ToxProfiler: Toxicity-target profiler based on chemical similarity

Mohamed Diwan M. AbdulHameed [a,b,*], Ruifeng Liu [a,b], Patric Schyman [a,b], Daniel Sachs [a,b], Zhen Xu [a,b], Valmik Desai [a,b], Anders Wallqvist [a,*]

[a] Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Development Command, Fort Detrick, MD 21702, USA
[b] The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD 20817, USA

ARTICLE INFO

ABSTRACT

Identifying the ability of a chemical to interact with toxicity targets, such as proteins in an adverse outcome pathway, is an essential step in drug discovery and risk assessment. Computational approaches to screen for chemical-toxicity target interaction can serve as a rapid alternative to traditional *in vitro/in vivo* methods. In this work, we have developed a chemical-similarity based protocol that predicts the potential of a chemical to interact with 64 established toxicity targets. In particular, we created a chemogenomics database from public data sources to identify target representatives, i.e., chemicals that are known to interact with the selected targets. We evaluated the performance of 2D and 3D similarity approaches in correctly ranking known interacting compounds using an external evaluation set from ChEMBL database. We found that the 2D approach outperforms the 3D approach in target prediction. Here, we developed a publically available toxicity profiler website (https://toxpro.bhsai.org/) using 2D similarity-based screening approach that allows user to obtain toxicity target profile for a set of query compounds. We utilized the profiler to screen 649 known acute and highly toxic chemicals with a Globally Harmonized System (GHS) score of less than 2. In this set, acetylcholinesterase was the most frequently occurring target underlying toxicity. The developed toxicity profiler tool provides a rapid means to screen for mechanisms underlying chemical toxicity.

1. Introduction

Animal-based toxicological evaluations are integral part of evaluations in the chemical and pharmaceutical/drug development industry and are essential to meet the regulatory safety requirements. Understanding the adverse effects associated with more than 80,000 chemicals produced world-wide to provide the associated risk assessments requires extensive testing.[1] Similarly, new drug/lead molecules have to be tested for their potential to cause adverse effects during the drug development process. Animal-based toxicological evaluations are low throughput and raise ethical considerations. The bulk of acute toxicity studies only identifies the lethal dose 50% (LD_{50}) but lacks insights into the mechanisms of toxicity.[2] To address these concerns, there is growing emphasis on developing non-animal alternative testing approaches. The concept of adverse outcome pathway (AOP) was developed to aid this effort.[3] An AOP is a conceptual framework and consists of molecular initiating events (MIE), key events (KE), and adverse outcome (AO).[4] This type of framework enables prediction of apical/phenotypic-level endpoints through mechanism-based testing approaches reflecting MIEs and KEs.[5] Molecular initiating events represent the first step in an AOP and capture the chemical interaction with a biomolecule, which in turn can be linked to adverse outcomes through alterations in biological pathway.[6] Such MIEs could involve specific protein targets or non-protein targets. Chemical inhibition of human ether-a-go-go-related gene (hERG) leading to cardiac arrhythmias is an example of a specific protein target-based MIE [7,8]. Disruption of mitochondrial membrane potential leading to liver steatosis is an example for non-protein target MIE.[5] In this paper, we refer to protein target-based MIEs, which are linked to adverse outcome as toxicity targets. Predicting the ability of a chemical to interact with toxicity targets will not only aid in understanding the adverse liability associated with them but will also provide insights into the underlying mechanisms of toxicity.

Pharmaceutical companies have compiled lists of such toxicity

targets, utilizing evidence from knock out models to clinical data, which are used to gain insights into the safety profile of drug candidates.[8] Profiling of lead compounds is then carried out against these toxicity targets using *in vitro* panels.[8] Given the significance of adverse effects through interaction with these toxicity targets, there exists a need/demand to understand the potential of chemicals to interact with them. Contract research organizations provide screening services against these toxicity targets.[9] In particular, the SafetyScreen44 panel (Eurofins discovery, St Charles, MO) is routinely reported in the literature for medicinal chemistry/lead optimization studies.[10] The concept of profiling chemicals across toxicity targets is useful in developing mechanism-based testing strategies for risk assessment community.

In contrast to experimental *in vitro* screening, computational virtual screening (VS), is faster, less expensive, and provides an alternative to animal testing.[11] This screening can be structure-based or ligand-based.[12] In structure-based VS, the structure of the active site of the target is known and used to identify compounds that interact with it by a technique called docking. This approach, however, has a number of limitations, including *1)* the lack of knowledge regarding the structure of many important toxicity targets and *2)* the challenges of properly treating target flexibility, ionization state of the side chains, and water-mediated interactions.[13] In contrast, ligand-based VS (LBVS) methods do not have these limitations because they only focus on active molecules (i.e., ligands or chemicals known to interact with the drug target of interest). Once these actives (reference compounds) are identified for a target, they are used as representatives of the target active site.

LBVS is based on the principle that similar chemicals exhibit similar interactions and biological activity.[13] This approach, also referred to as *read-across*, has been used widely in risk assessment focused on apical end points such as acute toxicity values of chemicals.[14,15] Here, we propose to use the similarity-based read-across approach to screen chemicals for their ability to interact with toxicity targets. Chemical similarity can be calculated based on a 2D or 3D approach. The 2D LBVS approach uses atom connectivity/fingerprints to calculate the similarity between the compounds in the chemical library (query) and the reference compounds, whereas the 3D approach uses the shapes and chemical features of the compounds.[16] Because the latter approach does not consider atomic connectivity, it is capable of scaffold hopping, i.e., it can find new active compounds with novel scaffolds.[17] Studies have reported 2D structural alerts for the toxicity targets,[18,19] whereas 2D or 3D similarity-based approaches that predict the potential of chemicals to interact with the known toxicity targets have not been extensively explored. Furthermore, it is not clear whether there is an advantage of using either 2D or 3D LBVS approaches for this set of targets.

In this work, we performed a systematic comparison of 2D and 3D similarity-based approaches in generating toxicity targets profile of chemicals. We created a compendium of 64 toxicity targets reported in literature that are known to causally link to phenotype-level adverse effects. We created the target representative dataset, i.e., reference compounds, using compound annotations from DrugBank and the Toxin and Toxin-Target Database (T3DB) database.[20,21] We compared the performance of 2D and 3D approaches and found that the 2D outperforms the 3D method for this class of targets. We generated toxicity target profiles for 649 known acute toxicants and found acetylcholinesterase to be the most frequent target among these compounds. We have made the 2D-based similarity approach publically available on the Toxicity Profiler website (https://toxpro.bhsai.org/). This website allows any user to submit sets of query compounds and obtain the associated toxicity targets profiles. This tool aids in understanding the mechanisms underlying toxicity and prioritizing compounds for detailed toxicological evaluation.

