molecule.one

Case Study

Deep-Learning-Guided Chemical Space Enables Rapid, Single-Iteration Hit Identification

Challenge

Small-molecule synthesis remains a central bottleneck in early drug discovery. Outsourced synthesis through CROs is costly, in-house chemistry often fails to scale, and commercially available or “on-demand” libraries frequently lack structural novelty or direct relevance to specific biological hypotheses. As a result, turning ideas into experimentally validated hits takes far too long.

Molecule.one’s solution

Using high-throughput experimentation (HTE) data as foundation, we built D2B-SpaceM1: the first generative, deep-learning chemical space trained on more than 300,000 experimentally observed reaction outcomes derived from Molecule.one’s MARIA™ platform. Unlike abstract virtual libraries, D2B-SpaceM1 is explicitly feasibility-anchored: every generated molecule is associated with a concrete, data-supported synthetic route, enabling immediate translation from *in silico* prioritization (AI predictive models, docking) to rapid direct-to-biology synthesis and testing experimental synthesis.

Notably, MARIA™ can operate at plate scale, enabling up to ~13,000 reactions per week with potential for further scaling. In addition, our synthesis models support uncommon yet effective conditions, such as Suzuki couplings in DMSO. This significantly broadens the range of makeable chemistry beyond standard constraints.

Case study #1 - CLK1

CLK1 is a protein kinase that regulates RNA processing inside cells, making it an emerging therapeutic target in cancer and neurodegenerative diseases.

Virtual Hit Prioritization

39 compounds were prioritized from D2B-SpaceM1 in collaboration with Inventro using their proprietary machine-learning models. An additional 18 compounds were selected from a suggested list of BB-derived analogs accessible through available building-block combinations and were independently confirmed as potentially active by Inventro's AI Predictive Models. Finally, 37 further BB-derived analogs were added to expand the set.

D2B Synthesis

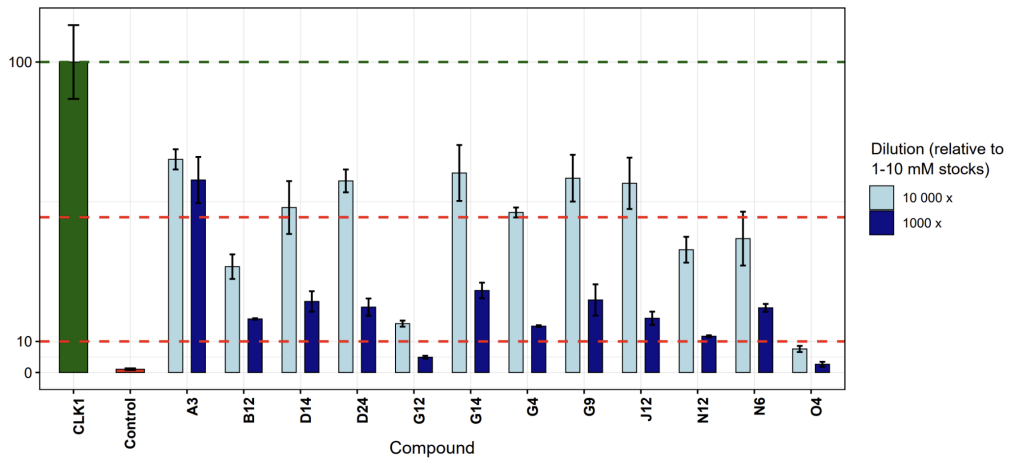
Selected molecules were advanced directly into microliter scale, plate-based synthesis, followed by minimal work-up and immediate biological testing. Across the CLK1 D2B grid ~93% of targeted products were successfully formed (87 out of 94), underscoring the robustness of feasibility-aware design at plate scale.

Primary and Secondary Screening (ADP-Glo)

Biological activity was assessed using the ADP-Glo™ luminescent kinase assay.

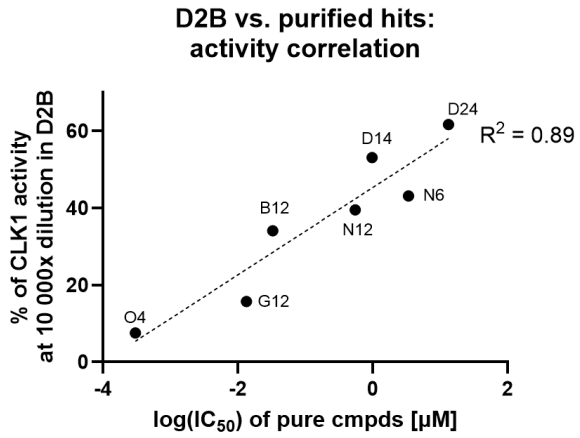
Primary screen: single-point measurement at 100× dilution identified 12 active compounds with <10% residual CLK1 activity.

Secondary screen: additional measurements at 1,000× and 10,000× dilutions confirmed 11 active compounds, corresponding to a hit rate of 18% for AI-prioritized compounds (12% across the full grid). ADP-Glo activity (% CLK1 residual activity) across 12 preselected D2B wells is depicted below.



Hit Confirmation

Based on activity and structural diversity, 7 compounds were selected for resynthesis, purification, and confirmatory testing. 7/7 purified compounds confirmed active in dose–response assays. Potency distribution spanned from µM to sub-nanomolar range.

