Efficient Drug Discovery Powered by High Accuracy A.I. Models

EXECUTIVE SUMMARY

The drug development process is notoriously lengthy and expensive. On average, a new drug requires more than 12 years and $2.8B to advance from discovery to the market. Artificial intelligence (A.I.) models for computer-aided drug discovery (CADD) are poised to transform the process through A.I.-informed methods for virtual screening, molecular design, and prediction of safety and efficacy.

Using proprietary experimental data from its high-throughput screening and chemistry cores, Southern Research is applying the latest A.I. technology in combination with next-generation molecular modeling and other chem and bioinformatics approaches to help our partner companies efficiently identify drug candidates with the greatest potential for success.

This white paper will discuss how A.I. models are shaping drug discovery, describe proprietary A.I. models developed at Southern Research, and present case studies of demonstrated success in drug discovery using the Southern Research A.I. drug discovery platform.

INTRODUCTION

The drug development process is a long, complex, and expensive operation—taking nearly 12 years and costing an average of $2.8 billion per drug.[1,2] Moreover, only one in ten drug candidates progress beyond Phase II clinical trials and regulatory approval.[2] In an effort to control costs associated with expensive laboratory analysis and bring therapies to the market sooner, drug developers are turning to A.I. to pinpoint which molecules are most likely to succeed. The payoffs of A.I. are evident: 100,000 compounds screened using wet lab methods to narrow the field to just 100 compounds for further analysis once took three months to complete. With A.I. virtual screening, the same process can be completed in two weeks and at a far lower cost.[3]

Biotech and pharma companies are rapidly investing in new A.I. tools to speed the pace of drug discovery. Recent changes in the FDA Modernization Act 2.0 of 2022[4] further opened the door to use of *in silico* models to fulfill regulatory requirements for approval of medical products. In fact, A.I.-generated and -supported drugs are making rapid progress; the first drugs discovered and designed through A.I. models have recently been granted approval.[5]

A.I. models are accelerating drug discovery at multiple points in the development process, but not all A.I. models operate with the same level of accuracy. To support rapid drug development, Southern Research has created best- or first-in-class A.I. prediction models. Trained with tightly controlled, in-house-generated experimental data, these highly accurate A.I. models have helped Southern Research and its partners identify novel preclinical drug candidates in a time- and cost-efficient manner.

AN OVERVIEW OF THE SOUTHERN RESEARCH INTEGRATED A.I. DRUG DISCOVERY PLATFORM

The integrated A.I. drug discovery platform at Southern Research incorporates decades of in-house, proprietary source data from wet lab experiments as well as high-throughput screening and medicinal chemistry cores. Using these data points, Southern Research has developed A.I. models for drug discovery that are either first-in-class or are more accurate than the best available models in literature.

Our unique blend of data from *in vitro* and *in vivo* assays against off-the-shelf and novel molecules has enabled Southern Research to establish models for a wide-ranging set of A.I. drug discovery use cases. To date, we have developed A.I. models to predict therapeutic efficacy, pharmacokinetic properties, and toxicology/drug safety. We leverage not only predictive A.I., but also generative A.I. and large language models to address challenges in drug discovery. Examples include accelerating receptor-based virtual screening for efficient exploration of billion-scale chemical space, enhancing molecular docking for phenotypically relevant predictions, and aiding medicinal chemistry via the multiparametric optimization of drug candidates.

