Towards a Computational Ecotoxicity Assay

Natasha Kamerlin,[†] Mickaël G. Delcey,[‡] Sergio Manzetti,[¶] and David van der Spoel[*,†]

†*Science for Life Laboratory, Department of Cell and Molecular Biology. Uppsala University, Box 596, SE-751 24 Uppsala, Sweden*

‡*Department of Chemistry - Ångström Laboratory, Uppsala University, SE-75120 Uppsala, Sweden*

¶*Fjordforsk A.S., Institute for Science and Technology, Midtun, 6894 Vangsnes, Norway*

E-mail: david.vanderspoel@icm.uu.se

Abstract

Thousands of anthropogenic chemicals are released into the environment each year, posing potential hazards to human and environmental health. Toxic chemicals may cause a variety of adverse health effects, triggering immediate symptoms or delayed effects over longer periods of time. It is thus crucial to develop methods that can rapidly screen and predict the toxicity of chemicals, to limit the potential harmful impacts of chemical pollutants. Computational methods are being increasingly used in toxicity predictions. Here, the method of molecular docking is assessed for screening potential toxicity of a variety of xenobiotic compounds, including pesticides, pharmaceuticals, pollutants and toxins deriving from the chemical industry. The method predicts the binding energy of the pollutants to a set of carefully selected receptors, under the assumption that toxicity in many cases is related to interference with biochemical pathways. The strength of the applied method lies in its rapid generation of interaction maps between potential toxins and the targeted enzymes, which could quickly yield molecular-level information and insight into potential perturbation pathways, aiding in the prioritisation

of chemicals for further tests. Two scoring functions are compared, Autodock Vina and the machine-learning scoring function RF-Score-VS. The results are promising, though hampered by the accuracy of the scoring functions. The strengths and weaknesses of the docking protocol are discussed, as well as future directions for improving the accuracy for the purpose of toxicity predictions.

Introduction

Environmental pollution and ecotoxic stress through habitat-destruction, further exacerbated by climate change, have led to the onset of the sixth great mass-extinction event.[1] Pollution affects all organisms in the environment including humans, through a range of mechanisms, from chronic toxicity via the respiratory function and the gastrointestinal channel to dermatological uptake. This may trigger the detoxification pathways of the body in an attempt to eliminate toxic compounds from the system.[2,3] At continuous low-dose exposure, pollutants may also incur long-termed effects.[4,5] There is, however, limited or no available toxicity data for tens of thousands of compounds to which organisms in the environment, including humans, are exposed, due to the high cost and laborious nature of traditional toxicity testing.[6,7] This underscores the need for approaches that can rapidly screen and predict the toxicity of chemicals and prevent them from being released in the environment.

In 2007, the National Research Council, Committee on Toxicity Testing and Assessment of Environmental Agents (U.S.A.), proposed a new framework, emphasising the importance of integrating $in\ vitro$-based and computational methods for evaluating toxicity.[8] In response, a series of programs have been initiated in the last decade, aimed at identifying assays relevant to toxicity and screening the biological activity of large numbers of candidate pollutants in a cost-efficient and timely manner. In particular, the Toxicity Forecasting (ToxCast) project of the Environmental Protection Agency (U.S.A.) has been instrumental in setting the stage for $in\ vitro$ high-throughput assays[9,10] In a significant effort, over 900 chemicals were employed in over 300 concentration-response assays to produce a large matrix with estimates of the chemical potency, in terms of

the AC_{50} parameter.[10] The launch of Phase III further broadened the chemical and assay space of ToxCast. Although the measured AC_{50} values have no direct toxicological interpretation and estimated values may have large standard errors,[11] high-throughput screening (HTS) nevertheless offers an important benchmark and a means of comparing chemical activities. Other ongoing programs for predicting the hazardous character of pollutants rely on computational measurements and properties of the chemical structures under consideration.[12–20] The most common approach is the Quantitative Structure Activity Relationship (QSAR) method, which is based on the chemical similarity principle. Although these algorithms present good predictions for various compounds, they give no information on the mechanism of action of pollutants, which often requires structural and molecular biology methods of study.

Several factors affect the toxicity of a given chemical, including its dose and route, frequency and duration of exposure. Many modes of molecular toxicity may be generalised as the binding between a chemical and a biomolecular target. The ability of a chemical to interact with a protein has, for a long time, been modelled within the pharmaceutical industry using the molecular docking formalism. By yielding a prediction of the binding affinity of chemicals to targets, putatively related to toxicity, docking could provide molecular-level information and insight into relevant interactions. Since the targets culpable for the underlying adverse effects remain unknown for most environmental chemicals, docking could provide information on the most likely targets. This could, for example, aid in designing $in\ vitro$ HTS bioassays. A multiple-target approach which considers all possible combinations of receptors and ligands could also yield a more complete picture of potential toxicity and the potential perturbation pathways. Furthermore, the docking output could also be combined with other computational approaches, for example to derive QSAR models to predict toxicity.[21–23]

The main result of this paper is a computational ecotoxicity assay (CETOXA), containing 65 prepared protein targets belonging to various protein families. The methods for developing this computational study have been investigated in a previous paper published by our group on an ecotoxicological case study on vitamin B deficiency in moose.[24] Here, we explore the feasibility

of applying a popular classical scoring function, Autodock Vina, and one of the most promising of the newer generation of machine-learning scoring functions, the Random Forest-based scoring function RF-Score-VS, for toxicity screening. The output from docking is a matrix with docking scores, representing an estimate of the binding energies, for all possible combinations of receptors and ligands. Of the two scoring functions considered, Vina shows promising results as it on average predicts higher affinities for active chemicals than for inactive ones. Furthermore, we also show that compound similarity is a strong predictor of binding affinity and that similar chemicals tend to show highest affinity for the same binding site of a given protein. Ultimately, the results may be integrated with existing toxicity and bioassay data to compare chemical profiles, to extract binding patterns and associations that may be key steps in triggering an adverse biological effect.

Methods

A flow-chart for the entire method of deriving the computational ecotoxicity assay is given in Fig. 1. The different steps in the flowchart are detailed below.

Data preparation

Receptors

Based on the assays present in Phases I & II of the ToxCast program,[10] 65 receptors were selected for which experimental structures are available in the Protein Data Bank (PDB). A complete list of the proteins can be found in Table S1 in the Supporting Information (SI), and the list includes kinases (22), phosphatases (5), proteases (6), G-protein-coupled receptors (9), nuclear receptors (11), and cytochromes P450 (4). These proteins represent a broad range of cellular functions that are critical for the survival and proliferation of cells.

PDB structures were chosen based on the availability of high-quality crystal structures. Only deposits with a resolution better than 2.85 Å were included. The protein structure quality was assessed using MolProbity.[25] Structures with an overall MolProbity score (representing a log-

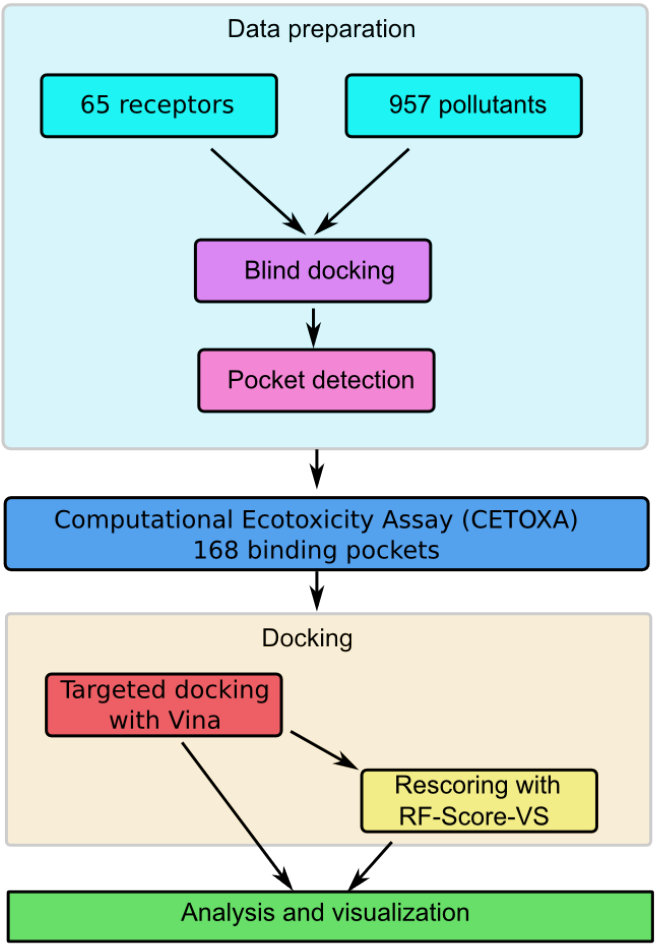


Figure 1: Flow-chart describing how the computational ecotoxicity assay is derived. See text for details.

weighted combination of the clashscore, rotamer, and Ramachandran evaluations) lower than the crystallographic resolution were kept. In many cases, better scores were found for the corresponding structure in the PDB-REDO databank, which contains re-refined and rebuilt PDB entries.[26–28]

Chains were removed as necessary, to ensure that each structure corresponds to the reported biologically-relevant assembly. Ligands, cofactors, and co-crystallised lysozyme molecules were also removed as needed (see Table S1). Zinc ions were kept, given their functional role in proteins, although they are not taken into account during docking. Common structural problems were corrected using Dock Prep[29] through the Chimera software. Missing non-terminal backbone residues were modelled using Modeller[30] and a local energy minimization was performed with the Amber

ff14SB force field[31] through Chimera to relieve potential atomic clashes.

Pollutants

The ligands in the ToxCast chemical library can be broadly separated into four chemical use groups: phthalates and alternative plasticizers, pesticides, pharmaceuticals (both marketed and failed), and other consumer use chemicals (such as food additives, soaps, and shampoos), all together comprising a structurally diverse chemical space. Ligands in reference 10 are denoted by PubChem codes and were downloaded using a script. About 1/3 of the compounds were available only as 2D chemical coordinates and a semi-automated procedure to make 3D coordinates was performed, using the `obgen` tool in OpenBabel[32] followed by manual curation of all the structures. Compounds with unsupported atoms, including metals and, $e.g.$, boron were removed, leaving 957 environmentally relevant compounds.

Both ligands and receptors were treated by scripts from AutodockTools[33] to prepare them for docking.

Identification of Binding Sites

Blind docking was employed to explore the protein surface for prospective binding sites.[34,35] For each potential chemical-target complex, a blind docking calculation was performed using QVina-W (an extension of QVina2, see below, specifically designed for blind docking),[36] with 64 independent runs per docking. Nine results were stored per complex, producing 9 x 957 bound ligand poses per target.