2. Materials and methods

The tasks involved in creating a toxicity targets profiler included: *1)* identifying a panel of toxicity targets, *2)* identifying target representatives, and *3)* evaluating and selecting a suitable chemical similarity approach.

2.1. Collection of toxicity targets panel

We used two major sources to create the toxicity targets panel in this work, i.e., the collaborative data published from four major pharmaceutical companies, and the U.S. National Tox21 collaborative program. Recently, major pharmaceutical companies came together and provided an essential set of *in vitro* toxicity targets panel that are routinely used in their drug discovery programs.[8] These targets are selected based on detailed weight of evidence, i.e., a chemical interaction with these targets will result in observable adverse outcome in humans. As suggested by Bowes et al., there is also a need to include additional targets from kinases and transporter classes.[8] Consequently, we surveyed the literature and added targets belonging to this class and known to be associated with toxicity.[22] The Tox21 program, a collaborative effort from several U.S. federal agencies, screens for chemicals that alter endocrine function and we selected 14 targets for inclusion in this work. [23] Two of the 14 targets were also present in the panel suggested by Bowes et al.[8] In total, we created a panel consisting of 64 toxicity targets.

2.2. Collection of target representatives

Target representatives are compounds that are known to interact with these targets. We used two publically available and well recognized reference datasets, DrugBank and the Toxin and Toxin-Target Database (T3DB), to create the target representative dataset.[20,21] We downloaded all 'drug-uniprot links' for the four protein classes (targets, enzyme, carrier, and transporter) from the DrugBank website (https://go.drugbank.com/; version 5.1.2). These tables provide the protein Uniprot IDs and corresponding list of small molecules known to interact with them. We downloaded the 'All toxin-target mechanisms of action and references' from T3DB website (http://www.t3db.ca/; version 2.0) and obtained the Uniprot IDs associated with Toxins. From these, we extracted data associated with 64 toxicity targets. We pre-processed the chemical data using Pipeline Pilot (Version 18.1.100.11) protocols and removed duplicate compounds, salts, mixtures, and standardized the molecules. Standardization refers to molecule pre-processing step wherein proper assignment of bond order, aromaticity, and hydrogens are done. Filter program (Version 3.0.1.2) from OpenEye software was used to remove macrocyclic compounds and non-drug like compounds. Finally, we created a chemogenomics matrix with 2,655 chemicals in rows and 64 toxicity targets as columns. The matrix has values 1 if there is a known interaction between the chemical and target otherwise it is 0. By parsing through each column in this matrix, we can obtain the list of target representatives for each of the toxicity targets.

2.3. Chemical similarity approach

Chemical similarity can be calculated based on a 2D or 3D approach. The 2D approach uses atom connectivity/fingerprints. We used Pipeline Pilot extended-connectivity fingerprints with a diameter of four chemical bonds (ECFP4) for 2D similarity calculations.[24] For 3D similarity search, we used the Rapid Overlay of Chemical Structures (ROCS) program (Version 3.2.2.2) .[25] ROCS is considered the *de facto* standard and a representative of this type of approach.[26] The 2D approach uses atom connectivity whereas the 3D approach uses shape and chemical group matching in 3D. The latter requires generation of a set of 3D conformers. We used the Omega program (OpenEye, Santa Fe, NM) (Version 3.0.1.2) to generate 3D conformations. We used the default parameters in Omega to generate the conformers. For all the target representatives, we generated multiple conformations with $nconf = 200$ option, i.e., up to 200 conformations were generated for each reference

molecule. In this work, for both 2D and 3D, we calculated the similarity between each query compound and a set of reference compounds (target representative set) known to interact with the toxicity target of interest, and then use the maximum similarity score (MAX) between them to represent the potential of the query to interact with the toxicity target.

2.4. Validation analysis

We evaluated the comparative performance of the 2D and 3D approaches by examining whether the target representatives were able to identify chemicals already known to interact with the target when mixed in with inactive chemicals. We collected the external test data set from the ChEMBL (Version ChEMBL 24) database. We downloaded the largest available bioactivity data of same type (IC_{50} or EC_{50} or K_i) for each target. Most of the targets (28 of 35) had IC_{50} values. Five targets (ADORA2A, CNR1, OPRK1, HTR2A, and HTR2B) had K_i values and two targets (OPRM1 and OPRD1) had EC_{50} values. Compounds with activity values ≤ 1 μM were labelled as actives and ≥ 10 μM were labelled as inactives. These are standard activity value cut-offs typically used in quantitative structure–activity relationship (QSAR) studies. We were able to collect external test dataset for 35 of 64 targets. The rest of the targets didn't have sufficient data for evaluation. We removed any overlapping compounds that were present in the target representative set. Because the number of actives was limited for some targets, we focused on interactions *per se* rather than on whether an interaction represented an agonistic or antagonistic activity. We pre-processed the data and removed duplicate compounds. If the duplicate compounds showed the same activity, we retained one; if they showed different activity, we removed both. We also removed the database compounds whose activity values were recorded as inconclusive. We performed screening for each target and calculated the performance using area under the receiver operating characteristic curve (AUC) values. Higher AUC values represent improved screening, whereas an AUC closer to 0.5 indicates that the results are close to random, i.e., the approach is not able to separate actives from inactives. We used the Omega program to generate conformation for each query molecule. We used the same set of query molecules for both 2D and 3D screening approaches.

2.5. Exemplar study

We collected a list of known, highly toxic molecules from the public datasets with Globally Harmonized Score (GHS) category 1 and 2.[27] The acute toxicity data for these molecules were obtained from rat oral LD_{50} studies. The molecules in this set were pre-processed as described for query molecules in the validation set above and created the toxicity targets profiles. This resulted in a matrix of 64 targets across 649 toxicants. The rows of the matrix have the MAX 2D similarity Tanimoto score between each query chemical and respective toxicity targets. We converted the Tanimoto score into Z-score. The Z-score of chemical *i* for a toxicity target *j* is given by

$$Z_{i,j} = \frac{X_{i,j} - \mu_i}{\sigma_i} \tag{1}$$

where $X_{i,j}$ is the MAX Tanimoto score for chemical *i* and toxicity target *j*; μ_i is the average of MAX Tanimoto scores for chemical *i* across all 64 toxicity targets; and σ_i is the standard deviation of the MAX Tanimoto score for chemical *i* across all 64 toxicity targets. In order to characterize the diversity of this set of 649 toxicants, we clustered them using the 'cluster molecules' component in pipeline pilot.