MACHINE LEARNING & A.I. MODELS

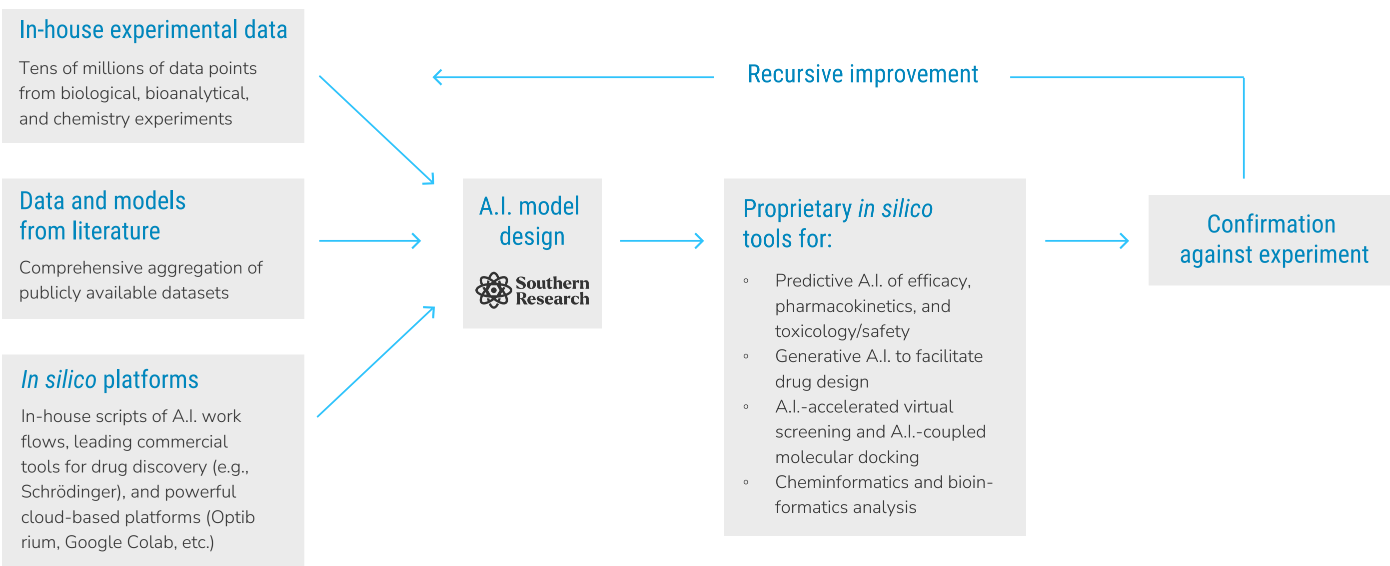
Southern Research has developed numerous A.I. models to streamline the drug development process and is always ready to develop or adapt new A.I. models to meet customized needs for upcoming drug discovery projects.

A.I. models exist to support:

- Efficacy predictions for various therapeutic areas covering infectious diseases, oncology, neurology, and other diseases to aid discovery of novel hits or drug repurposing for precision medicine
- Prediction of pharmacokinetic properties such as oral availability, metabolic stability, and brain permeability to aid the design of hit analogs suitable for *in vivo* studies
- Toxicology and safety predictions such as cardiotoxicity to de-risk further development of *in vivo* active leads into preclinical candidates
- Other tasks such as auto-enumeration of molecular structures, molecular modeling, prediction of protein structures (e.g., in conjunction with AlphaFold), and interpretation and use of medical data (e.g., BioGPT)

A.I. models and *in silico* tools can be used for a variety of standard and custom applications. For example, hit discovery studies can enable pre-selection of a compound library for assays performed in the high- throughput screening center and for low-throughput assays performed by other biology teams. Clients interested in lead optimization can use A.I. models and *in silico* tools to aid in the design of analogs of lead compounds that will be synthesized by our chemistry team. The efficacy of new analogs will be evaluated by our biology teams, and the pharmacokinetics of those analogs will be assessed by our bioanalytics team.

Figure 1. The Southern Research integrated A.I. drug discovery platform.



CREATING A.I. MODELS WITH HIGHER ACCURACY

Most commercial and academic A.I. models are trained using datasets obtained from multiple sources. A.I. models trained using data created by different methods, technologies, and conditions can yield results with lower accuracy. In addition to widely available data sources, Southern Research's A.I. models have been trained with high-precision, proprietary data generated from our prior and ongoing experiments. By reducing data with site-specific or method-specific variation, Southern Research models demonstrate higher theoretic accuracy based on the same scoring function used for A.I. models reported in literature **(Table 2)**.