Common binding sites were identified by clustering the 9 x 957 centres of mass of the bound ligands using the OPTICS algorithm,[37] a density-based clustering algorithm similar to DBSCAN, available in the Python machine learning library scikit-learn.[38] To reduce errors associated with inappropriate choice of binding site, multiple binding sites were considered. For each receptor, the centres of mass of up to four of the most populated clusters were stored. In most cases, fewer than four clusters were found by the algorithm (see SI, Table S1). An example of the procedure is

shown in Fig. 2. In this manner, blind docking is used as a method for pocket search.[39]

The receptors, with 1-4 binding sites each, define the computational ecotoxicity assay that can be used for *in silico* scoring of new and existing compounds.

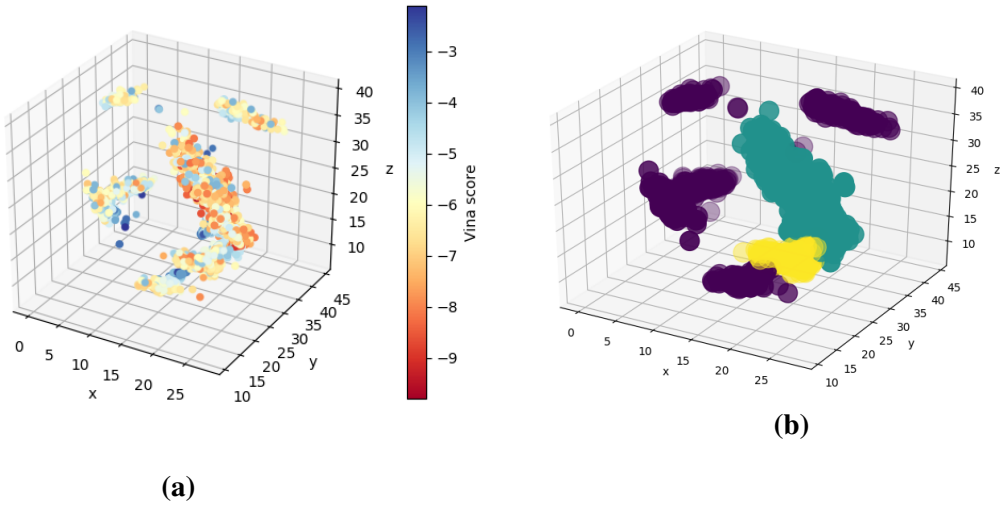


Figure 2: Example of receptor (PDB ID 1HFC) binding site classification, illustrating a) the centres of mass of the blind-docked chemicals and their corresponding Vina score and b) clustering performed by OPTICS. In this example, two clusters were identified, coloured green-blue (population 6708) and yellow (population 922), with the remainder of the (more spread-out) purple points unassigned.

Molecular Docking

AutoDock Vina[40] (or Vina) has been a popular choice for high-throughput screening, due to its efficiency and relatively high accuracy. Here, we use a revised version, Quick-Vina2 (or QVina2), which has improved on the local search algorithm, achieving significant speed-up in computation time without compromising accuracy.[41]

Re-docking co-crystallised ligands

In order to evaluate the docking protocol, preliminary docking simulations were performed in which each of the co-crystallised ligands was re-docked into the active site of its cognate protein target. The coordinates for the ligands were extracted from the PDB files, randomised using

`obconformer` in OpenBabel, and prepared for docking using AutodockTools. The search box dimension was set as in ref. 40. Re-docking was performed with QVina2, with an exhaustiveness level of 8, producing up to nine possible poses for each run. The symmetry-corrected RMSD was calculated using DockRMSD.[42]

Targeted Docking and Scoring

Using the computational ecotoxicity assay, targeted docking of the pollutants was performed on the identified binding sites for each receptor. A total of 160,776 targeted dockings and scorings were performed, using the same settings as during re-docking. The lowest score, corresponding to the strongest binding, of the 1-4 binding sites for each complex was kept, producing a cross-docking matrix of binding free energy predictions for 62,205 complexes, for all possible combinations of receptors and ligands.

Binding affinity prediction

For our purposes, the ability to accurately predict the binding pose is not as important as the ability to accurately rank chemicals based on the binding affinity. Enhancing the accuracy of scoring functions for predicting binding affinities or biological activity remains a challenge. It has been noted that most classical scoring functions suffer from limitations and are unable to accurately predict biological activities.[43] In order to move beyond the limitations of docking codes solvent effects may have to be included,[44–46] at a substantial additional computational cost.

Among the classical scoring functions currently available, Vina has one of the best scoring powers.[47] In recent years, however, a new class of scoring functions have emerged that use a non-parametric machine learning approach instead of imposing a predetermined functional form. Such machine–learning (ML) scoring functions have been found to show improvements over classical scoring functions, in terms of ranking compounds by binding affinity.[48–51] It should be noted that while there is at least one ML scoring function specifically designed to perform well at experimental pose prediction,[52] other ML scoring functions do not necessarily improve success rates.[48,53] For

the application of binding affinity predictions, however, the Random Forest-based scoring function RF-Score developed by Ballester $et\ al.$[54] has shown promising results, outperforming classical scoring functions such as the one used by Vina.[55] In particular, RF-Score-VS was adapted to virtual screening by training also on negative data ($i.e.$, known inactive ligands), although the improvement in performance appears to be less substantial when applied to new targets not included in the training set.[51]

Having generated an ensemble of viable docking poses with Vina, the top scored poses were rescored using the standalone version of RF-Score-VS. The Vina predicted binding affinities will be reported in terms of ΔG_{bind} and the result of RF-Score-VS rescoring in terms of pKd.

Results and discussion

Assessment of the docking protocol

Re-docking co-crystallised ligands

In order to assess the quality and reliability of the docking protocol, the co-crystallised ligands were redocked and the resulting pose compared to the experimental one. Pose generation error is commonly assessed by measuring the root mean square deviation (RMSD) of the predicted pose from the experimental binding orientation. The cumulative frequency curve of the RMSD between the native and predicted conformation is shown in Fig. 3.

With 2 Å as a commonly accepted RMSD cutoff value, the success rate ($i.e.$, fraction of predicted poses with an RMSD $\leq$ 2.0 Å) for the top pose of Vina is 50.0%. For comparison, a benchmark study by Wang $et\ al.$[47] found the average success rate for the top scored pose among academic docking programs to be 47.4%, with second best performance achieved by Vina (with a success rate of 49.0%). Our results are thus within the expected accuracy of the docking method.

Considering instead the best of the nine output poses ($i.e.$, the pose with the lowest RMSD), the success rate increases to 67.2 % (with 80% falling below 3.0 Å, see Fig. 3). In other words,

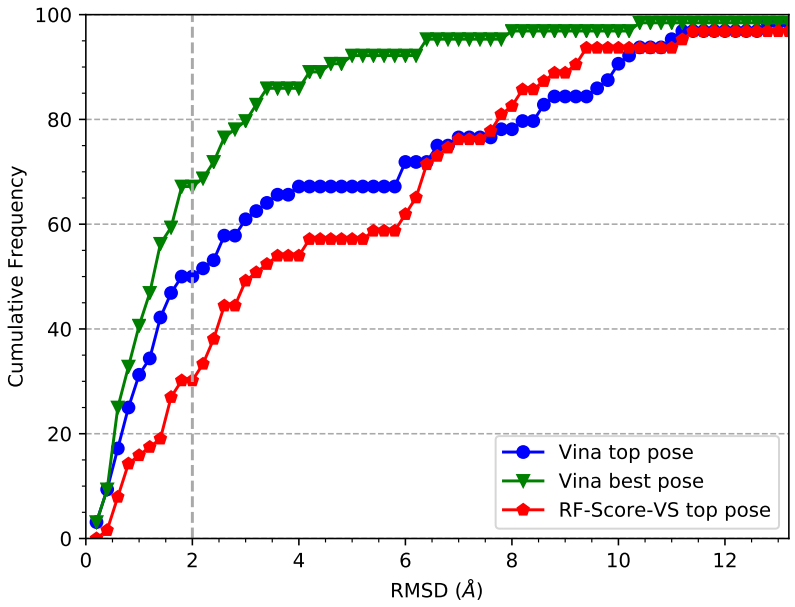


Figure 3: Cumulative frequency curve for the root mean square deviation (RMSD) between the native and predicted conformation of co-crystallised ligands, using the top scored pose and the best among all nine output poses from Vina ($i.e.$, the pose with lowest RMSD), as well as the RMSD for the predicted top pose after rescoring the nine output poses of Vina with RF-Score-VS.

Vina manages to find a pose close to the experimental binding orientation, but may fail to rank it as the best-scoring pose, consistent with previous studies.[39,47,56] It should be noted that a large pose generation error does not by default imply inaccurate binding affinity predictions. A study by Li $et\ al.$ found that there is a low correlation between pose generation error and binding affinity prediction error.[50] For our purposes, the latter is the more relevant quantity.

Rescoring with RF-Score-VS

The nine output poses of Vina were rescored using RF-Score-VS, giving a new top pose. This leads to a drop in success rate to 30.1% (Fig. 3). Note that it is known that machine-learning scoring functions do not necessarily outperform classical scoring functions in regards to pose prediction.[48,53]

Considering instead the reliability of the scoring functions in predicting binding affinities, a comparison of different docking programs using the PDBbind benchmark dataset[57,58] found that

Vina had the best Pearson correlation coefficient and Spearman ranking coefficient between the docked scores and experimental binding affinities, with values of 0.564 and 0.580, respectively.[47] In comparison, using the DUD-E database,[59] Vina had an effective Pearson correlation of 0.18 while RF-Score-VS had a correlation of 0.56 when both training and test sets contained data from all targets, and a more modest correlation of 0.2 when the training and test data were created independently.[51]

The performance of the scoring functions may also be judged by their ability to predict high scores for known binders. The crystal structure of the complexes were scored ($i.e.$, without docking) using Vina and RF-Score-VS. For a direct comparison between the two functions, the Vina score was converted to pKd units by using pKd$= -\log(e)/RT \cdot \Delta$G (where R is the gas constant and $T = 298.15$ K),[60] see Fig. 4. Values of pKd < 4 would suggest weakly bound compounds, while values above 10 would indicate tightly bound compounds.[61] As the activity cutoff, the value of pKd $= 6$ has previously been used to distinguish between active and inactive chemicals.[51] Vina

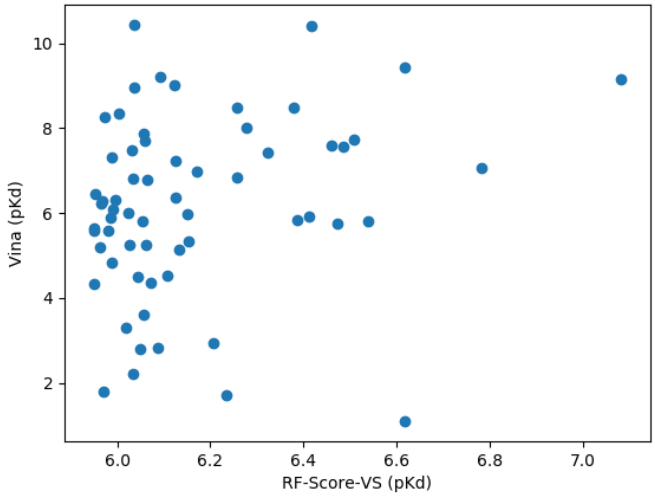


Figure 4: Correlation between RF-Score-VS estimated binding energies following rescoring and predictions from Vina scores converted to pKd units.

predicts the majority (75%) of ligands to have a predicted binding affinity pKd $>$ 5.2 (alt. $<$ -7.1 kcal/mol), with a mean binding affinity of pKd = 6.2 (alt. -8.4 kcal/mol). A number of the co-crystallised ligands are predicted to be weakly bound and some outliers even fall below pKd 4.

