Prospective Validation of HydraScreen: Virtual Screening and Hit Identification in IRAK1

Gintautas Kamuntavičius[1], Alvaro Prat[1], Tanya Paquet[1], Orestis Bastas[1], Hisham Abdel Aty[1], Quing Sun[2], Dan Rines[2], Marc Siladi[2], Roy Tal[1*], Povilas Norvaišas[1]

[1*]AI Chemistry, Ro5, 2801 Gateway Drive, Irving, 75063, TX, USA.
[2*]Strateos, 3565 Haven Ave Suite 3, Menlo Park, 94025, CA, USA.

*Corresponding author(s). E-mail(s): rtal@ro5.ai;
Contributing authors: gkamuntavicius@ro5.ai; aprat@ro5.ai; tpaquet@ro5.ai; obastas@ro5.ai; habdel-aty@ro5.ai; quing.sun@strateos.com; dan.rines@strateos.com; marc.siladi@strateos.com; pnorvaisas@ro5.ai;

Abstract

In this proof-of-concept study, we combine Ro5's digital drug discovery platforms *SpectraView* and *HydraScreen* with Strateos' robotic cloud labs capabilities to augment and accelerate target and hit identification. Using SpectraView to select IRAK1 as the target, we prospectively validate HydraScreen, a structure-based deep learning model. We demonstrate that HydraScreen could identify up to 23.8% of all hit compounds by screening only 1% of the compound library, simultaneously identifying the three of the most potent (nanomolar) scaffolds present in the library. All three nanomolar scaffolds identified in our project are novel for IRAK1 and lend themselves for future development. HydraScreen outperforms traditional methods in an unbiased prospective evaluation and offers advanced features such as ligand pose confidence scoring. Thus, SpectraView and HydraScreen are innovative tools which can aid and expedite the stages of early drug discovery.

Keywords: Artificial intelligence, machine learning, deep learning, drug discovery, SBDD, high-throughput screening, automated labs, IRAK1

1 Introduction

Drug discovery is a notoriously lengthy, expensive and inefficient process [10]. In recent years many of drug discovery's major challenges and bottlenecks were tackled by the digital transformation of pharma's legacy workflows [42]. Some of the most well known forms of digital transformation are the application of artificial intelligence (AI) [40, 6], lab automation [39, 38] and implementation of Big Data management systems [54]. These approaches hold the promise of reinventing drug discovery to usher in the new, digital era in pharma [42].

Artificial intelligence encompasses machine learning (ML) deep learning (DL) techniques, which have made a substantial impact across a range of applications in drug discovery including, protein-ligand activity prediction [24, 47], ADMET prediction [13, 52] and other areas. Traditional computation chemistry techniques such as docking [27], molecular dynamics [19, 2], and QSAR models [13] are either being augmented (e.g. DL-based docking [32], DL force fields [45]) or replaced by promising new AI techniques (e.g. Diffdock [9]). AI methods can offer previously unavailable solutions and perform better than traditional techniques [46, 6].

AI models require large volumes of high-quality data for their training [54]. The optimal setting for their application is therefore provided by automated robotic labs which can generate highly-reproducible and consistent data [1, 18, 38]. Such laboratories enable the full automation of the design-make-test-analyze (DMTA) cycle [35]. Some examples of automated DMTA have shown promise in protein engineering [36], synthetic biology [15], and chemical synthesis [14]. The coupling of automated data generation and AI-driven experimentation may introduce new paradigms of automated experimentation with no reliance on "man-in-the-loop" [36].

The large volume of biomedical data generated by high-throughput automated labs, as well as its increasing availability in the public domain [30] are creating new challenges. They are referred to as the "four Vs" of big data for drug discovery: volume (scale of data), velocity (growth of data), variety (diversity of sources), and veracity (uncertainty of data) [54]. Increasingly complex and voluminous data streams can no longer be managed manually, or require specialist knowledge. New data management and analysis systems have emerged that allow researchers to easily integrate, summarize and search biomedical data sets for hypothesis generation. Examples of such systems include knowledge graphs [53] and target identification and evaluation platforms [5]. The full potential of recent advances in AI and lab automation, coupled with the availability of high volumes of data may be fully realized by exploring their synergy.

Early drug discovery provides a perfect example of the challenges associated with such digital transformation. Target identification remains a manual process driven by specialized domain-knowledge [26], while traditional high-throughput screening (HTS) methods for hit identification rely on slow and costly unguided experimentation [55]. Previously, it was shown that the application of ML techniques could significantly accelerate HTS [11]. Such studies may be biased due to their retrospective nature and reliance on publicly available data to perform ML model benchmarking. Moreover, they tend to fall short in investigating the impact of model predictions on the final

selection of hits and hit series. In comparison, the prospective studies [55, 56], especially with AI methods [34], are few and typically do not offer comparison of different methods due to high experimental hit validation costs. The combination of these two paradigms is rare. Ultimately, the steps of target evaluation, hit identification and hit prioritization should be evaluated in the real-world context of a drug discovery program.

In this study, we demonstrate an augmented early drug discovery workflow, which integrates Ro5's target evaluation (*SpectraView*) and AI-driven virtual screening (*HydraScreen*) [**?**] tools alongside Strateos robotic cloud labs. Our goal is to show how the augmented drug discovery workflow can lead to data-driven decision-making in target evaluation and optimization of high-throughput screening. We perform data-driven target evaluation and perform prospective validation of the Ro5's structure-based deep learning model HydraScreen in an hit identification study. We evaluate HydraScreen against traditional techniques, such as molecular docking, QSAR models, and shape-based methods, to provide more context for its performance and a fair comparative analysis. Finally, we describe hits in the context of their further optimization.

2 Methods

2.1 Target Evaluation using SpectraView

Target selection and evaluation was performed using Ro5's proprietary target evaluation tool *SpectraView*. This tool allows data driven evaluation of prospective protein targets in drug discovery projects by drawing relevant contextual information from Ro5's Knowledge Graph. The evaluation criteria encompass both scientific (e.g. biological, chemical) and business (e.g. novelty, competition) considerations. The evaluation follows criteria commonly encountered in the drug discovery projects that are presented as questions raised by the researchers. Results from these queries are presented as interactive visualizations that allow exploration of different criteria for thorough target evaluation.

Ro5's Knowledge Graph consists of 4 main components:

- ontologies - databases of entities with unique identifiers (e.g. Ensemble, HGNC),
- unstructured (textual) data - > 34M PubMed abstracts and >90M patents from which entities and their relationships are extracted
- structured (database) data - relational databases with contextual information for each entity type (20+ databases, 8 entity types, >2M annotated compounds, >1Bn compound library, >500k annotated gene/protein targets, >300k 3D structures, >20k diseases, >15M assay results, >250k clinical trials)
- metadata and metrics - data origin metadata and custom metrics for data science analytics.

As such, theKnowledge Graph presents a comprehensive data resource for target evaluation.

Ro5 Knowledge Graph encompasses 12 entity types (Disease, Target, Mechanism, Compound, Species, Anatomical location, Cell line, Biomarker, Publication, Patent/Application, Author, Organization). Entity-to-entity edges are extracted for all of

these entity pairs. Additionally, full contextual information is preserved when processing textual information by introducing a Publication entity. As a result, conditional queries can be formulated for all combinations of extracted entities (e.g. Target - Diseases in the context of a Mechanism in a given publication). For all of the textual information, extensive metadata is preserved, including journal, author and affiliation information. This information is then used to enrich the Knowledge Graph with the corresponding entities (e.g. Author, Organization) which can be used business-oriented analyses. Finally, the Knowledge Graph is populated with metrics that allow quantitative evaluation of the graph structure and entity dynamics (e.g. network connectivity, edge emergence, point-wise mutual information, etc.). Altogether, such detailed representation of entities and their relationships present in the biomedical literature and patents allow an in-depth and up to date analysis of various scientific queries for the drug discovery purposes.

2.2 Strateos Cloud Lab

All of the *in vitro* experiments were performed at the Strateos Cloud Lab. The Strateos Cloud Lab comprises a collection of online software applications that integrate Strateos' automated chemistry and biology workstations, inventory management, data generation and data management. All experiments are coded in autoprotocol, an open-source standard develop by Strateos (www.autoprotocol.org), which coordinates instrument actions in specific work cells based on scientific intent. This platform allows scientists to configure experiments and experimental parameters, remotely initiate and oversee automated experiments, oversee protocol management and inventory, generate data, and access real-time outputs of experimental data in a closed loop fashion.

2.3 47k diversity library

A diverse library of 46743 compounds was employed as the primary screening resource. This compilation, chosen from a broader pool of 500,000 compounds through cheminformatics evaluation, with properties like scaffold diversity, good chemical quality, and favorable physicochemical attributes. Compounds prone to interference were systematically removed, aligning with the exclusion of Pan Assay Interference Compounds (PAINS) from screening libraries. Compounds stocks were stored at 10mM in DMSO. For screening compounds were arrayed 50 μl/well in Echo-qualified 384 well polypropylene microplates at room temperature for frequent use.

2.4 Strateos library ligand preparation & stereoisomer treatment

Strateos library compound SMILES were sanitized by removing salts and converting them into a canonical form. Stereoisomers of the same compound were treated as different ligands. For compounds with ≤ 4 unidentified stereocenters, all possible stereoisomers (up to 16) were generated and stored for further evaluation. For compounds with > 4 unidentified stereocenters, a random subset of 16 stereoisomers was generated and stored. This was done as a solution to exponential blowup that would occur for e.g. triterpenoids with > 10 unidentified stereocenters. This enumeration of

stereoisomers has increased the number of compounds to be considered from 46743 to 65546. The results for all *in silico* methods were collected for all of the stereoisomers. A final per-compound score was calculated by computing the mean value of the scores of the stereoisomers.

2.5 HydraScreen

HydraScreen is machine learning scoring function (MLSF) composed of a CNN-based deep learning framework designed for accurate prediction of protein-ligand affinity and pose scoring [**?**]. For a given target protein, HydraScreen estimates the affinity of a query ligand in a twofold process. First, a set of protein-ligand conformations are generated, creating a docked pose ensemble. Second, the affinity and pose estimation of each conformation is estimated, and a final aggregate affinity value is computed by means of a Boltzmann-like average of the protein-ligand conformational space. A schematic of the described procedure can be observed in Figure 1.

