Title

CACTI: an in silico chemical analysis tool through the integration of chemogenomic data and clustering analysis.

Journal

Journal of Cheminformatics, 16(1)

ISSN

1758-2946

Authors

Godinez-Macias, Karla

Winzeler, Elizabeth

Publication Date

2024-07-24

DOI

10.1186/s13321-024-00885-2

Godinez-Macias and Winzeler
Journal of Cheminformatics (2024) 16:84
https://doi.org/10.1186/s13321-024-00885-2

Journal of Cheminformatics

RESEARCH

Open Access

CACTI: an in silico chemical analysis tool through the integration of chemogenomic data and clustering analysis

Karla P. Godinez-Macias[1] and Elizabeth A. Winzeler[1*]

Abstract

It is well-accepted that knowledge of a small molecule's target can accelerate optimization. Although chemogenomic databases are helpful resources for predicting or finding compound interaction partners, they tend to be limited and poorly annotated. Furthermore, unlike genes, compound identifiers are often not standardized, and many synonyms may exist, especially in the biological literature, making batch analysis of compounds difficult. Here, we constructed an open-source annotation and target hypothesis prediction tool that explores some of the largest chemical and biological databases, mining these for both common name, synonyms, and structurally similar molecules. We used this Chemical Analysis and Clustering for Target Identification (CACTI) tool to analyze the Pathogen Box collection, an open-source set of 400 drug-like compounds active against a variety of microbial pathogens. Our analysis resulted in 4,315 new synonyms, 35,963 pieces of new information and target prediction hints for 58 members.

Scientific contributions

With the employment of this tool, a comprehensive report with known evidence, close analogs and drug-target prediction can be obtained for large-scale chemical libraries that will facilitate their evaluation and future target validation and optimization efforts.

Keywords Target prediction, Scaffold clustering, Machine learning, Neglected disease

Introduction

Understanding how a drug interacts with its biological target and whether its inhibition results in the desired phenotypic response is a critical step in drug discovery [1–3]. One area where this is particularly true is for infectious diseases where many of the hits that are advanced for drug discovery come from phenotypic, organismal screens as described in a recent review [4]. For malaria parasites, tens of thousands of compounds with parasite killing activity have been placed in the public domain. This has led to medicinal chemistry programs and molecules entering clinical trials [4, 5]. While understanding the mechanism of action or knowing the target is not strictly essential for a compound series to advance, knowing the target and ideally having a crystal structure for the target can make subsequent medicinal chemistry optimization much more efficient: fewer compounds need to be made and evaluated, and biochemical assays are typically more robust and lower cost. Furthermore, screening and assay costs can be many orders of magnitude lower for a biochemical target relative to whole organism work.

Because of the huge-cost savings that can come from knowing a target, substantial effort is often invested into target discovery. Biochemical and genetic approaches

*Correspondence:
Elizabeth A. Winzeler
ewinzeler@health.ucsd.edu
[1] Department of Pediatrics, University of California, San Diego, School of Medicine, La Jolla, CA 92093, USA

(e.g. in vitro evolution) are two examples of methods used for target identification. However, these tend to be lengthy and are resource-consuming [4, 5]. For malaria parasites, in vitro evolution has been a successful method but can often take six months or more [4]. It thus makes sense to accomplish as much computationally as possible. For example, molecular docking is commonly pursued to predict small molecule ligand binding to key therapeutic proteins prior to biochemical or phenotypic assays, and though it tends to have a high false positive rate [6], drug targets for some antimalarial inhibitors have been identified through this approach [7–9]. Similarly, virtual screening is a commonly applied. Singh et al. [10] describes how this approach can drastically reduce the chemical space to provide a focused library for subsequent biochemical high-throughput screening (HTS), although it is most reliable when knowledge of compound activity is available and libraries against a target of interest are accessible. Antimalarial compounds have also been elucidated through virtual screening [11–13]. Comparable in silico approaches also exist for target identification. For example, the tool TargetHunter by Wang et al. [14], is a web-based prediction approach that incorporates analog bioactivity data from ChEMBL [15]. Likewise, Chemmine, is an online resource by Backman et al. [16] that predicts targets based on similar records in PubChem database [17]. More recently, Jimenes-Vargas et al. [18] showed the power of artificial intelligence in drug discovery. Here, authors predict compound-target interactions based on different molecular description calculations and machine learning algorithms using ChEMBL records.

A constraint for these resources is the limit of one molecule per search and one chemogenomic database. Modern high-throughput screens may involve millions of compounds and thousands of primary hits. Having high-throughput methods to rapidly assess many hits can allow prioritization of compounds for resupply and resynthesis, which is particularly important for academic researchers that may have limited access to compound management groups or limited medicinal chemistry capacity. For example, in malaria parasites, many dihydrofolate reductase (DHFR) inhibitors have been compromised by widespread resistance [19]. It thus makes sense to identify phenotypic hits acting against DHFR prior to compound resupply. In fact, in silico approaches have proven to be a valuable in the analysis of large libraries, especially when various resources are utilized together to reduce the number of compound resupply and experimental testing needed [20, 21].

To address the need of an automated multi-compound analysis tool, particularly for neglected diseases, we constructed a pipeline named CACTI (Chemical Analysis and Target Identification), to provide comprehensive searches in chemogenomic databases and provide biological targets clues with single or bulk queries, integrating data not only from major databases such as ChEMBL and PubChem, but also from ligand-based databases (e.g. BindingDB [22]), as well as scientific and patent evidence (e.g. PubMed [23] and SureChEMBL [24]). Furthermore, to account for the difference in molecule identifiers across databases, we implemented a cross-reference method to map a given identifier based on chemical similarity scores and other known identifiers, or synonyms, in the expanded search. With this process, we provide a comprehensive large-scale study report incorporating all available identifiers for each small molecule, its close analogs, and available bioactivity data and/or mechanism of action (MoA) if known.

Materials and methods

Chemogenomic database selection and accessing

A common first step to evaluate a small molecule as a potential therapeutic, includes the exploration of chemical and biological space to identify known information and potential target. In contrast to most available tools that focus on one database for target-predictions [14, 25–27] we explored multiple chemogenomic databases as knowledge source. The solely criterion for database selection was the availability of REST API, a data request protocol for query and transfer. For chemical- and experimental-based information sources, we selected the well-established ChEMBL and PubChem databases. BindingDB was selected for protein–ligand evidence and, EMBL-EBI and PubMed as sources for comprehensive searches (Table 1). Although each bears unique data, we observed some overlap in data content, especially for literature evidence. However, it is important to note the indexing type and database specification prior to integration, as not every record is curated, and query fields could have different meaning. For example, experimental data in PubChem could be from biological assays (dose–response, phenotypic, or target-based) or biochemical assay (profiling or binding), while BindingDB refers to binding affinities assays solely.

