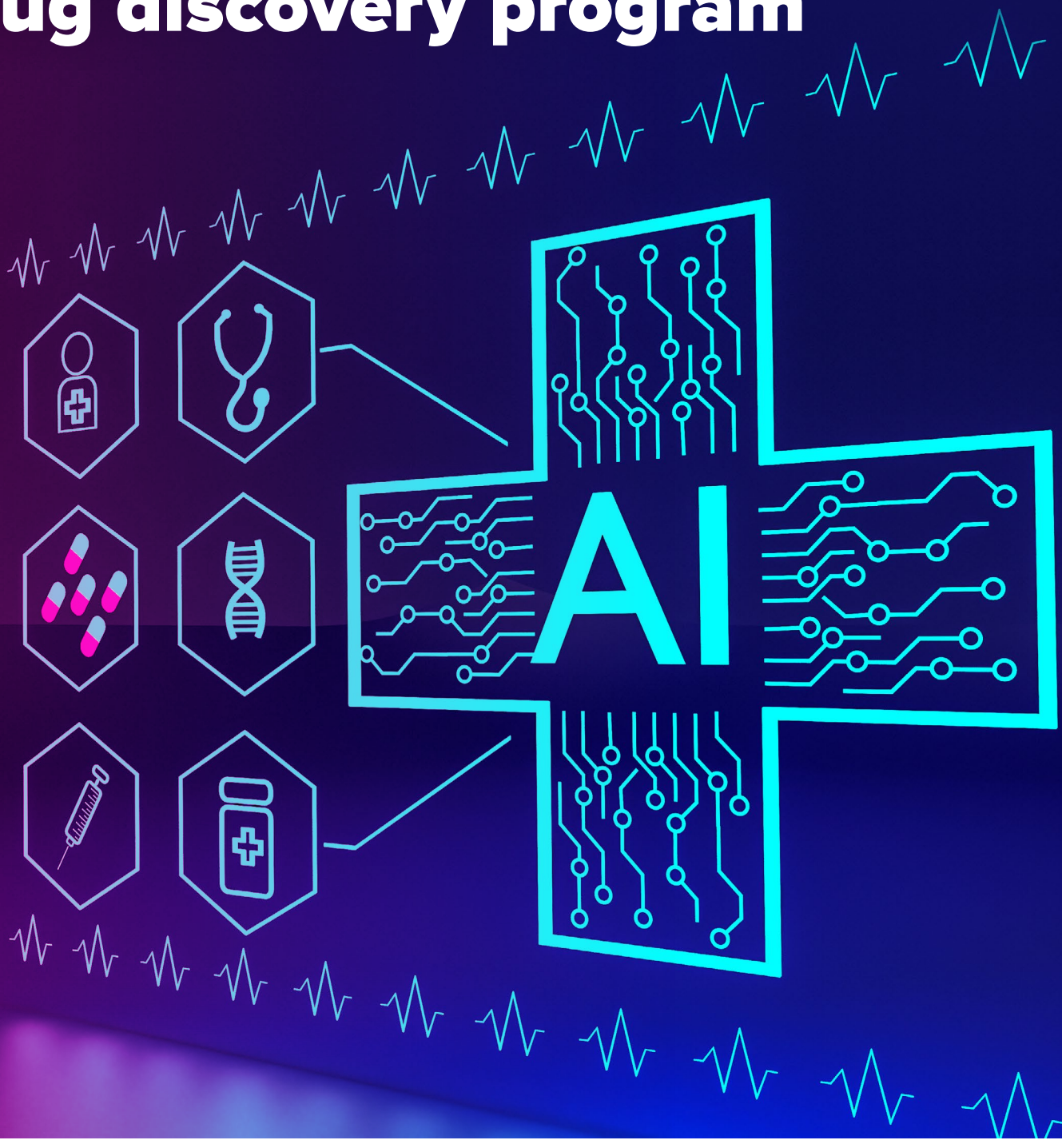
WHITE PAPER

How to build your AI-enabled drug discovery program



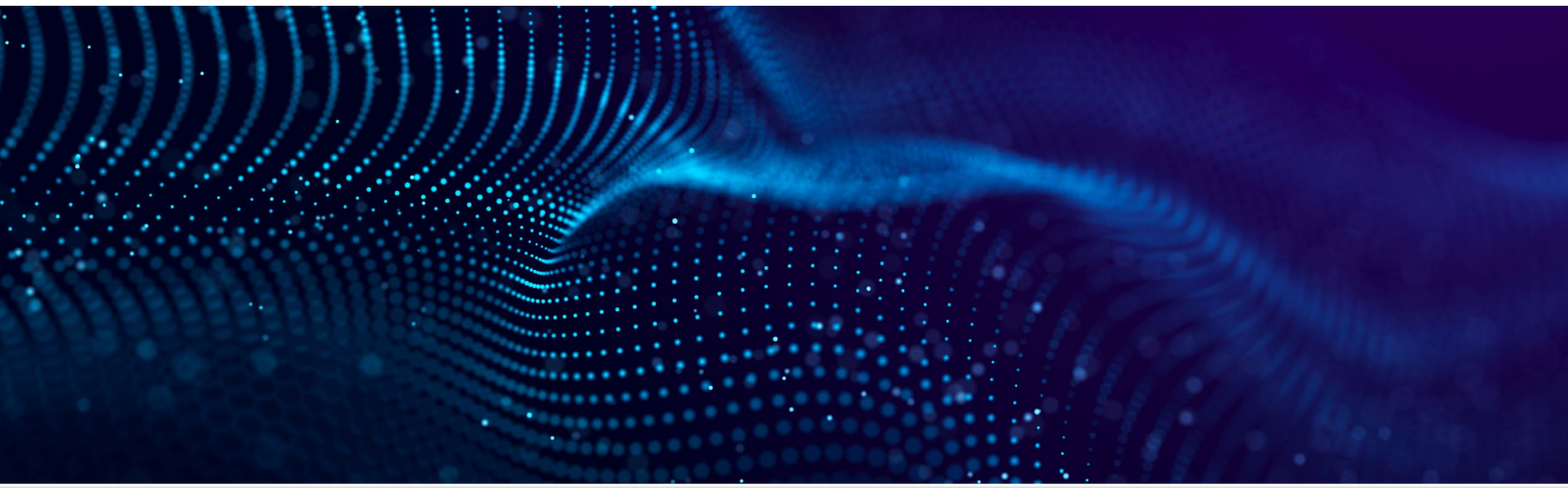
PUBLISHED BY:

FIERCE Life Sciences

PRESENTED BY:

MilliporeSigma

CONTENTS



Introduction

Although the pharmaceutical industry has been using machine learning, a type of artificial intelligence (AI), for decades, modern methods like generative AI are now transforming the industry in new ways (Duran & Chaudhuri, 2024) .

Using traditional methods, pharmaceutical companies invest, on average, more than $2 billion and 15 years of development time to bring one new medicine to market. While larger companies often hedge these investments with a rich product pipeline, revenue from licensing agreements and commercialized products, small companies struggle to develop new medicines because of the substantial costs and risks involved (Qureshi et al., 2023).

Today , even pharmaceutical companies without prior experience in AI have either started using AI technologies internally or are partnering with specialized AI companies to help them develop drugs

from bench to bedside. Indeed, the third-party investment in AI-enabled drug discovery doubled each year between 2017 and 2021, reaching $5.2 billion at the end of 2021. Some AI companies focus on AI tools that optimize a particular stage of small molecule drug discovery. In contrast, other companies develop end-to-end platforms with AI at the center of each step (Ayers et al., 2022; Kirkpatrick, 2022).

AI tools can help level the playing field for small- and medium-sized companies in small molecule drug discovery research by providing access to new disease targets; enhancing the properties of existing pipeline compounds; improving success rates of clinical trials; and accelerating the discovery process of new molecules while reducing overall costs (Ayers et al., 2022).

AI has already revolutionized the pharmaceutical industry, and we can only speculate about the ways AI will change drug discovery programs in the future. In this playbook, we outline considerations before implementing AI-enabled solutions in your current drug discovery programs.

Recent milestones of using AI tools in drug discovery include:

- **Early 2020 –** The first AI-designed molecule enters human clinical trials.
- **July 2021 –** AlphaFold predicts the protein structures for 330,000 proteins, including the entire human genome.
- **February 2022 –** The first AI-designed molecule based on an AI-discovered disease target enters phase 1 clinical trials.
- **January 2023 –** Generative AI creates and validates *de novo* antibodies *in silico*.
- **February 2023 –** The first AI-designed molecule is granted Orphan Drug Designation by the Food and Drug Administration (FDA).

