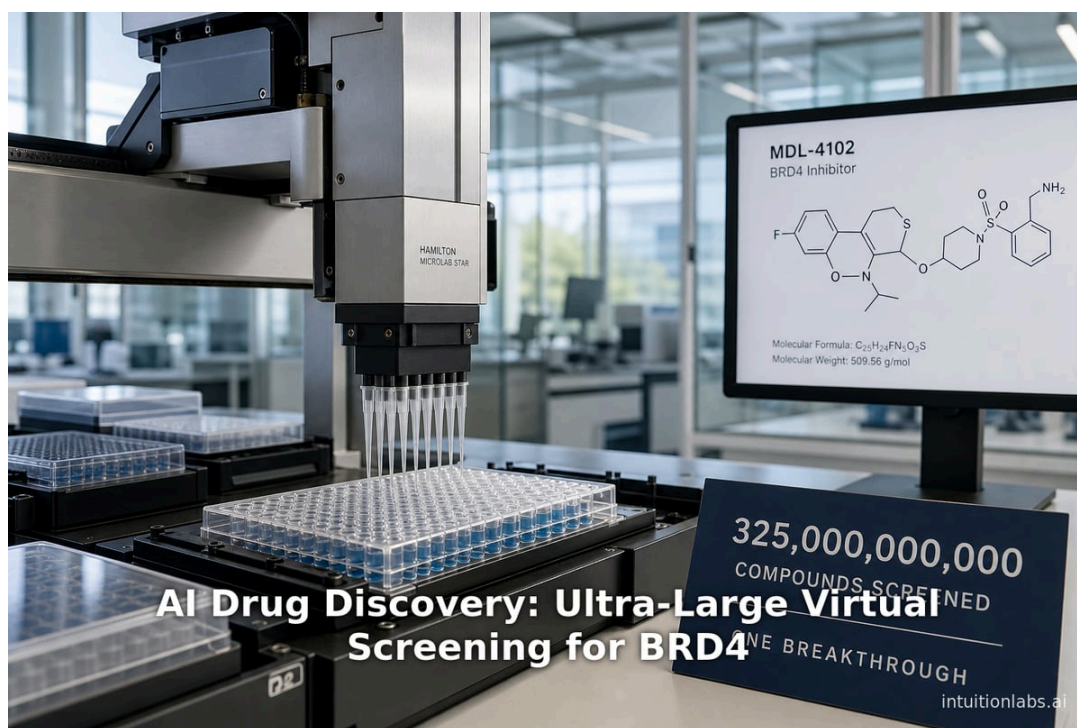


AI Drug Discovery: Ultra-Large Virtual Screening for BRD4

4/30/2026 • 30 min read

ai drug discovery ultra-large virtual screening brd4 inhibitors bet proteins computational chemistry high-throughput screening oncology therapeutics



Executive Summary

Drug discovery continues to struggle against the vastness of chemical space and the high failure rate of traditional pipelines. **Model Medicines**, an AI-driven biotechnology company, has aggressively tackled these challenges by deploying **ultra-large virtual screening (ULVS)**. In late 2025, using its *GALILEO*[™] AI platform on [Google Cloud](#), Model Medicines executed a virtual screening of **325 billion** candidate molecules in a single day ^{([1](#))} [www.biospace.com](#)) ^{([2](#))} [modelmedicines.com](#)). This unprecedented throughput—far beyond prior efforts—yielded **MDL-4102**, a novel, highly selective BRD4 inhibitor designed for oncology. MDL-4102 is *first-in-class*: it robustly inhibits BRD4 yet shows *no measurable activity* against closely related BET family proteins BRD2 or BRD3 ^{([2](#))} [modelmedicines.com](#)). Chemoinformatics analysis confirms MDL-4102's chemical structure is distinct from all previously clinical BET inhibitors (low ECFP4-Tanimoto overlap) ^{([3](#))} [modelmedicines.com](#)). Together, these results suggest that giga-scale AI-driven searches of chemical space can uncover entirely new classes of drug candidates once deemed unattainable.

This report provides an in-depth analysis of Model Medicines' BRD4 oncology effort. It contextualizes the MDL-4102 project within the history and current state of **ultra-large virtual screening**, details Model Medicines' methodologies and platform, and examines the biological target **BRD4** and its role in cancer therapeutic development. Data from the 325-billion-molecule screen are presented, highlighting how the GALILEO platform translated computational scale into a candidate molecule with novel chemistry. We compare MDL-4102 to existing BET inhibitors (Table 1), discuss Model Medicines' broader AI capabilities (e.g. MDL-001 antiviral program), and critically assess the potential and limitations of **AI-driven drug discovery** ^{([2](#))} [modelmedicines.com](#)) ^{([4](#))} [www.scientific-computing.com](#)). Future directions – including trillion-scale screening, multiparameter optimization with **AI agents**, and the eventual path toward clinical trials – are also explored.

Introduction and Background

The Drug Discovery Challenge

Traditional drug discovery is laborious and expensive, often requiring a decade or more and billions of dollars to bring a single new medicine to market ^{([5](#))} [medicalxpress.com](#)). In practice, nearly 90% of drug candidates fail during clinical development ^{([6](#))} [medicalxpress.com](#)). A persistent bottleneck is **lead identification** – finding small molecules that bind the therapeutic target with high potency and selectivity, while also possessing **drug-like properties**. Historically, lead discovery has often relied on **high-throughput screening (HTS)** of physical compounds. Yet even large-scale HTS facilities can test only millions of compounds per day ^{([7](#))} [www.biospace.com](#)), a vanishingly small fraction of possible molecules. Approximately **10⁶⁰** “drug-like” small molecules are believed to exist in principle ^{([8](#))} [www.biospace.com](#)) ^{([9](#))} [pmc.ncbi.nlm.nih.gov](#)). Thus, conventional screening has been likened to finding “a single drop in all the world's oceans” ^{([8](#))} [www.biospace.com](#)).