After database selection, we created a custom function to access and retrieve data from each source. Following the corresponding REST API web services, we constructed an html query link with the database domain + web service architecture + data of interest + parameters. Specifically, we used chembl_id_lookup, mechanism, document, compound_record, molecule/synonym and similarity data functions from ChEMBL. For PubChem we used compound's data SMILES, name, CID, PubMedID, PatentID, fastsimilarity_2d, synonyms and MolecularWeight functions. PubMed records were

Table 1 Approximate number of records by entity type for selected chemogenomic and literature databases

Database	Chemical Data	Experimental Data	Literature	Patent	Compound- Pathway/Gene
ChEMBL	22,399,743	20,334,684	88,630		15,398
SureChEMBL				> 17 M	
PubChem	115,036,406	304,160,942	38,884,316	43,059,414	Pathway: 240,631
BindingDB	1,164,246	2,707,699	39,818		9023
EMBL-EBI			34,135,964		
PubMed			~ 34 M		

found using the PMID and PMC xref from PubChem, and europepmc/search function from EMBL-EBI server. The command, GetTargetByCompound, was used to access BindingDB database. For exact URL query construction please refer to the database web-services [22–24, 28, 29].

Querying and standardization of chemogenomic databases using SMILES

We reasoned that, to investigate a set of query compounds, the first step was to explore hidden patterns among the various chemogenomic databases containing large datasets of small molecules and their bioactivity. However, these patterns are difficult to reveal when searching across databases, mainly due to the lack of indexing standardization and the existence of multiple equivalent representations of a chemical structure (SMILES [30]). For example, ethanol can be encoded with SMILES OCC, CCO and C(O)C. Furthermore, in ChEMBL the SMILES CCO corresponds to CHEMBL545 but C(O)C is not a valid query, and vice versa the notation C(O)C corresponds to CID702 in PubChem while CCO is not valid. Nevertheless, to test the hypothesis that using SMILES as identifier is sufficient for mapping the chemical space, we used the provided query SMILES as first input (Fig. S1A). To reveal the index identity of each query, we used the custom functions to cross-reference across databases using 100% similarity match, and further confirmed their exact identity by comparing Morgan fingerprints as described in our Methods.

Due to submission discrepancies across repositories, query identifiers were combined in a "synonym" (common names) column and filtered if they were (1) numerical with no indication of external source origin, (2) IUPAC name from an unreliable source, or (3) duplicated when converted to upper case and removal of special characters such as ":" or "–". The remaining synonyms were used to retrieve scientific evidence, patent evidence and additional useful information associated with the compound of interest. We reasoned that expanding the search to include these identifiers would provide more exhaustive research on the query compounds, improving the target-prediction steps when using the additional pieces of evidence. With this filtering, invalid or duplicated query records were removed and bioactivity data, naming synonyms, scholarly evidence, and chemical information across selected chemogenomic databases was integrated.

Chemical comparison through similarity calculations

In addition to mining with SMILES and common names, we expanded the search to include closely related analogs (Fig. S1B). We first used RDKIT v.2024.03.1 [31] to convert the query SMILES to a canonical form, generating a unique notation to be queried and compared with equal features. The standardized format (canonical) was used to identify analogs by transforming them to a binary representation called fingerprints (Morgan fingerprints), which allows chemical similarity calculations. These similarities were computed using the Tanimoto coefficient [32], T, which measures the degree of similarity between a query and target structure as follows:

$$T = \frac{N_{AB}}{N_A + N_B - N_{AB}}$$

where A represents the query fingerprint and B the target fingerprint, N_A and N_B represent the number of "1-bit" in A and B respectively and N_{AB} represents the "1-bit" shared in both. To ensure consistency, analogs were transformed to the same canonical form as the query and filtered if the T score was below 80% threshold (score ranging from 0 to 1 transformed to percentage). This conversion allowed us to further reduce discrepancies between SMILES conversion while identifying chemical analogs. We reasoned that by selecting close analogs with an 80% similarity, as opposed to the suggested 85% threshold by Patterson et al. study [33], we may be able to retain useful moieties related to antimalarial activity (such as functional groups), while maintaining a high degree of chemical similarity between the query and analog compound.

Although a target may have been reported for a given compound, the query compound might be similar but not identical and our goal was to capture these associations. To identify scaffold families, we constructed a similarity network (including the query dataset and searched analogs) by assigning nodes to chemical entities and edges between entities with a Tanimoto coefficient above 80%. With this network we were able to identify query drug-targets from analogs with known mechanisms of action or gene target, elucidating relationships that could be obscured otherwise.

Drug target identification

In order to provide drug-targets clues, we acquired and incorporated several large datasets with annotated compound-target pairs and created a module to search these sets for compounds closely related to query molecules (Fig. S1C). We used the Novartis Chemogenetic Library [34] assembly consisting of 4185 compounds annotated with their primary mammalian gene target, an in-house, consolidated file of 163 validated antimalarials originating from phenotypic screens with their drug target, a collection of 218 licensed well-validated therapeutic drugs extracted from well-established databases such as IUPHAR/BPS pharmacology database [35] or the FDA approved drugs list [36], and a set of 157 known antibacterial inhibitors [37] from which seven (azithromycin, mupirocin, dapsone, sulfalene, sulfadiazine, triclosan, sulfamethoxazole) overlap with the antimalarial set. The assembled list can be accessed at https://github.com/winzeler-lab/CACTI metadata folder.

To reduce biases when comparing large datasets with molecules of wide mass range, we implemented a method to partition the query and target dataset by molecular weight (default cutoff of 500 Da) and convert each into fingerprints for subsequent similarity calculations. We then implemented a function to allow selection of fingerprint conversion algorithm (RDKFingerprint or GetMorganFingerprintAsBitVect) and selection of binary vector size, to prevent biased structure pattern identifications. Once fingerprints were calculated, we created a NxM matrix, where N and M is the number of fingerprints in the dataset and calculated Tanimoto coefficients for every pair, followed by clustering molecules based on their score (default 0.8).

CACTI, a large-scale chemical compound analysis tool

As a final step we aimed to integrate the chemical querying and chemical comparison approaches to construct a compound analysis pipeline that will systematically (1) identify index identifiers for a set of given compounds, (2) consolidate relevant information known for index identifiers and identify close chemical analogs, and (3) identify scaffold families to provide target hypothesis with collected index annotations, analogs and known drug-target information. (Fig. 1). The pipeline, CACTI, is accessible at https://github.com/winzeler-lab/CACTI. Shortly, we built the customizable pipeline with python version 3.12 currently executable with UNIX, allowing for single or bulk queries and the selection of analyses and parameters to be performed. After submission validation of the query SMILES set (Fig. 1A), the first step