Among the latter, relatively poor binding affinities were predicted for a ligand involving halogen-bonding interactions (3W2S), which is not accounted for in most classical scoring functions, and three ligands involving metal-coordination (1HFC, 4G9L, and 4H3X). In comparison, none of the known binders are predicted to be weakly binding by RF-Score-VS. Yet, none are predicted to be particularly potent either, generally falling below pKd 6.6.

Evaluating the linear and monotonic relationship of the two scoring functions, the Pearson's correlation coefficient and Spearman's ranking coefficient are 0.23 and 0.26, respectively. It can be noted that chemicals with high scores from RF-Score-VS also tend to have high Vina scores while no correlation is found for more weakly binding chemicals.

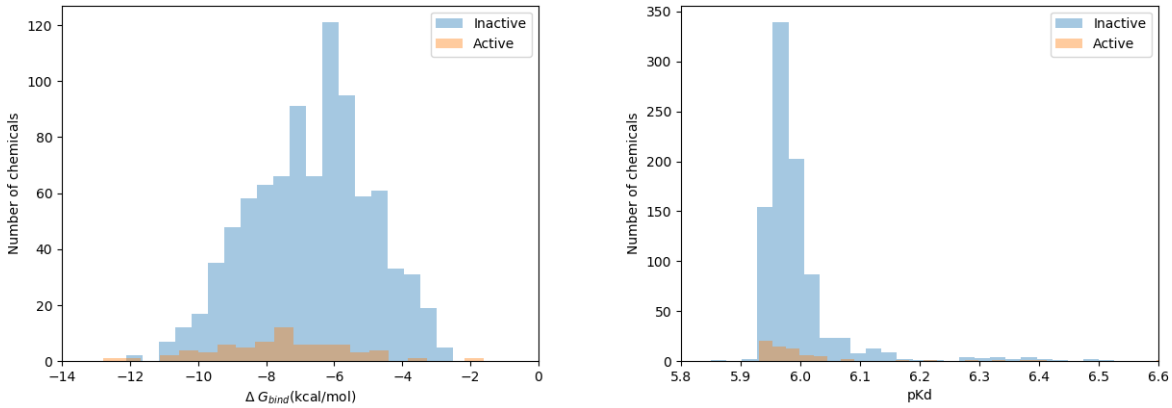
Identification of active chemicals

The pollutants from the ToxCast chemical library were docked to the 1-4 binding sites of each target, identified as shown in Fig. 2, and the Vina score for each complex was stored. The top poses were further rescored with RF-Score-VS.

For toxicity screening, the goal is to minimise the number of false negatives, meaning that no harmful chemicals should be classified as inactive. This poses a particular challenge since environmental chemicals can also elicit adverse effects from weak interactions.

To be able to make predictions on potential activity based solely on the scoring function, it would be necessary to identify a target-specific threshold score which could successfully delineate active chemicals from inactive ones while minimising the number of false negatives. Ideally, the active and inactive sets should be separated to such a degree to enable predictions by simply choosing an appropriate threshold value. Some promising results were found in one toxicity screening of environmentally pertinent chemicals, where two different molecular docking software (eHiTS and FRED) were compared and found to have the capacity to identify weakly active chemicals from inactive chemicals binding to rat estrogen receptors.[62] For the dataset investigated here, however, the two sets show significant overlap. As an example, Fig. 5 shows the sets of active and inactive chemicals binding to CYP 2C9, as determined by the recorded activities in the ToxCast

HTS, as a function of the Vina docking score (Fig. 5(a)) and the RF-Score-VS rescored potencies (Fig. 5(b)). It can be seen that neither the classical Vina scoring function nor the machine-learning scoring function RF-Score-VS are capable of distinguishing chemicals based on activity, making it impossible to use the docking score alone to separate likely active and inactive chemicals.



(a) **(b)**

Figure 5: For a set of active and inactive chemicals binding to CYP 2C9, distribution of the (a) Vina and (b) RF-Score-VS predicted binding affinities for each complex.

The application of molecular docking for toxicity screening is an intriguing method, however, the results indicate the limitations of using current scoring functions to predict chemical activity. Possible routes for improving the performance could be to investigate complexation in a more detailed fashion through, $e.g.$, molecular dynamics methods. Although implicit-solvent models remain popular, there is quite some evidence that they do not give sufficiently accurate results for predicting binding free energies.[63–66] Therefore, free energy calculations using explicit solvent will likely be required to get accurate estimates of binding energies.[67] In some cases, quantum-chemical estimates of binding-strength for toxin-receptor models[68,69] may yield additional insights as well. Finally, whereas the purpose of drug discovery is to increase the enrichment factor, focusing only on the best ranking compounds, toxicity screening using molecular docking should also consider weakly binding chemicals, as these may still elicit adverse biological effects. This requires additional considerations and places new requirements on scoring functions.[70]

Although a critical score cannot be set, Vina on average predicts better scores for active chem-

icals, illustrated as follows. The average binding affinity for active and inactive chemicals to each target which had at least 5 active chemicals was calculated and the set values subtracted, $\langle \Delta G_{bind,act} \rangle - \langle \Delta G_{bind,inact} \rangle$. The difference is plotted in Fig. 6 and can be seen to be mostly negative. These are encouraging results as Vina would appear to somewhat consistently predict stronger binding affinities for active chemicals than for inactive ones. A similar trend can be seen for RF-Score-VS, although showing a poorer performance, at best producing a difference of -0.27 kcal/mol.

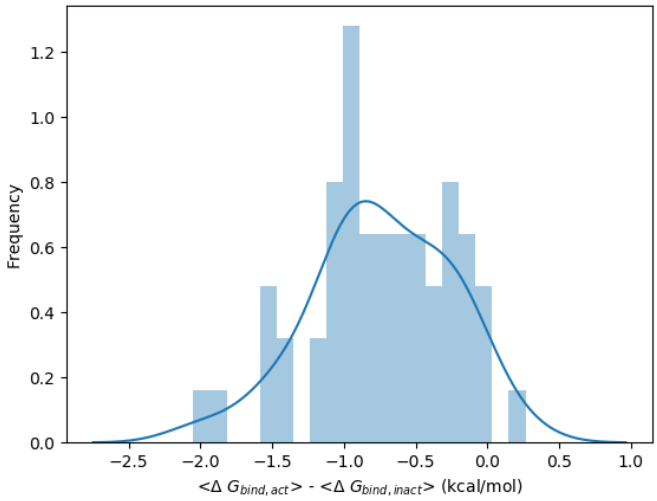


Figure 6: The difference in target-averaged predicted binding affinities for active and inactive chemicals, for targets with at least 5 active chemicals. The solid line shows the kernel density estimation by Seaborn, an estimate of the probability distribution from which the data is drawn.

Chemical similarity of ligands

It is generally assumed that structurally similar molecules exhibit similar biological activities.[71] This notion is, for example, used in drug discovery to search chemical libraries for matching compounds[72,73] and to generate information about common receptors.[74] Toxicological data gaps may also be filled by *read-across*, in which chemicals with known toxicity are used to predict the toxicity of untested chemicals, based on their chemical similarity.[75–77] It should be noted, however, that structurally similar compounds may in some cases show large differences in potency, due to

so-called activity cliffs.[78,79]

In this context, reproducibility of docking results may be evaluated by testing whether the docking program can reproduce the same binding pattern for structurally similar chemicals (see SI for details). The Tanimoto coefficient for pairs of chemicals was computed and the generated similarity matrix hierarchically clustered, see Fig. 7. A value of zero implies no shared chemical fragments and a value of one represents identity. The dendrogram in Fig. 7 is labeled by the four chemical categories used in ToxCast. Some degree of clustering may be seen within the chemical class groups, notably among pesticides and pharmaceuticals. Overall, however, the chemical library is structurally diverse with only few chemicals classified as highly similar.

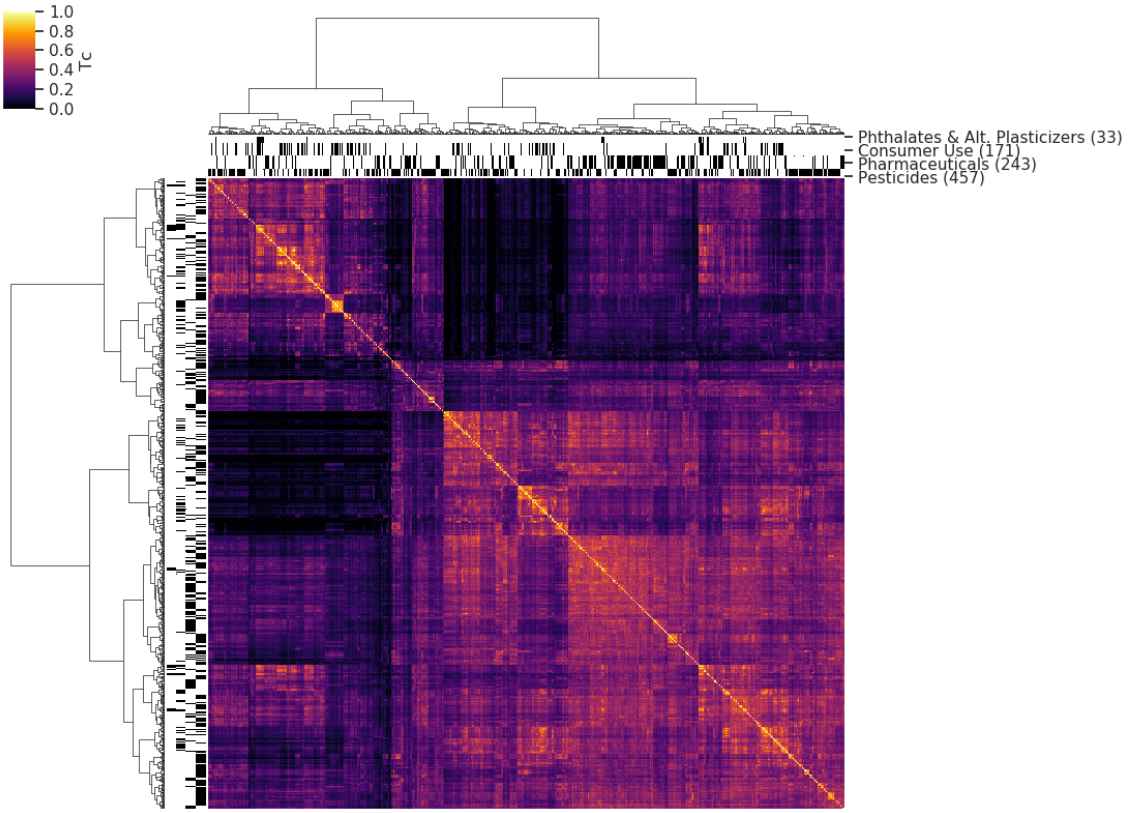


Figure 7: Hierarchically clustered chemical-chemical similarity matrix with the dendrogram showing the four chemical categories. A Tanimoto coefficient approaching 1.0 (yellow) indicates a high degree of similarity between the chemical pair whereas zero (dark purple) indicates no shared fragments.