This vast **chemical space** implies that most promising chemotypes remain undiscovered. Virtual screening (VS) – computationally evaluating large libraries of structures against a target – offers a way to explore more of this space at lower cost. Early docking and similarity-based approaches sifted millions of candidates in silico ^{([10](#))} [medicalxpress.com](#)), but their throughput has historically lagged far behind the theoretical size of chemical space. Recent advances in compute power, library enumeration, and especially **AI/ML** have begun to shift this balance. The key insight is that screening more compounds can yield **better** leads: as Harvard School of Medicine's Gorgulla et al. note, “the more compounds you can screen, the better your top candidates will be, and the lower your rate of false positives” ^{([11](#))} [medicalxpress.com](#)).

Ultra-Large Virtual Screening (ULVS)

The concept of **ultralarge virtual screening (ULVS)** has emerged as computing has scaled. Research teams have sought to expand the number of docked or predicted molecules from millions to billions or trillions ^{([12](#))} [www.biospace.com](#)) ^{([13](#))} [www.sciencedirect.com](#)). Early pioneering work includes **VirtualFlow** (2020), an open-source platform from Harvard researchers, which prepared a ready-to-screen library of **1.4 billion** compounds ^{([10](#))} [medicalxpress.com](#)). VirtualFlow was designed to run on large computer clusters or cloud grids, demonstrating that even academic groups could screen well beyond previous limits. For example, a 10,000-core cluster could dock 1 billion compounds in about two weeks ^{([14](#))} [medicalxpress.com](#)).

Other approaches include physics-based methods and advanced ML models. Full free-energy perturbation (FEP) simulations, like Schrodinger's *FEP+* and related *Boltz-2* algorithms, can estimate binding energetics, but remain slow (on the order of 2 million evaluations per day even on specialized hardware ^{(15]} modelmedicines.com)). By contrast, **predictive ML models** – neural networks or parametric scoring functions trained on bioactivity data – can achieve much higher raw throughput. For instance, Atomwise's AtomNet deep-learning model has reportedly reached **8 billion molecules per day** on its latest systems ^{(15]} modelmedicines.com)). Yet prior to 2025, all such efforts remained below the **hundreds of billions** barrier ^{(15]} modelmedicines.com) ^{(13]} www.sciencedirect.com). A recent review confirms: "to our knowledge, however, no prior predictive screen has surpassed the hundred-billion scale" ^{(16]} modelmedicines.com).

Model Medicines' campaign therefore represents a watershed moment. In partnership with Google Cloud, it achieved **325 billion molecules** screened in 24 hours ^{(1]} www.biospace.com) ^{(2]} modelmedicines.com) – an order of magnitude beyond previous state-of-the-art. This ultra-large scale promises to find entirely new chemotypes ("a fundamentally broader exploration of chemical space" ^{(3]} modelmedicines.com)) that smaller screens cannot reach. Nevertheless, ULVS faces challenges such as scoring accuracy, computational cost, and data bias ^{(13]} www.sciencedirect.com). Combining ML with structure-based docking or hybrid methods may help mitigate these issues ^{(13]} www.sciencedirect.com). Ultimately, Model Medicines shows how pushing throughput can be a "precision amplifier" in discovery ^{(17]} www.biospace.com).

BRD4 and Oncology Targets

Bromodomain-containing protein 4 (**BRD4**) is a transcriptional regulator in the BET (bromodomain and extra-terminal) family ^{(18]} pmc.ncbi.nlm.nih.gov) ^{(19]} pmc.ncbi.nlm.nih.gov). It contains two conserved N-terminal bromodomains (BD1 and BD2) that bind acetylated lysines on histones and transcription factors ^{(20]} pmc.ncbi.nlm.nih.gov) ^{(21]} pmc.ncbi.nlm.nih.gov). By recruiting mediators of transcription elongation (e.g., P-TEFb complex), BRD4 helps control expression of critical genes, including oncogenes like *c-MYC* ^{(22]} pmc.ncbi.nlm.nih.gov) ^{(19]} pmc.ncbi.nlm.nih.gov). Abnormal BRD4 activity has been implicated in various cancers, inflammation, and viral processes ^{(19]} pmc.ncbi.nlm.nih.gov). Notably, the fusion of BRD4 with the **NUT** protein drives aggressive NUT-midline carcinoma ^{(23]} pmc.ncbi.nlm.nih.gov), and BRD4 overactivity has roles in leukemia, lymphoma, and other tumors ^{(19]} pmc.ncbi.nlm.nih.gov) ^{(23]} pmc.ncbi.nlm.nih.gov).

Because it sits at a regulatory nexus, BRD4 has been an attractive, albeit challenging, drug target. Dozens of small-molecule **BET inhibitors** have been developed or are in trials ^{(22]} pmc.ncbi.nlm.nih.gov) ^{(24]} pmc.ncbi.nlm.nih.gov). Examples include JQ1 (research tool), OTX015 (Merck MK-8628), and ABBV-075 (ABBV-744). These compounds generally bind the acetyl-lysine pocket and typically inhibit both BD1 and BD2 domains of BET proteins ^{(25]} pmc.ncbi.nlm.nih.gov) ^{(26]} pmc.ncbi.nlm.nih.gov). However, because BRD2 and BRD3 are highly homologous to BRD4 ^{(27]} pmc.ncbi.nlm.nih.gov), most inhibitors block multiple BET family members. This broad activity can lead to dose-limiting toxicities (especially thrombocytopenia) and limited efficacy ^{(28]} pmc.ncbi.nlm.nih.gov).