February 2024 – The first AI-designed molecule is planned to start a phase 2a study to evaluate its safety and tolerability in adults with Idiopathic Pulmonary Fibrosis (Chun, 2023; ClinicalTrials.gov).

What to consider before implementing AI-enabled solutions for your small molecule drug discovery program

The exponential growth of AI tools and databases adds substantial complexity to the field. By providing what we believe are the most important considerations before integrating AI tools into your existing workflows, we hope to help you start the journey toward your AI-enabled drug discovery.

General considerations for AI-enabled solutions

Define objectives for your AI-enabled solution

AI tools offer great potential to use your resources more efficiently. Today, you can find a potential use case for AI tools in every step of the drug development process:

- Machine learning models can be trained on large datasets to understand biological mechanisms of diseases and, combined with deep neural networks, predict 3D structures of targets.
- Generative AI models can screen huge virtual libraries and design new compounds *de novo*.
- Decision trees and deep learning models can suggest a retrosynthesis strategy for drug candidates.
- Deep learning models can predict efficacy and toxicity of new molecules in pre-clinical and clinical development.

Natural language processing models can mine the scientific literature for regulatory approval and post-marketing analysis (Chun, 2023; Qureshi et al., 2023).

To successfully integrate AI tools into your existing workflows, identify the steps of your small molecule discovery program you would like to enhance and define specific, measurable, achievable, relevant and time-bound (SMART) objectives you want to achieve.

Build a cross-functional team

Successful integration of AI tools in small molecule drug discovery programs requires effective collaboration between subject matter experts of distinct fields: data scientists and engineers, drug discovery scientists, and subject matter experts for the software of the AI tools. Organizational siloes and inefficient communication between these teams are significant barriers to integrating AI-enabled solutions (Rodriguez et al., 2023).

Depending on the size of your current team and the AI tools you plan to implement, you may have to hire data scientists or upskill your in-house staff to:

- Operate the software of AI tools
- Train or update AI algorithms with your data
- Transform data that are fed into AI algorithms
- Evaluate the validity and reproducibility of data input and output
- Troubleshoot issues and fix bugs.

Build trust towards and promote responsible use of AI-enabled solutions

Lack of trust is a crucial barrier to adopting AI-based solutions, and opinions vary between *"AI is the future and will help us help us form new scientific laws"* and *"AI is a new hype that has yet to be proven"* (Rodriguez et al., 2023).

Most issues can be attributed to four areas:

- People do not fully understand the AI tools
- People do not trust the data generated by AI tools
- People do not see the value AI tools bring to their processes
- People are worried AI tools may replace them

To build trust, transparently communicate AI-enabled solutions' successes, failures, performance and limitations with your cross-functional team. Promote the responsible use of AI tools in your drug discovery program and in the industry (Rodriguez et al., 2023).

Technical considerations for AI-enabled solutions

Evaluate your existing data infrastructure

AI-enabled solutions are data-intensive, and limitations for harnessing their full potential are often rooted in the data storage infrastructure (Anderson, 2023).

Evaluate your existing data storage infrastructure for:

- **Storage capacity –** prepare for data volumes well beyond current estimates.
- **Speed, scalability, and adaptability –** ensure the training of AI tools and the predictions of AI models can be GPU-accelerated and scaled up efficiently.
- **Safety and privacy –** evaluate safety features like data encryption and access controls and make sure no personal or proprietary information is shared. Ensure compliance with ISO 27001, a widely accepted international standard for information security (ISO (2022)).
- **Interoperability –** evaluate the need to transform data to ensure the interoperability between cloud and local storage options (Anderson, 2023; Franklin, 2022).