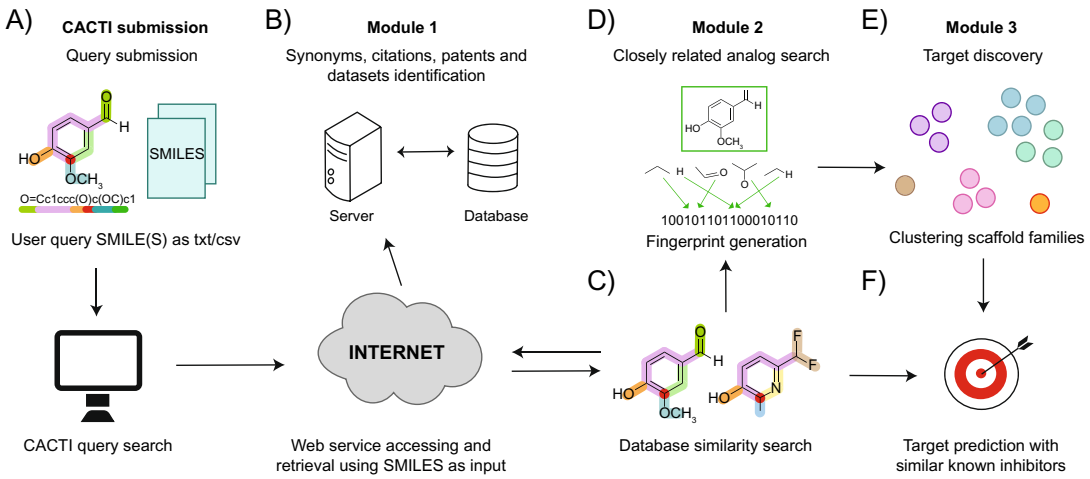

Fig. 1 CACTI drug target prediction workflow. Pipeline workflow steps are illustrated by modules where **A** represents the user SMILES set as input in a tabular or comma delimited, **B** refers to the database querying and retrieving steps using SMILES to identify synonyms, citation evidence, patents, membership in datasets and additional information for target prediction. Module 2 refers to **C** the identification of close analogs in chemogenomic databases to the query set, and **D** transformation from SMILES to binary fingerprint for scaffold comparison. In module 3, **E** clustering chemical entities based on similarity measures is performed and **F** target predictions are completed combining information from previous modules and validated known-target datasets

in the pipeline was to use the chemical and biological space exploration method to obtain the multiple identifiers from which a single compound is known, including IUPAC and common name(s), as well as peer-reviewed publications, deposited datasets, and patents associated to each common name (Fig. 1B). To identify analogs in the public domain that might have a known target, we then retrieved the list of similar scaffolds to the ones of interest (Fig. 1C) with at least 80% similarity. This step allowed us to increase the likelihood of accurately uncovering drug-target pairs or providing prediction starting points.

To better understand the query molecules and relationships within the dataset and public domain using the acquired genomic and bioactivity annotations, the next step in the pipeline includes the chemical comparison methods. SMILES for the query set and analogs are transformed to binary fingerprints (Fig. 1D) to allow comparisons between them, and clustered together to find core scaffolds that are similar to at least 80% to each other (Fig. 1E). Although this step seems redundant after querying for similar structures in Fig. 1C, performing this extra validation was valuable to discover if core scaffolds from the analog subset is shared by two or more query compounds. Lastly, to predict a drug-target under the molecular similarity principal (Fig. 1F), scholar references, known mechanisms of action or gene targets, and bioactivity data associated to compound members within each cluster are compared. It is important to mention that in case of compound analog reacting noncovalently against a biological target yet considered active, typically referred as pan-assay interference compounds (PAINS), target hypotheses may be misguided for a query compound based solely on the bioactivity annotations. Thus, careful inspection may be needed when evaluating members from the cluster and additional annotations may need to be considered.

Results

To facilitate the exploration of the chemical and biological space for large datasets, and to provide a better understanding of chemical structures and their impact on a biological system, we constructed a drug-target prediction tool using data mining and chemoinformatic techniques. Our overall objective was to automate the tedious molecule-by-molecule searching that occurs when hits from a high throughput phenotypic screen are initially evaluated, and to perform this search in a comprehensive, unbiased fashion, looking for information not only on the query molecules but also on closely-related analogs. Because we noted that molecules might have different common names in the literature, we also sought to systematically uncover synonyms. For example, the new imidazolopiperazine antimalarial named ganaplacide that is in phase III clinical trials has been known as KAF156, and GNF156 and substantial work has been published on the closely related scaffold named GNF179 which differs from KAF156 by a single halogen substitution. Our motivation was thus to provide a tool that would allow the user to focus energy on phenotypic screening hits that might have novel mechanisms of action and thus avoid compounds with a well-established mechanism or known drug resistance liabilities. In addition, we wished to create a tool that could alert the user to relevant information from work on other species. If a compound is a dihydrofolate reductase inhibitor in humans, it is likely also a dihydrofolate reductase inhibitor in malaria parasites [38].

As described in the methods, this tool, which we have named CACTI consists of 3 modules, was coded into python and accepts queries in the forms of tabular or comma-delimited files containing the SMILES of query molecule(s), as well as additional field columns such as the standard name, if desired. The output consists of Excel files from each selected analysis (module) with a comprehensive report for all query molecules with synonyms and scholar evidence mined from the selected knowledge databases including ChEMBL, PubChem, BindingDB and PubMed. The output also includes a complete list of close analogs to query compounds and their similarity scores from the network construction module. Lastly, it includes a report from a clustering module in which query compounds are matched to similar molecules from the annotated target set. These last two reports can be exported to external tools for network visualizations, such as Cytoscape [39]. CACTI can be accessed from the public GitHub repository and executed by creating a local copy and following instructions indicated in the repository.

Case study: assessment of drug-target querying using the Pathogen Box dataset

In order to assess the performance of the tool, we applied it to the Pathogen Box dataset [40] from the Medicines for Malaria Venture (Fig. 2A). This set is a collection of 400 diverse drug-like molecules, with an average molecular weight of 374.13 Da, active against several neglected tropical diseases. This set was physically assembled by the Medicines for Malaria Venture and shared with users working in the neglected disease space to provide drug-like starting points for oral drug discovery. Accordingly, only 32 compounds failed to comply the standard Lipinski Rule of Five [41]. The set includes 26 positive controls from various scaffold families, 24 of which are part of known drug-target pairs. Controls include well known drugs like the antimalarials doxycycline, primaquine