Table 2. In-house developed A.I. model for small molecule drug discovery compared to models found in the literature.

In-house developed A.I. model for small molecule drug discovery	Theoretic accuracy	Comparison to model in literature
Anti-SARS-CoV-2 activity	87%	70% accuracy[6]
Anti-Dengue activity	75%	First-in-class
Anti-Chikungunya virus activity	82%	First-in-class
Anti-encephalitis virus activity	84%	First-in-class
Anti-Influenza H1N1/H3N2 activity	75%	First-in-class
Efficiency for cystic fibrosis	74%	First-in-class
HO-1 enhancer potency	74%	First-in-class
IL-6 inhibition	87%	First-in-class
Solubility in intestinal fluid	90%	78% accuracy[7]
Liver microsomal half life	89%	87% accuracy[8]

Besides theoretic score, we care more about the actual performance of those A.I. models when applied to drug discovery projects in real cases. Two such examples are in-house developed A.I. models to predict intestinal fluid solubility and metabolic half-life, which were accurate across ~80% of compounds versus a ~60% accuracy rate with widely used non-A.I. driven methods **(Table 3)**. Applying in-house A.I. models have helped Southern Research medicinal chemists optimize drug candidates for desirable properties such as metabolic stability and oral bioavailability.

Table 3. Performance of Southern Research A.I. models for intestinal fluid solubility and metabolic half-life predictions in actual cases.

Case	Compound ID	Intestinal fluid solubility (A.I.)	Solubility (non-A.I. driven software)	Mouse liver microsomal half-life (A.I.)
Hit discovery for cancer immunotherapy	SRI-48269	Does not match	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48264	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48271	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48263	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48559	Prediction qualitatively matches assay data	Does not match	Prediction qualitatively matches assay data
Lead optimization of antiviral agents against VEEV	SRI-45916	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48295	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48288	Does not match	Does not match	Inconclusive
	SRI-48551	Prediction qualitatively matches assay data	Does not match	Prediction qualitatively matches assay data
	SRI-48582	Inconclusive	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48600	Inconclusive	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
Discovery of broad spectrum antiviral agents	SRI-48457	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Inconclusive
	SRI-46033	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-47534	Prediction qualitatively matches assay data	Does not match	Inconclusive
	SRI-48420	Prediction qualitatively matches assay data	Does not match	Does not match
Lead optimization of PROTACs to treat rare diseases	SRI-47578	Prediction qualitatively matches assay data	Does not match	Prediction qualitatively matches assay data
	SRI-47154	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Inconclusive
	SRI-47154	Prediction qualitatively matches assay data	Does not match	Prediction qualitatively matches assay data
	SRI-48069	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data	Prediction qualitatively matches assay data
	SRI-48140	Does not match	Does not match	Inconclusive
Accuracy in applied cases		~80%	~60%	**~80%**

Prediction qualitatively matches assay data | Inconclusive | Does not match

CASE STUDIES: OUR PROPRIETARY A.I. MODELS IN ACTION

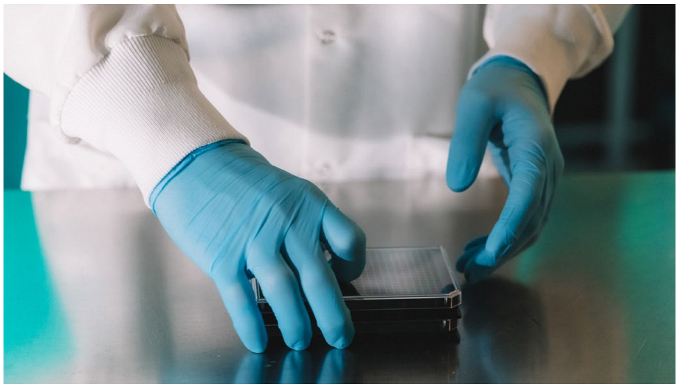
Southern Research has successfully used our A.I. models and *in silico* tools to facilitate different stages of the drug discovery pipeline from therapeutic target discovery to preclinical candidates. Here, we highlight a few examples of successful drug development projects supported by our preclinical integrated A.I. drug discovery platform.

