DESTRUCTURED DRUG DISCOVERY: HOW SEQUENCE-BASED AI SPEEDS AND EXPANDS THE SEARCH FOR NEW THERAPEUTICS

Scientists must venture through a vast chemical space to identify potential drugs: it's estimated that 10^{60} different small molecules obey the standard rules of bioavailability, a tally greater than the number of seconds that have elapsed since the Big Bang. Biologics, an important class of modern drugs that includes antibody therapeutics, expand this space further.[1,2] Identifying a promising drug candidate from this immense pool of possibilities requires years of research and millions of dollars of investment. Even then, the likelihood that a drug moves from Phase I trials to market approval ranges from only 8–12%.[3,4]

"Drug discovery is always a multiparameter optimization problem," says Lurong Pan, founder and CEO of Ainnocence. "If even one parameter is missing, the whole drug candidate will fail." To shorten drug development timelines, reduce costs, and get more treatments to patients, scientists require tools that help them balance this optimization problem and make sense of massive chemical spaces. Many within the field of drug discovery are betting on artificial intelligence models to be that tool.

That bet is beginning to pay off. More companies are relying on AI's predictive powers to inform drug discovery, and promising examples are emerging. For example, the small molecule rentosertib may soon become the first drug designed by generative AI methods to enter Phase 3 trials. It likely will not be the last.[5,6]

AI gives drug developers a new way to search for theoretical therapeutics. However, AI advances to date have largely focused on predicting the 3D structure of human protein targets and how drug molecules interact with them.[7–9] According to Pan, these structural simulations rely on expensive and imperfect data to train computationally expensive algorithms.

Pan believes the solution to this problem is to bypass structural modeling altogether and instead use proteins' amino acid sequences to guide AI models.

Sequence-based AI models may be able to discover new drug molecules faster and at lower computational cost than other approaches that depend on structural simulations.

Credit: Shutterstock

Unhindered by the cost and technical limitations of structure-based solutions, sequence-based models could speed up drug discovery.

A NEW GENERATION OF DRUG DISCOVERY TECHNOLOGIES

In 2018, the AlphaFold 1 AI program won a global competition to predict the 3D structures of proteins.[10] By 2020, some scientists considered the longstanding problem of predicting protein folding as solved, which fueled hopes of faster and cheaper drug discovery.[11]

The reality of protein chemistry is more complicated, however. Molecular simulations and AI models for drug discovery that rely on protein modeling often use "snapshots" of protein structure. These structural snapshots do not fully describe disordered regions, protein dynamics, or higher-order networks.[12] This is partially a technical problem: classical simulations struggle to incorporate quantum interactions and the behavior of water molecules, so confidence in the predicted structures and in identified drug candidates is reduced.[13,14]

Further, while predictive algorithms improve as they absorb more training data, the data used by structural AI models are more difficult to collect than data from experiments that measure protein-drug binding or other biological effects. "A binding assay is way cheaper than solving a structure in cryo-electron microscopy (cryo-EM) or crystallography and way more abundant," Pan says. (Collecting data with a binding assay can cost less than $1,000 per sample, compared with using cryo-EM instruments, which cost in the millions of dollars and are expensive to operate.)

Before AlphaFold's public launch, Pan began developing AI models that predicted whether a given biomolecule, like an antibody, would bind to a targeted protein according to the amino acid sequences of the biomolecule and target. By incorporating wet lab data into the model, Pan's group over time improved the predictive ability of the sequence-based model, which is now the core of Ainnocence's technology offerings.

Ainnocence uses these models to accelerate the discovery of new drugs. The models predict which molecules in a large database will bind most effectively to a protein of a given sequence. The approach bypasses 3D simulations entirely and can therefore screen billions of drug candidates in hours to days rather than in weeks or months, thus offering a workflow that can screen the entire proteome faster than alternative high-throughput in vitro approaches, atomic simulations, and 3D-structure-based AI methods.