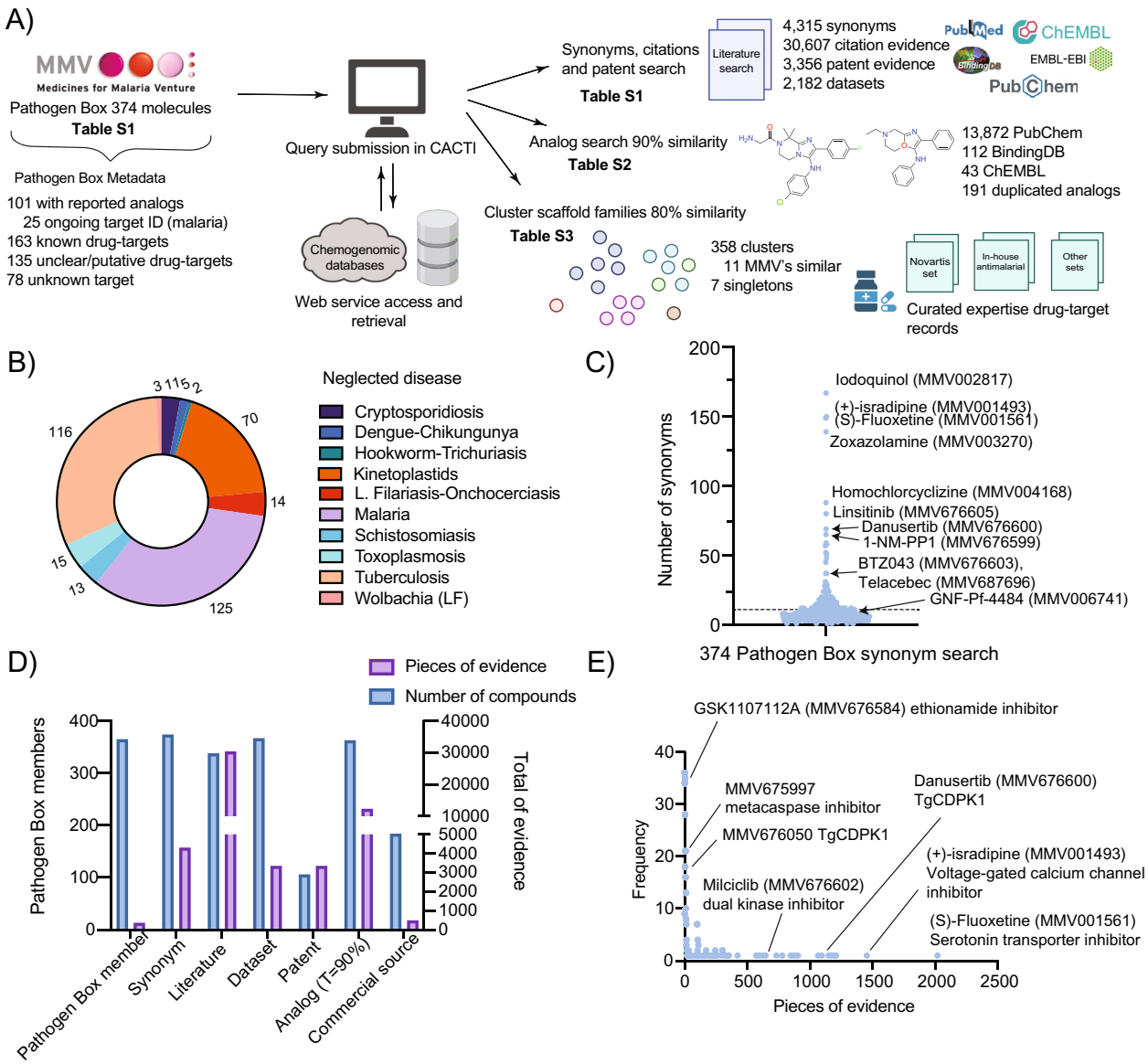

Fig. 2 Identification of synonyms, references, and close analogs for 374 Pathogen Box compounds. **A** Pathogen Box querying workflow. Results obtained from querying 374 Pathogen Box compounds against the pipeline. Supplementary tables related to each analysis are indicated along with the results from each module. **B** Membership of selected compounds. Number of compounds active against each neglected disease according to the Pathogen Box reference biological data [40]. **C** Synonyms identified. Number of names given for each Pathogen Box compound across different studies. Well established drugs belonging to the Pathogen Box are shown along with the common drug name. A dotted line denotes the average (n = 11) of synonym names per compound. **D** Statistics for identified information per compound. Identified categories for the 374 compounds are shown in the X axis. Left Y axis shows the number of Pathogen Box compounds found per category and the right Y axis shows the number of pieces of evidence per category. **E** Pieces of literature found per compound. Scatter plot showing the total pieces of evidence for each compound (X axis) against the frequency (Y axis). Known inhibitors with their inhibitory protein/enzyme are shown

and mefloquine [42], the antibacterial rifampicin [43], the antimicrobials pentamidine, nifurtimox and fluoxetine, the anti-schistosomiasis medicines, praziquantel and mebendazole. For these controls, we often found too much information, and these were thus excluded from further consideration. The remaining 374 compounds (Table S1) are ones that have reported activity in published phenotypic screens for neglected tropical disease, with the majority being antimalarial or antimycobacterial compounds [44–46] (Fig. 2B). Chemical properties for the 400 compounds and biological activity in multiple parasites and pharmacokinetic measurements, including cellular toxicity and pharmacokinetic measurements in HepG2 cells, can be found in ChEMBL-NTD

repository (https://chembl.gitbook.io/chembl-ntd) set 21.

Module 1 identifies 4,315 synonyms for the Pathogen Box compounds

A good literature search can prevent months of wasted effort, but searching biological databases such as PubMed is hampered by the fact that compounds frequently change names in biological publications. We assessed if we could find synonyms for the compounds in an automated way. Searching using CACTI identified 4315 synonyms for compounds in the Pathogen Box, that were not included in the initial Pathogen Box description, with an average of 11 names per compound (Table S1, Fig. 2C, D). For example, we found 10 synonyms for compound MMV006741, including GNF-Pf-4484, Maybridge3_001644, and BRD-K92165166-001-01-3, among others. Likewise, the automated synonym search revealed that one trivial name for MMV676600 is danusertib, a known pan-aurora kinase inhibitor [47]. Overall, we found 69 different names for MMV676600. These data also provided clues about the provenance of the compounds. The search revealed substantial overlap between the GSK TCAMS (Tres Cantos Anti-Malarial Set) library [48] (ChEMBL-NTD (https://chembl.gitbook.io/chembl-ntd), set 1) with 164 compounds present in both sets, 74 compounds overlapping with the St. Jude Biomedical library [49] (https://chembl.gitbook.io/chembl-ntd, set 3) and eight that were members of the Novartis GNF library [50] (https://chembl.gitbook.io/chembl-ntd, set 2). The automated search also revealed members from the Pathogen Box that are readily accessible to the scientific community. We found 206 compounds with suppliers' identification numbers for various commercial sources including AKOS, Zinc and MCULE. In general, module 1 was able to capture and identify diverse names from published studies and public libraries, as well as provide the unique identifiers for selected databases including PubChem (CID identifier) and ChEMBL (CHEMBL identifier).

Module 1 identifies new citations, datasets and patents for Pathogen Box molecules

No literature citations were provided with the initial Pathogen Box dataset. Therefore, we next used CACTI to comprehensively identify patent and literature citations using not just the MMV names but all synonyms (Fig. 1E). Unsurprisingly, searching with the MMV names often produced literature that described the Pathogen Box. However, searching with synonyms provided many more citations on the compounds. For example, MMV675997, is associated with 4 pieces of evidence that were deposited separately under several chemogenomic databases with different synonyms (MMV675997, CHEMBL1094051, BDBM50313769) (Table S1). Searching with these synonyms showed that this compound, which has a peptoid backbone and a nitrile P1′ warhead came from a library predicted to contain *T. brucei* Metacaspase enzyme inhibitor [51]. Likewise, the kinase inhibitor, danusertib (MMV676600) is associated with a total of 1195 pieces of evidence with citations indicating a possible role in inhibition of *Toxoplasma gondii* calcium-dependent protein kinase 1 CDPK1 (calcium-dependent protein kinase 1), as well as 100 pieces of patented evidence. From these, a CACTI PubMed search using PHA-739358, the additional synonym of MMV676600, yields three more literature references compared to MMV676600 (Table S1).

The literature search also showed that some of the compounds have activity in multiple species. For example, Murphy and colleagues [52] discovered that MMV676050 and MMV676182, both tested against cryptosporidiosis in the Pathogen Box activity profiling, are potent inhibitors targeting CDPK1 *T. gondii* with known crystal structure (PDB ID 3N51 [52]). Similarly, the literature search revealed that 162 members of the Pathogen Box showed evidence of a target validation in other models, such as cancer cell lines. For example, the chemotherapeutic milciclib (PHA-848125), currently in clinical trials, is a known dual cyclin-dependent and tropomyosin receptor kinase inhibitor [53, 54]. Overall, we found 30,425 new pieces of literature, 3356 patents and 2182 datasets associated with the Pathogen Box compounds. Though exploring the literature for different names may be straightforward for a small-scale study, as the study size increases, it becomes more difficult to capture these nuances. Therefore, the implemented method of acquiring and cross-referencing all known identifiers to retrieve data from the several chemogenomic databases provides a more detailed and complete report.