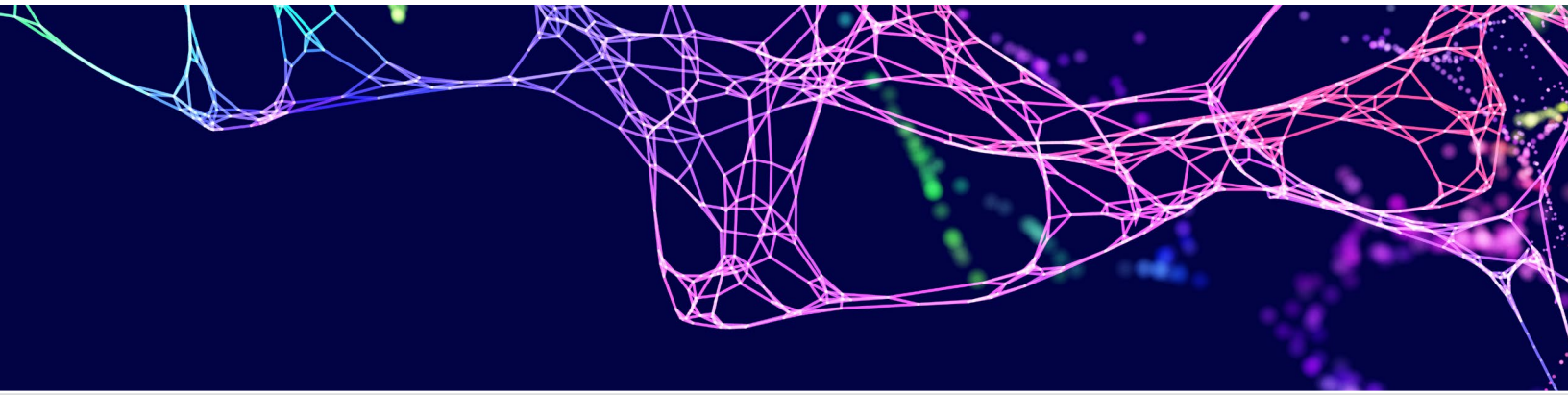
Assess the usability of AI tools

AI tools are often developed without broad use cases in mind. This is especially true for some open-source models. Frequently, these models are not regularly maintained, have complicated user interfaces, or cannot easily integrate into existing workflows.

Commercial tools are more expensive, but could offer important advantages:

- Proprietary datasets are often high-quality and sufficiently rich.
- Vendors may offer training and support to ensure their AI solution is optimally set up for their customer.

Assess the usability of open-source tools and consider purchasing commercial tools that are easy to use and maintain (Rodriguez et al., 2023).



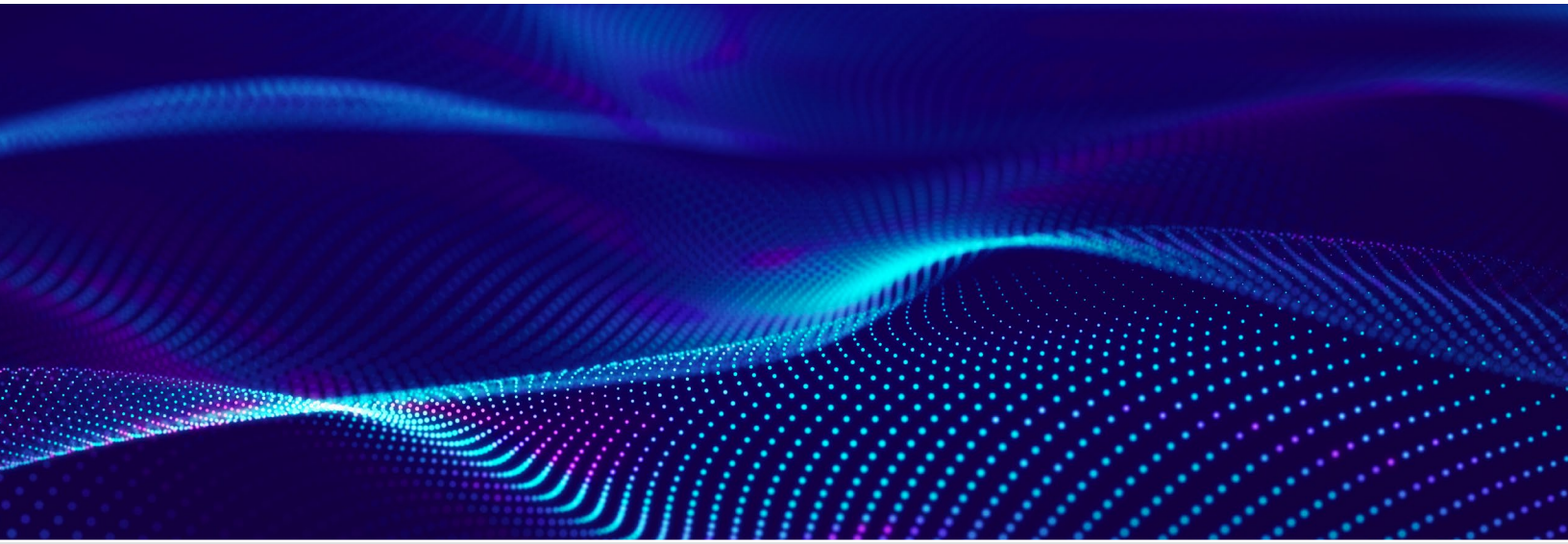
Scientific considerations for AI-enabled solutions