Docked poses are generated in a similar fashion to that outlined in [**?**]. Briefly, we use the open-source Smina software, a fork of the AutoDock Vina software with a number of improvements, to generate poses of a query ligand in the pocket of our target protein. For each protein-ligand pair, the docking process involves: (1) preparation of the protein structure, (2) preparation of the ligand structure, and (3) docking with Smina. For the protein, the following steps were taken to prepare it for docking: (1) solvent and ion deletion, (2) repair of truncated side-chains using Dunbrack 2010 rotamer library, (3) adding hydrogens (histidines were treated like other standard residues), (4) adding charges. Non-standard residues were changed to the nearest standard residue. For example, selenomethionine (MSE) → methionine (MET). For ligands, each ligand is sanitized through RDKit (2021.09.03), with hydrogens added prior to sanitization if the protonation state was incomplete or corrupt. We generate up to 20 poses per ligand query, and set the following Smina input parameters: (`num_modes`, 20), (`min_rmsd`, 1Å). Furthermore, we define the pocket by using the `autobox` option, passing in the reference crystal ligand pose (DL1) from 6BFN, and including all atoms within 4 Åof any atom in the native conformation.

In this study we primarily use HydraScreen to find potential hits amongst a large compound library, therefore we rely on its ranking power to discriminate between candidates that (i) do not successfully bind to the pocket and/or (ii) do not alter the function of the target (low affinity).

2.6 Benchmarks

We introduce a set of baselines for structure-based and ligand-based methods to better understand the contribution of HydraScreen with respect to traditional approaches.

Smina

In contrast to ranking our compounds according to our MLSF, Smina exploits a traditional structure-based approach. Herein, protein-ligand binding can be scored according to the maximum energy required to remove a ligand pose from the pocket (free energy). In order to score our compounds, we leverage our already generated

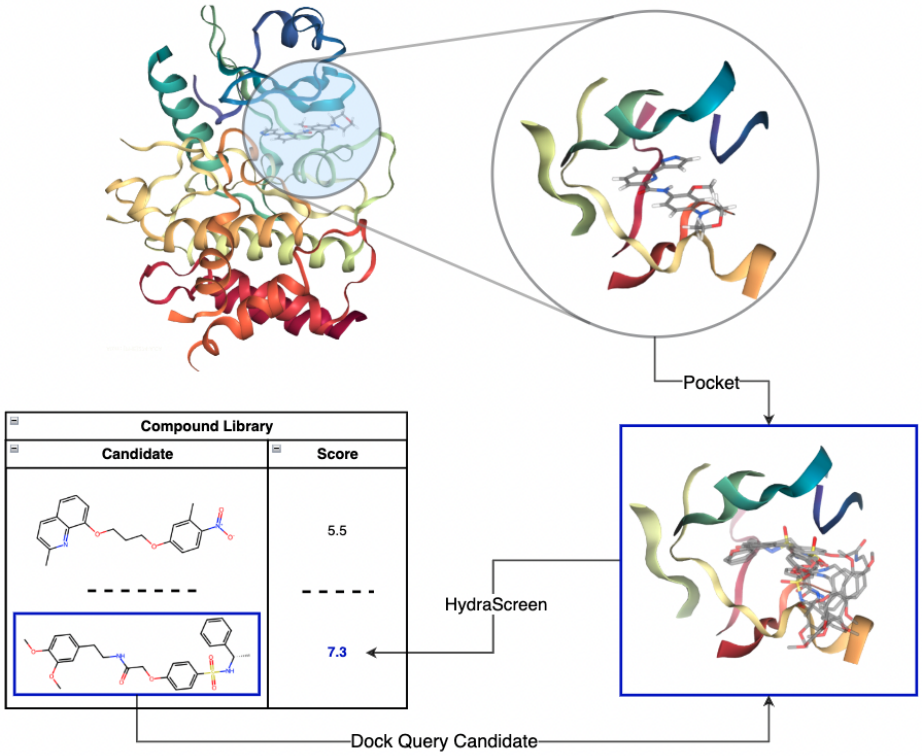


Fig. 1: End-to-end structure-based scoring via HydraScreen. Initially we extract IRAK1's pocket from PDB 6BFN, using the co-crystallised ligand (DL1) pose to identify the relevant residues (top). For each candidate in the compound library we create a pose ensemble in the identified pocket via docking. This docked ensemble is used as the input to HydraScreen which returns a score predicting the affinity of the query candidate.

poses and, for each docked ensemble, extract the largest free energy calculated by Smina (amongst all the poses).

DeCAF

Density-Encoded Canonically Aligned Fingerprint (*DeCAF*) [43] is a ligand-based approach that measures the similarity between two molecules. *DeCAF* can be used to rank compounds by rewarding similarity between the query candidate and the reference molecule (DL1). *DeCAF* score is computed by: (i) Finding the maximal common subgraph between the corresponding molecular graphs, represented as a coarse network of pharmacophore descriptors; (ii) Computing the modular product of the two graphical models and extracting the similarity between the maximal clique identified. In contrast to other shape-based methods like *USRCAT* [41], DeCAF does not require conformer generation.

Random Forest

We have trained a Random Forest (RF) classifier using publicly available IRAK1 data. We convert the available pKi and pIC_{50} values for IRAK1 to boolean values based on whether they are over or under the micromolar range $> 6pIC_{50}$ threshold. Out of 689 molecules available on PubChem, 142 were active and 547 were inactive. The inactive class was further up-sampled by 5K by using DeepCoy [23]. The compounds generated with DeepCoy were ensured to be structurally dissimilar to the actives while maintaining similar molecular weight as well as synthetic accessibility. The classified model was trained using Morgan fingerprints (ECFP4) generated using RDKit.

Pharmit

Pharmit [44] provides an online, interactive environment for the virtual screening of large compound databases using pharmacophores, molecular shape and energy minimization. Queries are specified in terms of a pharmacophore, a spatial arrangement of the essential features of an interaction, and molecular shape. Search results can be further ranked and filtered using energy minimization. Pharmit uses state-of-the-art sub-linear algorithms to provide interactive screening of millions of compounds. Queries typically take a few seconds to a few minutes depending on their complexity.

The pharmacophore hypothesis can be extracted from a co-crystallized structure; we used 6BFN for this to extract a 6-point pharmacophore hypothesis, later used in scoring the HTS compounds. In order to create a continuous score that can be used to rank the compounds, we extended the functionality of Pharmit to assess *how well* a compound matches a hypothesis, rather than a boolean match. This approach works by computing subsets of the original pharmacophore hypothesis and performing conformer matching on them, before combining the results of the subset matches to get a final score. Such a hypothesis-subset screening was made possible by the highly efficient nature of the Pharmit algorithm.

2.7 IRAK1 assay

Purified recombinant IRAK1-His (cat. # 40202) was purchased from BPS Bioscience Inc. (San Diego, CA, USA). Kinase tracer 236 (cat. # PR9078A) was purchased from Thermo Fisher Scientific Inc. (Waltham, MA, USA) Eu-W1024-anti-6xHis antibody (cat. #AD0400) and 384-well white ProxiPlates^{TM} (cat. # 6008289) were purchased from Perkin Elmer, Inc. (Waltham, MA, USA). Echo-qualified 384 well COC low dead volume source microplates (cat. #001-16128) and Echo-qualified 384 well polypropylene microplates (cat. #001-14615) were purchased from Beckman Coulter Inc.(Indianapolis, IN, USA).

LanthaScreen^{TM} Eu Kinase Binding Assay for IRAK1 The experimental method was developed based on the Invitrogen^{TM} IRAK1-GST LanthaScreen^{TM} binding assay. The assay was carried out in an enclosed workcell with subdued lighting. All reagents were prepared in the assay buffer (50mM HEPES, 10mM $MgCl_2$, 1mM EGTA, 0.01% Brij-35, 1mM DTT) and kept on ice. These included 2 x tracer 236 (0.2 μM), 2 x IRAK1 /antibody solution (20nM IRAK1-His, 4nM Eu-W1024-anti-6xHis antibody) and 2 x antibody solution (4nM Eu-W1024-anti-6xHis antibody). Five microliters of 2 x tracer 236 was dispensed into a 384-well white ProxiPlate^{TM},

followed by either 5μl of 2 x IRAK1/antibody solution or 5μl of 2 x antibody solution on a Tempest®dispenser (Formulatrix, Inc., Bedford, MA, USA). The plate was sealed on a Wasp plate sealer (KBiosciences Limited, Basildon, Essex, UK) and centrifuged at 1000 x g for 15 seconds on a HiG^{TM} automated centrifuge (BioNex Solutions Inc., San Jose, CA, USA) and incubated at room temperature for 30 minutes. The plate was then peeled and read on a PHERAstar®FSX (BMG LABTECH Inc., Cary, NC, USA) with a LanthaScreen^{TM} module at 340/615, 665 nm. The TR-FRET ratio (acceptor emission/donor emission x 10000) was used as the readout.

Biovalidation and pilot screen

Biovalidation was carried out with identical assay settings as for the anticipated production runs. Assay conditions and the instrument settings were tested for their performance within the acceptance criteria (Z' $\geq$ 0.5 (see eq. 1, where p and n refer to positive and negative control wells in the plates), no visible patterns on assay plates). Compounds from 2 library plates were dispensed at 10 nL/well in single point in columns 3 to 22 on assay plates (final concentration in assay at 10 μM) and 10 nL/well of DMSO was dispensed in columns 1, 2, 23 and 24 for controls. Ten nanoliter per well of DMSO was dispensed into all wells on positive and negative control plates. Compounds and DMSO were dispensed on an Echo 655 liquid handler in an Access workstation. For the kinase binding assay, the 2 x tracer solution was dispensed into all wells on all plates. For the assay plates, the 2 x IRAK1/antibody solution was dispensed into columns 1 and 3 to 23. The negative control plates have the same layout as the assay plates, with DMSO in place of the compounds. For the positive control plates, the 2 x antibody solution was used in place of the 2 x IRAK1/antibody solution in columns 3 to 22. Six plates were dispensed in total, including 2 assay plates, 2 negative control plates and 2 positive control plates. The compound dispense run and the binding assay run were both set up and launched in the Cloud Lab. The automated runs were carried out in the workcells and with the autoprotocols designated for production. Z', signal-to-background ratio and compound hit rate were analyzed as performance parameters.

$$\text{Z'} = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|} \tag{1}$$

Biovalidation was followed by a pilot screen with a plate number close to that in a production run for evaluation of the robustness of the assay, the automation scheduling and the data transfer. Compounds on 20 library plates were dispensed onto 20 assay plates. Two positive and two negative control plates were used in the same manner as in biovalidation. The screen was carried out with the same LOT of reagents, procedure, instrument settings and autoprotocols as in biovalidation. Z', signal-to-background ratio and compound hit rate were analyzed as performance parameters.

2.8 High-throughput screen (HTS)

Primary screen

The primary screen runs were performed with the same reagents and procedures as the pilot screen. Up to 40 plates were assayed per run. In total 153 plates and 46743

compounds were screened at 10 μM in single point. Plate QC was performed using manual inspection and Z' analysis (equation 1). Plates not passing with Z' $\geq$ 0.5 were run repeatedly.