Module 2 identifies 12,383 new closely related compounds in the public domain that can be used for SAR by inventory

An important step for compound evaluation and testing is determining if there are closely related analogs available in the public domain that can be sourced for testing. Performing a Tanimoto similarity search for the 374 uncharacterized Pathogen Box members resulted in 12,383 close analogs ($T = 90\%$) from the selected chemical knowledge databases (Table S2, Fig. 2D). In contrast, only eleven compounds (MMV393995, MMV676269, MMV676442, MMV021057, MMV202553, MMV688754, MMV099637, MMV676050, MMV676064, MMV676204 and MMV676398) out of the 374 had no obvious analogs in the public domain under this criterion. Most analogs identified belonged to PubChem (12,246

analogs), BindingDB (112), or ChEMBL (25) databases. On average, 34 closely related analogs were identified for each member of the query set with an average similarity of 94.44%. For example, we found 55 analogs of MMV019721, a recently discovered acetyl-coenzyme A synthetase *Plasmodium* inhibitor [55]. Many of these analogs are commercially available.

Interestingly, we observed eight closely-related analogs that were identified twice when querying for the 374 Pathogen Box compounds (Table S2, Fig. S2). Finding a close analog for two different query compounds suggest the presence of similar core scaffolds and likely similar biological targets. For example, the compounds MMV595321 and MMV676477, share five different analogs (CID's 136636721, 136636828, 156278160, 156278182 and 156285860), most of which are part of an antiparasitic inhibitor optimization effort (patent WO-2021077102-A1). Another analog (CID 44526919), related to MMV023969 (TCMDC-134161) and MMV024035 (TCMDC-134227), inhibits *P. falciparum* growth by up to 97% ($IC_{50} < 2$ μM) [56]. Thus the related Pathogen Box compounds could be attractive drug starting points.

Module 3 for new target discovery hypotheses—guilt by association approaches

One of the most resource-consuming steps in drug discovery is target deconvolution. To create target predictions, we utilized the information collected from previous modules and clustered the 374 Pathogen Box compounds and 156 close analogs, with sets of 4,716 molecules with known mechanisms of action to create hypotheses about the function of Pathogen Box compounds. The "known target" set was derived from multiple sources, including 4,185 compound-target pairs assembled by Canham et al. [34], 150 known antibacterial drug-target pairs [37], and a collection set of 381 antimalarial compound-target pairs extracted from established databases such as IUPHAR [35] and PDB Ligands [57], in addition to validated target-ligand pairs originating from the TCAMS and GNF Novartis Malaria Box phenotypic screening libraries. Although the assembled library of known target pairs was enriched for antimalarial inhibitors, many have activity across the parasitic and bacterial diseases included in the study, such as atovaquone and doxycycline. It is worth noting that targets for 152 Pathogen Box compounds (Table S1) were already known and were already included in the known target set. Clustering the 374 and their 156 close analogs with 4,716 known targets resulted in 77 clusters using a similarity threshold of $T = 80\%$ (Table S3), with an average of 4 compounds per cluster, and 264 singletons (Fig. 3A). Out of the 77 clusters, twenty-five had two or more Pathogen Box molecules and 71 clusters contained one or more MOA/target or putative target annotations. From the 71, 20 have at least one Pathogen Box compound with unknown or unclear mechanism. We did not find a MOA/target prediction for six clusters (Fig. 3B).

To assign targets for the Pathogen Box compounds, we inspected members from each cluster and assigned a potential target/function for the "unknown" Pathogen Box compound based on their counterpart target evidence. From this, we found targets appear repeatedly at rates higher than expected (hypergeometric mean function $= 4.4 \times 10^{-3} - 2.6 \times 10^{-48}$) (Fig. 3C). For example, three clusters [55, 56] contain predicted dihydroorotate dehydrogenase (DHODH) inhibitors and six contain predicted cytochrome b inhibitors, of which cluster 62 have one Pathogen Box compound (MMV687807) with no known reported target in *Plasmodium* or *Mycobacterium*. We also found 9 clusters whose members contain an analog or known target set compound from two or more different annotations. This includes cluster 338 where close analogs to MMV687700 are known salicyl-AMP inhibitor and other members from the known drug target set are known human adenosine receptor agonists. Despite this apparent incongruity, all compounds in cluster 338 contain the basic adenosine scaffold and are adenosine analogs. Another example includes cluster 177 where MMV688283 and MMV687246 (a PfCDPK5 inhibitor [58]) are clustered with TCMDC-138293, a known *Plasmodium* DHODH inhibitor. This may suggest some polypharmacology or potential weaknesses in the automated annotations.

We also found cases where the predicted target was assigned based on evidence from a different parasite genus to the one initially tested. The diazine scaffold family (cluster 30) has two compounds that have demonstrated activity against parasite and mammalian methionine aminopeptidase-1b, an enzyme predicted to catalyze the removal of the N-terminal initiator methionine during protein synthesis in parasites and mammals (MMV084997/GNF-Pf-359). The group also contains the uncharacterized, anti-kinetoplastid Pathogen Box compounds MMV658993 and MMV658988. This association creates a hypothesis that MMV658993 and MMV658988 may target methionine aminopeptidase-1b in kinetoplasts.

Discussion

Neglected diseases (parasitic diseases, and to some extent, rare diseases), attract less commercial interest and significantly less overall funding than other diseases, such as cancer or diabetes. On the other hand, much of the small molecule data for well-funded diseases such as cancer lies in the private, well-curated databases that