CASE STUDY #1: A.I.-AIDED DRUG DISCOVERY FOR RENAL CELL CARCINOMA

Clear cell renal cell carcinoma (ccRCC) is a highly vascularized type of kidney tumor.[9] Patients carrying certain mutations of proteins in the oxygen sensing pathway may exhibit resistance to the standard-of-care therapeutic Belzutifan.[10]

To build a predictive A.I. model, 1,353 phenotypically active and 1,399 inactive compounds against RCC were sourced from literature **(Figure 2)**. Leveraging this model, we conducted virtual screening for compounds that would bind to the mutant protein of interest and additionally screened for phenotypic killing effect on RCC cells. We identified fourteen compounds to advance for testing via in-house experimental oncology assays.

We ultimately identified and advanced two novel compounds that showed early signs of efficacy against Belzutifan-resistant ccRCC. Our use of A.I. early in the discovery process allowed for a rate of experimental validation exceeding prior studies that did not leverage a predictive A.I. model, decreasing associated experiment costs and accelerating our development timeline.



CASE STUDY #2: A.I.-AIDED SCREEN TO DISCOVER ANTI-FLAVIVIRAL AGENTS

Leveraging in-house antiviral screening data, an A.I. model was developed and successfully used to discover anti-flaviviral compounds **(Figure 3)**. These compounds are being evaluated for use against mosquito-borne flaviviruses, Dengue (DENV) and ZIKA.

Development of the model began by conducting an in-house high-throughput screening campaign of 200,000 compounds for anti-Dengue virus (DENV) activity in a cell-based assay. Data from 620 active compounds and 7,300 inactive compounds were used to train an A.I. model with a deep neural network algorithm to predict anti-DENV activity.

Two chemical scaffolds of active compounds from the original 200,000 were selected for further development. The A.I. model suggested 48 additional compounds to screen, helping to identify a third chemical scaffold of highly potent anti-DENV activity. A conventional approach to screening 100,000 compounds would, on average, typically yield 1 highly potent scaffold. The A.I. aided approach yields an average of 1 highly potent scaffold per 50 compounds. Once enough data is generated for training A.I. models, discovering additional scaffolds will be accelerated approximately 2,000-fold.

Figure 2. A.I.-aided identification of compounds with potential for activity against Belzutifan-resistant RCC.