Binding patterns were assessed for each pair of chemicals by computing the mean absolute deviation (MAD) in binding affinity to the various targets. A low MAD value would suggest

similar binding behaviour to the different targets and vice versa. These were correlated to the Tanimoto coefficient by colour-encoding the clustered similarity matrix with the MAD values (Fig. 8). A qualitative comparison of the two matrices reveals that clusters along the diagonal in Fig.

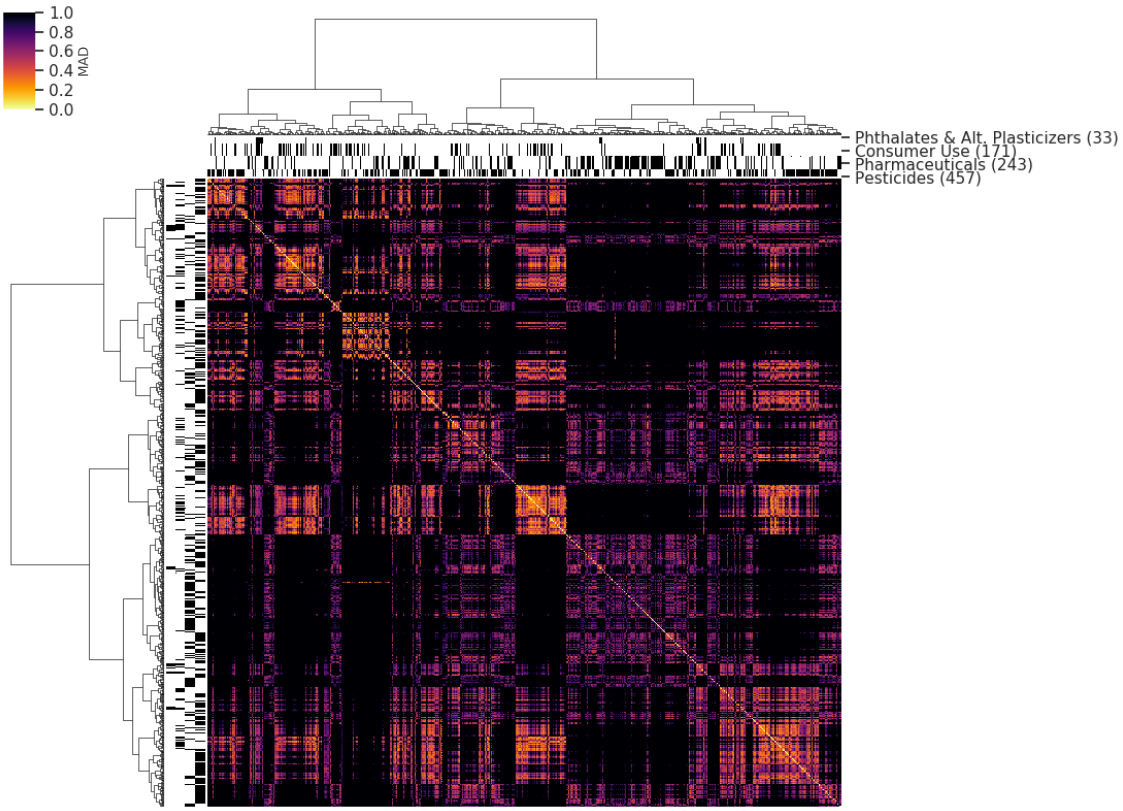


Figure 8: Clustered chemical-chemical similarity matrix colour-encoded by the mean absolute deviation in Vina-predicted binding affinities. Note that the errors range between 0.0-7.7 kcal/mol, however, the upper limit of the colormap has for clarity been anchored to 1.0.

7, corresponding to blocks of chemicals with shared fragments, also interact with similar binding affinities with the different protein targets, as indicated by their lower MAD values in Fig. 8. Low MAD values are also observed for pairs of highly dissimilar chemicals, notably for a group of consumer use chemicals and pesticides. These would appear to be chemicals which bind weakly to all targets, typically with an average Vina score $>$ -6.0 kcal/mol.

A more rigorous approach was employed for a quantitative comparison, by looking at the spread in binding affinities among similar compounds. For each query compound in the pollutant library, a subset of N_{sim} similar chemicals was stored, selected based on the corresponding Tan-

imoto threshold (see SI). If $N_{sim} > 3$, the standard deviation in binding affinity for this subset, std(ΔG_{bind}), was calculated (Fig. 9). For comparison, a subset of least similar chemicals as well as randomly chosen chemicals was also stored. As reference, the target-based standard deviation of all chemicals and the overall standard deviation among all docked complexes (vertical line in Fig. 9) are shown. It can be seen that compound similarity is generally a strong predictor of

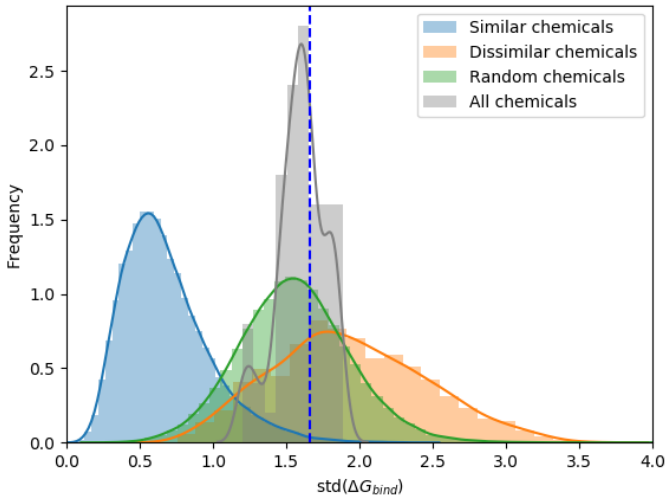


Figure 9: Distribution of standard deviations in binding affinities for subsets of chemicals identified as highly similar to each query molecule. As reference, the plot includes the distribution for subsets of highly dissimilar and randomly chosen chemicals, as well as the standard deviation of all chemicals docked to each target and overall standard deviation of all complexes (vertical dashed line). The solid lines show the corresponding kernel density estimations.

binding affinity, with Vina predicting more tightly packed scores for the most similar compounds. Conversely, the chemicals least similar to the query molecule also have a wider spread in affinities compared to the query. It should be noted that the latter subset may still contain chemicals that are similar to one another.

Having compared patterns in binding affinity, it is also of interest to investigate whether similar chemicals show highest affinity for the same binding site of a given protein. At the onset, each protein was assigned up to 4 docking sites. Using the same subsets as above, for each protein with at least 2 docking sites and each query molecule binding with the lowest Vina score to site A, the fraction $pA = N_{sim,A}/N_{sim}$ of similar chemicals also binding best to A was calculated.

The value was normalized by the number of hits on each binding site, $i.e.$, $pAtot = N_{tot,A}/N_{tot}$, to remove any potential bias for a more promiscuous binding site. The histogram of $pA/pAtot$ is shown in Fig. 10, where a value greater than 1 indicates that the compounds are more likely to bind to the same site. It can be seen that Vina predicts structurally similar chemicals to preferentially bind to the same site compared to randomly chosen chemicals, with a distribution skewed towards larger $pA/pAtot$ values, whereas the least similar ones, which may be significantly different in terms of size and functional groups present, have a distribution distinctly skewed towards zero, $i.e.$, dissimilar chemicals prefer to bind to an alternative site.

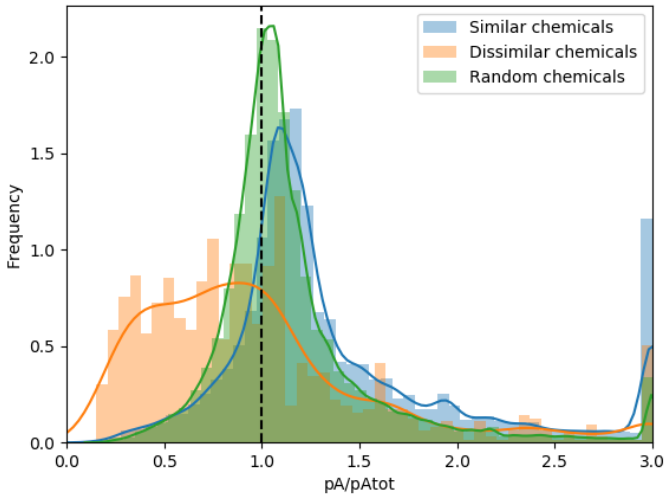


Figure 10: Histogram of the normalized fraction of chemicals binding to the same site as a given query molecule, for molecules identified as similar or dissimilar, and for randomly chosen chemicals. The solid lines show the corresponding kernel density estimations. Note that values greater than 3.0 in the distribution were set to the limit, for clarity.

The fact that binding is in general largely correlated with structural similarity motivates the use of chemical similarity in predicting toxicity, as mentioned earlier. Based on available toxicity data, this would allow us to identify protein-ligand associations and predict a set of chemical fragments likely to contribute to the toxicity of a given molecule, which could be highly relevant in drug design. An example of the implementation is shown in SI.

Hierarchical clustering of binding affinities

Hierarchical clustering was performed using the Seaborn package in Python, with Euclidean distance as the similarity metric and Ward variance minimization algorithm as the linkage method. Figure 11 shows the clustered matrix of binding affinities (heatmap) as predicted by Vina, illustrating sectors ranging from low to high scores. The corresponding heatmap rescored using RF-Score-VS is shown in SI. Similar to what was seen above for re-docking, RF-Score-VS when applied to an independent dataset appears to underestimate the binding affinities of tightly binding ligands and overestimate the affinities of weakly binding ligands.