Primary screen data analysis

With the collected fluorescence data we have performed per-plate fluorescence normalization. Normalization entailed scaling the reported fluorescence in the ratio channel in between negative control (DMSO) and positive control (Staurosporine) so that 0 normalized fluorescence value would correspond to the mean of the negative control value and 100 normalized fluorescence ratio channel value would correspond to the positive control fluorescence, as seen is Equation 2. Mean values from the 32 negative control (μ_{DMSO}), and 32 positive control (μ_{SS} - Staurosporine) wells in each plate were used in the normalization.

$$ratio_{norm} = \frac{ratio_{raw} - \mu_{DMSO}}{\mu_{SS} - \mu_{DMSO}} \tag{2}$$

Normalized fluorescence ratio reported relative inhibition of IRAK1, where 100 corresponds to the inhibition with staurosporine. The distribution of normalized fluorescence ratio values is presented in Figure 2. Only normalized fluorescence ratio channel values were used in further analysis. The threshold for hit selection was chosen to be 50% inhibition of IRAK1 relative to staurosporine. Using this threshold 353 hit compounds were identified.

Single-dose hit confirmation

For single-dose hit confirmation we have selected 10 plates with most hits and assayed them at 10 μM in duplication. The experiments showed high consistency with Z' values above 0.6 in all plates and high correlation between the replicates.

Compound clustering

The 353 hits identified via HTS were subsequently clustered by their structural similarity using the Louvain algorithm [4]. The algorithm identifies clusters ("communities") within a graph of related compounds that is constructed using compound Tanimoto similarity (TS). The Louvain algorithm was chosen for its compatibility with Tanimoto similarity and robustness in terms of the number of clusters in the dataset. In total, 200 unique clusters were identified, 160 of which were singletons. 5 compounds with the greatest ligand efficiency (LE) values were selected from each cluster to form a diversified set of 283 hits.

Hit dose-response

A diversified set of 283 hits was selected for dose response assays. Each compound was assayed in an 8-point curve with an approximately 4-fold (subject to Echo dispense volume limits) dilution starting at 30 μM in triplicates. The exact concentrations are 30, 7.5, 1.875, 0.469, 0.117, 0.029, 0.007, 0.002 μM. On each plate, three replicates of a staurosporine titration curve starting at 3 μM were assayed in parallel as the reference.

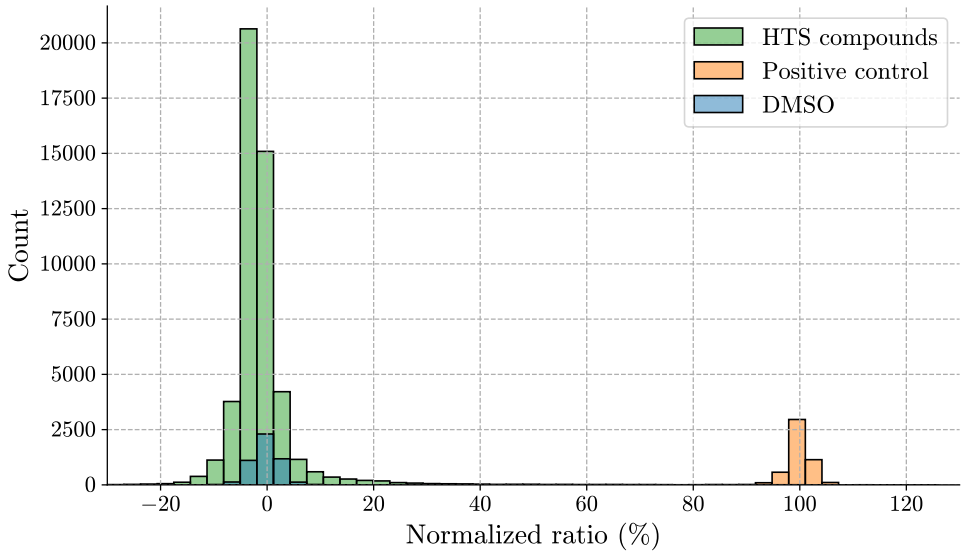


Fig. 2: Normalized fluorescence values in the ratio channel from IRAK1 HTS. Values from the individual compounds from the library, positive and negative control wells are represented in different colors. Here 0% corresponds to the mean normalized fluorescence ratio in negative control wells and 100% to normalized fluorescence ratio in positive control wells across the whole library. Positive control represents IRAK1 inhibition with staurosporine.

Hit dose-response data analysis

The IC_{50} of each dose-response curve was derived by fitting a four-parameter logistic (4PL) model, shown in Equation 3. The variables of the model are shown in Equation 3:

- A: Minimum asymptote. It's the response value when x approaches infinity.
- D: Maximum asymptote. It's the response value when x is very small or close to zero.
- B: Slope factor or Hill's slope. It describes the steepness of the curve.
- C: Inflection point. The concentration of the analyte that gives half-maximal response.

$$f(x) = A + \frac{D - A}{1 + \left(\frac{x}{C}\right)^B} \tag{3}$$

The 4PL model was fitted for each compound with data points for all three replicates all at once. The resulting in a pIC_{50} distribution as seen in Figure 3. As an additional quality control, fits for all submicromolar compounds were manually inspected. In 7 cases the model fits were erroneous and the IC_{50} values were reduced

to $30\mu M$, the highest concentration of the assay. Whenever the model produced fits with IC_{50} values higher than $30\mu M$, the largest measured concentration in the assay, the IC_{50} values were made to be equal to $30\mu M$, or equivalently, ~ 4.52 pIC_{50}.

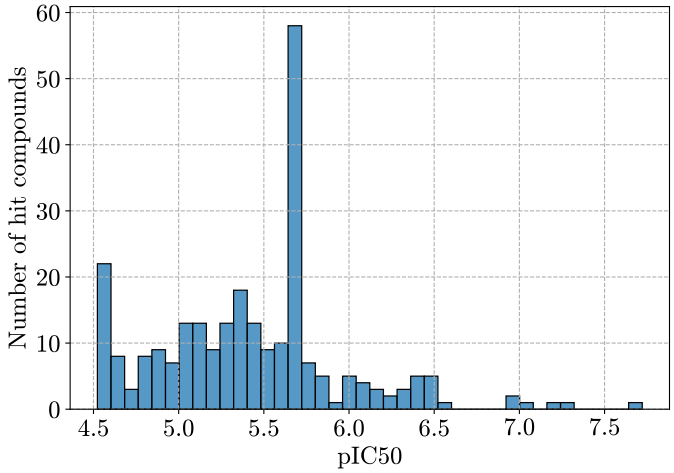


Fig. 3: Distribution of $\log_{10}$-normalized IC_{50} values (pIC_{50}) for the 283 hit compounds for which the dose-response data was collected.

3 Results

3.1 Target evaluation using SpectraView

3.2 Target evaluation goals and criteria

Multiple protein targets were considered for the joint Ro5-Strateos project. The targets were proposed by Strateos based on the availability of scalable assays and interest from their potential customers. In order to perform a thorough assessment of each target, we employed Ro5's target evaluation tool: SpectraView. Our aim was to identify a therapeutically relevant and commercially viable target for a drug discovery project, which could also be used for the prospective validation of Ro5's HydraScreen model. SpectraView relies on Ro5's integrated Knowledge Graph to serve information from multiple data sources (see methods section 2.1) following these queries:

- Availability of a crystal structure
- Existing biochemical data
- Existing drugs and most potent compounds
- Publication count and trends
- Novelty/Traction balance
- Target-disease associations
- Translation from academia to industry

- Competitive landscape

Table 1: Targets considered for the Ro5-Strateos project and some of the data considered in target evaluation. Data from RSCB PDB [28], PubChem [25] and DrugBank [51] as of start of the project on 2022-01.

Target	Crystal structures, n	Data points, thousands	Max affinity, nM	FDA approved drugs, n
JAK1	44	6.5	< 0.01	5
JAK2	115	10.0	< 0.01	5
JAK3	38	6.0	< 0.001	5
TYK2	38	3.5	< 0.7	1
IRAK1	1	1.3	< 5.6	0 / 1 inv. [a]
FGFR1	59	7.0	0.2	5
FGFR2	37	2.1	0.1	7
FGFR3	4	4.5	0.1	9
FGFR4	28	2.0	0.1	6
RIPK2	24	0.2	1.3	0 / 1 inv.
VGFR2 (KDR)	45	18.0	0.02	2
TAK1 (MAP3K7)	19	0.3	1	-

[a] Early clinical studies of IRAK1 inhibitor R835 [29].

3.3 Target evaluation

One of the main considerations when selecting a target in drug discovery is its novelty/confidence trade-off [26]. Most of the targets considered for the project were very well-studied, as marked by the volume of PubMed publications mentioning them (e.g. 800 articles mentioning KDR were published each year, see Suppl. Fig. A1), availability of crystal structures, biochemical assay data and approved or investigational drugs (Tab. 1). We have focused on the relatively less established targets with lower volume of publications, fewer data points and only few known high activity compounds - IRAK1, FGFR3 and TAK1.

The availability of a crystal structure and biochemical assay data were crucial criteria when selecting a target for virtual screening and subsequently the prospective validation of HydraScreen. The crystal structure is necessary to generate poses between the ligand and binding site for predictions by HydraScreen and docking, while public assay data would be used in *DeCAF* pharmacophore-based comparison and for building ligand-based QSAR models. All of the considered targets had at least 1 crystal structure (Table 1). IRAK1, one of the least established targets, had a recent publication with its crystal structure resolved [48] (6BFN) and $1.3k$ biochemical assay data points. Moreover, IRAK1 was not present in HydraScreen's training data set, allowing the method's unbiased and fair prospective validation. IRAK1 thus satisfied the minimal requirements for selection, while also being the most underexplored target in the selection.