2.6. Implementation of the ToxProfiler web tool

The implementation of ToxProfiler consists of the front view, database, and controller. The front view is implemented with PrimeFace 7.0 library and BootsFaces 1.3.0 library. On the database side, we used

PostgresSQL 12.0. All submitted queries and results from the program are stored in the database. Users can submit up to 10,000 query compounds. The controller is written in Java and runs in JDK 1.8. It handles interaction with the user from file uploading to job submission. When a job gets submitted, the controller stores a record in the database and later executes the job when resources are available. The system uses Pipeline Pilot to manage job execution and uses Linux server for computational functions. After the job is finished, the user gets an email notification with a link to the result page. The result page provides the job name, description, raw outputs, heat maps, and a table of compounds along with analytic results. Multiple records in the table can be selected and more details on the results will show up when the "show detail" button is clicked.

3. Results and discussion

Chemical interaction with toxicity targets serves as the MIE for adverse outcome. Predicting the potential of the chemical to interact with such toxicity targets will help in understanding the potential mechanisms and possible adverse effects associated with chemical exposure.

3.1. Toxicity targets panel and target representatives

The first requirement for creating a chemical similarity-based toxicity targets profiler is to identify a panel of toxicity targets and a set of reference compounds that are known to interact with these targets, i.e., target representatives. Our panel includes 64 toxicity targets identified from the literature as discussed in the methods section. Fig. 1 shows the different protein classes represented in our toxicity targets panel. Supplementary Table 1 provides the list of all 64 targets along with their UniProt identifiers,[28] gene symbol, and protein class. A majority of the targets in our panel are G-protein coupled receptors (GPCRs) (24 of 64) followed by nuclear receptors (14 of 64) and ion channels (9 of 64). It is to be expected as GPCRs play a major physiological role and nearly 50% of approved drugs are known to interact with a protein in the GPCR family.[8] Chemical interaction with nuclear receptors leads to endocrine disruption and subsequent adverse health effects.[23] It should be noted that 44 of the 64 targets are being used in commercial *in vitro* safety screening panel such as Eurofins Cerep SafetyScreen44.[10] Developing an *in silico* profiler for these targets provides a complementary tool to the experimental *in vitro* approach. Furthermore, it can aid in prioritizing larger sets of compounds to be selected for more detailed *in vitro* screening evaluation.

Target representatives are reference compounds that are known to interact with the target of interest. Once a set of target representatives are identified it allows us to utilize similarity-based screening approaches. Similarity is calculated between target representatives
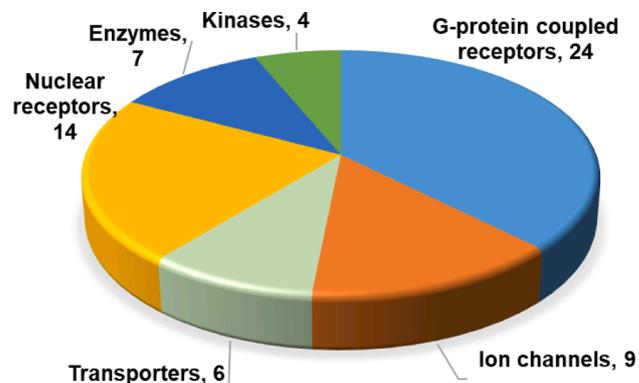
Fig. 1. Protein classes of the 64 toxicity targets. Toxicity targets were grouped into six major protein classes and the number of targets per class is given.

(reference set) and query set (screening set). If the compounds in the query set are found to be similar to any one of the compound in the target representative set above a certain threshold then the query compound is predicted to interact with associated target.

As described in the Methods section, we used two publically available data resources (DrugBank and T3DB) to collect known chemical-target interaction sets. After pre-processing, we obtained a final chemical-toxicity target matrix with 2,655 chemicals and 64 toxicity targets. Each row in the matrix is a chemical and each column is a toxicity target. The matrix has values 1 if there is a known interaction between the chemical and target otherwise it is 0. Overall, it is a sparse matrix with 1,300 chemicals (49%) interacting with only one target and 407 chemicals interacting with two targets. Fig. 2**A** summarizes the matrix as a network where the red circles represent the toxicity targets and blue circles denote chemicals. Forty-three of the 64 targets have more than 50 chemicals as target representatives, fourteen targets have between 10 and 50 chemicals as target representatives, and seven targets have between 3 and 10 chemicals as target representatives. Aryl hydrocarbon receptor (AHR), estrogen receptors-1,2 (ESR1,ESR2), and pregnane X receptor (NR1I2) have the highest number (≥300) of target representatives (Fig. 2**B**). Vasopressin receptor 1A (AVPR1A) and voltage gated potassium channel protein (KCNQ1) have the lowest number (three) of target representatives.

3.2. Evaluation of 2D and 3D similarity-based approaches

We next focused on selecting the similarity-based approach to develop the *in-silico* profiler. Given the target representative set, we evaluated the performance of 2D and 3D approach in retrieving actives from inactives for each target. For this, we retrieved experimentally reported actives and inactives associated with 35 targets from ChEMBL

and utilized them as external validation dataset.[29] Each compound in the external validation set was sorted based on the maximal similarity to the respective target representative set (reference compounds) and AUC values were calculated as a measure of performance. The AUC values reflects the ability of the similarity approach in retrieving the actives from inactive compounds with a higher AUC value indicating better performance. The performance of an approach is considered to be random if the AUC values are closer to 0.5. Using the same set of target representatives and external validation set allows us to directly compare the performance of 2D and 3D similarity-based approach. As mentioned in the Methods Section, for 2D we used ECFP4 fingerprints and for 3D used the Omega-generated conformations.