Demand high-quality datasets and mature AI tools

The quality of predictive models that screen for molecule design, property prediction, retrosynthesis and other applications are critical for the successful integration of AI-enabled solutions into your small molecule discovery workflows. Consider the following aspects:

- **Size –** Datasets for training algorithms and screening must be sufficiently robust and diverse. The exact size depends on your specific application. As a rule of thumb, larger is better if, and only if, the data are independently distributed and not biased.
- **Bias –** Datasets must reduce bias towards age, gender, race, geographic location or other essential factors.
- **Data privacy and security –** Datasets must transparently address concerns around collecting, storing and sharing sensitive data, like patient information or proprietary data.
- **Consistency –** Datasets for training AI tools and screening must be labeled consistently.
- **Transparency –** Datasets must communicate transparently the sources of data and the methods by which AI tools were trained.
- **Ownership –** Datasets must transparently communicate who owns the input and output data.
- **Customizability –** Datasets to train AI models must be customizable with your own proprietary datasets to closely link output and predictions to your field.

Open-source databases exist for many applications in the early phase of the drug discovery pipeline. They may, however, lack quality in one or more of the above. Consider purchasing proprietary databases if they meet the quality requirements for your specific use case (Rodriguez et al., 2023).



Assess the efficiency and effectiveness of the AI-enabled solution

AI tools could potentially assist in all stages of your small molecule drug discovery workflow, from target selection and validation to post-marketing surveillance (Kate et al., 2023). These tools are especially useful in the early stages of drug discovery.

- Deep-learning algorithms can learn from vast amounts of data and quickly sort through vast virtual chemical spaces (10^{60} molecules) to suggest a small number of potential drug candidates.
- Machine learning algorithms, like random forests, k-nearest neighbors and support vector machines, are traditionally used to predict quantitative structure-property relationships (QSPR) and other important molecular properties.

Neural networks can outperform human-encoded reaction rules for retrosynthesis analyses and reaction predictions (Hu et al., 2023).

Carefully consider if the AI tool can help achieve your goals more efficiently and effectively than traditional strategies.

Financial considerations for AI-enabled solutions

Estimate the return on investment

AI tools have the potential to create a significant return on investment. The main drivers are more efficient use of existing resources, reduced rate of failures along the clinical development program and the identification of new molecular targets and optimized drugs. For example, AI tools can:

- Reduce the number of long and expensive laboratory experiments (or in some cases replace them entirely)
- Develop improved hypotheses and better select molecules with desired properties such as efficacy, toxicity and immunogenicity
- Generate novel therapeutic structures for previously undruggable targets (Rodriguez et al., 2023)

The benefits need to be contrasted with the investment in implementing AI-enabled solutions. The main investments are:

- **Data storage –** Implementing AI-based solutions will probably require updating your existing data storage solutions and buying on-demand cloud-solutions.
- **Software as a Service (SaaS) licenses –** Working with multiple vendors and potentially disparate tools can add substantial costs. Streamlining SaaS licenses and working with a single vendor and platform that can accomplish various tasks could lower the investment burden.
- **Training and hiring a cross-functional team –** Experts in data science, software development and technical drug discovery need to be trained or hired to collaborate successfully.
- **Maintenance and updates –** Databases and AI models must be maintained properly and updated regularly to produce high-quality output and predictions.
- **Materials –** New compounds suggested by AI tools must be synthesized and evaluated in the laboratory.