Additional evidence was needed to substantiate IRAK1's choice for a drug discovery program in terms of its therapeutic links. In contrast to many other kinases,

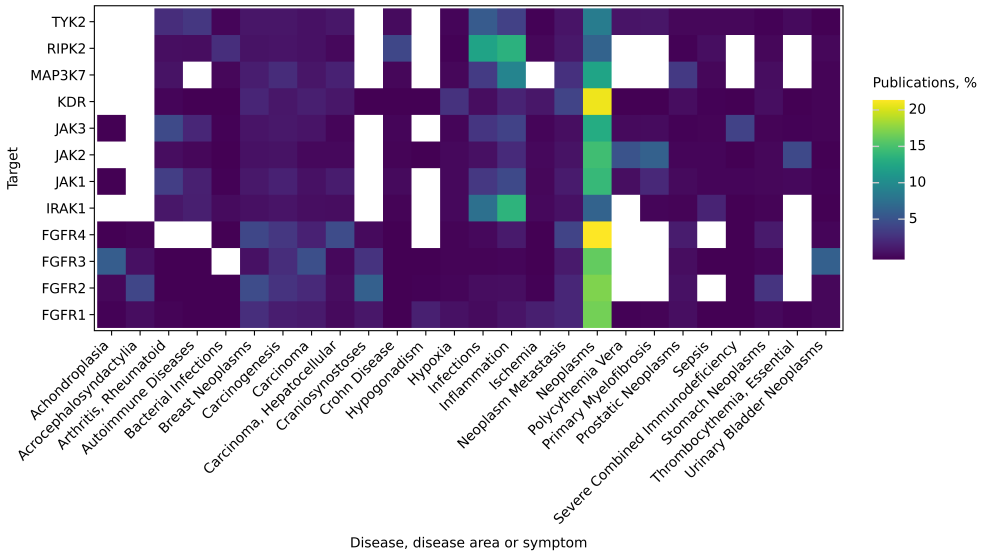


Fig. 4: Diseases, disease areas and symptoms co-mentioned with each of the considered targets. Colors represent the fraction of PubMed-indexed publications per disease for each of the targets.

IRAK1 is primarily associated with inflammation (Fig. 4, e.g. [21]) and not cancers. It's only recently that IRAK1 has been implicated in multiple cancers, including breast cancer [49], lymphoma [16] and acute myeloid leukemia [20]. The combination of fewer publications and emerging new therapeutic links provides additional support for IRAK1's selection.

Finally, IRAK1 was assessed in terms of the potential competitors in the drug development effort. We have analyzed the competitive landscape by querying the publications and patents held by major pharma companies as well as the most potent drugs and compounds reported in the public domain. We have found only few PubMed-indexed publications with affiliations linked to major pharma companies: Johnson and Johnson - 4, Genentech - 2, Roche - 2, GlaxoSmithKline - 2, Pfizer - 2, Novartis - 1, Rigel - 2 (Fig. A3). This was a promising finding, given much greater academic interest in IRAK1 with 637 publications (Fig. A5). Similarly, we have looked into patent and patent applications (Fig. A4). There were two notable entities - Dana Farber Cancer Institute and Yissum Research and Development Company of the Hebrew University both holding 14 patents mentioning IRAK1. No major pharma companies were found to hold patents linked to IRAK1. We have then looked into the most potent compounds active against IRAK1. Only few nanomolar compounds have been reported for IRAK1 (e.g. JH-X-119-01 is 9nM, [16]). IRAK1 has been shown to be an off-target of an active metabolite R406 of FDA approved drug Fostamatinib developed by Rigel pharma for the treatment of chronic immune thrombocytopeni [37]. Rigel pharma has recently started pre-clinical and clinical studies of IRAK1/4 inhibitor R835 which

has shown potential in murine models of multiple inflammatory diseases, including arthritis and lupus, however, has not yet received FDA approval [29]. The combination of largely academic research in IRAK1 with only recently emerging interest by pharma companies (Fig. A5), especially the supporting pre-clinical and clinical work [29, 37] provides corroborative evidence for its potential as a prospective drug target. The lack of any FDA-approved IRAK1 targeting drugs leaves an opportunity for the development of novel small molecule inhibitors. Altogether, the balance between the novelty and sufficient support in terms of biochemical and biological rationale as well as competitive considerations make IRAK1 an attractive target to be pursued in the Ro5-Strateos study.

3.4 Identification of IRAK1 hits using HydraScreen

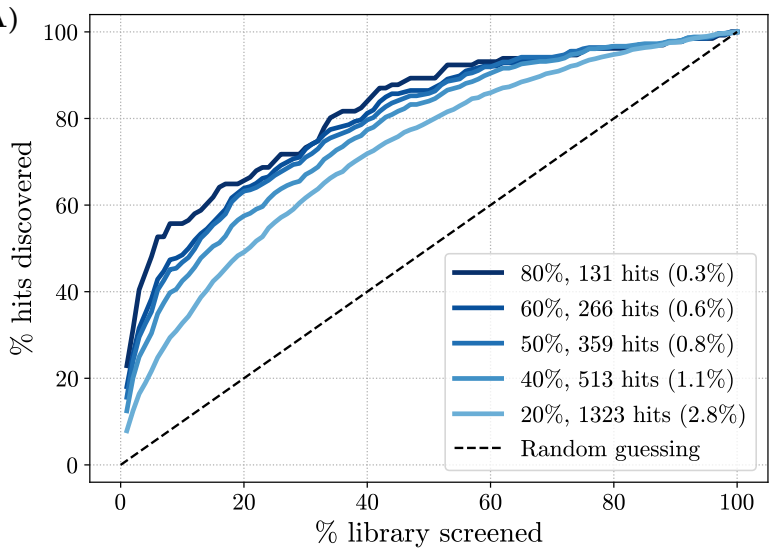
3.4.1 HydraScreen virtual screen

Following the selection of IRAK1 as a target using SpectraView, we performed *in silico* virtual screening for IRAK1, subsequently followed by experimental hit identification via HTS. The goal of this stage of the project was to perform prospective evaluation of HydraScreen's performance using *in vitro* data collected by Strateos HTS in comparison to traditional, industry-standard methods like *Smina* [27] (molecular docking), *DeCAF* [43] (pharmacophore modeling) and a random forest model trained on IRAK1 assay data that is publically available. Together, these results will provide a comprehensive and unbiased evaluation of HydraScreen as a virtual screening method.

In order to prospectively evaluate Ro5's HydraScreen model's performance in hit identification, we performed a virtual screen of the entire Strateos 47k compound diversity library (see Methods 2.5). HydraScreen predictions were then used to rank the library and select the top 1% (470) compounds to be considered *in silico* hits. Strateos subsequently performed an *in vitro* HTS with the same library (see Methods 2.8), which returned 353 hit compounds. These compounds were compared to the ones ranked in the top 1% by HydraScreen. In total, 57 compounds were discovered by HydraScreen that were also identified in the HTS, constituting a 15.9% hit discovery rate via virtual screening.

Next we have investigated the impact that different IRAK1 activity thresholds for hit selection in HTS can have for hit identification in the HydraScreen virtual screen (Figure 5). This is an important consideration, because both the virtual *in silico* and the high-throughput *in vitro* screens rely on arbitrary thresholds for hit selection [56, 55]. Here, we consider the comparison of virtual screening predictions against the HTS results for each individual compound in the ranking generated by HydraScreen. Virtual screening hit recovery rate for HydraScreen was estimated as a proportion of hits identified per number of compounds screened or tested. Standard HTS protocols randomly test compounds from the library (i.e. in the order in which they are stored); therefore, the hit recovery rate of traditional HTS is roughly proportional to the percentage of the library screened, as marked by the diagonal black line in Figure 5. Any method that is able to prioritize active compounds over the inactive ones would provide a better hit recovery rate than random sampling (i.e. above the diagonal line in Figure 5).

(A)



(B)

Ratio (%)	**Library screened (%)**			
	1	5	20	50
80	23.8	48.4	66.7	90.5
60	18.5	38.5	64.6	86.9
50	15.9	35.4	63.7	86.1
40	12.6	30.4	57.8	84.2
20	7.7	21.5	49.1	79.3

Fig. 5: **(A)** HydraScreen hit discovery rate (% of hits discovered per library screened) for different IRAK1 inhibition thresholds in HTS (marked by lines of in the shades of blue). For each IRAK1 inhibition threshold the number of hits identified in HTS is presented together with the overall HTS hit rate. Dashed black line represents random compound ranking. Supporting data is presented in table **(B)**.

We found that ranking the compound library according to HydraScreen predictions greatly increases hit discovery rates in virtual screening for IRAK1. This result is consistent for any number of top compounds selected in the ranking and any >0% threshold of relative IRAK1 inhibition in HTS. Using the 50% IRAK1 inhibition threshold, as was used in the *in vitro* experiment, HydraScreen identified 35.4% of the hit compounds within the top 5% and 63.7% with 20% of the ranking (Fig. 5B). Notably, close to 90% of the hits can be identified with only 50% of the ranking (see Figure 5B). Generally, HydraScreen exhibits better performance at higher IRAK1 assay inhibition thresholds. For example, HydraScreen identified 23.8% (30 out of 126)

of hits at the top 1% of the compound ranking when using 80% relative inhibition threshold of IRAK1 (Fig. 5B).

The number of distinct highly active scaffolds identified in HTS can be more relevant than the total number of hits. Greater variety of scaffolds provides medicinal chemists with more opportunities for lead series development, which is crucial at the later stages of drug discovery [22]. Therefore, in addition to hit discovery rates, HydraScreen's performance was also assessed in terms of its ability to prioritize highly active compounds that are also structurally diverse. This would provide evidence both for models ability to predict ligand potency, and lack of structural biases in the prioritization of the compounds.

To select a diverse, representative and unbiased set of compounds to screen in the secondary assay, the 353 hits from HTS were clustered by their structural similarity using the Louvain algorithm. In total, 200 unique clusters were identified, 160 of which had single compound member. Scaffolds associated with each cluster were identified using maximum common substructure (MCS) analysis. Five compounds with the greatest ligand efficiency (LE) values were selected from each cluster to form a diversified set of 283 hits from 200 distinct scaffolds. For these 283 diversified hit compounds, Strateos collected dose-response data (see methods 2.8). Based on their pIC_{50} ($-log_{10}$ transformed IC_{50}) activity values, hits and their corresponding scaffolds were grouped into micromolar, high nanomolar and nanomolar groups (Table 2). We identified 5 nanomolar and 25 high nanomolar hits, while the rest possessed micromolar activity. Scaffolds were labeled based on the most active compound in each cluster. Out of the 200 total scaffolds, 15 were labeled as high nanomolar and 3 as nanomolar. We will refer to the union of high nanomolar and nanomolar compounds as *sub-micromolar*.

Table 2: Dose-response assay results for 283 diversified hits. Compounds and scaffolds were labeled as micromolar, high nanomolar and nanomolar based on the their pIC50 values. For scaffolds, the highest activity found in the correspond cluster of compounds was used as a label.

	pIC50 range	**Compounds**	**Scaffolds**
Micromolar	< 6	253	182
High nanomolar	$6 \leq x < 7$	25	15
Nanomolar	≥ 7	5	3

The dose-response data was used to evaluate HydraScreen's performance in terms of discovery of highly active scaffolds (Fig. 7). We considered a scaffold to be "discovered" by a model if at least one compound from the corresponding cluster was ranked by the model in top X% of the library. Notably, HydraScreen successfully recovered compounds belonging to all 3 nanomolar scaffolds within the top 1% of the library. Within the top-ranked 2%, HydraScreen recovered 8/18 of the submicromolar scaffolds. The remaining 10 scaffolds were present in the top 50% of the ranked compounds.