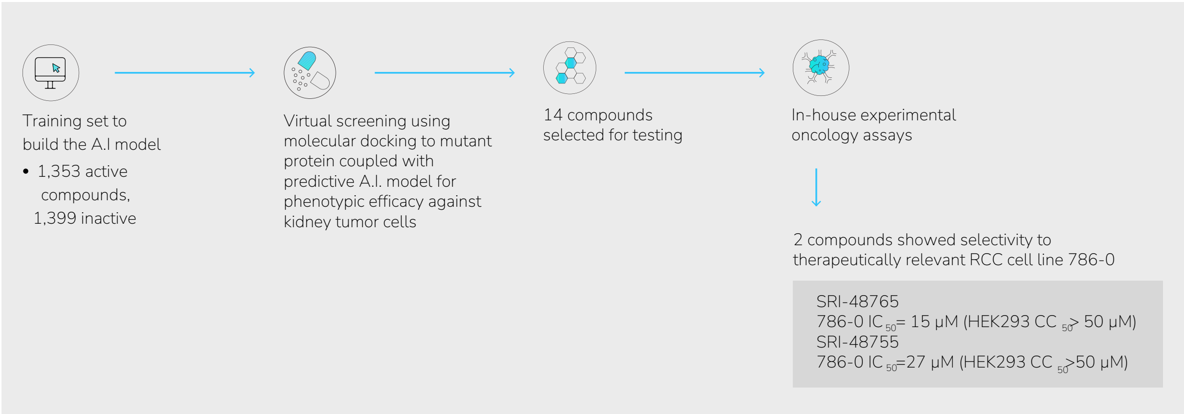
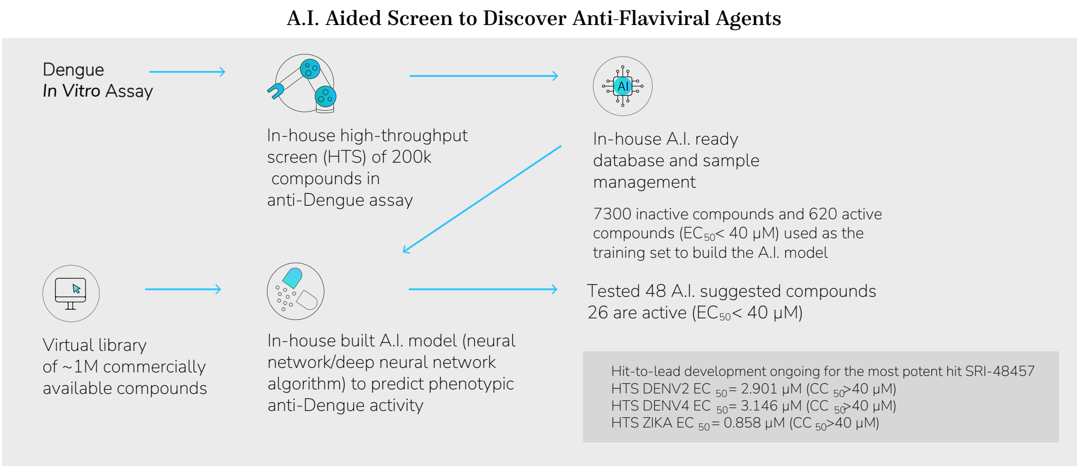


Figure 3. A.I. aided screen to discover anti-flaviviral compounds.

