be using more reliable models, and we will also see more results from AI-developed molecules that are currently in or soon to enter late-stage clinical trials. These data – from both preclinical and clinical research – will help define the role of virtual screening alongside traditional tools. The adoption of virtual screening tools will continue to grow as their reliability is proven. There's also great potential not just to identify new drug targets but to use computational tools in other ways, such as to virtually assess ADMET and toxicity.

Ultimately, virtual screening will transition to a primary, highly reliable hit-finding engine for early-stage drug discovery.