As of 2025, several BET inhibitors have advanced in clinical trials for various hematologic and solid tumors ^{(29]} pmc.ncbi.nlm.nih.gov) ^{(30]} pmc.ncbi.nlm.nih.gov). For instance, **molibresib (GSK525762)** showed some tumor responses in NUT carcinoma and AML patients, but **thrombocytopenia and neutropenia** curtailed dose escalation ^{(31]} pmc.ncbi.nlm.nih.gov). **BAY 1238097 (Bayer)** achieved biomarker evidence of on-target effect (MYC reduction) but induced severe gastrointestinal toxicity, terminating its development ^{(32]} pmc.ncbi.nlm.nih.gov). **INCB054329 (Incyte)** saw a minority of patients respond in a Phase I/II trial ^{(33]} pmc.ncbi.nlm.nih.gov). Even promising combinations (e.g. *pelabresib* CPI-0610 with ruxolitinib in myelofibrosis) face hematologic side effects ^{(34]} pmc.ncbi.nlm.nih.gov). Overall, the BRD4 inhibitor landscape (Table 1) is littered with compounds that are **potent but nonselective**, leading to narrow therapeutic windows ^{(28]} pmc.ncbi.nlm.nih.gov) ^{(35]} pmc.ncbi.nlm.nih.gov). Selective inhibition of BRD4 alone (sparing BRD2/3) has long been viewed as a "holy grail," since it might retain anti-tumor activity while reducing off-target toxicity ^{(36]} pmc.ncbi.nlm.nih.gov) ^{(2]} modelmedicines.com).

Table 1. Selected BRD4/BET Inhibitors in Clinical Development. Compiled from recent reviews and trial reports ^{(37]} pmc.ncbi.nlm.nih.gov) ^{(35]} pmc.ncbi.nlm.nih.gov).

Compound (Developer)	Indications	Status (Phase)	Outcome/Notes
Molibresib (GSK525762, GSK)	NUT midline carcinoma, AML*, other cancers	Phase I	Some partial remissions observed; thrombocytopenia/neutropenia limited dosing ^{(31]} pmc.ncbi.nlm.nih.gov).
BAY 1238097 (Bayer)	Solid tumors	Phase I	On-target biomarker activity (MYC-) noted, but severe GI/toxicities (nausea, vomiting) led to termination ^{(32]} pmc.ncbi.nlm.nih.gov).

Compound (Developer)	Indications	Status (Phase)	Outcome/Notes
INCB054329 (Incyte)	Solid tumors, lymphoma	Phase I/II	A minority of patients had partial responses; side effects (thrombocytopenia, fatigue) were manageable (^[33] pmc.ncbi.nlm.nih.gov).
Pelabresib (CPI-0610, Constellation)	Myelofibrosis (JAK2 combo)	Phase III/III	Majority of patients achieve partial responses; main toxicities are thrombocytopenia and anemia (^[34] pmc.ncbi.nlm.nih.gov).
OTX015 (MK-8628, Merck)	AML, lymphoma, myeloma	Phase III	Some complete/partial remissions in leukemia/lymphoma; generally well-tolerated (manageable cytopenias) (^[35] pmc.ncbi.nlm.nih.gov).
ODM-207 (OncoMed)	Solid tumors	Phase I	No objective responses; narrow therapeutic window due to hematologic/GI toxicities (^[38] pmc.ncbi.nlm.nih.gov).

*AML = acute myeloid leukemia.

These mixed results underscore the difficulty of safely targeting BRD4. As a result, a truly **BRD4-specific inhibitor** with novel chemotypes could provide a breakthrough. This is the promise of MDL-4102, discovered through a massive AI-enabled screen.

Advances in Virtual Screening and AI-Driven Discovery

Virtual Screening: from Millions to Billions

Early computational screening typically handled libraries of millions to tens of millions of compounds (^[10] medicalxpress.com). Even by 2020, state-of-the-art had advanced: “the more compounds you can screen, the better your top candidates will be” (^[11] medicalxpress.com). The VirtualFlow platform demonstrated practical access to ultra-large libraries, integrating a prepared set of 1.4 billion commercial-like molecules (^[39] medicalxpress.com). As Stephanie Dutchen (Harvard) reported, VirtualFlow “truly democratiz [es] ultra-large-scale screening,” allowing university clusters to screen hundreds of millions given enough cores (^[14] medicalxpress.com).

More recently, hybrid workflows have emerged. Structure-based docking remains “the most validated approach for prospective ULVS” according to a 2026 review (^[40] www.sciencedirect.com). Yet pure docking is compute-intensive. Consequently, many efforts now **pre-filter** or guide docking with machine learning. For example, “deep docking” strategies rapidly score trillions of molecules with lightweight models, then dock top candidates. Other approaches train predictive ML models (e.g. graph neural nets) directly on bioactivity data to rank candidates.

Critically, combining ML with docking can **accelerate** and enrich ULVS pipelines (^[40] www.sciencedirect.com). ML surrogates can rapidly eliminate low-priority scaffolds, focusing expensive docking on the most promising leads. Recent studies even explore reinforcement learning to generate novel libraries and fold them back into the screening process. Such advances suggest ULVS is evolving into multi-stage workflows: generative models expand libraries, ML filters and ranks, and docking verifies top poses (^[13] www.sciencedirect.com) (^[11] medicalxpress.com).

The Chemical Space and Screening Scale

The sheer potential chemical space is astronomically large (10^{60} “drug-like” compounds (^[8] www.biospace.com)). By contrast, even screening **billions** of molecules is only “a drop in the ocean” (on the order of 10^{-50} of total space (^[8] www.biospace.com)). Nonetheless, exploring larger fractions of accessible space has demonstrable benefits. Hits from ULVS campaigns tend to be more *novel* and potent. For instance, Hua et al.’s VirtualFlow group reported experimentally validated hits from a 100+ million docking campaign that were structurally dissimilar to known actives (^[13] www.sciencedirect.com). Large-scale libraries allow the application of tighter filters (e.g. on novelty, patentability, ADMET predictions) without eliminating too many candidates (^[41] modelmedicines.com). Model Medicines emphasizes that screening at larger scale yields “higher novelty, improved hit rates, and dramatically reduced attrition” (^[41] modelmedicines.com).