Fig. 3 shows the AUC values for both approaches in the external validation study. Table 1 provides the list of 35 targets, AUC values, and percentage difference between the two approaches. Overall, we find that in comparison with the 2D approach, the 3D approach performed poorly in our external validation analysis. The 2D approach performed better than 3D approach for 86% (31 of 35) of the targets evaluated. For two of the four targets where 3D performs slightly better, either approach had AUC values less than 0.6. For six targets, the 3D approach had AUC value ≤ 0.5 whereas the 2D approach had an AUC value greater than 0.5 for all the targets. In 46% (16 of 35) targets, the 3D approach had AUC values less than 0.6 whereas for the 2D approach only 0.09% (3 of 35) targets had such low AUC values. For 21 targets the 2D approach had a greater than 10% difference, whereas the 3D approach did not achieve a 10% difference in any of the targets (Table 1). Previous studies also report similar or better performances of 2D compared to 3D similarity-based approaches against certain targets.[30,31] As the goal of this comparative analysis was to select a similarity-search approach to be implemented in the specific 64-target toxicity profiler developed here, we performed this evaluation because we did not find a corresponding
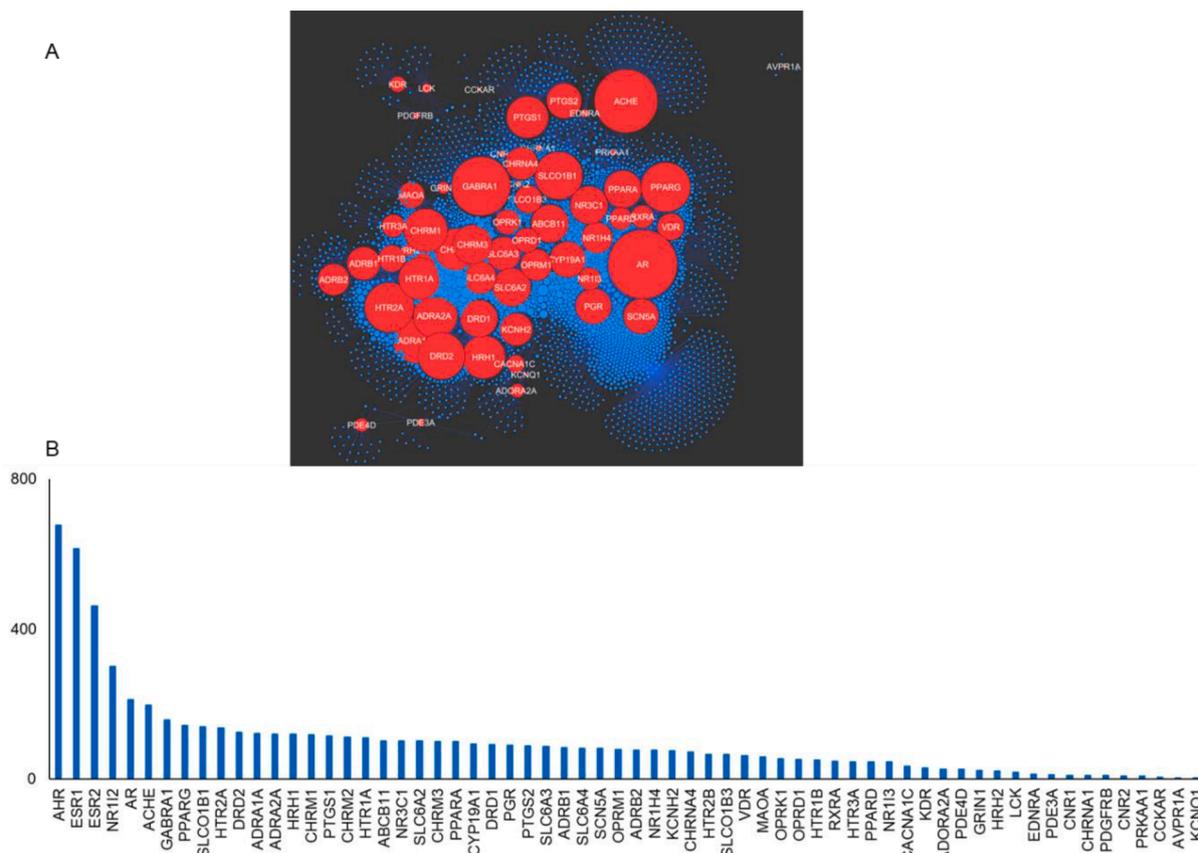
Fig. 2. A) Network representation of the toxicity target-chemical matrix; B) Number of target representatives per toxicity target. The size of the node corresponds to the number of interactions associated with them.
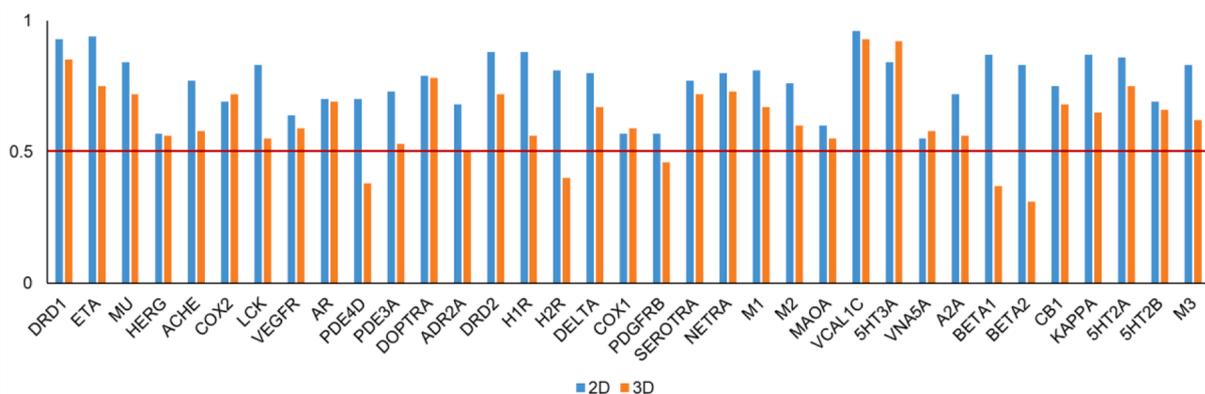
Fig. 3. Receiver operating characteristic area under the curve (AUC) values using the 2D and 3D similarity approach across 35 targets in the external validation set. The names of each target are given in Table 1.

Table 1
Receiver operating characteristic area under the curve (AUC) values from 2D and 3D approach across 35 targets in external validation analysis.