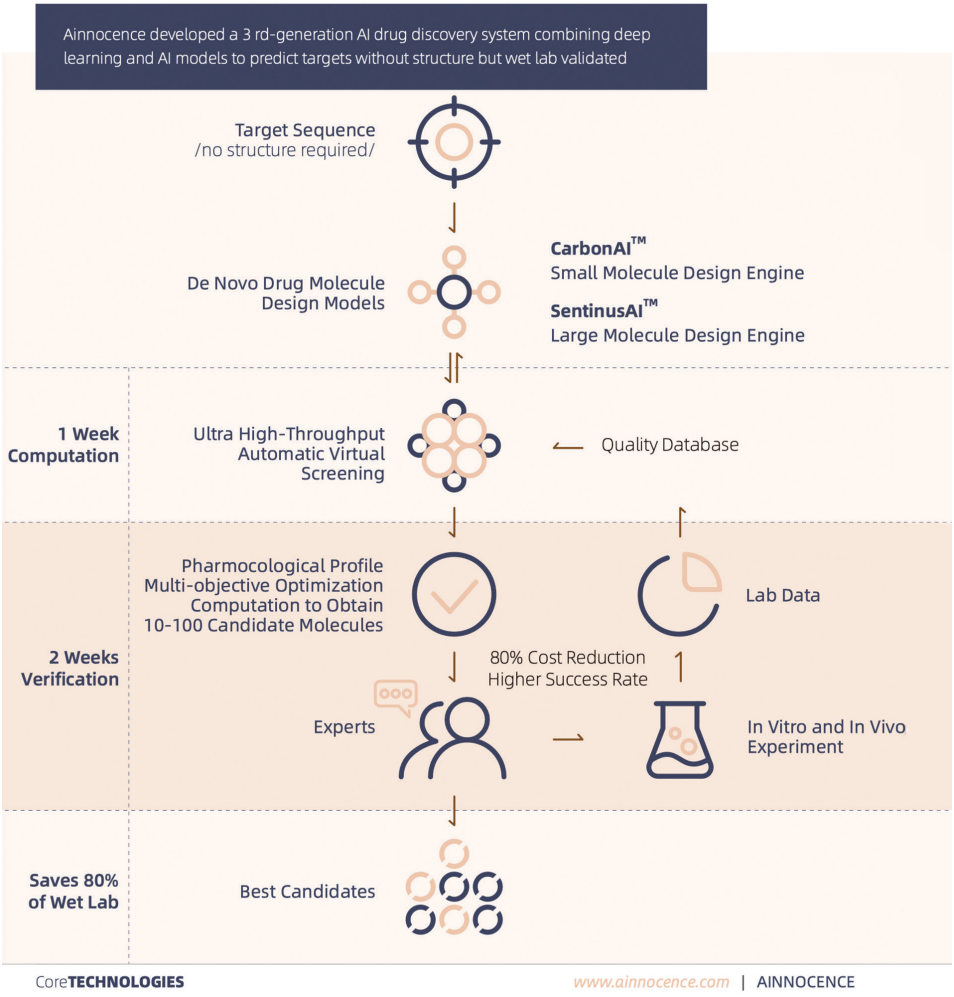

Figure 1. Ainnocence says its fast workflow reduces wet lab time and costs by 80% thanks to virtual screening based on protein target sequence.

Credit: Ainnocence

The workflow (figure 1) begins with the sequence of a specific protein target. The AI model then screens up to billions of drug candidates—from antibodies to small molecules—to find dozens of potential leads based on those candidates' predicted ability to bind the target protein. Ainnocence scientists test each lead in wet lab experiments and feed that new data back into the model. After one or more cycles of screening and validation, Ainnocence researchers can recommend highly functional drug candidates. Each cycle retrains the model: according to Ainnocnece, the final candidates' binding affinities can improve between 10-fold and 1,000-fold relative to the starting point. Within three rounds, Ainnocence validates 10-80% of candidates in wet lab experiments.

Overall, the Ainnocence team estimates that their process reduces early drug discovery expenses by 80%.

In August 2025, Ainnocence internal benchmarks reached higher Spearman correlation scores than those of previously published sequence-based models.[15] Spearman correlations measure how well AI predictions align with experimental results, with $\rho = 0$ being random and $\rho = 1$ being perfect. Ainnocence's overall score of $\rho = 0.441$ was similar to models that incorporate 3D structure information despite requiring significantly fewer computational resources.

A PANDEMIC PROOF OF CONCEPT

One of Ainnocence's first applications of its sequence-based AI model was the identification of antibodies capable of targeting and neutralizing multiple viral SARS-CoV-2 strains.[16]

In the early months of the COVID-19 pandemic, some scientists hypothesized that SARS-CoV-2 antibodies isolated from people previously infected with the virus could help other patients fight active infections. But viral pathogens mutate frequently. With each mutation come new chemistries and new opportunities to subvert immune defenses. It is difficult to discover and design antibodies that are therapeutically effective against diverse strains. "But AI is a different story," Pan says.

Ainnocence researchers first fine-tuned their sequence-based model with molecules known to bind the spike protein in various strains of SARS-CoV-2: commercial antibodies and an antibody isolated from a SARS-CoV-1 patient. The AI then sifted through databases of antibody sequences, evaluating whether those antibodies could bind and or even neutralize the virus.

This workflow yielded 50 AI-designed antibodies that could bind various SARS-CoV-2 strains without requiring any information about the structure of the virus's spike protein.[17] Of those 50, laboratory experiments showed that 10 neutralized the Delta variant of the virus while 1 neutralized the Omicron strain, which had not yet emerged when the antibodies were designed.