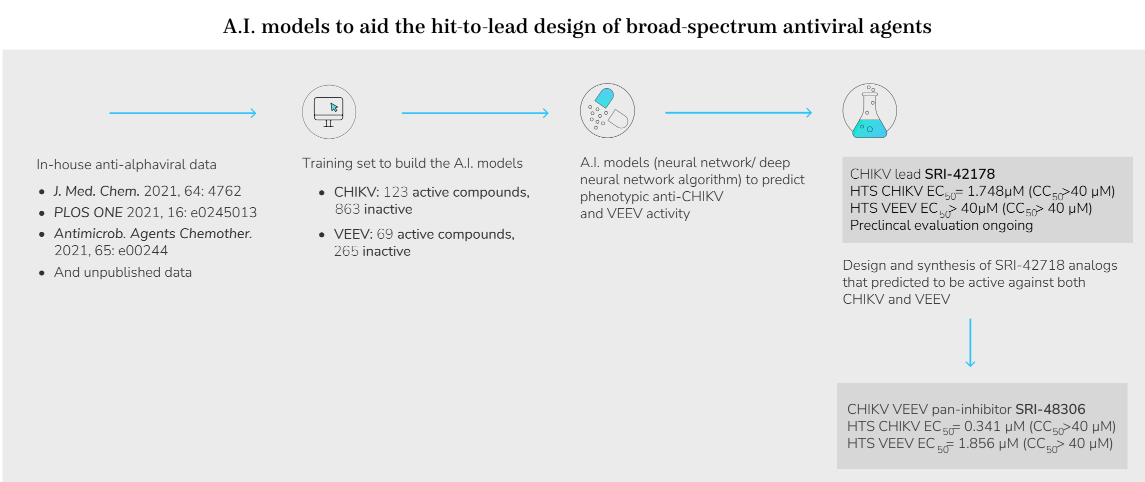


CASE STUDY #3:A.I.-AIDED DESIGN OF ANTI-ALPHAVIRAL AGENTS

Alphaviruses, like Chikungunya virus (CHIKV) and Venezuelan Equine Encephalitis Virus (VEEV), are life-threatening, mosquito-borne viruses that currently lack targeted treatments. Southern Research molecular modeling tools and in-house A.I. models have enabled the development of broader antiviral spectrum analogs of a CHIKV lead compound **(Figure 4)**. Data from more than 1,000 compounds with anti-CHIKV/VEEV activity observed through in-house cell-based assays were used to train A.I. models with a deep neural network algorithm to predict anti-CHIKV/VEEV activities. New compounds that could be effective for both CHIKV and VEEV were designed based on molecular modeling of viral protein binding specificity and A.I. prediction of phenotypic antiviral activity.

Assay data showed that, compared to the CHIKV-only lead compound, newly designed analogs also have equal potency against VEEV. The combination of molecular modeling and A.I. models successfully designed compounds with the desirable antiviral spectrum. This approach to drug development can provide an effective means for rapid response to emerging epidemics.

Figure 4. Molecular modeling and in-house A.I. models combine to identify compounds with anti-alphaviral properties and design analogs predicted to have broader spectrum activity.



CONCLUSION

The efficiency of the drug discovery process will improve with the help of A.I. technology. Southern Research is poised to partner with companies to accelerate their preclinical drug discovery efforts by leveraging their extensive wet lab capabilities, advanced *in silico* tools, and high accuracy A.I. models. With the expertise and resources to design custom solutions and develop new A.I. models, Southern Research can help bring new drugs to the market faster.

REFERENCES

1. Chan, H.C. Stephen, et al. Advancing Drug Discovery via Artificial Intelligence. Trends in Pharmacological Sciences. 2019 Aug; 40(8): 592-604.
2. Paul, Debleena, et al. Artificial intelligence in drug discovery and development. Drug Discovery Today. 2021 Jan; 26(1): 80-93.
3. Martis, E. A., Radhakrishnan, R., & Badve, R. R. (2011). High-throughput screening: the hits and leads of drug discovery-an overview. Journal of Applied Pharmaceutical Science. 01: 02-10.
4. Adashi, Eli Y. MD, MS, et al. The FDA Modernization Act 2.0: Drug Testing in Animals is Rendered Optional." The American Journal of Medicine. 18 Apr 2023.
5. Philippidis, A. "Insilco Gains FDA's First Orphan Drug Designation for AI Candidate." Genetic Engineering & Biotechnology News. 10 Feb. 2023.
6. KC, G.B., Bocci, G., Verma, S., et al. A machine learning platform to estimate anti-SARS-CoV-2 activities. Nature Machine Intelligence. 2021 May; 3: 527-535.
7. Fagerberg, J.H., Karlsson, E., Ulander, J. et al. Computational prediction of drug solubility in fasted simulated and aspirated human intestinal fluid. Pharmaceutical Research. 2015 Feb; 32: 578-589.
8. Perryman, A.L., Stratton, T.P., Ekins, S. et al. Predicting mouse liver microsomal stability with "pruned" machine learning models and public data. Pharmaceutical Research. 2016 Feb; 33: 433-449.
9. Kruk, L., et.al. Inflammatory Networks in Renal Cell Carcinoma. Cancers. 2023 Apr; 15(8): 2212.
10. Kaelin, W. Von Hippel-Lindau disease: insights into oxygen sensing, protein degradation and cancer. Journal of Clinical Investigation. 2022 Sep 15; 132(18): e162480.