No	NAME	SYMBOL	2D	3D	% Diff
1	D(1A) dopamine receptor	DRD1	0.93	0.85	9
2	Endothelin-1 receptor	EDNRA	0.94	0.75	25
3	Mu-type opioid receptor	OPRM1	0.84	0.72	17
4	Potassium voltage-gated channel subfamily H member 2	KCNH2	0.57	0.56	2
5	Acetylcholinesterase	ACHE	0.77	0.58	33
6	Prostaglandin G/H synthase 2	PTGS2	0.69	0.72	-4
7	Tyrosine-protein kinase Lck	LCK	0.83	0.55	51
8	Vascular endothelial growth factor receptor 2	KDR	0.64	0.59	8
9	Androgen receptor	AR	0.70	0.69	1
10	cAMP-specific 3′,5′-cyclic phosphodiesterase 4D	PDE4D	0.70	0.38	84
11	cGMP-inhibited 3′,5′-cyclic phosphodiesterase A	PDE3A	0.73	0.53	38
12	Sodium-dependent dopamine transporter	SLC6A3	0.79	0.78	1
13	Alpha-2A adrenergic receptor	ADRA2A	0.68	0.50	36
14	D(2) dopamine receptor	DRD2	0.88	0.72	22
15	Histamine H1 receptor	HRH1	0.88	0.56	57
16	Histamine H2 receptor	HRH2	0.81	0.40	103
17	Delta-type opioid receptor	OPRD1	0.80	0.67	19
18	Prostaglandin G/H synthase 1	PTGS1	0.57	0.59	-3
19	Platelet-derived growth factor receptor beta	PDGFRB	0.57	0.46	24
20	Sodium-dependent serotonin transporter	SLC6A4	0.77	0.72	7
21	Sodium-dependent noradrenaline transporter	SLC6A2	0.80	0.73	10
22	Muscarinic acetylcholine receptor M1	CHRM1	0.81	0.67	21
23	Muscarinic acetylcholine receptor M2	CHRM2	0.76	0.60	27
24	Amine oxidase [flavin-containing] A	MAOA	0.60	0.55	9
25	Voltage-dependent L-type calcium channel alpha-1C	CACNA1C	0.96	0.93	3
26	5-hydroxytryptamine receptor 3A	HTR3A	0.84	0.92	-9
27	Sodium channel protein type 5 subunit alpha	SCN5A	0.55	0.58	-5
28	Adenosine receptor A2a	ADORA2A	0.72	0.56	29
29	Beta-1 adrenergic receptor	ADRB1	0.87	0.37	135
30	Beta-2 adrenergic receptor	ADRB2	0.83	0.31	168
31	Cannabinoid receptor 1	CNR1	0.75	0.68	10
32	Kappa-type opioid receptor	OPRK1	0.87	0.65	34
33	5-hydroxytryptamine receptor 2A	HTR2A	0.86	0.75	15
34	5-hydroxytryptamine receptor 2B	HTR2B	0.69	0.66	5
35	Muscarinic acetylcholine receptor M3	CHRM3	0.83	0.62	34

evaluation of the selected target set. The current results are limited to the selected target representative set, and the result do reflect a generalized comparison of a 2D versus a 3D approach, as the results depend on

the selected targets and decoy set. In addition to an improved performance, the 2D approach is also faster than 3D as it doesn't require generation of conformers and overlap calculation. Based on these results, we implemented the 2D approach using ECFP4 fingerprint as the similarity search approach.

3.3. ToxProfiler – Web-based tool

We created ToxProfiler, a web-based tool to computationally screen query chemicals and predict their potential to interact with toxicity targets, i.e., MIE for adverse outcomes. In summary, we used 2,655 chemicals collected from public chemogenomics databases as target representatives to represent 64 toxicity targets and used 2D similarity approach to compare the similarity between the query molecules with the reference set of molecules. The ToxProfiler website is publically available online at https://toxpro.bhsai.org/ and allows user to upload query chemicals in SMILES format. Users are also given an option to draw the compound of interest and add it as a query molecule. The status of the submitted jobs can be tracked in the web site. Users can view all of the results on the ToxProfiler website or download them for offline analysis. Fig. 4 shows the snapshots of output obtained from ToxProfiler web site. For each query compound ToxProfiler displays as output its name, structure, and a toxicity targets profile bar. The toxicity targets profile is color coded based on the similarity to target representatives. Based on the Z-score range greater than 1.96, 1.645–1.96, less than 1.645 it is marked in red, yellow, and green, respectively. Scrolling over each profile bar renders the detailed color-coded result for each of the 64 toxicity targets (Fig. 4**B**). The web site also provides a heat map (Fig. 4**C**) as output, summarizing the results for all submitted query molecules as one snap shot. The raw results, which is a matrix of query chemicals in columns and 64 toxicity targets in rows along with Z-scores, can be downloaded as text file for storage or further analysis. In addition to screening all 64 toxicity targets, an option is provided to subset, screen, and profile the toxicity targets used in SafetyScreen44. This option can then be used to compare the results between computational and external experimental safety profile screening.