As new strains of SARS-CoV-2 emerged, existing commercial antibodies became outdated and were unable to bind or neutralize the virus.[18] But Ainnocence's model was able to predict antibodies effective at binding and neutralizing new variants. Pan suggests that AI models can perceive patterns that relate one strain's targets to another. "AI could learn evolutionary patterns," she says. That bodes well for applying such AI-based tools in future outbreaks.

During their study to identify SARS-CoV-2 antibodies, the Ainnocence team also compared the computational cost of its sequence-based methods with two structure-based methods. Of the latter, one predicts binding affinities from 3D structural data and one calculates free energy values using molecular dynamics simulations. The Ainnocence team began with an antibody containing a 20-amino acid region of interest and generated virtual libraries with one to four possible point mutations in that region. The libraries grow exponentially, from approximately 10^2 variants (one mutation site) to 10^8 (four mutation sites). They then tasked the three models with screening these libraries. Ainnocence's platform searched these libraries significantly faster than both structure-based models. For example, it searched the one-site mutation library in under 2 min—approximately 5 times as fast as the 3D-structure-based method and 2,800 times as fast as the model using molecular dynamics.

BEYOND ANTIBODIES

Since the SARS-CoV-2 experiment, Ainnocence has applied its sequence-based AI modeling approach to over 40 therapeutic protein design projects. Antibody design projects use Ainnocence's **SentinusAI** platform.

The company has analogous platforms for predicting the effects of small molecules, cell therapies, and messenger RNA vaccines.

CarbonAI, a small-molecule prediction model, considers the amino acid sequence of the protein that the small molecule is meant to target. The model searches as many as billions of small molecules for hits capable of detecting binding pockets based solely on target sequence.

For cell-based cancer therapies, **CellulaAI** helps design binding molecules that can target multiple protein fragments at the surface of cancer cells, including neoantigens specific to a particular patient's tumor.

Each platform is built using curated databases and allows Ainnocence to learn from initial hits to further optimize candidates. These platforms are also enhanced by incorporating a diverse yet tailored collection of properties. For example, the right antibody must generate an adequate immune response and avoid off-target binding. Small molecules must satisfy standards for absorption, distribution, metabolism, and excretion to be considered safe and effective. All proposed drug candidates must be synthesizable and have functional groups that are physically stable during manufacturing. According to Ainnocence, its AI models incorporate over 30 different factors, none of which

rely on 3D structure. Each factor acts as a scoring parameter to narrow down candidates.

Ainnocence's applies this approach to models built from alternative datasets. For example, the platform **NatmolAI** screens hundreds of thousands of natural molecules for potential use as "nutraceuticals" or repurposable drugs. **BioSynthAI** extends Ainnocence's tool to synthetic biology, where microbial strains and enzymes can be optimized for use in pharma and industrial manufacturing.

WHAT'S NEXT?

Biochemists have long suspected that protein sequences hold enough information to predict which molecules can drug them. Biology's central dogma posits that the sequences of DNA, RNA, and proteins encode all biochemical processes. There is no fundamental reason that an AI model could not learn to predict function from sequence. Until recently, there simply wasn't enough data to make good predictions about binding affinity and other properties. But "the number of data has increased dramatically," Pan says. "There's a turning point."

She likens this transition to the evolution of other AI models like ChatGPT and AlphaFold, which found success years after their underlying algorithms were invented. ChatGPT's language model was developed in 2015 but reached its own turning point in 2022. "Only at that time were the data enough to reach a threshold of accuracy," Pan says. Ainnocence continues to improve its protein language model with open databases and peer-reviewed analyses, as well as by generating more wet lab data.

Given the success of Ainnocence's sequence-only approach, Pan foresees opportunities to complement innovations that support other stages of drug discovery and development. In early 2025, Ainnocence announced a partnership with Obatala, a company that develops humanized microfluidic chips.[19] The two firms hope to reduce both the need for costly animal experiments and the cost of drug discovery and development overall.

Sequence-based models are more scalable than structure-based models for sifting through large databases for drug candidates. They excel at optimizing across multiple parameters or multiple targets. Still, Pan acknowledges that structure-based AI has a role in drug discovery. Structural simulations help decode biochemical function in a more qualitative and visual way. When scientists know a target's structure, they can derive insights about its fundamental biochemistry. "Both sequence and structure have great value, and they could really answer questions together," Pan says.

"If you look at the string of a small molecule or a sequence of a protein, you can't imagine chemically how these letters or symbols relate to chemical and biological properties. It's just not possible for human scientists," Pan says. "But AI could imagine it."