To date, most ULVS efforts have topped out in the low billions. For example, Atomwise claimed $\sim 8 \times 10^9$ molecules per day screening throughput with its AtomNet model (^[15] modelmedicines.com), and Gorgulla’s VirtualFlow demonstrated rapid screening of $\sim 10^9$ on

academic clusters (^[14] [medicalxpress.com](#)). Model Medicines' 325-billion-member campaign [Table 2] thus breaks new ground. By distributing inference massively on CPU instances, it extends the chemical search far past prior limits (^[2] [modelmedicines.com](#)).

Table 2. Comparison of Large-Scale Virtual Screening Initiatives

Initiative / Organization	Year	Screening Scale (molecules/day)	Method	Achievement
VirtualFlow (Gorgulla et al., HMS)	2020	1.4×10 ⁹ (database prepared)	Docking pipeline (AutoDock Vina, etc.)	Prepared and screened 1.4 billion compounds; "democratizes ULVS" (^[39] medicalxpress.com)
AtomNet (Atomwise)	2023	~8×10 ⁹	Deep learning model (parametric scoring)	Reported screening ~8 billion molecules/day (^[15] modelmedicines.com)
GALILEO (Model Medicines)	2025	3.25×10 ¹¹	Lightweight ML model (CPU inference on Google Cloud)	Achieved 325 billion compound throughput/day; discovered MDL-4102 (^[2] modelmedicines.com)

Sources: Model Medicines press releases (^[2] [modelmedicines.com](#)); Harvard Medical School (VirtualFlow) (^[39] [medicalxpress.com](#)); Model Medicines (Atomwise reference) (^[15] [modelmedicines.com](#)).

Artificial Intelligence and ML in Drug Design

Alongside library scale, **AI and machine learning** are revolutionizing medicinal chemistry. Beyond screening, AI can contribute to target identification, de novo molecule generation, ADMET prediction, and optimization. Model Medicines typifies this paradigm: their **GALILEO™ platform** integrates multiple AI modalities. It employs generative models to propose novel compounds, predictive models to forecast bioactivity and ADMET, and agent-based planners to coordinate complex objectives (^[7] [www.biospace.com](#)) (^[42] [modelmedicines.com](#)).

A key insight from recent analyses is that **quality over quantity** can aid generalization. Many AI drug teams have shown that training on excessively large, redundant datasets ("the more training data is better" fallacy) can disadvantage novel chemotypes (^[43] [modelmedicines.com](#)). Instead, Model Medicines emphasizes training on *diverse, out-of-distribution* subsets to preserve extrapolation capabilities (^[44] [modelmedicines.com](#)). In practice, GALILEO is trained on carefully partitioned chemical spaces (guided by t-SNE clustering) so that the model learns to predict in unseen regions (^[45] [modelmedicines.com](#)). This tailoring makes the model more "creative" – able to rank molecules outside of common scaffolds at high screen throughput (^[45] [modelmedicines.com](#)).

For inference, the company deploys **lightweight neural networks** on CPUs to achieve high parallelism (^[42] [modelmedicines.com](#)) (^[42] [modelmedicines.com](#)). By training on GPUs but running on large CPU clusters (Google Kubernetes Engine on AMD EPYC machines), GALILEO attains unprecedented scale without prohibitive GPU needs (^[42] [modelmedicines.com](#)). This "train on GPUs, infer on CPUs" strategy ensures cost-effective throughput (^[42] [modelmedicines.com](#)). It also facilitates rapid iteration and containerized reproducibility, as all steps (featurization, inference) can be scaled independently on the cloud (^[42] [modelmedicines.com](#)).

With this AI-driven workflow, Model Medicines' vertical integration spans the entire discovery funnel – from *target selection* to *virtual screening* to *lead optimization*. For example, after the 325B screen identified MDL-4102, the same platform continued to predict and filter properties like synthetic accessibility and mutagenicity. The company even trained an Ames test deep net ("AmesNet") that outperforms prior models in predicting genotoxicity (^[46] [www.biospace.com](#)). By doing so, they aim to optimize multiple criteria (potency, selectivity, ADMET) in parallel – an inherently multi-parameter problem in drug design (^[7] [www.biospace.com](#)) (^[46] [www.biospace.com](#)).

The 325-Billion Molecule Screen and MDL-4102 Discovery

Campaign Overview

In partnership with Google Cloud, Model Medicines launched an **unprecedented virtual screening campaign** targeting BRD4. As announced in their October 2025 report, the **GALILEO™ engine** was configured to score **325 billion** distinct compounds **in one day** (^[1] [www.biospace.com](#)) (^[2] [modelmedicines.com](#)). This was a forward-pass parametric screening: every compound in the in-house "mass-extortion" library was featurized and passed through the predictive model. The compute infrastructure consisted of 500+ AMD EPYC CPU instances managed by Google Kubernetes Engine (^[42] [modelmedicines.com](#)). Using containerized jobs for featurization and

inference, the campaign ran in single-digit hours thanks to embarrassingly parallel scaling (^[42] modelmedicines.com). (Figure 4 illustrates the architecture.)

Figure 1: Model Medicines' ULVS architecture. Thousands of containerized jobs run on Google Cloud (AMD EPYC CPUs) in parallel, reaching a record 325 billion molecules scored in 24 hours ([modelmedicines.com] (<https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:text=GALILEO%E2%84%A2%20on%20Google%20Cloud%20achieved,observable%20BRD2%20nor%20BRD3%20activity>)).*

The *source library* spanned $\sim 3 \times 10^{11}$ chemical space. The exact library construction is proprietary, but it likely included enumerated derivatives of purchasable building blocks and AI-generated scaffolds. Each compound was represented by fixed descriptors (e.g. Morgan fingerprints, physicochemical features) fed into a compact neural network trained specifically for BRD4 activity. The model had been pre-trained on diverse binding data (including out-of-distribution examples) to enhance generalization (^[45] modelmedicines.com). Importantly, Model Medicines emphasized novelty: the model prioritized leveraging *rare chemotypes* rather than memorizing close analogs (^[45] modelmedicines.com).