3.4. Exemplar analysis

We screened 649 known acute and highly toxic chemicals with a Globally Harmonized System (GHS) score of less than 2 using the ToxProfiler. It should be noted that GHS score of less than 2 represents compounds with LD_{50} values lower than 50 mg/kg. We performed clustering analysis to understand the composition/diversity of 649 chemicals using FCFP4 fingerprints. We found 12 clusters and Supplementary Fig. 1 provides the structure cluster centers, illustrating the diversity of this set of chemicals. Fig. 5 shows the toxicity targets profile
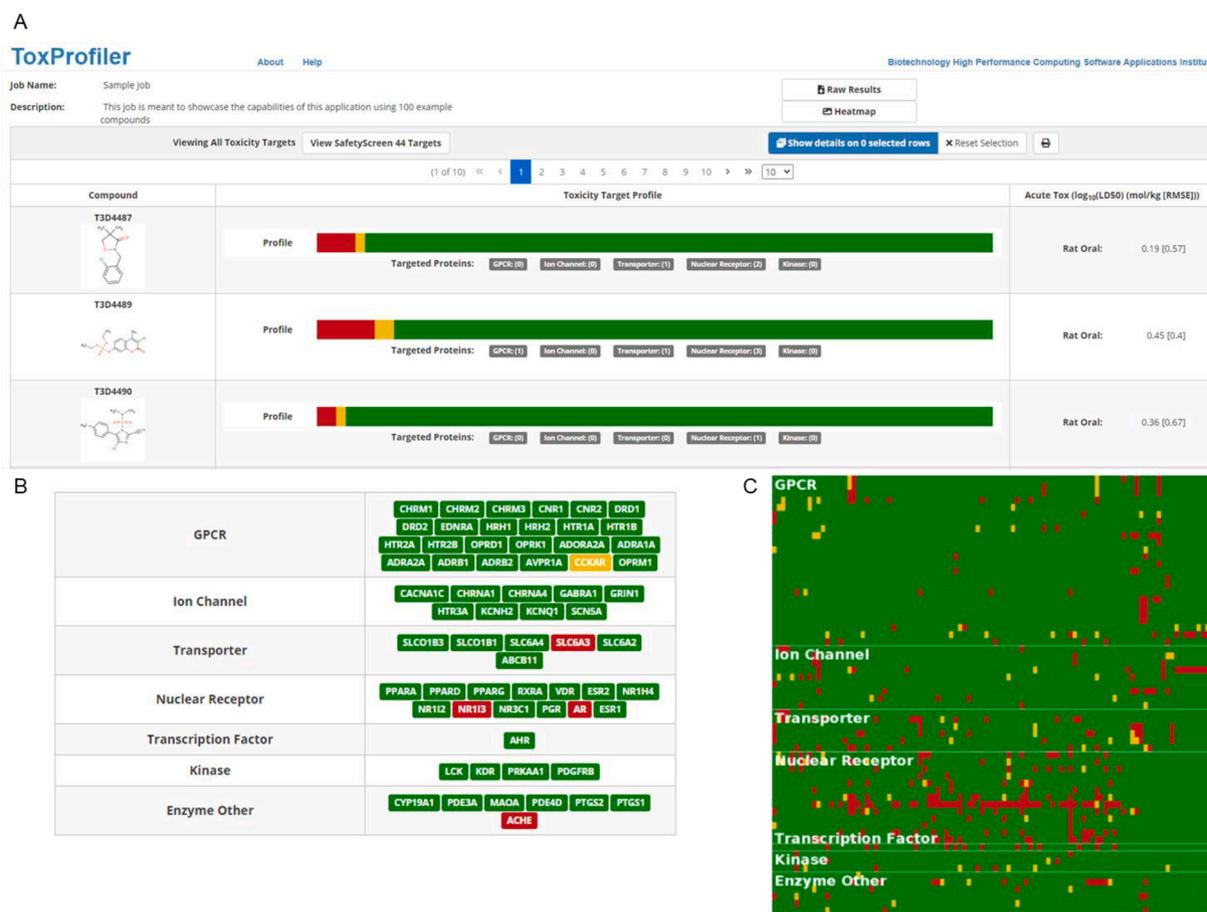
Fig. 4. A) Screenshot of typical result generated from ToxProfiler web tool. Each chemical in the user uploaded query set is present in each row. The molecular structure of chemical query is displayed along with a summary of toxicity target profile and rat oral acute toxicity values. The summary of the toxicity target profile is present as a bar plot. The red, yellow, and green color in the bar plot represents the number of predicted interactions, potential interactions, and lack of interactions with 64 toxicity targets respectively. B) Results can be visualized for each chemical where predicted interaction with toxicity target is highlighted. C) Summary result for a set of query molecules can be visualized as a heat map. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for these compounds. We found acetylcholinesterase to be the most frequent target of these highly toxic chemicals (Fig. 5). Two hundred-ninety of the 649 chemicals (45%) were predicted to interact with acetylcholinesterase. Nuclear receptors like pregnane X receptor (NR1I2), estrogen receptors-1 and 2 (ESR1, ESR2), and aryl hydrocarbon receptors are the other most frequent targets of these chemicals. Targets belonging to the kinase class are predicted to have the least interactions with these toxicants (Fig. 5, **Supplementary Table 2**). These results show the potential utility of this tool. For example, through this screening we can obtain some mechanistic insights into the nature of the query dataset. The identification of acetylcholinesterase highlights the prevalence of organophosphates and carbamates in this set of known acute toxic compounds. It should also be noted that some frequent hits like pregnane X receptor and aryl hydrocarbon receptor are to be expected as they are known to interact with many chemicals due to their role as xenobiotic sensors.[32] DTXSID5021099 is a chemical among the 649 acute toxicants in the exemplar study set which is also not present in our reference set. ToxProfiler tool predicts it to interact with seven toxicity targets, including voltage dependent calcium channel (gene symbol: CACNA1C) and muscarinic receptors (gene symbol: CHRM1, CHRM2, and CHRM3) (**Supplementary Table 3**). We reviewed the literature and found that this compound is indeed known to interact with voltage dependent calcium channel.[33] Katayama et al. reports that compounds with calcium channel antagonist activity also have interactions with muscarinic receptors.[34] This is in agreement with the predictions made using the ToxProfiler tool. It should be noted that the

above exemplar set is limited to acutely toxic compounds. Analysis of other manifestations of toxicity, such as chronic, requires different datasets and will be performed as a part of future work.

Overall, we have created a computational tool that performs similarity-based screening and predicts the potential of query chemicals to interact with toxicity targets. This approach can be considered as an online read-across tool focusing on chemical-target interactions/MIE profiler. There are previous works by Allen et al. and Mellor et al. focusing on structural alerts and MIEs.[18,35] Another publically available read-across tool like GenRA focuses on *in vivo* toxicity end points.[36] Our new tool will be useful to toxicologists as it will be complementary to the previous work in this area and it focuses on chemical-target interaction and utilizes chemical similarity approach.