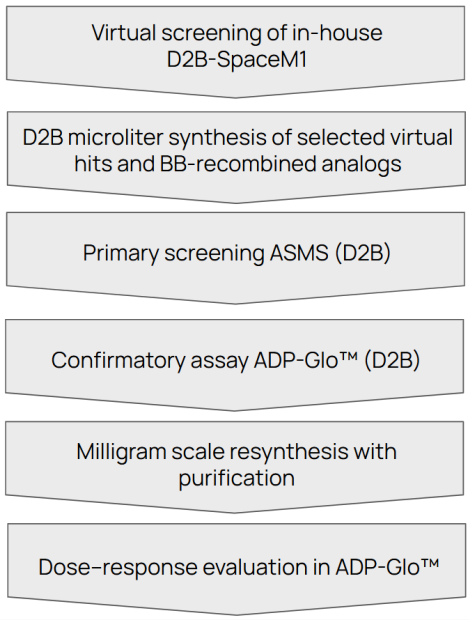


A strong correlation (R^2 = **0.89**) was observed between CLK1 inhibition measured directly from D2B crudes and IC_{50} values determined for purified compounds. This demonstrates that crude D2B data are sufficiently predictive to support confident hit selection without iterative purification cycles.

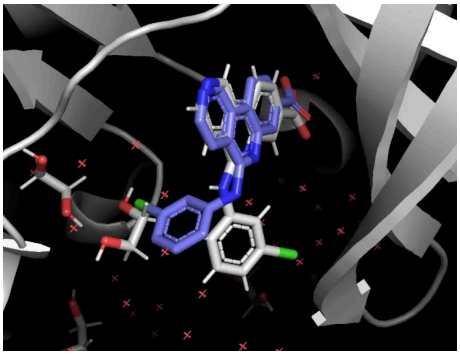
The most potent compound exhibiting sub-nanomolar activity is an analog suggested based on available building-block combinations (confirmed by the ML model). HitID was completed within 4 weeks.

Case study #2 - CK2

CK2 is a protein kinase implicated in cell survival and stress-response pathways, making it a relevant target in oncology and inflammation. The enzyme possesses a highly conserved ATP-binding pocket, which constitutes the main catalytic site and has been extensively explored for the development of small-molecule inhibitors.



Virtual Screening and D2B Execution



For CK2, candidates were selected from D2B-SpaceM1

using consensus docking (AutoDock Vina + Boltz2) combined with deliberate diversity sampling. Docking-based virtual screening yielded a set of virtual hits. For their synthesis, a defined set of building blocks (BBs) was required, from which additional BB-recombined analogs were subsequently designed.

The resulting grid was advanced directly into D2B plate synthesis in microliter scale, followed by minimal work-up. Out of 1,152 targeted reactions, 887 products were successfully formed, corresponding to a ~77% synthesis success rate, in close agreement with ML-predicted feasibility.

ASMS Screening

When exploring large compound collections, early screening methods must balance throughput, cost, and sensitivity. Affinity-selection mass spectrometry (ASMS) is well suited for this purpose, enabling direct detection of protein–ligand interactions without labeling or compound purification. Importantly, ASMS aligns well with D2B workflows, as it supports screening of pooled compounds directly from crude reaction mixtures. Accordingly, the D2B product set was first evaluated using ASMS. 10 CK2 binders were identified by ASMS, corresponding to a hit rate of ~1.1%, exceeding typical ASMS benchmarks. Notably, 90% of the ASMS hits were BB-recombined analogs generated by reusing building blocks originally selected for the virtual hit set, with 8 out of 10 being novel structures.

Confirmatory assay

The orthogonal biochemical ADP-Glo™ luminescent kinase assay was used to confirm the activity of selected hits. Ten compounds were initially tested at a single concentration using D2B crude reaction mixtures with Silmitasertib as positive control. 3 out of 10 ASMS hits exhibited inhibitory activity as D2B post reaction crudes activity and were therefore resynthesized and purified at milligram scale. Dose–response evaluation was subsequently done, revealing one compound with an IC_{50} of approximately 200 nM and two compounds with IC_{50} values below 80 nM. Exact potency values will be determined upon retesting at lower concentrations;

nevertheless, the most potent compounds reduced enzyme activity to below 5% at an 80 nM concentration and one structure remains highly novel.

	CLK1	CK2
Hit selection	ML predictive model by Inventro	Consensus docking with diversity sampling
Synthesis success rate	93 %	77 %
Hit rate	18 % (AI-prioritized), 12 % (full grid)	ASMS: 1.1% (ASMS typical ~0.4% to ~1%)
Potency of hits	Sub-nM top, followed by low µM activity	3 hits with IC_{50} in nM range
Compounds made (in one week)	~100	~1000

Collectively, the data show that our chemistry—diverse, feasibility-anchored SpaceM1 and rapid HTE execution—streamlines DMTA cycles and cuts resource use in early discovery allowing for hit identification within weeks not months, while requiring minimal FTE involvement. We repeatedly observe that close analogs often achieve higher potency than the originally designed molecules. This effect reflects what can be described as engineered serendipity: while computational methods are inherently approximate, systematic analog exploration—enabled by low-cost, plate-scale synthesis—provides tolerance to design error and increases the probability of discovering high-quality hits. As a result, this approach is particularly well suited for challenging or poorly characterized targets.