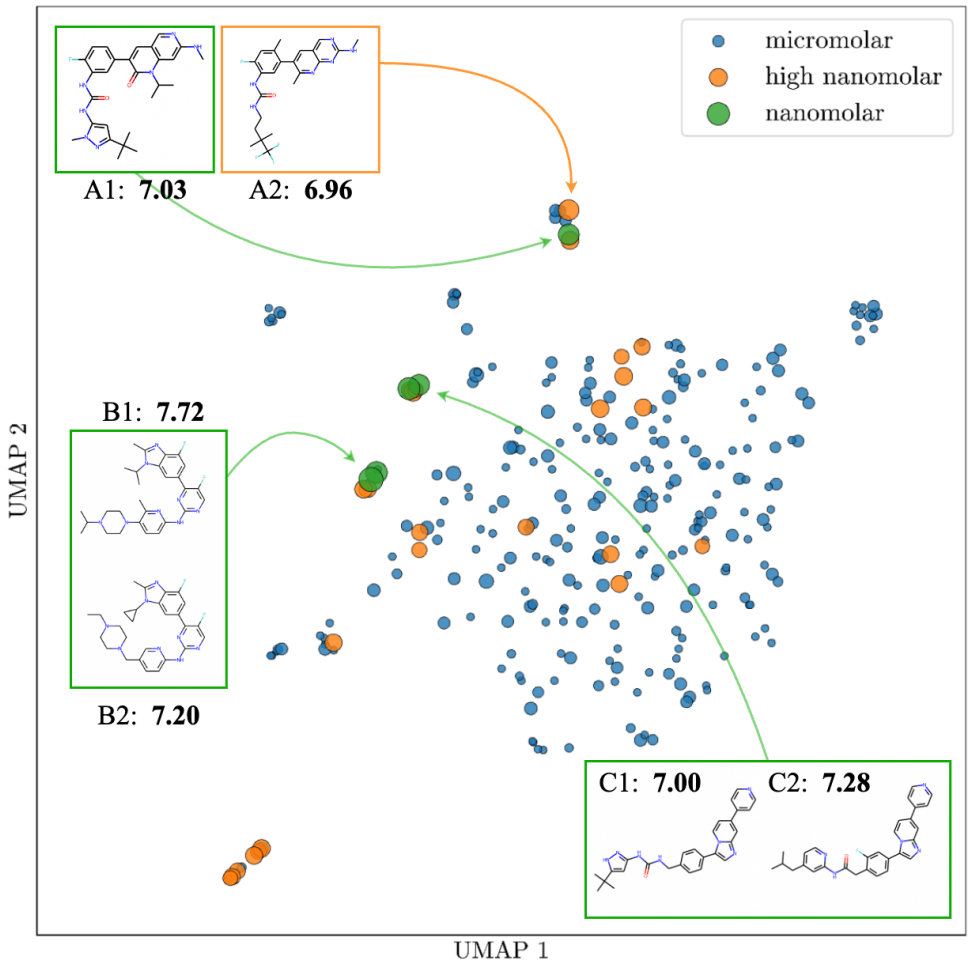


Fig. 6: UMAP embedding of 283 hit compounds from HTS screen. The space in the plot represents relative similarity of the compounds. Nanomolar compounds from the three nanomolar scaffolds are highlighted with their pIC_{50} values indicated underneath.

3.4.2 HydraScreen comparison against other virtual screening techniques

Virtual screening can be performed using a range of different techniques [31]. It is therefore relevant to evaluate HydraScreen's performance in comparison to different traditional methods. In parallel to the HydraScreen virtual screen, we also prospectively generated predictions using *Smina* docking [27], *DeCAF* 2D pharmacophore matching [43], *Pharmit* 3D pharmacophore hypothesis matching [44] and a random forest model trained on ECFP4 fingerprints. The comparison between methods was

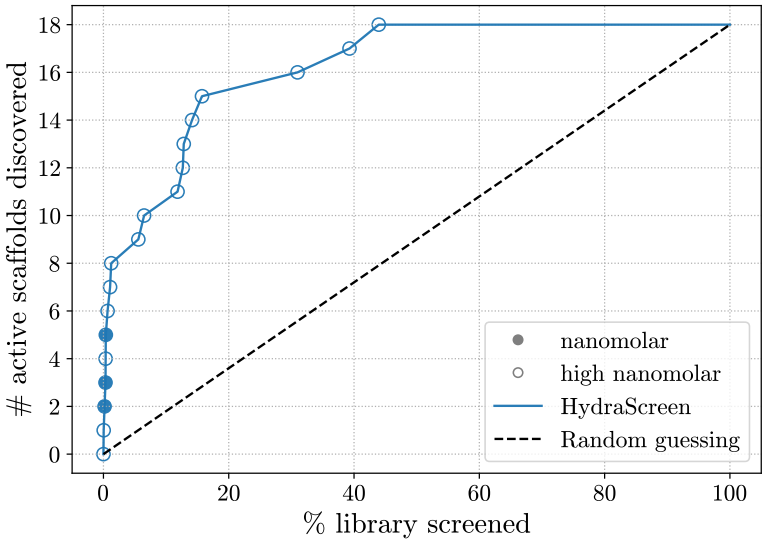
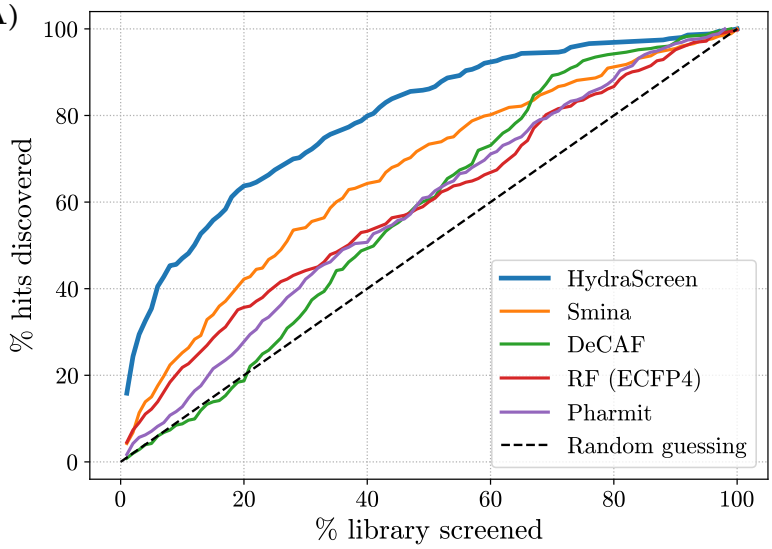


Fig. 7: HydraScreen distinct scaffold discovery rate (number of distinct scaffolds discovered per library screened). Dashed black line represents random compound ranking. Filled and empty circles represent nanomolar and high nanomolar scaffolds correspondingly.

made at the 50% IRAK1 inhibition threshold used in *in vitro* HTS, with 353 hits identified in total. HydraScreen considerably outperforms the traditional techniques and its hit identification rate is consistently higher for any selected percentage of the top ranked compounds. At the top 1% ranking the model provides 3.5x better performance than *Smina*, 3.2x better than *RF* with ECFP4 (Fig. 8B) and ∼20 better performance than pharmacophore-based methods *Pharmit* and *DeCAF*. Both *Smina* docking and random forest perform similarly, and outperforming the pharmacophore based methods. However, for low ranked compounds (>70% ranking) *DeCAF* exhibits better performance than docking.

Following the same methodology used for evaluating HydraScreen in terms of its ability to rank distinct active scaffolds, we compared different virtual screening methods. As before, we considered a scaffold "discovered" if at least one compound from that scaffold was ranked in the corresponding library screening range. In comparison to other methods, HydraScreen exhibits superior scaffold discovery rates (Fig. 9A). Within the top 1% of the ranked compounds, HydraScreen discovered all three nanomolar and, in total, 6 out of 18 submicromolar scaffolds (Fig. 9B). In comparison, docking ranked the three nanomolar scaffolds only at 18% of the ranking, *Pharmit* at 27% and random forest at 30%. Random forest ranked one of the nanomolar scaffolds as the top 10 compound (0.02%). This scaffold is has a highly similar analogue in the public IRAK1 data that the RF model has trained on. This scaffold and the analogue are discussed in more detail in Section 3.5.1.

(A)



(B)

Method	Library screened (%)			
	1	5	20	50
HydraScreen	**15.9**	**35.4**	**63.7**	**86.1**
Smina	4.2	15.0	42.2	73.4
DeCAF	0.8	4.2	19.0	60.6
Random Forest	4.5	12.2	35.7	60.1
Pharmit	1.7	7.1	27.8	61.2

Fig. 8: **(A)** Hit discovery rates provided by different methods in IRAK1 virtual screen. Dashed black line corresponds to random compound ranking. Supporting data is presented in table **(B)**.

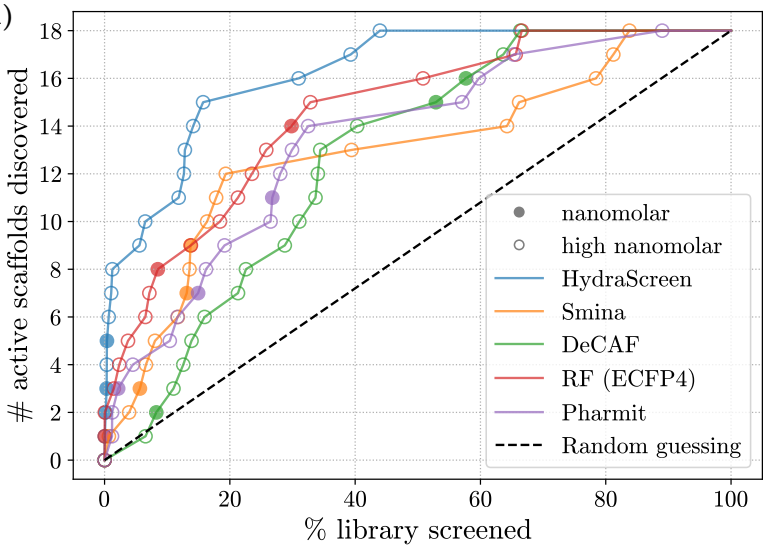
3.5 IRAK1 hits

IRAK1 hit identification via HTS experimentally discovered 283 hit compounds, representing 200 distinct structural scaffolds. In the last stage of the project, we evaluated these compounds and scaffolds in terms of their novelty, physio-chemical properties and IRAK1 binding modes.