Logistic considerations for AI-enabled solutions

Establish the required infrastructure to support AI-based solutions

The infrastructure necessary for AI-enabled solutions in your small molecule discovery workflow may differ significantly from your traditional infrastructure. Depending on the AI-enabled solution, you may need:

- Servers to store large amounts of data
- GPU-accelerated hardware for computers
- Fast inter- and intranet connections
- Security systems for access control and identity confirmation
- Separate office space for data scientists close to the production environment
- Digital archives for documentation and licenses

Source raw materials for testing AI-designed compounds

Despite the ability of AI tools to design drugs virtually and predict their efficacy, toxicity, and other important properties, AI-designed compounds still need to be tested using traditional laboratory methods to confirm the predictions and ultimately obtain regulatory approval. However, *in silico* screening and compound design results in fewer molecules that need to be synthesized in the lab (or procured) reducing the consumption of laboratory chemicals and supplies.

Assess the need for adjusting the amount of raw materials, laboratory equipment, and work hours when using AI-enabled solutions in your small-molecule discovery workflows.

How AIDDISON™ can assist in your small molecule drug discovery program

What is AIDDISON™?

AIDDISON™ is a SaaS-based drug discovery platform that combines AI, machine learning, and computer-aided drug design (CADD) to efficiently and effectively design ligand-based and structure-based drugs. It integrates all facets of virtual screening and is a valuable tool for lead discovery and optimization (Merck KGaA, Darmstadt, Germany 2024).

What are the benefits of AIDDISON™?

The benefits of the AIDDISON™ software include:

- Fully-integrated software platform eliminates the need to switch between tools and data reformatting.
- Integrated machine learning models trained on proprietary and proven experimental assay data from the pharmaceutical industry that can be updated as project needs change.
- Easily design *de novo* small molecules using generative methods with drug-like properties.
- Secure and scalable on-demand cloud services for faster results.
- ISO 270001 certification for secure end-to-end data flow.
- Intuitive user interface allows all levels of expertise to quickly learn the software.
- Effective decision support at each step of the process from molecule design to chemical synthesis.
- Customizable platform to train AI/ML models on your proprietary data or to enumerate chemical space to focus on your specific therapeutic targets.(Merck KGaA, Darmstadt, Germany 2024)

What are the features of AIDDISON™?

Features of AIDDISON™ include:

- **De novo molecular design –** The AI-based generative model, trained on three decades of validated R&D data, can design novel, drug-like molecules and optimize for synthetic accessibility, ADME-Toxicity properties and more.
- **Similarity and pharmacophore search –** The software can rapidly screen 2D structures and pharmacophores among more than 64 billion virtual molecules, with access to millions of commercially-available starting materials.
- **Shape-based search –** The software can identify, filter and sort drug candidates by closely matching 3D molecule structures to a reference ligand and cluster optimal subsets of molecules.
- **Molecular docking –** The software can confirm molecular interactions by aligning the desired drug candidates with the protein's active site in 3D (Merck KGaA, Darmstadt, Germany 2024).

Checklist to begin building your AI-enabled drug discovery program

- ☐ Define objectives for your AI-enabled solution
- ☐ Build a cross-functional team
- ☐ Build trust towards and promote responsible use of AI-enabled solutions
- ☐ Evaluate your existing data storage infrastructure
- ☐ Assess the usability of AI tools
- ☐ Demand high-quality datasets and mature AI tools
- ☐ Assess the efficiency and effectiveness of the AI-enabled solution
- ☐ Estimate the return on investment
- ☐ Establish the required infrastructure to support AI-based solutions
- ☐ Source raw materials for testing AI-designed compounds
- ☐ Reach out to aiddison@milliporesigma.com to learn more about how AIDDISON™ can be integrated into your small molecule drug discovery workflow.

About MilliporeSigma

Before researchers can make scientific breakthroughs, they must have access to state-of-the-art tools, services and expertise to perform experiments and engineer new products. At MilliporeSigma, that's where we come in. We offer one of the broadest portfolios in the industry for scientists. Researchers around the world rely on us to do their best science — from familiar experiments to novel approaches and technologies.