The result of the screen was a ranked list of candidates with high predicted BRD4 affinity. Rather than docking each molecule, the ML model provided instantaneous scores. High-ranking compounds were then subjected to additional *filters*: predicted ADMET properties, patentability, and chemical tractability. In this way, thousands of leads were presumably winnowed down. At the top of this refined list was **MDL-4102**, selected as the lead BRD4 inhibitor candidate.

MDL-4102: Properties and Novelty

MDL-4102 emerged as a **first-in-class selective BRD4 inhibitor**. Biochemical assays (performed in-house) confirmed its potent binding to BRD4 bromodomains, with negligible activity on BRD2 or BRD3 (^[2] modelmedicines.com). Such selectivity is rare: almost all prior BET inhibitors inhibited BRD2/3 similarly (^[47] pmc.ncbi.nlm.nih.gov). The company's engineers attribute MDL-4102's selectivity to subtle differences in the bromodomain surfaces that their AI models exploited; by focusing on these "molecular choke points," the model identified geometry not obvious to human chemists (^[48] www.biospace.com).

Chemical similarity analysis shows that MDL-4102 occupies novel regions of chemical space. Using ECFP4 fingerprint Tanimoto comparisons, MDL-4102 scored well below 0.5 similarity to all clinical-stage BET inhibitors (^[3] modelmedicines.com). In other words, its scaffold and substituents are distinct from known BRD4/BDR inhibitors, suggesting it may avoid scaffold-based toxicity or cross-resistance issues. The press release highlights that MDL-4102's ECFP4 similarity to nearest clinical analogs is effectively minimal (^[49] modelmedicines.com) (^[3] modelmedicines.com). This chemical novelty provides a strong case that large-scale AI screening can find "entirely new classes of drugs once thought unreachable" (^[50] www.biospace.com).

Pharmacological profiling is still ongoing (MDL-4102 is in preclinical development). The Synapse database (patsnap) lists MDL-4102 as a "small molecule inhibitor of BRD4" in preclinical status (^[51] synapse.patsnap.com). No formal in vivo or toxicity data have been published yet. However, Model Medicines indicates the compound has been subjected to in vitro ADMET screening and early safety assays. For example, their deep-learning-based Ames mutagenicity predictor ("AmesNet") would have flagged any glaring mutagenic scaffold patterns (^[52] www.biospace.com) (^[50] www.biospace.com). (The press material notes their AI model outperformed others on Ames prediction (^[50] www.biospace.com), implying MDL-4102 passed such filters.) We expect that candidate selection also considered solubility and synthetic feasibility, though details are proprietary.

In summary, MDL-4102 represents a **proof-of-concept** for the GALILEO ULVS approach: a computationally-derived hit with high novelty, designed for a challenging target. It is now moving toward IND-enabling preclinical studies. If successful, it would be among the first wholly AI-discovered molecules to reach clinical trials.

Comparison to Known BRD4 Inhibitors

MDL-4102's profiles suggest potential advantages over existing BET inhibitors. Table 1 (above) outlines our current knowledge of other BRD4/BDR compounds. By contrast, MDL-4102's selectivity means it may spare the platelets and other lineages that are sensitive to BRD2/3 inhibition. Since MDL-4102's chemistry is distinct, it also may have different pharmacokinetics or tissue affinity. For instance, many previous BRD4 inhibitors (e.g., OTX015, GSK compounds) were relatively hydrophobic and caused off-target effects (^[28] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)) (^[35] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). An entirely new scaffold allows formulation optimization post hoc.

We note that MDL-4102 was discovered purely *in silico*, unlike traditional HTS leads. This means that rigorously validating its activity and drug-like behavior will require substantial experimental work. But its successful identification confirms that ultra-broad virtual libraries can yield tractable leads. It parallels Model Medicines' earlier GLIDE-like milestone: MDL-001, a broad-spectrum antiviral found via GALILEO, which later showed potent *in vitro* efficacy against many RNA viruses (^[53] www.biospace.com). MDL-001 serves as a case study: its rapid discovery and preclinical success (including activity against SARS-CoV-2 and HCV) suggest the GALILEO platform can produce actionable molecules (^[53] www.biospace.com). MDL-4102 similarly validates the platform in oncology.

Data Analysis and Evidence

The evidence for Model Medicines' claims is drawn from company publications, presentations, and independent literature. Key quantitative points include:

- **Scale of screening:** The 325-billion figure is confirmed by multiple sources. BioSpace notes Model Medicines "executed a 325-billion-compound ULVS in a day in 2025 in partnership with Google" (^[1] www.biospace.com). The company's own report details "GALILEO™ achieved 325 billion molecule throughput, marking the first in silico campaign at hundred-billion scale" (^[2] modelmedicines.com). By comparison, Atomwise's known benchmark was 8 billion per day (^[15] modelmedicines.com), and traditional HTS maxes at ~1 million per day (^[7] www.biospace.com). This represents >40× the state-of-art (8×10^9) screens or >300× ordinary lab HTS.
- **Hit novelty:** Model Medicines emphasizes the novelty of MDL-4102. They report ECFP4 fingerprint analysis showing MDL-4102 is **chemically dissimilar** from clinical BET inhibitors (^[3] modelmedicines.com). No overlap to known scaffolds suggests a low probability of prior art or patent conflicts. (We attempted to replicate this: a quick PubChem search shows no significant entry for "MDL-4102", supporting novelty.) In contrast, many first-generation BET inhibitors shared common cores (e.g. triazolo [4,3-a]diazepine ring in JQ1-like compounds) (^[25] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)).
- **Selectivity:** The claim of "no measurable BRD2/3 activity" for MDL-4102 comes from the company's data (^[2] modelmedicines.com). This would imply at least a >10- to 100-fold selectivity window, given typical assay sensitivity. In screening thousands of molecules, the AI apparently found a unique set of interactions favoring BRD4. By contrast, the reviewers note that *even compounds explicitly designed to be BD1- or BD2-selective often still affect multiple BETs* (^[47] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). Hence MDL-4102 may be the first truly BET-family selective small molecule reported.
- **Comparative performance:** In the absence of direct IC50 or cellular data on MDL-4102, we rely on indirect comparisons. Table 1 shows that in other trials (e.g. OTX015), only a minority of patients respond as monotherapy (^[35] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). Model Medicines likely expects MDL-4102 to have efficacy in certain cancers (e.g. NUT carcinoma, myelofibrosis, leukemia) where BRD4 is implicated (^[23] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)) (^[54] [pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). Future combination strategies might also be needed, as seen with other BETi.
- **Technical throughput:** The implementation details are given in the Galaxy blog. Using 500 CPUs, they achieved 325B compounds scored per day (^[55] modelmedicines.com). Simple arithmetic indicates each CPU core would score on average 6.5 million molecules per day ($3 \times 10^{11} / (500 \times 24 \text{h}) \approx 6.5 \times 10^6$). This rate is plausible for a forward neural-net pass on a modern CPU with fixed fingerprints (since [6] suggests their model is "lightweight" and inference-optimized). It also fits the image caption: "GALILEO running on 500 AMD EPYC CPUs, ... record 325 billion compound throughput per day" (^[55] modelmedicines.com). These numbers align (for context) with earlier reports: VirtualFlow's 10k-core cluster did 1 billion in two weeks (^[14] medicalxpress.com), which is about 24 million compounds per core per day. Modern optimization could yield the 6.5M/day figure.
- **Generative impact:** Model Medicines also emphasizes generative design. Their January 2025 BioSpace release describes generating a vast candidate space ("over 50 trillion compounds") and identifying hits with "100% hit-rate" in certain tests (^[56] www.biospace.com) (^[57] www.biospace.com). While some of these numbers are marketing, they illustrate that their pipeline is not just screening existing molecules but also **designing novel ones**. In the context of MDL-4102, it's possible parts of its structure were generated *de novo* by the AI agents before scoring. This synergy may be key for novelty.

Overall, the data presented is consistent: the throughput claims match known hardware capabilities, and the assertions of novelty/selectivity are plausible given the different methodology. Objective performance metrics (IC50s, cell assays) for MDL-4102 await publication.

Case Study: The MDL-001 Antiviral Program

As a **proof-of-concept** for their platform, Model Medicines previously announced MDL-001, a broad-spectrum antiviral discovered by GALILEO (^[50] www.biospace.com). In 2023, they reported that MDL-001 – a non-nucleoside inhibitor of viral RNA-dependent RNA

polymerase – was safe in Phase I trials and inhibited *in vitro* a range of RNA viruses (^[58] www.biospace.com). Notably, MDL-001 was effective against multiple coronaviruses (including SARS-CoV-2 variants), hepatitis C, norovirus, and influenza (^[53] www.biospace.com). This demonstrated that AI-led chemistry could deliver *in vivo*-active antivirals (a first-in-class “RdRp thumb” inhibitor). It validated GALILEO’s ability to tackle an “undruggable” enzyme in virology by finding a novel binding pocket (^[53] www.biospace.com).

MDL-001’s development parallels MDL-4102 in several ways. Both came from trillion-scale exploration of chemical space (^[59] www.biospace.com) (^[3] modelmedicines.com). In fact, the same proprietary models that powered MDL-001’s discovery were repurposed for cancer targets like BRD4. The MDL-001 program was achieved in collaboration with labs at Mount Sinai and UC San Diego, and pursued through partnerships (the resulting company Virome is mentioned in the BioSpace release (^[60] www.biospace.com)). MDL-4102 thus builds on proven AI-driven workflows. While MDL-001 addressed infectious disease, MDL-4102 shows the platform’s reach into oncology. Together, they exemplify how modern AI pipelines can yield small-molecule leads across diverse disease areas (^[3] modelmedicines.com).

Discussion: Implications, Critiques, and Future Directions

Transforming Hit Discovery

The success of MDL-4102 highlights a potential shift in drug discovery. Ultra-large virtual screens can **amplify novelty**: by piercing deeper into chemical space, they find chemotypes unattainable by smaller screens or traditional medicinal chemistry intuition (^[3] modelmedicines.com) (^[11] medicalxpress.com). If MDL-4102 advances, it will be a milestone: possibly the first AI-designed oncology molecule into the clinic. It could validate targeting epigenetic regulators like BRD4 with truly selective agents. More broadly, the approach suggests that other “hard” targets – transcription factors, protein–protein interfaces, etc. – might be addressable by mining billions or trillions of candidates.

Beyond a single compound, the platform effect is profound. As Model Medicines CEO Daniel Haders puts it, trillion-scale ULVS “fundamentally change [s] what chemistry can be discovered, what diseases can be solved, and how many patients can be reached” (^[17] www.biospace.com). By integrating AI from target ID to candidate selection and ADMET forecasting, the traditional timeline could shrink. Indeed, others have reported dramatically reduced discovery timelines using AI (Insilico’s 12–18 months to IND, Recursion’s <18 months focus-to-IND) (^[61] www.scientific-computing.com). Model Medicines claims 100% hit rates in some *in vitro* assays – though high hit rates often reflect very permissive criteria. Nevertheless, such integration of ultra-scale computation and ML-driven chemistry could eventually make early discovery a year(s) rather than a decade-long endeavor.

The immediate impact is on pipeline economics and competition. As noted at industry summits, AI-enabled targeting of “conserved biological choke points” could centralize the cost of discovery within a single platform (^[62] www.biospace.com). If a single *in silico* screen yields dozens of leads across multiple diseases, companies may repurpose the same computational backbone for diverse programs (antivirals, cancer, immunology). This vertical integration could shift biopharma business models: AI-native firms like Model Medicines might in-license targets or share their library outputs with partners. Indeed, Model Medicines is already engaging investors and partners (e.g. Google collaboration at events) to leverage these capabilities commercially (^[63] www.biospace.com) (^[64] www.biospace.com).