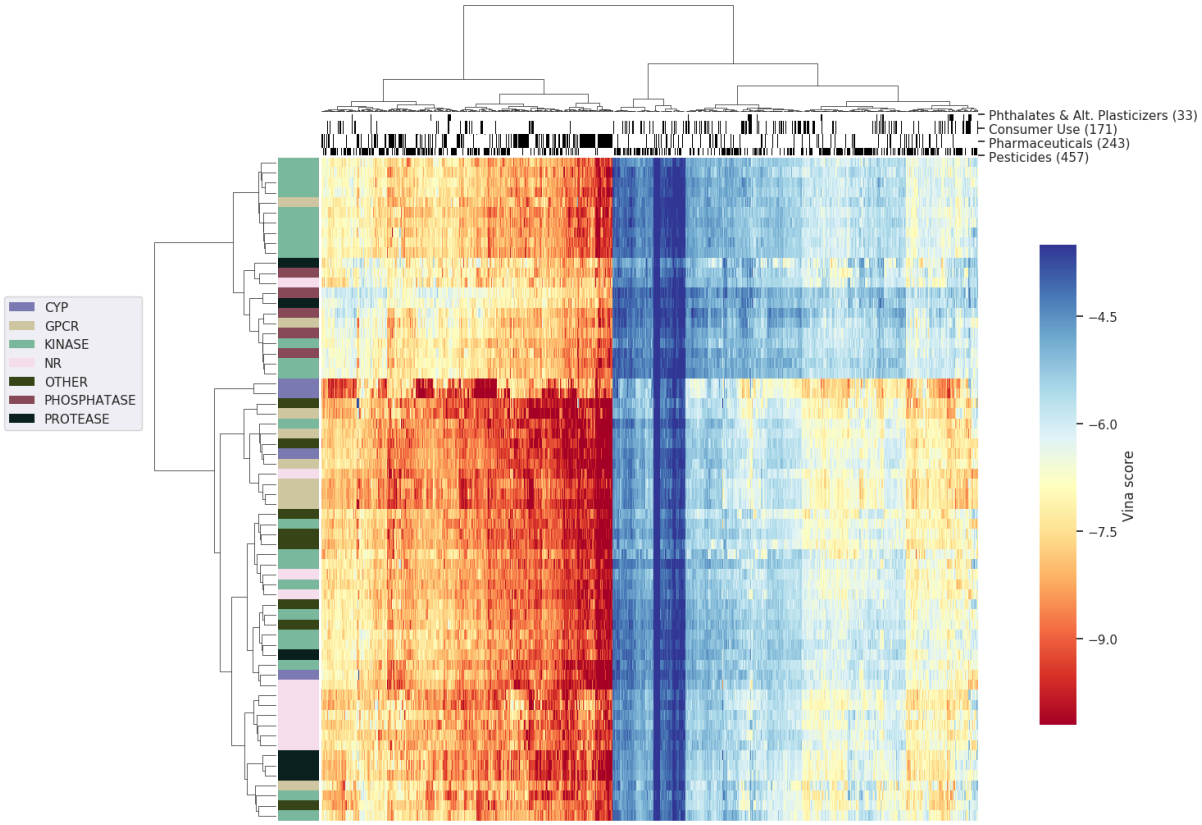


Figure 11: Hierarchical clustering of 957 chemicals by 65 receptors, based on the Vina predicted binding free energies. Left bar indicates clusters of seven protein families and top bars indicate clusters of four chemical use groups, following the ToxCast classification.

Some prominent protein family clusters can be seen, particularly homogeneous clusters of NR, Kinase, and GPCR. Among the various protein families, CYP and GPCR are predicted to be the most affected, with over half the chemicals binding with an affinity $<$ -8 kcal/mol to CYP 1A1,

CYP 1A2, and the muscarinic acetylcholine receptors. Comparing chemicals in the different chemical use categories, consumer use chemicals and phthalates and alternative plasticisers tend to have low predicted binding affinities while pesticides and pharmaceuticals show a broader distribution of Vina scores. This can be seen more clearly in Fig. 12, which shows the binding free energy, averaged over all receptors, for pollutants within each chemical use group.

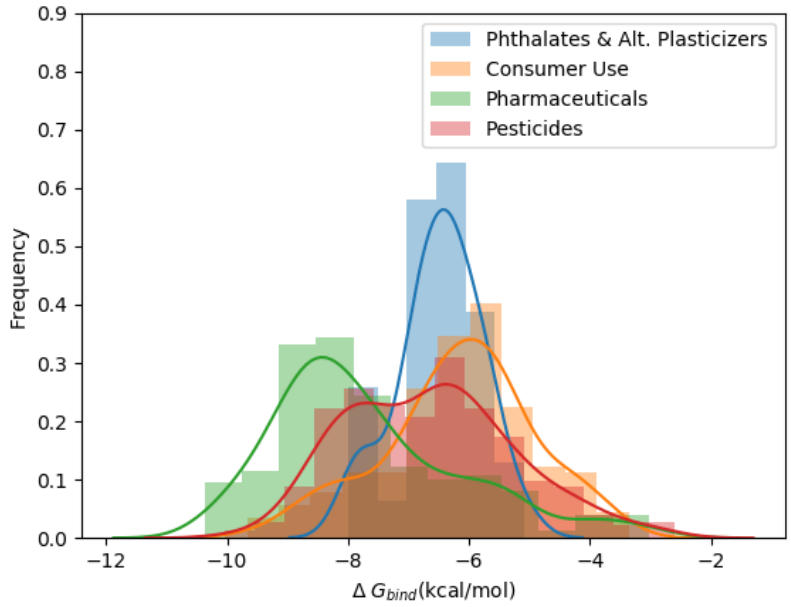


Figure 12: Histogram of predicted ligand binding free energies for each chemical averaged over all receptors, shown separately for each chemical class (phthalates and alternative plasticizers, consumer use, pharmaceuticals, and pesticides). Solid lines show the kernel density estimates.

A wide band of pesticides, pharmaceuticals, and consumer use products in Fig. 11, corresponding to 10% of all chemicals considered, have low binding affinities ($\Delta G_{bind} >$ -6 kcal/mol) to all protein targets. Among the chemicals that show specificity by binding to only a few targets with higher affinities are Mirex, a persistent organic pollutant, binding to the farnesoid X receptor α (ΔG_{bind} = -8.4 kcal/mol) and PI3K α (ΔG_{bind} = -7.7 kcal/mol); Pentachlorophenol, a pesticide and disinfectant, binding to CYP 1A1 and 1A2 (ΔG_{bind} = -7.6 and -7.7 kcal/mol, respectively); Lindane, an insecticide, pediculicide, and scabicide, binding to CYP 1A2 (ΔG_{bind} = -7.5 kcal/mol). Interestingly, Mirex has been found to inhibit particularly Adenosine triphosphatase,

a protein which is central for liver function.[80] This enzyme shares structural motifs with Phosphoinositide 3-kinase, which binds Mirex as calculated by CETOXA with a free energy of -7.7 kcal/mol. The common features between these two proteins suggest that the docking procedure detects structural similarities of particular enzyme motifs, even though these two proteins pertain quite different 3D structures. The second important finding is the docking result associated with the Farnesoid receptor. Indeed, this receptor is critical for liver function and has a central role in the detoxification mechanism in hepatocytes. This protein is also related to carcinogenesis and its inhibition causes considerable liver and gallbladder complications.[80–82] The result that Mirex binds to the Farnesoid receptor is thus relevant for the biochemistry of this receptor, as it is for the case of Adenosine triphosphatase. Another interesting finding is that pentachlorophenol binds strongly to CYP1A1 and -1A2. Both these enzymes use O-deethylation to oxidize phenolic substances and both are expressed in the liver and perform detoxification in the liver microsomes of pyrimidine-like and phenolic xenobiotics such as caffeine, alpha-napthoflavone and 7-ethoxyocourmarine and phenacetin.[83] The results also indicate that lindane may be destined for detoxification by CYP1A2. Interestingly, lindane is a hexa-chlorinated compound that binds particularly well to CYP1A2, as CYP1A2 has a 30% higher affinity towards halogenated compounds compared to CYP1A1. This indicates that the docking procedure in this case has the potential to distinguish enzymes with very similar specificity with however a 30% preference difference towards halogenated aromatics.[83]

Among the plasticizers, one phthalate (butyl benzyl phthalate) and three alternative plasticizers (dipropylene glycol dibenzoate, pentane-1,5-diyl dibenzoate, and hexane-1,6-diyl dibenzoate) showed promiscuous behaviour, targeting over 1/3 of the receptors with $\Delta G_{bind} <$ -8 kcal/mol. Among the consumer use chemicals identified as potentially highly promiscuous are the surfactants perfluorodecanoic acid and perfluoroundecanoic acid, as well as the widely used synthetic food dyes allura red and FD&C Blue no. 1. Among the pesticides targetting multiple targets with high binding affinities are Famoxadone, Novaluron, and Prosulfon. The interactions of Famoxadone with multiple protein targets has been confirmed in the literature. In a study by ,[84] Famoxadone was found to inhibit both mitochondrial enzymes as well as the cytochrome system (particularly

CYPBc1[85]).

The heatmap notably contains a cluster of about 28 pharmaceuticals with high Vina scores, binding to multiple targets. These include failed or terminated drugs, such as the thiazolidinedione based and non-thiazolidinedione based antihyperglycemic agents Troglitazone and Farglitazar, which are specific ligands for peroxisome proliferator-activated receptors (PPAR). Troglitazone has been found to have hepatotoxic effects,[86] while Farglitazar did not reach past phase III clinical trials. Zamifenacin in this list was also identified to have highly promiscuous biological activity in the ToxCast HTS.[10] Although drugs are typically designed to interact with a specific target, unintended drug-target interactions are one of the major challenges in drug design, associated with side-effects and high failure rates.[87,88]

Assessment of Vina high scores

The highest binding affinity per chemical in the heat map was extracted and the top 20 chemical-target pairs are shown in Table 1.

Table 1: Top 20 Vina scores.

Target	**Chemical**	ΔG_{bind} **(kcal/mol)**
CYP 1A2	7,12-Dimethylbenz[a]anthracene	-15.6
CYP 1A2	Benzo[b]fluoranthene	-15.2
CYP 1A2	Benz[a]anthracene	-14.2
CYP 1A1	PharmaGSID_47330	-13.4
CYP 1A2	Fluroanthene	-12.9
CYP 1A2	Pyrene	-12.9
CYP 1A2	1-Hydroxypyrene	-12.9
COX2	SB236057A pharma	-12.9
CYP 2C9	GSK163929B	-12.8
AT1	Milbemycin A3	-12.7
Sirt2	SB243213A pharma	-12.6

COX2	sodium 3-(4-{[1-(3-ethoxyphenyl)-2-(4-methyl phenyl)-1H-imidazol-4-yl]carbonyl}piperazin-1-yl)naphthalene-1-carboxylate (pharma)	-12.5
ADRB2	Farglitazar	-12.4
mGluR1	SB413217A	-12.4
COX2	FD&C Blue no. 1	-12.4
CYP 1A2	C.I. Solvent yellow 14	-12.2
mGluR1	3-pyridinecarboxamide, 2-(2,1,3-benzoxadiazol-5-yloxy)-N-[[4-(1-hydroxy-1-methylethyl)phenyl]methyl]- (pharma)	-12.2
mGluR1	3-({(3R,4R)-6-[(5-fluoro-1,3-benzothiazol-2-yl)methoxy]-4-hydroxy-3,4-dihydro-2H-chromen-3-yl}methyl)benzoic acid, CP-85958	-12.2
AurA	1-[(3S)-4-[{[3,5-bis(trifluoromethyl)phenyl] carbonyl}(methyl)amino]-3-(3,4-dichlorophenyl)butyl]-4-cyclohexyl-1-azoniabicyclo[2.2.2]octane methanesulfonate	-12.2
COX2	SB413217A	-12.2

This table contains the polycyclic aromatic hydrocarbons (PAH) 7,12-Dimethylbenz[a]anthracene, Benzo[b]fluoranthene, Benz[a]anthracene, Fluroanthene, Pyrene, and 1-Hydroxypyrene in complex with cytochrome P450 1A1 and 1A2. PAHs are among the most ubiquitous environmental carcinogens and known to be potent P450 inhibitors.[89–91] 7,12-Dimethylbenz[a]anthracene and benz[a]anthracene are two short and linear PAHs, which both exert genotoxicity and toxicity by the triple arrangement of their aromatic rings. They expose their lateral electrons for oxidation by detoxification enzymes and are then converted to their diol-epoxide forms which react with DNA.[92,93] These compounds are ranked as a top compound to target CYP1A2 (Table 1), which reflects well with empirical studies[94] showing a high preference of CYP1A2 towards ethoxyresorufin. Ethoxyresorufin is geometrically similar to these compounds, suggesting the docking system recognizes the conserved molecular volume and geometry. There are no studies on fluroanthene and benzo[b]fluoranthene which could confirm their interaction with CYP1A2, however, by their elongated geometries, it is highly feasible that CYP1A2 is the detoxification enzyme

for these two compounds, as both are known to be genotoxic.[95,96] A further study showed that CYP1A2 is activated by benzo[b]fluroanthene, however the expression of CYP1A2 was not confirmed.[96] In the prediction, the docking procedure indicated that CYP1A2 was the preferred target for benzo[b]fluroanthene, which indicates that physiological responses may be different than cellular responses when considered on a cohort of individuals exposed to the compound via air-particulate matter.