Our more than 28,000 employees believe science offers unlimited possibilities. This sentiment fuels each new innovation. We know our cutting-edge products, services and digital offerings create solutions for people around the world and drive sustainable futures for subsequent generations.

The life science business of Merck KGaA, Darmstadt, Germany, operates as MilliporeSigma in the US and Canada. Merck KGaA, Darmstadt, Germany is a global science and technology company with more than 64,000 employees in 66 countries.

Explore our offerings on SigmaAldrich.com.

References:

- Anderson, K. (2023). *Choosing the right storage solution for AI*. Telnyx LLC. https://telnyx.com/resources/storage-for-ai
- Ayers, M., Jayatunga, M., Goldader, J., & Meier, C. (2022). *Adopting AI in Drug Discovery*. Boston Consulting Group. https://www.bcg.com/publications/2022/adopting-ai-in-pharmaceutical-discovery
- Chun, M. (2023). *How Artificial Intelligence is Revolutionizing Drug Discovery*. Petrie-Flom Center; Havard Law School. https://blog.petrieflom.law.harvard.edu/2023/03/20/how-artificial-intelligence-is-revolutionizing-drug-discovery/)
- ClinicalTrials.gov. *Study Evaluating INS018_055 Administered Orally to Subjects With Idiopathic Pulmonary Fibrosis*. Retrieved 21.Feburary from https://clinicaltrials.gov/study/NCT05975983?term=INS018_055%20&rank=1
- Duran, T., & Chaudhuri, B. (2024). Where Might Artificial Intelligence Be Going in Pharmaceutical Development? *Molecular Pharmaceutics*, *21*(3), 993-995. https://doi.org/10.1021/acs.molpharmaceut.4c00112
- Franklin, A. (2022). *Choosing the Right Storage for Enterprise AI Workloads*. NVIDIA Developer. https://developer.nvidia.com/blog/choosing-the-right-storage-for-enterprise-ai-workloads/
- Hu, W., Liu, Y., Chen, X., Chai, W., Chen, H., Wang, H., & Wang, G. (2023). *Deep Learning Methods for Small Molecule Drug Discovery: A Survey*.
- Kate, A., Seth, E., Singh, A., Chakole, C. M., Chauhan, M. K., Singh, R. K., Maddalwar, S., & Mishra, M. (2023). Artificial Intelligence for Computer-Aided Drug Discovery. *Drug Res (Stuttg)*, *73*(7), 369-377. https://doi.org/10.1055/a-2076-3359
- Kirkpatrick, P. (2022). Artificial intelligence makes a splash in small-molecule drug discovery. *Nature Reviews Drug Discovery*, *21*, 175-176.
- Merck KGaA, Darmstadt, Germany (2024) and/or one of its affiliates. *AIDDISON™ AI-Powered Drug Discovery*. https://www.sigmaaldrich.com/AT/en/services/software-and-digital-platforms/aiddison-ai-powered-drug-discovery#form
- Qureshi, R., Irfan, M., Gondal, T. M., Khan, S., Wu, J., Hadi, M. U., Heymach, J., Le, X., Yan, H., & Alam, T. (2023). AI in drug discovery and its clinical relevance. *Heliyon*, *9*(7), e17575. https://doi.org/10.1016/j.heliyon.2023.e17575
- Rodriguez, A., Meier, C., Serazin, E., Gooch, J., Aggarwal, P., Steene, A., Jayatunga, M., Nayak, S., & Randall, W. (2023). *Unlocking the Potential of AI in Drug Discovery*. Boston Consulting Group. https://www.bcg.com/publications/2023/unlocking-the-potential-of-ai-in-drug-discovery
- Standardization, I. O. f. (2022). *Information security, cybersecurity and privacy protection – Information security management systems (ISO/IEC 27001:2022)*. https://www.iso.org/standard/27001
- Andrew Rusinko, Mohammad Rezaei, Lukas Friedrich, Hans-Peter Buchstaller, Daniel Kuhn, and Ashwini Ghogare. AIDDISON: Empowering Drug Discovery with AI/ML and CADD Tools in a Secure, Web-Based SaaS Platform.
- *Journal of Chemical Information and Modeling* 2024 64 (1), 3-8, DOI: 10.1021/acs.jcim.3c01016