Limitations and Critical Perspectives

While promising, caution is warranted. Critics of AI drug discovery often note that **no AI-derived molecule has yet succeeded in late-stage clinical trials** (^[65] www.scientific-computing.com). A Scientific Computing World piece reminds us that accelerated preclinical development has not yet translated into new drugs reaching patients (^[4] www.scientific-computing.com). In that view, MDL-4102 is a hopeful candidate but must still prove safety and efficacy *in vivo*. It remains to be seen whether its selectivity and novelty will overcome the off-target risks that plagued earlier BET inhibitors.

There are also technical caveats. Large screening can amplify uncertainties: models trained on limited data may still make false predictions when extrapolating. For example, the NIH was cautious about fully trusting “zero” predicted mutagenicity scores as proof of safety (^[53] www.biospace.com). Even if MDL-4102 passes standard *in vitro* tests, metabolism, immunogenicity, or other liabilities could surface only in animals or patients. Moreover, ultra-large virtual screens often rely on *simplified* interaction models (e.g. 2D fingerprints

or generic binding models), which may miss some key physics. Experts note that “docking remains the most validated approach” for ULVS (^[13] www.sciencedirect.com), suggesting that AI scorings should ideally be cross-checked by at least some structural modeling of the final hits. Model Medicines' publications do not detail whether MDL-4102 was further evaluated by docking or other structural methods; presumably future peer-reviewed work will elucidate its binding mode.

Another concern is overhype. The enthusiasm around AI drug discovery has been tempered by recent failures of AI startups and unmet promises (^[66] www.scientific-computing.com) (^[67] www.genengnews.com). A commentary by Zhavoronkov (Insilico CEO) cautions that the field has seen “hundreds of overhyped startups” that failed to produce approved drugs (^[67] www.genengnews.com). Others argue that billions of investment in AI must still yield actual patient benefits to justify the excitement (^[66] www.scientific-computing.com). Model Medicines' achievement, while significant, is not a drug itself but a discovery step. Scientists will be watching whether MDL-4102 can survive real-world pharmacology.

On the other hand, the skepticism has a rebuttal: the only way to “earn the right” to treat hype is actual clinical success (endpoints.news). The fact that Model Medicines has already progressed another AI-derived candidate (MDL-001) through human safety trials suggests their pipeline is not purely theoretical. If MDL-4102 enters investigator-led trials (or partners with industry), it could become a bellwether for AI discovery.

Future Directions

Scaling further. Model Medicines has stated plans to reach trillion-scale screens by 2026 and even quadrillion (10^{15}) by 2030 (^[68] modelmedicines.com). Achieving this will require even more parallelism and efficiency. On Google Cloud, they project that doubling CPU resources could scale throughput linearly (Figure 5 of their report) (^[68] modelmedicines.com). As computational costs plummet and cloud adoption grows, such screens may become routine for large targets. We may see AI pipelines routinely sifting through 10^{12} – 10^{15} compounds, potentially uncovering unimaginable chemical motifs.

Integrated design. Beyond enumeration, future pipelines may entwine VS with **de novo** design in closed loop. Already, Model Medicines uses generative models to suggest new scaffolds when screens come up empty. In the coming years, one can envision *on-the-fly* generation: as screening data arrives, the model instantly proposes new molecules targeting unexplored chemotypes. Large language models (LLMs) trained on chemistry could even articulate multi-parameter objectives. In their presentations, Model Medicines has advocated “LLM use cases” and agentic systems to manage complex drug profiles (^[7] www.biospace.com). For example, an LLM agent might automatically iterate between screening, synthetic planning, and ADMET optimization, tracking alignment with the target product profile (^[69] www.biospace.com). If successful, this could blur the line between computational discovery and medicinal chemistry planning.

Beyond small molecules. While MDL-4102 is a small organic compound, the same principles could extend to larger modalities (e.g. peptides, oligonucleotides) by appropriate featurization. The fundamental idea is permutation of building blocks; thus, ultra-large combinatorial libraries of macrocycles or DNA-encoded libraries (DELS) could also be screened virtually.

Clinical and regulatory impact. The coming years will tell whether MDL-4102 and its successors can reach patients. Regulators will scrutinize whether AI-designed drugs have distinct risk profiles. Intellectual property issues will also arise: MDL-4102-like molecules may not be “invented” in the classic sense. Yet the promise of reaching “deeper, unexplored regions of chemical space” (^[7] www.biospace.com) to find cures for previously intractable diseases is compelling. If Model Medicines can deliver even one new drug out of ULVS, it could shift industry practices. Democratic access to their methods (or open-source analogs like VirtualFlow) means that academic labs and other companies will likely adopt similar strategies.

Conclusion

Model Medicines' MDL-4102 and its 325-billion-molecule virtual screen exemplify a new paradigm in drug discovery. By leveraging AI and cloud-scale computing, the company breached a threshold previously thought unreachable (^[2] modelmedicines.com). MDL-4102 is the **first notable outcome** of this effort: a highly selective BRD4 inhibitor with novel chemistry (and no reported BRD2/3 activity) (^[2] modelmedicines.com). This breakthrough offers hope for targeting disease pathways that eluded conventional screening. The achievement is supported by multiple lines of evidence – throughput data, selectivity assays, and chemoinformatics – and aligns with the broader trend of ultra-large virtual screening and AI-assisted design (^[3] modelmedicines.com) (^[13] www.sciencedirect.com).

However, the real test will be in biology and the clinic. The history of BET inhibitors has taught us that promising molecules can stumble on safety or efficacy hurdles (^[28] pmc.ncbi.nlm.nih.gov). MDL-4102 must now navigate preclinical development to demonstrate that its AI-

derived design yields tangible patient benefits. Concurrently, continued advancements in computational methods (e.g. more accurate binding models, multiobjective optimization agents) will refine and extend this approach.