The potentially genotoxic and carcinogenic[97] food dye C.I. Solvent yellow 14 is predicted to bind strongly to both P450 1A1 and 1A2, with an affinity of -11.7 kcal/mol and -12.2 kcal/mol, respectively. Although 1A1 and 1A2 are highly homologous, CYP1A1 is considered to be most efficient in metabolizing C.I. Solvent yellow 14, whereas other CYPs, including CYP1A2, were previously found to be almost ineffective.[98] The discontinued PPAR agonist Farglitazar is predicted to bind strongly to the beta-2 adrenergic receptor, although it was developed to treat type 2 diabetes.[99] It has a predicted binding affinity to PPAR α and PPAR γ of -9.4 kcal/mol and -11.3 kcal/mol, respectively. In other words, many of the Vina high scores can be correlated with known toxic compounds.

Assessment of known interactions

Based on literature data and as considered in the ToxCast publications,[9,10] we assessed whether Vina succeeds to predict high scores for chemicals that are known to interact with certain proteins. Some of these known complexes and their predicted binding affinities are listed in Table 2.

Bisphenol A[100] and 2,2-bis(4-hydroxyphenyl)-1,1,1-trichloro ethane,[101] a metabolite of the pesticide methoxychlor, are known estrogen receptor agonists. Perfluorooctanoic acid (PFOA) and perfluorooctane sulfonic acid (PFOS) are chemicals with widespread use and with multiple reported toxicities, generally thought to be triggered by activating PPAR α.[102,103] In the HTS screening in ref. 9, both were active for PPAR γ, but only PFOA was active for PPAR α. In our case, the predicted binding affinity is relatively high also for PPAR γ. It should be noted that one $in\ vivo$ study found that neither compound activates PPAR γ,[104] illustrating potential risks in extrapolat-

ing *in vitro* or *in silico* results to *in vivo*. The pesticide lactofen[105] is also an expected PPAR activator found to be positive for PPAR assays in ToxCast.[9] The pharmaceuticals CP-471358 and CP-544439 are known inhibitors of matrix metalloproteinases MMP2, MMP9, and MMP13,[106,107] and these complexes all have high Vina scores. However, relatively low predicted binding affinities are found for the complexes involving the androgen receptor (AR). The pesticides linuron, prochloraz, and vinclozolin are known to cause toxicity by interacting with AR as antagonists.[108] Binding to the ligand binding domain of nuclear receptors is known to cause structural and dynamic changes in the protein.[109] For example, one study used computer simulations to follow the long-time scale conformational fluctuations of AR interacting with different ligands and found that agonists and antagonists induce distinct conformational changes.[110] Since the androgen receptor used for docking here corresponds to an agonist complex,[111] with an associated redocking score of -8.9 kcal/mol, the low binding affinities of the (antagonist) pesticides may be due to the shortcomings of neglecting protein flexibility. This issue can be addressed in the future by expanding the computational ecotoxicity assay to include multiple protein structures for the same receptor, with both agonist and antagonist co-crystallised conformations to account for multiple binding modes.

The HTS screening in ref. 10 in some cases missed known active compounds. Zamifenacin is a muscarinic antagonist[112] that did not show activity for the M1 assay. 5,5-diphenylhydantoin is a known substrate and inhibitor for many CYPs,[113] but was not found to inhibit any of the CYPs in the HTS screening. These chemicals show high predicted binding affinities to their respective targets. Additional chemical-target interactions that were considered, including estrogenic and nonestrogenic compounds within the EPA's Endocrine Disruptor Program, can be found in SI Tables S2 and S3.

Table 2: Known chemical-protein complexes, based on literature data (see text), and their corresponding binding affinities as predicted by Vina.

Target	**Chemical**	**ΔG_{bind} (kcal/mol)**
Estrogen receptor	Bisphenol A	-8.7
	2,2-bis(4-hydroxyphenyl)-1,1,1-trichloroethane	-9.3

PPAR α	Perfluorooctanoic acid (PFOA)	-9.4
	Perfluorooctane sulfonic acid (PFOS)	-10.3
	Lactofen	-8.9
PPAR γ	Perfluorooctanoic acid (PFOA)	-8.5
	Perfluorooctane sulfonic acid (PFOS)	-8.8
	Lactofen	-9.3
MMP3	Perfluorooctane sulfonic acid (PFOS)	-9.0
MMP13	Perfluorooctane sulfonic acid (PFOS)	-9.2
Androgen receptor	Linuron	-7.2
	Prochloraz	-6.7
	Vinclozolin	-6.2
MMP9	CP-471358 pharma	-9.5
	CP-544439 pharma	-10.2
MMP13	CP-471358 pharma	-9.1
	CP-544439 pharma	-9.9
M1 muscarinic receptor	Zamifenacin	-10.5
M2 muscarinic receptor		-8.6
M4 muscarinic receptor		-10.8
CYP 1A1	5,5-diphenylhydantoin	-9.0
CYP 1A2		-9.0
CYP 2C9		-8.9
CYP 2D6		-9.5

Assessment of Vina low scores

Considering potentially missed interactions and weakly binding active chemicals, the 20 least promiscuous chemicals in the docking simulations were checked for measured activity in the HTS assays. Of these 1300 complexes with low predicted binding affinities, only 8 were reported as active in ToxCast. These assays include a number of nuclear receptors which, as noted above, may undergo conformational changes upon ligand binding. In one case, sodium nitrite in complex with

monoamine oxidase A (MAO), the low predicted binding affinity may be due to the mechanism of inhibition, as nitrates would appear to inhibit MAO through oxidation of the SH-groups,[114] a mechanism not captured in current docking methods.

Potential refinement of the protocol

In this work, we assessed the strengths and weaknesses in employing molecular docking for screening potential toxicity of xenobiotic compounds. This approach allows the generation of interaction maps between potential toxins and targets linked to perturbation pathways in a a cost-efficient and timely manner, yielding molecular-level information. While neither the classical Vina scoring function nor the machine-learning scoring function RF-Score-VS considered here were capable of distinguishing active chemicals, Vina on average predicted higher affinities for active chemicals and succeeded in identifying several interactions that could be confirmed in the literature. To be able to achieve higher docking accuracy with fewer false negatives, the method can be further developed by considering, for example, protein side-chain conformational changes, covalent interactions, charge redistribution, and bio-transformation products. A further refinement could be to consider multiple structures per target to take into account multiple binding modes. More exact methods may be required to distinguish binders from non-binders and the obvious challenge there remains to balance computational cost and accuracy. Finally, current scoring functions are limited in their ability to predict biological activity, underscoring the need to consider other properties of the ligand or to explore new scoring functions optimised on minimising the number of false negatives.

Acknowledgement

The authors are grateful for funding from eSSENCE - The e-Science Collaboration (Uppsala-Lund-Umeå, Sweden). Samuel Ferrer is acknowledged for enthusiastic discussions.

Supporting Information Available

Additional background text including tables and figures describing details of the computational ecotoxicity assay. This information is available free of charge via the Internet at http://pubs.acs.org. The protein structures as well as the ToxCast ligands prepared for docking using QVina2 are available at http://github.com/dspoel/cetoxa.

References

(1) Ceballos, G.; Ehrlich, P. R.; Dirzo, R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci. U.S.A* **2017**, *114*, E6089–E6096.

(2) Manzetti, S.; van der Spoel, E. R.; van der Spoel, D. Chemical properties, environmental fate, and degradation of seven classes of pollutants. *Chem. Res. Toxicol.* **2014**, *27*, 713–737.

(3) Guengerich, F. P. *Cytochrome P450: Structure, mechanism, and biochemistry*; Springer US: Boston, MA, 2005; pp 377–530.

(4) Lee, D.-H.; Steffes, M. W.; Sjödin, A.; Jones, R. S.; Needham, L. L.; Jacobs, D. R. Low dose of some persistent organic pollutants predicts type 2 diabetes: a nested case-control study. *Environ. Health Perspect.* **2010**, *118*, 1235–1242.

(5) Ray, D. E.; Richards, P. G. The potential for toxic effects of chronic, low-dose exposure to organophosphates. *Toxicol. Lett.* **2001**, *120*, 343–351.

(6) Judson, R.; Richard, A.; Dix, D. J.; Houck, K.; Martin, M.; Kavlock, R.; Dellarco, V.; Henry, T.; Holderman, T.; Sayre, P.; Tan, S.; Carpenter, T.; Smith, E. The toxicity data landscape for environmental chemicals. *Environ. Health Perspect.* **2009**, *117*, 685–695.

(7) Hartung, T.; Rovida, C. Chemical regulators have overreached. *Nature* **2009**, *460*, 1080–1081.

(8) Council, N. R. *Toxicity testing in the 21st century: a vision and a strategy*; Washington, DC: The National Academies Press, 2007.

(9) Judson, R. S.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Mortensen, H. M.; Reif, D. M.; Rotroff, D. M.; Shah, I.; Richard, A. M.; Dix, D. J. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ. Health Perspect.* **2010**, *118*, 485–492.

(10) Sipes, N. S.; Martin, M. T.; Kothiya, P.; Reif, D. M.; Judson, R. S.; Richard, A. M.; Houck, K. A.; Dix, D. J.; Kavlock, R. J.; ; Knudsen, T. B. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem. Res. Toxicol.* **2013**, *26*, 878–895.

(11) Shockley, K. R. Quantitative high-throughput screening data analysis: challenges and recent advances. *Drug Discov. Today* **2015**, *20*, 296–300.