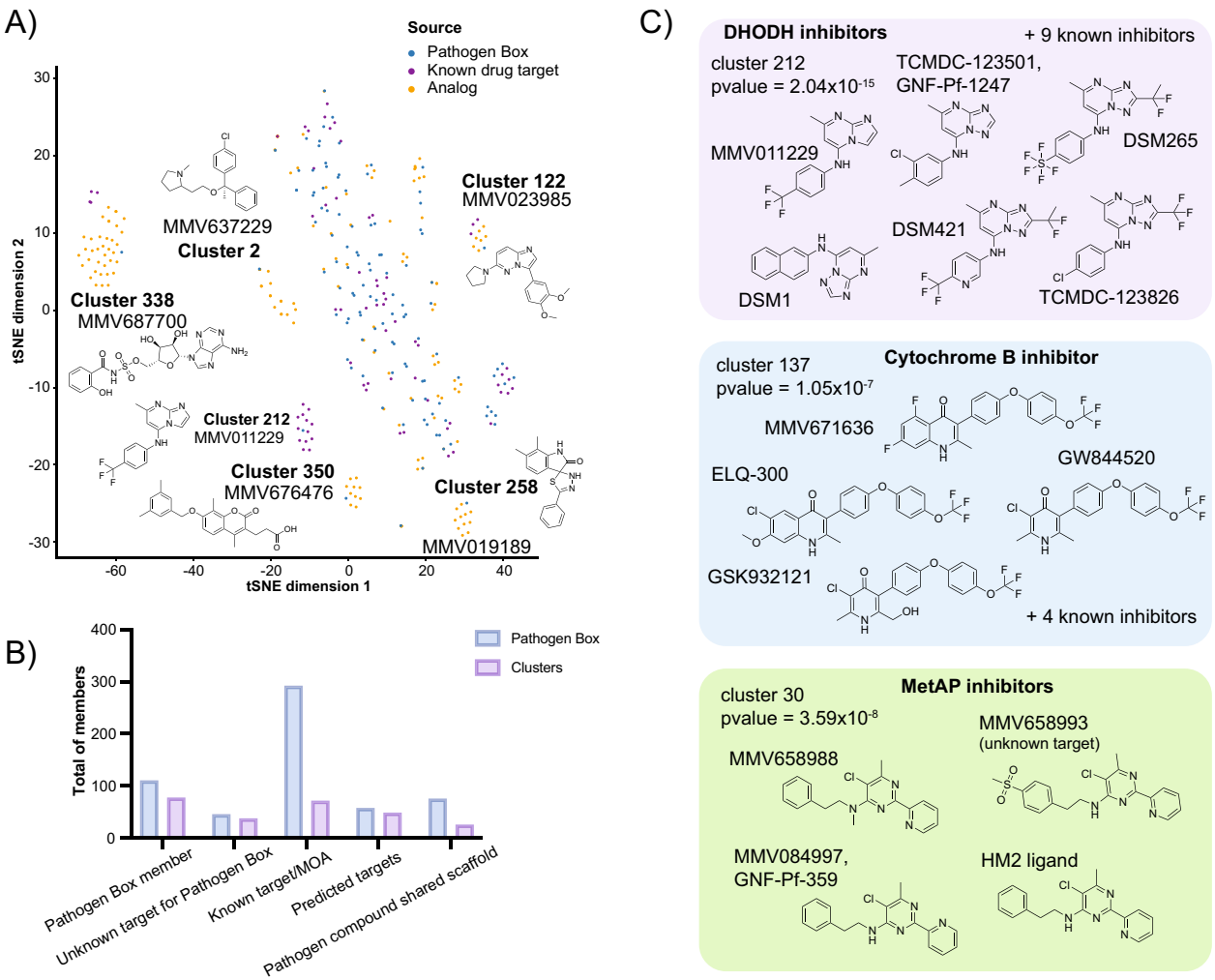

Fig. 3 Chemical clustering and target prediction. **A** Scaffold clustering. Structural similarity calculations with Pathogen Box library and known drug targets, as determined with RDKIT fingerprints. Clusters were assigned using 80% Tanimoto similarity threshold. Visualization of chemical space was performed using scikit-learn t-SNE function. **B** Scaffold clustering statistics. Total number of Pathogen Box members (light blue) and clusters (light purple) identified for the different categories shown in X axis. Total of known drug-target annotations, and predicted drug-targets, among all clusters are included in the summary categories. **C** Overrepresented clusters with known target. Examples of clusters with drug-targets found at rates greater than expected according to the hypergeometric mean function. Structures for cluster members are included

are maintained by pharmaceutical companies. These databases are often one of a company's biggest intellectual property assets. In contrast, for neglected disease, a larger proportion of the drug discovery data will be generated by academic researchers or public–private partnerships where much of it will ultimately be placed in the public domain. The availability of well curated public screening datasets (e.g. such as the Pathogen Box dataset) creates both opportunities to make new discoveries (e.g. for drug repurposing and target discovery) as well as challenges that relate to the dispersed and disorganized nature of the data. To make the task of assessing diverse data more accessible to neglected disease drug discovery researchers, we constructed a tool to pragmatically assess the major chemical databases and identify all data available for a query compound, as well as similarities between small molecule inhibitors in any system. With this approach a comprehensive report is generated and may be used to redirect laboratory resources in a more focused way, in addition to reducing the amount of work needed for drug-target or chemical optimization efforts.

Our approach has its limitations. First, our algorithm makes no assessment of literature quality. A manuscript reporting that a given compound binds a target may not rigorously assess the quality or strength of the binding. This may happen when a researcher develops

a biochemical assay for a target and then tests a library, such as the Pathogen Box library, against this target: In this case the best compounds from this exercise may be reported as potential inhibitors of the target in a publication. Our computational algorithms cannot catch these nuances and are out of scope; although a thorough report on a particular query set will be obtained, a human review is needed to revise and confirm the veracity/strength of acquired evidence. Perhaps not surprisingly, the data-querying strategy across multiple chemogenomic repositories is especially helpful when querying for small molecules that have been evaluated for activity against multiple disease indications. For example, compounds advancing to clinical-trials or those having a validated drug-target, are query entries that when searched provide the most complete information.

Another limitation is that our approach somewhat relies on the assumption that a compound will have the same target in *Trypanosoma cruzi* as in *Toxoplasma gondii*, despite the two species belonging to different eukaryotic phyla. Nevertheless, previous studies have highlighted that drug-target pairs tend to be conserved across species. A compound like methotrexate likely acts against dihydrofolate reductase in human and in malaria parasites. Cladosporin targets lysyl-tRNA synthetase in both yeast and malaria parasites [59]. There are hundreds of other examples of target conservation across species. Species selectivity thus comes from natural or engineered specificity, as well as innate ability of different species to detoxify compounds. On the other hand, just because a compound targets the electron transport chain in malaria parasites doesn't mean that it targets the electron transport chain in *M. tuberculosis* (e.g. such as in ELQ-300, with unreported target in tuberculosis) and predictions and caution is needed. As a future improvement to the pipeline, similar to gene ontology strategies (GO terms), querying and comparison of medical subject headings (MeSH terms) associated to compounds of interest (and analogs) can be implemented to mitigate the need to rely on similar structure function assumptions. This could increase the confidence of CACTI target identification hypotheses.

Another concern is that we relied on chemical scaffold clustering: This chemical comparison depends on the chemical fragment partition that is initially selected, resulting in potential arbitrary cluster assignment. In addition, clustering is less suitable for natural products, which tend to have more complex structures than smaller molecules. Increased size means partition methods favor fragment partitions that break down molecules and ignore the R groups that may be vital for the pharmacological effect of the natural product. Thus, although the provided scholarly report will be helpful, the clustering approach as scripted is less suitable for their evaluation. Nevertheless, we found that close analogs with similar drug-target matches did tend to group together even if the larger group often split into more than one cluster. We are confident that the established prediction method provides an initial hint for further target validation efforts.

A limitation with this approach includes the predictive component of the analysis. Though it is accepted that similar molecules tend to have similar mechanisms of action or drug-targets, experimental validation is needed to confirm binding to the desired target. The assumption that one small molecule will have just one target is more fraught as compounds get larger and where two different pharmacophores (e.g. an ATP-like molecule and a naphthalene) may be incorporated into one small molecule. In addition, for promiscuous inhibitors, like kinase inhibitors, our approach may provide limited resolution.

Another concern is that one component of this tool relies completely on network connectivity and database server availability (e.g. open query-request for selected databases) to query and retrieve literature and analogs from the public domain. To partially address this need, a future avenue includes the use of a periodic "database dump" that can be locally accessed and queried in lieu of internet access, when connectivity is unreliable or inaccessible.

Overall, we believe the identification of literature evidence as well as close analogs obtained through CACTI will aid researchers in early stages of drug discovery pipeline, and consequently allow more work on structure–activity relationship analysis. Similarly, the target prediction module implemented in this tool will serve as starting points for subsequent drug-target validation efforts and support the translation of genomic data into effective new drugs through the comparison of scaffold families. This automated tool shows a promising approach to quickly investigate multiple chemical entities in a single query, and prioritize hits for further exploration, especially in academic settings where compound management and resupply is limited.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-024-00885-2.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Acknowledgements
This work was funded by a Grant from the Bill and Melinda Gates Foundation (INV-039628). The authors thank MMV for their contribution of the Pathogen Box collection and making it available to the community.