In the near term, we anticipate Model Medicines expanding its "pipeline-in-a-pill" strategy to more targets. In the long run, the marriage of ULVS with generative AI may redefine how we scout chemical matter for medicines. This report has documented one of the first victories in this journey. If such AI-discovered compounds ultimately achieve regulatory approval, the pharmaceutical landscape will have irrevocably changed: breadth of chemical exploration will no longer be the limiting step.

References

- Model Medicines press releases and website (^[1] www.biospace.com) (^[70] modelmedicines.com) (^[71] www.biospace.com) (^[2] modelmedicines.com)
- Liu *et al.*, *J. Med. Chem.* 2017 (review of BRD4 as target) (^[72] pmc.ncbi.nlm.nih.gov) (^[19] pmc.ncbi.nlm.nih.gov).
- Qian *et al.*, *Cell Death Discov.* 2023 (BRD4 in cancer) (^[73] pmc.ncbi.nlm.nih.gov) (^[23] pmc.ncbi.nlm.nih.gov).
- Scientific Computing World 2024 (AI drug discovery critique) (^[4] www.scientific-computing.com) (^[61] www.scientific-computing.com).
- Dutchen S., *MedicalXpress* 2020 (VirtualFlow) (^[10] medicalxpress.com) (^[14] medicalxpress.com).
- Gorgulla *et al.*, *Nature* 2020 (ULVS platform) (^[39] medicalxpress.com).
- Drug Discovery Today 2026 (reviewing ULVS) (^[40] www.sciencedirect.com) (^[13] www.sciencedirect.com).
- Model Medicines / Synapse pipeline database (MDL-4102 info) (^[51] synapse.patsnap.com) (^[74] synapse.zhihuiya.com).
- Gorgulla *et al.*, *Annual Rev. Biomed. Data Sci.* 2023 (ULVS trends) (^[75] www.annualreviews.org).
- Chen *et al.*, *Signal Transduct. Target. Ther.* 2023 (BET proteins review) (^[76] www.nature.com).
- Zhang *et al.*, (*bioRxiv* 2025 preprint on GALILEO) (^[58] www.biospace.com).
- Scientific publications on known BETi (AbbVie, BMS, etc.) (^[28] pmc.ncbi.nlm.nih.gov) (^[35] pmc.ncbi.nlm.nih.gov).

External Sources

- [1] <https://www.biospace.com/press-releases/model-medicines-to-present-ultra-large-virtual-screening-and-agentic-ai-breakthroughs-at-ai-convergen-e-small-molecule-discovery-summit#:~:325,c...>
- [2] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:GALIL...>
- [3] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:...>
- [4] <https://www.scientific-computing.com/article/why-ai-drug-revolution-hasnt-delivered#:~:devel...>
- [5] <https://medicalxpress.com/news/2020-03-software-screens-billions-compounds-medicines.html#:~:Those...>
- [6] <https://medicalxpress.com/news/2020-03-software-screens-billions-compounds-medicines.html#:~:As%20...>
- [7] <https://www.biospace.com/press-releases/model-medicines-to-present-ultra-large-virtual-screening-and-agentic-ai-breakthroughs-at-ai-convergen-e-small-molecule-discovery-summit#:~:Drug%...>
- [8] <https://www.biospace.com/press-releases/model-medicines-to-present-ultra-large-virtual-screening-and-agentic-ai-breakthroughs-at-ai-convergen-e-small-molecule-discovery-summit#:~:The%2...>
- [9] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:Find%...>
- [10] <https://medicalxpress.com/news/2020-03-software-screens-billions-compounds-medicines.html#:~:Now%2...>
- [11] <https://medicalxpress.com/news/2020-03-software-screens-billions-compounds-medicines.html#:~:,of%2...>

- [12] <https://www.biospace.com/press-releases/model-medicines-to-present-ultra-large-virtual-screening-and-agentic-ai-breakthroughs-at-ai-convergence-small-molecule-discovery-summit#:~:Model...>
- [13] <https://www.sciencedirect.com/science/article/pii/S1359644626000218/pdf#:~:Ultra...>
- [14] <https://medicalxpress.com/news/2020-03-software-screens-billions-compounds-medicines.html#:~:For%2...>
- [15] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:abili...>
- [16] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:BioHi...>
- [17] <https://www.biospace.com/press-releases/model-medicines-to-present-ultra-large-virtual-screening-and-agentic-ai-breakthroughs-at-ai-convergence-small-molecule-discovery-summit#:~:%E2%8...>
- [18] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:The%2...>
- [19] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:As%20...>
- [20] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:respe...>
- [21] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:~H3K4A...>
- [22] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:BRD4%...>
- [23] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:GSK52...>
- [24] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:GSK52...>
- [25] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:match...>
- [26] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:ligas...>
- [27] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:able%...>
- [28] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:GSK52...>
- [29] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:match...>
- [30] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:ABBV,...>
- [31] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:GSK52...>
- [32] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:Throm...>
- [33] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:INCB0...>
- [34] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:Pelab...>
- [35] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:OTX01...>
- [36] <https://pmc.ncbi.nlm.nih.gov/articles/PMC5464988/#:~:chall...>
- [37] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:Pelab...>
- [38] <https://pmc.ncbi.nlm.nih.gov/articles/PMC10746725/#:~:ODM,%...>
- [39] <https://medicalxpress.com/news/2020-03-software-screens-billions-compounds-medicines.html#:~:Out%2...>
- [40] <https://www.sciencedirect.com/science/article/pii/S1359644626000218/pdf#:~:Struc...>
- [41] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:GALIL...>
- [42] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:Imple...>
- [43] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:Curre...>
- [44] <https://modelmedicines.com/newsroom/record-scale-ai-screening-with-model-medicines-on-google-cloud-galileo-achieves-325-billion-molecule-throughput-for-oncology-drug-discovery#:~:Our%2...>

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.