(12) Thomas, R. S. et al. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol. Sci.* **2019**, *169*, 317–332.

(13) Rusyn, I.; Daston, G. P. Computational toxicology: realizing the promise of the toxicity testing in the 21st century. *Environ. Health Perspect.* **2010**, *118*, 1047–1050.

(14) Hubal, E. A. C.; Richard, A. M.; Shah, I.; Gallagher, J.; Kavlock, R.; Blancato, J.; Edwards, S. W. Exposure science and the U.S. EPA National Center for Computational Toxicology. *J. Expo. Sci. Environ. Epidemiol.* **2010**, *20*, 231–236.

(15) Helma, C.; Vorgrimmler, D.; Gebele, D.; Gütlein, M.; Engeli, B.; Zarn, J.; Schilter, B.; Piparo, E. L. Modeling chronic toxicity: a comparison of experimental variability with (Q)SAR/read-across predictions. *Front. Pharmacol* **2018**, *9*, 413.

(16) Manganelli, S.; Roncaglioni, A.; Mansouri, K.; Judson, R. S.; Benfenati, E.; Manganaro, A.;

Ruiz, P. Development, validation and integration of in silico models to identify androgen active chemicals. *Chemosphere* **2019**, *220*, 204–215.

(17) Strope, C.; Mansouri, K.; Clewell, H.; Rabinowitz, J.; Stevens, C.; Wambaugh, J. High-throughput in-silico prediction of ionization equilibria for pharmacokinetic modeling. *Sci. Total Environ.* **2018**, *615*, 150–160.

(18) Zang, Q.; Mansouri, K.; Williams, A.; Judson, R.; Allen, D.; Casey, W.; Kleinstreuer, N. In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. *J. Chem. Inf. Model* **2017**, *57*, 36–49.

(19) Browne, P.; Judson, R.; Casey, W.; Kleinstreuer, N.; Thomas, R. Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ. Sci. Technol.* **2015**, *49*, 8804–8814.

(20) Chushak, Y. G.; Shows, H. W.; Gearhart, J. M.; Pangburn, H. A. In silico identification of protein targets for chemical neurotoxins using ToxCast in vitro data and read-across within the QSAR toolbox. *Toxicol. Res.* **2018**, *7*, 423–431.

(21) Yao, Z.; Lin, Z.; Wang, T.; Tian, D.; Zou, X.; Gao, Y.; Yin, D. Using molecular docking-based binding energy to predict toxicity of binary mixture with different binding sites. *Chemosphere* **2013**, *92*, 1169–1176.

(22) Zou, X.; Lin, Z.; Deng, Z.; Yin, D.; Zhang, Y. The joint effects of sulfonamides and their potentiator on Photobacterium phosphoreum: differences between the acute and chronic mixture toxicity mechanisms. *Chemosphere* **2012**, *86*, 30–35.

(23) Li, F.; Xie, Q.; Li, X.; Li, N.; Chi, P.; Chen, J.; Wang, Z.; Hao, C. Hormone activity of hydroxylated polybrominated diphenyl ethers on human thyroid receptor-β: In vitro and in silico investigations. *Environ. Health Perspect.* **2010**, *118*, 602–606.

(24) Yildirim, A.; Zhang, J.; Manzetti, S.; van der Spoel, D. Binding of pollutants to biomolecules: a simulation study. *Chem. Res. Toxicol.* **2016**, *29*, 1679–1688.

(25) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall III, W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375–W383.

(26) Joosten, R. P. et al. PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J. Appl. Cryst.* **2009**, *42*, 376–384.

(27) Joosten, R. P.; Joosten, K.; Murshudov, G. N.; Perrakis, A. PDB_REDO: constructive validation, more than just looking for errors. *Acta Cryst.* **2012**, *D68*, 484–496.

(28) Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **2014**, *1*, 213–220.

(29) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *13*, 1605–12.

(30) Fiser, A.; Do, R. K.; Sali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.

(31) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput.* **2015**, *11*, 3696–3713.

(32) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.

(33) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.;

Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

(34) Hetényi, C.; van der Spoel, D. Efficient docking of peptide ligands to proteins. *Protein Sci.* **2002**, *11*, 1729–1737.

(35) Hetényi, C.; van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Letters* **2006**, *580*, 1447–1450.

(36) Hassan, N. M.; Alhossary, A. A.; Mu, Y. M.; Kwoh, C.-K. Protein-ligand blind docking using QuickVina-W with inter-process spatio-temporal integration. *Scientific Reports* **2017**, *7*, 15451.

(37) Ankerst, M.; Breunig, M. M.; Kriegel, H. P.; Sander, J. OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD Record* **1999**, *28*, 149–60.

(38) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(39) Hetényi, C.; van der Spoel, D. Toward prediction of functional protein pockets using blind docking and pocket search algorithms. *Protein Sci.* **2011**, *20*, 880–893.

(40) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, 455–461.

(41) Alhossary, A.; Handoko, S. D.; Mu, Y.; Kwoh, C.-K. Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics* **2015**, *31*, 2214–2216.

(42) Bell, E. W.; Zhang, Y. DockRMSD: an open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism. *J. Cheminform.* **2019**, *11*, 40.

(43) Xu, W.; Lucke, A. J.; Fairlie, D. P. Comparing sixteen scoring functions for predicting biological activities of ligands for protein targets. *J. Mol. Graph. Model.* **2015**, *57*, 76–88.

(44) Jeszenöi, N.; Horváth, I.; Bálint, M.; van der Spoel, D.; Hetényi, C. Mobility-based prediction of hydration structures of protein surfaces. *Bioinformatics* **2015**, *31*, 1959–1965.

(45) Jeszenöi, N.; Bálint, M.; Horváth, I.; van der Spoel, D.; Hetényi, C. Exploration of Interfacial Hydration Networks of Target Ligand Complexes. *J. Chem. Inf. Model.* **2016**, *56*, 148–158.

(46) Bálint, M.; Jeszenöi, N.; Horváth, I.; van der Spoel, D.; Hetényi, C. Systematic exploration of multiple drug binding sites. *J. Cheminf.* **2017**, *9*, 65.

(47) Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964.

(48) Ain, Q. U.; Aleksandrova, A.; Roessler, F. D.; Ballester, P. J. Machine–learning scoring functions to improve structure–based binding affinity prediction and virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 405–424.

(49) Li, H.; Leung, K. S.; Wong, M. H.; Ballester, P. J. Improving AutoDock Vina using random forest: the growing accuracy of binding affinity prediction by the effective exploitation of larger data sets. *Mol. Inform.* **2015**, *34*, 115–126.

(50) Li, H.; Leung, K. S.; Wong, M. H.; Ballester, P. J. Correcting the impact of docking pose generation error on binding affinity prediction. *BMC Bioinformatics* **2016**, *17*, 308.

(51) Wójcik, M.; Ballester, P. J.; Siedlecki, P. Performance of machine-learning scoring functions in structure-based virtual screening. *Scientific Reports* **2017**, *7*, 46710.

(52) Wang, W.; He, W.; Zhou, X.; Chen, X. Optimization of molecular docking scores with support vector rank regression. *Proteins* **2013**, *81*, 1386–1398.

(53) Gabel, J.; Desaphy, J.; Rognan, D. Beware of machine learning-based scoring functions – on the danger of developing black boxes. *J. Chem. Inf. Model* **2014**, *54*, 2807–2815.

(54) Ballester, P. J.; Mitchell, J. B. O. A machine learning approach to predicting protein?ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010**, *26*, 1169–1175.

(55) Li, H.; Leung, K. S.; Wong, M. H.; Ballester, P. J. *The use of random forest to predict binding affinity in docking. In: Ortuño F., Rojas I. (eds) Bioinformatics and Biomedical Engineering. IWBBIO 2015*; Springer International Publishing, Cham., 2015.

(56) Ramirez, D.; Caballero, J. Is it Reliable to take the molecular docking top scoring position as the best solution without considering available structural data? *Molecules* **2018**, *23*, 1038.

(57) Liu, Y.; Li, L.; Han, J.; Li, J.; Liu, Z.; Zhao, W.; Nie, Y.; Liu,; Wang, R. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* **2015**, *31*, 405–412.

(58) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977–2980.

(59) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of eseful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594.

(60) Hongjian, L.; Leung, K.-S.; Ballester, P. J.; Wong, M.-H. istar: a web platform for large-scale protein-ligand docking. *PLOS ONE* **2014**, *9*, e85678.

(61) Li, H.; Peng, J.; Sidorov, P.; Leung, Y.; Leung, K.-S.; Wong, M.-H.; Lu, G.; Ballester, P. J. Classical scoring functions for docking are unable to exploit large volumes of structural and interaction data. *Bioinformatics* **2019**, *35*, 3989–3995.

(62) Rabinowitz, J. R.; Little, S. B.; Laws, S. C.; Goldsmith, M.-R. Molecular modeling for screening environmental chemicals for estrogenicity: use of the toxicant-target approach. *Chem. Res. Toxicol.* **2009**, *22*, 1594–1602.

(63) Genheden, S.; Mikulskis, P.; Hu, L.; Kongsted, J.; Söderhjelm, P.; Ryde, U. Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J. Am. Chem. Soc.* **2011**, *133*, 13081–13092.

(64) Zhang, H.; Tan, T.; van der Spoel, D. Generalized Born and explicit solvent models for free energy calculations in organic solvents: cyclodextrin dimerization. *J. Chem. Theory Comput.* **2015**, *11*, 5103–5113.

(65) Zhang, H.; Yin, C.; Yan, H.; van der Spoel, D. Evaluation of generalized Born models for large scale affinity prediction of cyclodextrin host–guest complexes. *J Chem Inf Model.* **2016**, *56*, 2080–2092.

(66) Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D. Comparison of implicit and explicit solvent models for the calculation of solvation free energy in organic solvents. *J. Chem. Theory Comput.* **2017**, *13*, 1034–1043.

(67) Cournia, Z.; Allen, B.; Sherman, W. Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.

(68) Manzetti, S. Bonding of butylparaben, bis (2-ethylhexyl)-phthalate, and perfluorooctanesulfonic acid to DNA: Comparison with benzo [a] pyrene shows low probability for strong noncovalent DNA intercalation. *Chem. Res. Toxicol.* **2018**, *31*, 22–36.

(69) Manzetti, S.; Lu, T. Addendum: solvation energies of butylparaben, benzo [a] pyrene diol epoxide, perfluorooctanesulfonic acid, and DEHP in complex with DNA bases. *Chem. Res. Toxicol.* **2018**, *31*, 639–640.

(70) Rabinowitz, J. R.; Goldsmith, M.-R.; Little, S. B.; Pasquinelli, M. A. Computational molecular modeling for evaluating the toxicity of environmental chemicals: prioritizing bioassay requirements. *Environ. Health Perspect.* **2008**, *116*, 573–577.

(71) Johnson, M.; Maggiora, G. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons, New York, 1990.

(72) Dean, P. *Molecular similarity in drug design*; Dean, PM, 1995.