REFERENCES:

1. Jean-Louis Reymond, "The Chemical Space Project," *Acc. Chem. Res.* 48, no. 3 (March 17, 2015): 722–30, https://doi.org/10.1021/ar500432k.
2. Regine S. Bohacek, Colin McMartin, and Wayne C. Guida, "The Art and Practice of Structure-based Drug Design: A Molecular Modeling Perspective," *Med. Res. Rev.* 16, no. 1 (Jan. 1996): 3–50, https://doi.org/10.1002/(SICI)1098-1128(199601)16:1%3C3::AID-MED1%3E3.0.CO;2-6.
3. Biotechnology Innovation Organization, QLS Advisors, Pharma Intelligence, *Clinical Development Success Rates and Contributing Factors 2011–2020,* Feb. 2021, https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf.
4. Congressional Budget Office, *Research and Development in the Pharmaceutical Industry,* April 2021, https://www.cbo.gov/publication/57126.
5. Zuojun Xu et al., "A Generative AI-Discovered TNIK Inhibitor for Idiopathic Pulmonary Fibrosis: A Randomized Phase 2a Trial," *Nat. Med.* (Aug. 2025): 1–9, https://doi.org/10.1038/s41591-025-03743-2.
6. Anthony King, "As AI-Designed Drug Looks to Pass Final Hurdle Will this Tech Change Drug Discovery Forever?" *Chemistry World,* July 31, 2025, https://www.chemistryworld.com/news/as-ai-designed-drug-looks-to-pass-final-hurdle-will-this-tech-change-drug-discovery-forever/4021894.article.
7. Ewen Callaway, "Major AlphaFold Upgrade Offers Boost for Drug Discovery," *Nature* 629, no. 8012 (May 8, 2024): 509–10, https://doi.org/10.1038/d41586-024-01383-z.
8. Sarfaraz K. Niazi and Mariam Zamara, "Computer-Aided Drug design and Drug Discovery: A Prospective Analysis," *Pharmaceuticals* 17, no. 1 (2024): 22, https://doi.org/10.3390/ph17010022.
9. Josh Abramson et al., "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature* 630, no. 8016 (June 13, 2024): 493–500, https://doi.org/10.1038/s41586-024-07487-w.
10. Protein Structure Prediction Center, 13th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction, 2018, https://predictioncenter.org/casp13/index.cgi.
11. Yasemin Saplakoglu, "How AI Revolutionized Protein Science, but Didn't End It," Quanta Magazine, Sept. 26, 2024, https://www.quantamagazine.org/how-ai-revolutionized-protein-science-but-didnt-end-it-20240626/.
12. Apolline Bruley et al., "Digging into the 3D Structure Predictions of AlphaFold2 with Low Confidence: Disorder and Beyond," *Biomolecules* 12, no. 10 (Oct. 2022): 1467, https://doi.org/10.3390/biom12101467.
13. Weitang Li et al., "A Hybrid Quantum Computing Pipeline for Real World Drug Discovery," *Sci. Rep.* 14, no. 1 (2024): 16942, https://doi.org/10.1038/s41598-024-67897-8.
14. Balázs Zoltán Zsidó and Csaba Hetényi, "Water in Drug Design: Pitfalls and Good Practices," *Expert Opin. Drug Discovery* 20, no. 6 (2025): 745–64, https://doi.org/10.1080/17460441.2025.2497912.
15. "Ainnocence Achieves New Milestone with Protein Foundation Models,"

news release, Aug. 23, 2025, https://ainnocence.com/ainnocence-achieves-new-milestone-with-protein-foundation-models/.

16. "Ainnocence Demonstrates Breakthrough AI Technology for Rapid Pandemic Response," news release, Jan. 29, 2025, https://ainnocence.com/ainnocence-demonstrates-breakthrough-ai-technology-for-rapid-pandemic-response/.
17. Yue Kang, Kevin Jin, and Lurong Pan, "AI Designed, Mutation Resistant Broad Neutralizing Antibodies Against Multiple SARS-CoV-2 Strains," *Sci. Rep.* 15, no. 1 (2025): 15533, https://doi.org/10.1038/s41598-025-98979-w.
18. "Casirivimab and Imdevimab—High Risk of Treatment Failure Due to Circulation of SARS-CoV-2 Omicron Variant," Government of Canada, Recalls and Safety alerts, Jan. 7, 2022, https://recalls-rappels.canada.ca/en/alert-recall/casirivimab-and-imdevimab-high-risk-treatment-failure-due-circulation-sars-cov-2.
19. "Obatala Sciences and Ainnocence Announce Partnership to Transform Drug Discovery Through AI and Human Biological Models," news release, April 14, 2025, https://ainnocence.com/obatala-sciences-and-ainnocence-announce-partnership-to-transform-drug-discovery-through-ai-and-human-biological-models/.

AMERICAN CHEMICAL SOCIETY
1155 Sixteenth Street, NW,
Washington, DC 20036