3.5. Conclusion

Identifying the potential of a chemical to interact with toxicity targets is a critical step in drug discovery and risk assessment. We have created ToxProfiler as a web-based tool to screen chemicals against 64 toxicity targets and identify the potential of the chemicals to interact with these targets as MIEs in an Adverse Outcome Pathway. We used the chemical similarity principle to generate the profile of toxicity targets. We showed that the 2D approach performs better than the 3D similarity approach for this select set of toxicity targets. ToxProfiler is publically available at https://toxpro.bhsai.org/ and can be used for prospective analysis of toxicity, performing risk evaluation, and identifying
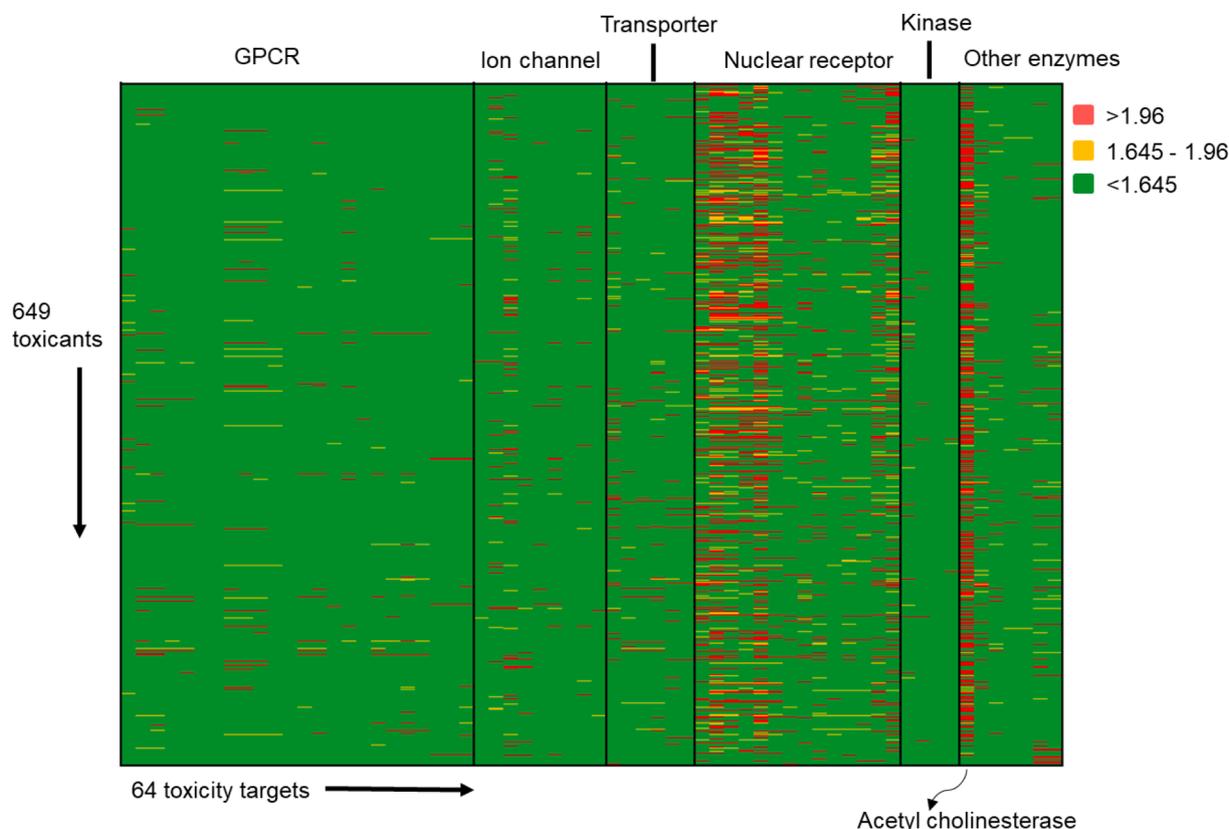
Fig. 5. Toxicity target profile generated for 649 toxic chemicals from ToxProfiler. 649 acute toxicants are present in the rows and 64 toxicity targets are present in the columns grouped as per protein class. Red color represents predicted interactions, yellow represents possible interactions, and green represent lack of interactions. Acetylcholinesterase has the highest number of predicted interactions and is highlighted in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

potential mechanisms associated with toxicity.

CRediT authorship contribution statement

Mohamed Diwan M. AbdulHameed: Conceptualization, Formal analysis, Investigation, Methodology, Data curation, Software, Validation, Writing - original draft, Writing - review & editing, Project administration. **Ruifeng Liu:** Conceptualization, Software, Resources. **Patric Schyman:** Conceptualization, Resources. **Daniel Sachs:** Data curation, Software. **Zhen Xu:** Data curation, Software. **Valmik Desai:** Resources, Software, Writing - review & editing. **Anders Wallqvist:** Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comtox.2021.100162.

References

[1] R.J. Kavlock, G. Ankley, J. Blancato, M. Breen, R. Conolly, D. Dix, K. Houck, E. Hubal, R. Judson, J. Rabinowitz, A. Richard, R.W. Setzer, I. Shah, D. Villeneuve, E. Weber, Computational toxicology–a state of the science mini review, Toxicol. Sci. 103 (2008) 14–27.

[2] J. Strickland, A.J. Clippinger, J. Brown, D. Allen, A. Jacobs, J. Matheson, A. Lowit, E.N. Reinke, M.S. Johnson, M.J. Quinn Jr., D. Mattie, S.C. Fitzpatrick, S. Ahir, N. Kleinstreuer, W. Casey, Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies, Regul. Toxicol. Pharm. 94 (2018) 183–196.

[3] G.T. Ankley, R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrrano, J.E. Tietge, D.L. Villeneuve, Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment, Environ. Toxicol. Chem. 29 (2010) 730–741.

[4] N.O. Oki, M.D. Nelms, S.M. Bell, H.M. Mortensen, S.W. Edwards, Accelerating Adverse Outcome Pathway Development Using Publicly Available Data Sources, Curr. Environ. Health Rep. 3 (2016) 53–63.

[5] M.D.M. AbdulHameed, V.R. Pannala, A. Wallqvist, Mining Public Toxicogenomic Data Reveals Insights and Challenges in Delineating Liver Steatosis Adverse Outcome Pathways, Front. Genet. 10 (2019) 1007.

[6] T.E. Allen, J.M. Goodman, S. Gutsell, P.J. Russell, Defining molecular initiating events in the adverse outcome pathway framework for risk assessment, Chem. Res. Toxicol. 27 (2014) 2100–2112.

[7] P. Schyman, R. Liu, A. Wallqvist, General Purpose 2D and 3D Similarity Approach to Identify hERG Blockers, J. Chem. Inf. Model. 56 (2016) 213–222.

[8] J. Bowes, A.J. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldron, S. Whitebread, Reducing safety-related drug attrition: the use of in vitro pharmacological profiling, Nat Rev Drug Discov 11 (2012) 909–922.