(73) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular similarity in medicinal chemistry. *J. Med. Chem.* **2014**, *57*, 3186–3204.

(74) Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, L. J.; Bork, P. Drug target identification using side-effect similarity. *Science* **2008**, *321*, 263–266.

(75) Wu, S.; Fisher, J.; Naciff, J.; Laufersweiler, M.; Lester, C.; Daston, G.; Blackburn, K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem. Res. Toxicol.* **2013**, *49*, 1840–1861.

(76) Zhu, H.; Bouhifd, M.; Donley, E.; Egnash, L.; Kleinstreuer, N.; Kroese, E.; Liu, Z.; Luechtefeld, T.; Palmer, J.; Pamies, D.; Shen, J.; Strauss, V.; Wu, S.; Hartung, T. Supporting read-across using biological data. *ALTEX - Alternatives to animal experimentation* **2013**, *33*, 167–182.

(77) Russo, D. P.; Strickland, J.; Karmaus, A. L.; Wang, W.; Shende, S.; Hartung, T.; Aleksunes, L. M.; Zhu, H. Nonanimal models for acute toxicity evaluations: applying data-driven profiling and read-across. *Environ. Health Perspect.* **2019**, *127*, 47001.

(78) Hu, Y.; Stumpfe, D.; Bajorath, J. Advancing the activity cliff concept. *F1000Research* **2013**, *2*, 199.

(79) Stumpfe, D.; Hu, H.; Bajorath, J. Evolving concept of activity cliffs. *ACS Omega* **2019**, *4*, 14360–14368.

(80) Curtis, L. R.; Mehendale, H. M. Hepatobiliary dysfunction and inhibition of adenosine triphosphatase activity of bile canaliculi-enriched fractions following in vivo mirex, photomirex, and chlordecone exposures. *Toxicol. Appl. Pharmacol.* **1981**, *61*, 429–440.

(81) Yang, F.; Huang, X.; Yi, T.; Yen, Y.; Moore, D. D.; Huang, W. Spontaneous development of liver tumors in the absence of the bile acid receptor farnesoid X receptor. *Cancer Res.* **2007**, *67*, 863–867.

(82) Ananthanarayanan, M.; Balasubramanian, N.; Makishima, M.; Mangelsdorf, D. J.; Suchy, F. J. Human bile salt export pump promoter is transactivated by the farnesoid X receptor/bile acid receptor. *J. Biol. Chem.* **2001**, *276*, 28857–28865.

(83) Tassaneeyakul, W.; Birkett, D. J.; Veronese, M. E.; McManus, M. E.; Tukey, R. H.; Quattrochi, L. C.; Gelboin, H. V.; Miners, J. O. Specificity of substrate and inhibitor probes for human cytochromes P450 1A1 and 1A2. *J. Pharmacol. Exp. Ther.* **1993**, *265*, 401–407.

(84) Jordan, D. B.; Livingston, R. S.; Bisaha, J. J.; Duncan, K. E.; Pember, S. O.; Picollelli, M. A.; Schwartz, R. S.; Sternberg, J. A.; Tang, X.-S. Mode of action of famoxadone. *Pesticide Sci.* **1999**, *55*, 105–118.

(85) Gao, X.; Wen, X.; Yu, C.; Esser, L.; Tsao, S.; Quinn, B.; Zhang, L.; Yu, L.; Xia, D. The crystal structure of mitochondrial cytochrome bc1 in complex with famoxadone: the role of aromatic- aromatic interaction in inhibition. *Biochemistry.* **2002**, *41*, 11692–11702.

(86) Smith, M. T. Mechanisms of Troglitazone hepatotoxicity. *Chem. Res. Toxicol.* **2003**, *16*, 679–687.

(87) Keiser, M. J. et al. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181.

(88) Reddy, C. K.; Das, A.; Jayaram, B. Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 41–47.

(89) Liu, J.; Sridhar, J.; Foroozesh, M. Cytochrome P450 family 1 inhibitors and structure-activity relationships. *Molecules* **2013**, *12*, 14470–14495.

(90) Shimada, T.; Guengerich, F. P. Inhibition of human cytochrome P450 1A1-, 1A2-, and 1B1-mediated activation of procarcinogens to genotoxic metabolites by polycyclic aromatic hydrocarbons. *Chem. Res. Toxicol.* **2006**, *19*, 288–294.

(91) Shimada, T. Inhibition of carcinogen-activating cytochrome P450 enzymes by xenobiotic chemicals in relation to antimutagenicity and anticarcinogenicity. *Toxicol. Res.* **2017**, *33*, 79–96.

(92) Manzetti, S. Polycyclic aromatic hydrocarbons in the environment: environmental fate and transformation. *Polycycl. Arom. Comp.* **2013**, *33*, 311–330.

(93) Manzetti, S. *Chemical and electronic properties of polycyclic aromatic hydrocarbons: a review*; Handbook of polycyclic aromatic hydrocarbons: chemistry, occurrence and health issues; 2012; Vol. 423; pp 423–435.

(94) Mise, M.; Hashizume, T.; Komuro, S. Characterization of substrate specificity of dog CYP1A2 using CYP1A2-deficient and wild-type dog liver microsomes. *Drug Metabol. Dispos.* **2008**, *36*, 1903–1908.

(95) Palmqvist, A.; Selck, H.; Rasmussen, L. J.; Forbes, V. E. Biotransformation and genotoxicity of fluoranthene in the deposit-feeding polychaete Capitella sp. I. *Environ. Toxicol. Chem. Internat. J.* **2003**, *22*, 2977–2985.

(96) Chiarolini, A.; Teresa Donato, M.; Jose Gomez Lechon, M.; Pala, M.; Valerio, F.; Ferro, M. Comparison of rat hepatocyte and differentiated hepatoma cell line cultures as bio indicators of CYP 1A1 inducers in urban air. *Biomarkers* **1997**, *2*, 279–285.

(97) Fonovich, T. Sudan dyes: are they dangerous for human health? *Drug and Chemical Toxicology* **2013**, *36*, 343–352.

(98) Stiborová, M.; Martínek, V.; Rýdlová, H.; Hodek, P.; Frei, E. Sudan I is a potential carcinogen for humans. *Cancer Research* **2002**, *62*, 5678.

(99) McHutchison, J. et al. Farglitazar lacks antifibrotic activity in patients with chronic hepatitis C infection. *Gastroenterology* **2010**, *138*, 1365–1373.

(100) Chapin, R. E.; Adams, J.; Boekelheide, K.; Gray, L. E.; Hayward, S. W.; Lees, P. S.; McIntyre, B. S.; Portier, K. M.; Schnorr, T. M.; Selevan, S. G.; Vandenbergh, J. G.; Woskie, S. R. NTP-CERHR expert panel report on the reproductive and developmental toxicity of bisphenol A. *Birth Defects Res. B Dev. Reprod. Toxicol.* **2008**, *83*, 157–395.

(101) Gaido, K. W.; Leonard, L. S.; Maness, S. C.; Hall, J. M.; McDonnell, D. P.; Saville, B.; Safe, S. Differential interaction of the methoxychlor metabolite 2,2-bis-(p-hydroxyphenyl)-1,1,1-trichloroethane with estrogen receptors alpha and beta. *Endocrinology* **1999**, *104*, 5746–5753.

(102) DeWitt, J. C.; Shnyra, A.; Badr, M. Z.; Loveless, S. .; Hoban, D.; Frame, S. R.; Cunard, R.; Anderson, S. E.; Meade, B. J.; Peden-Adams, M. M.; Luebke, R. W.; Luster, M. I. Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha. *Crit. Rev. Toxicol.* **2009**, *39*, 76–94.

(103) Lau, C.; Butenhoff, J. L.; Rogers, J. M. The developmental toxicity of perfluoroalkyl acids and their derivatives. *Toxicol. Appl. Pharmacol.* **2004**, *198*, 231–241.

(104) Takacs, M. L.; Abbott, B. D. Activation of mouse and human peroxisome proliferator-activated receptors (α, β/δ, γ) by perfluorooctanoic acid and perfluorooctane sulfonate. *Toxicological Sciences* **2007**, *95*, 108–117.

(105) Butler, E. G.; Tanaka, T.; Ichida, T.; Maruyama, H.; Leber, A. P.; Williams, G. M. Induction of hepatic peroxisome proliferation in mice by lactofen, a diphenyl ether herbicide. *Toxicol. Appl. Pharmacol.* **1988**, *93*, 72–80.

(106) Planting, A.; van der Gaast, A.; Schoffski, P.; Bartkowski, M.; Verheij, C.; Noe, D.; Ferrante, K.; Verweij, J. A phase I and pharmacologic study of the matrix metalloproteinase inhibitor CP- 471,358 in patients with advanced solid tumors. *Cancer Chemother. Pharmacol.* **2005**, *55*, 136–142.

(107) Dalvie, D.; Cosker, T.; Boyden, T.; Zhou, S.; Schroeder, C.; Potchoiba, M. J. Metabolism distribution and excretion of a matrix metalloproteinase-13 inhibitor, 4-[4-(4-fluorophenoxy)- benzenesulfonylamino]tetrahydropyran-4-carboxylic acid hydroxyamide (CP-544439), in rats and dogs: assessment of the metabolic profile of CP-544439 in plasma and urine of humans. *Drug Metab. Dispos.* **2008**, *36*, 1869–1883.

(108) Wilson, V. S.; Blystone, C. R.; Hotchkiss, A. K.; Rider, C. V.; Gray, L. E. Diverse mechanisms of anti-androgen action: impact on male rat reproductive tract development. *Int. J. Androl.* **2008**, *31*, 178–187.

(109) Gao, W.; Bohl, C. E.; Dalton, J. T. Chemistry and structural biology of androgen receptor. *Chem. Rev.* **2008**, *105*, 3352–3370.

(110) Singam, E. R. A.; Tachachartvanich, P.; La Merrill, M. A.; Smith, M. T.; Durkin, K. A. Structural dynamics of agonist and antagonist binding to the androgen receptor. *J. Phys. Chem. B* **2019**, *123*, 7657–7666.

(111) Asano, M. et al. Synthesis and biological evaluation of novel selective androgen receptor modulators (SARMs). Part II: Optimization of 4-(pyrrolidin-1-yl)benzonitrile derivatives. *Bioorganic & Medicinal Chemistry Letters* **2017**, *27*, 1897–1901.

(112) Watson, N.; Reddy, H.; Stefanich, E.; Eglen, R. M. Characterization of the interaction of zamifenacin at muscarinic receptors in vitro. *Eur. J. Pharmacol.* **1995**, *285*, 135–142.

(113) Luo, G. et al. CYP3A4 induction by drugs: correlation between a pregnane X receptor reporter gene assay and CYP3A4 expression in human hepatocytes. *Drug Metab. Dispos.* **2002**, *30*, 795–804.

(114) Needleman, P.; Jakschik, B. A.; Johnson, J., E. M. Sulfhydryl requirement for relaxation of vascular smooth muscle. *J. Pharmacol. exp. Ther.* **1973**, *187*, 324–331.

Graphical TOC Entry