3.5.1 Hit novelty and properties

In order to assess the uniqueness of the 283 diversified hits, we compared them against IRAK1 actives available in PubChem. Out of the 689 compounds reported to be active against IRAK1, 141 had sub-micromolar activity (>6 pIC_{50}). For each of the 283 hits, we found the nearest neighbor in the set of IRAK1 actives and scaffolds based on

(A)



(B)

Method	Library screened (%)			
	1	5	20	50
HydraScreen	**5**	**8**	**15**	**18**
Smina	1	2	12	13
DeCAF	0	0	6	14
Random Forest	2	5	10	15
Pharmit	1	4	9	14

Fig. 9: **(A)** Scaffold discovery rates provided by different methods in IRAK1 virtual screen. Nanomolar and high nanomolar scaffolds are marked by filled and empty circles correspondingly. Dashed black line corresponds to random compound ranking. Supporting data is presented in table **(B)**.

their Tanimoto Similarity (TS). The number of neighbors above a certain similarity threshold is reported in Table 3. We observe that the vast majority of compounds are distinct from publicly known actives, not having related chemical matter with a TS <0.4 in the public domain. Only 39 compounds, corresponding to 27 distinct scaffolds, exhibit >0.4 TS. Focusing on the nanomolar hit compounds, only 1 of the 3 scaffolds had a similar (TS < 0.4) active compound in the public domain; the closest structure was the Pan-RAF inhibitor LY3009120 [17] with a TS of 0.82. A whole cell-based kinase screen revealed that LY3009120 displayed some IRAK1 inhibition (390 nM IC_{50}), though it was not the primary target of the compound.

The 30 sub-micromolar hit compounds represented 18 distinct scaffolds, with the 6 most active compounds spanning 3 of these as indicated in figure 6. The 6 most

TS threshold	Hits	Scaffolds	Nanomolar scaffolds
Total	283	200	3
> 0.4	21	13	1
> 0.6	5	3	1
> 0.8	1	1	1
> 0.9	0	0	0

Table 3: Numbers of hits and scaffolds that have at least one neighbor in the IRAK1 public dataset that is more similar than the specified threshold.

active compounds are synthetically tractable, with synthetic accessibility scores in a similar range to that of catalogue compounds (2-3) [12]. They border on the upper end of the Lipinski rule of 5 [8] with regards to molecular weight (466 to 521 g/mol) and Crippen LogP values of 4.7 to 6 [50]. The high molecular weight and LogP nature will have to be further assessed during a medicinal chemistry program.

3.5.2 HydraScreen hit compound binding modes

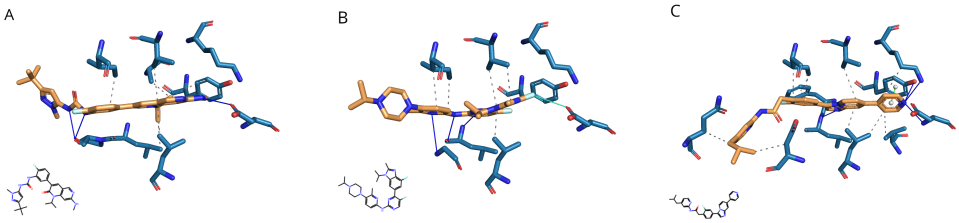


Fig. 10: IRAK1-ligand poses with the highest HydraScreen confidence for selected nanomolar hits A1, B1, and C2, represented in panels A, B, and C respectively. PLIP protein-ligand interactions are shown with grey dashes (hydrophobic interactions), blue lines (H-bonds), and green dashes connecting white spheres (pi-pi stacking).

In addition to ranking compounds according to their predicted affinity against a target, HydraScreen provides insight into the likely binding modes of the compounds by predicting pose confidence for the docked poses. To describe the HydaraScreen predicted binding for the most active compound of each of the nanomolar scaffolds (compounds A1, B1, and C2, Fig. 6), the IRAK1-ligand interactions of the predicted highest confidence pose were identified using the protein–ligand interaction profiler, PLIP [3]. Across the highest confidence poses, the sequential aromatic heterocycles of the compounds were situated towards the back of the ATP binding pocket, with hydrophobic interactions with valine (V226), leucine (L347), and isoleucine (I218) residues (Fig. 10). The central heterocycles of compounds B1 and C2 forms hydrogen bonds (H-bonds) with the hinge region, whereas the urea in A1 forms H-bonds to the backbone. Both A1 and B1 interacts with the carbonyl of aspartic acid D358 in the

back of the pocket, respectively through an H-bond and halogen bond. On the other hand, the highest confidence pose of compound C2 highlights a pi-stacking interaction with the gatekeeper residue tyrosine Y288, as well as H-bonds to both Y288 and the catalytic lysine K239. Across the compounds, aliphatic sp3-rich motifs are situated toward the solvent exposed region of the pocket.

The HydraScreen-provided insight into the compound poses and the different interactions of scaffold motifs aids further compound design by highlighting areas and interactions to exploit not only around a specific scaffold, but also from one scaffold to another. The hit compound activity, novelty, and ample positions to tailor, render them attractive scaffolds for further SAR exploration and subsequent hit-to-lead development.

4 Discussion

Prospective validation of HydraScreen and hit identification in IRAK1

In this study, we propose an augmented drug discovery workflow that relies on Ro5's AI and data science platform while utilizing Stateos' robotic labs capabilities. We show how target evaluation driven by Ro5's SpectraView allowed for the selection of IRAK1 serine-threonine kinase target with emerging therapeutic links to cancers and inflammation. We provide evidence for HydraScreen's virtual screening performance. Notably, Ro5's HydraScreen provides high hit discovery rates in virtual screening, with upwards of 15.9% hits and all of the 3 nanomolar scaffolds ranked in the top 1%. Moreover, the performance increased with stricter thresholds for experimental hit selection - 23.8% hit rate in top 1% of the ranked compounds with 80% relative inhibition of IRAK1 threshold. All of the distinct nanomolar and high nanomolar scaffolds were ranked in the top 50% of the compounds. HydraScreen is able to prioritize highly active compounds and does not exhibit biases in terms of treatment of structurally diverse compounds. Importantly, HydraScreen model has not been trained on IRAK1 data, so these results also reflect on the model's ability to generalize to an unseen target.

HydraScreen was shown to be superior to traditional, industry-standard methods like *Smina* [27] (molecular docking), *DeCAF* [43] (pharmacophore modeling) and a random forest model in terms of hit and scaffold discovery rates. HydraScreen offers state-of-the-art performance in line with the most recent AI models available for protein-ligand binding activity prediction [33]. Beyond the high hit rate, HydraScreen uniquely provides ligand pose confidence scores [**?**] which can be used when considering modifications of the most potent hits in the later hit to lead or lead optimization stages of a drug discovery program.

In addition, this study successfully identified potent and novel IRAK1 inhibitors. The most potent 5 nanomolar hits represent 3 distinct scaffolds, which are synthetically accessible, 2 of which are novel when compared to known IRAK1 actives. The scope to optimize these compound scaffolds was strengthened through the identification of analogues in the HTS, which revealed additional compounds with varying degrees of activity for all 3 nanomolar scaffolds (A: 8; B: 5; C: 32). Although the

hits have high Crippen LogP, the lack of correlation between the single point fluorescence readout and LogP suggested that activity is not driven by lipophilicity. The general high molecular weight and lipophilicity of the compounds will have to be further explored during a medicinal chemistry program. One of the identified nanomolar scaffolds exhibits high similarity to a known Pan-RAF inhibitor LY3009120 [17], while the other two do not have drug analogues. These compounds and associated series show promise for further development.

The most important contribution of this study is the prospective validation of HydraScreen for virtual screening. Prospective virtual screening evaluation studies reported in the literature usually experimentally test only a small fraction of the library, well below 1% - a dozen of compounds in total (median value of 44 based on 401 studies) [55]. The median of hit rates reported in these studies is approximately $\sim 11.76\%$ for 385 studies across all target classes and $\sim 9.61\%$ for 67 studies focusing on kinases [55]. However, the reported hit rates are prone to bias due to a small test size. A comparable study that used a support-vector machine (SVM) model for virtual screening in IRAK1 reported a 2.83% hit rate (1/38) with the most potent compounds reaching $2\mu M$ activity [7]. In contrast, we screened the entire 47k library to provide a robust assessment of HydraScreen and report a hit discovery rate of upwards of 15.9% for the top 1% (470) of tested compounds, which is well above the median and substantially higher than the comparable IRAK1 virtual screening study [7]. Furthermore, HydraScreen achieves even better hit rates (up to 23.8%, greater than the 3rd quartile 23.5% of hit rates reported in [55]) for higher IRAK1 inhibition thresholds. This is an important observation, since HTS assays are inherently noisy and their own intrinsic false-positive rate [56]. HydraScreen's evaluation at stricter IRAK1 inhibition thresholds is potentially more representative of its true performance due to a higher confidence in the hits selected from the assay (i.e. lower false-positive rate).

Summary

This study provides compelling evidence for the effectiveness Ro5's innovative tools, SpectraView and HydraScreen in early stage drug discovery.. By leveraging Ro5's HydraScreen and Strateos' automated labs, we show how AI-driven virtual screening with HydraScreen could offer high hit discovery rates and reduce experimental costs. In the top 1% of the ranked compounds, HydraScreen identified all three nanomolar classes with favorable chemical properties and almost a quarter of the total actives in the library. The unbiased, prospective evaluation of HydraScreenand comparison against industry-standard methods supports the reliability and robustness of our findings. Ro5's SpectraView and HydraScreen provide innovative methods that can expedite the early stages of drug discovery.

5 Acknowledgements

We would like to thank Sarah Flatters for conducting a thorough review of this manuscript and providing useful comments and suggestions.

References

[1] Milad Abolhasani and Eugenia Kumacheva. The rise of self-driving labs in chemical and materials sciences. *Nature Synthesis*, 2:483–492, 1 2023.

[2] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindah. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.

[3] Melissa F Adasme, Katja L Linnemann, Sarah Naomi Bolz, Florian Kaiser, Sebastian Salentin, V Joachim Haupt, and Michael Schroeder. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Research*, 49(W1):W530–W534, 05 2021.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[5] Denise Carvalho-Silva, Andrea Pierleoni, Miguel Pignatelli, Chuang Kee Ong, Luca Fumis, Nikiforos Karamanis, Miguel Carmona, Adam Faulconbridge, Andrew Hercules, Elaine McAuley, Alfredo Miranda, Gareth Peat, Michaela Spitzer, Jeffrey Barrett, David G. Hulcoop, Eliseo Papa, Gautier Koscielny, and Ian Dunham. Open targets platform: New developments and updates two years on. *Nucleic Acids Research*, 47:D1056–D1065, 1 2019.