Author contributions
E.A.W. conceived experiments, provided funding, provided supervision, wrote the manuscript, reviewed the manuscript, constructed supplemental files, performed analysis and designed figures. K.P.G.-M. wrote the first draft of the manuscript, reviewed the manuscript, constructed supplemental files, performed analysis and constructed figures.

Data availability
Data is provided within the manuscript or supplementary information files. CACTI is accessible via GitHub (https://github.com/winzeler-lab/CACTI) for download and personal use.

Declarations

Competing interests
The authors declare no competing interests.

Received: 5 March 2024 Accepted: 14 July 2024
Published online: 24 July 2024

References

1. Tabei Y, Pauwels E, Stoven V, Takemoto K, Yamanishi Y (2012) Identification of chemogenomic features from drug-target interaction networks using interpretable classifiers. Bioinformatics 28:i487–i494
2. Schenone M, Dancik V, Wagner BK, Clemons PA (2013) Target identification and mechanism of action in chemical biology and drug discovery. Nat Chem Biol 9:232–240
3. Zhou W, Wang Y, Lu A, Zhang G (2016) Systems pharmacology in small molecular drug discovery. Int J Mol Sci 17:246
4. Rao SPS, Manjunatha UH, Mikolajczak S, Ashigbie PG, Diagana TT (2023) Drug discovery for parasitic diseases: powered by technology, enabled by pharmacology, informed by clinical science. Trends Parasitol 39:260–271
5. Hovlid ML, Winzeler EA (2016) Phenotypic screens in antimalarial drug discovery. Trends Parasitol 32:697–707
6. Sink R, Gobec S, Pecar S, Zega A (2010) False positives in the early stages of drug discovery. Curr Med Chem 17:4231–4255
7. Chaniad P, Mungthin M, Payaka A, Viriyavejakul P, Punsawad C (2021) Antimalarial properties and molecular docking analysis of compounds from *Dioscorea bulbifera* L. as new antimalarial agent candidates. BMC Complement Med Ther 21:144
8. Owoloye A, Enejoh OA, Akanbi OM, Bankole OM (2020) Molecular docking analysis of *Plasmodium falciparum* dihydroorotate dehydrogenase towards the design of effective inhibitors. Bioinformation 16:672–678
9. Owoloye AJ, Ligali FC, Enejoh OA, Musa AZ, Aina O, Idowu ET, Oyebola KM (2022) Molecular docking, simulation and binding free energy analysis of small molecules as PfHT1 inhibitors. PLoS ONE 17:e0268269
10. Singh N, Chaput L, Villoutreix BO (2021) Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. Brief Bioinform 22:1790–1818
11. de Sousa ACC, Combrinck JM, Maepa K, Egan TJ (2020) Virtual screening as a tool to discover new beta-haematin inhibitors with activity against malaria parasites. Sci Rep 10:3374
12. Godara P, Reddy KS, Sahu W, Naik B, Srivastava V, Das R, Mahor A, Kumar P, Giri R, Anirudh J, Tak H, Banavath HN, Bhatt TK, Goyal AK, Prusty D (2023) Structure-based virtual screening against multiple *Plasmodium falciparum* kinases reveals antimalarial compounds. Mol Divers
13. Uddin A, Gupta S, Mohammad T, Shahi D, Hussain A, Alajmi MF, El-Seedi HR, Hassan I, Singh S, Abid M (2022) Target-based virtual screening of natural compounds identifies a potent antimalarial with selective falcipain-2 inhibitory activity. Front Pharmacol 13:850176
14. Wang L, Ma C, Wipf P, Liu H, Su W, Xie XQ (2013) TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. AAPS J 15:395–406
15. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100-1107
16. Backman TW, Cao Y, Girke T (2011) ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Res 39:W486-491
17. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH (2016) PubChem substance and compound databases. Nucleic Acids Res 44:D1202-1213
18. Jimenes-Vargas K, Pazos A, Munteanu CR, Perez-Castillo Y, Tejera E (2024) Prediction of compound-target interaction using several artificial intelligence algorithms and comparison with a consensus-based strategy. J Cheminform 16:27
19. Gregson A, Plowe CV (2005) Mechanisms of resistance of malaria parasites to antifolates. Pharmacol Rev 57:117–145
20. Yang SQ, Zhang LX, Ge YJ, Zhang JW, Hu JX, Shen CY, Lu AP, Hou TJ, Cao DS (2023) In-silico target prediction by ensemble chemogenomic model based on multi-scale information of chemical structures and protein sequences. J Cheminform 15:48
21. Ji KY, Liu C, Liu ZQ, Deng YF, Hou TJ, Cao DS (2023) Comprehensive assessment of nine target prediction web services: which should we choose for target fishing? Brief Bioinform 24
22. Chen X, Liu M, Gilson MK (2001) BindingDB: a web-accessible molecular recognition database. Comb Chem High Throughput Screen 4:719–725
23. Coordinators NR (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44:D7-19
24. Papadatos G, Davies M, Dedman N, Chambers J, Gaulton A, Siddle J, Koks R, Irvine SA, Pettersson J, Goncharoff N, Hersey A, Overington JP (2016) SureChEMBL: a large-scale, chemically annotated patent document database. Nucleic Acids Res 44:D1220-1228
25. Shaikh F, Tai HK, Desai N, Siu SWI (2021) LigTMap: ligand and structure-based target identification and activity prediction for small molecular compounds. J Cheminform 13:44
26. Nickel J, Gohlke BO, Erehman J, Banerjee P, Rong WW, Goede A, Dunkel M, Preissner R (2014) SuperPred: update on drug classification and target prediction. Nucleic Acids Res 42:W26-31
27. Awale M, Reymond JL (2019) Polypharmacology browser PPB2: target prediction combining nearest neighbors with machine learning. J Chem Inf Model 59:10–17
28. Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, Bellis L, Overington JP (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res 43:W612-620
29. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R (2013) Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41:W597-600
30. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28(1):31–36
31. RDKit. RDKit: Open-source cheminformatics
32. Bajusz D, Racz A, Heberger K (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform 7:20
33. Patterson DE, Cramer RD, Ferguson AM, Clark RD, Weinberger LE (1996) Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors. J Med Chem 39:3049–3059
34. Canham SM, Wang Y, Cornett A, Auld DS, Baeschlin DK, Patoor M, Skaanderup PR, Honda A, Llamas L, Wendel G, Mapa FA, Aspesi P Jr, Labbe-Giguere N, Gamber GG, Palacios DS, Schuffenhauer A, Deng Z, Nigsch F, Frederiksen M, Bushell SM, Rothman D, Jain RK, Hemmerle H, Briner K, Porter JA, Tallarico JA, Jenkins JL (2020) Systematic chemogenetic library assembly. Cell Chem Biol 27:1124–1129
35. Harding SD, Armstrong JF, Faccenda E, Southan C, Alexander SPH, Davenport AP, Spedding M, Davies JA (2023) The IUPHAR/BPS Guide to PHARMACOLOGY in 2024. Nucleic Acids Res 52:D1438–D1449
36. Food Drug Administration, F. Drugs@FDA: FDA-Approved Drugs
37. Mugumbate G, Overington JP (2015) The relationship between target-class and the physicochemical properties of antibacterial drugs. Bioorg Med Chem 23:5218–5224