[9] S. Whitebread, J. Hamon, D. Bojanic, L. Urban, Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development, Drug Discov Today 10 (2005) 1421–1433.

[10] J. Sniecikowska, M. Gluch-Lutwin, A. Bucki, A. Wieckowska, A. Siwek, M. Jastrzebska-Wiesek, A. Partyka, D. Wilczynska, K. Pytka, K. Pociecha, A. Cios, E. Wyska, A. Wesolowska, M. Pawlowski, M.A. Varney, A. Newman-Tancredi, M. Kolaczkowski, Novel Aryloxyethyl Derivatives of 1-(1-Benzoylpiperidin-4-yl) methanamine as the Extracellular Regulated Kinases 1/2 (ERK1/2) Phosphorylation-Preferring Serotonin 5-HT1A Receptor-Biased Agonists with Robust Antidepressant-like Activity, J. Med. Chem. 62 (2019) 2750–2771.

[11] S. Ekins, Progress in computational toxicology, J. Pharmacol. Toxicol. Methods 69 (2014) 115–140.

[12] D.M. Kruger, A. Evers, Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors, ChemMedChem 5 (2010) 148–158.

[13] M.D. AbdulHameed, S. Chaudhury, N. Singh, H. Sun, A. Wallqvist, G.J. Tawa, Exploring polypharmacology using a ROCS-based target fishing approach, J. Chem. Inf. Model. 52 (2012) 492–505.

[14] M.D. Nelms, G. Ates, J.C. Madden, M. Vinken, M.T. Cronin, V. Rogiers, S.J. Enoch, Proposal of an in silico profiler for categorisation of repeat dose toxicity data of hair dyes, Arch. Toxicol. 89 (2015) 733–741.

[15] P. Grace, H. George, P. Prachi, S. Imran, Navigating through the minefield of read-across tools: A review of in silico tools for grouping, Comput. Toxicol. 3 (2017) 1–18.

[16] D. Rognan, Chemogenomic approaches to rational drug design, Br. J. Pharmacol. 152 (2007) 38–52.

[17] T.S. Rush 3rd, J.A. Grant, L. Mosyak, A. Nicholls, A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction, J. Med. Chem. 48 (2005) 1489–1495.

[18] T.E. Allen, S. Liggi, J.M. Goodman, S. Gutsell, P.J. Russell, Using Molecular Initiating Events To Generate 2D Structure-Activity Relationships for Toxicity Screening, Chem. Res. Toxicol. 29 (2016) 1611–1627.

[19] R. Liu, X. Yu, A. Wallqvist, Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries, J Cheminform 7 (2015) 4.

[20] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082.

[21] D. Wishart, D. Arndt, A. Pon, T. Sajed, A.C. Guo, Y. Djoumbou, C. Knox, M. Wilson, Y. Liang, J. Grant, Y. Liu, S.A. Goldansaz, S.M. Rappaport, T3DB: the toxic exposome database, Nucleic Acids Res. 43 (2015) D928–D934.

[22] T. Force, K.L. Kolaja, Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes, Nat Rev Drug Discov 10 (2011) 111–126.

[23] R.R. Tice, C.P. Austin, R.J. Kavlock, J.R. Bucher, Improving the human hazard characterization of chemicals: a Tox21 update, Environ. Health Perspect. 121 (2013) 756–765.

[24] D. Rogers, M. Hahn, Extended-connectivity fingerprints, J. Chem. Inf. Model. 50 (2010) 742–754.

[25] P.C. Hawkins, A.G. Skillman, A. Nicholls, Comparison of shape-matching and docking as virtual screening tools, J. Med. Chem. 50 (2007) 74–82.

[26] A. Nicholls, G.B. McGaughey, R.P. Sheridan, A.C. Good, G. Warren, M. Mathieu, S. W. Muchmore, S.P. Brown, J.A. Grant, J.A. Haigh, N. Nevins, A.N. Jain, B. Kelley, Molecular shape and medicinal chemistry: a perspective, J. Med. Chem. 53 (2010) 3862–3886.

[27] R. Liu, M. Madore, K.P. Glover, M.G. Feasel, A. Wallqvist, Assessing Deep and Shallow Learning Methods for Quantitative Prediction of Acute Chemical Toxicity, Toxicol. Sci. 164 (2018) 512–526.

[28] C. UniProt, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res. 47 (2019) D506–D515.

[29] D. Mendez, A. Gaulton, A.P. Bento, J. Chambers, M. De Veij, E. Felix, M. P. Magarinos, J.F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Maranon, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, A.R. Leach, ChEMBL: towards direct deposition of bioassay data, Nucleic Acids Res. 47 (2019) D930–D940.

[30] G.B. McGaughey, R.P. Sheridan, C.I. Bayly, J.C. Culberson, C. Kreatsoulas, S. Lindsley, V. Maiorov, J.F. Truchon, W.D. Cornell, Comparison of topological, shape, and docking methods in virtual screening, J. Chem. Inf. Model. 47 (2007) 1504–1519.

[31] R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods? Drug Discov Today 7 (2002) 903–911.

[32] M.D. AbdulHameed, D.L. Ippolito, A. Wallqvist, Predicting Rat and Human Pregnane X Receptor Activators Using Bayesian Classification Models, Chem. Res. Toxicol. 29 (2016) 1729–1740.

[33] Y. Okamiya, T. Kishimoto, K. Aoki, K. Sunakawa, K. Hoshina, T. Takeshita, Effect of palonidipine hydrochloride (TC-81), a novel calcium antagonist, on the canine coronary artery, Nihon Yakurigaku Zasshi 102 (1993) 23–33.

[34] S. Katayama, S. Kito, R. Miyoshi, H. Matsubayashi, Effects of calcium antagonists on muscarinic receptor subtypes in the rat brain, Brain Res. 422 (1987) 168–171.

[35] C.L. Mellor, F.P. Steinmetz, M.T. Cronin, Using Molecular Initiating Events to Develop a Structural Alert Based Screening Workflow for Nuclear Receptor Ligands Associated with Hepatic Steatosis, Chem. Res. Toxicol. 29 (2016) 203–212.

[36] G. Helman, I. Shah, A.J. Williams, J. Edwards, J. Dunne, G. Patlewicz, Generalized Read-Across (GenRA): A workflow implemented into the EPA CompTox Chemicals Dashboard, ALTEX 36 (2019) 462–465.