[6] H. C.Stephen Chan, Hanbin Shan, Thamani Dahoun, H. Vogel, and Shuguang Yuan. Advancing drug discovery via artificial intelligence, 8 2019.

[7] Jinxin Che, Ruiwei Feng, Jian Gao, Hongyun Yu, Qinjie Weng, Qiaojun He, Xiaowu Dong, Jian Wu, and Bo Yang. Evaluation of artificial intelligence in participating structure-based virtual screening for identifying novel interleukin-1 receptor associated kinase-1 inhibitors. *Frontiers in Oncology*, 10, 9 2020.

[8] Beryl W. Dominy Paul J. Feeney Christopher A. Lipinski, Franco Lombardo. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23, 1997.

[9] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. 10 2022.

[10] Michael Dickson and Jean Gagnon. The cost of new drug discovery and development. *Discovery medicine*, 4:172–9, 06 2004.

[11] Gabriel H.S. Dreiman, Magda Bictash, Paul V. Fish, Lewis Griffin, and Fredrik Svensson. Changing the hts paradigm: Ai-driven iterative screening for hit finding. *SLAS Discovery*, 26:257–262, 2 2021.

[12] Schuffenhauer A. Ertl, P. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of*

Cheminformatics, 8, 2009.

[13] Cheng Fang, Ye Wang, Richard Grater, Sudarshan Kapadnis, Cheryl Black, Patrick Trapa, and Simone Sciabola. Prospective validation of machine learning algorithms for absorption, distribution, metabolism, and excretion prediction: An industrial perspective. *Journal of Chemical Information and Modeling*, 63(11):3263–3274, 2023. PMID: 37216672.

[14] Elizabeth Farrant. Automation of synthesis in medicinal chemistry: Progress and challenges. *ACS Medicinal Chemistry Letters*, 11:1506–1513, 8 2020.

[15] Robert P Goldman, Robert Moseley, Nicholas Roehner, Breschine Cummins, Justin D Vrana, Katie J Clowers, Daniel Bryce, Jacob Beal, Matthew Dehaven, Joshua Nowak, Trissha Higa, Vanessa Biggers, Peter Lee, Jeremy P Hunt, Lorraine Mosqueda, Steven B Haase, Mark Weston, George Zheng, Anastasia Deckard, Shweta Gopaulakrishnan, Joseph F Stubbs, Niall I Gaffney, Matthew W Vaughn, Narendra Maheshri, Ekaterina Mikhalev, Bryan Bartley, Richard Markeloff, Tom Mitchell, Tramy Nguyen, Daniel Sumorok, Nicholas Walczak, Chris Myers, Zach Zundel, Benjamin Hatch, James Scholz, and John Colonna-Romano. Highly-automated, high-throughput replication of yeast-based logic circuit design assessments, 2022.

[16] John M. Hatcher, Guang Yang, Li Wang, Scott B. Ficarro, Sara Buhrlage, Hao Wu, Jarrod A. Marto, Steven P. Treon, and Nathanael S. Gray. Discovery of a selective, covalent irak1 inhibitor with antiproliferative activity in myd88 mutated b-cell lymphoma. *ACS Medicinal Chemistry Letters*, 11:2238–2243, 11 2020.

[17] James R. Henry, Michael D. Kaufman, Sheng-Bin Peng, Yu Mi Ahn, Timothy M. Caldwell, Lakshminarayana Vogeti, Hanumaiah Telikepalli, Wei-Ping Lu, Molly M. Hood, Thomas J. Rutkoski, Bryan D. Smith, Subha Vogeti, David Miller, Scott C. Wise, Lawrence Chun, Xiaoyi Zhang, Youyan Zhang, Lisa Kays, Philip A. Hipskind, Aaron D. Wrobleski, Karen L. Lobb, Julia M. Clay, Jeffrey D. Cohen, Jennie L. Walgren, Denis McCann, Phenil Patel, David K. Clawson, Sherry Guo, Danalyn Manglicmot, Chris Groshong, Cheyenne Logan, James J. Starling, and Daniel L. Flynn. Discovery of 1-(3,3-dimethylbutyl)-3-(2-fluoro-4-methyl-5-(7-methyl-2-(methylamino)pyrido[2,3-d]pyrimidin-6-yl)phenyl)urea (ly3009120) as a pan-raf inhibitor with minimal paradoxical activation and activity against braf or ras mutant tumor cells. *Journal of Medicinal Chemistry*, 58:4165–4179, 2015.

[18] Ian Holland and Jamie A. Davies. Automation in the life science research laboratory, 11 2020.

[19] Scott A. Hollingsworth and Ron O. Dror. Molecular dynamics simulation for all, 9 2018.

[20] Mona M. Hosseini, Stephen E. Kurtz, Sherif Abdelhamed, Shawn Mahmood, Monika A. Davare, Andy Kaempf, Johannes Elferich, Jason E. McDermott, Tao Liu, Samuel H. Payne, Ujwal Shinde, Karin D. Rodland, Motomi Mori, Brian J. Druker, Jack W. Singer, and Anupriya Agarwal. Inhibition of interleukin-1 receptor-associated kinase-1 is a therapeutic strategy for acute myeloid leukemia subtypes. *Leukemia*, 32:2374–2387, 11 2018.

[21] Muhammad Jahangir Hossen, Woo Seok Yang, Daewon Kim, Adithan Aravinthan, Jong Hoon Kim, and Jae Youl Cho. Thymoquinone: An irak1 inhibitor with in vivo and in vitro anti-inflammatory activities. *Scientific Reports*, 7, 2 2017.

[22] J. P. Hughes, S. S. Rees, S. B. Kalindjian, and K. L. Philpott. Principles of early drug discovery, 3 2011.

[23] Fergus Imrie, Anthony R Bradley, and Charlotte M Deane. Generating property-matched decoy molecules using deep learning. *Bioinformatics*, 02 2021.

[24] José Jiménez, Miha Škalič, Gerard Martínez-Rosell, and Gianni De Fabritiis. Kdeep: Protein-ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of Chemical Information and Modeling*, 58:287–296, 2 2018.

[25] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A. Shoemaker, Paul A. Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E. Bolton. Pubchem 2019 update: Improved access to chemical data. *Nucleic Acids Research*, 47:D1102–D1109, 1 2019.

[26] Jonathan Knowles and Gianni Gromo. Target selection in drug discovery, 1 2003.

[27] David Ryan Koes, Matthew P. Baumgartner, and Carlos J. Camacho. Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise. *Journal of Chemical Information and Modeling*, 53:1893–1904, 8 2013.

[28] Andrei Kouranov, Lei Xie, Joanna de la Cruz, L. Chen, John Westbrook, Philip E. Bourne, and Helen M. Berman. The rcsb pdb information portal for structural genomics. *Nucleic acids research*, 34, 2006.

[29] C Lamagna, M Chan, E Tai, S Siu, R Frances, S Yi, C Young, V Markovtsov, Y Chen, L Chou, G Park, E Masuda, and V Taylor. Op0133 preclinical efficacy of r835, a novel irak1/4 dual inhibitor, in rodent models of joint inflammation. *Annals of the Rheumatic Diseases*, 79:86, 2020.

[30] Zhihai Liu, Yan Li, Li Han, Jie Li, Jie Liu, Zhixiong Zhao, Wei Nie, Yuchen Liu, and Renxiao Wang. Pdb-wide collection of binding data: Current status of the pdbbind database. *Bioinformatics*, 31:405–412, 7 2015.

[31] Eduardo Habib Bechelane Maia, Letícia Cristina Assis, Tiago Alves de Oliveira, Alisson Marques da Silva, and Alex Gutterres Taranto. Structure-based virtual screening: From classical to artificial intelligence, 4 2020.

[32] Andrew T. McNutt, Paul Francoeur, Rishal Aggarwal, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. Gnina 1.0: molecular docking with deep learning. *Journal of Cheminformatics*, 13, 12 2021.

[33] Rocco Meli, Garrett M. Morris, and Philip C. Biggin. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: A review. *Frontiers in Bioinformatics*, 2, 6 2022.

[34] N. Arul Murugan, Gnana Ruba Priya, G. Narahari Sastry, and Stefano Markidis. Artificial intelligence in virtual screening: Models versus experiments, 7 2022.

[35] Alleyn T. Plowright, Craig Johnstone, Jan Kihlberg, Jonas Pettersson, Graeme Robb, and Richard A. Thompson. Hypothesis driven drug design: Improving quality and effectiveness of the design-make-test-analyse cycle, 1 2012.

[36] Jacob T Rapp, Bennett J Bremer, and Philip A Romero. Self-driving laboratories to autonomously navigate the protein fitness landscape. *bioRxiv*, 2023.

[37] Michael G. Rolf, Jon O. Curwen, Margaret Veldman-Jones, Cath Eberlein, Jianyan Wang, Alex Harmer, Caroline J. Hellawell, and Martin Braddock. In vitro pharmacological profiling of r406 identifies molecular targets underlying the clinical effects of fostamatinib. *Pharmacology Research and Perspectives*, 3, 10 2015.

[38] Semion K. Saikin, Christoph Kreisbeck, Dennis Sheberla, Jill S. Becker, and A. Aspuru-Guzik. Closed-loop discovery platform integration is needed for artificial intelligence to make an impact in drug discovery, 1 2019.

[39] Gisbert Schneider. Automating drug discovery. *Nature Reviews Drug Discovery*, 17:97–113, 2018.

[40] Petra Schneider, W. Patrick Walters, Alleyn T. Plowright, Norman Sieroka, Jennifer Listgarten, Robert A. Goodnow, Jasmin Fisher, Johanna M. Jansen, José S. Duca, Thomas S. Rush, Matthias Zentgraf, John Edward Hill, Elizabeth Krutoholow, Matthias Kohler, Jeff Blaney, Kimito Funatsu, Chris Luebkemann, and Gisbert Schneider. Rethinking drug design in the artificial intelligence era, 5 2020.

[41] Adrian M. Schreyer and Tom Blundell. Usrcat: Real-time ultrafast shape recognition with pharmacophoric constraints. *Journal of Cheminformatics*, 4, 11 2012.

[42] Valentin Steinwandter, Daniel Borchert, and Christoph Herwig. Data science tools and applications on the way to pharma 4.0, 9 2019.

[43] Marta M. Stepniewska-Dziubinska, Piotr Zielenkiewicz, and Pawel Siedlecki. Decaf-discrimination, comparison, alignment tool for 2d pharmacophores. *Molecules (Basel, Switzerland)*, 22, 7 2017.