38. Raimondi MV, Randazzo O, La Franca M, Barone G, Vignoni E, Rossi D, Collina S (2019) DHFR inhibitors: reading the past for discovering novel anticancer agents. Molecules 24:1140
39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504
40. Venture, M. f. M. Pathogen Box
41. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
42. Wong W, Bai XC, Sleebs BE, Triglia T, Brown A, Thompson JK, Jackson KE, Hanssen E, Marapana DS, Fernandez IS, Ralph SA, Cowman AF, Scheres SHW, Baum J (2017) Mefloquine targets the *Plasmodium falciparum* 80S ribosome to inhibit protein synthesis. Nat Microbiol 2:17031
43. Cheng J, Ma X, Krausz KW, Idle JR, Gonzalez FJ (2009) Rifampicin-activated human pregnane X receptor and CYP3A4 induction enhance acetaminophen-induced toxicity. Drug Metab Dispos 37:1611–1621
44. Tiash S, Saunders J, Hart CJS, Ryan JH, Riches AG, Skinner-Adams TS (2020) An image-based Pathogen Box screen identifies new compounds with anti-Giardia activity and highlights the importance of assay choice in phenotypic drug discovery. Int J Parasitol Drugs Drug Resist 12:60–67
45. Dennis ASM, Rosling JEO, Lehane AM, Kirk K (2018) Diverse antimalarials from whole-cell phenotypic screens disrupt malaria parasite ion and volume homeostasis. Sci Rep 8:8795
46. Veale CGL (2019) Unpacking the pathogen box-an open source tool for fighting neglected tropical disease. ChemMedChem 14:386–453
47. Meulenbeld HJ, Mathijssen RH, Verweij J, de Wit R, de Jonge MJ (2012) Danusertib, an aurora kinase inhibitor. Expert Opin Investig Drugs 21:383–393
48. Gamo FJ, Sanz LM, Vidal J, de Cozar C, Alvarez E, Lavandera JL, Vanderwall DE, Green DV, Kumar V, Hasan S, Brown JR, Peishoff CE, Cardon LR, Garcia-Bustos JF (2010) Thousands of chemical starting points for antimalarial lead identification. Nature 465:305–310
49. Guiguemde WA, Shelat AA, Bouck D, Duffy S, Crowther GJ, Davis PH, Smithson DC, Connelly M, Clark J, Zhu F, Jimenez-Diaz MB, Martinez MS, Wilson EB, Tripathi AK, Gut J, Sharlow ER, Bathurst I, El Mazouni F, Fowble JW, Forquer I, McGinley PL, Castro S, Angulo-Barturen I, Ferrer S, Rosenthal PJ, Derisi JL, Sullivan DJ, Lazo JS, Roos DS, Riscoe MK, Phillips MA, Rathod PK, Van Voorhis WC, Avery VM, Guy RK (2010) Chemical genetics of *Plasmodium falciparum*. Nature 465:311–315
50. Gagaring K, Borboa R, Francek C, Chen Z, Buenviaje J, Plouffe D, Winzeler E, Brinker A, Diagana T, Taylor J, Glynne R, Chatterjee A, Kuhen K (2010) Novartis-GNF Malaria Box. Genomics Institute of the Novartis Research Foundation (GNF) and Novartis Institute for Tropical Disease
51. Berg M, Van der Veken P, Joossens J, Muthusamy V, Breugelmans M, Moss CX, Rudolf J, Cos P, Coombs GH, Maes L, Haemers A, Mottram JC, Augustyns K (2010) Design and evaluation of *Trypanosoma brucei* metacaspase inhibitors. Bioorg Med Chem Lett 20:2001–2006
52. Murphy RC, Ojo KK, Larson ET, Castellanos-Gonzalez A, Perera BG, Keyloun KR, Kim JE, Bhandari JG, Muller NR, Verlinde CL, White AC Jr, Merritt EA, Van Voorhis WC, Maly DJ (2010) Discovery of potent and selective inhibitors of calcium-dependent protein kinase 1 (CDPK1) from *C. parvum* and *T. gondii*. ACS Med Chem Lett 1:331–335
53. Albanese C, Alzani R, Amboldi N, Avanzi N, Ballinari D, Brasca MG, Festuccia C, Fiorentini F, Locatelli G, Pastori W, Patton V, Roletto F, Colotta F, Galvani A, Isacchi A, Moll J, Pesenti E, Mercurio C, Ciomei M (2010) Dual targeting of CDK and tropomyosin receptor kinase families by the oral inhibitor PHA-848125, an agent with broad-spectrum antitumor efficacy. Mol Cancer Ther 9:2243–2254
54. Weiss GJ, Hidalgo M, Borad MJ, Laheru D, Tibes R, Ramanathan RK, Blaydorn L, Jameson G, Jimeno A, Isaacs JD, Scaburri A, Pacciarini MA, Fiorentini F, Ciomei M, Von Hoff DD (2012) Phase I study of the safety, tolerability and pharmacokinetics of PHA-848125AC, a dual tropomyosin receptor kinase A and cyclin-dependent kinase inhibitor, in patients with advanced solid malignancies. Invest New Drugs 30:2334–2343
55. Summers RL, Pasaje CFA, Pisco JP, Striepen J, Luth MR, Kumpornsin K, Carpenter EF, Munro JT, Lin D, Plater A, Punekar AS, Shepherd AM, Shepherd SM, Vanaerschot M, Murithi JM, Rubiano K, Akidil A, Ottilie S, Mittal N, Dilmore AH, Won M, Mandt REK, McGowen K, Owen E, Walpole C, Llinas M, Lee MCS, Winzeler EA, Fidock DA, Gilbert IH, Wirth DF, Niles JC, Baragana B, Lukens AK (2022) Chemogenomics identifies acetyl-coenzyme A synthetase as a target for malaria treatment and prevention. Cell Chem Biol 29(191–201):e198
56. Crowther GJ, Hillesland HK, Keyloun KR, Reid MC, Lafuente-Monasterio MJ, Ghidelli-Disse S, Leonard SE, He P, Jones JC, Krahn MM, Mo JS, Dasari KS, Fox AM, Boesche M, El Bakkouri M, Rivas KL, Leroy D, Hui R, Drewes G, Maly DJ, Van Voorhis WC, Ojo KK (2016) Biochemical Screening of five protein kinases from *Plasmodium falciparum* against 14,000 cell-active compounds. PLoS ONE 11:e0149996
57. Wang R, Fang X, Lu Y, Wang S (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 47:2977–2980
58. Rout S, Mahapatra RK (2019) In silico analysis of plasmodium falciparum CDPK5 protein through molecular modeling, docking and dynamics. J Theor Biol 461:254–267
59. Hoepfner D, McNamara CW, Lim CS, Studer C, Riedl R, Aust T, McCormack SL, Plouffe DM, Meister S, Schuierer S, Plikat U, Hartmann N, Staedtler F, Cotesta S, Schmitt EK, Petersen F, Supek F, Glynne RJ, Tallarico JA, Porter JA, Fishman MC, Bodenreider C, Diagana TT, Movva NR, Winzeler EA (2012) Selective and specific inhibition of the plasmodium falciparum lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. Cell Host Microbe 11:654–663

Publisher's Note