[44] Jocelyn Sunseri and David Ryan Koes. Pharmit: interactive exploration of chemical space. *Nucleic Acids Research*, 44(W1):W442–W448, 04 2016.

[45] Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus Robert Müller. Machine learning force fields, 8 2021.

[46] Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, and Shanrong Zhao. Applications of machine learning in drug discovery and development, 6 2019.

[47] Izhar Wallach, Michael Dzamba, and Abraham Heifets. Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. 10 2015.

[48] Li Wang, Qi Qiao, Ryan Ferrao, Chen Shen, John M. Hatcher, Sara J. Buhrlage, Nathanael S. Gray, and Hao Wu. Crystal structure of human irak1. *Proceedings of the National Academy of Sciences of the United States of America*, 114:13507–13512, 12 2017.

[49] Zhen Ning Wee, Siti Maryam J.M. Yatim, Vera K. Kohlbauer, Min Feng, Jian Yuan Goh, Bao Yi, Puay Leng Lee, Songjing Zhang, Pan Pan Wang, Elgene Lim, Wai Leong Tam, Yu Cai, Henrik J. Ditzel, Dave S.B. Hoon, Ern Yu Tan, and Qiang Yu. Irak1 is a therapeutic target that drives breast cancer metastasis

and resistance to paclitaxel. *Nature Communications*, 6, 10 2015.

[50] Scott A. Wildman and Gordon M. Crippen. Prediction of physicochemical parameters by atomic contributions. *Journal of Chemical Information and Computer Sciences*, 39, 1999.

[51] David S. Wishart, Yannick D. Feunang, An C. Guo, Elvis J. Lo, Ana Marcu, Jason R. Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam MacIejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, DIana Le, Allison Pon, Craig Knox, and Michael Wilson. Drugbank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Research*, 46:D1074–D1082, 1 2018.

[52] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, Andrew Palmer, Volker Settels, Tommi Jaakkola, Klavs Jensen, and Regina Barzilay. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019. PMID: 31361484.

[53] Xiangxiang Zeng, Xinqi Tu, Yuansheng Liu, Xiangzheng Fu, and Yansen Su. Toward better drug discovery with knowledge graph, 2 2022.

[54] Hao Zhu. Big data and artificial intelligence modeling for drug discovery. *Annual Review of Pharmacology and Toxicology*, 2019.

[55] Hui Zhu, Yulin Zhang, Wei Li, and Niu Huang. A comprehensive survey of prospective structure-based virtual screening for early drug discovery in the past fifteen years, 12 2022.

[56] Tian Zhu, Shuyi Cao, Pin Chih Su, Ram Patel, Darshan Shah, Heta B. Chokshi, Richard Szukala, Michael E. Johnson, and Kirk E. Hevener. Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis, 9 2013.

Appendix A SpectraView

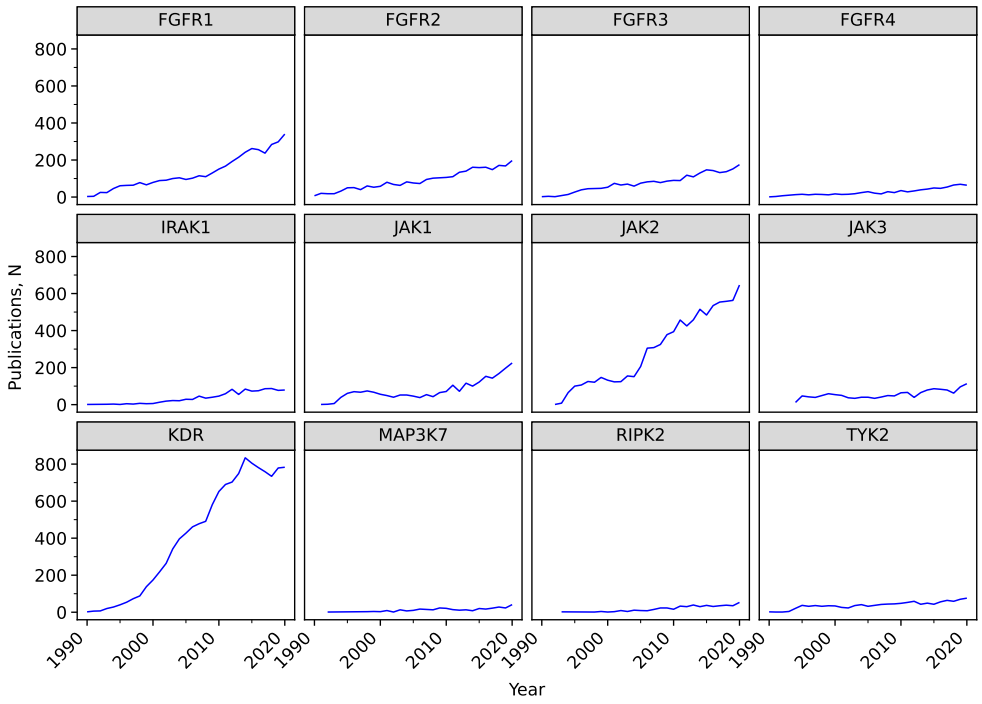


Fig. A1: PubMed-indexed publication trends over the last 3 decades for each of the considered targets.

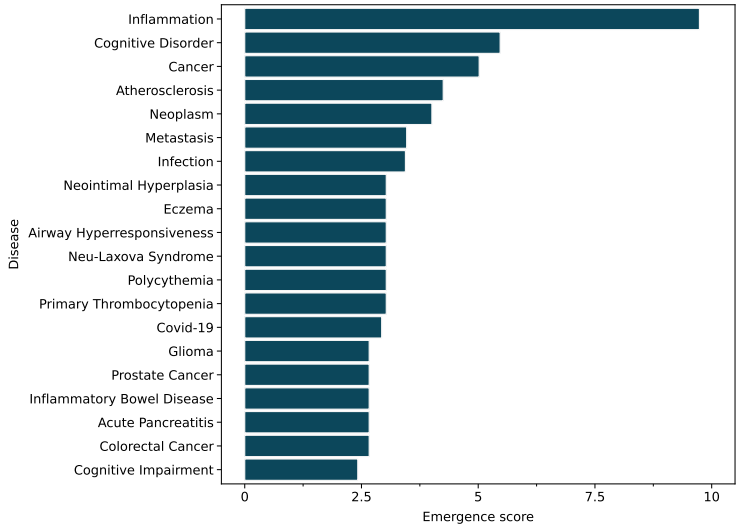


Fig. A2: Emergence scores (ES) for different diseases linked to IRAK1. The emergence score is calculated as a maximum increase in the number of publications per disease scaled by the total publication volume and recency.

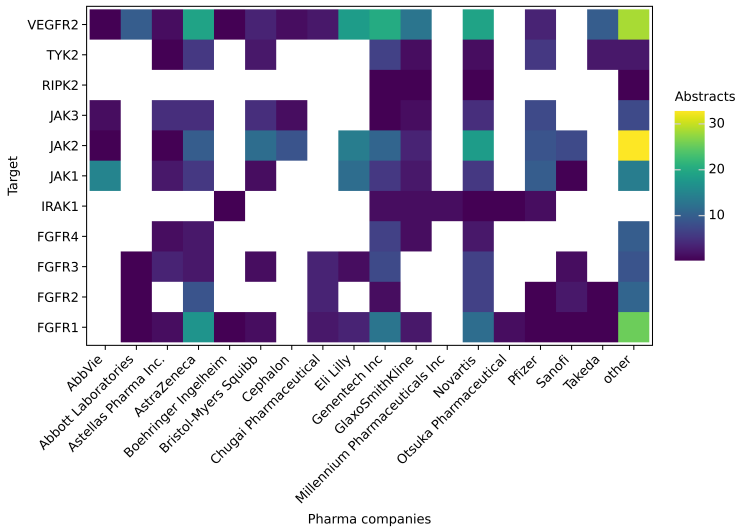


Fig. A3: Total number of PubMed-indexed publications with pharma company affiliations that at least one of the considered targets.

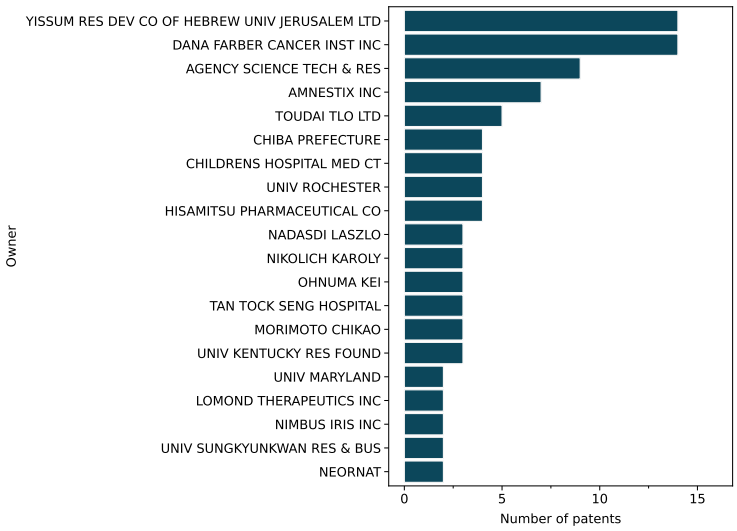


Fig. A4: Top organizations that own patents or patent applications that mention IRAK1.

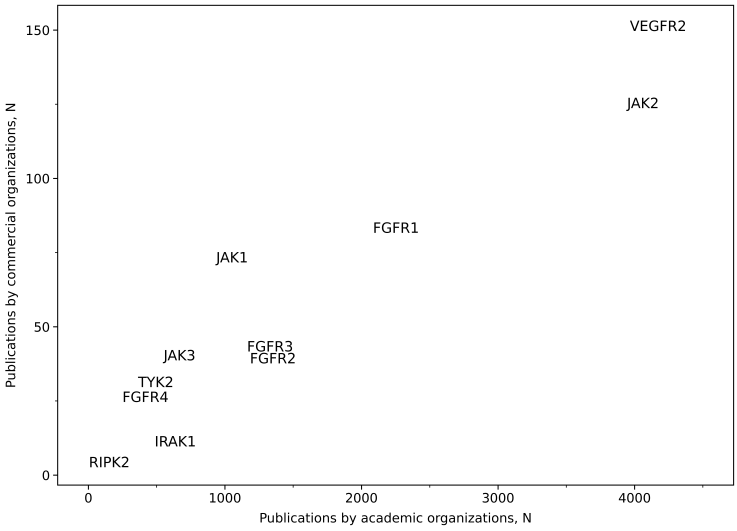


Fig. A5: The correlation between the number of PubMed-indexed publications affiliated with academic or pharma organizations for each of the considered targets.